unesco
United Nations
Educational, Scientific
and Cultural Organization

ADAPTATION FUND

GLOFCA
GLACIER LAKE OUTBURST FLOODS IN CENTRAL ASIA

REDUCING VULNERABILITIES OF POPULATIONS IN THE CENTRAL ASIA REGION FROM GLACIER LAKE OUTBURST FLOODS IN A CHANGING CLIMATE (GLOFCA)

GLOFCA Glacial Lake Inventory User Manual

GLOFCA Glacial Lake Inventory: User Manual

Section 1: Glacial lake inventory development

The GLOFCA glacial lake inventory is generated for Central Asia (Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan) using a standardised and largely automated approach. The inventory can be updated annually, or even more regularly for critical regions and compliments existing national lake mapping initiatives. The glacial lake mapping analytical toolbox backend algorithm (Normalised Difference Water Index, NDWI-G) is used to derive the inventory using open source Python libraries and freely available Sentinel-2 optical satellite data.

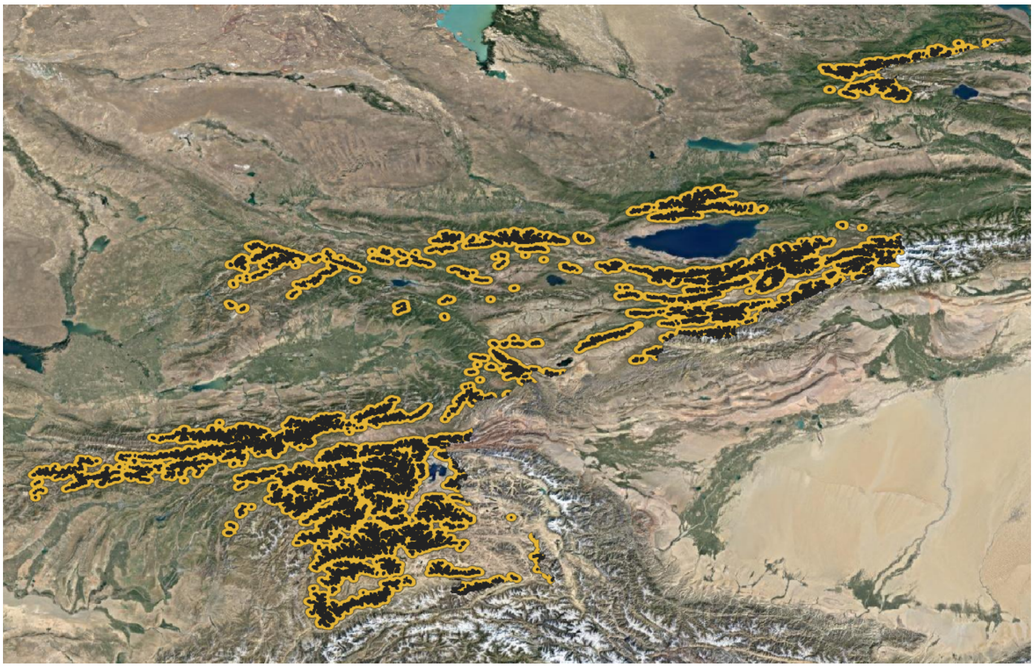
Section 1.1: How the regional glacial lake inventory was generated?

***Step 1*: *Get the glacier extents* (7 km buffer)**

The glacier extents in the four target countries (Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan) can be extracted from the Randolf Glacier Inventory (RGI) using QGIS software, see Fig. 1. To extract the glacier extents from the RGI in QGIS, use *Vector-> Geoprocessing Tools -> Clip.* To create the buffer region in QGIS, use *Vector-> Geoprocessing Tools -> Buffer*

a) The region covering the four target countries (OpenStreetMap, visualised in QGIS)

b) Glacier extents (black) in the four target countries extracted from the RGI (overlaid on OpenStreetMap).



c) The 7 km buffer region (shown as yellow, approx. area of 150,000 km^2) in and around the glaciers (shown as black) in the four target countries.

Figure 1. 7 km buffer region in and around the glacier extents in the four target countries.

Refer to the following book for more details on RGI:

Kargel, J.S., G.J. Leonard, M.P. Bishop, A. Kaab, B. Raup (Eds), 2014, Global Land Ice Measurements from Space (Springer-Praxis). 33 chapters, 876 pages. ISBN: 978-3-540-79817-0

Step 2*: *Download all the 69 Sentinel-2 tiles:

Download all the Sentinel-2 tiles (Level 1C, "Top of Atmosphere" product) that cover the 7 km glacier buffer region. For generating the GLOFCA glacial lake inventory, the data from one full month (August 2021, both S2A and S2B satellites, approx. 500GB) is downloaded using the Sencast toolbox. Instructions on installing the Sencast toolbox and downloading the Sentinel-2 data are in: https://renkulab.io/projects/odermatt/sentinel-hindcast

The 69 Sentinel-2 tiles are as follows: 'T44TNR', 'T42TXL', 'T42TXM', 'T43TCG', 'T42SYF', 'T43TDE', 'T43SDD', 'T44TMN', 'T43TGJ', 'T42SXJ', 'T43TDH', 'T44TLM', 'T43SBD', 'T43TCF', 'T42TWK', 'T42SXH', 'T43TGH', 'T43TGF', 'T43SCB', 'T43SEB', 'T43SCC', 'T43TEF', 'T44TKM', 'T42SUH', 'T42TXN', 'T42TYN', 'T44TLR', 'T44TMQ', 'T43TDG', 'T44TLL', 'T43TFF', 'T42SUJ', 'T44TMM', 'T43TGG', 'T43SDB', 'T43SCD', 'T43TCE', 'T43TFJ', 'T42SYG', 'T42SYH', 'T42SVJ', 'T44TMR', 'T43TCH', 'T42TYK', 'T44TKL', 'T43SBA', 'T43TBE', 'T44TLN', 'T43TEH', 'T43TFH', 'T43TEE', 'T42SXG', 'T43TBG', 'T42SYJ', 'T42SVH', 'T42TWM', 'T43TDF', 'T43SDC', 'T42TXK', 'T43TFG', 'T43TEG', 'T42SWH', 'T42SWJ', 'T43TFE', 'T43SBB', 'T43SBC', 'T44TLQ', 'T42TYM', 'T42TWL'

For all the 69 tiles, all data from August 2021 should be downloaded (*.SAFE* file extension) and stored in the following folder:

<INP_PATH_RASTERS>/rasters/sentinel2_L1C/2021/all_tiles/

Note: The variable *INP_PATH_RASTERS* can be set in the file *parameters.py.*

***Step 3*: *Pre-processing the Sentinel-2 data*:**

The completely cloudy or mostly cloudy Sentinel-2 tiles must be removed. Also, remove the tiles with almost zero coverage (invalid tiles), see Fig. 2 for an example.

