Living Planet Fellowship

BI-MONTHLY REPORT

Investigating Greenland's ice marginal lakes under a changing climate (GrIML)

Penelope How

Geological Survey of Denmark and Greenland (GEUS)

Prepared by	**Penelope How**
Approved by	**Signe Bech Andersen & Andreas Ahlstrøm**
Reference	**BR_GrIML**
Date of Issue	**2022-03-06**

1 OBJECTIVES OF WORK

GrIML officially started in January 2022, and thus the first two months of operations are reported in this bi-monthly report. The over-arching aim of GrIML is to examine ice marginal lake changes across Greenland using a multi-method and multi-sensor remote sensing approach, refined with in situ validation; with the following project objectives:

1. Integrate pre-existing SAR, optical and DEM lake detection methods (developed during the ESA CCI project, How et al., 2021) into a unified and open-source processing chain that can be implemented in cloud processing platforms
2. Examine changes in ice marginal lake abundance and extent across Greenland through the production of pan-Greenland inventories for selected years in the satellite era
3. Form detailed time-series of the dynamics of selected ice marginal lakes (using the finest temporal resolution possible), and examine changes in GLOF events over time
4. Evaluate the remotely-sensed lake extents against in situ observations

Given that the project kick-off date (KO) differs from the date proposed in the GrIML project outline, a revised timeline is provided in Fig. 1, which is adapted from the Gantt chart presented in the Living Fellowship application.

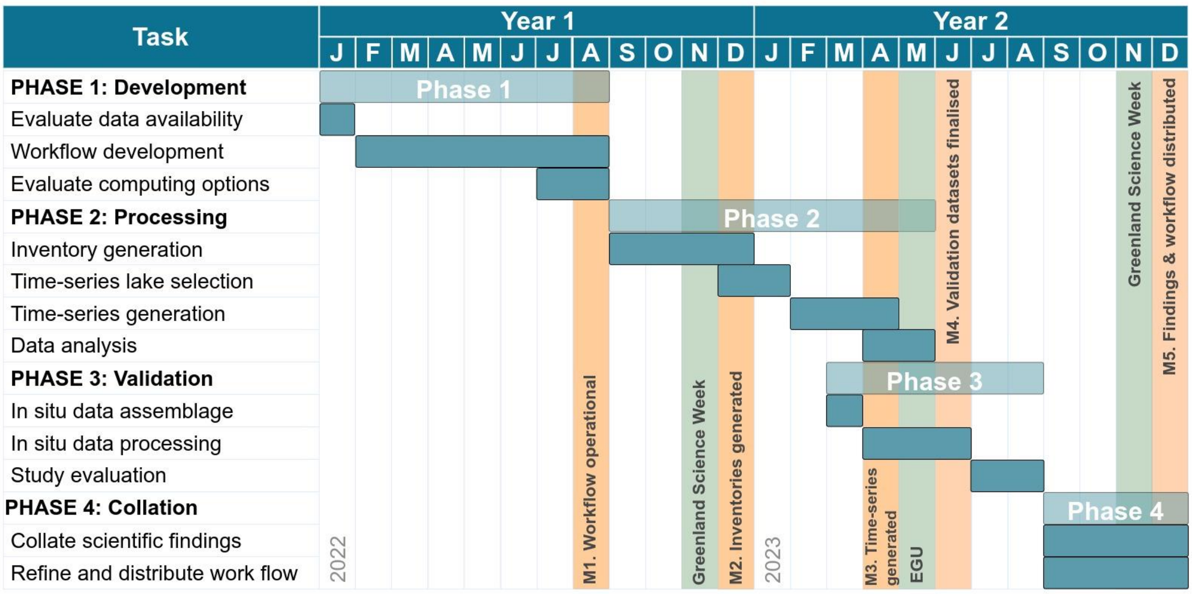

Figure 1. Revised Gantt chart showing the work plan for GrIML.

