

ABoVE: Alaska Lake and Pond Occurrence

Documentation Revision Date: 2025-03-09

Dataset Version: 1

Summary

The Alaska Lake and Pond Occurrence Dataset (ALPOD) is a spatially explicit map of lakes and ponds across Alaska and their seasonal fluctuations. The core product is an open water occurrence raster that: (a) separates lakes and ponds from other components of the landscape (e.g., rivers and wetlands); (b) is built from Sentinel-2 imagery and has 10-m resolution; and (c) records the percentages of time that each pixel was open water and attached to a lake or pond during the 2016-2021 ice-free seasons at near-daily temporal resolution. The number of water bodies depends on the chosen occurrence threshold, but a conservative estimate is that ALPOD maps over 800,000 lakes and ponds larger than 0.001 km². The lake occurrence rasters are tiled by UTM zone and latitude. ALPOD also includes a vector product defined using a 25% occurrence threshold. ALPOD was created using a U-Net lake identification model and manual inspection to produce a maximum possible lake extent mask. This mask serves as the region of interest for an adaptive NDWI threshold water classification algorithm written in Google Earth Engine (GEE), which was used to classify open water within the maximum lake mask in every available cloud- and ice-free Sentinel-2 image during the study period. ALPOD is suitable for investigations of individual water bodies as well as lake and pond patterns across Alaska. The data are provided in GeoTIFF and shapefile formats.

This dataset includes eight files in cloud optimized GeoTIFF format and one shapefile in a compressed Zip file.

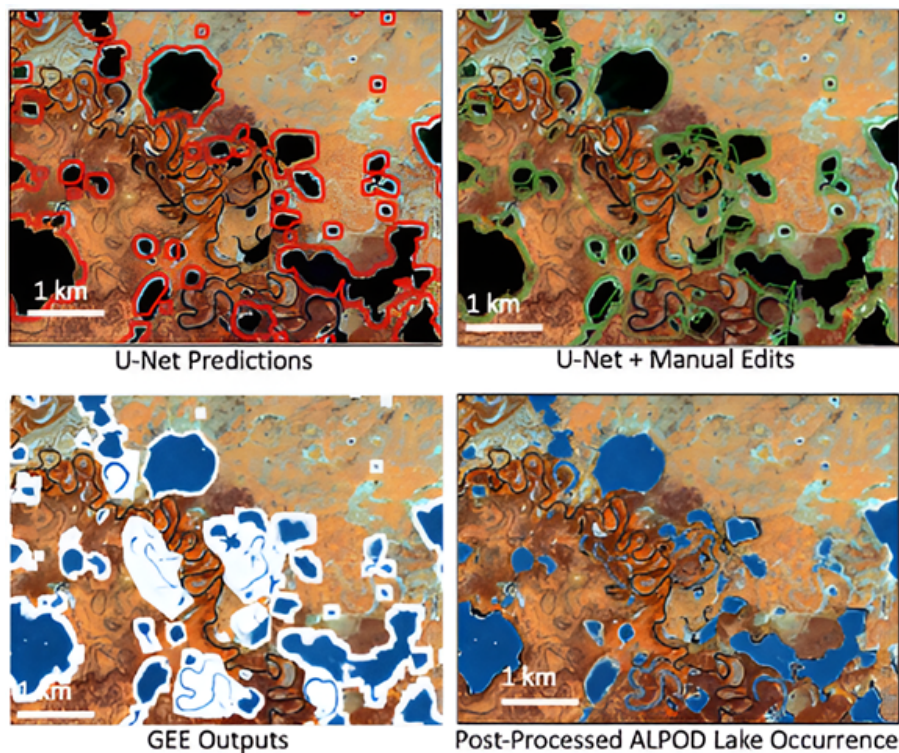


Figure 1. Dataset construction progression starting with the U-Net predictions (upper left). Manual edits to the U-Net outputs produce the maximum possible lake extent mask (upper right). The lower left panel shows the output of the Google Earth Engine (GEE) processing that classifies open water in every available image within the maximum lake mask. Lastly, the final lake occurrence image is produced by applying a watershed segmentation algorithm to the GEE outputs. Note the areas within the maximum possible lake extent mask (green) that are excluded from the final lake occurrence dataset. These areas had low open-water occurrence values and were filtered by the watershed segmentation post-processing algorithm.

Citation

Levenson, E., S.W. Cooley, and A. Mullen. 2025. ABoVE: Alaska Lake and Pond Occurrence. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2399>

Table of Contents

1. Dataset Overview
2. Data Characteristics
3. Application and Derivation
4. Quality Assessment
5. Data Acquisition, Materials, and Methods

1. Dataset Overview

The Alaska Lake and Pond Occurrence Dataset (ALPOD) is a spatially explicit map of lakes and ponds across Alaska and their seasonal fluctuations. ALPOD is a new dataset of lakes and ponds larger than 0.001 km² produced for the state of Alaska. The core product is an open water occurrence raster that: (a) separates lakes and ponds from other components of the landscape (e.g., rivers and wetlands); (b) is built from Sentinel-2 imagery and has 10-m resolution; and (c) records the percentages of time that each pixel was open water and attached to a lake or pond during the 2016-2021 ice-free seasons at near-daily temporal resolution.

