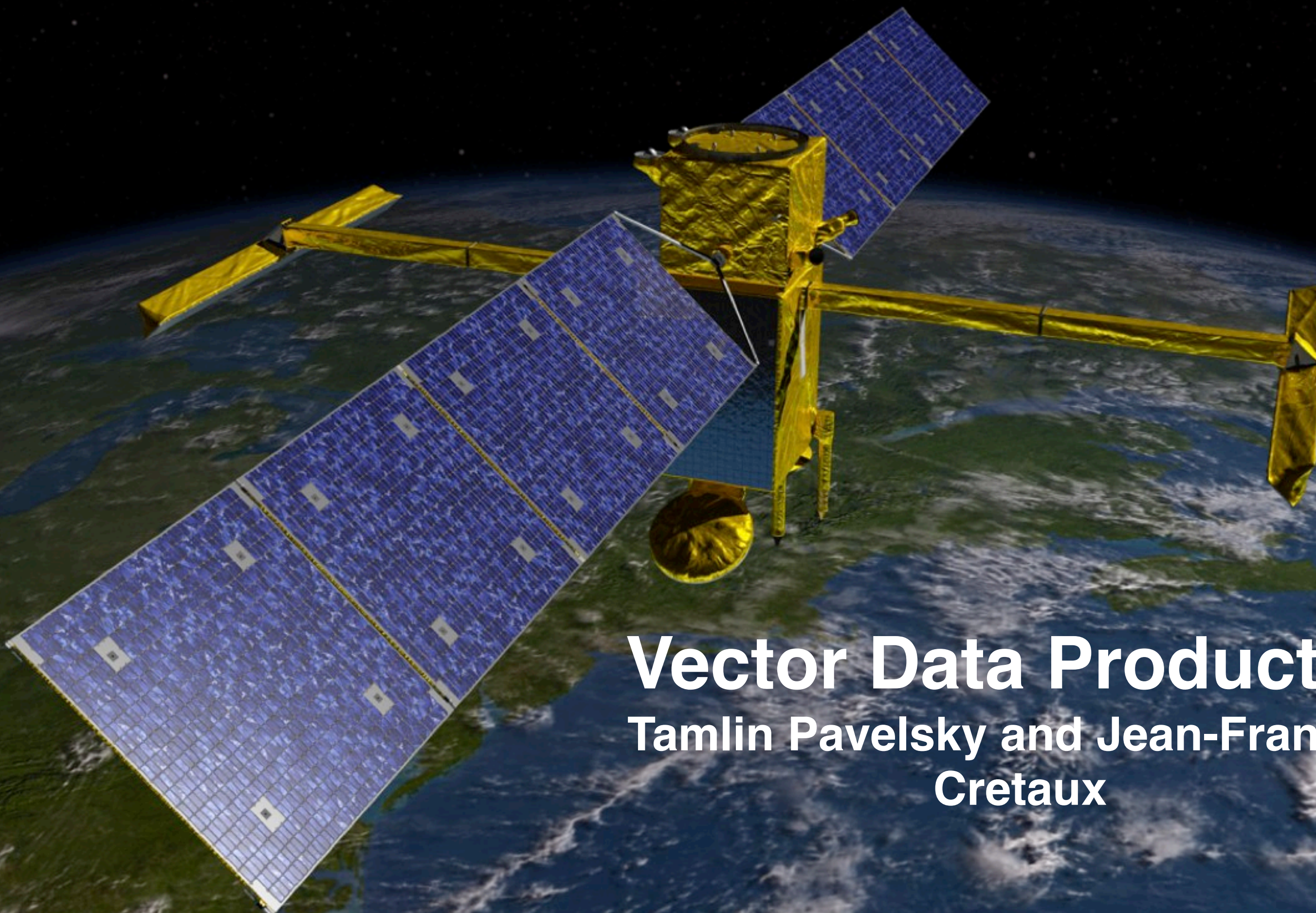
Surface Water and Ocean Topography (SWOT) Mission



Vector Data Products

Tamlin Pavelsky and Jean-Francois Cretaux

SWOT Science Team Meeting, June 13-16, 2016

Proposed Vector Data Products

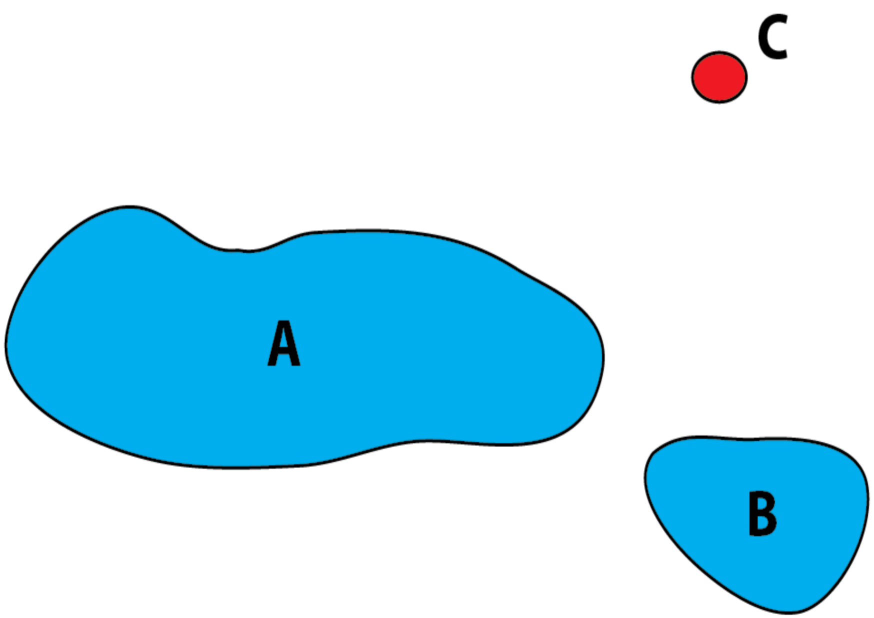
- Pass-based product:
  - Produced separately for every SWOT overpass
- Cycle-based product:
  - Produced from all overpasses for one 21-day cycle
  - A monthly product could be produced using a similar method
- For both datasets, there needs to be an online tool allowing extraction of attribute data in tabular form.

- Both of these products will depend on *a priori* lake/river masks that must be generated before launch.

A Priori Lake Information

- We already have a great deal of information about where lakes are, globally.
  - Global Lakes and Wetlands Database (Lehner and Doll, 2004)
  - Landsat-derive databases (**Sheng, in prep**; Verpoorter et al, 2014)
- Before launch, we should develop an *a priori* lakes database that contains:
  - All lake features likely to be detectable by SWOT, with each lake having a unique identification code
  - A nominal height extracted from a DEM or altimetry data
  - A flag for whether each feature is likely to experience ice cover.
- This *a priori* mask should be updated during the mission using SWOT data.