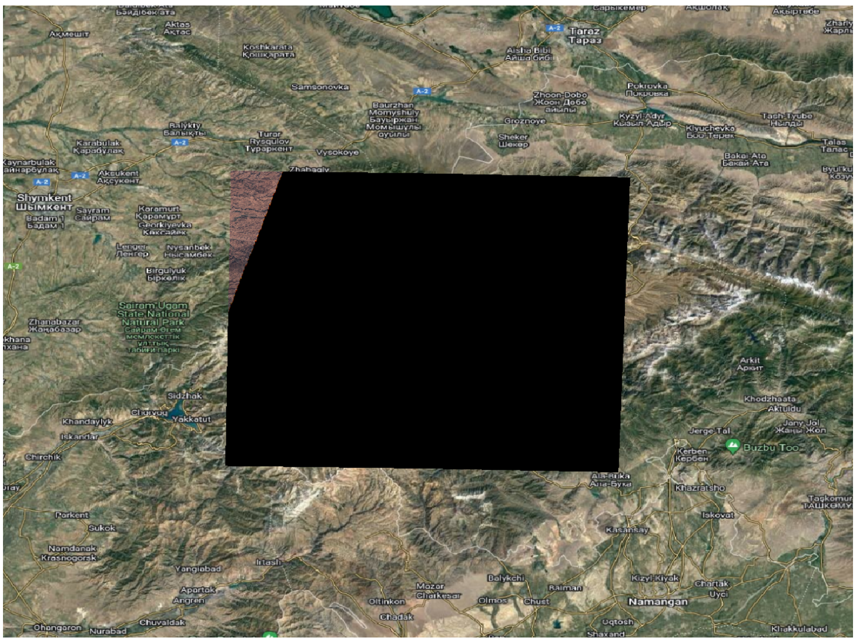
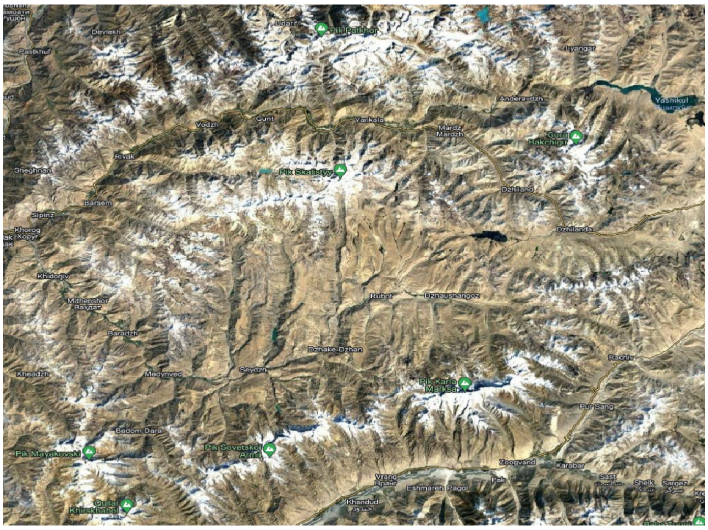


Figure 2. Example of an invalid Sentinel-2 tile (with almost zero coverage).

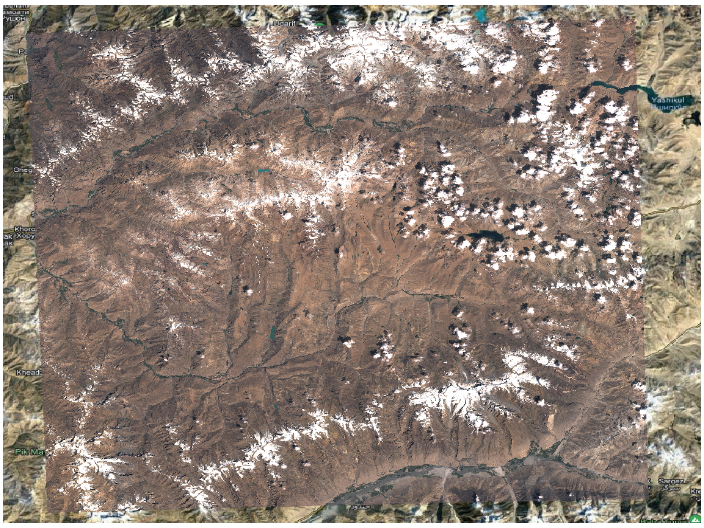
***Step 4*: *NDWI L1C product generation*:**

To generate the GLOFCA regional glacial lake inventory, we rely on water body mapping using NDWI-G (Normalised Difference Water Index), see Fig 3. For more details on NDWI-G index, refer to:

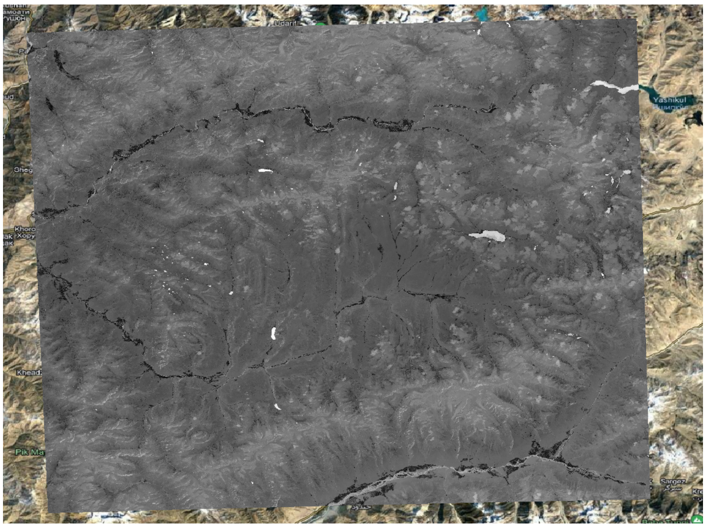
S. K. McFEETERS (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17:7, 1425-1432, DOI: 10.1080/01431169608948714



a) Region in and around the Sentinel-2 tile T43SBB (as in OpenStreetMap).



b) The T43SBB True Colour Image (TCI, L1C, 06 Aug 2021) overlaid on OpenStreetMap.



c) NDWI-G (T43SBB, 06 Aug 2021, L1C).

d) Thresholded NDWI-G (T >= 0.28) L1C product.

e) Thresholded NDWI-G (T >= 0.36) product.

Figure 3. Example water body mapping results for the Sentinel-2 tile T43SBB (06 Aug 2021).

Steps to be followed to generate the NDWI L1C product:

- First, the NDWI L1C products (raster, *.tif* format, one per Sentinel-2 L1C product that remains after the pre-processing steps) are generated for each tile by running the *jupyter* notebook:
  create_L1C_NDWI_product.ipynb
  The threshold value can be set in *parameters.py* using the variable *NDWI_G_THRESH*. The NDWI L1C products will be generated in the folder:
  <INP_PATH_RASTERS>/optical_gl/datasets_All/L1C/NDWI-Gen/CA_GA_2021_atlas/NDWI_G/NDWI_G_THRESH

- Second, the L1C products are accumulated to perform the multi-temporal analysis by running the *jupyter* notebook:
  multi_temporal_analysis.ipynb
  The accumulated NDWI products (raster, *.tif* format, one per Sentinel-2 tile) will be generated in the folder:
  <FINAL_PRODS_FLDR_MIRROR>/CA_GA_2021_atlas/sentinel2/Exp_NDWI_<NDWI_G_THRESH>/gpred/monthly_products
  Corresponding lake polygons (vector, *.shp* format, one per tile) will be generated in the folder:
  <FINAL_PRODS_FLDR_MIRROR>/CA_GA_2021_atlas/sentinel2/Exp_NDWI_<NDWI_G_THRESH>/gpred/lake_polygons

- Third, the 69 vector shape files need to be merged to a single vector shape file (vector, *.shp* format, one for the whole region). This is done by running the *jupyter* notebook: *merge_shp.ipynb*
  The following merged shape file will be generated:
  <FINAL_PRODS_FLDR_MIRROR>/CA_GA_2021_atlas/sentinel2/lake_polygons_merged/lake_polygons_merged.shp

Note 1: Python code to generate the NDWI L1C products, perform multi-temporal analysis and merging shape files is available at the online repository:
https://github.com/czarmanu/glofca/
under the sub-folder "lake_inventory_generation".