GrIML is currently in Phase 1, focusing on workflow development (as shown in Figure 1). The objectives of this phase are as follows:

a. Evaluate processing platforms/packages for workflow processing, including the availability of suitable satellite image data
b. Migrate and adapt the ice marginal lake workflow from How et al. (2021)
c. Assess options for cloud computing integration into the revised workflow

The end of Phase 1 (KO+8) culminates in the finalization of the workflow, with the first milestone (M1) being the operational workflow.

2 WORK PERFORMED

2.1 Methods

The original ice marginal lake workflow was developed as part of the ESA Glaciers CCI, under the Option 6 project 'An Inventory of Ice-Marginal Lakes in Greenland' The resulting 2017 inventory from this project was presented in How et al. (2021), hosted in the CEDA archive (Wiesmann et al., 2021). The associated processing approach consisted of three independent workflows for identifying water bodies, as summarised in Table I.

Table I. The three workflows adopted for the ESA Glaciers CCI 2017 ice marginal lake inventory

Workflow	Data source	Processing software
SAR backscatter intensity classification	Sentinel-1 GRD (Ground Range Detected) imagery	GAMMA software (ANSI)
Multi-spectral indices optical classification	Sentinel-2 TOA (Top-Of-Atmosphere) L1C imagery	ArcPy (Python), gdal (Python)
Hydrological sink basin detection	ArcticDEM 10 m mosaic	ArcPy (Python)

The purpose of the workflow migration in GrIML Phase 1 is to unify the three workflows presented in Table I. It is intended to facilitate this unification using Python (version 3.8 and above). By doing so, the new workflow has the potential to be a self-validating methodology that eases the work load for heavy data downloads, data compiling, and post-processing steps (such as metadata generation). Operational monitoring capabilities could also be a possibility under such a workflow.

The original workflows were heavily reliant on licensed software, so one of the focuses of this unification is to source alternative open-source software/packages. It is intended that the workflow and datasets produced from this will adhere to the FAIR (Findability, Accessibility, Interoperability and Reusability) Data Principles (Williamson et al., 2019). Version control is also essential for the new workflow in order to demonstrate transparency in the workflow development process, and encourage its development beyond the timeframe of the Living Planet Fellowship.

The following criteria were used to evaluate suitable processing software and packages for the new workflow:

a. Open-source
b. Long-standing, community-driven development (to limit the risk of dependencies becoming unsupported in the future)
c. Python API (Application Programming Interface) possibilities
d. Little or no reliance on external software (to ensure easy set-up)
e. Satellite image processing capabilities
f. Satellite image data availability (in the case of cloud computing options)
g. Avoidance of heavy downloads and storage of big datasets
h. Parallelization capabilities (for fast processing and data handling, if needed)

2.2 Data

Selected datasets were used for testing the capabilities of candidate software and packages. Several bounding boxes of assorted sizes were used to test the capabilities of large-scale processing, ranging from individual lake catchments to regional AOIs over the SW margin of the Greenland Ice Sheet. Tests for conducted using datasets acquired between 1 July and 31 August 2017, in order to overlap with the ESA Glaciers CCI 2017 ice marginal lake inventory and provide validation.

The GIMP (Greenland Mapping Project) ice and ocean mask was used to test filtering capabilities for removing ocean and on-ice water bodies (Howat et al., 2014). Ocean and on-ice areas were initially masked out using with the GEE copy of the GIMP ice and ocean mask (https://developers.google.com/earth-engine/datasets/catalog/OSU_GIMP_2000_ICE_OCEAN_MASK) to save computing time. The modified GIMP ice margin used in How et al. (2021) was also adopted to test the capabilities of ice marginal lake classification based on ice margin proximity.