Example Lakes for Illustration



Lake A: 24 km^2
Lake B: 7.5 km^2
Lake C: 0.8 km^2

Lakes A & B in a priori mask
Lake C not in a priori mask

Pass-based Lake Vector Product

Begin with A Priori Mask and SWOT Pixel Cloud Product

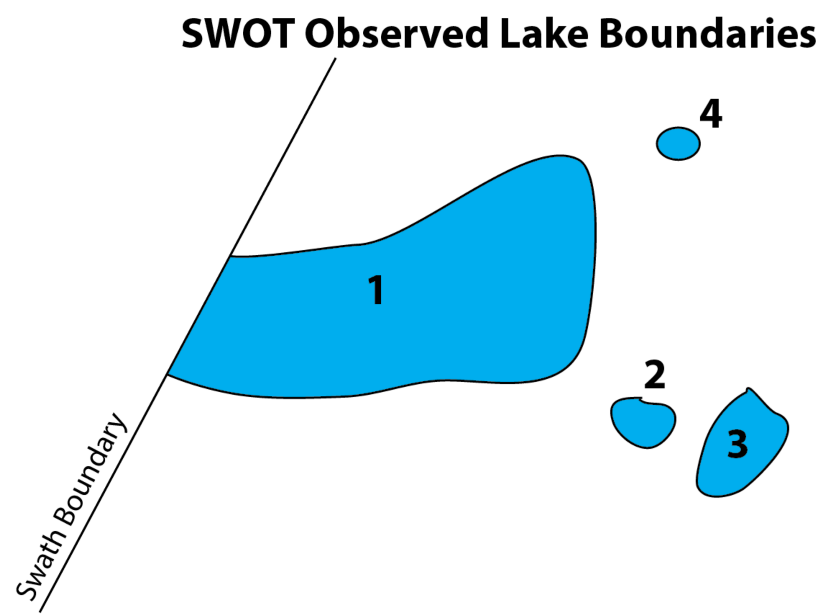
Pass-based Lake Vector Product

Begin with A Priori Mask and SWOT Pixel Cloud Product

↓

Classify lake regions in SWOT Data

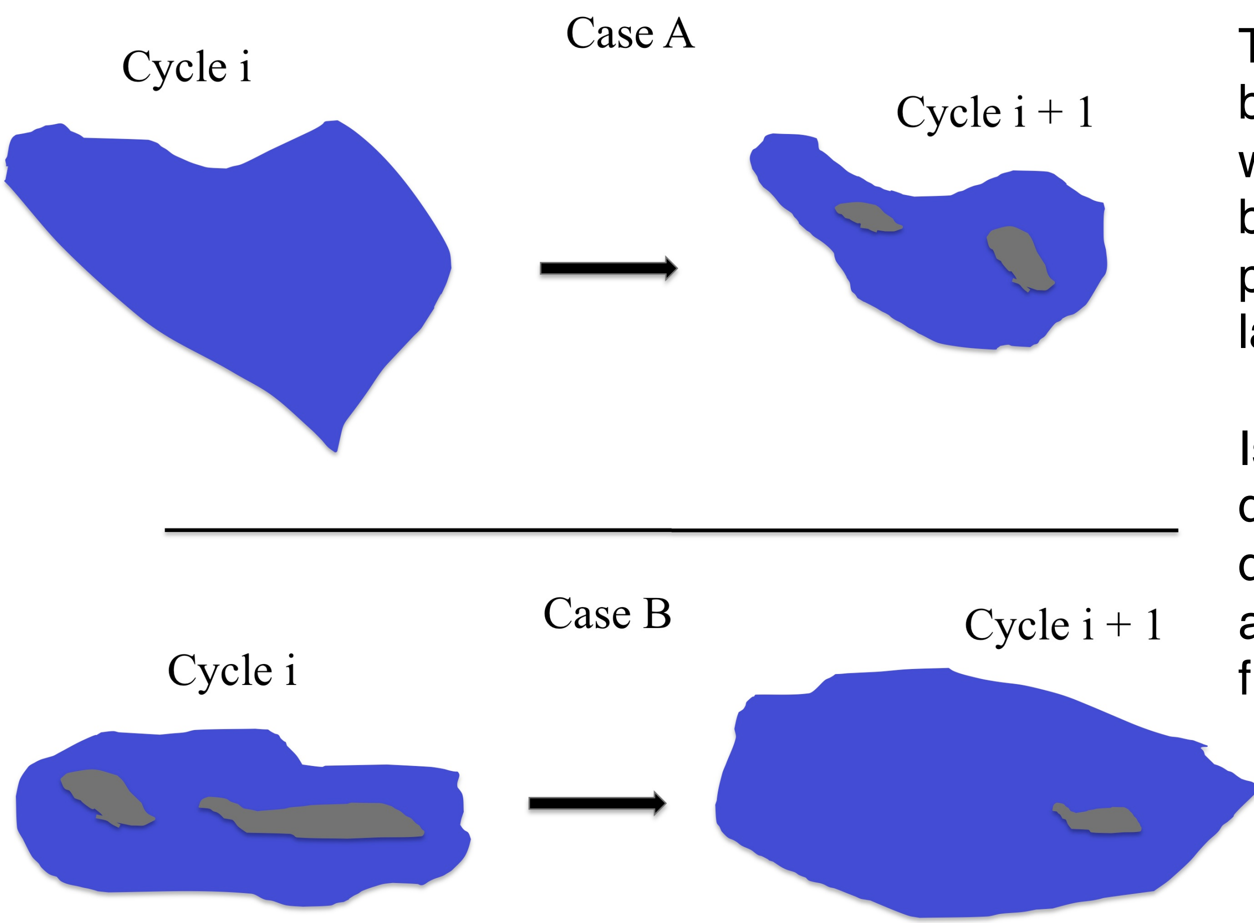
- **--Low detection threshold for lakes in a priori mask**
- **--Higher threshold for lakes outside mask**
- **--Ice flag calculated from SWOT and A Priori Information**



Pass-Based Lake Vector Product

Islands in lakes

Many lakes have islands, and some of these islands will appear and disappear depending on water level.



The full lake boundary will be stored as one polygon, while island boundaries will be stored as separate polygons in a different layer.

Islands will be detected directly from the SWOT data, but also taking advantage of information from the a priori lake mask.