Note 2: The variables *INP_PATH_RASTERS and FINAL_PRODS_FLDR_MIRROR* can be set in the file *parameters.py*. Since the whole Sentinel-2 data from August 2021 sums up to around 500GB, we recommend storing the satellite images in an external hard drive and choosing *INP_PATH_RASTERS* as a folder in the external drive. However, *FINAL_PRODS_FLDR_MIRROR* must be a folder in the internal drive (creating polygons from raster functionality works only when executed from an internal drive).

Step 5: Post-processing steps in QGIS:
The following steps should be executed in the specified order (make sure that the *EPSG:4326* projection is set in QGIS):

- Load the shape file:
  <FINAL_PRODS_FLDR_MIRROR>/CA_GA_2021_atlas/sentinel2/lake_polygons_merged/lake_polygons_merged.shp
- Fix the invalid geometries in the shape file using the QGIS Processing toolbox (*Processing -> Toolbox -> Fix Geometries*)
- Remove the lakes outside the 7 km buffer region by clipping (*Vector-> Geoprocessing Tools -> Clip*)
- Duplicate removal needed for regions where multiple Sentinel-2 tiles overlap. This can be done by dissolving (*Vector-> Geoprocessing Tools -> Dissolve*) followed by converting multipart to single parts (*Vector-> Geometry Tools -> Multipart to single parts*)
- Smooth the lake polygons using the QGIS Processing toolbox (*Processing -> Toolbox -> Smooth*)
- Delete the fields "DN" and "area" from the lake polygons vector (*Open attribute table -> Toggle editing mode -> Delete selected features*)

- Assign new attributes to the lake polygons vector shape file:
    - Create new attribute: **Year**
        - Right click on the shape file (in "layers panel"): Click "*Open Attribute Table*"
        - Click "*Toggle editing mode*" in the attribute table window
        - Click "*Open field calculator*"
        - *Create a new field*
            - *Output field name*: "Year"
            - *Output field type*: Choose "Whole number (integer)" or "Integer(32 bit)"
            - *Output field length*: 10, precision: 3
            - In "*Expression*" box, write "2021"
        - Click "*OK*"
        - Click "*Toggle editing mode*" and "*save*"
    - Create new attribute (in QGIS): **Month**
        - Click "*Toggle editing mode*"
        - Click "*Open field calculator*"
        - *Create a new field*
            - *Output field name*: "Month"
            - *Output field type*: Choose "Whole number (integer)" or "Integer(32 bit)"
            - *Output field length*: 10, precision: 3
            - In "*Expression*" box, write "8"
        - Click "*OK*"
        - Click "*Toggle editing mode*" and "*save*"
    - Create new attribute: **Latitude**
        - Click "*Toggle editing mode*"
        - Click "*Open field calculator*"
        - *Create a new field*
            - *Output field name*: "Latitude"
            - *Output field type*: Choose "*Decimal number (real)*"
            - *Output field length*: 10, precision: 3
            - In "*Expression*" box, write: "y(transform($geometry, layer_property(@layer_name, 'crs'), 'EPSG:4326'))
        - Click "*OK*"
        - Click "*Toggle editing mode*" and "*save*"
    - Create new attribute: **Longitude**
        - Click "*Toggle editing mode*"
        - Click "*Open field calculator*"
        - *Create a new field*
            - *Output field name*: "Longitude"
            - *Output field type*: Choose "*Decimal number (real)*"
            - *Output field length*: 10, precision: 3
            - In "*Expression*" box, write: "x(transform($geometry, layer_property(@layer_name, 'crs'), 'EPSG:4326'))
        - Click "*OK*"
        - Click "*Toggle editing mode*" and "*save*"

    - Create new attribute: **Lake Area**
        - Click "*Toggle editing mode*"
        - Click "*Open field calculator*"
        - *Create a new field*
            - *Output field name*: "Area"
            - *Output field type*: Choose "*Decimal number (real)*"
            - *Output field length*: 10, precision: 3
            - In "*Expression*" box, write: "$area"
        - Click "*OK*"
    - Click "*Toggle editing mode*" and "*save*"
- Filter out very small lakes (area <2500 m^2)
    - Right click on the lake polygon vector file in the layers panel
    - In "Provider Specific Filter Expression" use the expression "Area" >= 2500 and click OK
- Filter out big lakes (area > 10km^2)
    - Right click on the lake polygon vector file in the layers panel
    - In "Provider Specific Filter Expression" use the expression "Area" <= 10000000 and click OK

Note: Sentinel-2 has a spatial resolution of 10m in the best case. Hence, only the lakes with a surface area above 0.01 sq. km. can be detected with good accuracy. Note also that no manual editing of the lake outlines is done and hence the whole processing is 100% reproducible.

Section 1.2: How to select the NDWI thresholds?

Careful NDWI threshold selection is needed based on expert judgement. We use a base threshold of 0.28 for all 69 tiles. Then, for some selected tiles, the threshold is adapted (as per instructions in section 1.3) after a manual visual inspection of the lake outlines. During the visual inspection, if over-detection of lake outlines is noted for a Sentinel-2 tile, that means the NDWI threshold chosen is smaller than what is actually required. Hence, use a higher threshold to improve the results. Similarly, if the outlines are under-detected, the chosen NDWI threshold is higher than what is required. A lower threshold is actually better in such a case.

Section 1.3: How to adapt the NDWI thresholds only for a few selected tiles?

- *Step 1*: In *parameters.py,* comment out the variable *S2_TILES* with 69 tiles and set the tiles for which the threshold needs to be adapted. For example, if adaptation is needed for the tile T43TFH, assign S2_TILES = ["T43TFH"]. In case the threshold of more than one tile need to be adapted (say, T43TDH and T42TXM), assign S2_TILES = ["T43TDH", "T42TXM"]
- *Step 2*: In *parameters.py,* choose the new threshold by setting a new value for the parameter NDWI_G_THRESH
- *Step 3:* Run the jupyter notebook *"create_L1C_NDWI_product.ipynb"* after setting the new parameters (steps 1 and 2). This step will regenerate the NDWI L1C products (only for the tiles for which a threshold adaptation is needed)

- *Step 4: Run the jupyter notebook "multi_temporal_analysis.ipynb".* This step will regenerate the accumulated products (only for the tiles for which a threshold adaptation is needed)
- *Step 5: Run the jupyter notebook "merge_shp.ipynb".* This step will regenerate the polygons for the selected tiles (according to the new threshold). This will also regenerate the merged lake polygons. Note that, for the selected tiles, the polygons generated will correspond to the new threshold and for the remaining tiles, the old threshold applies.

Section 2: Spatial and temporal coverage of the regional glacial lake inventory

The GLOFCA regional glacial lake inventory includes all glacial lakes detected within the 7 km buffer region from August 2021. See Fig. 4 for the spatial coverage.