The number of water bodies depends on the chosen occurrence threshold, but a conservative estimate is that ALPOD maps over 800,000 lakes and ponds larger than 0.001 km². The lake occurrence rasters are tiled by UTM zone and latitude. ALPOD also includes a vector product defined using a 25% occurrence threshold. ALPOD was created using a U-Net lake identification model and manual inspection to produce a maximum possible lake extent mask. This mask serves as the region of interest for an adaptive NDWI threshold water classification algorithm written in Google Earth Engine (GEE), which was used to classify open water within the maximum lake mask in every available cloud and ice-free Sentinel-2 image during the study period. ALPOD is suitable for investigations of individual water bodies as well as lake and pond patterns across Alaska.

Project: [Arctic-Boreal Vulnerability Experiment](#)

The Arctic-Boreal Vulnerability Experiment (ABoVE) is a NASA Terrestrial Ecology Program field campaign being conducted in Alaska and western Canada, for 8 to 10 years, starting in 2015. Research for ABoVE links field-based, process-level studies with geospatial data products derived from airborne and satellite sensors, providing a foundation for improving the analysis, and modeling capabilities needed to understand and predict ecosystem responses to, and societal implications of, climate change in the Arctic and Boreal regions.

Related Publication

Levenson, E. S., Cooley, S., Mullen, A., Webb, E. E., & Watts, J. (2025). Glacial history modifies permafrost controls on the distribution of lakes and ponds. *Geophysical Research Letters*, 52(4), e2024GL112771. <https://doi.org/10.1029/2024GL112771>

Acknowledgments

This work was supported by the NASA New Investigators Program (grant 80NSSC21K0920) and Future Investigators in Earth and Space Sciences and Technology (grant 80NSSC23K1556). A. Mullen and J. Watts acknowledge funding from NASA CSDA (grant 80NSSC21K1161). E. Webb is supported by the National Science Foundation (grant 2317873).

2. Data Characteristics

Spatial Coverage: Alaska

ABoVE Reference Locations:

Domain: Core and extended
State/Territory: Alaska
Grid Cells: Most "A" cells in Alaska

Spatial Resolution: 10 m

Temporal Coverage: 2016-05-01 to 2021-09-30

Temporal Resolution: one-time estimate

Site Boundaries: Latitude and longitude are given in decimal degrees.

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
Alask	-179.36957	-129.22095	71.6586	50.98534

Data File Information

This dataset includes eight files in GeoTIFF format and one shapefile in a compressed Zip file.

Properties of the GeoTiff files:

- coordinate system: geographic coordinates using WGS84 datum (EPSG: 4326)
- spatial resolution: 0.00008983 x 0.00008983 degrees (approximately 10 x 10 m)
- pixel values: 0 to 100, units = percent
- no data value = 255
- bands = 1

Properties of the shapefile:

- coordinate system: NAD 83 / Alaska Albers; unit = m ; EPSG 3338

Table 1. Data Files and Descriptions

File Name	Descriptions
AK<#>N_LakeOccurrence.tif	Six cloud optimized GeoTIFF (.tif) files where <#> is the UTM zone number, 2-7. Each pixel represents the percentage of time that location was open water and attached to a lake or pond during the 2016-2021 ice-free seasons.
AKSouth_LakeOccurrence_a.tif, AKSouth_LakeOccurrence_b.tif	The files provide everything in Alaska south of 60 degrees. The longitudinal border separating the "a" and "b" tiles is arbitrary and exists to limit the maximum file size of the GeoTiffs.
ALPODlakes.shp	One shapefile contained in a compressed Zip archive (ALPODlakes.zip) that provides the area (m ²) and perimeter (m) of the individual lakes.

3. Application and Derivation

This novel high-resolution dataset of the seasonal wetting and drying patterns of Alaskan lakes and ponds enables new investigations into hydrologic and ecological change, biogeochemical cycling, and lake permafrost interactions across Alaska.

4. Quality Assessment

In order to assess the accuracy and precision of ALPOD, the study area was split into 25 km² tiles. 100 tiles were randomly sampled and the lake dataset was compared to the ESRI base map imagery, which is primarily constructed from very high-resolution aerial and satellite imagery. An uncertainty assessment showed that ALPOD's F1 score for classifying lakes and ponds larger than 0.001 km² was 0.999 which suggests that ALPOD reliably maps lake extents. An analysis of the spatial patterns of lake abundance and size uses a 25% occurrence threshold (i.e., open water in at least 25% of observations) to include intermittently inundated depressions, as these may be useful for understanding the potential mechanisms explaining the distribution of lakes. Although the occurrence dataset includes ponds as small as 0.0001 km², water bodies smaller than 0.001 km² were more likely to be omitted or false positives. Therefore, water bodies smaller than 0.001 km² were excluded. After the low-end truncation, the surface area of errors of omission and commission was manually evaluated and recorded.