Pass-based Lake Vector Product

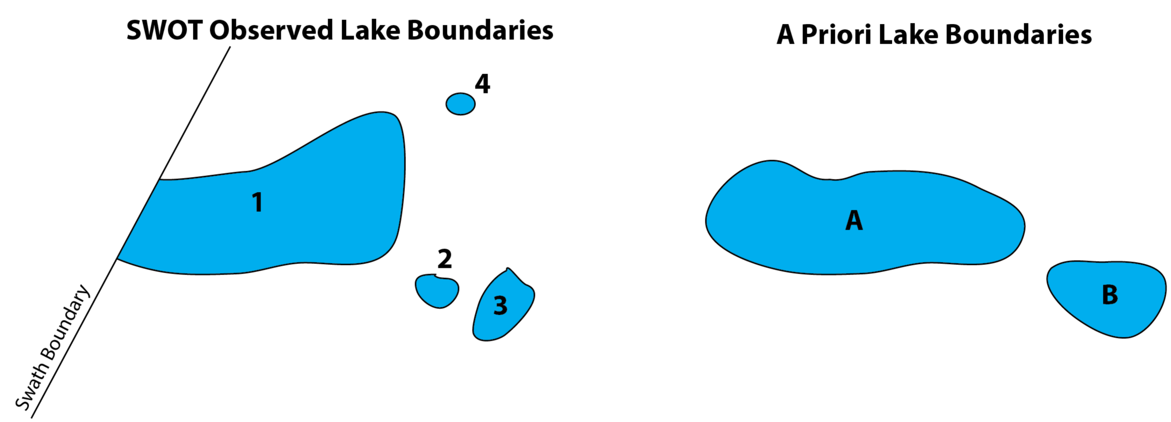
Begin with A Priori Mask and SWOT Pixel Cloud [Slant Range Image]

↓

Classify lake regions in SWOT Data
- --Low detection threshold for lakes in a priori mask
- --Higher threshold for lakes outside mask
- --Ice flag calculated from SWOT and A Priori Information

↓

Link SWOT lake regions to lakes in a priori mask
- --Flag SWOT lake regions not completely covered by swath
- --Flag Lake Regions that cover more than 1 a priori lake
  - --Assign ID for largest intersected a priori lake to SWOT regions



Pass-based Lake Vector Product

Begin with A Priori Mask and SWOT Pixel Cloud [Slant Range Image]

↓

Classify lake regions in SWOT Data
- --Low detection threshold for lakes in a priori mask
- --Higher threshold for lakes outside mask
- --Ice flag calculated from SWOT and A Priori Information

↓

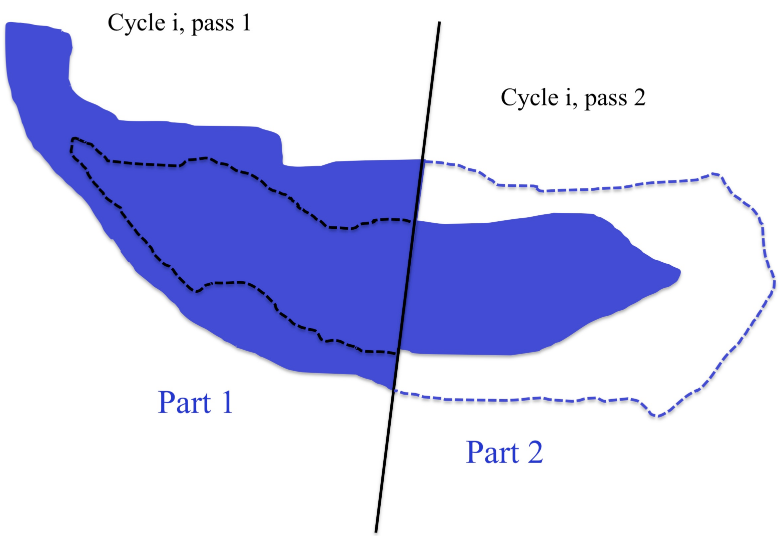
Link SWOT lake regions to lakes in a priori mask
- --Flag SWOT lake regions not completely covered by swath
- --Flag Lake Regions that cover more than 1 a priori lake
  - --Assign ID for largest intersected a priori lake to SWOT region

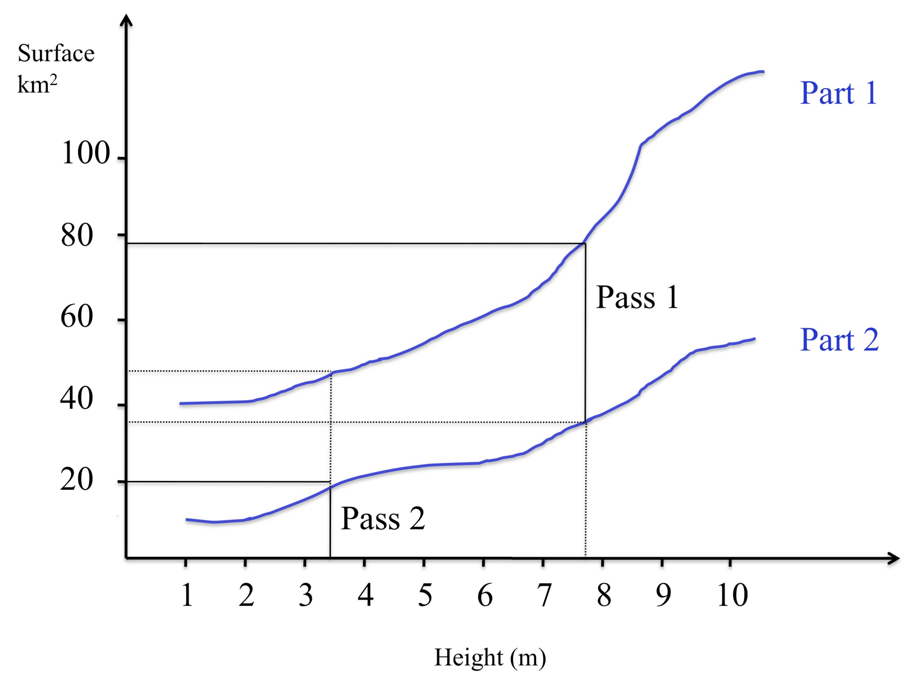
↓

Calculate lake area & uncertainty
Modify lake boundaries to match lake area
Calculate lake height & uncertainty

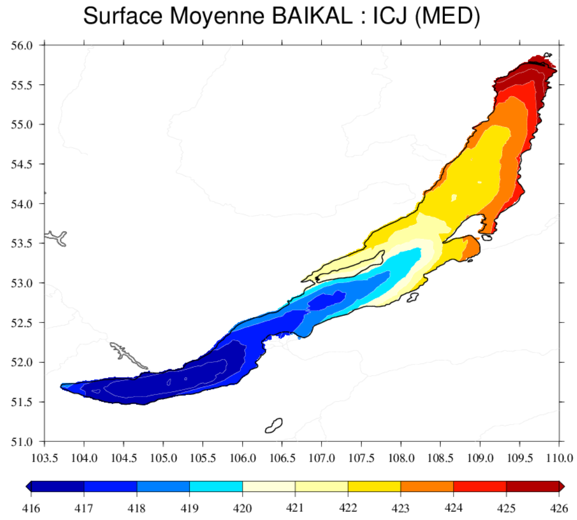
Lake Area from Partially Observed Lakes

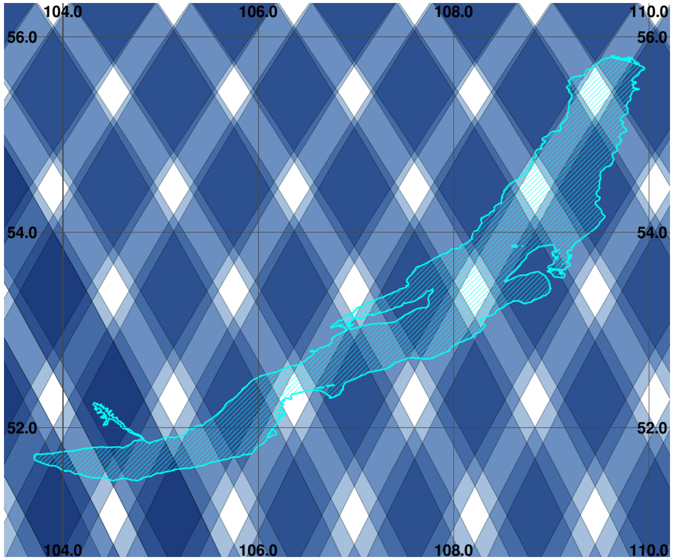
- Many lakes will be only partially observed in any given swath.
- Rating curves be developed between height and area for each individual swath.
- Thus, when height is observed in Part 1, total lake area can be estimated by summing the measured area in Part 1 and the estimated area from the rating curve for Part 2.