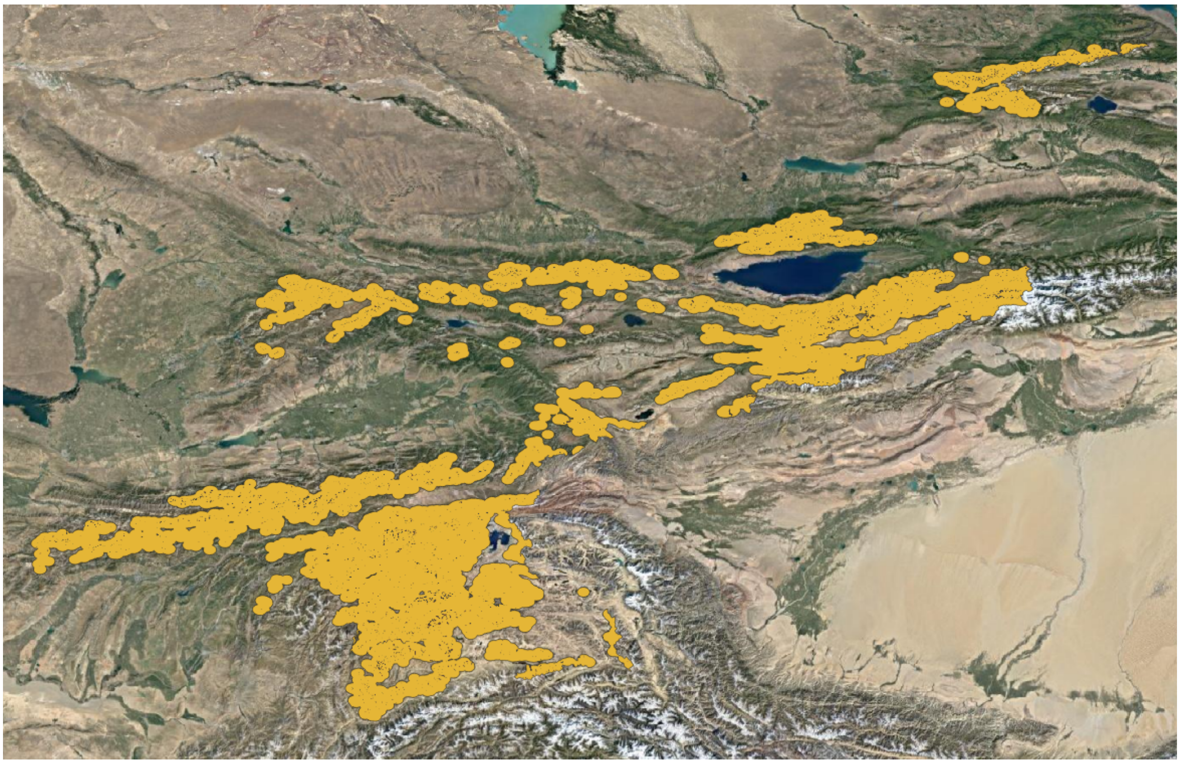


Figure 4. Spatial coverage of the regional glacial lake inventory.

Section 3: Output format

The GLOFCA glacial lake inventory is provided as a vector (.*shp* file) with the following attributes:

- Year
- Month
- Latitude
- Longitude
- Area

Example attributes are shown in Fig. 5. Note: the unit of the lake area is sq. m.

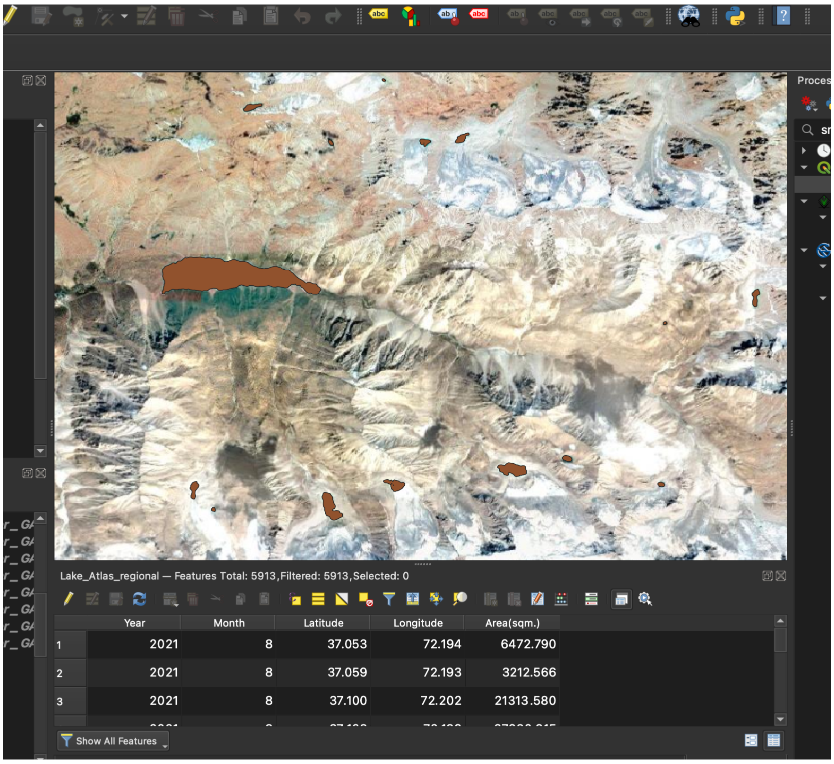


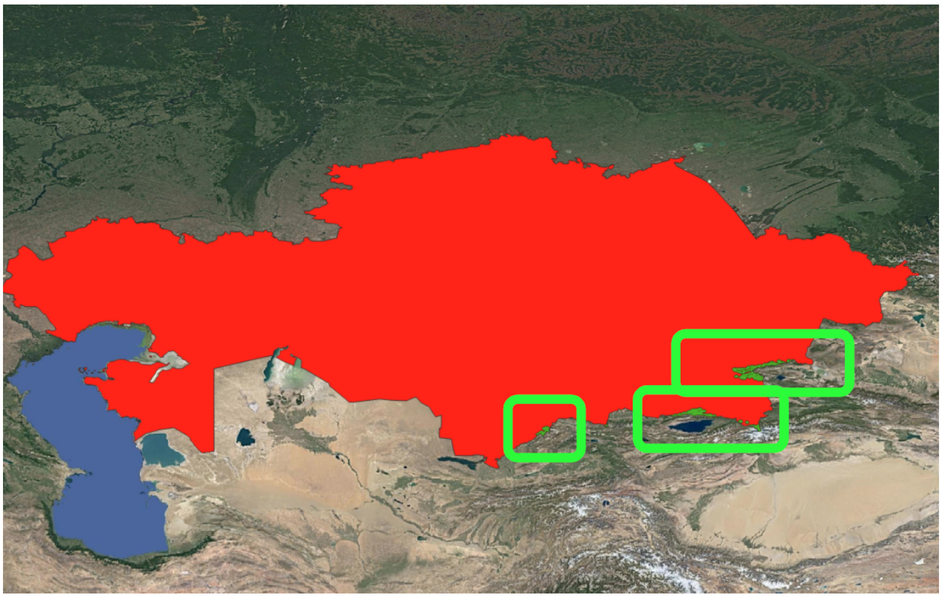
Figure 5. Example lake attributes (Year, Month, Latitude, Longitude, Area) as in the *Attribute Table* in QGIS.

Section 4: How the GLOFCA national glacial lake inventories are generated from the regional glacial lake inventory?

