DM

A Computational Software Tool for Satellite Laser Altimetry Data Processing, Analysis and Visualisation

MASTER DISSERTATION

Bruno Miguel Pereira da Silva
MASTER IN INFORMATICS ENGINEERING

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

April | 2021

A Computational Software Tool for Satellite Laser Altimetry Data Processing, Analysis and Visualisation

MASTER DISSERTATION

Bruno Miguel Pereira da Silva
MASTER IN INFORMATICS ENGINEERING

SUPERVISOR
Luiz Carlos Guerreiro Lopes

CO-SUPERVISOR
Pedro Filipe Pereira Campos

A Computational Software Tool for Satellite Laser Altimetry Data Processing, Analysis and Visualisation

Bruno Miguel Pereira da Silva
Licentiate (5-year degree) in Informatics Engineering

A thesis presented to the University of Madeira
in fulfillment of the requirements for the degree
of Master of Science in Informatics Engineering

Supervisor:
Prof. Dr. Luiz Carlos Guerreiro Lopes

Co-Supervisor:
Prof. Dr. Pedro Filipe Pereira Campos

Evaluation Committee:
Prof. Dr. Eduardo Miguel Dias Marques, FCEE/UMa (Committee Chair)
Prof. Dr. Filipe Magno de Gouveia Quintal, FCEE/UMa
Prof. Dr. Luiz Carlos Guerreiro Lopes, FCEE/UMa

Funchal, Portugal, April 2021

"Visualization gives you answers to questions you didn't know you had."

— Ben Schneiderman ([15], p. 67)

Acknowledgements

I would like to present my deepest appreciation to the great deal of support and assistance provided by my supervisor, Prof. Dr. Luiz Guerreiro Lopes to the work presented in this thesis. Your expertise, dedication, insightful feedback and motivation were invaluable for the research and methodology, engaging me in new ideas that ultimately brought my work to a higher level, while always giving me intellectual freedom.

I am also very thankful to my co-supervisor Prof. Dr. Pedro Filipe Campos for the relevant suggestions and comments that ultimately helped to enhance the final version of this work.

The valuable comments and suggestions of the members of the evaluation committee are also acknowledged, which have also helped to improve the final version of this thesis.

I would also to thank the Portuguese Foundation for Science and Technology (FCT) through LARSyS – FCT Pluriannual funding 2020–2023, for providing partial funding for the presentations and publications resulting from this work.

And, of course, my eternal gratitude to my parents who supported me and gave me the opportunity to obtain my diploma that finally culminated in my master's degree. All this could not be achieved without them.

Contents

Abstract

NASA's Ice, Cloud, and Land Elevation Satellite (ICESat), the first global laser altimetry satellite, was operated between 2003 and 2009 with the primary mission objective of measuring Earth's ice sheet mass balance, namely sea-ice thickness and ice sheet elevations. In addition of polar regions coverage, around the globe data about cloud propriety information, vegetation canopy structure and topographic data were also recorded. It has proven to be a very successful mission, operating beyond its initial 5-year goal and saw its data applied effectively in many scientific models outside its initial application scope.

The ICESat-2 satellite, the follow-up of the ICESat mission, with a more capable light detection and ranging (LIDAR) instrument, was launched in 2018 and represented an advancement over the laser technology of the first ICESat mission, firing laser pulses at 10 kHz rate, instead of the previous 40 Hz. This fast-firing laser technology allows the ICESat-2's LIDAR, called Advanced Topographic Laser Altimeter System (ATLAS), to take measurements approximately at every 0.7 meters along the satellite's track on Earth's surface instead the 170 meters of the previous ICESat mission. Also, each transmitted laser pulse is split in six individual beams, arranged in three pairs (each pair having a strong and a weak beam) and separated by 3 km apart, providing a multi-beam profiling of the surface.

The main objective of this thesis was the development of a software tool, called ICEComb, that allows scientists and researchers to visualise and process, in a suitable way, the available altimetry data from the ICESat mission and the ICESat-2 satellite, a follow-up of the previous mission, which was launched on the 15th of September 2018 and is currently operational.

ICEComb is a web-based software tool that offers its users the ability to access the available data from both missions, visualise them interactively on a geographic map, store the data products locally, explore data in a detailed, efficient and meaningful way, and provide satellite altimetry data processing by implementing different algorithms and statistical procedures, thus providing an easy-to-use environment for analysis, interpretation and processing of satellite laser altimetry data.

The need to create a new tool for processing and visualising the ICESat and ICESat-2 data products was derived primarily from the fact that existing solutions only provide access to a limited amount of information contained in the datasets of both missions and do not allow its processing and analysis in the same interface.

The developed tool was built using well-known and well-documented technologies in order to facilitate the incorporation of new functionalities and features to it and allow to extend its application to data obtained from other satellite altimetry missions.

Keywords: Satellite laser altimetry; LIDAR; ICESat/GLAS; ICESat-2/ ATLAS; Data visualisation; Software tool design.

Resumo

O primeiro satélite de altimetria a laser global da NASA, designado de ICESat (*Ice, Cloud, and land Elevation Satellite*), foi operado entre 2003 e 2009 com o objetivo principal da missão em medir as variações de massa do manto de gelo da Terra, ou seja, as mudança de elevação do manto de gelo e espessura do gelo marinho. Além da cobertura das regiões polares, foram recolhidos dados à volta do globo sobre propriedades das nuvens, estrutura da cobertura da vegetação e dados topográficos. Provou ser uma missão de muito sucesso, operando para além da sua meta inicial de cinco anos e viu os seus dados aplicados de forma eficaz em muitos modelos científicos fora do seu escopo inicial.

O satélite ICESat-2, lançado em 2018 no seguimento da missão ICESat, contém um instrumento LIDAR (LIght Detection And Ranging) mais capaz e representou um avanço da tecnologia laser da primeira missão ICESat, passando a disparar impulsos laser a uma taxa de 10 kHz em vez dos anteriores 40 Hz. Esta tecnologia laser de disparo rápido permite que o sistema ATLAS (*Advanced Topographic Laser Altimeter System*) faça medições aproximadamente a cada 0,7 metros ao longo da trilha do satélite sobre a superfície da Terra, em vez dos anteriores 170 metros da missão ICESat. Além disso, cada impulso laser transmitido é dividido em seis feixes individuais, dispostos em três pares (compostos por um feixe forte e um feixe fraco) e separados por 3 km de distância, fornecendo um perfil de feixe múltiplo da superfície.

O principal objetivo deste trabalho foi o desenvolvimento de uma ferramenta de software, denominada ICEComb, que permite aos cientistas e investigadores processar e visualizar, de forma adequada, os dados disponíveis de altimetria laser por satélite, nomeadamente dados da missão ICESat e do satélite ICESat-2, a missão que a sucedeu, lançado a 15 de setembro de 2018 e que se encontra ainda em operação.

ICEComb é uma ferramenta baseada na web que oferece aos utilizadores finais uma aplicação para análise e interpretação de dados de altimetria a laser por satélite com a capacidade de aceder aos dados disponíveis de ambas as missões, visualizá-los interativamente num mapa geográfico, armazenar os registros de dados localmente, explorar os dados de forma eficiente, detalhada e significativa e realizar o processamento de dados de altimetria de satélite através dos diferentes algoritmos e procedimentos estatísticos implementados, proporcionando assim um ambiente de software de fácil utilização.

A necessidade da criação de uma nova ferramenta para o processamento e a visualização dos produtos de dados dos satélites ICESat e ICESat-2 foi derivada principalmente do facto de que as soluções existentes apenas fornecem acesso a uma quantidade limitada de informação contida nos conjuntos de dados de ambas as missões, além de não permitirem o seu processamento e análise na mesma interface.

A ferramenta desenvolvida foi construída com tecnologias bem conhecidas e bem documentadas, de forma a facilitar a incorporação de novas funcionalidades e fontes de dados, permitindo assim estender a sua aplicabilidade a dados de outras missões de altimetria laser por satélite.

Palavras-chave: Altimetria laser por satélite; LIDAR; ICESat/GLAS; ICESat-2/ATLAS; Visualização de dados; Desenvolvimento de ferramentas de software.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

National Aeronautics and Space Administration (NASA)'s experimental scientific satellite ICESat was launched on the 13th of January 2003 as part of NASA's EOS (Earth Observing System) program. The satellite operated for seven years in a 600 km orbit with 94° inclination [28], two more years than the initial five-year mission goal, and was retired in February 2010 after completing 19 successful laser-operations campaigns [33], the last valid data having been gathered in October of 2009.

The Geoscience Laser Altimeter System (GLAS) the primary mission instrument [2, 3] onboard the ICESat satellite, was constituted by three different lasers that operated sequentially during the duration of the mission, one at a time—complemented by a dual-frequency Global Positioning System (GPS) receiver [27] and a customised star-tracker attitude determination system [29]—and was the first spaceborne laser-ranging (LiDAR) system designed for continuous, near-global observation of the Earth, covering the latitude range from 86°N to 86°S.

The GLAS lasers emitted short pulses of 4 ns with a repetition frequency of 40 Hz at both the wavelengths of 1064 nm (near-infrared light, used for surface altimetry and measurement of dense cloud heights) and 532 nm (visible green light, used for profiling atmospheric clouds and aerosols) [30], and operated intermittently, having been fired only at strategic time periods due to the operational and energy constraints resulting from the premature failure of the first of the lasers and the rapid energy decline of the second one [2].

Due the small footprint of the laser echo ($\sim$70 m), the GLAS altimeter provided high accuracy earth surface elevation data with few corrections to be applied.

Although the satellite's main objective was to observe ice elevation changes on both Greenland and Antarctica [28], the obtained data was widely used on far more subjects than the initial purpose. A variety of application fields, such as ice sheet elevation, land topography, cloud and aerosol height distribution, vegetation canopy characteristics and height, surface reflectivity [38], urban environment monitoring [13], and water level monitoring in rivers and natural and artificial lakes [5, 35] indicate how wide is the scope of applicability of the ICESat data.

The groundbreaking success of the first ICESat mission was proceeded by the launch, in September of 2018, of a follow-up mission called ICESat-2 [23], with an updated and more capable LIDAR instrument designated by ATLAS [19], which is currently operational. This instrument provides a measurement concept different from the LiDAR on ICESat, providing a faster repetition rate of 10 kHz, resulting in a separation of about 0.7 m between data points along the satellite's track [23], compared with about 170 m for the previous mission, in addition to using much lower energy pulses [36].

Besides that, each transmitted laser pulse is split into six individual beams, arranged in three pairs composed of a strong and a weak beam and separated by about 3 km cross-track [23], providing a multi-beam mapping of the Earth's surface. Thus, the measurements obtained have minimal gaps and provide higher fidelity of the topography obtained along the satellite's ground tracks [18].

The main science objective of ICESat-2 is to continue the ICESat mission and quantify changes in Earth's glaciers, ice sheets, sea ice and vegetation [1]. Having the first mission as an example, new scientific models will certainly emerge from the use of the ICESat-2 data. Studies performed in regions where in situ measurements are scarce or almost non-existent, such as remote areas of Africa and the Amazon, can benefit from data of this near-global satellite mission. The potential application of laser altimetry data from ICESat-2 for hydrological studies, namely to study the water level fluctuations in lakes, reservoirs and rivers in remote regions of different continents is a good example of its applicability.

Such data can also potentially be used to study air quality in regions affected by dust storms from the Sahara and the Sahel deserts, such as the part of the North Atlantic Ocean corresponding to the biogeographical region of Macaronesia, which includes the archipelagos of Madeira and Canary Islands.

When comparing satellite radar and laser altimetry data, it is observed that laser altimetry offers higher precision and accuracy data due to the its finer range bin resolution. Over Greenland's ice sheet, for example, the laser elevation measurements from the GLAS altimeter in the ICESat mission, had a precision from 0.14 to 0.59 m, in comparison with a precision from 0.28 to 2.06 m of the Envisat radar altimeter measurements [6].

Although there are some applications and Application Programmer Interfaces (APIs) to handle satellite data, there is a lack of tools that allows to access and visualise all the raw data collected by the ICESat and ICESat-2 satellite missions based on the data gathering location.

Furthermore, it is important to have the possibility of visually representing data on a map and to be able to process it directly in order to streamline the analysis and validation of scientific data.

A reference tool to visualise ICESat and ICESat-2 data based on the data collection location is the OpenAltimetry platform [14]. This NASA funded collaborative project, involving the Scripps Institution of Oceanography (SIO), San Diego Supercomputer Center (SDSC), National Snow and Ice Data Center (NSIDC) and the University Navstar Consortium (UNAVCO), was created to provide a tool that allows access to altimetry-specific data from ICESat and ICESat-2 missions through a web based interactive interface. However, by only focusing on altimetric laser data, this tool leaves out access to a lot of other still important information contained in the data products.

1.2 Objectives

The objective of this thesis was the development of a computational software tool to process and display data from the ICESat and ICESat-2 missions, in a meaningfully way, on top of a high detailed web mapping service and based on the geographic location from where data was collected.

The developed tool offers the possibility to browse the raw data contained in the datasets, represented using visual aids like graphs, tables and data conversions, in a geographic-related rich environment in order to facilitate data interpretation and discoverability. It also allows the selection and processing of altimetry data, based on well-documented methodologies, as well is able to perform statistical analysis and has data export capabilities.

The developed tool also fulfils other important requirements. It provides a friendly and rich user interface, offers expected performance levels while handling large datasets, provides a simple data access model, and allows a simple software deployment procedure.

The aim of the approach adopted was to simplify the knowledge correlation and the scientific models' validation and conception. After all, when dealing with satellite data, the data itself is only half of the equation and having the ability to quickly and precisely identify the location, and easily characterise its surroundings, is crucial for the interpretation of the obtained results.

1.3 Thesis structure

The first chapter begins with an explanation of the motivation of this study and its final goal. A characterisation of the principles of satellite altimetry is presented in Chapter 2, where the more common satellite radar altimeters are compared with the more recent laser altimeters, and the specifications of the GLAS and ATLAS laser altimeters are presented. This chapter finishes with a brief explanation of the principles of the spaceborne (laser) altimetry.

Chapter 3 contains the specific procedures used to select, identify and analyse information about the data obtained from the satellite laser altimetry, as well some of those challenges. The chapter ends with a short review of the state-of-the-art tools that allow to browse satellite laser altimetry data.

The ICEComb solution is presented in Chapter 4, where the tool is described extensively, from its architecture, the technology used, implementation, functionalities of the user interface in both the Client and Server applications. It also includes the interpretation and explanation of the results obtained with the tool developed.

Three papers emerged from Chapter 4. The first one is an abstract that was presented orally in the EGU (European Geosciences Union) General Assembly, and the second is an extended abstract accepted for oral presentation in the IEEE International Geoscience and Remote Sensing Symposium, indicated below:

1. Silva, B., Lopes, L. G., and Campos, P., ICEComb – A new software tool for satellite laser altimetry data processing and visualisation, *EGU General Assembly 2021* (vEGU21), online, 19–30 Apr. 2021, EGU21-13727, 2021. doi:10.5194/egusphere-egu21-13727.

2. Silva, B., Lopes, L. G., and Campos, P., A new web-based software tool for ICESat and ICESat-2 laser altimetry data processing and visualization, *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS 2021), Brussels, 12–14 July 2021, Article 4214, IEEE, 2021.

The third one is a manuscript in preparation, with a more detailed description, application, and assessment of the developed tool, to be submitted for possible publication in an open-access journal ranked Q1 in JCR 2020:

3. Silva, B., Lopes, L. G., and Campos, P., Description and assessment of a new tool for the processing, analysis and visualisation of satellite laser altimetry data: A case study in the world's largest coastal lagoon system (manuscript in preparation).

Finally, Chapter 5 includes the summary and reflection on the research, new knowledge contributions and recommendations for forthcoming work on this subject.

Chapter 2

Principles of satellite altimetry

2.1 Overview

In a simple phrase, satellite altimetry is a technique for observing the Earth surface from space by using precise measurements of a satellite's altitude and its distance to the surface. This space-to-earth measurement technique is used to accurately measure and understand the Earth's geometric shape, orientation, gravitational field (known as Geodesy science) and how these elements vary over time. These methods are also applicable to other planets of the solar system and natural satellites or moons.

Spaceborne radar altimeters are established tools for mapping the ice sheet mass balance, global sea level changes and terrain and ocean surface topography. The operational principal of a radar altimeter is to send a signal pulse to the surface and measure the return time. Delays or interference caused by the different layers of the atmosphere, climate variations and other factors are determined, and corrections are applied accordingly. By combining the altimetry data with the precise spacecraft location, it is possible to ascertain the surface heights within a few centimetres.

Satellite altimetry is an particular remote sensing approach because it can contribute with more than surface observations. By measuring the surface topography and its change over time, altimeters can provide information on the Earth's gravity field, study the ocean floor, analyse the ocean's heat and salt content, the land topography, cloud and aerosol height distribution, vegetation canopy characteristics and more.

2.2 Radar vs. laser

There are two main satellite altimetry methods, altimetry based on radar and the altimetry based on laser. The working principal is the same, both send a pulse of energy and use the round-trip time to determine the distance, while the first uses radar radiation and the second laser radiation.

By having a long history of being used on altimetry missions, radar-based altimeters are a more refined technology thus are more robust and offer lengthier operation, and longer up-time, when compared with laser-based altimeters. One factor attributed to the increased wear verified on laser altimeters is the use of high-powered lasers that generally are not very efficient, when compared to radar signal emitters, and require a lot of energy to operate.

When comparing the vertical accuracy, radar altimetry can offer accuracy values around 5 cm while laser altimeters offer accuracy of around 10 cm. In terms of work surfaces, laser altimeters have the advantage as measurements can be made over water, ice and land, while radar altimeters work best over water and ice. Weather factors will influence laser altimetry, such as clouds layers and blowing snow due the scattering effect on the laser pulses (Figure 2.1), while radar altimetry is virtually weather independent.

Radar-retrieved elevations inaccuracies are manly caused by errors in the radar retrieval signal due imprecise surface penetration models, slope-induced error (due its larger footprint) and limitations in retrieving a precise geolocated elevation. While with laser altimetry, the major errors are due inaccurate modelling of the saturation and scattering effects in the returning signal as well as inaccuracies in the pointing angle of the received laser signal. Also, the slope-induced error in laser altimetry is not as relevant because of the small footprint of the laser pulse, due its narrow divergence.

Calculations based in surface slope variation in ice sheets over Greenland and Antarctic found that the ICESat laser altimeter precision would vary from 14 to 59 cm while the radar altimeter precision in the ERS-2 satellite varied between 59 cm and 3.7 m [6]. The ICESat laser altimeter footprint was up to 70 m whereas the ERS-2 radar altimeter had an effective footprint between 1.6 and 10.8 km.

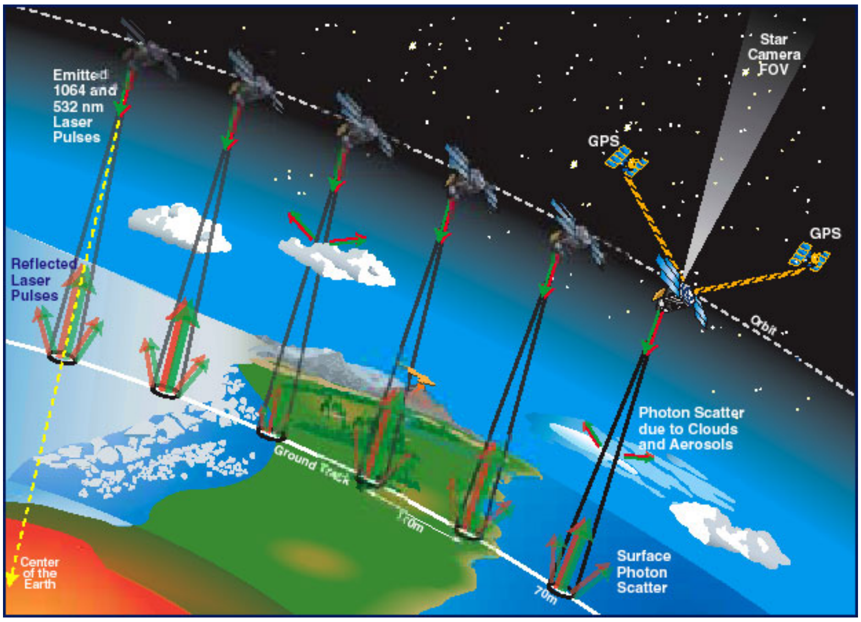


Figure 2.1: Schematic illustration of the ICESat laser altimeter, taking measurement while orbiting the Earth. Credits: NASA.

Table 2.1: Radar Altimetry Satellite Missions

Mission	Operational	Accuracy	Altitude (km)
Skylab	1973 for 9 months	100 m	435
GEOS-3	1975 to 1978	20 cm	840
SEASAT	1978 for 3 months	10 cm	762
GEOSAT	1986 to 1990	5 cm	800
TOPEX/Poseidon	1992 to 2006	4.2 cm	1 337
ERS-1	1991 to 2000	5 cm	780
ERS-2	1995 to 2003	5 cm	780
GFO-1	1998 to 2008	5 cm	800
Jason-1	2001 to 2013	3.3 cm	1 336
ENVISAT	2002 to 2012	3 cm	782
Jason-2	2008 to 2019	3.3 cm	1 336
CryoSat-2	2010 to present	3 cm	700
SARAL	2013 to present	8 mm	800
Sentinel-3A	2016 to present	3 cm	814
Jason-3	2016 to present	3.3 cm	1 340
Sentinel-3B	2018 to present	3 cm	814
Jason-CS (Sentinel-6)	2020 to present	3 cm	1 336

Table 2.2: Laser Altimetry Satellite Missions

Mission	Operational	Accuracy	Altitude (km)
MGS (Mars)	1996 to 2006	2–30 m	378
ICESat (Earth)	2003 to 2011	10 cm	600
SELENE (Moon)	2007 to 2009	5 m	100
ICESat-2 (Earth)	2018 to present	<10 cm	481
Gaofen 7 (Earth)	2019 to present	14 cm	500

2.3 GLAS altimeter

The GLAS is the main instrument aboard the ICESat satellite, it measures continuously along ground tracks at a rate of 40 Hz (40 times a second). The surface spot produced by the laser has approximately 70 m in diameter that are separated by nearly 170 m along track (see Figure 2.1).

GLAS uses lasers with two wavelengths, on with 1064 nm (near-infrared light) and another with 532 nm (green light). The first wavelength is more targeted for surface applications since it offers a better cloud penetration, while the second wavelength is mainly used for atmosphere applications. The 1064 nm wavelengths laser uses an analog detection scheme, based on waveforms [2], while the 532 nm uses photon counting.

2.4 ATLAS altimeter

The ATLAS is the altimetry instrument aboard the ICESAT-2. This is the successor of the GLAS altimeter and is the most sophisticated Earth-observing laser instrument deployed in orbit by NASA. The along track measuring rate is now 10 kHz (10 000 pulses per second, instead of the 40 pulses per second in the previous mission), resulting in a ∼70 cm separation of each measurement along the satellite's path [1]. The laser footprint is roughly 17 m.

The measurement approach used in the ATLAS altimeter is essentially different from that used in the ICESat mission. The instrument only transmits laser pulses at a wavelength of 532 nm and uses only the detection scheme based on photon-counting [23], instead of relying on full-waveform. In addition, each transmitted laser pulse is then spitted into six individual beams, arranged in three pairs of one "strong" and one "weak" beam (see Figure 2.2). The ratio between a "strong" beam and the corresponding "weak" beam is about 1:4.

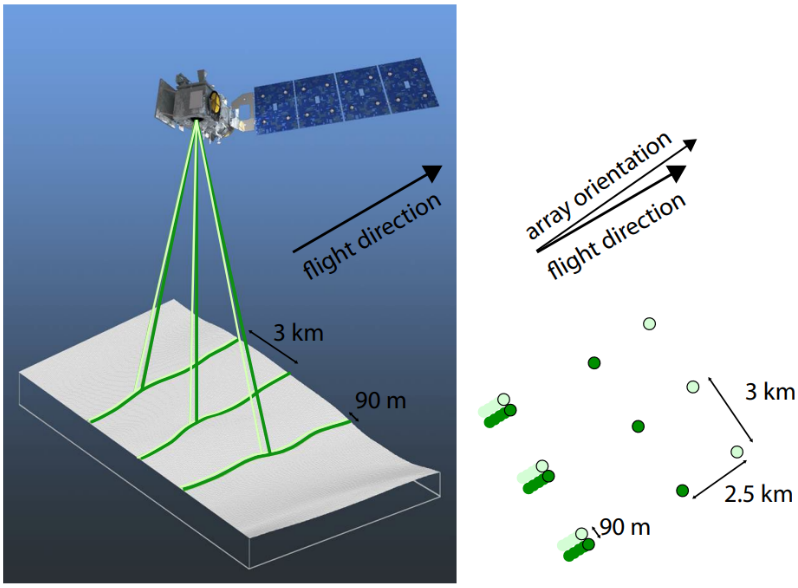


Figure 2.2: ICESat-2 laser beam surface pattern. The dark and light green colours represent respectively the "strong" and "weak" beam footprints.
Source: NASA.

The beam separation scheme is illustrated on the right side of Figure 2.2. Each pair of laser beams is around 3.3 km apart, across the track, while the "strong" and the "weak" beams are separated by about 90 m (in each pair of beams). In addition, the "strong" and "weak" beams have an along track offset of about 2.5 km.