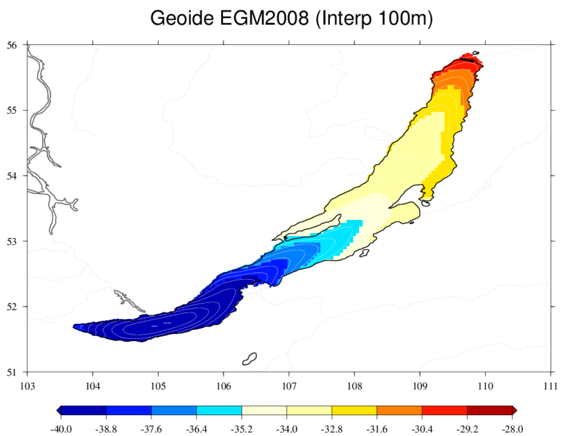


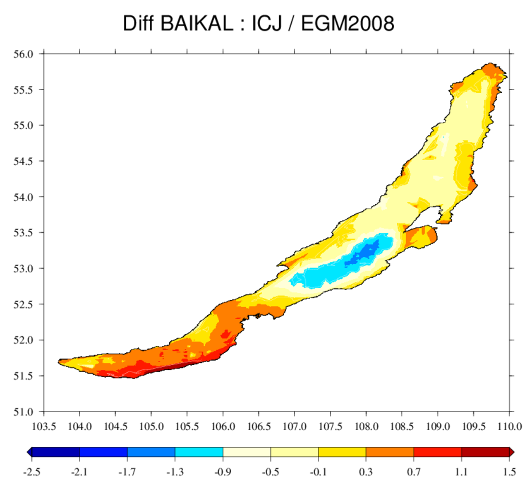


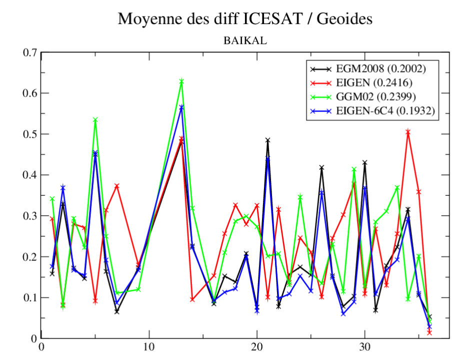
Geoid error corrections is crucial









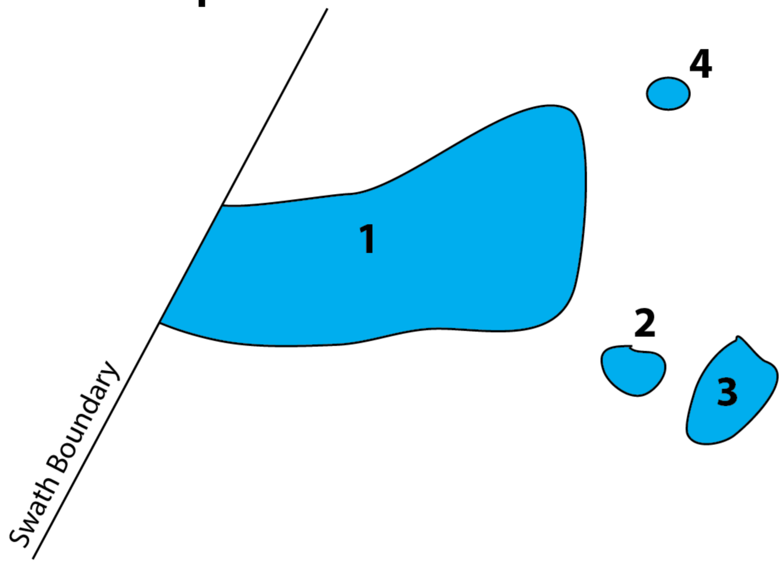


Pass-Based Lake Vector Product

Attributes:

- Height based on the average height of inundated pixels (not including laid-over pixels), relative to best known geoid.
- Area value defined as the total observed area for the SWOT-observed region
- A no-layover area value for the total area used to calculate the height, and various flags (ice, rain, etc.)
- Flag indicating whether the a priori lake area falls entirely within the swath
- Flag indicating if the SWOT-observed region intersects multiple a priori lakes.

Spatial Data for Swath Lake Product



Attribute Table:

SWOT ID	Apriori ID	Obs. Area	Obs. Area Error	Height	Height Error	Ice Flag	Multi Flag	Partial Flag
1	A	22	0.7	144.32	0.01	0	0	1
2	B	1.9	0.3	143.11	0.05	0	0	0
3	B	3.5	0.4	143.25	0.03	0	0	0
4	No	0.8	0.2	150.65	0.12	0	0	0