2.3 Results

2.3.1 Migration of the workflow

The original workflow from the ESA Glaciers CCI has been migrated to a new unified workflow, hereafter referred to as the GrIML workflow. The initial focus is to develop a similar version to the original workflow for comparison and validation. The following packages were selected based on the described criteria in Section 2.1.:

a. Google Earth Engine (GEE) Python API
b. Shapely
c. Geopandas

GEE was selected based on its cloud computing capabilities, along with its compatibility to the Sentinel-1 and Sentinel-2 image collections, and the ArcticDEM mosaic dataset. It is already established as a readily used platform for glacial lake classification (including ice marginal lakes) in regional studies (e.g. Zhang et al., 2018;

Rick et al., 2022) and global inventories (e.g. Shugar et al., 2020). In addition, a particular draw to GEE is its pre-processed Sentinel-1 backscatter image collection, eliminating the need for backscatter processing (which is typically performed through external software such as the ESA SNAP toolbox).

GEE is operated through its Python API, rather than the Javascript portal (at https://code.earthengine.google.com), in order to unify the GrIML workflow under a Python framework, and reduce the reliance on third party portals and data holdings (such as Google Drive).

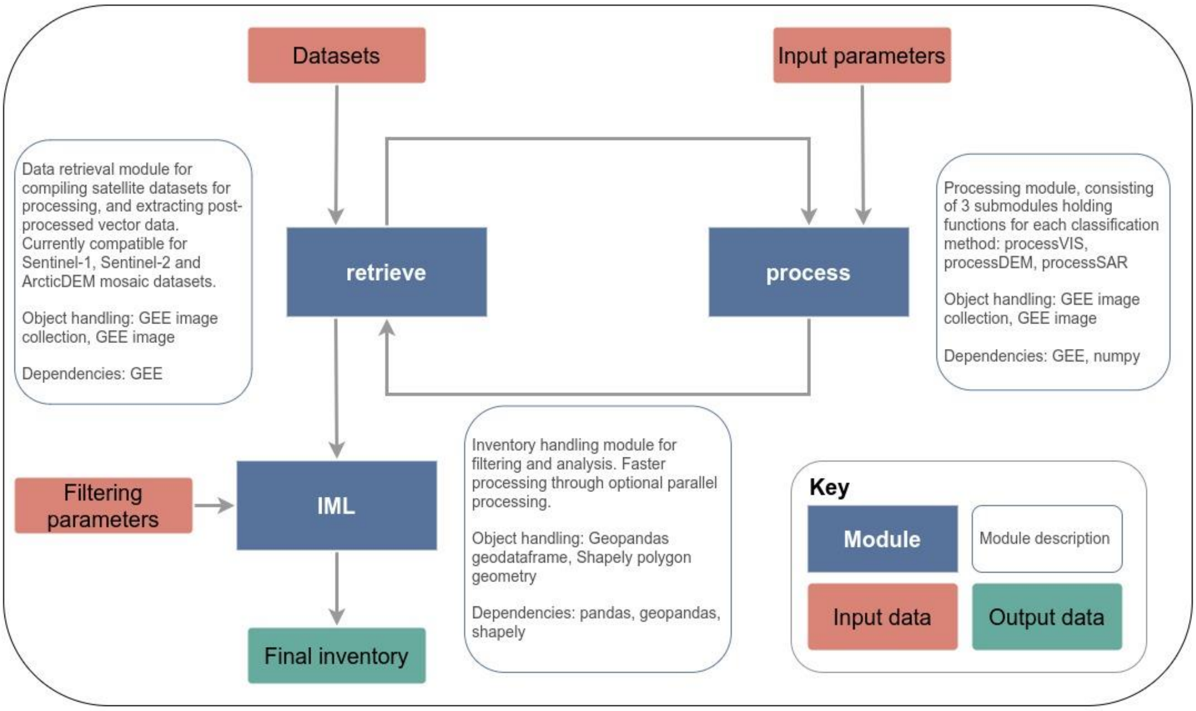

Figure 2. The current GrIML workflow modules

The GEE Python API is used in the GrIML workflow for retrieving the required satellite imagery, and performing basic processes that are necessary before classification, such as acquisition filtering, computing average pixel values, image masking, and smoothing (Fig. 2). Subsequently, the GEE Python API is used to perform the SAR imagery, optical imagery, and DEM classification methods, under the processing sub-modules outlined in Table II. The classified water bodies are then exported to a local version, from which post-processing steps can be carried out.