When comparing with the previous mission, the ATLAS altimeter it is much more complex, with its high sampling rate and six-beam configuration. This configuration generates an amount of data plenty enough that the ATLAS altimeter is unable to assign unique time markers to the photon data. This information is added later by a on-board software system that uses range windows from 500 meters to 6 000 meters, depending on the type of surface and topography [16].

2.5 Height calculation

An advantage of satellite altimetry is the possibility of measuring remote or inaccessible locations, so the first major use of this measurement technique was to determine the topography of the ocean's surface globally. These new tool offered a broad range of applications in areas such as oceanography, geodesy and geophysics, and without the difficulty and complexity of performing in situ measurements.

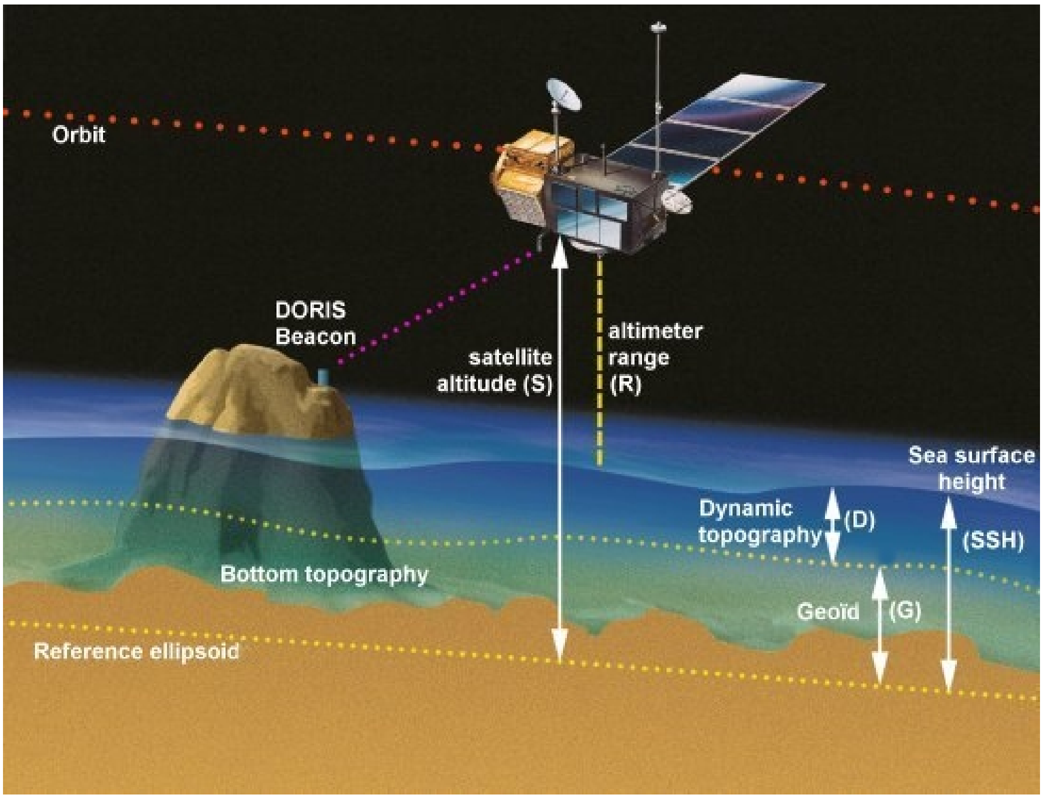


Figure 2.3: Satellite altimeter measurement principles.
Source: CNES – Centre National d'Études Spatiales.

When analysing Figure 2.3, the Sea Surface Height (SSH) can be determined by subtracting the range of the altimeter (R) from the satellite altitude (S):

$$SSH = S - R. \tag{2.1}$$

However, this surface height calculation is not sufficient for oceanographic applications due the superposition of geophysical effects, including the geoid undulation, tidal height variation and the ocean surface response to atmospheric pressure, which affect the height calculation [10]. To remove these external effects and determine the dynamic sea surface height ($dSSH$), represented by D in Figure 2.3, the geoid undulation (h_g), the tidal height variations (h_t) and the ocean surface response to atmospheric pressure loading (h_a) must be subtracted from the sea surface height (SSH):

$$dSSH = SSH - h_g - h_t - h_a. \tag{2.2}$$

Afterwards, the mean sea level (MSL), often abbreviated to sea level, can be determined by averaging the $dSSH$ along the satellite track:

$$MSL = \frac{\sum_{i=1}^{N} dSSH_i}{N}, \tag{2.3}$$

where N is the number of specimens along the track.

2.5.1 Geoid model

The geoid is a mathematical model of the Earth's mass distribution based on gravitational measurements and by ignoring currents and ocean tides. The resulting geoidal surface is irregular due to the uneven distribution of mass and rotation the Earth. An accurate model of the Earth's geoid is crucial to accurately determine sea level variations and map the topographic characteristics of the planet.

Although the geoid has been considered an important concept in the geodesy and geophysical sciences for almost 200 years, it was only in the late twentieth century, with the advances in satellite geodesy, that its model was defined with high precision.

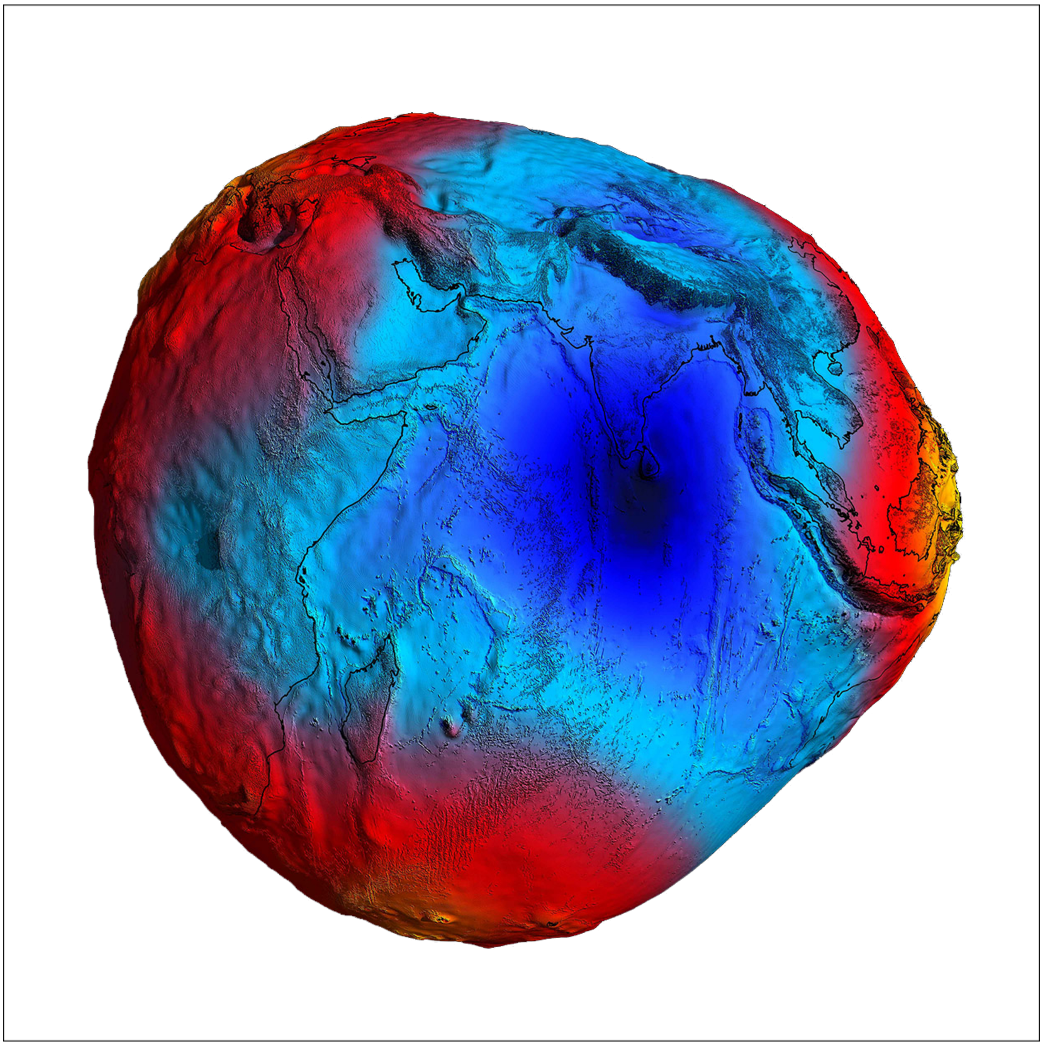


Figure 2.4: 2011 geoid model from ESA's Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission. The height deviations from an ideal geoid is represented on a scale of values from −100 m (low) to +100 m (high), denoted by shades of blue to yellow, passing through red.

Source: ESA – European Space Agency.

Chapter 3

Methodology

The main objective of the thesis was the development of a tool in order to process, analyse and visualise altimetry data from the ICESat and ICESat-2 missions. With that in mind, the first step was to understand the data generated by the satellites and how to interface with it, analyse which tools that currently exists and perform similar tasks, and finally the new solution's design and architecture.

3.1 Know your data

One of the first steps to face the challenges proposed by the thesis was the understanding of the data collected and processed in the ICESat and ICESat-2 missions. Questions such as how was the data organised, what type of data was collected and its parameters, in which file format are the data available, its internal structure and what tools and APIs do exist to access them, are answered here.

3.1.1 Data organisation

The data from both missions are organised into different datasets, or data products, that group the measured/processed altimetry data by a common subject, as shown in Table 3.1 and Table 3.2.

In the titles of both ICESat and ICESat-2 datasets, there is a reference to the level of data processing, that is indicated in the beginning of the title and represented by a letter "L" followed by a number/letter.

Table 3.1: GLAS/ICESat Datasets and Parameters

ID	Title	Parameter(s)
GLAH01	GLAS/ICESat L1A Global Altimetry Data	Waveform
GLAH02	GLAS/ICESat L1A Global Atmosphere Data	Sensor counts
GLAH03	GLAS/ICESat L1A Global Engineering Data	Total temperature
GLAH04	GLAS/ICESat L1A Global Laser Pointing Data	Attitude characteristics, viewing geometry
GLAH05	GLAS/ICESat L1B Global Waveform-based Range Corrections Data	Glacier elevation / ice sheet elevation, glacier topography / ice sheet topography, ice sheets, terrain elevation
GLAH06	GLAS/ICESat L1B Global Elevation Data	Glacier elevation / ice sheet elevation, glacier topography / ice sheet topography, ice roughness, ice sheets, reflectance, sea surface height, sea surface slope, surface roughness
GLAH07	GLAS/ICESat L1B Global Backscatter Data	Aerosol backscatter, scattering, terrain elevation
GLAH08	GLAS/ICESat L2 Global Planetary Boundary Layer and Elevated Aerosol Layer Heights	Aerosol particle properties, planetary boundary layer height
GLAH09	GLAS/ICESat L2 Global Cloud Heights for Multi-layer Clouds	Cloud height, cloud vertical distribution
GLAH10	GLAS/ICESat L2 Global Aerosol Vertical Structure Data	Aerosol backscatter, aerosol extinction, cloud reflectance, transmittance
GLAH11	GLAS/ICESat L2 Global Thin Cloud/Aerosol Optical Depths Data	Aerosol optical depth/thickness, cloud optical depth/thickness
GLAH12	GLAS/ICESat L2 Global Antarctic and Greenland Ice Sheet Altimetry Data	Glacier elevation / ice sheet elevation, glacier topography / ice sheet topography, ice sheets, reflectance
GLAH13	GLAS/ICESat L2 Sea Ice Altimetry Data	Ice roughness, reflectance, sea ice elevation
GLAH14	GLAS/ICESat L2 Global Land Surface Altimetry Data	Reflectance, terrain elevation
GLAH15	GLAS/ICESat L2 Ocean Altimetry Data	Reflectance, sea surface height, sea surface slope

Table 3.2: ATLAS/ICESat-2 Datasets and Parameters

ID	Title	Parameter(s)
ATL02	ATLAS/ICESat-2 L1B Converted Telemetry Data	Engineering telemetry ancillary data
ATL03	ATLAS/ICESat-2 L2A Global Geolocated Photon Data	Terrain elevation
ATL04	ATLAS/ICESat-2 L2A Normalized Relative Backscatter Profiles	Lidar backscatter
ATL06	ATLAS/ICESat-2 L3A Land Ice Height	Glacier elevation / ice sheet elevation
ATL07	ATLAS/ICESat-2 L3A Sea Ice Height	Sea ice elevation
ATL08	ATLAS/ICESat-2 L3A Land and Vegetation Height	Canopy height, terrain elevation
ATL09	ATLAS/ICESat-2 L3A Calibrated Backscatter Profiles and Atmospheric Layer Characteristics	Cloud properties, lidar backscatter
ATL10	ATLAS/ICESat-2 L3A Sea Ice Freeboar	Freeboard
ATL11	ATLAS/ICESat-2 L3B Annual Land Ice Height	Glacier elevation / ice sheet elevation
ATL12	ATLAS/ICESat-2 L3A Ocean Surface Height	Sea surface height
ATL13	ATLAS/ICESat-2 L3A Inland Water Surface Height	Elevation
ATL16	ATLAS/ICESat-2 L3B Weekly Gridded Atmosphere	Aerosol fraction, apparent surface reflectivity, blowing snow, cloud fraction, cloud optical depth / thickness
ATL17	ATLAS/ICESat-2 L3B Monthly Gridded Atmosphere	Aerosol fraction, apparent surface reflectivity, blowing snow, cloud fraction, cloud optical depth/thickness
ATL20	ATLAS/ICESat-2 L3B Daily and Monthly Gridded Sea Ice Freeboard	Freeboard

The lower-level products typically work with more fundamental data, obtained directly from the instrumentation, therefore indicating individual events, while the higher-level data products work with data aggregations from the several sources in order to determine quantities of geophysical interest. Figure 3.1 illustrates the ICESat-2 data products relations and their corresponding levels.

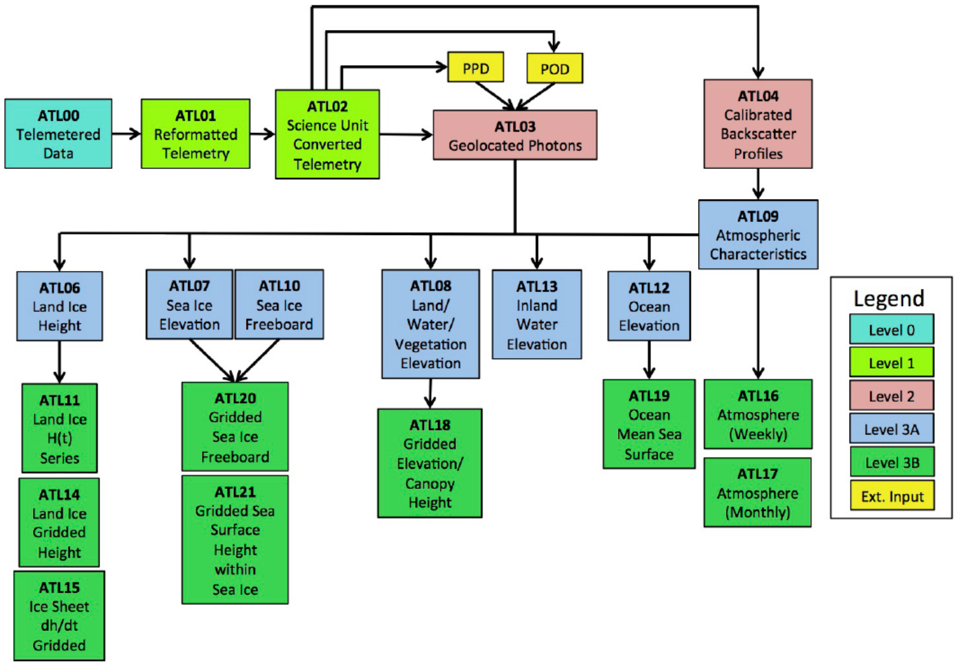


Figure 3.1: ICESat-2 data processing flow. The external inputs are the PPD (Precise Pointing Determination) and POD (Precise Orbit Determination). Source: NSIDC.

3.1.2 Data description

NASA's NSIDC website is a crucial resource for downloading the different datasets from both the ICESat satellites missions and access to detailed technical documentation such as user guides and the different data dictionaries. These documents describe in detail all data points contained in the files for each dataset, indicating parameters such as the label, datatype and dimension, long or standard names of the data point, units used, description, example value, source and the type of coordinates used.

The data dictionary information is also transcribed internally in each data set file, offering direct and immediate access to the description, as explained in the following section.

3.1.3 Data format (HDF5)

Another step in understanding the data was to learn about its support format. In the latest versions of the ICESat data products files, NASA decided to abandon the distribution in binary file format and instead adopted the use of HDF (Hierarchical Data Format) version 5 - or simply HDF5. This decision was due the fact that the various future missions, including the ICESat-2 and Soil Moisture Active Passive (SMAP) missions, would also use the HDF5 file format.

This file format is designed to store and organise large amounts of data, supporting many different data models including multidimensional arrays, raster images, tables, datasets (homogeneous type of multidimensional arrays) and groups (containers that can contain datasets and other groups). Files can contain a mix of related objects that can be accessed as a group or as an individual object.

HDF files are of self-describing nature allowing the interpretation of its structure and contents without resorting to external information. The HDF5 format is open and widely supported by a large community, and its storage mechanism is considered to be a scalable, versatile, and fast.

3.1.3.1 HDF5 data model and file structure

As the name implies, Hierarchical Data Format files are organised in a hierarchical structure that contains two primary items:

- Groups: a collection structures (folder like elements) that can contain other groups or datasets.
- Datasets: the data supporting elements. These dataset can contain one or many values (as an array) or one or many array of values (as a multidimensional array).

The self-describing nature of the HDF5 file is achieved by having a metadata repository associated to each file, group, and dataset, which contains a set of attributes that characterises the data itself. In terms of structure, these attributes do not necessarily represent a new data type since they are handled as a one value only dataset.

In Figure 3.2 the data path is visible above each element. The root element is represented by a forward slash (/) and as you browse each data element, its name is added to the data path, similar to what happens with file paths in a computer file system.

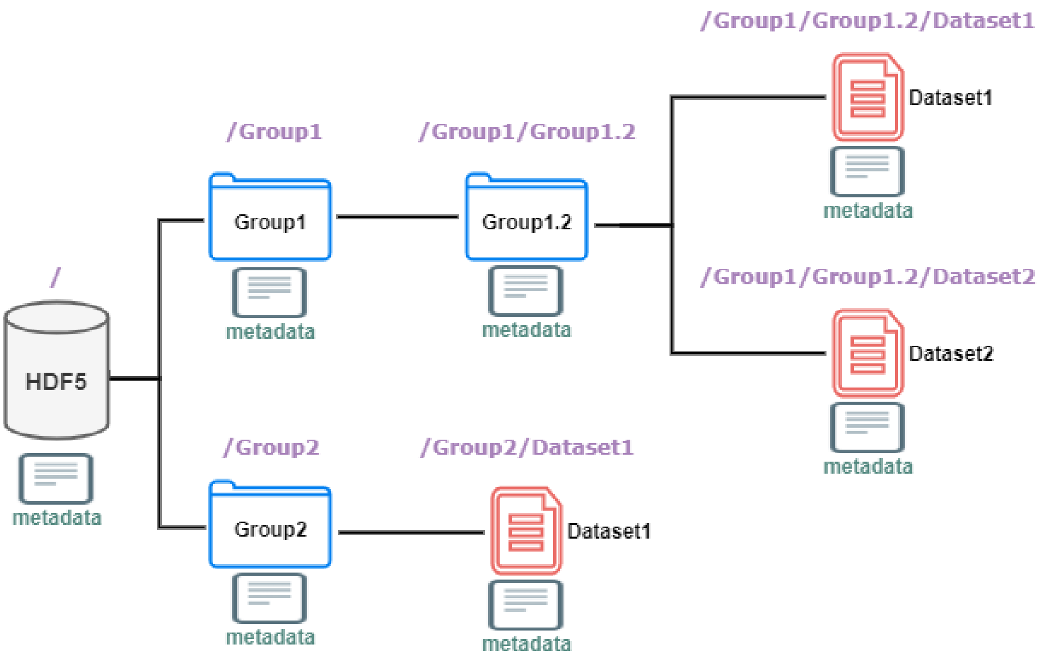


Figure 3.2: HDF5 data model and file structure.

The metadata attributes are accessed in a similar way, to do so, simply add the attribute name in front of the element. For example, to access "*attribute1*" at the root location, the data path for that attribute would be "*/attribute1*", in "*Group2*" it would be "*/Group2/attribute1*" and for the "*Dataset1*" inside the "*Group2*" the attribute data path would be "*/Group2/Dataset1/attribute1*".

3.1.3.2 HDF5 viewer

HDFView is the NSIDC recommended tool for viewing raw data from the ICESat and ICESat-2 data products. It allows the user to view data in a similar way as a spreadsheet and offers the possibility to generate plots and render images. The tool is open source and requires the Java SE Runtime instalation to run.

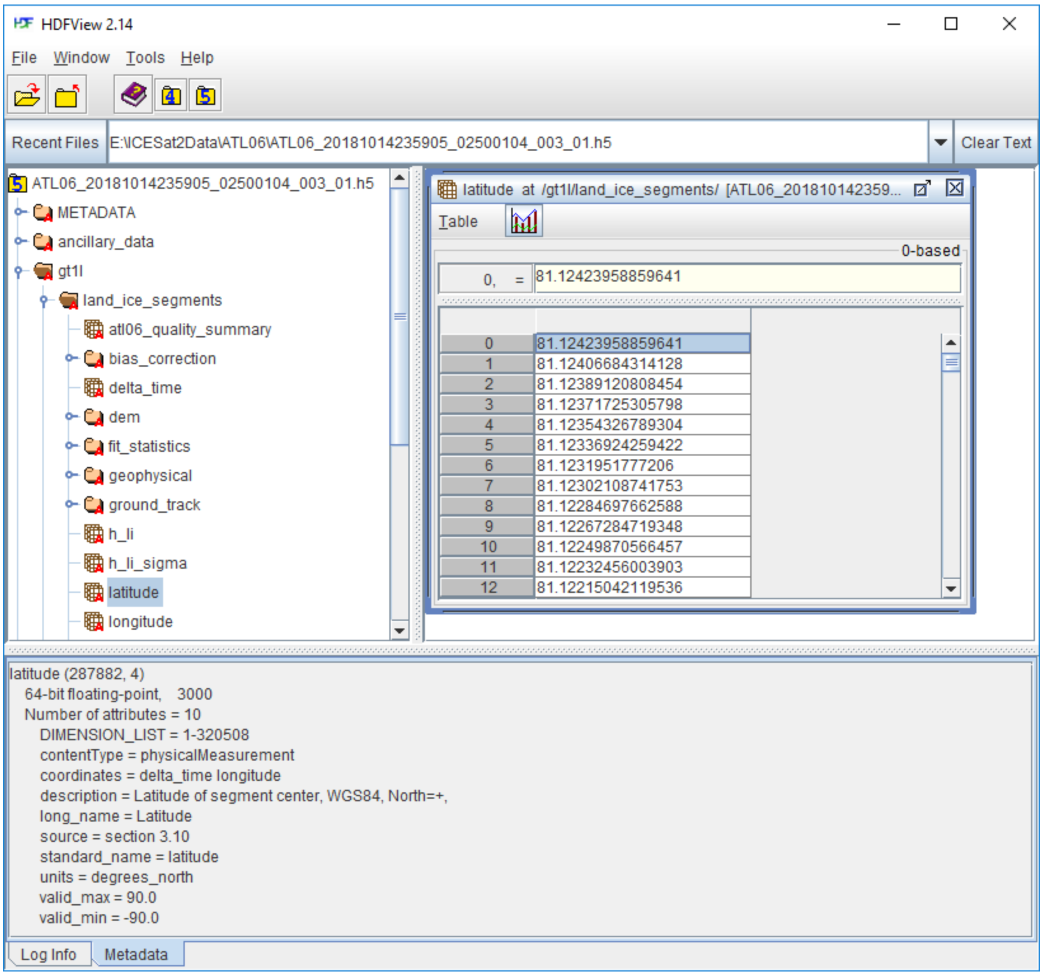


Figure 3.3: HDF5View GUI. The tree structure is visible on the left and is used to navigate through the data while the raw data is presented on the right side. In the bottom of the window the metadata for the selected dataset (latitude) is visible, where the attributes are separated by an equal sign from their values.

3.1.4 Data volume

A major obstacle to be overcome in the processing of satellite data is obviously the volume of data, especially for long missions. Although the ICESat-2 mission has only been in operation for about three years, on the date of this writing, the new ATLAS altimeter has already produced a larger volume of data than the seven years of operation of the previous ICESat mission. This is due to the fact that ATLAS works five times faster than the GLAS altimeter and with six beams, instead of one, as already been mentioned before.

The data volume information presented in Table 3.3 is considered as final since there are no more data revisions scheduled for the ICESat products while the ICESat-2 mission is still undergoing, the data in Table 3.4 will change every time there is a new release.

The information collected on the data products from the ICESat and ICESat-2 missions were carried out on 3rd of May of 2021.