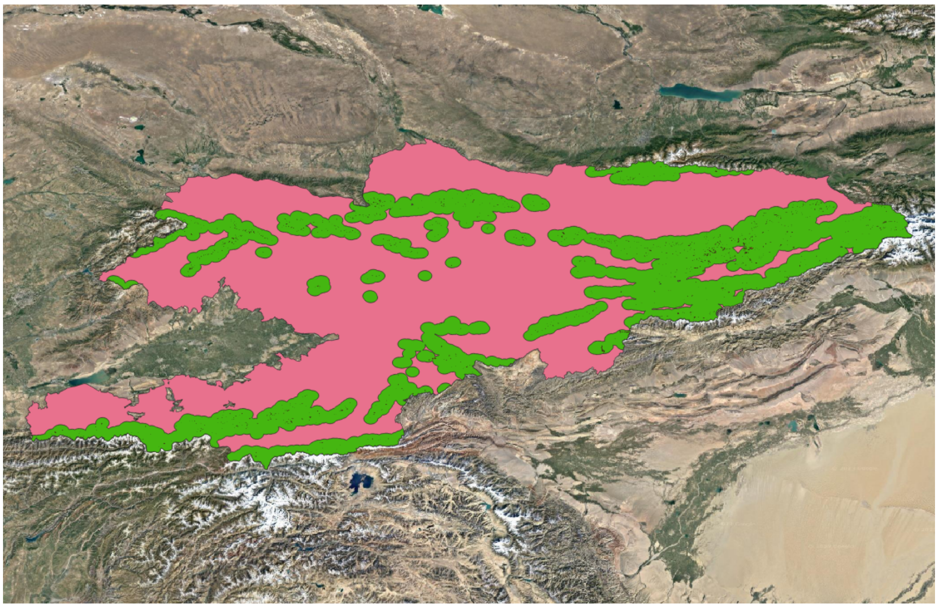
To generate the national lake inventory, follow the steps in the provided order:

- Load the regional lake inventory shape file in QGIS
- Load the national boundary shape file in QGIS
- Clip the lakes within the target country from the regional glacial lake inventory. In QGIS, use *Vector-> Geoprocessing Tools -> Clip*
- See Fig. 6 for example results

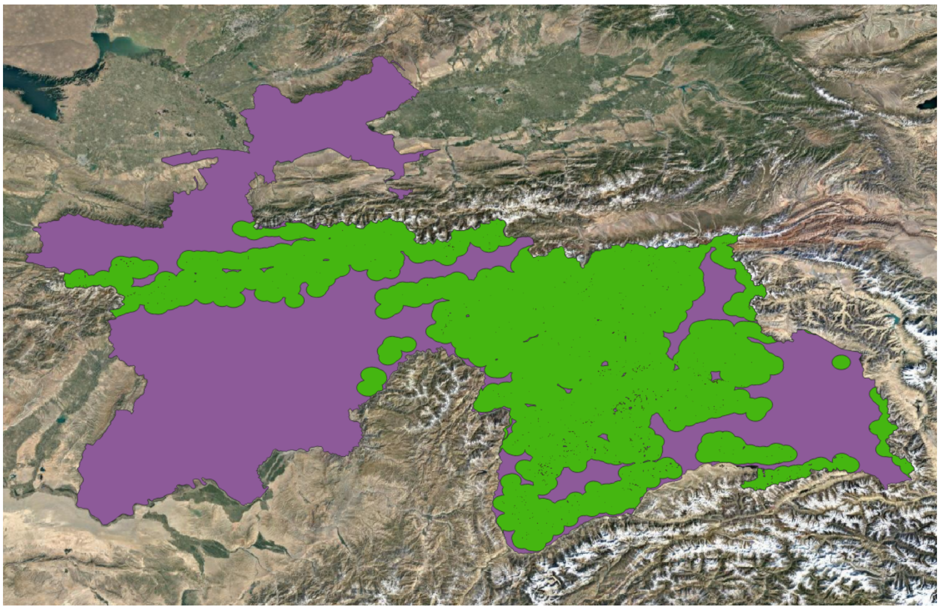
  Note: the sub-figures 5a, 5b, 5c and 5d show unofficial country boundaries just for the purpose of illustration. We recommend to use the official country boundary polygons while clipping the country-scale lake inventories from the regional lake inventory.



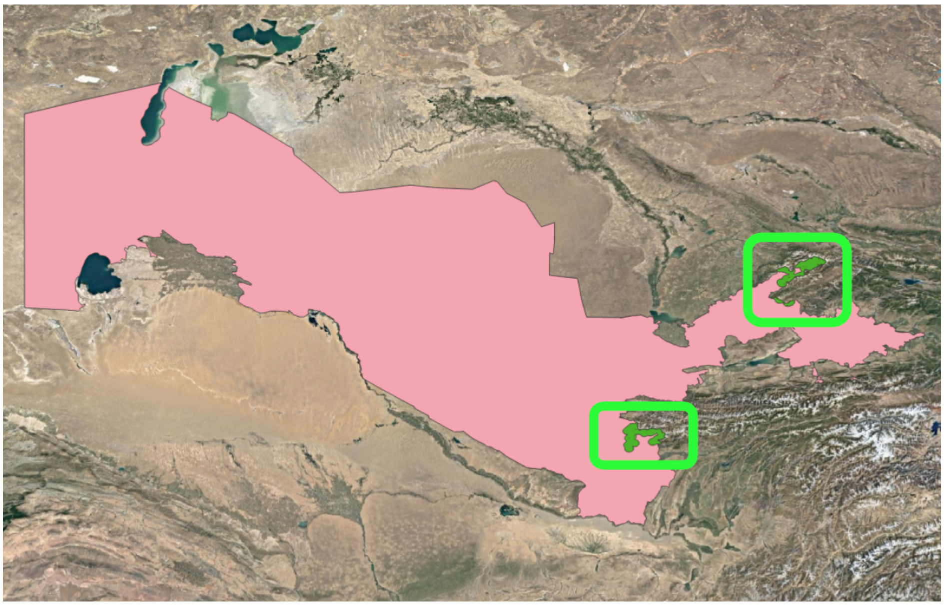
a) Country glacial lake inventory coverage: Kazakhstan. The glaciated regions (green, 7km buffer) with lakes overlaid are highlighted using yellow rectangles.



b) Country glacial lake inventory coverage: Kyrgyzstan. The glaciated regions (green, 7km buffer) with lakes overlaid are shown.



c) Country glacial lake inventory coverage: Tajikistan. The glaciated regions (green, 7km buffer) with lakes overlaid are shown.



d) Country glacial lake inventory coverage: Uzbekistan. The glaciated regions (green, 7km buffer) with lakes overlaid are highlighted using yellow rectangles.

Figure 6. Example illustration of the country-scale GLOFCA glacial lake inventories for all the four target countries.