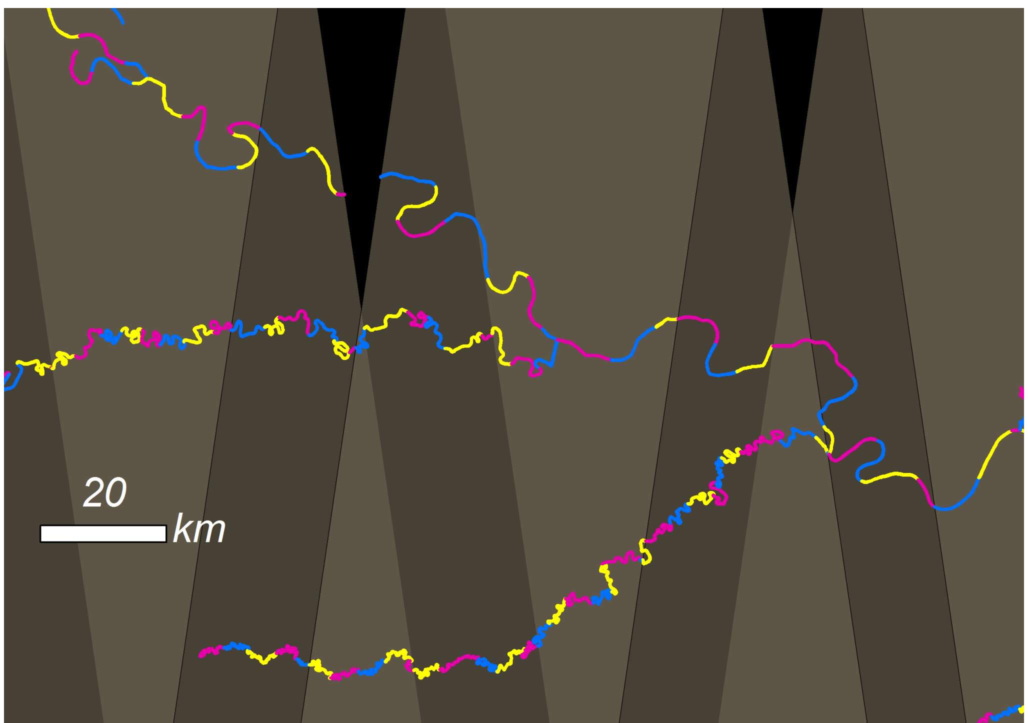
Pass-Based Lake Vector Product

- Lake boundaries will vary from pass to pass and will be restricted to match the calculated area value for each pass.
- Lake heights will be calculated using whatever portion of the lake is observed, and an associated error will be calculated.
- Ice flags will be calculated using SWOT data informed by a priori information (e.g. ice will never be assigned to lakes in the central Amazon basin).
- Initially, no storage change value will be produced for the pass-based product. We propose that after at least a year, the science team produce a series of height-storage anomaly rating curves based on the cycle-based product described below.
  - With these rating curves, it will be possible to estimate storage change in the swath-based product even for lakes that are not completely observed.

A Priori River Information

The a priori river dataset will be modified during the mission as rivers move and change in planform.

How to handle this kind of multitemporal information will be the subject of an algorithm team meeting in September.



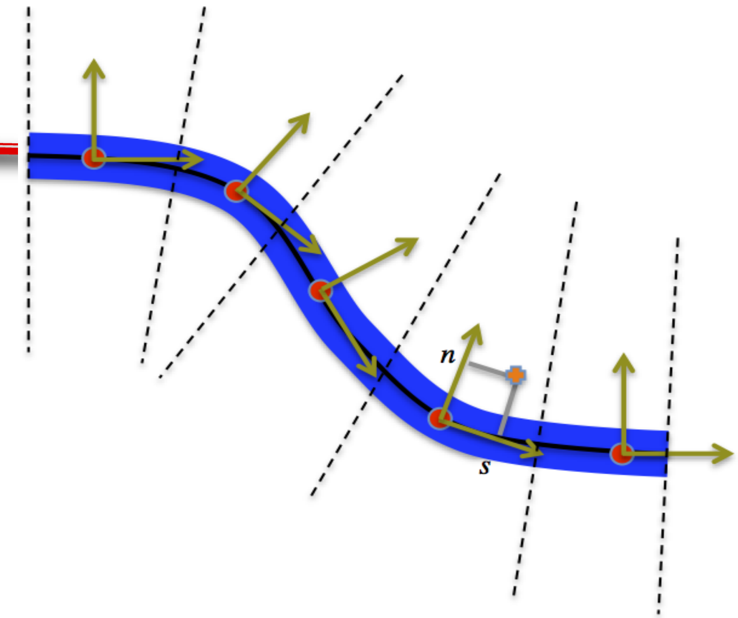
Goal: Provide a system to allow dynamic recalculation of a priori river reach boundaries that can take in a large number of different variables, including:

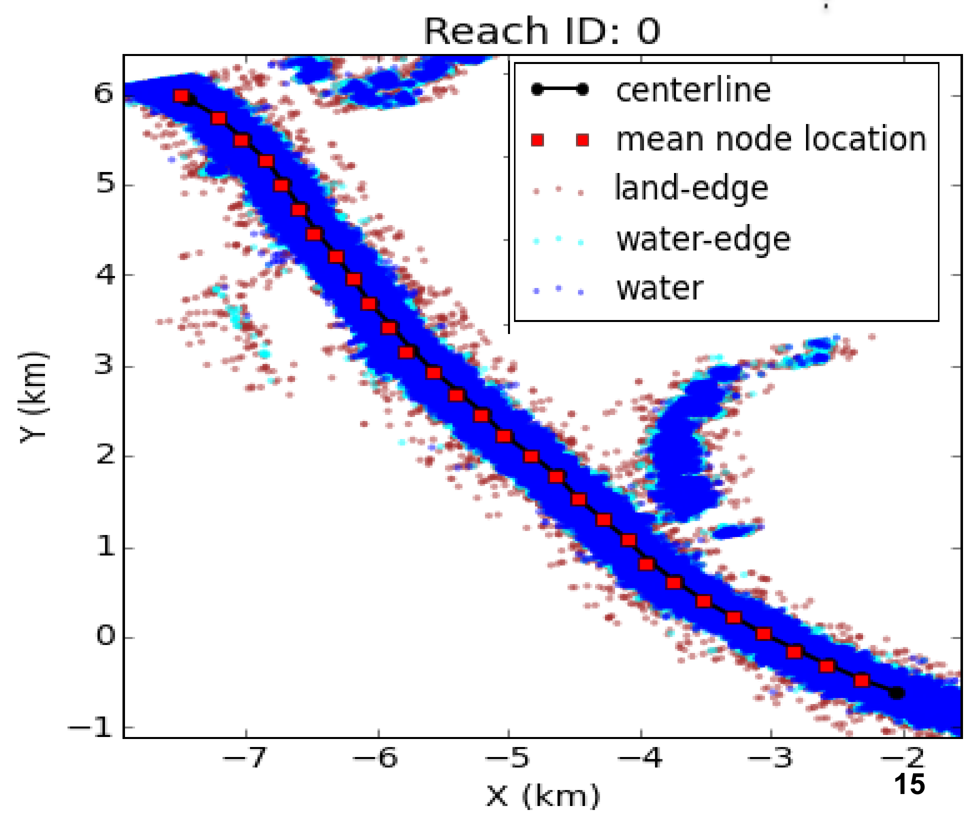
1. River centerline location
2. River width
3. River slope
4. Tributary junctions/changes in basin area
5. Braiding index
6. Reservoir/Lake extents
7. Dam locations
8. Stream gauge locations
9. SWOT swath boundaries
10. Sinuoisity

Etc, etc.

RiverObs

- Takes pixel cloud and estimates quantities over sub-reaches (nodes)
- Uses center-line and reach definitions from prior database
- Center-line and nodes
  - Associates pixels to closest node (exclude those beyond some threshold)
  - Assigns an along-river and cross-river coordinate
  - Computes node-level metrics: width and height
    - ♦ Width estimated 3 different ways
      - Based on total pixel area (more robust to geolocation errors)
      - Based on STD around centerline
      - Based on maximum distance (Peak-to-peak)
  - Estimates reach slope using nodes
- Outputs reach-level vector product
  - Reach length, slope, avg. height
  - Center-line location (mean node coordinates)
  - Other useful quantities (viewing geometry, inundated area…)





Pass-based River Product Attributes (1/4)

Attributes Included in Pass-Based River Vector Products:

1. River centerline: calculated on a per-pass basis from the SWOT pixel cloud.
2. Time-averaged centerline: to be calculated after the first year of operation using an algorithm defined by the ADT water classification WG and the hydrology product WG. To be used following the first year of operation as the default option for distribution of SWOT timeseries. Methods and details for this will be clarified in the functional flow.
3. Landsat-based centerline: this will be used during the first year of operation as the default option for distribution of SWOT timeseries.