Table II. GrIML process sub-modules

	process.dem	**process.sar**	**process.vis**
Description	DEM sink detection classification	SAR backscatter water body classification	Multi-spectral indices optical classification
Current functionality	1 Elevation and slope retrieval 2 Topographic sink dilation 3 Sink filling	1 Image mosaicking 2 Speckle filtering 3 Backscatter thresholding	1 Cloud masking 2 Pixel resampling 3 NDWI, MNDWI, AWEIsh, AWEInsh, and BRIGHT index generation 4 Classification decision tree
Future development	Hydrological analysis	-	Topographic shadow masking

The local version of the GEE-derived classifications are constructed into geometries from the outputted vector coordinates, using the Shapely and Geopandas Python packages (Fig. 2). These packages were chosen given that they are open-source, widely used for vector dataset processing, and with the option to develop with parallel processing capabilities to handle large datasets. Shapely is used to initially construct polygon geometries from the coordinates, which are then used to populate a Geopandas Geodataframe. Metadata population, filtering, and aggregation are performed using the functionality of the Geodataframe; of which the final inventory output can be readily exported as a shapefile or spreadsheet.

The GrIML workflow is hosted as a GitHub repository, which offers version controlling through Git, promotes collaboration with other users, and can be a registered as a persistent entity with a DOI identifier. A basic skeleton copy of this is provided in the public project repository at https://github.com/PennyHow/GrIML, whilst the fully formed workflow will be hosted privately until the final outputs of the project are released.