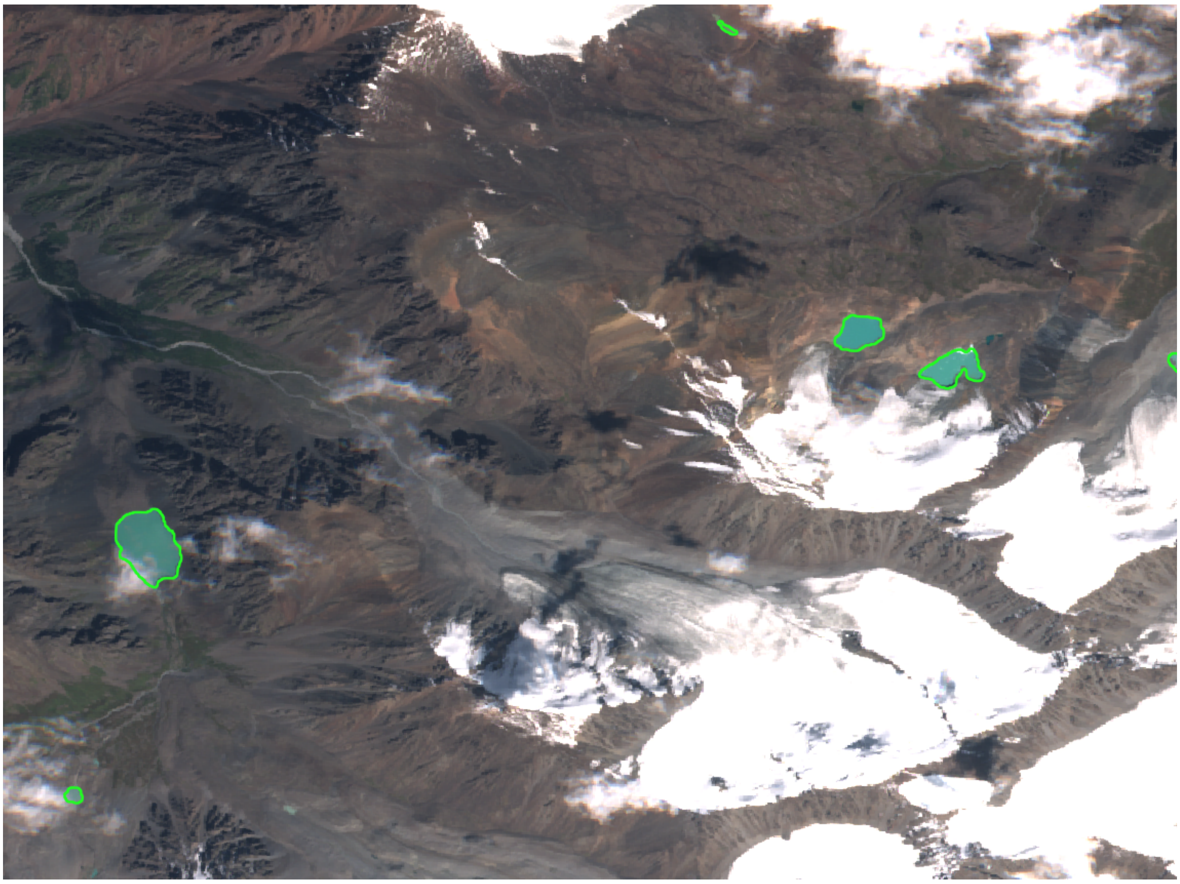
Section 5: How to generate glacial lake inventory for a future year?

Follow the instructions below to generate a new glacial lake inventory for a future year, say 2025?

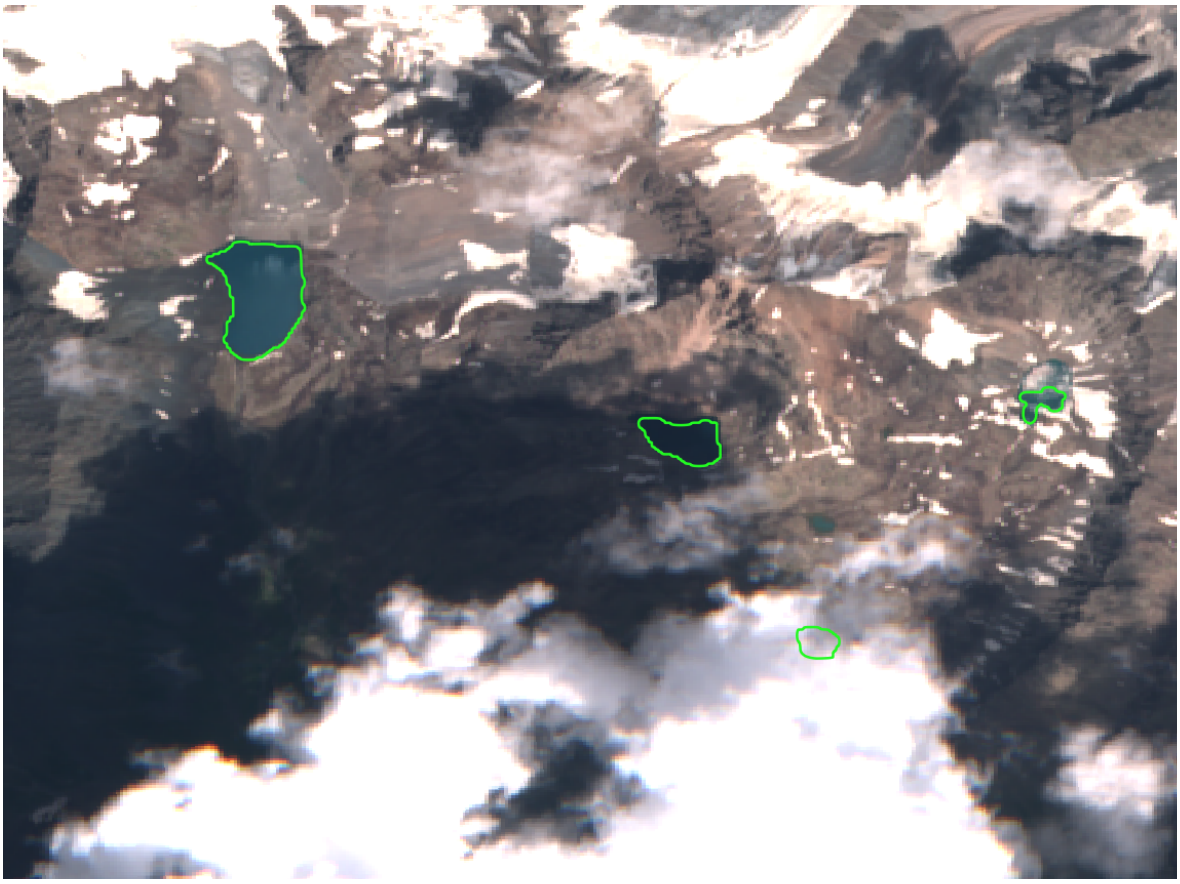
- Sentinel-2 data need to be downloaded for all 69 tiles from the year for which the inventory need to be generated. We recommend downloading the satellite imagery from August.
- Follow the instructions in Section 1.1 to create NDWI for all 69 tiles which involves
  - NDWI L1C product generation
  - NDWI monthly accumulation
- Choose the correct NDWI threshold for each tile by following instructions in sections 1.2 and 1.3
  - Start with a base threshold of 0.28
  - Adjust thresholds per-tile after analysing the results
    - Increase the threshold if over detection
    - Reduce the threshold if under-detection
    - Keep the threshold unchanged if the detection is good
- Perform the QGIS post-processing steps as described in Section 1.1