Pass-based River Product Attributes (2/4)

4. Reach ID: should be unique and constant in time. The ID should be informative (e.g. first digits will identify a basin endorheic/exorheic, then next digits a lower strahler order or any other classification). A database of reach IDs will be provided by the Science Team. Reach ID will also be assigned for ephemeral rivers.
5. Reach length [m]
6. Reach-averaged height [m]. Heights will be calculated to the best-available geoid
7. Reach-averaged height uncertainty [m]
8. Reach-averaged slope [cm/km]
9. Reach-averaged slope uncertainty [cm/km]
10. Reach-averaged inundated area [m^2]. Note that reach average width can be obtained by dividing inundated area by reach length.
11. Reach-averaged inundated area uncertainty [m^2]

Pass-based River Product Attributes (3/4)

12. Reach-averaged discharge [m^3/s]. Different discharge values would be computed and provided for several different algorithms.
13. Reach-averaged discharge uncertainty [m^3/s]. To be computed as described in the discharge error budget.
14. Parameters used to compute reach-averaged discharge (time-invariant). These will be provided for each discharge algorithm.
15. Island detection flag. Four states are envisioned: 0: islands are definitely present; 1: islands might be present; 2: if there are no islands; 3: presence of islands unknown. It is not known whether SWOT data can reliably detect the presence of islands (especially sand bars). Thus, this classification will come from a priori datasets, viz. GRWL.
16. River planform classification: This classification could follow the same scheme as used here for data product specification: single channel, simple multi-channel, braided or anastomosing, and river with floodplain interaction. This will be based on a combination of Landsat scenes and SWOT data.

Pass-based River Product Attributes (4/4)

17. Quality and other flags (snow/ice/high rate precipitation impacting the signal). Corrections
    - Wet troposphere delay
    - Ionosphere delay
    - Roll correction
18. Connectivity
    - Upstream reach ID(s). A special code will indicate upstream-most reaches.
    - Downstream reach ID(s). A special code will be included for rivers emptying into the ocean.
19. Spatial
    - Distance from outlet [km].
    - Cross-track distance (with regard to the reach center) [m]
20. Layover flag. Three states are envisioned: 0: there is no layover; 1: there is layover; 2: there might be layover.

How do we produce cycle-based products?

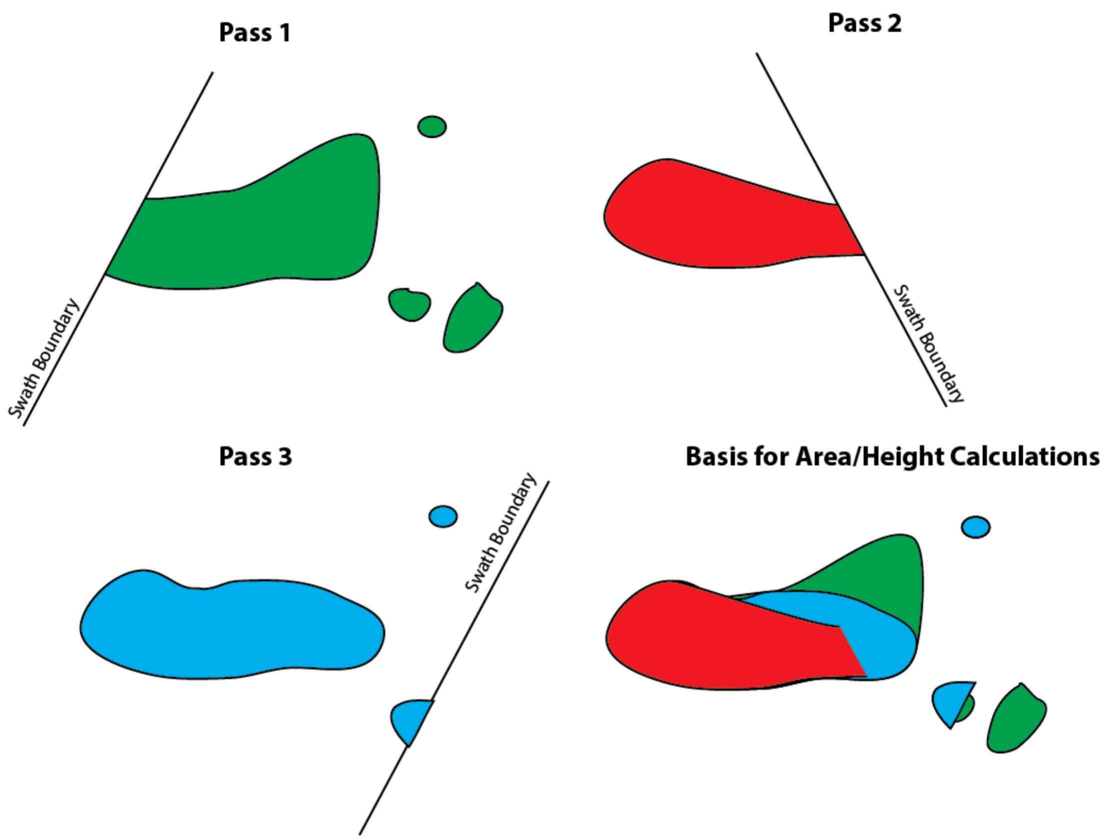
- First question: do we use pass-based vector products, or do we use the pixel cloud data?
  - Pixel cloud data may allow more intelligent algorithms
- We have several requirements:
  - Algorithms must be fully automated and must produce sensible data very close to 100% of the time.
  - Multitemporal data must be available for individual water bodies; a user must be able to download a time series for a lake or river reach.
  - Boundaries of water bodies must be able to vary from cycle to cycle.
- Development of a cycle-based product from one or more overpasses is often non-trivial. We can imagine at least 7 different cases, which we will explore here.

Cycle Lake Vector Product

Begin with A Priori Mask and all pixel cloud data for cycle [pass-based product?]

↓

Calculate merged lake regions based on stacked pass-by-pass information



Cycle Lake Vector Product

- Storage change will be computed relative to a baseline height-area pairing defined by the first SWOT cycle, at least until the first reprocessing.
- How to deal with lakes that are separate in the a priori mask but that become connected remains TBD.
- For lakes with partially diffuse or indeterminate boundaries (e.g. patchy wetlands), a best effort will be made to determine a suitable area based on the contour of 50% inundation. Efforts will be made to factor the resulting uncertainty into the error accounting.
- All calculated values will be assigned to a database associated with the *a priori* mask. In other words, every 21 days, each lake in the mask will have a new height, area, and storage change value (with associated errors and flags) added to its associated database.
- The *a priori* mask will be updated regularly (where the meaning of regularly is TBD), and upon update storage change values for any added lakes will be computed for all prior observations.