5. Data Acquisition, Materials, and Methods

To create ALPOD (refer to Figures 1-2), a mask was produced of the maximum potential lake inundation extent that excludes river channels and persistently dry areas. Sentinel-2 images were accessed for every location in the study area, prioritizing cloud-free images captured near the seasonal maximum water extent based on visual inspection. Next, lake locations were predicted in these images using a U-Net lake identifier model (Mullen et al., 2023). The U-Net model is a convolutional neural network that represents features such as edges and frequencies across scales, which is advantageous for separating lakes from components of the landscape with similar reflectance properties, such as rivers. The predictions were buffered by 60 meters and the U-Net predictions were manually edited in QGIS by visualizing the shapefile over the previously downloaded Sentinel-2 imagery. Then, the U-Net results were merged with the Prior Lake Database (Wang et al., 2023) to ensure that the largest lakes were included in the mask. At this stage, both errors of omission and commission were fixed by deleting misclassified water bodies and producing polygons to include missed lakes. This manual edit was conducted at a consistent scale of 1:50,000 to finalize the maximum possible lake extent mask before uploading it as an asset to Google Earth Engine (GEE) for water classification.

The lake occurrence dataset was produced using a water classification algorithm based on adaptive NDWI thresholding (Cooley et al., 2017; 2019). First, the study site was split into 160 km² tiles to run a water classification script. For a given tile, a collection of cloud-masked Sentinel-2 mosaics were produced for every available date during the ice-free months of 2016-2021. The mosaic was clipped to the maximum possible lake extent mask and mosaics were filtered out covering less than 70% of the ROI. The lake mask was vectorized at 25% open-water occurrence threshold and exported to a shapefile. The number of water bodies depends on the chosen occurrence threshold, but a conservative estimate is that ALPOD maps over 800,000 lakes and ponds larger than 0.001 km².

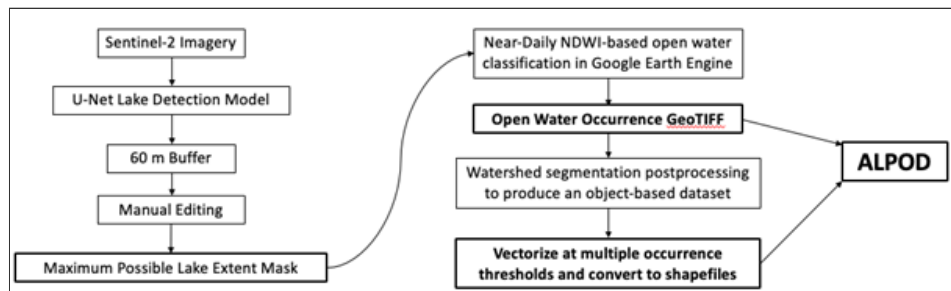


Figure 2. Methodological approach to constructing the ALPOD dataset.

Refer to Levenson et al. (2025 in review) for additional details.

6. References

Cooley, S. W., L. C. Smith, J. C. Ryan, L. H. Pitcher, and T. M. Pavelsky. 2019. Arctic Boreal lake dynamics revealed using CubeSat imagery. *Geophysical Research Letters* 46:2111–2120. <https://doi.org/10.1029/2018GL081584>

Cooley, S., L. Smith, L. Stepan, and J. Mascaro. 2017. Tracking dynamic northern surface water changes with high-frequency Planet CubeSat imagery. *Remote Sensing* 9:1306. <https://doi.org/10.3390/rs9121306>

Levenson, E. S., Cooley, S., Mullen, A., Webb, E. E., & Watts, J. (2025). Glacial history modifies permafrost controls on the distribution of lakes and ponds. *Geophysical Research Letters*, 52(4), e2024GL112771. <https://doi.org/10.1029/2024GL112771>

Mullen, A. L., J. D. Watts, B. M. Rogers, M. L. Carroll, C. D. Elder, J. Noomah, Z. Williams, J. A. Caraballo-Vega, A. Bredder, E. Rickenbaugh, E. Levenson, S. W. Cooley, J. K. Y. Hung, G. Fiske, S. Potter, Y. Yang, C. E. Miller, S. M. Natali, T. A. Douglas, and E. D. Kyzivat. 2023. Using high-resolution satellite imagery and deep learning to track dynamic seasonality in small water bodies. *Geophysical Research Letters* 50:e2022GL102327. <https://doi.org/10.1029/2022GL102327>

Wang, J., C. Pottier, C. Cazals, M. Battude, Y. Sheng, C. Song, M. S. Sikder, X. Yang, L. Ke, M. Gosset, R. R. A. Oliveira, M. Grippa, F. Girard, G. H. Allen, S. Biancamaria, L. Smith, J.-F. Créaux, and T. M. Pavelsky. 2023. The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake mask and operational auxiliaries. (preprint). Authorea; December 14, 2023.
<https://doi.org/10.22541/au.170258987.72387777/v1>