Table 3.3: GLAS/ICESat data volume by dataset

DataSet	Latest version	Number of files	Total data
GLAH01	33	34 391	470 GB
GLAH02	33	4 306	1 126.4 GB
GLAH03	33	4 306	37.1 GB
GLAH04	33	4 298	227.9 GB
GLAH05	34	34 274	170 GB
GLAH06	34	34 208	84 GB
GLAH07	33	4 317	1 228.8 GB
GLAH08	33	252	2.1 GB
GLAH09	33	515	25 GB
GLAH10	33	262	5.5 GB
GLAH11	33	643	22 GB
GLAH12	34	637	15 GB
GLAH13	34	642	18 GB
GLAH14	34	642	57 GB
GLAH15	34	642	58 GB
Total		**124 335**	**3.46 TB**

Table 3.4: ATLAS/ICESat-2 data volume by dataset

DataSet	Latest Version	Number of files	Total data
ATL02	3	170 250	100 TB
ATL03	3	156 927	170 TB
ATL04	3	11 161	9.3 TB
ATL06	3	76 900	2.9 TB
ATL07	3	21 336	1.5 TB
ATL08	3	119 100	3.8 TB
ATL09	3	10 912	25 TB
ATL10	3	20 318	0.94 TB
ATL11	2	7 779	0.12 TB
ATL12	3	2 822	630 GB
ATL13	3	2 817	70 GB
ATL16	2	67	200 MB
ATL17	2	17	92 MB
ATL20	1	20	140 MB
Total		**600 426**	**314.24 TB**

3.2 State-of-the-art

While there are several tools for accessing the ICESat and ICESat-2 data products (notably the HDF5 files), these tools typically only provide raw data access, without any visual representation of where the data originated from.

However, when navigating data obtained from satellite altimetry, having an quick and easy way to characterise the measured geographical location is a fundamental action for the interpretation, selection and validation of the data.

Two tools that attempt to address these issues are discussed in this section.

3.2.1 ICESatProcessor

The ICESatProcessor is an application that uses MATLAB subroutines for extraction and visualisation of altimetry data for water bodies, obtained from the ICESat satellite mission. The water level results calculated by the application were then validated by comparison with levels acquired by remote sensing and data from gauged stations belonging to the Brazilian National Water Agency (Agência Nacional de Águas).

The water level obtained from ICESat data approached the levels measured by the Brazilian National Water Agency gauged stations, within the mean squared error of less than 30 cm [24].

To use the ICESatProcessor, user must input the following information:

1. The study period (the ICESat mission date interval);
2. The study site (as latitude and longitude limits—equivalent to the Southwest and Northeast bounds—or the region limits as a Geographic Information System polygon);
3. Optionally a SRTM DEM (Shuttle Radar Topography Mission Digital Elevation Model) criterium for the removal of outliers.

3.2.1.1 Setup and case study

Before presenting a case study for ICESatProcessor, there are some tasks that must be executed before running the application:

1. Download the data product GLA14 – GLAS/ICESat L2 Global Land Surface Altimetry Data and maintained the sub-folder structure presented on the NSIDC web site, meaning, the folder must contain sub-folders with the date of the data colection in the format (YYYY.MM.DD).
2. Install the MATLAB Compiler Runtime version 8.0.

Completing these tasks, the program can now be executed and the screen shown in Figure 3.4 will be displayed.

Now, a step-by-step setup of the case study at Guaíba River region, located in the state of Rio Grande do Sul, Brazil, provided by the author as part of the application package:

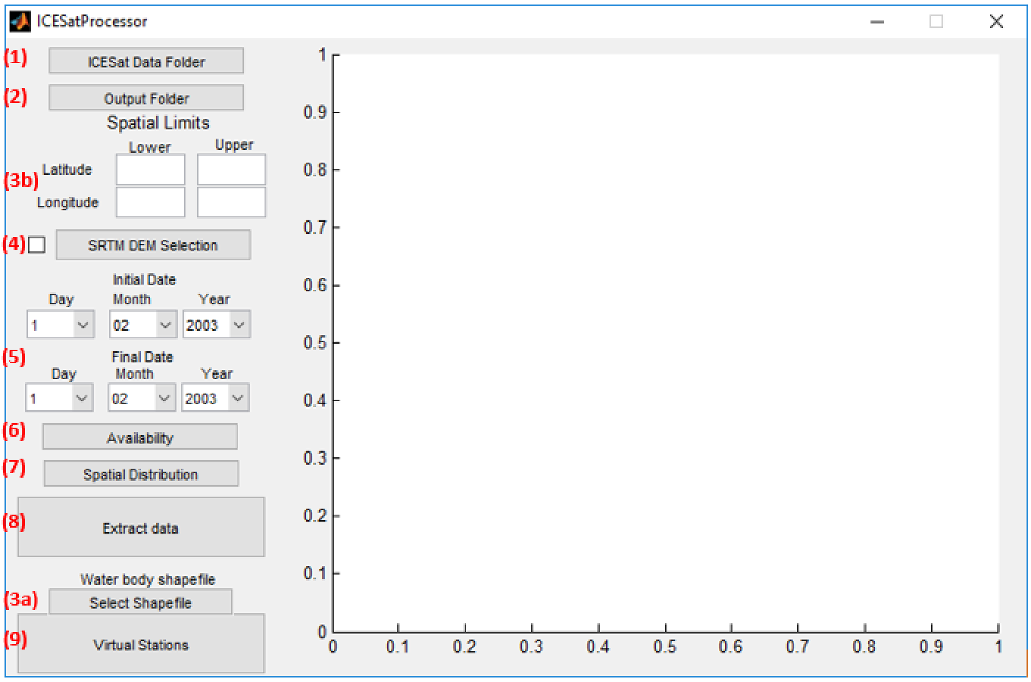


Figure 3.4: ICESatProcessor GUI.

1. Select the ICESat Data Folder by clicking the corresponding button and picking the location of the GLAH14 dataset.

2. Select the Output Folder for the result of the processing.

3. Select the "Water body shapefile" by clicking on Select Shapefile (3a) by select the provided "Guaiba.shp" file (sample polygon). By doing so, the shape is presented on the main preview window and the "Spatial Limits" are automatically set (3b) with the coordinates (bounds in this case) of the selected shape (corresponding of the study area of the Guaíba River).

4. Optionally the SRTM DEM Selection can be activated by selecting the checkbox and the provided "srtm_26_19.asc" file for outliers removal. More grids may be downloaded at the SRTM 90m DEM Digital Elevation Database web site (http://srtm.csi.cgiar.org/, 2021).

5. Select the ICESat data obtained period by indicating the "Initial Date" and "Final Date".

6. Check ICESat data availability for the selected period by clicking the Availability button, since there isn't data for every day of the satellite campaign. For this example, the author advises to choose the following interval: 08/11/2003 to 08/11/2004.

7. The satellite orbits over the selected region can be displayed by clicking the Spatial Distribution button. The data points are drawn on the main preview window over the shape and the file "Results.txt" is generated containing the "longitude", "latitude", "elevation", "year", "month" and "day" of the data points.

8. Now the data extracting may be executed by clicking the Extract Data button generating 3 text files: "READ-ME.txt"; "Results.txt" and "ICESat and Srtm.txt".

    (a) The "READ-ME.txt" is basically a log file indicating which data granule (or file) has data about the selected region, how many observations (data points) and how many points were removed based on the elevation flag and saturation flag.

    (b) "Results.txt" is the same as obtained by clicking the Spatial Distribution button.

    (c) Finally, the "ICESat and Srtm.txt" file contains the "longitude", "latitude", "elevation of ICESat", "SRTM elevation", "elevation used", "year", "month" and "day".

9. The Virtual Stations button action is similar as the Extract Data, it generates and additional 2 text files named "ResultsShapefile.txt" and "VirtualStations.txt".

    (a) The "ResultsShapefile.txt" is similar as the "Results.txt" file but only contains the observations inside the polygon and an additional column named "ID" stating the number of the polygon in which the virtual station was made.

    (b) The "VirtualStations.txt" contains the data points removed by the criteria set by the author.

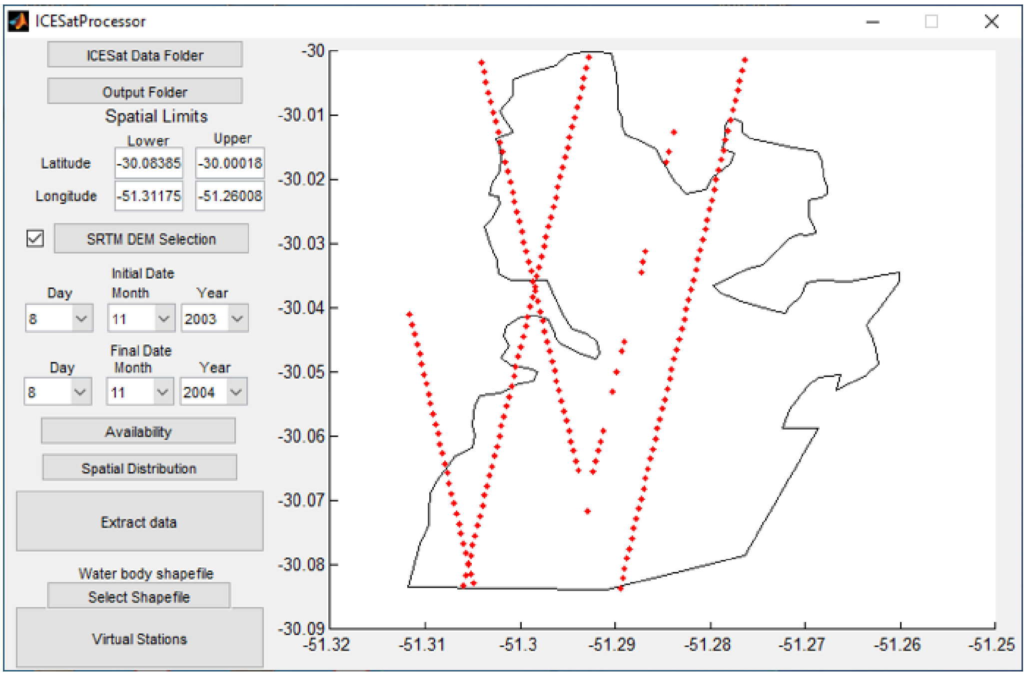


Figure 3.5: ICESatProcessor GUI after loading the Guaíba River case study.

3.2.1.2 Conclusion

This approach demonstrated that the use of NASA's ICESat satellite altimetry data for hydrological applications can be applied effectively and with a small deviation from the in situ measurements. This opens the possibility to apply the same method on other remote locations, with no local measurement stations, and determine the water elevation with some degree of certainty.

The developed application facilitates the ICESat data processing, but requires pre-calculation tasks, namely, manually determine the spatial limits of the study area, or the use of a third party tools to draw the shape of the study region, and pre-determine the date interval for the passage of the satellite over that region.

These are critical tasks that will determine the calculation error and success. Also, the lack of error checking, an incomplete step-up process and the need to perform more than ten steps to get a result makes this a non user-friendly tool and a somewhat frustrating experience.

Another problem is that the data output file is formatted as plain text and with no direct machine-readable format, instead of the standard comma-separated values (CSV) file format that can be directly opened in a spreadsheet.

Also, according to the tests carried out, the ICESatProcessor cannot read the whole GLAH14 dataset. The application does not return any error message and does not allow any calculation to be performed. To overcome this issue it is necessary to break the dataset in smaller chunks, by selecting data corresponding to a shorter period of time and removing the remaining data folders.

3.2.2 OpenAltimetry project

The OpenAltimetry project (https://openaltimetry.org/) is a web based interface that overlays altimetry-specific data from both the ICESat and ICESat-2 missions on top of a interactive geographic map. This collaborative project was funded by NASA and involves the Scripps Institution of Oceanography (SIO), San Diego Supercomputer Center (SDSC), NSIDC and the University Navstar Consortium (UNAVCO) [14].

The objective of this tool is to facilitate the discovery and visualisation of surface elevation data in an appealing way, without the need to rely on external tools and without handling multiple and scattered data products.

3.2.2.1 Interface and Architecture

OpenAltimetry uses the OpenLayer's geographical map where a mix of reference and actual satellite ground tracks are displayed, and offers access to the sampling segments for each sensor beam on which user can visualise track information, sample point data, view energy waveform, photon information, and other related data.

When viewing elevation data from the ICESat/GLAS datasets (Figure 3.6), the only data products used by the OpenAltimetry tool are the "GLAH06 – L1B Global Elevation Data" for the elevation data, and the "GLAH01 – L1A Global Altimetry" for the waveform data—although this last one is not mentioned on their website.

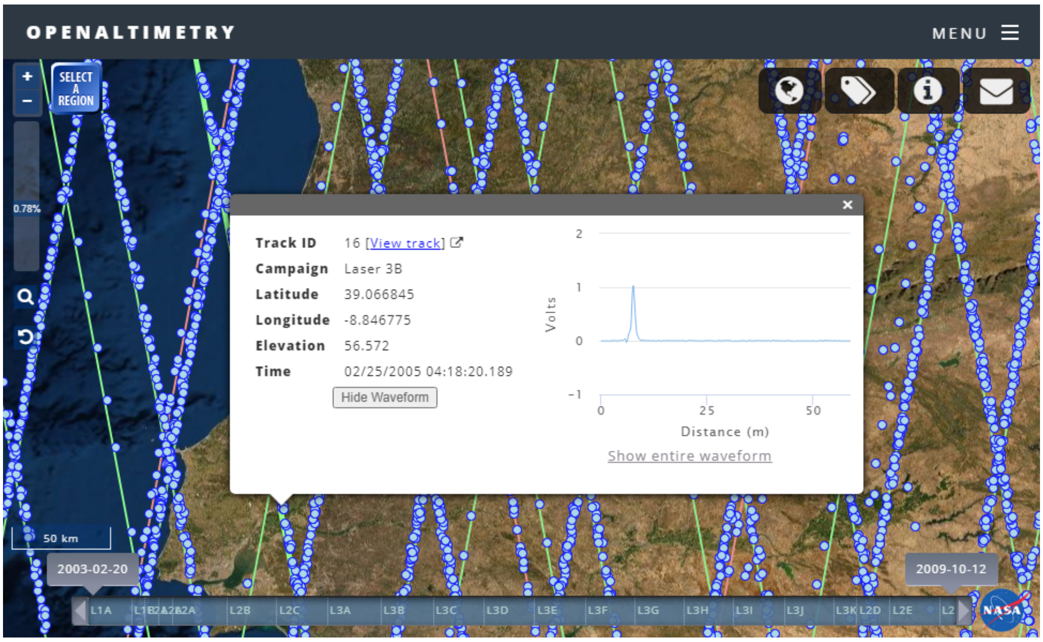


Figure 3.6: View of ICESat GLAS elevation data on OpenAltimetry.

ICESat-2 data availability is still expanding and new data is made available while the mission progresses. In the case of this mission, the data products access is more comprehensive, it includes data from the "ATL06 – L3A Land Ice Height", "ATL07 – L3A Sea Ice Height", "ATL08 – L3A Land and Vegetation Height", "ATL10 – L3A Sea Ice Freeboar", "ATL12 – L3A Ocean Surface Height" and "ATL13 – L3A Inland Water Surface Height".

In Figure 3.7 is possible to view the six data points corresponding the six laser beams of the ATLAS altimeter, each represented in a different colour. The elevation information of a coordinate and its respective photon data is also visible.

OpenAltimetry's web interface was build using OpenLayers mapping library and is modelled after NASA's Earth Observing System Data and Information System (EOSDIS) Worldview system.

The OpenAltimetry project was developed using a three-tier architecture, the presentation, the service and the data tier. The data tier consists on highly optimised PostgreSQL database with split-level storage and a disassociated object-based storage for HDF5 files so data of both the ICESat and ICESat-2 missions are stored on their original format [21].

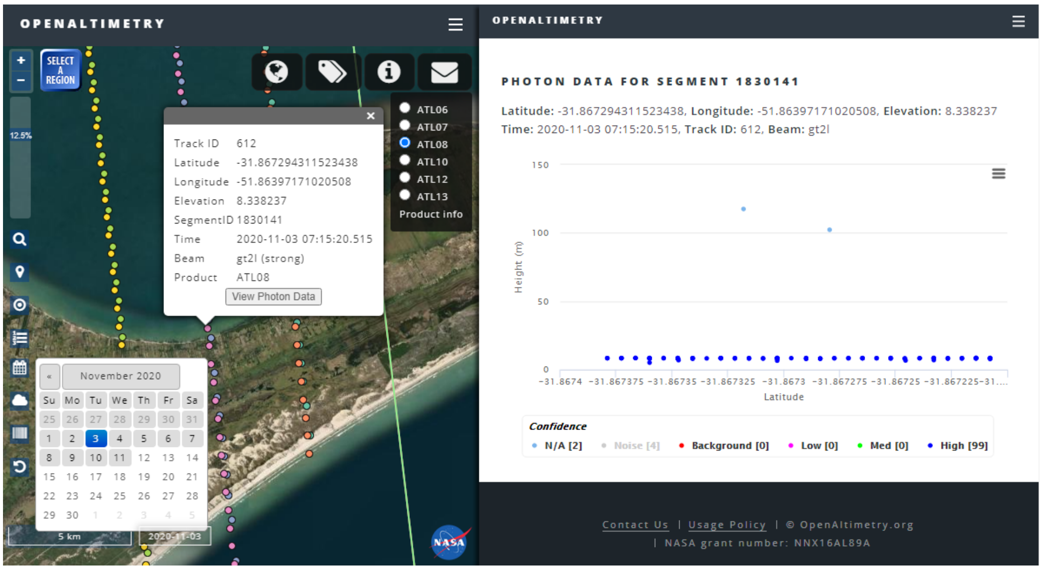


Figure 3.7: View of ICESat-2 ATLAS elevation and photon data on OpenAltimetry.

3.2.2.2 Conclusion

Without having familiarity with the data, users tend to find difficult to navigate the different datasets. The OpenAltimetry platform manages to closes the gap of promoting and facilitating data access to ICESat and ICESat-2 products in a straightforward and accessible way wile using a visual environment based on a geographical map to display the data.

However, by only focusing in some altimetry-specific data, this tool leaves out the access of many other, but still important, information existing in the data products.

Data from atmosphere segments, cloud layers, backscatter, aerosols, reflectivity characteristics, optical depth, and data from geophysical and planetary boundary layer, as well as some raw altimetry data, corrections and flags, are some examples of data that exists in the ICESat and ICESat-2 data products but that aren't accessible in the OpenAltimetry platform.

Chapter 4

Results and discussion

This chapter describes the solution developed to achieve the main objective of this work, a computational software tool that allows scientists to process, analyse and visualise, in a suitable way, the available satellite laser altimetry data, namely from the ICESat and the ICESat-2 satellite missions.

The created tool has two main parts, a server application to handle the data from the satellite missions in a decentralised manner and a user interface to present the data on top of a interactive map. This offers a visual representation of where the data was collected and the possibility to access the recorded data.

4.1 The ICEComb tool

In order to create a richer visual environment while accessing data from both the ICESat and ICESat-2 missions, a web interface based on the graphical component of the Google Maps API [9] was chosen. This mapping API was chosen since its a scalable and modern web platform that provides user access to updated satellite imagery, high-resolution aerial photography and high-detail low altitude images, in some areas down to street level, while offering high performance.

Likewise, the Google Maps API also offers the possibility to use different map view formats, such as Google Earth satellite images and a physical map based on terrain information, and access to a set of other tools in the form of Client Libraries with APIs such as Distance Matrix and Elevation.

The first allows to calculate the distance between geographic coordinates, while the second provides elevation data for the Earth's surface and depth for the ocean floor.

The software tool developed, called ICEComb (www.icecomb.org), has an architecture composed of two main components, a Data Server and a Web Engine that interfaces with the user through a Web Browser, as displayed in Figure 4.1. This Client/Server approach [26] permits to share the processing demands and grants better performance overall.

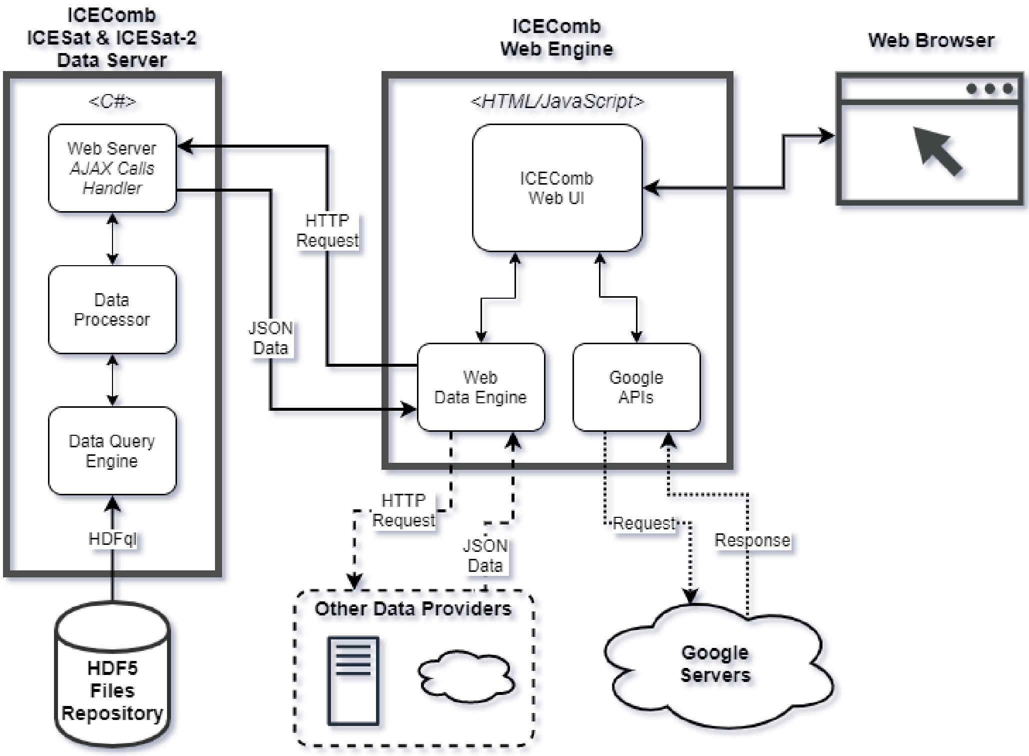


Figure 4.1: ICEComb architecture overview.

The ICEComb tool was developed using standard languages and open APIs to allow the possibility to be extended and, for instance, allow data mashup by aggregating other data providers on the same interface, such as other earth observing system satellites missions. This opens the possibility for end-users to expand the research context of a specific geographic area or study.

4.2 ICEComb data server

Data Server actions are limited to data file handling operations such as filtering, querying and gathering information, while all other activities related to data analysis and presentation for end user, handling of data requests and responses, and other interface tasks are offloaded to the client side.

The Data Server was designed with the RESTful architecture pattern [31] in mind, so requests are made through the use of JavaScript URI libraries, payloads are formatted in JSON (JavaScript Object Notation), and the server does not maintain the client's state information.

The server was developed using the C#language in the latest .NET Framework (version 4.8) and in the x64 bit architecture. The choice in this programming language was based on it being a fast modern programming language, scalable and with a rich library.

4.2.1 HDFql – Hierarchical Data Format query language

ICESat data files are stored using the HDF5 format which was designed to manage very large numerical data collections. However, applications using HDF (Hierarchical Data Format) and HDF5 data manipulation APIs tend to be hard to implement due to their low level of abstraction.

The official API for handling these type of files is the "HDF5 API", from the HDF Group, a non-profit organisation with the aim to create a scientific data file format with an architecture-independent software library.

The HDFql (Hierarchical Data Format query language) API is also made available free of charge and royalty-free, and is the first data interface to allow HDF5 data access through a high-level language with a syntax similar to SQL (Structured Query Language), that is also the standard for query languages and Web applications. For these reasons, this API was chosen instead of the official API to handle the HDF5 files.

Firstly released in 2016, HDFql was developed and is maintained by a team of international and interdisciplinary experts and engineers from universities of the United States and Europe, large research organisations like CERN, and several top corporations.

The main objective of HDFql is to reduce the complicity associated with handling HDF5 files. Because it is based on a declarative language, it allows to focus on what data to obtain instead of how to get it. In contrast, the official HDF5 API is a low-level language with more than 400 functions, which has been one of the main difficulties when it comes to creating applications that handle HDF5 files.

Listing 4.1: HDF API VS HDFql API.

```
// C# HDF API
BinaryWriter file = new BinaryWriter(new FileStream("result.bin", FileMode.Create));
H5DataSetId datasetID;
H5DataSpaceId dataspaceID;
int [,]data;
int i, j, rows;
datasetID = H5D.open(fileID, "my_dataset");  // we assume that the file was opened previously
dataspaceID = H5D.getSpace(datasetID);
rows = H5S.getSimpleExtentNPoints(dataspaceID) / 1024;
data = new int[rows, 1024];
H5D.read(datasetID, H5T.H5Type.NATIVE_INT, new H5Array(data));
Order(data);  // call hypothetical function that orders "data" in an ascending way
for(i = 0; i < rows; i++)
   for(j = 0; j < 1024; j++)
      file.Write((int) data[i, j]);
file.Close();
H5S.close(dataspaceID);
H5D.close(datasetID);

...........................................................

// C# HDFql API
 HDFql.Execute("SELECT FROM my_dataset ORDER ASC INTO BINARY FILE result.bin");
```

Credits: HDF Group.

