# Challenges to merging multitemporal geospatial data: glaciers

Bruce H. RAUP, Adina RACOVITEANU, Richard L. ARMSTRONG, Siri Jodha Singh KHALSA

braup@nsidc.org

National Snow and Ice Data Center
University of Colorado Boulder

## Summary

**Problem:** **We have glacier outlines in a database with unique IDs. When new outlines are produced, it is challenging to assign the same ID to each new outline that covers an existing outline that has an ID already. Glaciers may have changed size or broken into pieces between the two datasets.**

**Solution:** **We wrote software that spatially compares the new outlines to the old and assigns either existing or new IDs based on topological rules.**

### Background on the GLIMS Initiative

GLIMS (Global Land Ice Measurements from Space) is an effort to map the world's glaciers at high resolution. Analysis is done by a globally distributed group of researchers. Important points:

- The database can contain multitemporal outlines and other data for each glacier, requiring a single glacier ID to tie the records together
- A unique ID for each glacier is imperative for obtaining statistics or performing change analysis.

### Software used

This code is written in Python with the module “osgeo” (osgeo.ogr, osgeo.osr) to read and write shapefiles, as well as to calculate polygon areas, intersections, and centroids.

### Approach

Polygons are compared by computing degree of overlap:

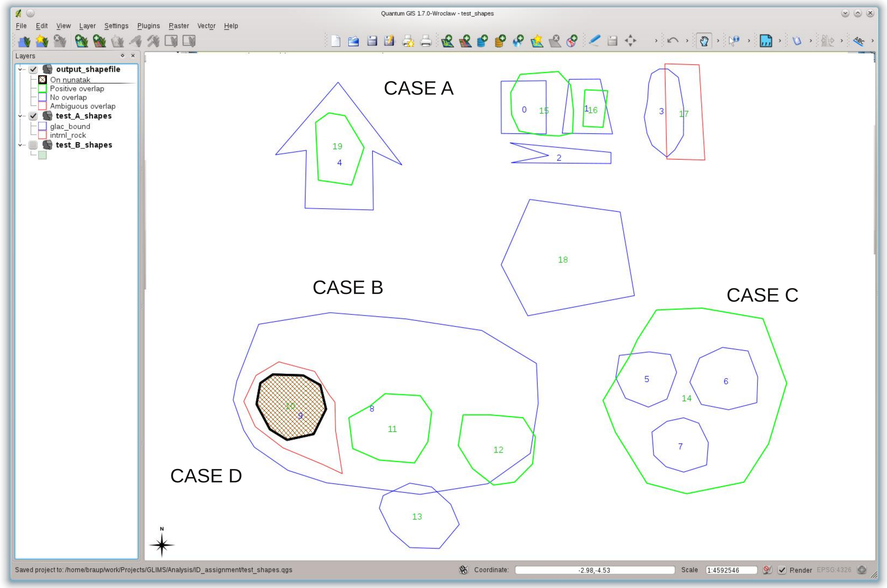
$$lap(A,B) = \frac{|A \cap B|}{|A|}$$

$$maxlap(A,B) = maxlap(B,A) = max(lap(A,B), lap(B,A))$$

The overlap fraction, or “maxlap” ranges from 0 (no overlap) to 1 (complete overlap or containment). A threshold is applied to this overlap fraction:

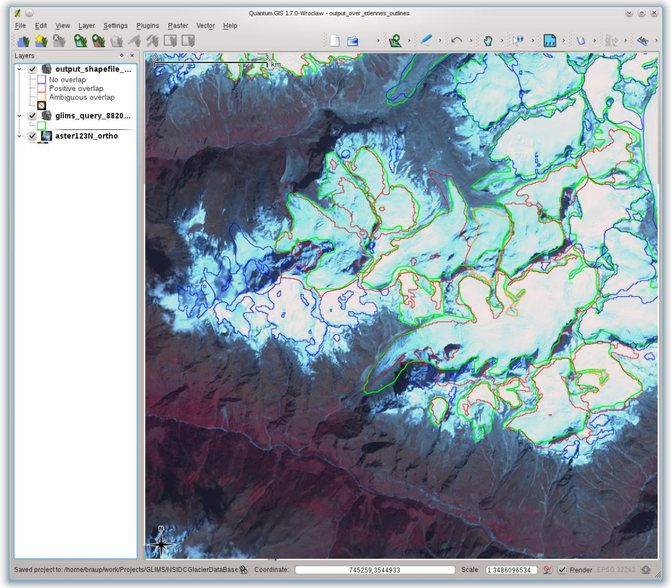
| Value of overlap fraction | Action |
|---|---|
| [0, 0.3) | No overlap; warning issued |
| [0.3, 0.7) | Ambiguous overlap; warning issued |
| [0.7, 1.0] | Definite overlap |