2.3.2 Current workflow performance

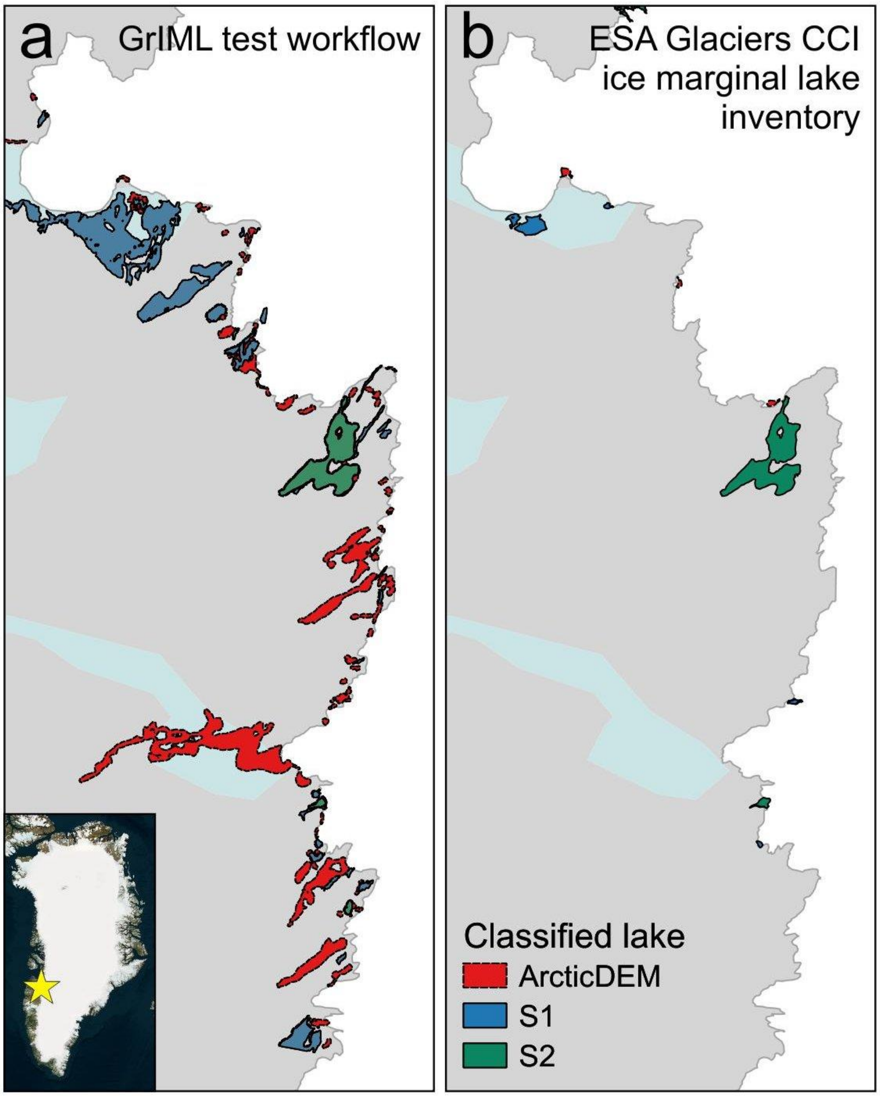

Figure 3. Preliminary lake classifications for an AOI in SW Greenland, showing the output from the current GrIML workflow (a) compared to the ice marginal lake classifications from the ESA Glaciers CCI inventory (b).

The current workflow is being tested on a selected area of interest (2394 km^2) at the SW margin of the Greenland Ice Sheet (hereafter referred to as the AOI) (Fig. 3a). GEE is used to retrieve the satellite imagery and perform the water classification methods remotely, following which the results are exported to the local environment for filtering and metadata generation.

Topographic sinks were computed from the ArcticDEM mosaic using a moving window to determine and dilate low elevation features, producing 3992 sinks for potential lake presence. 43 Sentinel-1 scenes in HH polarization (horizontal transmit, horizontal received) were used in the SAR classification, producing 3788 identified water bodies after averaging, speckle filtering and thresholding. 106 Sentinel-2 scenes were used in the optical classification, after filtering scenes by cloud cover (up to 50%).

3329 water bodies were classified from these after band resampling and averaging, and spectral indices classification. In total, 11,109 potential and directly classified water bodies were detected before filtering. The water bodies were subsequently exported from GEE's cloud processing platform to the local environment as vectorized polygons.

In the local environment, the water bodies were filtered by a size threshold of 0.05 km^2, leaving 1973 water bodies (removing 9136). A spatial filter was then applied using the modified version of the GIMP ice mask, where water bodies within a 1 km buffer of the ice margin were retained. In all, 100 water bodies were classified from the AOI, of which 70 unique water bodies (i.e., discounting multiple classifications from multiple methods) (Fig. 3a). The total processing time is 3 minutes (with an internet link speed of 144 Mb/s, using a Linux environment operating on an Intel Core i7 2.80 GHz quad core with 32 GB RAM).

Fewer ice marginal lakes were identified according to the ESA Glaciers CCI ice marginal lake inventory, with only 12 lakes present in the AOI compared to 70 classified with the GrIML workflow (Fig. 3b). This is primarily due to the manual filtering performed on the ESA Glaciers CCI inventory in post-processing. However, this comparison is useful in demonstrating the need for additional filtering in the GrIML workflow.

2.3.3 Other cloud computing options

Other cloud computing platforms were explored during this period, with the possibility to include alternative options as add-on modules to the GrIML workflow. The overall vision is to offer the primary GrIML workflow as an open-source package for public use, with optional add-on functionality from licensed cloud computing platforms should they be available to the user.

The SentinelHub Processing Python API was trialled under a short-term free license, to test whether the platform is suitable and compatible with the GrIML workflow. Sentinel-1 and Sentinel-2 image collections over the SW region of the Greenland Ice Sheet were used in the trial, which are accessible through SentinelHub as readily available image collections.

Generally, the platform is suitable for GrIML with a straightforward migration of the workflow. Tests were performed using the Processing Python API, however it would be more suitable to use the BatchProcessing API for rapid Greenland-wide processing that would promote the use of the GrIML workflow at an operational level. SentinelHub provides access to additional datasets (such as ASTER, Landsat and PLANET imagery) that would extend the inventories beyond the Sentinel era, and thus improve GrIML's time-series analysis. However, one limitation is the absence of a readily available high-resolution time-series DEM dataset, such as the ArcticDEM strip data. However, an external dataset can be ingested into SentinelHub via an AWS S3 (Amazon Web Services Simple Storage Service) bucket. Therefore, ingesting ArcticDEM strip data into SentinelHub is a possibility.