Listing 4.1 represents the following scenario: reading an entire HDF5 dataset and writing it in ascending order to a binary file called "result.bin", written in C#, first using the official HDF5 API and then the HDFql API. As shown, one HDFql instruction can replace dozens of lines of code written in the official API, greatly simplifying programming and code readability.

At its core, HDFql does not implement new low-level methods for accessing the HDF5 files, but instead it is built using the official HDF5 API.

Although based on a SQL-like syntax, HDFql it is still a very typed language. In order to be able to execute queries and process its results efficiently, a number of mechanisms were developed to optimise and generalise the Data Query Engine.

To accomplish a more efficiently execution of HDFql queries, the Data Processor and the Data Query Engine must be tightly coupled because the Data Processor needs to know what type and size the query result data will be. To satisfy this requirement, without having to write individual code for all types of queries, the server executes a set of routines created to automatically check the involved HDF5 datasets for their variable type and size to act accordingly. This is done effortlessly thanks to the use of a new variable type in C# called "dynamic".

First made available in C# 4.0 and the Microsoft .NET Framework 4.5, the dynamic variable type avoids compile-time type checking and instead it resolves its type at run time, based on the assigned value. This versatility allows the creation of a single algorithm to handle data from any data type and any data size.

4.2.1.1 Memory leak

During the development of the ICEComb server a memory leak was detected while using the HDFql API. In the proprietary data structure called "cursors", that allows a way to traverse result sets, it was noticed that the memory used by the server application was not being released when the execution stopped. This was more noticeable after accessing a large amount of files.

After reporting this issue the HDFql support, we started testing code together and realised that the problem was not in the ICEComb server application but instead in the HDFql API source code itself, namely in the cursor dispose method called *CursorClear()*. As result, this work was an indirect contributor to the HDFql API and the first beta tester of the current version of HDFql (2.3.0), with the fix implemented, before its official release in early February.

Table 4.1: ICEComb Server memory usage.

Task	Cursor dispose method test	
	Before correction	After correction
Program start	6.7 MB	6.6 MB
Access 100 files	32.3 MB	21.3 MB
Access 1 000 files	170.0 MB	21.2 MB
Access 9 228 files	1 428.2 MB	21.2 MB
Access 9 228 files (2nd run)	2 539.6 MB	21.3 MB

Table 4.1 illustrates the memory used by the ICEComb Server while accessing a test sample of 9 228 files, before and after correcting the memory leak. The tasks are cumulative and the measurements were made in the Windows Task Manager and after the server execution was finished, so that they correspond to the state of inactivity and not at peak execution.

The specific use scenario of the ICEComb Server, more specifically the need to make thousands of calls to access different files, was an ideal test case for discovering this bug that previously had not been detected. The way the ICEComb Server was implemented was even the target of some curiosity on the part of the HDFql contact, who showed interest in understanding the purpose of the tool and how the ICEComb Server dealt with the challenges of having to access such a high number of files.

4.2.2 User interface

The ICEComb server application has a graphical interface, shown in Figure 4.2, which is organised in two parts. On the left are the server configuration parameters, where elements such as the location of the error log file and the data from the ICESat and ICESat-2 missions are set. On the right side there is a presentation area where general information is displayed, such as server and the HDFql versions and a text-area with some server status information.

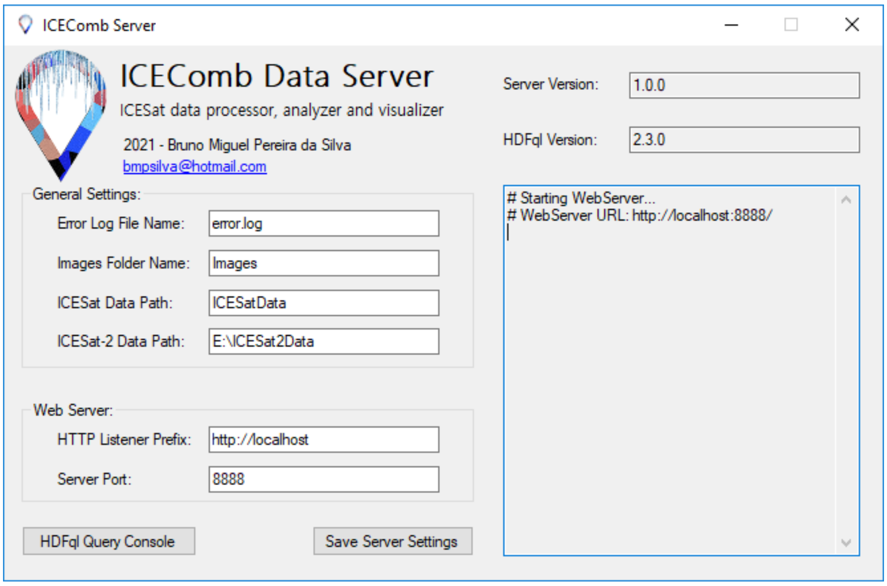


Figure 4.2: ICEComb Server User Interface.

4.2.2.1 HDFql query console

In addition to the Visual Studio debugger, another way to test queries results would be directly on the ICEComb client, but until then the data needed to go through several layers of code source (from the HDFql API to the Data Query Engine and the Data Processor modules) that could obfuscate the test results. Therefore, a quick and immediate way to test HDFql queries was needed and the ICEComb HDFql Query Console was created to fill that gap.

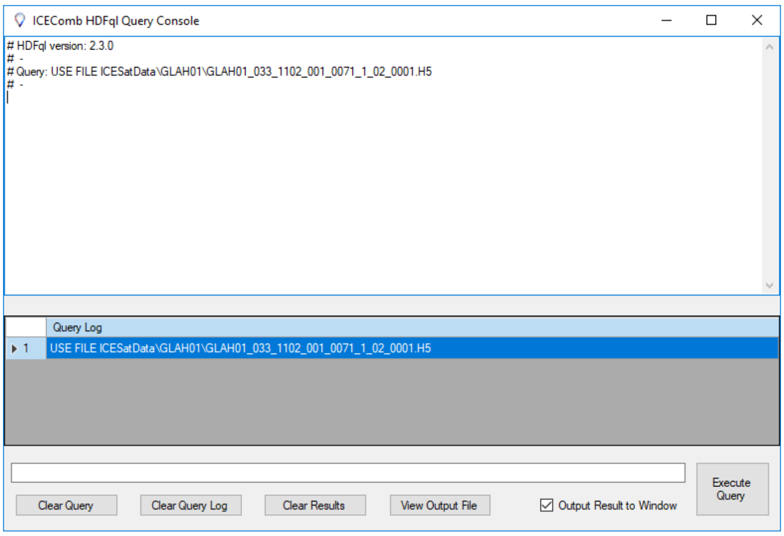


Figure 4.3: ICEComb HDFql Query Console User Interface.

4.3 ICEComb client

The target of this section will be the Client part of the ICEComb solution. Here the user interface is presentment and all functionalities that work together to allow the different main actions such as the data product's selection and filter, the different user interface (UI) options, the data viewing capabilities and finally the data processing section.

By having the ICEComb Client developed in JavaScript, a core technology for the web, the ICEComb solution is able to offload some computation to the client. This approach relegates all data access and processing operations to the data provider, leaving the client responsible to handle all user interface aspects, such as data requests, aggregation, map display and data representation.

Figure 4.4 illustrates the ICEComb Client user interface, which is composed of two main areas, the map on the left and the input controls on the right, grouped in three tabs. On the top of the tab section there is an option to hide it to provide a wider map area, and the data granules execution progress is shown.

4.3.1 Data selection

As visible in Figure 4.4, the first tab is the Data Selection tab and contains the data selection and filtering options. Users start by selecting the data product to display (i.e., the satellite mission and desired dataset) and optionally a date interval filter.

Then there is the Data Granules Management system that allows users to set the allocation of the amount of data granules that are processed each time. Granule files can be browsed by lots (by batch size) or all files at once. The parameters defined here have a high impact on performance.

In the "Dataset Measured Parameters" area, users may filter the granules by its data quality by selecting the QA (Quality Assessment) information parameters, that are obtained from the granule's metadata.

For the ICESat mission, these parameters reside on the granules metadata files (in XML format) and contain the evaluation of the instruments or geophysical parameters expressed in the data as well as specific tests

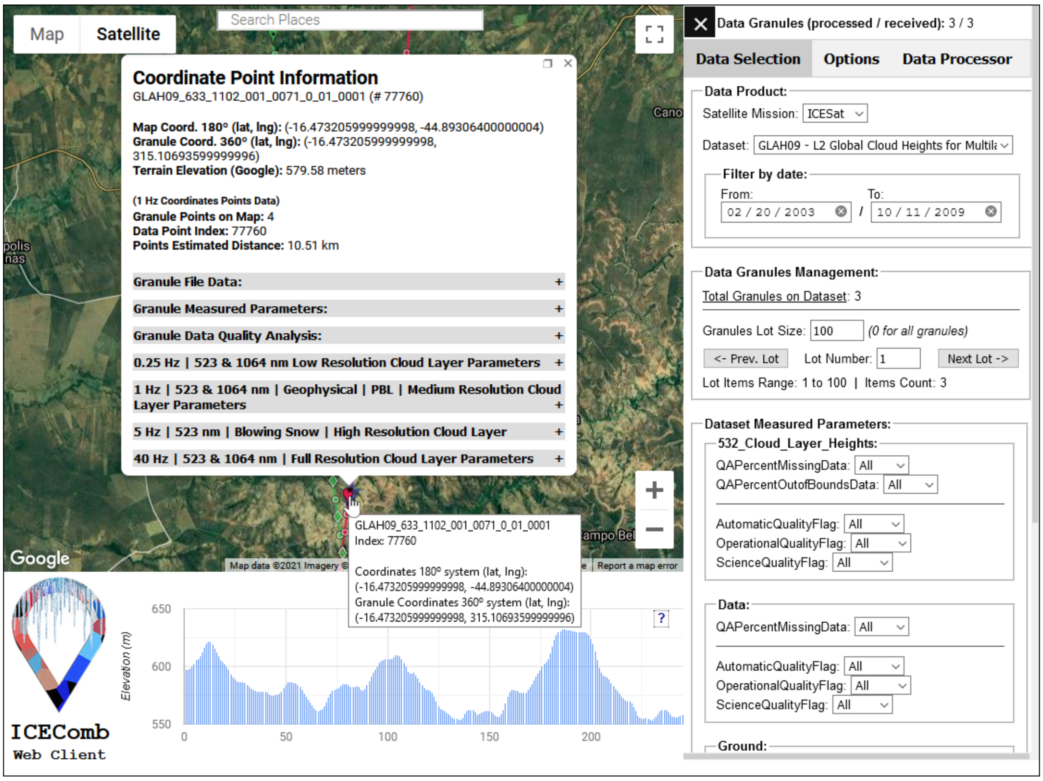


Figure 4.4: ICEComb Client User Interface.

assessment flags. The type and quantity of these parameters are specific to each dataset but usually contain one QAStats (QA Statistic) and three QAFlags per parameter. The QAStats consists in a quantitative value while the QAFlags values are qualitative.

For all the ICESat-2 granules there are only two general quality parameters available, namely the "Granule Quality" and the "Failure Reason". Both parameters are of qualitative type and they are available inside the metadata portion of the granule file.

In the external metadata files of the ICESat-2 granules (also in XML format) there is an area called "DataQuality" but it only contains the description "TBD", which usually means "to be determined". This is probably due the fact that the ICESat-2 data products are still very recent (as of now, it's only in its third revision), so maybe they are not mature enough to be able to determine more data quality assessments parameters.

4.3.2 UI options

The Options Tab, visible in Figure 4.5, contains the general UI options for the ICEComb Client, that are common to both satellite missions and all datasets. In this area, users can control the behaviour of the user interface, the marked elements on the map, as well as to view and define various parameters related to the data handling, map parameters and tool behaviour.

4.3.2.1 Map bounds

One part of the mechanism created to limit, thus optimise the amount of data sent by the ICEComb server, and therefore the amount of data processed by the ICEComb client, is called Map Bounds (or map bounding area). This works by defining two geographical locations, the southwest and the northeast location points (with latitudinal and longitudinal values) and that are used to instruct the ICEComb Server to exclude any data point outside the area defined by those two locations.

The default geographic bound mechanism is the "Map Bounds", meaning that the southwest and the northeast points are automatically determined from the map. But users can also manually set both the bound points by changing the Bound Mechanism to "Shape Bounds" and then by drawing a rectangle on the map to define the valid area for location coordinates. The process is analogous to the Map Bounds system, that is, is also based in the area defined by the southwest and the northeast points, but this time obtained from the drawn shape.

When selected, the "Shape Bounds" option, a set of controls are displayed in a toolbar on the top-centre area of the map (replacing the "Search Places" text box) and allow the execution of three actions, move the shape (hand icon), draw the shape (square icon) and eliminate the shape ("X" icon), as presented in Figure 4.5. After the shape is drawn, in addition to be able to move it around the map, user can also change its size by dragging on of the circle controls in the shape's limits.

Another available options is the ability to "Extend Bounds", meaning that this option will instruct the server to select locations beyond the limits established by the map or by the shape. Since the ground track are represented with the location data obtained from the granule, the "Extend

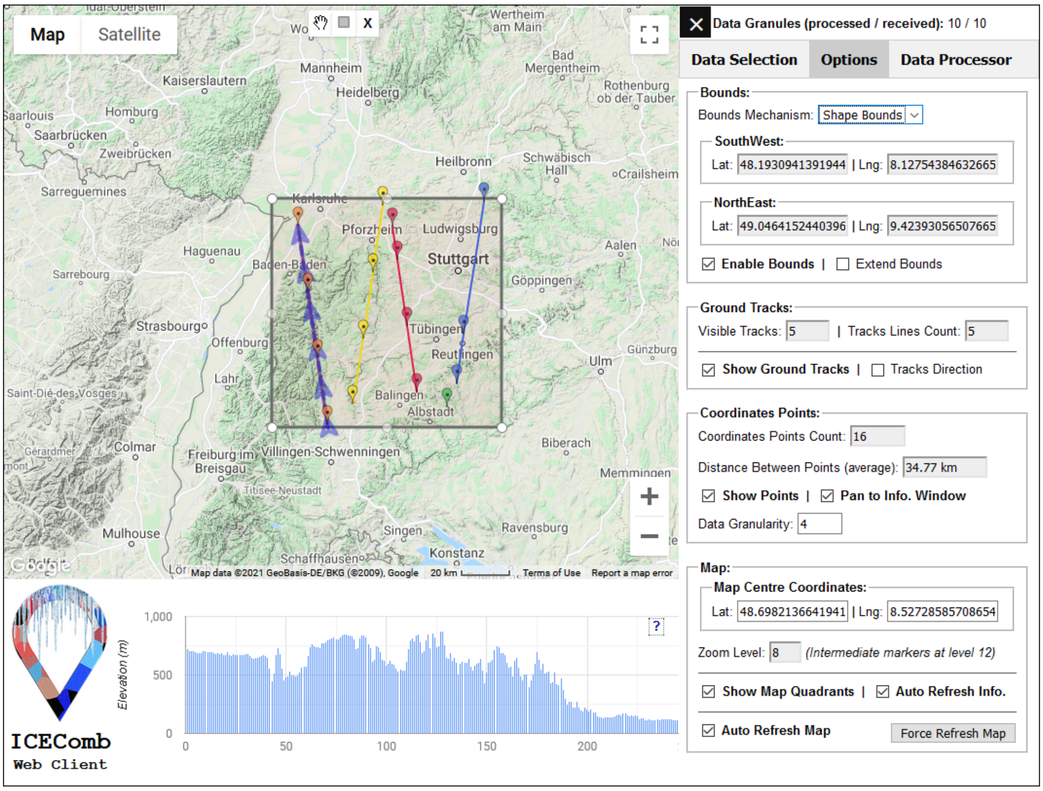


Figure 4.5: ICEComb Client – Map Bounds Mechanism using Shape Bounds.

Bounds" option is useful to draw the ground tracks lines right up to the limit border (map or shape), instead of it being drawn just up to the last received location. When this option is active, the server sends up to two additional geographical locations, one beyond the southwest limit and one beyond the northeast limit, depending if those location exists in the granule. This is a purely cosmetic option and has little or no impact in performance.

4.3.2.2 Ground tracks

The ground track representation in ICEComb is defined by the coordinates of the granule data (thus the coordinates of the collected data), rather than by the an estimated satellite path. This means that the represented ground tracks represent the actual laser's track. Each track is represented by a different colour, one colour by each data granule. This technique was adopted so that the ICEComb tool had a visual representation approach centred on the collected data, meaning that users can quickly identify, at any time, the surface data path and the granule file that originated it.

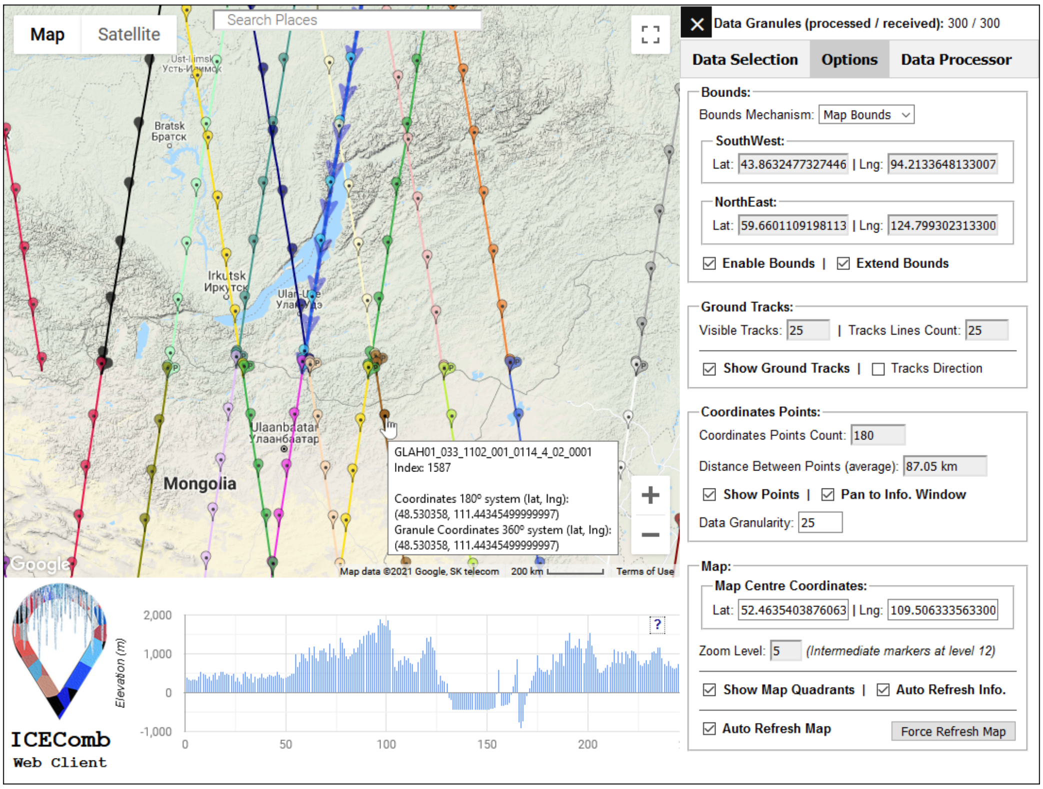


Figure 4.6: ICEComb ground tracks and coordinates points.

For representing a ground tracks line the ICEComb client must have at least two geolocations (latitude and longitude coordinates pairs) belonging to the track. This is not a problem for low map zoom levels since the map area covers a lot of geolocations available for granule file. In the other hand, high map zoom levels are problematic due its small visible area. Under these circumstances, the visible map area may contain only one geolocation per ground track or even none at all.

To resolve this issue, a specific procedure to handle high zoom ratios was created and is illustrated in Figure 4.7. Up to the zoom level 12, the map bound are defined by the southwest and the northeast bounds, but for zoom levels higher than 12, the map bounding system enters in a different working mechanism. For map zoom level 13 and beyond, the system keeps the using the southwest and the northeast bounds from the level 12 as reference, offsetting them as the map moves to maintain the correct reference proportions. In other words, the southwest and the northeast bounds from the level 12 are the highest bounding mechanism available.

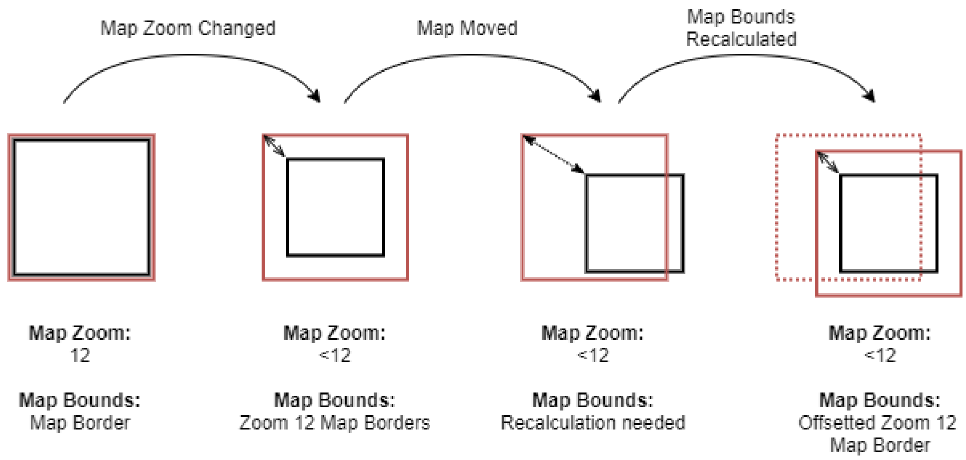


Figure 4.7: Diagram for the high zoom map bounds mechanism.

The recalculation of the new map bounds is done by adjusting the southwest and the northeast locations to the new map position by calculating their offset. This is done by determining the distance (in meters) and the heading (in degrees) from the borders from the zoom 12 map, to the current map borders as initial reference. To calculate the location displacement, it is necessary to determine the distance travelled and its heading, then use these values to calculate the offset of the deviation. Theses calculations are executed by taking advantage of the methods existing in the Googles Geometry Library accessible in the Google Maps API.

In the Ground Tracks section UI (Figure 4.6), users are presented with the number of visual tracks on the map and the count of track lines. The first represent the number of different satellite track on the map while the second represent the number of track fragments. This last parameter can be greater than the number of tracks, due to the fact that the representation of the trajectory of a satellite orbiting around the Earth, appears to be sinusoidal when mapped onto a flat plane.

Therefore, depending on the characteristics of the data product, the selected map area and zoom level, one ground track may cross that area more than once. Each track segment is drawn with the same colour (since only one colour is used per granule file). As the ICEComb tool has specified 21 unique colours with optimised contrast between them, the 22nd track will have the same colour as the first track, and so on.

Other options available for the ground tracks include the possibility to toggle on and off the display of the tracks and an option to show the track direction. When this option is active, arrows indicating the direction are added along the track, based on the satellite heading for that specific granule.

To help to characterise the ground elevation underneath the track, the ICEComb tool can produce a 256 samples ground elevation plot, as shown on Figure 4.6, below the map area, by simply clicking on the desired track line. The plotted track is marked with a thicker blue line with the indication of the plotted direction, that is also based on the satellite heading for that track. Data for this functionality is obtained from the Google Elevation Service that provides elevation data for the surface of the Earth, including depth locations on the ocean floor. When the exact elevation is not available, the service interpolates data from four nearest locations and returns its average.

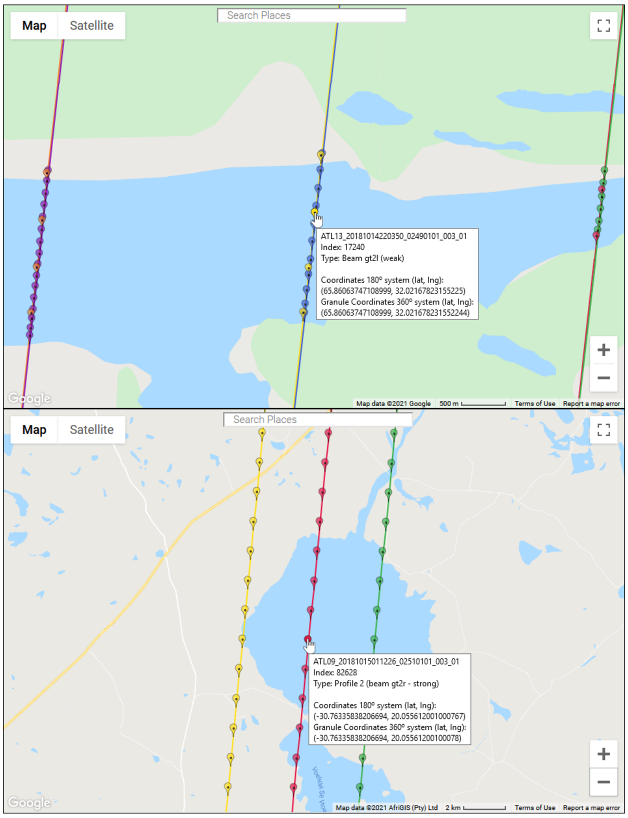


Figure 4.8: ICESat-2 ground tracks view. On the top, a six ground track representation, three main laser pulses with the weak and strong beams (ATL13). On the bottom, a three ground track representation, representing three profiles with data from the strong beam (ATL09).