Case 1: A Single, Complete Observation

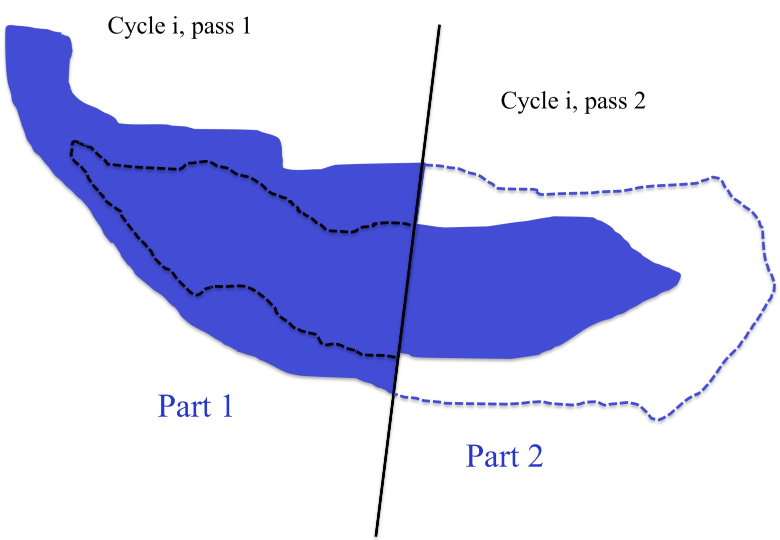
- In Case 1, a water body is observed only once during the cycle, and it is observed in its entirety.
- Likely to occur in tropical and some mid-latitude locations; small lakes and many river reaches.
- The cycle-based vector is identical to the pass-based vector in this case, and is thus (relatively) simple to produce.

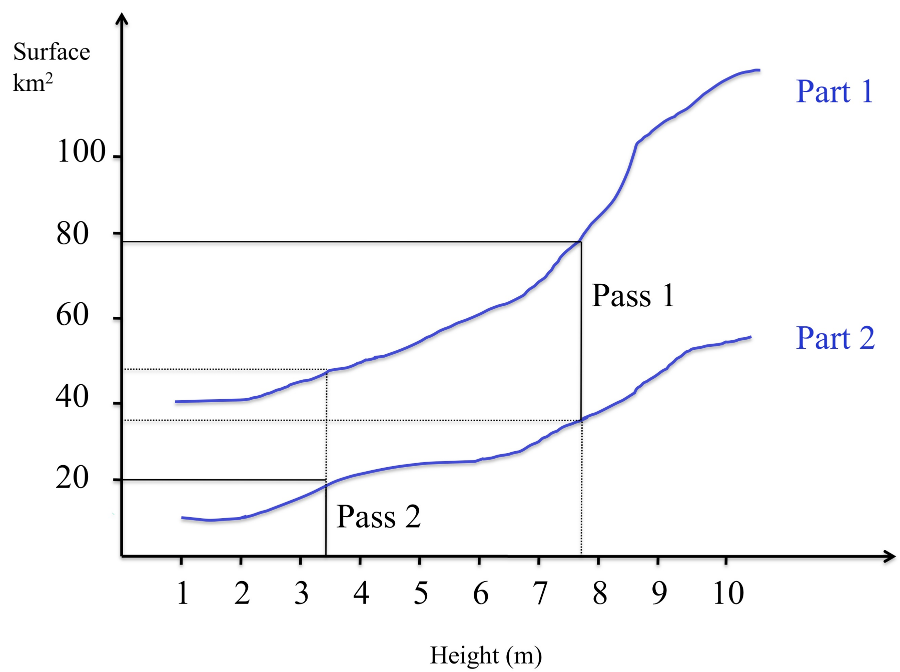
Case 2: Multiple Complete Observations

- In Case 2, a water body is observed in its entirety two or more times during a 21-day cycle.
- Simplest algorithm: take the average height of all pixels within the water body. Determine water body boundaries using the same method used for pass-based data.
  - Doesn't allow intelligent handling of data quality and errors.
  - For example, if on one overpass a water body is heavily laid over, while on another it is not, we don't want to treat these two observations identically.
- An intelligent algorithm has not yet been defined to combine data from multiple SWOT overpasses to produce one vector product.

Case 3: Multiple Observations, Some or All Incomplete

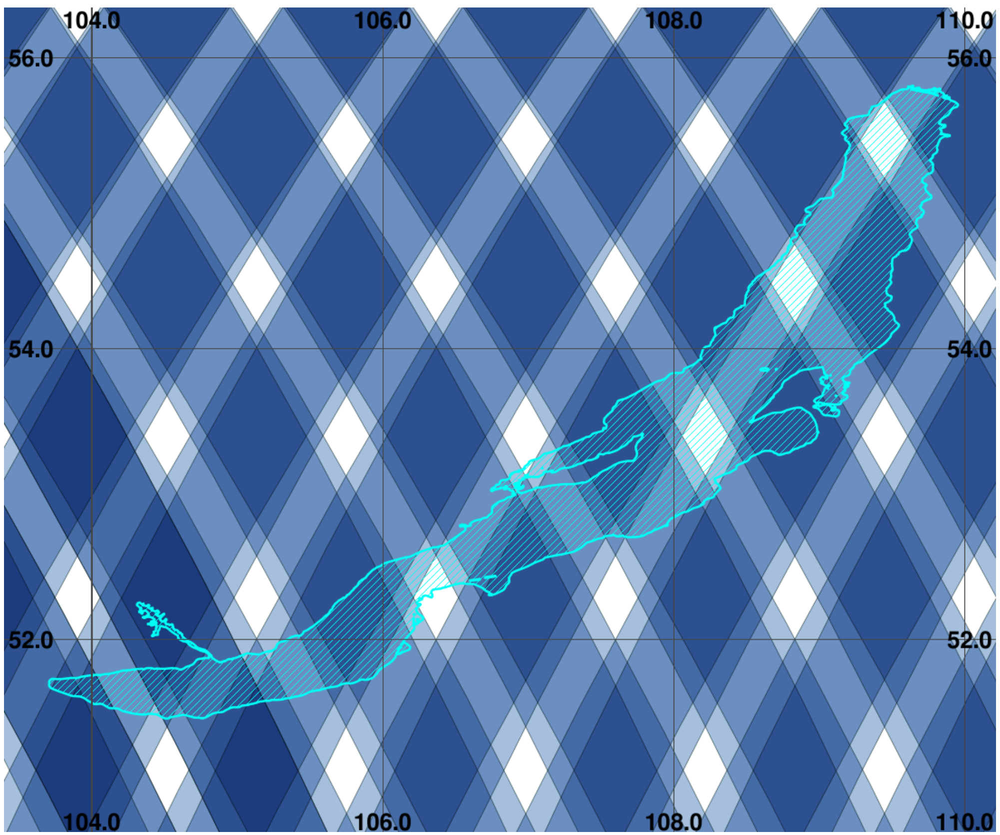
- In Case 3, some of the observations are only partial due to SWOT coverage, rain, dark water, etc.
- One possible solution: develop rating curves like those used in pass-based products.
  - Calculate average lake height for each pass based on observations, then average these averages.