Overall, the performance and functionality of SentinelHub would be a valuable addition to the GrIML workflow, facilitating faster processing and access to a greater range of datasets. Use of SentinelHub in GrIML is subject to funding as an add-on component to the Living Planet Fellowship, as stipulated in the project proposal. If funded, it would be an invaluable addition to the project.

3 CONCLUSIONS

An initial workflow has been developed to classify ice marginal lakes from SAR and optical satellite imagery, and DEMs, based on the classification workflow from the ESA Glaciers CCI, under the Option 6 project 'An Inventory of Ice-Marginal Lakes in Greenland' (How et al., 2021). This workflow harnesses open-source, cloud-based processing to classify water bodies, under a unified Python processing chain that limits the reliance on heavy downloads and data storage.

The workflow has been tested on an AOI at the SW margin of the Greenland Ice Sheet (2394 km^2), identifying 70 potential ice marginal lakes in 3 minutes of processing time. Testing will now commence for larger regions to identify challenges in big data processing and look at adopting procedures to promote processing efficiency, such as parallel processing.

Comparison to the ESA Glaciers CCI ice marginal lake inventory highlights the need for stringent filtering measurements to remove mis-classifications and water bodies that do not share a common boundary with the ice margin. In the first instance, classification of the terrestrial ice margin will be explored in order to better classify lakes with a shared ice margin boundary; and move away from reliance on ice margin delineations with ill-matched acquisitions.

A trial of SentinelHub was also used to test its suitability for classifying ice marginal lakes, concluding that it would be a valuable addition that would allow access to additional datasets and processing capabilities. However, inclusion of SentinelHub is dependent on add-on funding. Add-on funding was also sought for fieldwork validation, which would greatly refine the workflow, and facilitate the examination of ice marginal lakes that have a direct impact on society and infrastructure. The funding for this add-on work needs to be addressed soon to ensure its successful implementation in GrIML.

4 REFERENCES

How, P. *et al.* Greenland-wide inventory of ice marginal lakes using a multi-method approach. *Sci Rep.* **11,** 4481 (2021). https://doi.org/10.1038/s41598-021-83509-1

Howat, I.M., Negrete, A., and Smith, B.E. The Greenland Ice Mapping Project (GIMP) land classification and surface elevation datasets. *Cryosphere* **8,** 1509-1518 (2014). https://doi.org/10.5194/tc-8-1509-2014

Rick, B. *et al.* Dam type and lake location characterize ice-marginal lake area change in Alaska and NW Canada between 1984 and 2019. *Cryosphere* **16,** 297-314 (2022). https://doi.org/10.5194/tc-16-297-2022

Shugar, D.H. *et al.* Rapid worldwide growth of glacial lakes since 1990. *Nat. Clim. Change* **10,** 939-945 (2020). https://doi.org/10.1038/s41558-020-0855-4

Wiesmann, A. *et al.* ESA Glaciers Climate Change Initiative (Glaciers_cci): 2017 inventory of ice marginal lakes in Greenland (IIML), v1. *Centre for Environmental Data Analysis (2021).* http://doi.org/10.5285/7ea7540135f441369716ef867d217519

Wilkinson, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Zhang, Mm., Chen, F., and Tian, Bs. An automated method for glacial lake mapping in High Mountain Asia using Landsat 8 imagery. *J. Mt. Sci.* **15,** 13–24 (2018). https://doi.org/10.1007/s11629-017-4518-5

5 PUBLICATIONS

- Carrivick, J.L., How, P., Sutherland, J., Cornford, S., Lea, J., Tweed, F., Grimes, M., and Mallalieu, J. Ice-marginal proglacial lakes across Greenland: present status and a possible future. *In Review*.

- Investigating Greenland's ice marginal lakes under a changing climate (GrIML), presented to the GEUS board of directors under the Glaciology and Climate department's ongoing activities.