The ground tracks representation mechanism is identical between the ICESat and ICESat-2 data products. The main distinction is that each data products from the ICESat mission contains data for a single ground track, while the split of the laser bream in the ALTAS altimeter in the ICESat-2 mission, generates three or six ground tracks, depending on the data product.

The ATL04, ATL09, ATL11 will generate tree track, due the profiling nature of these datasets (that combine the weak and strong data in one profile track or just use one beam for the profile), while the remaining datasets will generate the expected six tracks (three beams split in weak and strong pairs), as showned in Figure 4.8.

On ICESat-2, the ground tracks are identify from left to right—along the direction of the satellite—with the letters "GT", meaning ground track, followed by the beam number (from 1 to 3) and the indication of the weak beam with the letter "L" (left beam) and strong beam with the letter "R" (right beam), as shown in Table 4.2.

Table 4.2: ICESat-2 ATLAS Ground Track Pattern.

Beam 1		Beam 2		Beam 3	
Weak	Strong	Weak	Strong	Weak	Strong
GT1L	GT1R	GT2L	GT2R	GT3L	GT3R

4.3.2.3 Coordinates points

As seen before, satellite data obtained from a given location, has associated with it the respective latitude and longitude coordinates (geolocation), which correspond to the exact geographic point where data was collected. This geolocation is used by the ICEComb client to mark on the map the location where the sample took place, using specific shapes, and also using them as an access point to view the collected data. These map elements are referred as Coordinates Points.

On the ICESat mission, five different coordinate points were identified, four based on the sampling rate of the laser pulses (see Table 4.3) and one that was added to mark the last 1 Hz point of the ground track.

Table 4.3: ICESat data sampling rates by dataset.

Dataset	0.0625 Hz (16 sec.)	0.25 Hz (4 sec.)	1 Hz	4 Hz	5 Hz	10 Hz	40 Hz
GLAH01			×				×
GLAH02			×		×		×
GLAH03	×	×	×	×			×
GLAH04			×			×	×
GLAH05			×				×
GLAH06			×				×
GLAH07			×		×		×
GLAH08		×	×		×		
GLAH09		×	×		×		×
GLAH10		×	×				
GLAH11		×	×				×
GLAH12			×				×
GLAH13			×				×
GLAH14			×				×
GLAH15			×				×

The GLAH03 and GLAH04 datasets are not represented on the ICEComb tool because these granules do not contain any surface or atmosphere data, but rather "Global Engineering Data" (satellite housekeeping data used to calibrate data values) and "Global Laser Pointing Data" (data from the spacecraft star-tracker, laser reference system, and additional spacecraft instrument data to calculate precise laser pointing), respectively. Consequently, since the 0.0625 Hz, 4 Hz and 10 Hz data rates are limited to those datasets, they are not contemplated in the ICEComb tool.

Coordinates from the 1 Hz data rate are used to create the pin shaped marker and are the points from which the satellite ground tracks are drawn.

The other data rates markers are represented in Table 4.4 and indicated as follows:

- data points that contain 0.25 Hz data are represented by a pin with a black star on the top-right side;
- 5 Hz data markers are represented by a diamond marker;
- 40 Hz data markers are symbolised by a circle.

There is also a special marker, shaped as a 1 Hz pin but tilted to the right with a "P" character on the centre, that represents a predicted location (calculated assuming satellite nadir position) to indicate the termination of the granule ground track and that also matches to the initial coordinate of the next granule on the same track. This marker only contains the general information and it is used to provide a ground track reference, after the last marked 1 Hz coordinate, to allow the representation the intermediates data markers (namely the 5 Hz and 40 Hz data rates).

Table 4.4: ICESat data products marker's icons and meaning.

				
0.25 Hz	1 Hz	Lasts 1 Hz marker	5 Hz	40 Hz

To optimise the query of the geolocations in the ICESat data products, it was chosen the 1 Hz data as the main lookup location dataset, since is the common frequency between all data products. This approach offers a normalised way of accessing the location data for all data products but also the advantage of using a dataset that is 5 to 40 times smaller than the 5 Hz or the 40 Hz alternatives.

The geolocations below the 1 Hz are referred as "intermediate markers" and they are only displayed when there is a map zoom level lower or equal to 12. This systems offers performance advantages in the server, from lowering the data transfer to the clients, and in the client, since it lowers the graphical processing need by reducing the number of marker's displayed in the map.

Because data is stored differently, the ICESat-2 data products have a different approach for marking the Coordinates Points. The internal data is not grouped by data rates but rather it is stored by its availability, meaning that data is only recorded if it is collected with a degree of certainty, instead of being recorded in a expected and fixed rate. The only exception is the ATL09 data product that has its data grouped at 1 Hz and 25 Hz. Another difference is that in the ICESat-2 data products there are no "intermediate markers", due the same reason.

Table 4.5: ICESat-2 data products marker's icons and meaning.

	
General / 25 Hz	1 Hz

The marker's icons must follow the same colour pattern as the ground tracks. Since the ground tracks follow a colour palette of 21 colours and the fact that they can personalised in the code, it was a problem to have the markers available in all different colours. To resolve this issue while keeping the colours palette flexible, the ICEComb client uses two distinct mechanism to create the marker's icons on demand and "on the fly".

One method depends on the Dynamic Icons functionality from the Google Charts API, that is used to dynamically generate the icons for the pin, the tilted pin with the letter "P" and the pin with a black star, by are using set o parameters, including the colour.

Since the circle and the diamond icons were not available in that API, it was necessary to implemented a dynamic icon generator engine for the ICEComb tool. The generator takes advantages of the new HTML element introduced in the HTML5 called canvas - that allows the creation of 2D drawings via JavaScript.

In a similar way as in the Dynamic Icons in the Google Charts API, it was created a function that receives the icons parameters, mainly the shape and colour, that use a hidden canvas element on the ICEComb page to draw the icon. Then the canvas is converted to an image, in the PNG format, and sent to the map for representation. This image is also stored in memory for future use to avoiding the repetition of process again. This way, there is no need to regenerate icons in the new colours externally if the colour palette changes.

In the group UI corresponding to the coordinate points available in the Options tab (Figure 4.6), users can view the total of the coordinate points marked on the map as well as the average distance between all marked points. Note that the total of the coordinate points marked includes markers that are not currently visible, as is the case of the marker added with the option "Extend Bounds".

Also, in the case of the ICESat data products, the intermediate markers (below 1 Hz) are not included in the coordinate points count, as well in the average distance calculation.

There is also an option to toggle the display of coordinate points on the map and an option to pan (i.e., adjust to fit in the map) to the Coordinates Point Information Window. This window is where users can view the data extracted from the granule file for that specific location point. The Coordinates Point Information Window will be described in more details in a dedicated section, further in the document.

When the "Pan to Info. Window" is active, once the Information Window is opened, the map area automatically adjusts (i.e., moves) to fully contain on the window for ease of view.

The last parameter is called "Data Granularity" and its used to define the number of Coordinates Points the client receives per dataset. This parameters is applied to the ICESat and ICESat-2 missions and works as a data divisor, meaning that to increase the amount of geolocations received by the client, the value of the parameter has to decrease, and vice-versa.

For example, if the value is set to 1, the server will return the total amount of the valid geolocation data (the same as sending the data in steps of 1), if set to 2, the server will return about half of the valid geolocation data (the same as sending the data in steps of 2), and so on. The parameter's value is an integer that can be set manually but there is also an automatic system that adapts the data granularity parameter accordingly to the map zoom level and the characteristics of the data product.

The "Data Granularity", the "intermediate markers" and the "Bounds Mechanism" are the main parameters to limit the number of Coordinates Points that are received from the server and provide a positive impact on the system performance by reducing the amount of transferred data and drawn on the map. Another optimisation implemented for the coordinate points (including intermediate markers), is to first check if the marker's geolocation is within the visible area of the map before decide to draw it.

The intermediate markers have an additional visibility parameter that works in a similar way as the "Data Granularity", meaning that its visibility changes based in the map zoom level.

The first zoom level for the intermediate markers is 12 and it is shown 1/4 of the available markers, at zoom level 13 it is shown 1/2 of the intermediate markers, and lastly all intermediate markers are displayed in zoom level 14, as long as they are inside the visible map area.

4.3.2.4 Map area

The map area is the main element of the user interface, as this is where most of the user interactions takes place and where data is presented. First, the ground tracks are drawn, allowing the visualisation of the projection of the satellite data sensor trajectory on the Earth's surface, and then the data points are marked along the tracks. These points are also used to access to the recorded data for that geolocation simply by clicking on them.

The Google Maps API is the base of the graphical component of the Web Engine. It not only handles the presentation of a high detailed satellite images and geographical map, with elevation contour lines, but also is used to handle presentation of the visual elements on top of the map.

This API also offers improved places labels (such as countries, capitals, seas and lakes), country borders contours and road information when comparing with other mapping libraries, and without relying on external libraries. An example of the amount of details offered by the Google Map's satellite imagery system can be observed in Figure 4.9.

In some areas the Google maps API allows the possibility to tilt the map for 45°, allowing an almost 3D view of the surface and facilitating the depth perception of some aspects of the terrain. In this mode users also have the possibility to rotate the view in 90° increments in either direction. The tilt and rotate controls are accessible above the zoom control of the map, on the bottom right (see Figure 4.9).

For studies that require a correct representation of the polar regions Google Maps API may not the best tool for the job since the map relays on the Mercator projection, consequently landmasses such as Greenland and Antarctica turn up larger than they actually are. Although Google introduced a 3D Globe Mode in the public version of Google Maps, released on August 2, 2018, which solves the polar regions representation accuracy, developers cannot use of this mode as this feature is not yet accessible in the Google Maps API.

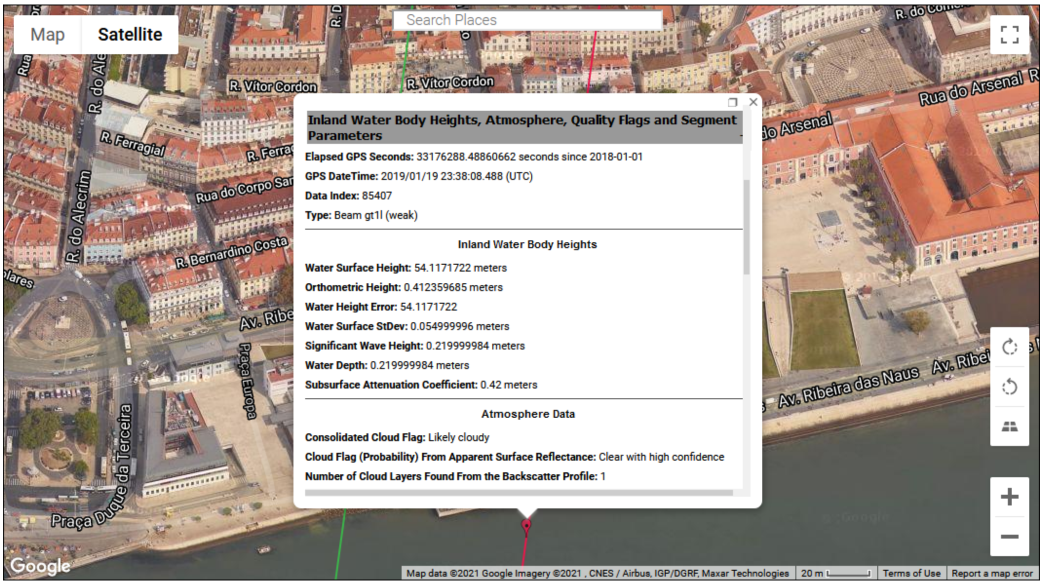


Figure 4.9: View of the Inland Water Body Heights data, from ICESat-2's ATL12 data product with geolocation in downtown Lisbon, Portugal.

Besides being used to characterise the ground track elevation profiles, the Google's Elevation Service is also used to extend the web map functionalities by offering the ability to determine the elevation of any point in the map simply by clicking on it.

Figure 4.10 also highlights a very useful functionality available on the ICEComb interface that is the possibility to search and jump to any locations in the world map. This allows to execute general searches like for countries, capitals, regions and streets but also more specific elements such as watercourses, lakes, basins, oceans, mountains, etc.

In the UI options for the map (the last area in the Options tab in Figure 4.6), there is an area called "Map Centre Coordinates" where users have the ability to check the coordinates of the centre of the map, and bellow it to see the current map zoom level. The "Map Centre Coordinates" also has the possibility to have a latitude and longitude coordinate entered manually by the user so that he can take the map to that location.

The map centre and zoom level are always updated and maintained in the URL of the ICEComb client, enabling an easy way to maintain the state to the current map area while bookmarking or sharing the client link, for example.

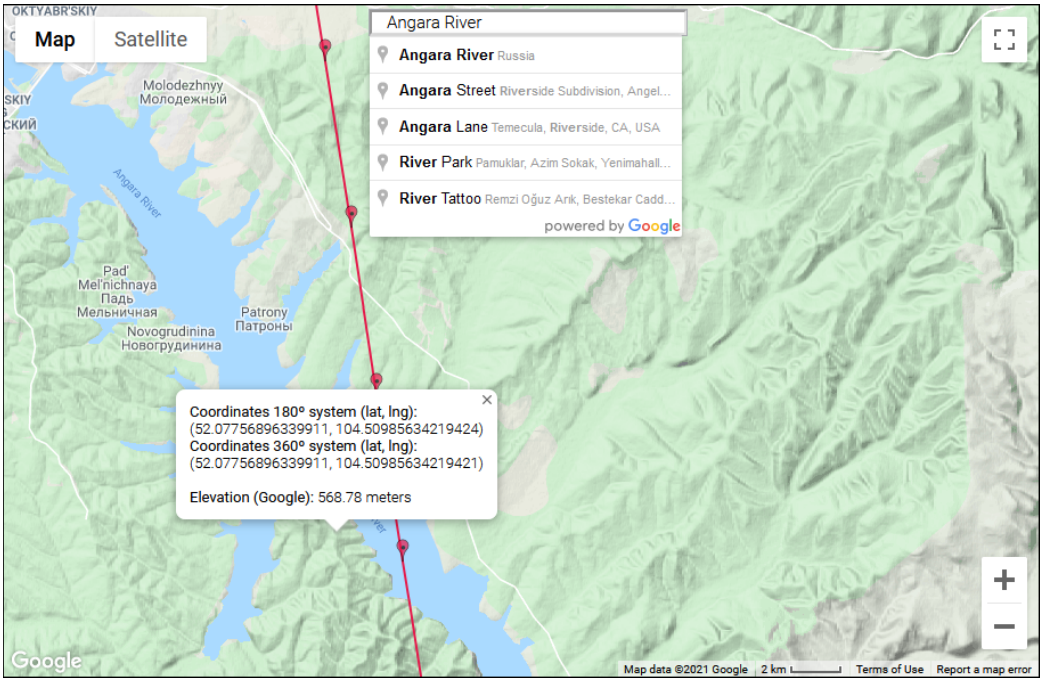


Figure 4.10: Elevation measurement taken close to the bank of the Angara River in Siberia, near the border with Mongolia.

There is an option to disable the coordinate quadrants, that consists in two reference lines that separates the world map in four latitude and longitude areas from the zero coordinate, for a quick visual reference of the geolocation coordinates.

There is an option called "Auto Refresh Info." that allows to toggle on or off the continuous refresh of the UI while loading data from server. When disabled, the UI counters will only update when all actions are finished rather then the current progress. This option offers better performance in clients with low end systems.

Finally the "auto Refresh Map" that toggles on or off the refresh of the ground track and markers, when moving the map or changing the zoom level. With this option disabled, the only way to refresh the map is manually by using the "Force Refresh Map" button.

Coordinate point information window

The Coordinate Point Information (CPI) window is where the contents of the data products from the ICESat and ICESat-2 missions can be viewed.

Data is accessible by clicking on a Coordinate Point icon (Figure 4.4), that has its contents organised in five different sections, the title, information about the geolocation, information about the coordinate point, information about the granule file and data and finally the the data gathered for that location, as represented in Figure 4.11.

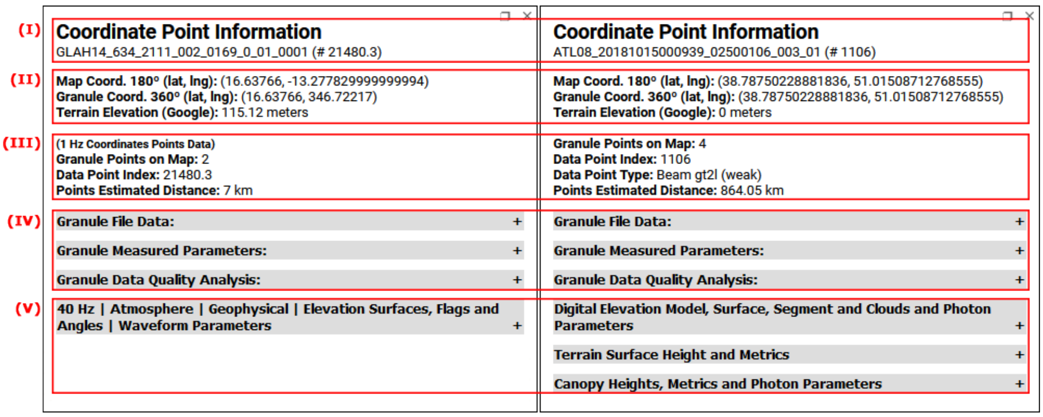


Figure 4.11: Coordinate Point Information Window sections: (left) CPI window from an ICESat data product (GLAH14), and (right) CPI window from an ICESat-2 data product (ATL08).

As evidenced in Figure 4.11, the CPI window structure is similar between the ICESat and ICESat-2 data. The differences are mostly due to characteristics in the internal data structure of the ICESat and ICESat-2 data products and the process in which their altimeters collect data.

Title (I)

The title section contains the name of the granule file name where the data point originate and in front of it the data index number, corresponding to the entry number in the corresponding datasets. In the ICESat data products, the data index number have some peculiarities that are addressed in the section III.

Geolocation Information (II)

The geographic coordinate pairs of the latitude and longitude (geolocation) of the data point is shown both using the 180° coordinate system as well as the 360°. While the ICESat-2 data products and the Google Maps API both uses the 180° longitude representation, the ICESat data products have their location data stored in the 360º longitude representation, thereby both 180º and 360º geographic coordinates systems were chosen to be display at all times.

The geolocation information was also enriched with the use of the Google Elevation service to offer an elevation reference for each location point.

Coordinate Point Information (III)

Due the fact that the Ground Tracks for the ICEComb data products are focused on the 1 Hz data, this section's data only represents data for the 1 Hz Coordinate Points (i.e., do not include the Intermediate Markers).

As the ICESat-2 data does not use different levels of geolocations datasets for marking the Coordinate Points on the map, the information presented here takes into account all Coordinate Points currently presented in the map.

This has implications on the values of the number of Granule Points on Map, the Data Point Index and the calculation of the Points Estimated Distance.

In Figure 4.11, the ICESat CPI window is referent to a 40 Hz Intermediate Marker (IM). Consequently, the Data Point Index has two parts, separated by a dot (21480.3). The first part corresponds to the index for the 1 Hz data (21480) while the second corresponds to the third 40 Hz index from that 1 Hz point (IM position), and the Intermediate Marker Index can be calculated using the following formula:

$$IM\_Index_{IM\_Position} = (1Hz\_Index \times IM\_DataRate) + IM\_Position. \quad (4.1)$$

Considering the example of the previous paragraph,

$$IM\_Index_3 = (21480 \times 40) + 3 = 859203. \quad (4.2)$$

In this example, the $IM\_DataRate$ used is 40, so the $IM\_Position$ ranges from 0 (first element) to 39 (last element), for each 1 Hz point. It is not necessary to manually calculate the data indexes for IMs, as these are available for consultation in the corresponding data section.

Since the ICESAT-2 data can be originate from one of the six laser beams, or in one of the three sets of profiles, the data origin inside the granule file has to be included.

As such, the ICESAT-2 data points have a dedicated files called “Data Point Type” to indicate the data source. For the ICESat data products this is not necessary, since it uses only one data gathering source.

Granule File Information (IV)

This section presents details about the granule file and its data content. The first element is the “Granule File Data” (see Figure 4.12) and contains general information about the HDF5 file, such as the file name, size and data revision, information about the satellite cycle, track and orbit number, and date-time information about the beginning and end of the measurements contained in the file. In this section there is also the option to download the full granule file and its respective XML metadata file.

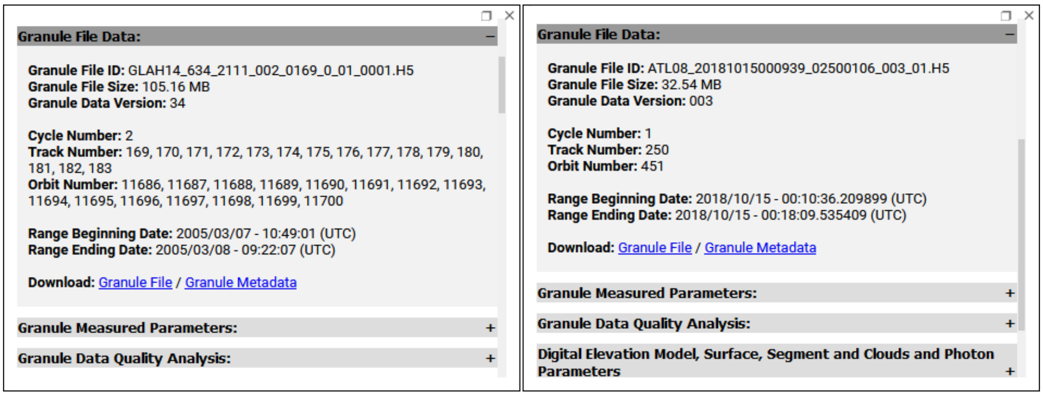


Figure 4.12: Granule File Data elements: (left) ICESat data product (GLAH14), and (right) ICESat-2 data product (ATL08).

A granule file can contain data from different cycles, tracks and orbits of the satellite when belongs to a dataset that was created by aggregating data from various other datasets. That is the case of the GLAH14 that uses data from GLAH06 (Global Elevation Data) in combination with GLAH05 (Global Waveform-based Range Corrections Data) to create a data product focused on land surface altimetry data.

The second element in the Granule File Information section is the Granule Measured Parameters where the information about the data quality of the granule can be found (refer to Figure 4.13). As said before, the information contained in this element varies from dataset.

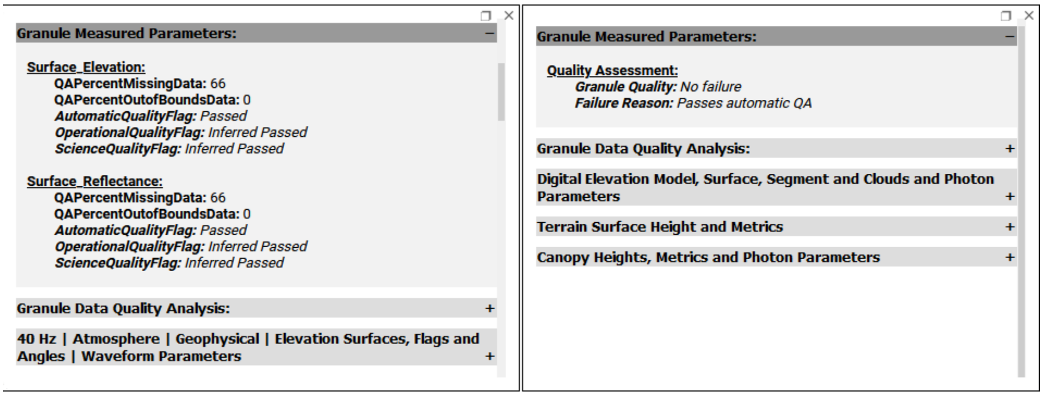


Figure 4.13: Granule Measured Parameters elements: (left) ICESat data product (GLAH14), and (right) ICESat-2 data product (ATL08).

The last element is the Granule Data Quality Analysis that contains a set of predetermined, structured, computer-parseable and inventory-level metadata analysis that are presented as a set of images (Figure 4.14).

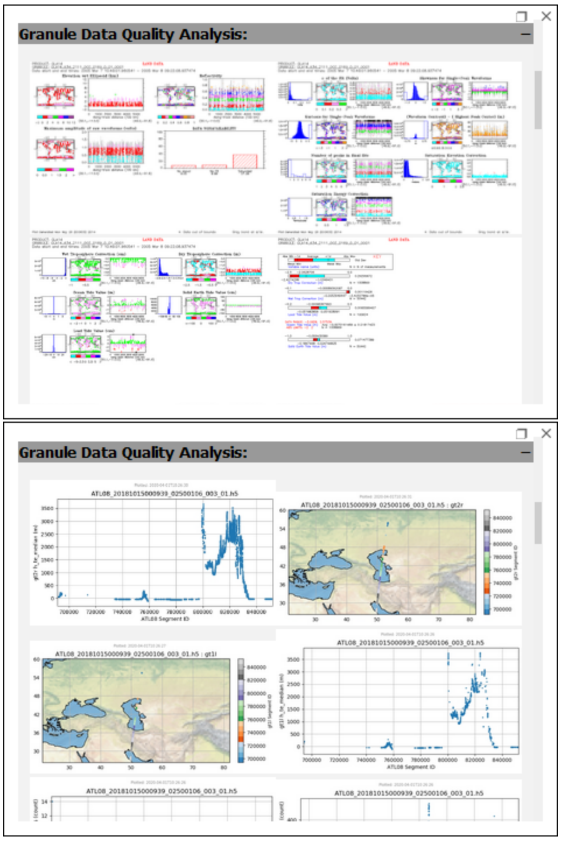


Figure 4.14: Data Quality Analysis elements: (top) ICESat data product (GLAH14), and (bottom) ICESat-2 data product (ATL08).