Section 6: Some example lakes

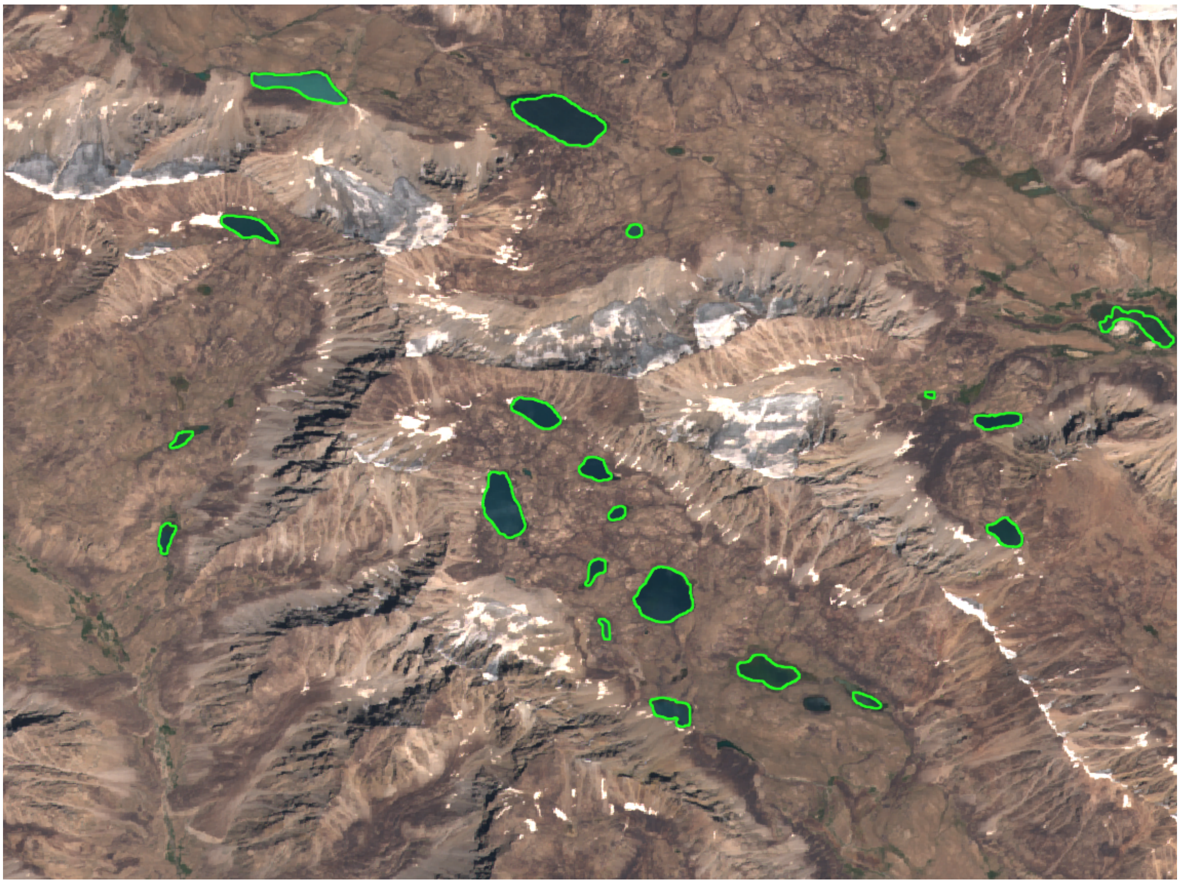
Some example lakes in the GLOFCA regional lake inventory are shown in Fig. 7.



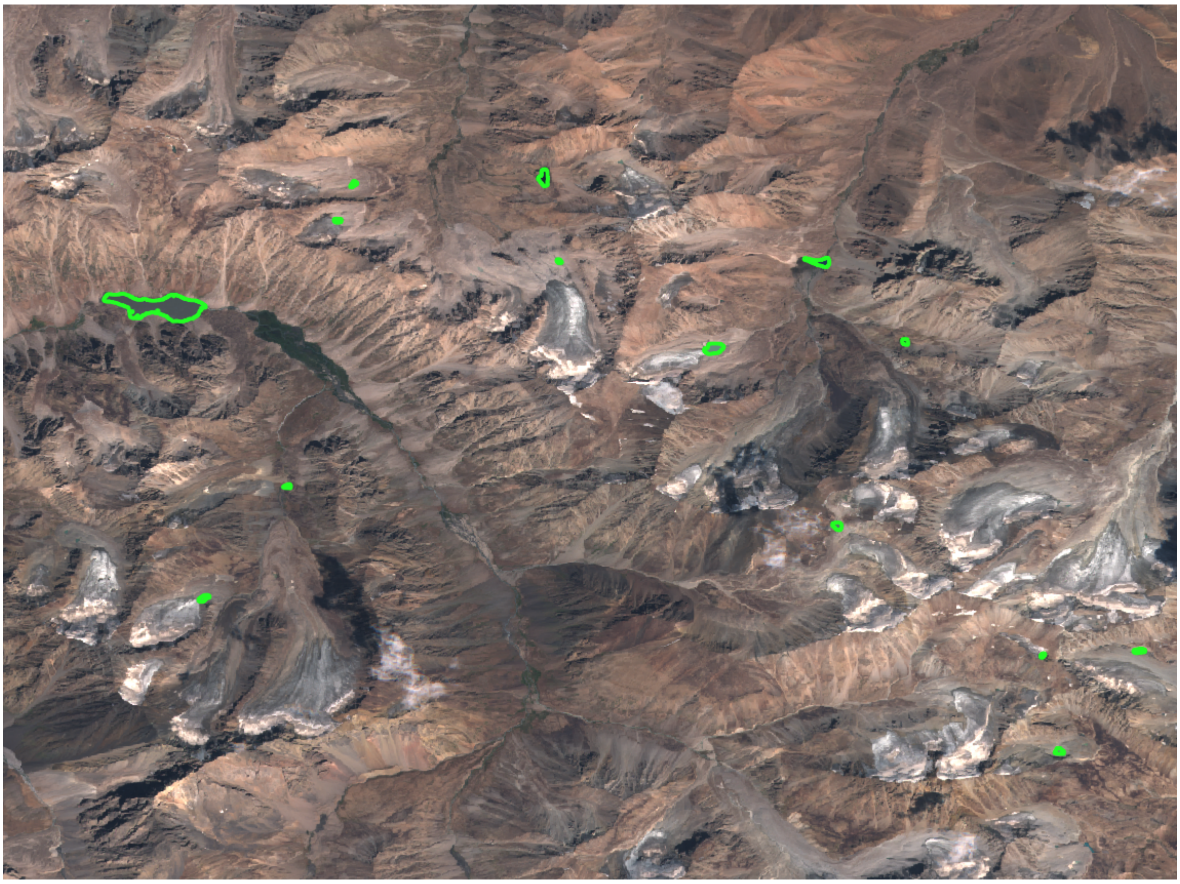
a) Example lakes (green outline) from Kazakhstan (overlaid on Sentinel-2 TCI from 23 Aug 2021, T43TFH)



b) Example lakes (green outline) from Kyrgyzstan (overlaid on Sentinel-2 TCI from 6 Aug 2021, T43TDH)



c) Example lakes (green outline) from Tajikistan (overlaid on Sentinel-2 TCI from 16 Aug 2021, T43SBB).



d) Example lakes (green outline) from Uzbekistan (overlaid on Sentinel-2 TCI from 19 Aug 2021, T42TXM)

Figure 7. Example lakes from Kazakhstan (a), Kyrgyzstan (b), Tajikistan (c) and Uzbekistan (d) overlaid on the Sentinel-2 TCI (True Colour Image).

Section 7: Qualitative comparison with existing lake inventory

For a selected region in Tajikistan, a qualitative comparison of our glacier inventory results (from August 2021, green) with Wang et al. (2020, blue) is shown in Fig. 8. Note that the existing lake inventory is from 2018, however, the GLOFCA lake inventory is from 2021. It can be seen that some lakes that were not present in the existing inventory have been detected by our algorithm.

Note: a direct comparison with the 2018 inventory has limitations as the GLOFCA inventory is based on satellite images from 2021. From 2018 to 2021, the lakes might have changed shapes or even disappeared. Additionally, new lakes might have formed. In addition, both algorithms use different thresholds. Moreover, the definition of a glacial lake is different in both approaches. Refer to the following publication for more details on the 2018 inventory:

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., and Tang, Z.: Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images, Earth Syst. Sci. Data, 12, 2169–2182, https://doi.org/10.5194/essd-12-2169-2020, 2020.