Case 4: Water Body is Never Fully Observed

In some cases, lakes (in particular) will fall partially within the SWOT mask and partially outside. This will occur most often for large lakes, but will sometimes occur for small lakes.

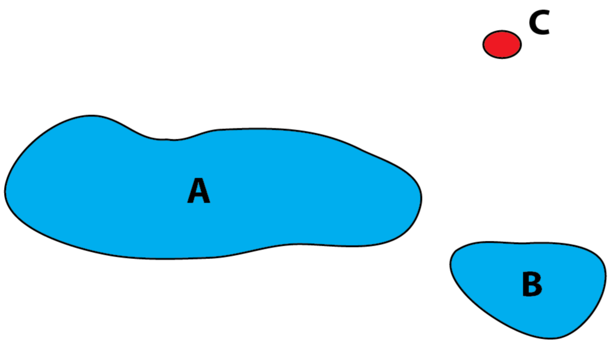


For small lakes, we propose to simply use the a priori lake mask area.

For large lakes, we may be able to merge information from the a priori mask with

Case 6: New Water Body Appears

- Add new provisional ID
- Calculate height, area, slope, etc.
- Do not immediately calculate storage change, discharge, etc.
- At next reprocessing, go back and fully merge into database.



Lake A: 24 km^2
Lake B: 7.5 km^2
Lake C: 0.8 km^2

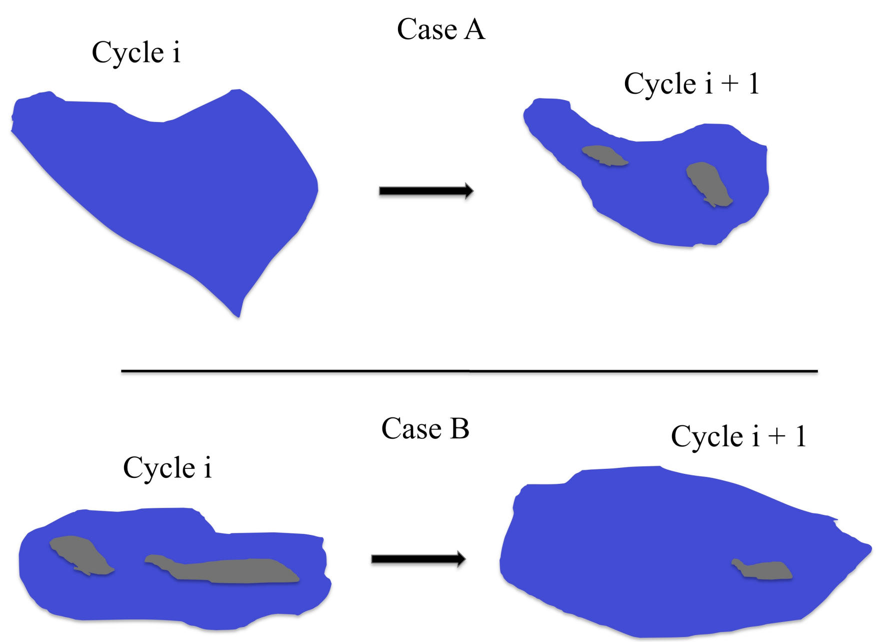
Lakes A & B in a priori mask
Lake C not in a priori mask

Case 6: Changing Water Body Topology

- Water bodies that are separate in one SWOT overpass are merged in another overpass.
    - In this case, we will consult the a priori database.
    - If one water body in the a priori database is split into two water bodies in one or more images, then we will simply count all of the individual water bodies within the one larger a priori water body.
    - If two or more a priori water bodies merge into one water body, then we will provide the average height of all water bodies. It is unclear what we should do with area.
- This circumstance needs to be flagged in the final database.

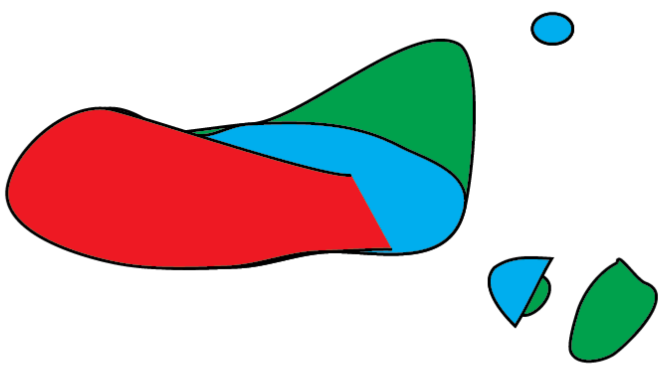
Case 7: Islands in lakes

- As with pass-based products, we will need to have a separate vector layer that shows the extent of islands.
- Key question: what to do if islands are only present in some overpasses due to rising/falling lake levels?
- For now, we will include polygons showing maximum extent of islands. This could be substantially improved.

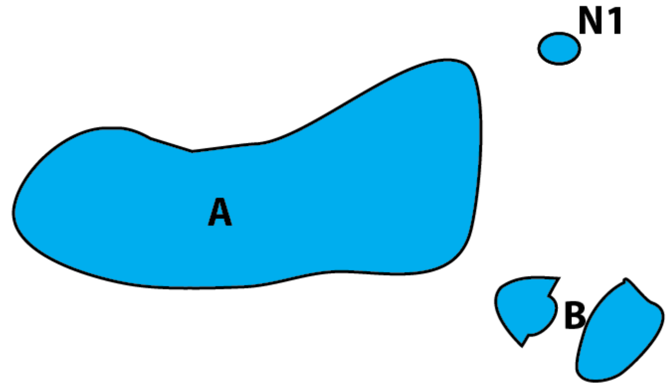


Cycle Lake Vector Product

Basis for Area/Height Calculations



Final Product Appearance



Final Product Attribute Table

Apriori ID	Obs. Area (km^2)	Obs. Area Error	Height (m)	Height Error	Storage Δ (m^3)	Storage Error	Ice Flag	Multi Flag
A	28	1.7	144.32	0.01	10920000	17000	0	0
N1	0.8	0.2	150.65	0.12			0	0
B	3.8	0.6	143.11	0.05	2415000	30000	0	0

Queryable Water Body Databases

- It is essential that scientists be able to obtain multitemporal SWOT height, area, and storage change values (along with associated flags) on all lakes & river reaches in the a priori masks WITHOUT downloading many separate vector files.
- It should also be possible to obtain vectors for an arbitrarily defined region.
- These goals are likely to be best achieved using a web-based tool that allows users to specify the region to be queried, the length of time, and which lake/river attributes should be output.
- This tool should be applicable to both the pass-based product and the cycle-based product.

Backup Slides

What Is a Lake?

- A **Lake** is provisionally defined as:
  - any water body large enough to be observed by SWOT (> (100-250 m)^2)
  - not substantially vegetated*
  - has a discrete and identifiable shoreline
  - Does not exhibit characteristics normally associated with rivers (e.g. persistent slope not forced by wind, channelized flow).