These studies were produced by the Science Team responsible for generating the data products and the type of test executed varies by dataset.

In order to facilitate the images consultation and the data that they present, a visualisation gallery was added with functions that allows users to view images in detail, zoom in, download and easily browse the image collection, as exampled in Figure 4.15 and Figure 4.16.

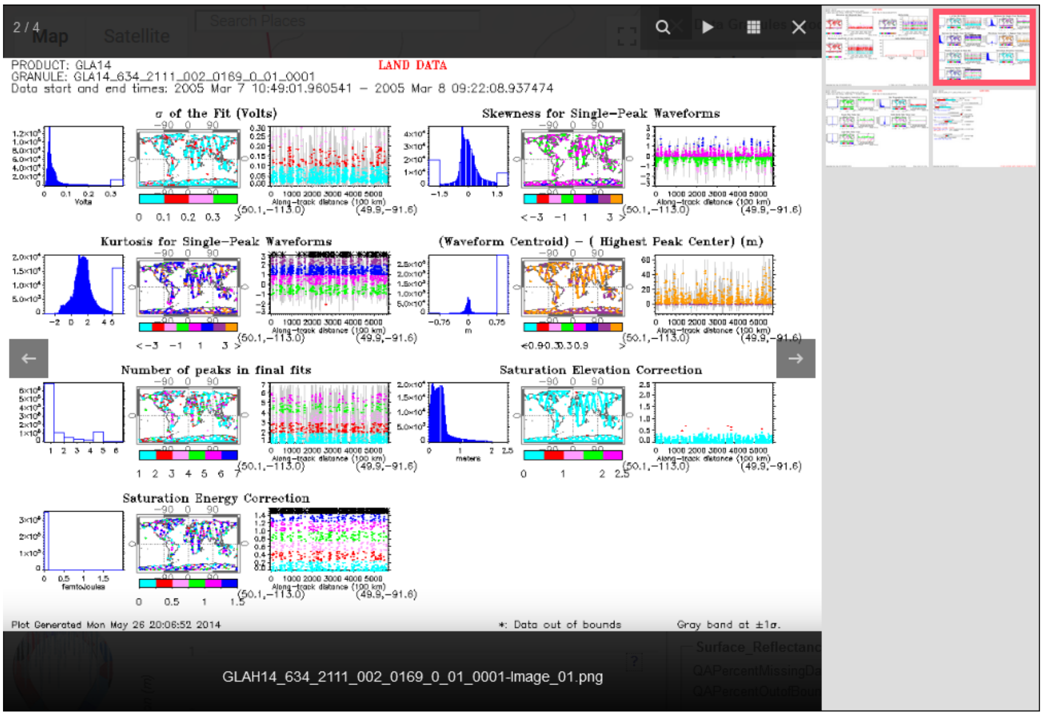


Figure 4.15: Data Quality Analysis elements zoomed in, showing an ICESat data product (GLAH14).

Recorded data (V)

The data collected by the satellite for the specific Coordinate Point is presented in this section, organised in several sub-elements, depending on the data contained in a given dataset.

In all sub-elements in the Recorded Data there is a common header that contains the a chronological reference of the data acquisition, the data index (the entry number in the datasets) and in the case of the ICESat-2 mission, the beam identification or the corresponding data profile, as shown in Figure 4.17.

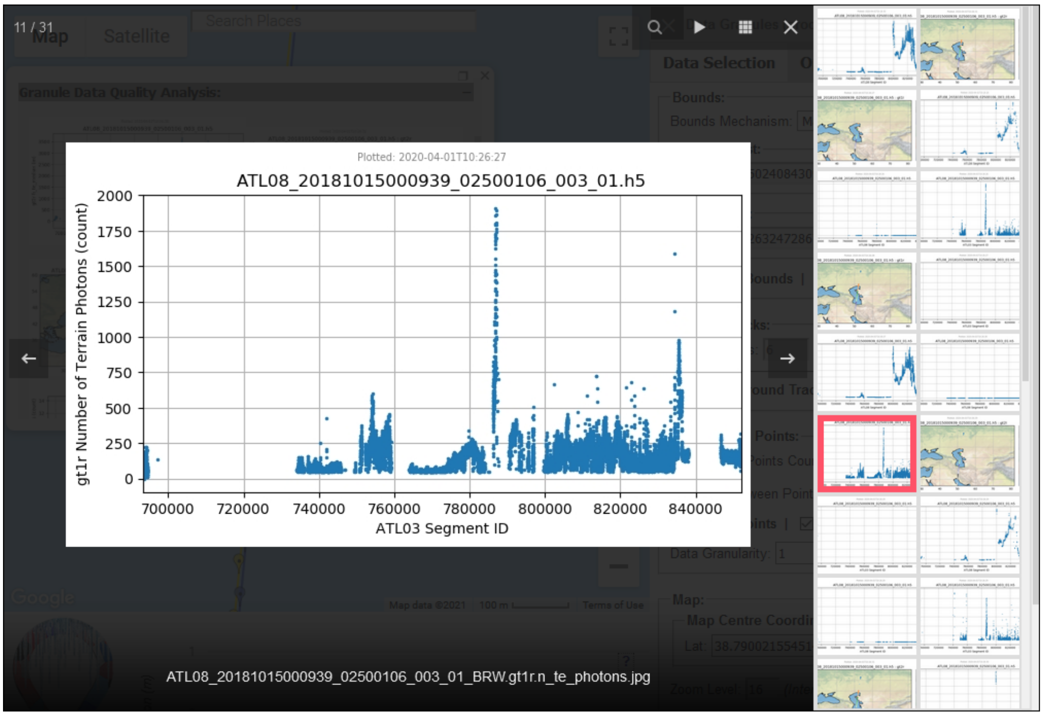


Figure 4.16: Data Quality Analysis image elements zoomed in. Example from an ICESat-2 data product (ATL08).

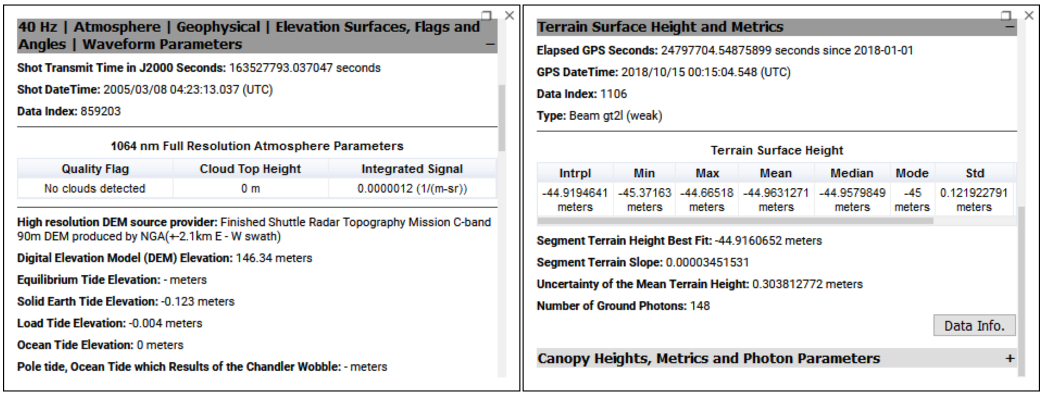


Figure 4.17: Recorded Data elements: (left) ICESat data product (GLAH14), and (right) ICESat-2 data product (ATL08).

Presenting the data indexes is a useful feature that allows users to backtrack the displayed data precisely to the raw data entry in the granule's HDF5 file.

Several data presentation components have been created to best present data in a meaningful way, from the simple single data entry, represented by the data identifier, its value and units, to more complex structures as is the case of tables, charts and also by using data translation or conversion mechanisms.

The single data entry format and an example of data entry are shown below:

```
<data_identifier>:  <data_value> <units>
Digital Elevation Model (DEM): 146.34 meters
```

The corresponding HDF5 data path and attributes are indicate below:

```
Data path: /Data_40HZ/Geophysical/d_DEM_elv
units:  meters
_FillValue:  1.7976931348623157E308
long_name:  DEM Elevation
description:  Elevation at the footprint location from the
SRTM30 (GTOPO30 + SRTM) Digital Elevation Model (DEM). The
reference frame for the DEM elevation was changed to the
TOPEX/Poseidon ellipsoid to make it consistent with the GLAS
elevations.
```

In order to make the meaning of <*data_identifier*> as clear as possible, this attribute is defined by using HDF5 dataset's metadata attribute "long_name" and sometimes combined with information contained the "description" attribute.

The <*data_value*> is the HDF5 dataset value, for the corresponding data index, but before it can be display it has to be compared with the value of the attribute _FillValue, that defined with values are considered null in a HDF5 dataset. If is equal, then the value is replaced with the hyphen character "-" (representing no available data), otherwise the HDF5 dataset's value is used.

The <*units*> element is optional since not all values have a corresponding unit to display.

There is also a type of HDF5 dataset that is based on flags, that usually have a translation associated, below an example and the HDF5 dataset data:

```
<data_identifier>:  <data_value>
High resolution DEM source provider:  Finished' Shuttle Radar
Topography Mission C-band 90 m DEM produced by NGA (+−2.1km
E-W swath)
```

The corresponding HDF5 data path and attributes are indicate below:

Data path: `/Data_40HZ/Geophysical/i_DEM_hires_src_1`

long_name: `High Resolution Source Flag`

flag_values: `0,1,2,3,4,5,6`

flag_meanings: `none pre_srtm srtm greenland antarctica cded_90 olap_srtm`

description: `High resolution DEM source provider; Indicates if no high res source available, unfinished research Shuttle Radar Topography Mission (SRTM) C-band 90m DEM produced by JPL (+−1.1km E-W swath), finished' SRTM C-band 90 m DEM produced by NGA (+−2.1km E-W swath), ICESat Greenland V1 1km DEM, ICESat Antarctica V1 500m DEM, 90m Canadian Digital Elevation Data (CDED), Data in overlap area of SRTM and CDED. Value reported is from finished SRTM C-band 90 m DEM.`

As shown in the example above, the data is not presented using its flag value (from 0 to 6) but instead the ICEComb tool translated the flag value into its description thus substantially facilitating the data interpretation. Also, for greater context, the translation table was created manually by using the information in the *description* attribute instead of the data available in the *flag_meanings* attribute, which has a more limited context.

Tables are important elements in data visualisation as they allow to group related data and visualise it in a common structure to quickly interpret their meaning. Tables in the Recorded Data section obey the same principles as the single data entry, meaning that the headers also are chosen to maximise the interpretation, the respective units are added to the values and if the data has a meaning list associated then a translation also takes place. There is also boolean values that are resolve into symbols, again, to facilitate readability. This functionalities are visible in Figure 4.18.

In Figure 4.19, a presentation of a table that contains data aggregated in multiple rows, and each rows with a different meaning. In these cases, a column with the row number is added to the table, with a corresponding row description presented underneath it.

Depending on the values types added to the table, they can also be converted to scientific notation (when numbers with decimal places are very small or very large) or metric prefixes (when integer numbers are very large) in order to take less space on screen and also improve readability.

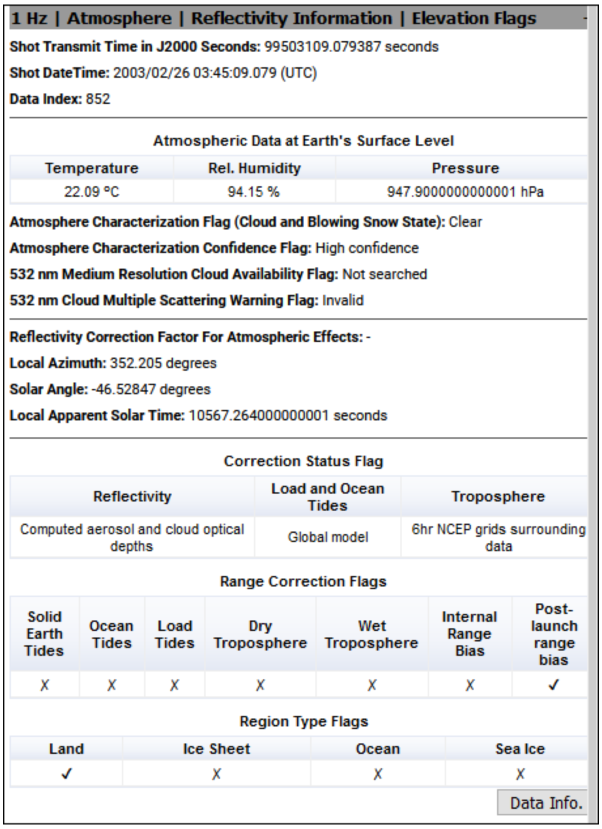
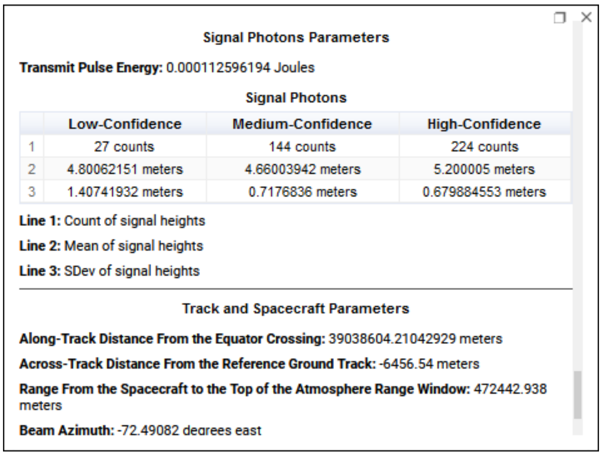


Figure 4.18: Coordinate Point Information tables: example from an ICESat data product (GLAH06).

Chart representations allows data interpretations that otherwise wold not be possible. In the ICEComb tool there are two essential charts types, the line chart and the histogram. Both these elements were also optimised for the use with satellite altimetry data by also implementing the same convention principle as the one used with the tables (scientific notation or metric prefixes) but most importantly the processing of infinite numbers or invalid values (using the aforementioned *_FillValue* attribute to validate data).

To correctly process the infinite and the invalid values, the chart mechanism breaks the chart representation every time one of these values is encountered, and starts again when encounters valid values, while maintaining a linear x-axis, as shown in Figure 4.20, on the left. This is important because in this way there is a correct representation of the encountered data on a axis that can represent height or distance, for example.



Signal Photons Parameters

Transmit Pulse Energy: 0.000112596194 Joules

Signal Photons

	Low-Confidence	Medium-Confidence	High-Confidence
1	27 counts	144 counts	224 counts
2	4.80062151 meters	4.66003942 meters	5.200005 meters
3	1.40741932 meters	0.7176836 meters	0.679884553 meters

Line 1: Count of signal heights
Line 2: Mean of signal heights
Line 3: SDev of signal heights

Track and Spacecraft Parameters

Along-Track Distance From the Equator Crossing: 39038604.21042929 meters
Across-Track Distance From the Reference Ground Track: -6456.54 meters
Range From the Spacecraft to the Top of the Atmosphere Range Window: 472442.938 meters
Beam Azimuth: -72.49082 degrees east

Figure 4.19: CPI table with data aggregation: example from an ICESat-2 data product (ATL09).

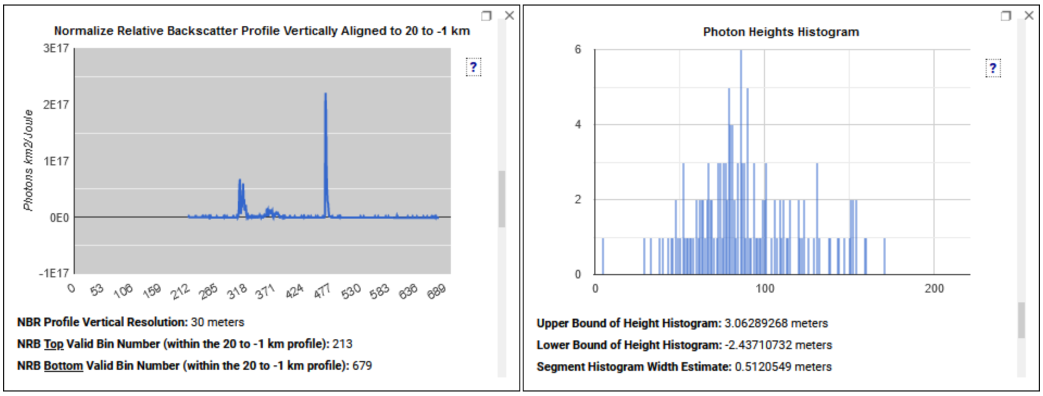


Figure 4.20: Coordinate Point Information chart types: examples of (left) a linear chart and (right) a histogram, with data from ICESat-2 data products ATL04 and ATL07.

Charts data can sometimes have long x-axes, which causes the lines or columns to be very compact, not allowing a correctly visualisation of some details such as data variations. To tackle this issue, all charts in the ICEComb tool have the possibility to dig the represented data by allowing user to drag a portion of the chart for zoom in, as shown in Figure 4.21. Also, all represented values can be consulted by simply placing the mouse pointer at any point in the graph, together with their units.

Users can also view detailed information about the HDF data records that were used to generate the displayed data, such as the HDF data paths, the HDF dataset, the long name, the units, valid maximum and minimum

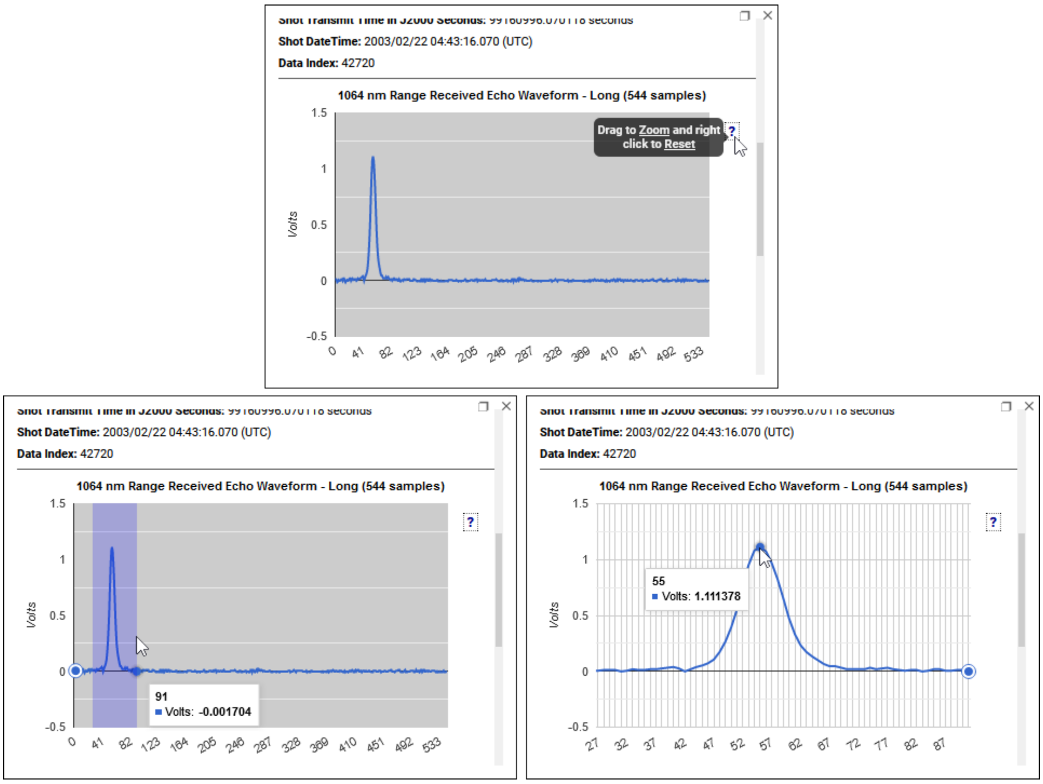


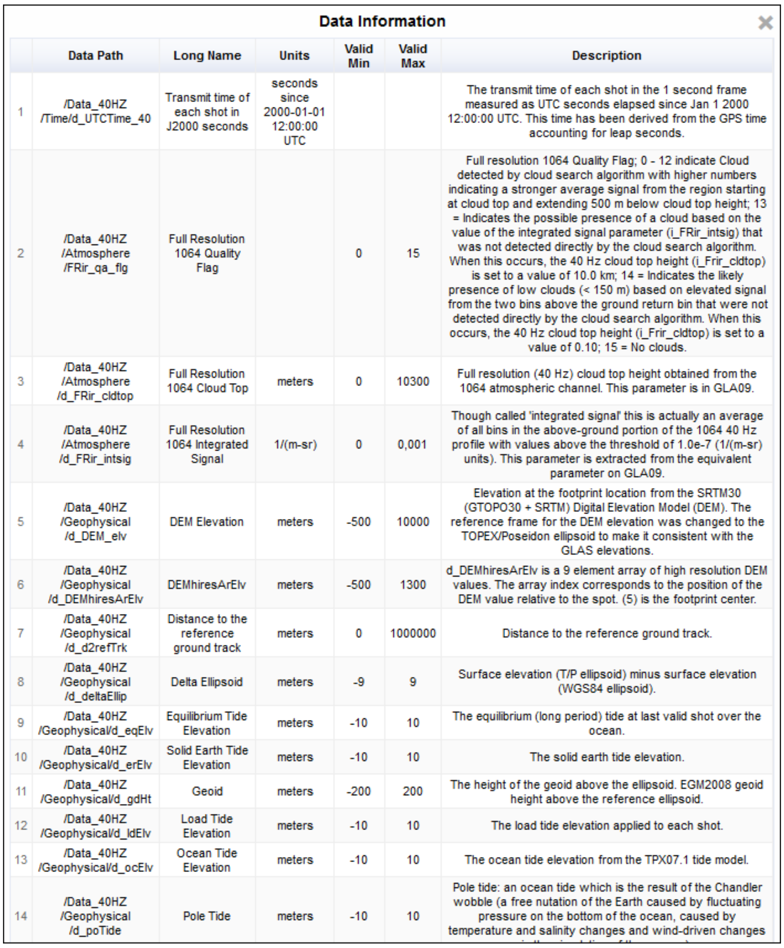
Figure 4.21: Coordinate Point Information charts interactivity: example from an ICESat data product (GLAH01).

values and data description, as is displayed in Figure 4.22. This information is accessible by clicking on the button with the label "Data Info." located at the end of each data element presented in the Recorded Data section, as represented in the Figure 4.18, for example.

Blank values represent HDF metadata attributes that are not present or do not apply to a given HDF dataset.

The last functionality in the Recorded Data section is the ability to maximise the CPI window, allowing to have a wider view of the data existing in all the section. The maximise button is available in the top right area of the CPI window, next to the close button, and when in maximised mode, all section of the CPI window are automatically expanded (see Figure 4.23).

Among the ICESat and ICESat-2 data products, the ICEComb tool implemented access to more than 1 900 HDF dataset data paths with more than 49 400 associated HDF metadata attributes, that were used to produce the presentation of 707 individual data entries, 176 tables and 52 charts.