The 2018 lake inventory can be downloaded from: https://doi.org/10.12072/casnw.064.2019.db

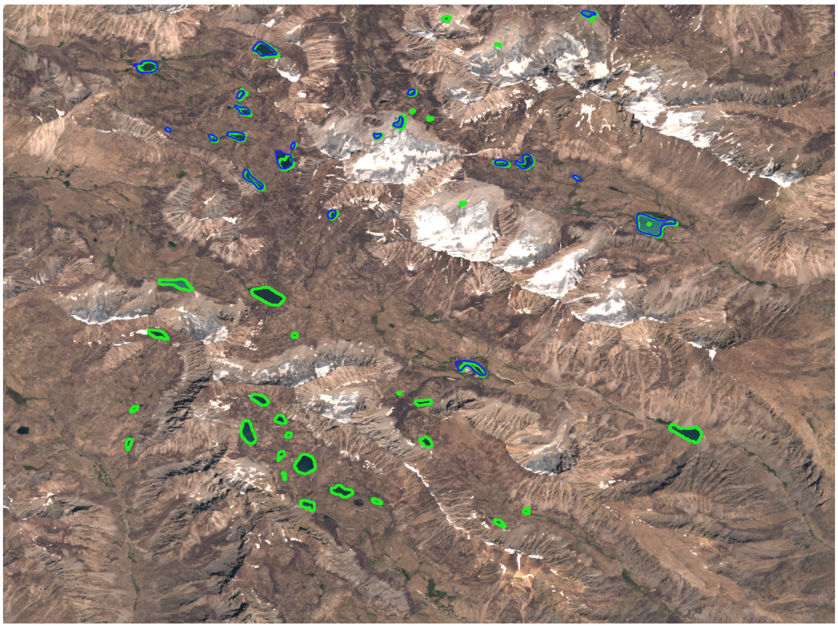
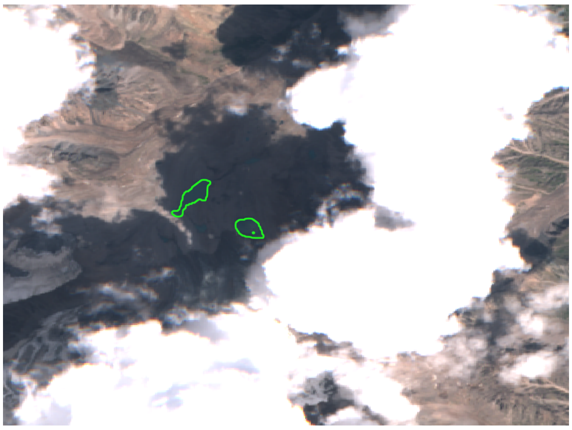


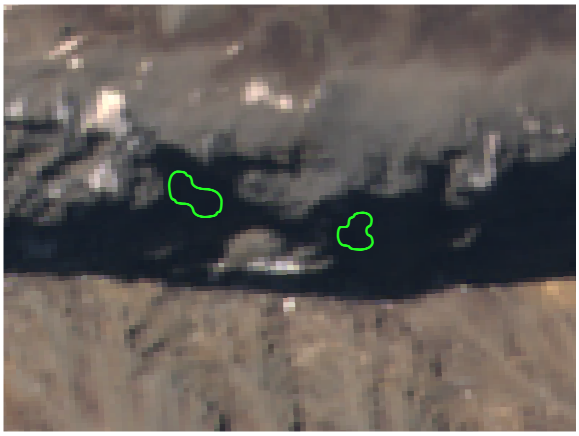
Figure 8. Qualitative comparison of GLOFCA glacial lake inventory (green) with existing lake inventory (blue, Zhang et al., 2020).

Section 8: Performance in cloudy scenarios

Some examples of the lakes in GLOFCA inventory detected in cloudy scenarios are shown in Fig. 9. Our algorithm performs well even in the presence of light cloud shadows, see Fig 9a. This good performance is due to the multi-temporal analysis that fuses the detections from multiple L1C NDWI products within August 2021. However, we also noticed false positives in the case of darker cloud- and cast shadows, see Fig. 9b. Such false detections need to be manually eliminated.



a) Examples of good detection even in cloud shadows.



b) Example false positive detections during darker and consistent shadows.

Figure 9. Performance in the presence of cloud- and cast shadows.

Section 9: Other false positives

Since our algorithm is fundamentally a water-body detection technique, all the water bodies (including river segments) are detected that lie within the 7 km buffer region in and around the glacier boundary, see Fig. 10 for an example from Tajikistan. Such false positives should be eliminated manually.

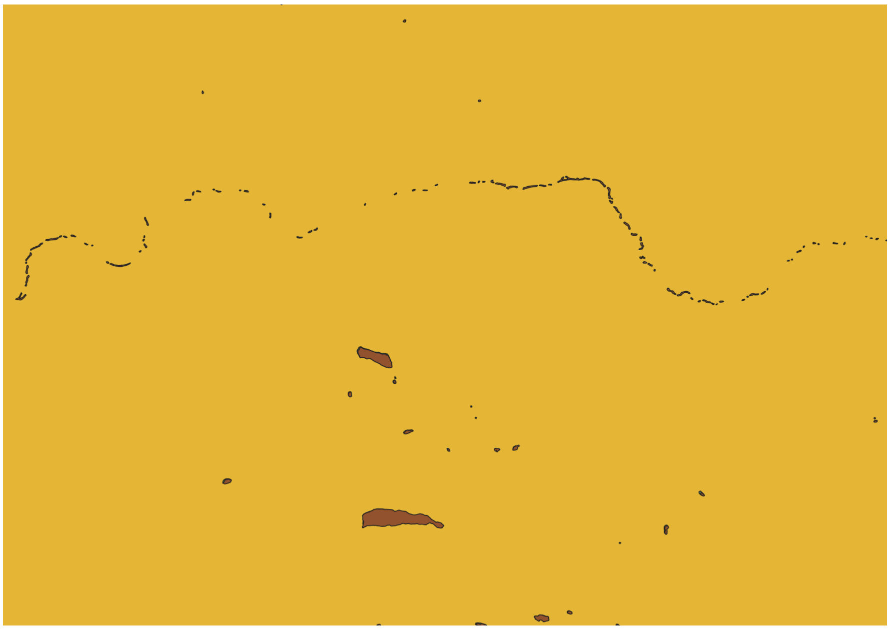


Figure 10. River segments detected as false positives (example from Tajikistan).

We suggest removing the false positive river segments as follows:

- Load the GLOFCA glacial lake inventory shape file in *Google Earth*
- *Manually* spot the rivers segments
- Delete the river segments (one by one) by:
    - Zooming to the river segment
    - Right click (on the river segment) -> Delete
- Save the modified glacial lake inventory*:*
    - "*Save place as*"
    - Choose "*file type*" as "kmz"

Section 10: Indications to local experts to undertake final validation of the national lake inventories

Validation of the outline for each lake in the inventory could be done as per the following instructions:

- Load the lake inventory shape file in *Google Earth* and compare the lake outlines against the images from 2021 (if possible, compare against the images from August 2021 itself since we used Sentinel-2 data from August 2021 to generate the lake inventory.). In the *Google Earth* platform, manually delete the false positives (wrong detections due to shadows, river segments detected as lakes etc.). Other high-resolution satellite imagery also may be used for validation (if already available). However, we strongly recommend using the images from 2021 August.
- In case of false negatives (miss detections) and wrong detections (cases where the detected lake outline has less than 50% overlap with the real lake outline that can be judged from the *Google Earth* imagery or the existing 2018 inventory), perform a manual correction of the lake outlines.
- We checked (qualitative visual verification) for most of the lakes in pilot sites. As the next step, a local expert should thoroughly check each lake and do necessary adjustments.