## Simplified test case

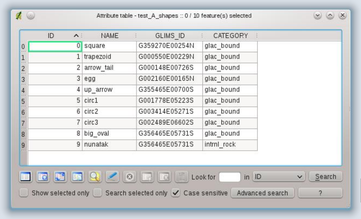



Test case illustrating various cases (labeled above and below). Case A: one-to-one overlap; Case B: one-to-many overlap; Case C: many-to-one overlap; Case D: small glacier on nunatak (rock outcrop within a glacier).

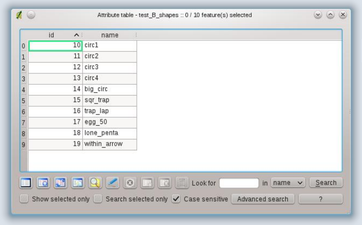
## Real-world test case



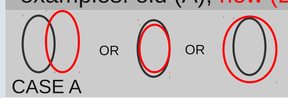
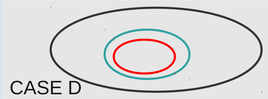
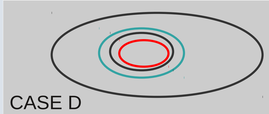
Result of algorithm applied to real data. Green polygons are existing set of outlines. New outlines are color-coded according to completeness of computed overlap with existing outlines. Some errors in the representation of the new outlines led to wrong computed overlaps (blue polygons that overlap green ones).

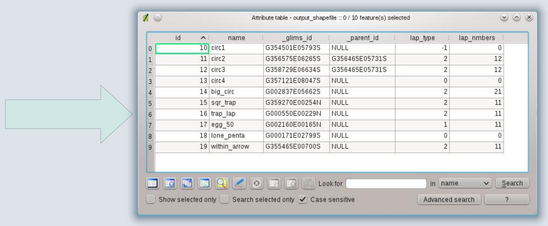


Attributes of existing dataset (“A” shapes)



Attributes of new dataset (“B” shapes)

| Shape relationship examples: old (A); new (B) | Interpretation and Action |
|---|---|
|  CASE A | Simple one-to-one overlap. Threshold on percent overlap and 1. assign same ID, or 2. assign same ID and warn of misregistration, or 3. assign new ID and warn of ambiguity. |
|  CASE B | Multiple overlap, where glacier has broken into pieces. Threshold on percent overlap as above. Assign new IDs to B shapes, and associate ID of **A shape** with B shapes as “parent ice mass”. |
|  CASE C | Multiple overlap, where glacier pieces have merged, or been reanalyzed at coarser resolution. Assign new ID to B shape, and back-populate this new ID into “parent ice mass” fields of **A shapes**. |
|  CASE D | Small glacier (B shape) is contained within a large glacier but is on a rock outcrop (nunatak) within that large glacier. Assign a new ID. |
|  CASE D | Same as above, but there is an existing glacier outline (**A shape**) on the rock outcrop already with its own ID. Assign the same ID as this to the B shape. |



Merged attributes of new (B) shapes.

### Summary and Future Work

- We have written software that assigned IDs to new glacier outlines based on their spatial relationships to existing outlines.
- The software works for simple cases and some complex cases.
- In the future, the software needs to
  - read the existing outlines from the definitive database, rather than only use shapefiles;
  - Handle the case of multiple sets of existing outlines.



CIRES University of Colorado Boulder