Data Information

	Data Path	Long Name	Units	Valid Min	Valid Max	Description
1	/Data_40HZ /Time/d_UTCTime_40	Transmit time of each shot in J2000 seconds	seconds since 2000-01-01 12:00:00 UTC			The transmit time of each shot in the 1 second frame measured as UTC seconds elapsed since Jan 1 2000 12:00:00 UTC. This time has been derived from the GPS time accounting for leap seconds.
2	/Data_40HZ /Atmosphere /FRir_qa_flg	Full Resolution 1064 Quality Flag		0	15	Full resolution 1064 Quality Flag; 0 - 12 indicate Cloud detected by cloud search algorithm with higher numbers indicating a stronger average signal from the region starting at cloud top and extending 500 m below cloud top height; 13 = Indicates the possible presence of a cloud based on the value of the integrated signal parameter (i_FRir_intsig) that was not detected directly by the cloud search algorithm. When this occurs, the 40 Hz cloud top height (i_Frir_cldtop) is set to a value of 10.0 km; 14 = Indicates the likely presence of low clouds (< 150 m) based on elevated signal from the two bins above the ground return bin that were not detected directly by the cloud search algorithm. When this occurs, the 40 Hz cloud top height (i_Frir_cldtop) is set to a value of 0.10; 15 = No clouds.
3	/Data_40HZ /Atmosphere /d_FRir_cldtop	Full Resolution 1064 Cloud Top	meters	0	10300	Full resolution (40 Hz) cloud top height obtained from the 1064 atmospheric channel. This parameter is in GLA09.
4	/Data_40HZ /Atmosphere /d_FRir_intsig	Full Resolution 1064 Integrated Signal	1/(m-sr)	0	0,001	Though called 'integrated signal' this is actually an average of all bins in the above-ground portion of the 1064 40 Hz profile with values above the threshold of 1.0e-7 (1/(m-sr) units). This parameter is extracted from the equivalent parameter on GLA09.
5	/Data_40HZ /Geophysical /d_DEM_elv	DEM Elevation	meters	-500	10000	Elevation at the footprint location from the SRTM30 (GTOPO30 + SRTM) Digital Elevation Model (DEM). The reference frame for the DEM elevation was changed to the TOPEX/Poseidon ellipsoid to make it consistent with the GLAS elevations.
6	/Data_40HZ /Geophysical /d_DEMhiresArElv	DEMhiresArElv	meters	-500	1300	d_DEMhiresArElv is a 9 element array of high resolution DEM values. The array index corresponds to the position of the DEM value relative to the spot. (5) is the footprint center.
7	/Data_40HZ /Geophysical /d_d2refTrk	Distance to the reference ground track	meters	0	1000000	Distance to the reference ground track.
8	/Data_40HZ /Geophysical /d_deltaEllip	Delta Ellipsoid	meters	-9	9	Surface elevation (T/P ellipsoid) minus surface elevation (WGS84 ellipsoid).
9	/Data_40HZ /Geophysical/d_eqElv	Equilibrium Tide Elevation	meters	-10	10	The equilibrium (long period) tide at last valid shot over the ocean.
10	/Data_40HZ /Geophysical/d_erElv	Solid Earth Tide Elevation	meters	-10	10	The solid earth tide elevation.
11	/Data_40HZ /Geophysical/d_gdHt	Geoid	meters	-200	200	The height of the geoid above the ellipsoid. EGM2008 geoid height above the reference ellipsoid.
12	/Data_40HZ /Geophysical/d_ldElv	Load Tide Elevation	meters	-10	10	The load tide elevation applied to each shot.
13	/Data_40HZ /Geophysical/d_ocElv	Ocean Tide Elevation	meters	-10	10	The ocean tide elevation from the TPX07.1 tide model.
14	/Data_40HZ /Geophysical /d_poTide	Pole Tide	meters	-10	10	Pole tide: an ocean tide which is the result of the Chandler wobble (a free nutation of the Earth caused by fluctuating pressure on the bottom of the ocean, caused by temperature and salinity changes and wind-driven changes

Figure 4.22: Coordinate Point Information Data Information window: example from a ICESat data product (GLAH14).

4.3.3 Data processor

The last tab on the ICEComb Client user interface holds the Data Processor tab, illustrated in Figure 4.24. It allows the study of geographic areas using the altimetry data from the ICESat and ICESat-2 missions and the build-in Elevation Data Processor.

In order to operate the Data Processor, the user first must select the Coordinate Points that will be part of the study area (i.e, the sample), then the tool extracts the point's relevant elevation data—that are dataset dependent—to performs statistical studies and to be applied to algorithms to remove the observations considered abnormal within the sample (i.e., the outliers), based on parameters defined by the user.

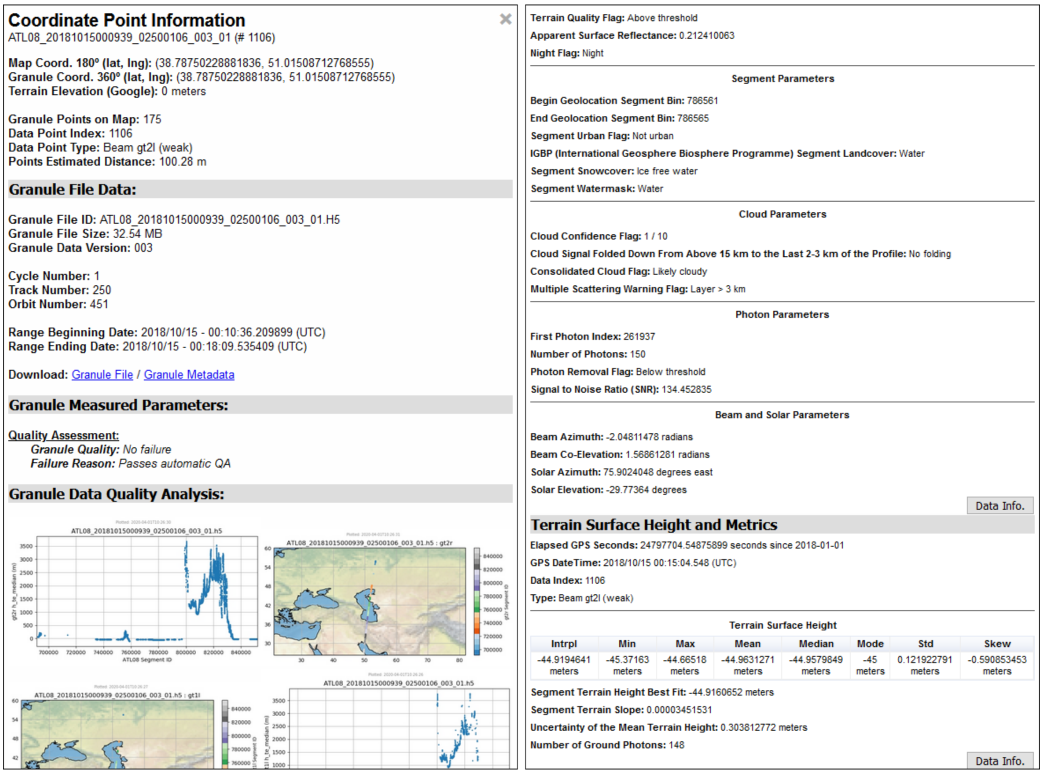


Figure 4.23: Coordinate Point Information window maximised: example from an ICESat-2 data product (ATL08).

The selection of the Coordinates Points can be done in two ways, by selecting points one by one (Single Point selection) or by indicating a Coordinates Points range (Interval selection), so the tool automatically determine the points in between. After starting the Interval selection mode, user just have to click in two points (the first and last point) of the desires track as exemplified in Figure 4.24.

Single Coordinates Points can be selected from different ground tracks (different granule files), any marker type (any data rate) and duplicated points will be automatically ignored by the Coordinates Points Selection algorithm. When selecting a interval of points, both selected points must belong to the same granule (regardless of the marker type) and the same beam type (in the case of ICESat-2 data) so all intermediate markers can be added to the selection list, even those that are not visible to the user.

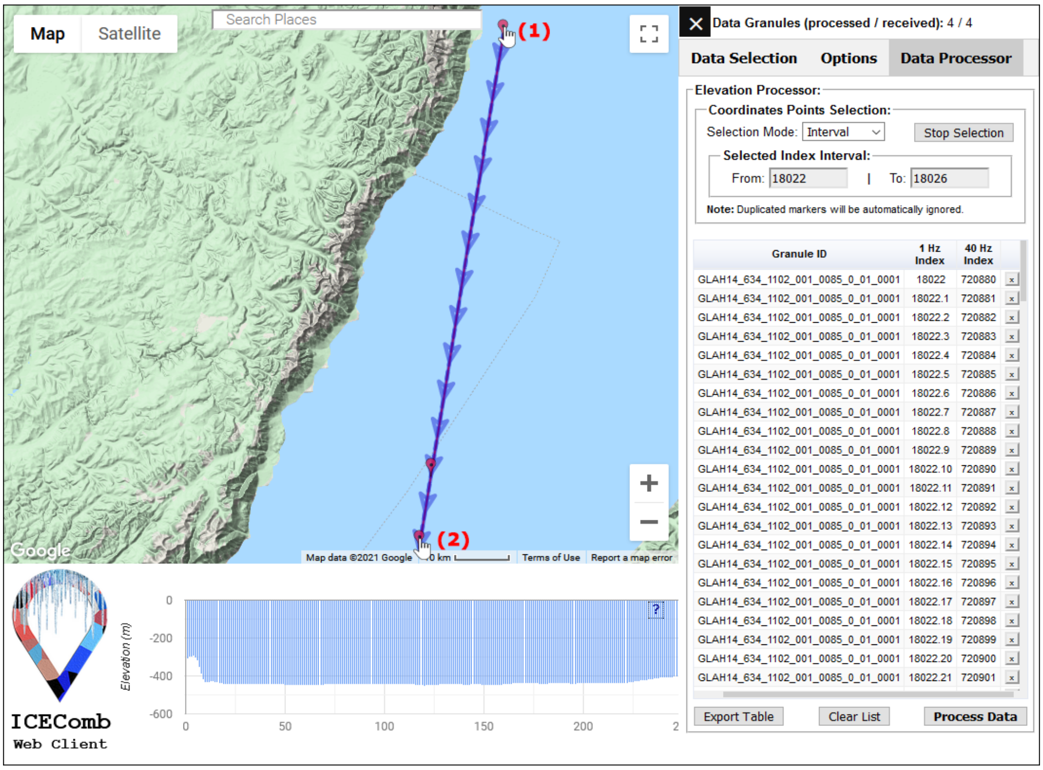


Figure 4.24: ICEComb Data Processor – Coordinates Points Selection Mechanism.

Below the "Coordinates Points Selection" section there is a table with the selected Coordinates Points, that contains information such as the Granule ID (granule file name), the data indexes (for the ICESat data products, the 1 Hz and 40 Hz indexes), and an option to remove data entry one by one (by row). This is done by clicking in the "x" button in the rightmost column of the table, or all at once by clicking in the "Clear List" button. There is also an option to export the selected points to a comma-separated values (CSV) file. After selecting the Coordinates Points, users can click on the "Process Data" button in order to view the Elevation Data Processor window.

Figure 4.25 shows the data extracted from the selection of coordinate points from a GLAH14 data product from the ICESat mission. The elevation data table content was extracted from the respective granule file except the column of the Elevation with reference to the World Geodetic System 1984 (WGS84) and the Earth Gravitational Model 2008 (EGM2008) datum, which is calculated using the following formula:

$$Elev.\,WGS84 - EGM08 = Surface\ Elev. - Geoid - Delta\ Ellipsoid. \quad (4.3)$$

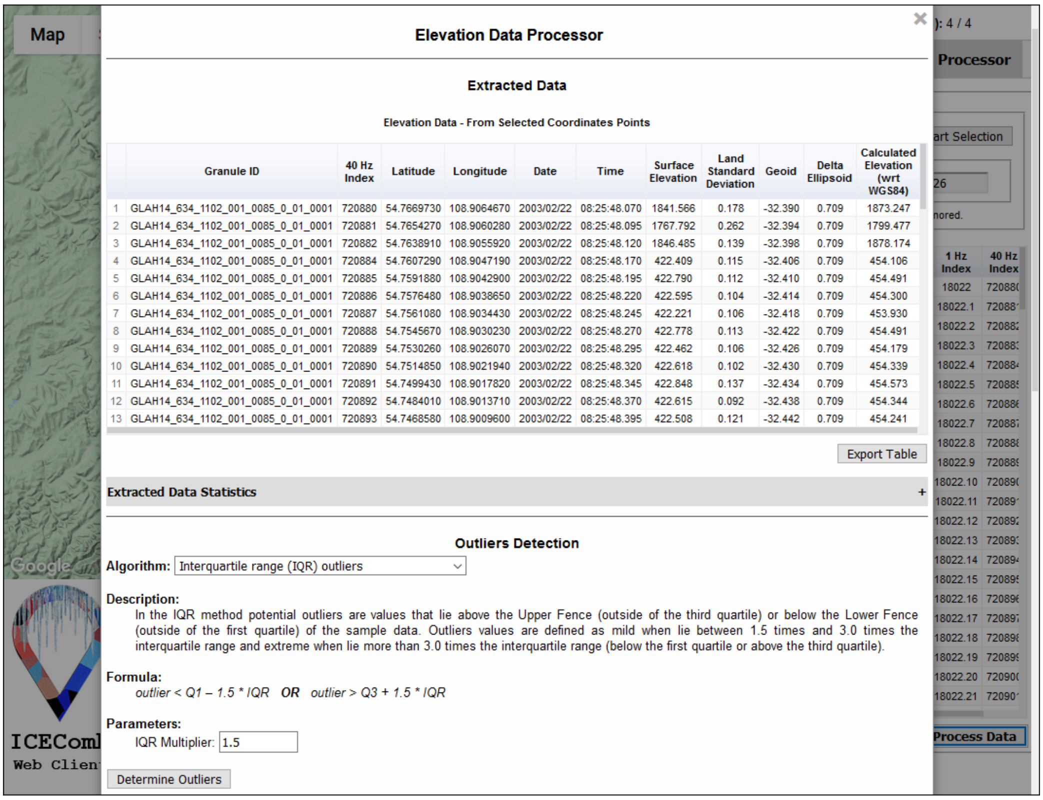
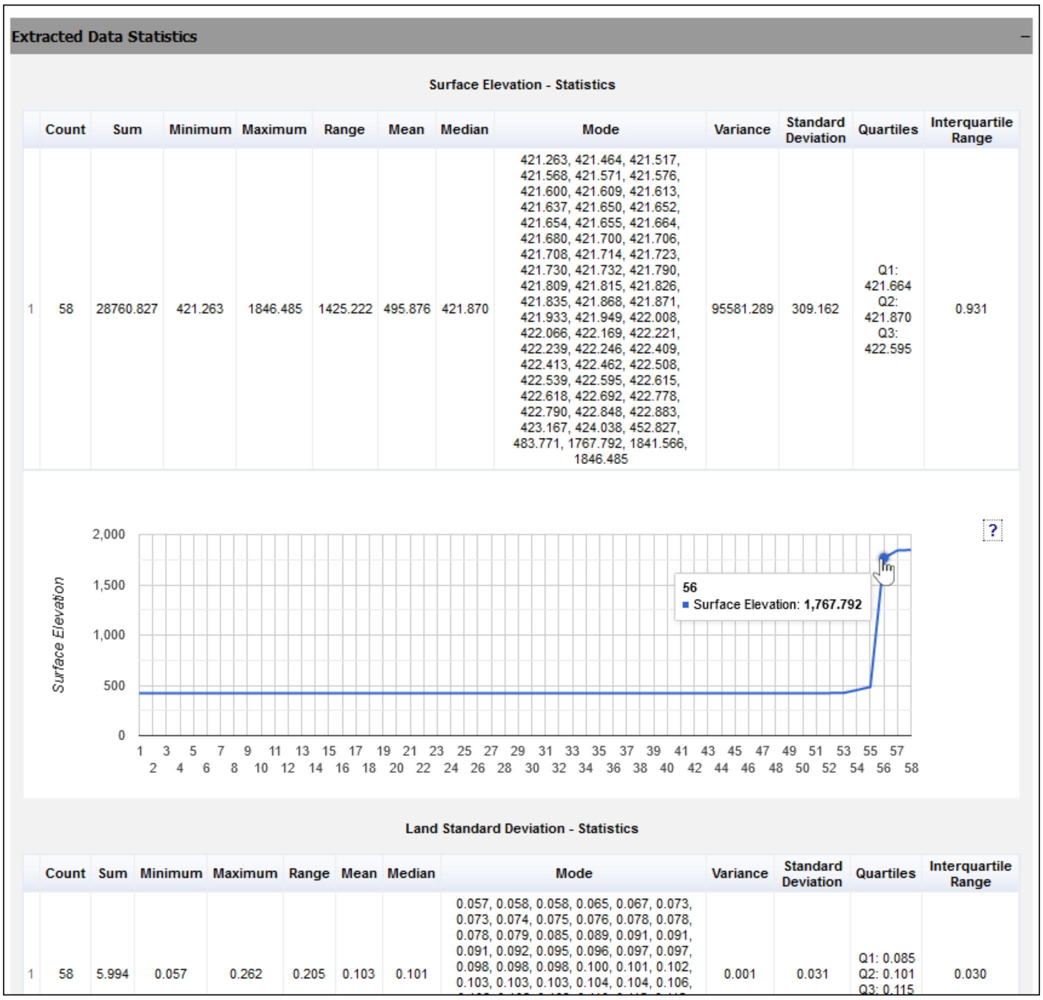


Figure 4.25: ICEComb Data Processor – Elevation Data Processor Main Window.

The extracted data is then subjected to statistical analysis where are determined elements like the count, sum, minimum, maximum, range, mean, median, mode, variance, standard deviation, quartiles and the interquartile range.

A plot for each data element is generated and accessible in the section “Extracted Data Statistics”, as exemplified in Figure 4.26. A similar section is also created to analyse the final data, after processing the outliers.

Then final step is the selection of the outlier detection algorithm and its parameters in the “Outliers Detection” section. The Data Processor have four different outliers removal algorithms implemented that users can choose from. While selecting the algorithm, users are presented with the description of the method, its formula or pseudocode and the available parameters to tweak the algorithm’s result (as in Figure 4.25). Each algorithms will be described in more details further in this section.



Extracted Data Statistics

Surface Elevation - Statistics

	Count	Sum	Minimum	Maximum	Range	Mean	Median	Mode	Variance	Standard Deviation	Quartiles	Interquartile Range
1	58	28760.827	421.263	1846.485	1425.222	495.876	421.870	421.263, 421.464, 421.517, 421.568, 421.571, 421.576, 421.600, 421.609, 421.613, 421.637, 421.650, 421.652, 421.654, 421.655, 421.664, 421.680, 421.700, 421.706, 421.708, 421.714, 421.723, 421.730, 421.732, 421.790, 421.809, 421.815, 421.826, 421.835, 421.868, 421.871, 421.933, 421.949, 422.008, 422.066, 422.169, 422.221, 422.239, 422.246, 422.409, 422.413, 422.462, 422.508, 422.539, 422.595, 422.615, 422.618, 422.692, 422.778, 422.790, 422.848, 422.883, 423.167, 424.038, 452.827, 483.771, 1767.792, 1841.566, 1846.485	95581.289	309.162	Q1: 421.664 Q2: 421.870 Q3: 422.595	0.931

Land Standard Deviation - Statistics

	Count	Sum	Minimum	Maximum	Range	Mean	Median	Mode	Variance	Standard Deviation	Quartiles	Interquartile Range
1	58	5.994	0.057	0.262	0.205	0.103	0.101	0.057, 0.058, 0.058, 0.065, 0.067, 0.073, 0.073, 0.074, 0.075, 0.076, 0.078, 0.078, 0.078, 0.079, 0.085, 0.089, 0.091, 0.091, 0.091, 0.092, 0.095, 0.096, 0.097, 0.097, 0.098, 0.098, 0.098, 0.100, 0.101, 0.102, 0.103, 0.103, 0.103, 0.104, 0.104, 0.106, [illegible]	0.001	0.031	Q1: 0.085 Q2: 0.101 Q3: 0.115	0.030

Figure 4.26: ICEComb Data Processor – Extracted Data Statistics section.

Figure 4.27 shows the “Processed Data” section with the results of the selected algorithm in two tables, one with the final data (i.e., the inliers) and another with the removed data (i.e., the outliers).

Both the Extracted Data (before processing) and the Processed Data (after outliers removal) tables can be extracted in to a CSV file.

4.3.3.1 Outliers detection algorithms

There are four different outliers detection algorithms implemented in the ICEComb tool that can be used with both the ICEsat and the ICESat-2 data. The parameters in the algorithms can be adjusted to narrow or widen the outliers filtering, based on the study characteristics or selected data.

Processed Data

Outliers Detection Algorithm: ***Interquartile range (IQR) outliers***

Elevation Data - After Removing Surface Elevation Outliers

	Granule ID	40 Hz Index	Latitude	Longitude	Date	Time	Surface Elevation	Land Standard Deviation	Geoid	Delta Ellipsoid	Calculated Elevation (wrt WGS84)
1	GLAH14_634_1102_001_0085_0_01_0001	720884	54.7607290	108.9047190	2003/02/22	08:25:48.170	422.409	0.115	-32.406	0.709	454.106
2	GLAH14_634_1102_001_0085_0_01_0001	720885	54.7591880	108.9042900	2003/02/22	08:25:48.195	422.790	0.112	-32.410	0.709	454.491
3	GLAH14_634_1102_001_0085_0_01_0001	720886	54.7576480	108.9038650	2003/02/22	08:25:48.220	422.595	0.104	-32.414	0.709	454.300
4	GLAH14_634_1102_001_0085_0_01_0001	720887	54.7561080	108.9034430	2003/02/22	08:25:48.245	422.221	0.106	-32.418	0.709	453.930
5	GLAH14_634_1102_001_0085_0_01_0001	720888	54.7545670	108.9030230	2003/02/22	08:25:48.270	422.778	0.113	-32.422	0.709	454.491
6	GLAH14_634_1102_001_0085_0_01_0001	720889	54.7530260	108.9026070	2003/02/22	08:25:48.295	422.462	0.106	-32.426	0.709	454.179
7	GLAH14_634_1102_001_0085_0_01_0001	720890	54.7514850	108.9021940	2003/02/22	08:25:48.320	422.618	0.102	-32.430	0.709	454.339
8	GLAH14_634_1102_001_0085_0_01_0001	720891	54.7499430	108.9017820	2003/02/22	08:25:48.345	422.848	0.137	-32.434	0.709	454.573
9	GLAH14_634_1102_001_0085_0_01_0001	720892	54.7484010	108.9013710	2003/02/22	08:25:48.370	422.615	0.092	-32.438	0.709	454.344
10	GLAH14_634_1102_001_0085_0_01_0001	720893	54.7468580	108.9009600	2003/02/22	08:25:48.395	422.508	0.121	-32.442	0.709	454.241
11	GLAH14_634_1102_001_0085_0_01_0001	720894	54.7453140	108.9005480	2003/02/22	08:25:48.420	422.539	0.106	-32.446	0.709	454.276
12	GLAH14_634_1102_001_0085_0_01_0001	720895	54.7437710	108.9001340	2003/02/22	08:25:48.445	422.692	0.129	-32.450	0.709	454.433
13	GLAH14_634_1102_001_0085_0_01_0001	720896	54.7422270	108.8997190	2003/02/22	08:25:48.470	423.167	0.128	-32.454	0.709	454.912

Export Table

Outliers Removed

	Granule ID	40 Hz Index	Latitude	Longitude	Date	Time	Surface Elevation	Land Standard Deviation	Geoid	Delta Ellipsoid	Calculated Elevation (wrt WGS84)
1	GLAH14_634_1102_001_0085_0_01_0001	720880	54.7669730	108.9064670	2003/02/22	08:25:48.070	1841.566	0.178	-32.390	0.709	1873.247
2	GLAH14_634_1102_001_0085_0_01_0001	720881	54.7654270	108.9060280	2003/02/22	08:25:48.095	1767.792	0.262	-32.394	0.709	1799.477
3	GLAH14_634_1102_001_0085_0_01_0001	720882	54.7638910	108.9055920	2003/02/22	08:25:48.120	1846.485	0.139	-32.398	0.709	1878.174
4	GLAH14_634_1102_001_0085_0_01_0001	720898	54.7391470	108.8988810	2003/02/22	08:25:48.520	452.827	0.125	-32.462	0.709	484.580
5	GLAH14_634_1102_001_0085_0_01_0001	720961	54.0259260	108.7034610	2003/02/22	08:26:00.095	483.771	0.098	-33.260	0.709	516.322
6	GLAH14_634_1102_001_0085_0_01_0001	721007	53.9549070	108.6842680	2003/02/22	08:26:01.245	424.038	0.143	-33.250	0.709	456.579

Processed Data Statistics +

Figure 4.27: ICEComb Data Processor – Processed Data section.

All selected algorithms were tested and are described in scientific papers in studied that used ICESat or ICESat-2 data.

Interquartile range (IQR) outliers

In the IQR method, possible outliers are values that are above the Upper Fence (outside of the third quartile) or below the Lower Fence (outside of the first quartile) of the sample data.

Outliers values are defined as mild when lie between 1.5× and 3.0× the interquartile range and extreme when lie more than 3.0 times the interquartile range (beneath the first quartile or over the third quartile).

This method was applied with success to ICESat-2 data to remove outliers before averaging the data used to determine reliable lake surface heights (LSH) over some lakes in China [37].

Algorithm Parameters: IQR Multiplier.

Equation:

$$IQR = Q3 - Q1. \tag{4.4}$$

$$\text{outliers:} \begin{cases} \text{height data} < Q1 - (1.5 \times IQR); \\ \text{height data} > Q3 + (1.5 \times IQR). \end{cases} \tag{4.5}$$

Sigma rejection criteria

The 2-sigma rejection criteria express a heuristic on which values are taken to lie between two standard deviations (σ) of the mean, while the remaining values are considered as outliers. In other words, values that are outside the mean plus or minus two times the standard deviation are considered as outliers.

This was one of the methods used to remove spurious observations in data originated in the ICESat mission over inland water surfaces [22], also to remove outliers when comparing results between radar and satellite altimetry in a study about lake and reservoir volume variations [7].

Algorithm parameters: σ Multiplier.

Equation:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}. \tag{4.6}$$

$$\text{outliers:} \begin{cases} \text{height data} < mean - (2 \times \sigma); \\ \text{height data} > mean + (2 \times \sigma). \end{cases} \tag{4.7}$$

Residual and standard deviation fence

This method excludes values (outliers) whose absolute value of the residual (deviation of the value from the mean) exceeds 2.5 times the standard deviation (StD) of the residuals. It is an iterative process that will only stop when all the residuals fulfil the premise.

This algorithm was used to determine large lake level fluctuation/trends using ICESat data [34].

Algorithm parameters: StD Multiplier.

Listing 4.2: Residual and standard deviation fence pseudocode.

```
Given:
    values - A set of observations.

Return:
    values - The values set without outilers

mean = calculate the mean of the values set
outlierFound = true // assumes there are outliers

while outlierFound = true do
    // reset the outliersFound flag
    outlierFound := false
    residuals := a set of the calculated residuals
    residualsStD := calculate the standard deviation of the residuals
    for every residual in residuals do
        if absolute residual > 2.5 * residualsStD then
            // outlier found
            outlierFound := true
            values := values without the value that originated the outlier residual
        end if
    end for
end while

return values // set without outilers
```

RANdom SAmple Consensus algorithm (RANSAC)

This learning algorithm uses random sampling of the data to classify the data elements as outliers and inliers. The RANSAC paradigm fits a line repeatedly through a random pair of points belonging to the dataset within a fit threshold, fit points ratio, and a number of interactions [11]. The model used here is the 2D line fitting, so the determined line should be estimated using only the inliers and not be contaminated by outliers.

There have been several applications of this method to remove data outliers with different error threshold (maximum deviation) parameter. In a study using ICESat data to derive elevation changes of Tibetan lakes, it was used a 15 cm error threshold [25] (0.15 in the ICEComb Data Processor), corresponding to the GLAS vertical accuracy, while in a study of water level variations in the Hulun (or Dalai) Lake, in China's autonomous region of Inner Mongolia, a 10 cm (0.10 in ICEComb) error threshold was applied [17], corresponding to the accuracy of ICESat measurements over water.

ALGORITHM PARAMETERS: Error threshold (maximum deviation); Model fit points acceptance ratio (ratio to accept the model); Number of iterations to try the model (number of loops).

Listing 4.3: RANSAC pseudocode.

```
Given:
    data - A set of observations.
    model - A model to explain observed data points.
    n - Minimum number of data points required to estimate
         model parameters.
    k - Maximum number of iterations allowed in the
         algorithm.
    t - Threshold value to determine data points that are
         fit well by model.
    d - Number of close data points required to assert that
         a model fits well to data.

```

```
Return:
    bestFit - model parameters which best fit the data (or null if no good model is found)

iterations = 0
bestFit = null
bestErr = something really large

while iterations < k do
    maybeInliers := n randomly selected values from data
    maybeModel := model parameters fitted to maybeInliers
    alsoInliers := empty set
    for every point in data not in maybeInliers do
        if point fits maybeModel with an error < t then
            add point to alsoInliers
        end if
    end for
    if the number of elements in alsoInliers is > d then
        // This implies that we may have found a good model
        // now test how good it is.
        betterModel := model parameters fitted to all points in maybeInliers and alsoInliers
        thisErr := a measure of how well betterModel fits these points
        if thisErr < bestErr then
            bestFit := betterModel
            bestErr := thisErr
        end if
    end if
    increment iterations
end while

return bestFit
```

Credits: Wikipedia.

Figure 4.28 illustrates the application of the ICEComb Data Processor to ICESat water level observations over Lake Mai-Ndombe, a large and shallow freshwater lake located within the Ngiri-Tumba-Maindombe area, one of the largest Ramsar wetlands of international importance, situated in the Cuvette Centrale region of the Congo Basin.

The RANSAC algorithm was used to discard outliers by applying a 2D line fitting model with a 15 cm threshold, equivalent to the ICESat's GLAS vertical accuracy.

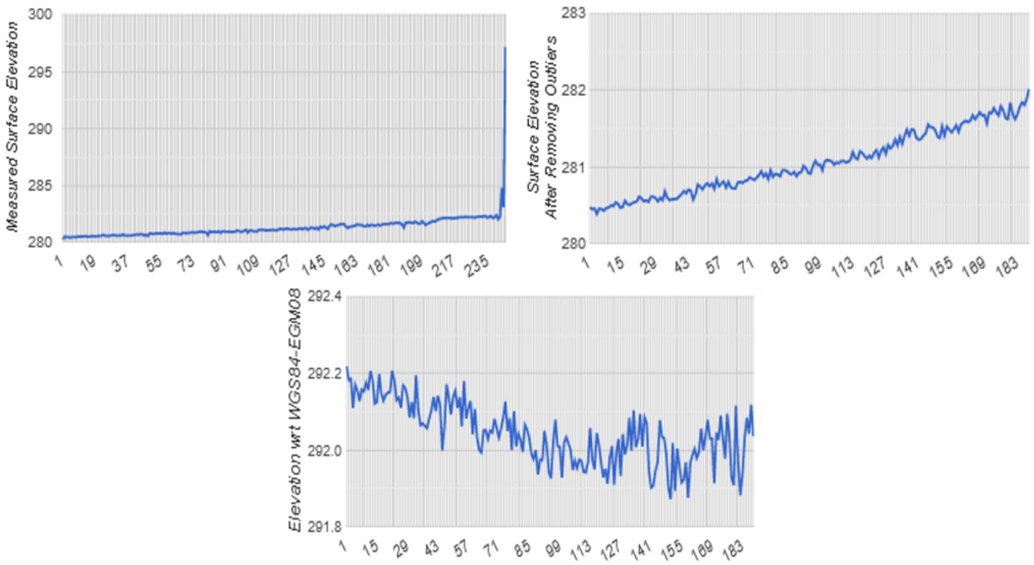


Figure 4.28: GLAH14 elevation data plots from the ICESat ground track 275 (10/27/2004) over Lake Mai-Ndombe. Extracted surface elevation (top left), filtered surface elevation after outliers removal (top right), and calculated surface elevation based on EGM-2008 and WGS84 datum (bottom centre).

4.4 Performance analysis

The characteristics and the amount of data to be handled by the ICEComb tool is one the most important factors to determine the tool efficiency.

In order to measure the tool performance, two deployment scenarios were evaluated. In the first scenario, the server and client applications are running on the same system (Table 4.6), and in the second one, the server and client applications are running in separated systems (Table 4.7). Both test systems were equipped with a 4-core, 8-thread CPUs, 16 GB of RAM, SSD storage, Windows 10 operating system, and Mozilla Firefox as the web browser.

The selected data chunk was composed by 9 228 data products (9 228 .H5 granules and 9 228 .XML files) from the ICESat / GLAH01 dataset, which amounted to approximately 136 GB of total data.

Table 4.6: Performance when running the server and client applications in the same system

Actions	Server Application		Client Application
	RAM	CPU Usage	Load time (s)
Init. + 100 files	39.70 MB	3.57%	2.77
9 228 files 1st run	41.20 MB	7.24%	69.72
9 228 files 2nd run	42.10 MB	12.03%	37.18
9 228 files 3rd run	42.90 MB	11.98%	36.82

Table 4.7: Performance when running the server and client applications in separate systems

Actions	Server Application		Client Application
	RAM	CPU Usage	Load time (s)
Init. + 100 files	39.60 MB	3.28%	1.83
9 228 files 1st run	39.10 MB	6.63%	59.17
9 228 files 2nd run	39.48 MB	12.42%	26.72
9 228 files 3rd run	41.16 MB	12.44%	26.57

Taking into account the obtained results, the main bottleneck for processing data in the ICEComb tool is the disk IOPS (input/output operations per second) and not the CPU cycles, indicating that the storage technology plays a major role on performance. In addition, data demonstrated that the server application is taking advantage of the Windows caching system and offers, on average, about 39% of performance after the first run of the test, which indicates that the amount of available memory is also important.

The performance analysis carried out shown that the ICEComb tool can process 9 228 files from the ICESat / GLAH01 dataset in about 37 seconds when running the client and server in the same system. This result clearly confirms that the optimisations implemented in the server and client applications were very effective and that the developed tool can be used in a single system with a very adequate performance.

The optimal scenario for running the ICEComb tool is to have the client and server applications in separated systems. In this case, the second run of the 9 228 files took about 26 seconds, an improvement of about 39% when comparing with both applications running on the same system. Overall, running the ICEComb tool server and client on separated systems offers 36.73% additional performance in data load time on average (see Table 4.8).

Table 4.8: Comparison between running the server and client applications in separate systems and running in the same system

Actions	Server Application		Client Application
	RAM	CPU Usage	Load time
Init. + 100 files	+0.25%	+8.84%	+51.37%
9 228 files 1st run	+5.37%	+9.20%	+17.83%
9 228 files 2nd run	+6.64%	-3.14%	+39.15%
9 228 files 3rd run	+4.23%	-3.70%	+38.58%
Average	+4.12%	+2.80%	+36.73%

4.5 Discussion

The results presented previously clearly outlined that the benefits of having a tool to visually navigate the ICESat and ICESat-2 data products quickly and efficiently is an invaluable resource for the scientific community. There are few examples of currently available tools that use a map to interface with data from the ICESat and ICESat-2 missions, such as the NASA Earthdata Search, NASA's WorldView and the OpenAltimetry project [14, 21].

The NASA Earthdata Search is a web-based client for managing and distribute data from multiple Earth-observing satellites missions. Therefore, its main focus is on advanced data products filtering and sub-setting, on which the main objective is the data granules' download, while using a geographical map as a reference. NASA's WorldView is a tool to allow end-users access to view and download NASA remote sensing imagery [20]. Both these NASA tools only offers access to the spatial coverage of the ICESat and ICESat-2 missions but not to data values presentation.

In contrast, the OpenAltimetry project was created to specifically allow the visualisation of altimetry data by its geographical location [14], from the ICESat mission initially and later added support for ICESat-2, on a map-based interface using the OpenLayers mapping library.

But both the ICESat and ICESat-2 datasets contain more than altimetry-specific data. Data from atmosphere segments, cloud layers, backscatter profiles, aerosols, elevation flags, reflectivity information, optical depth, waveform parameters, planetary boundary layer and other geophysical parameters are some examples of data that are not possible to be viewed on the OpenAltimetry platform.

That is why there is still a need for a much broader tool for visualising ICESat and ICESat-2 data, and this is where the ICEComb tool fits in. The primarily objective of ICEComb it to offer the access to the whole ICESat and ICESat-2 data products while facilitating the satellite data location and visualisation, providing additionally an instant data analysis functionality through its data processor.

To improve the data presentation capabilities of the ICESat and ICESat-2 missions, the ICEComb tool uses a combination of text, tables, graphs and data converters to enable a clear interpretation of data product contents. This lowers the entry barrier for new users for this type of data while offering analytical tools for advanced data users.

The ability to view, process and analyse satellite laser altimetry data in great detail are features that no other tool currently offers.

Although the ICEComb tool uses a web-based architecture, the intention is that it be mainly used locally in academic and research centres, rather than in the public Internet (even though it can be perfectly used on it). This approach allowed to reduce the development complexity of some parts of the ICEComb tool, as well its deployment, by providing a simple "one click" ready-to-use solution. Furthermore, this also facilitates the expansion of the tool by reducing the complexity of developing new features, offering at the same time the advantages of a web system, such as the distribution of the workload between a server and a client.

Chapter 5

Conclusions and future work

5.1 Conclusions

Both ICESat and ICESat-2 missions provided many important insights into the value of satellite laser altimetry and demonstrated to be a very useful tool for examining and understanding changes on the Earth's surface. They enabled researchers and scientists to quantify the seasonal and annual contributions to the sea level rise by the ice sheets reductions, the pace of mass balance changes, the size of the global biomass, and other measurements, with unprecedented accuracy and in a global scale.

Handling the substantial data volume generated by satellite missions is challenging. Moreover, the fact that data is directly coupled to the gathered location information, by having a visual reference of the satellite track and the possibility to evaluate the characteristics of that location, is a fundamental feature to comprehend and develop new scientific models.

To produce a viable solution that offers both ease of use and performance, requires the development of tools that are based on well-known and user-friendly platforms as well as the creation of processes based on programming models that accommodate some level of parallelism.

The need to create a new tool for the processing and visualisation of ICESat and ICESat-2 data products derived primarily from the fact that the existing solutions only offer access to a limited amount of information contained in that datasets.

The ICEComb software tool allows users to browse the raw data contained in the satellite data products while representing it using visual aids such as graphs and tables in a rich geospatial data environment, facilitating data interpretation and discoverability.

The feature set offered by the mapping tool used to represent the world map play an important role in the interpretation of the satellite data, since the characterisation of the geographic location of the data collection points is often vital in many research areas.

ICEComb is intended to be mainly used by scientists and researchers to help their work when using the ICESat and ICESat-2 data. It offers a ready-to-use system to quickly access the raw gathered data in a visually engaging fashion, without the requirement to have previous knowledge of the format, parameters or structure of the data products.

The built-in Elevation Data Processor is a valuable scientific tool, as it allows the immediate study of geographic areas using altimetry data, without the need to use external tools, while having access to implemented data filtering methods proven by scientific studies.

The design and the implementation of the ICEComb tool were created using standardised languages and open source software that facilitate the possibility of its expansion, allowing the integration of new models that coupled with the ICESat and ICESat-2 data allow to enrich the context of studies developed using this type of data.

In addition, the fact that the ICEComb solution has only two main elements—a server program and a web client—means that the tool can almost be seen as a standalone application.

This allows the ICEComb tool to be easily deployed to a single computer, for quick access to data from ICESat and ICESat-2 missions, without the need to allocate high-performance server hardware, that may not always be available. Although the data storage required for all ICESat and ICESat-2 data products is quite large.

5.2 Main contributions

This thesis offers an innovative tool that is able to greatly reduce the complexity of interpreting satellite altimetry data in the scientific community. The tool developed was designed to be flexible, open source, and easy to use and deploy, making it unique when comparing with available similar tools.

In the field of data visualisation, the principal contributions of the work developed in the scope of this thesis are:

- A tool with a rich user interface where the map has high detailed satellite imagery, geographical map view with elevation contour lines, improved map places, country borders contours, labels and road information with the option of search and jump to a specific location.

- A ground tracks representation system based on the collection locations rather than on the estimated satellite path, with the ability to profile the elevation along-the-track and indication of the satellite direction.

- A dynamic location marker icon system based on the type of the available data and colour coded to the Ground track and in turn, to the data product.

With regard to data analysis, the main contributions of this thesis work are indicate below:

- Meaningful data presentation including data identifiers, values and units, as well as the use of more complex structures to organise data such as tables, interactive charts and the implementation of data conversions and translations mechanisms.

- The ability to access to the predetermined, structured, computer-parseable and inventory-level metadata analysis available in each granule file, as well the possibility to filter data products by its quality assessment.

- The opportunity to access information about the raw internal data paths, form the granule files, including metadata elements that help to characterise the presented data in more detail. This functionality has the additional advantage to help the development of new tools that use raw data as an input by facilitating the understanding of the internal structure of data products.

Finally, in the domain of data processing, the main contributions can be summarised as follows:

- The data points selection mechanism for data processing is a versatile, fast and precise way of effectively select the study area, by selecting the data points directly from the map. The data points can be added or removed freely from any granule file, while sill being possible to view their contents.

- Access in a single tool to several data processing algorithms that were described in scientific papers an tested using ICESat or ICESat-2 data. The data processor also offers the possibility to tweak the parameters of the algorithms and a set of statistical analysis to aid with the data interpretation and algorithm selection.

5.3 Future work

Although the ICEComb tool fulfils the requirements to act as capable visual tool to explore and process the ICESat and ICESat-2 data, scientific data can always be manipulated, analysed and viewed in numerous ways.

The fact that the technologies used to develop the ICEComb tool are scalable and well documented opens the possibility to effortlessly expand the scope of the application and offer new functionalities such as:

- At the moment granules data are presented individually, but relations between data of different data products do exist. Linking data from different data products would allow to complement information about a particular area in more detail, without the need to go through the different datasets.

- An option to aggregate data from both satellite missions, allowing to mix ICESat and ICESat-2 data in a single view, providing access to altimetry data from a much wider timeline.

- A function to create a subset and automatically extract data from data products in an area of the map. The extracted data would be useful, for example, to be applied in to an external processing tool, without the need for any additional filtering to be implemented.
- Perform adjustments to the tool's performance, namely a further improvement in the execution offload between the client and the server, and other algorithms optimisation.
- Aggregate data from other data sources, outside of satellite altimetry, allowing the simultaneous access to multiple sources of information that describe a common phenomenon (data mashup). The Google Elevation Service is an example of an external tool already integrated in the ICEComb tool.

Finally, as a last suggestion, the tool developed could be extended to include support to other satellite altimetry missions. For example, data from satellite radar altimetry missions, such as Envisat, SARAL/AltiKa and Sentinel-3, could be easily integrated in the ICEComb tool.

References

[1] Abdalati, W., Zwally, H., Bindschadler, R., Csatho, B., Farrell, S., Fricker, H., Harding, D., Kwok, R., Lefsky, M., Markus, T., Marshak, A., Neumann, T., Palm, S., Schutz, B., Smith, B., Spinhirne, J., and Webb, C. The ICESat-2 laser altimetry mission. *Proceedings of the IEEE 98*, 5 (2010), 735–751.

[2] Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., Yi, D., and Liiva, P. Geoscience Laser Altimeter System (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters 32*, 21 (2005), L21S02.

[3] Abshire, J. B., Sun, X., Riris, H., Sirota, M., McGarry, J., Palm, S., Ketchum, E. A., and Follas, R. B. Geoscience Laser Altimeter System (GLAS) on the ICESat mission: Pre-launch and on-orbit measurement performance. In *2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)* (Baltimore, MD, 2003), IEEE, pp. 1534–1536.

[4] Bae, S., Smith, N., and Schutz, B. E. The GLAS Algorithm Theoretical Basis Document for Precision Attitude Determination (PAD). Technical Memorandum NASA/TM-2013-208641/Vol 10, NASA Goddard Space Flight Center, Greenbelt, MD, March 2013.

[5] Baghdadi, N., Lemarquand, N., Abdallah, H., and Bailly, J. S. The relevance of GLAS/ICESat elevation data for the monitoring of river networks. *Remote Sensing 3*, 4 (2011), 708–720.

[6] Brenner, A. C., P.DiMarzio, J., and Zwally, H. J. Precision and accuracy of satellite radar and laser altimeter data over the continental

ice sheets. *IEEE Transactions on Geoscience and Remote Sensing 45*, 2 (2007), 321–331.

[7] Carabajal, C. C., and Boy, J.-P. Lake and reservoir volume variations in South America from radar altimetry, ICESat laser altimetry, and GRACE time-variable gravity. *Advances in Space Research 68*, 2 (2021), 652–671.

[8] Chelton, D. B., Ries, J. C., Haines, B. J., Fu, L.-L., and Callahan, P. S. Satellite altimetry. In *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave, Eds. Academic Press, San Diego, CA, 2001, pp. 1–131.

[9] Dincer, A., and Uraz, B. *Google Maps JavaScript API Cookbook*. Packt Publishing, Birmingham, UK, 2013.

[10] Duerr, R. E., Weaver, R. L., and Kaminski, M. Data acceptance procedures and levels of service at the National Snow and Ice Data Center. In *2010 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2010, July 25–30, 2010, Honolulu, HI* (Honolulu, HI, 2010), IEEE, pp. 2322–2325.

[11] Fischler, M., and Bolles, R. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395.

[12] Gallaher, D., Weaver, R. L. S., Stroeve, J., and Swick, R. Data access tools – Filling the usability gap in cryosphere data. In *2008 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008, July 6–11, 2008, Boston, MA* (Boston, MA, 2008), vol. 4, IEEE, pp. IV–287–IV–290.

[13] Gong, P., Li, Z., Huang, H., Sun, G., and Wang, L. ICESat GLAS data for urban environment monitoring. *IEEE Transactions on Geoscience and Remote Sensing 49*, 3 (2011), 1158–1172.

[14] Khalsa, S. J. S., Borsa, A., Nandigam, V., Phan, M., Lin, K., Crosby, C., Fricker, H., Baru, C., and Lopez, L.

OpenAltimetry−rapid analysis and visualization of spaceborne altimeter data. *Earth Science Informatics* (Sep 2020).

[15] Kirk, A. *Data Visualization: A Successful Design Process.* Packt Publishing, Birmingham, UK, 2012.

[16] Leigh, H. W., Magruder, L. A., Carabajal, C. C., Saba, J. L., and McGarry, J. F. Development of onboard digital elevation and relief databases for ICESat-2. *IEEE Transactions on Geoscience and Remote Sensing 53*, 4 (2015), 2011–2020.

[17] Li, C., Wang, J., Hu, R., Yin, S., Bao, Y., and Li, Y. ICESat/GLAS-derived changes in the water level of Hulun Lake, Inner Mongolia, from 2003 to 2009. *Frontiers of Earth Science 12*, 2 (2018), 420–430.

[18] Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., and Zwally, H. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment 190* (2017), 260–273.

[19] Martino, A. J., Bock, M. R., Gosmeyer, C., Field, C., Neumann, T. A., III, D. W. H., III, R. L. J., Dabney, P. W., Webb, C. E., and Lee, J. ICE, CLOUD, and Land Elevation Satellite (ICESat-2) Project: Algorithm Theoretical Basis Document (ATBD) for ATL02 (Level-1B) Data Product Processing. ICESat-2-SIPS-SPEC-0150, NASA Goddard Space Flight Center, Greenbelt, MD, March 2020.

[20] Murphy, K. J., Davies, D. K., Michael, K., Justice, C. O., Schmaltz, J. E., Boller, R., McLemore, B. D., Ding, F., Vollmer, B., and Wong, M. M. LANCE, NASA's Land, Atmosphere Near real-time Capability for EOS. In *Time-Sensitive Remote Sensing*, C. D. Lippitt, D. A. Stow, and L. L. Coulter, Eds. Springer, New York, 2015, pp. 113–127.

[21] Nandigam, V., Lin, K., Phan, M., Borsa, A., Khalsa, S. J. S., and Crosby, C. J. Rapid access and visualization of spaceborne altimetry data from ICESat and ICESat-2. In *2018 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018, July 22–27, 2018, Valencia, Spain* (2018), IEEE, pp. 5270–5273.

[22] O'Loughlin, F. E., Neal, J., Yamazaki, D., and Bates, P. D. ICESat-derived inland water surface spot heights. *Water Resources Research 52*, 4 (2016), 3276–3284.

[23] Parrish, C. E., Magruder, L. A., Neuenschwander, A. L., Forfinski-Sarkozi, N., Alonzo, M., and Jasinski, M. Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance. *Remote Sensing 11*, 14 (2019), 1634.

[24] Passaia, O. A., and de Paiva, R. C. D. Processamento de dados do ICESat para aplicações hidrológicas. In *Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto, XVIII SBSR, 28–31 May 2017, Santos, SP, Brazil* (2017), INPE, pp. 934–940.

[25] Phan, V. H., Lindenbergh, R., and Menenti, M. ICESat derived elevation changes of tibetan lakes between 2003 and 2009. *International Journal of Applied Earth Observation and Geoinformation 17* (2012), 12–22.

[26] Saternos, C. *Client–Server Web Apps with JavaScript and Java.* O'Reilly, Sebastopol, CA, 2014.

[27] Schutz, B. E., Bae, S., Smith, N., and Sirota, M. Precision orbit and attitude determination for ICESat. *Advances in Astronautical Sciences 132* (2008), 775–791.

[28] Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P. Overview of the ICESat mission. *Geophysical Research Letters 32*, 21 (2005), L21S01.

[29] Smith, N., Fowell, R., Bae, S., and Schutz, B. Improved star tracker instrument magnitude prediction from ICESat flight telemetry. *Advances in Astronautical Sciences 141* (2011), 639–654.

[30] Spinhirne, J. D., Palm, S. P., Hart, W. D., Hlavka, D. L., and Welton, E. J. Cloud and aerosol measurements from GLAS: Overview and initial results. *Geophysical Research Letters 32*, 22 (2005), L22S03.

[31] Subramanian, H., and Raj, P. *Hands-On RESTful API Design Patterns and Best Practices*. Packt Publishing, Birmingham, UK, 2019.

[32] Sun, X., Jester, P. L., Abshire, J. B., and Chang, E. S. Performance of the GLAS space lidar receiver through its seven-year space mission. In *CLEO:2011 – Laser Applications to Photonic Applications* (2011), Optical Society of America. Washington, D.C., paper ATuA2.

[33] Wang, X., Cheng, X., Gong, P., Huang, H., Li, Z., and Li, X. Earth science applications of ICESat/GLAS: a review. *International Journal of Remote Sensing 32*, 23 (2011), 8837–8864.

[34] Wang, X., Gong, P., Zhao, Y., Xu, Y., Cheng, X., Niu, Z., Luo, Z., Huang, H., Sun, F., and Li, X. Water-level changes in China's large lakes determined from ICESat/GLAS data. *Remote Sensing of Environment 132* (2013), 131–144.

[35] Ye, Z., Liu, H., Chen, Y., Shu, S., Wu, Q., and Wang, S. Analysis of water level variation of lakes and reservoirs in Xinjiang, China using ICESat laser altimetry data (2003–2009). *PLOS ONE 12*, 9 (2017), e0183800.

[36] Yu, A., Stephen, M., Li, S., Shaw, G., Seas, A., Dowdye, E., Troupaki, E., Liiva, P., Poulis, D., and Mascetti, K. Space laser transmitter development for ICESat-2 mission. *Proceedings of SPIE 7578* (Feb 2010).

[37] Yuan, C., Gong, P., and Bai, Y. Performance assessment of ICESat-2 laser altimeter data for water-level measurement over lakes and reservoirs in China. *Remote Sensing 12*, 5 (2020), 770.

[38] Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock,

D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., and Thomas, R. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics 34*, 3 (2002), 405–445.