Extraction and Delineation of Lake-Associated Catchments Using HydroBASINS and CCILakes Datasets

# 1. Introduction

In this study, we developed a reproducible geospatial pipeline to associate lakes from the CCILakes dataset with their corresponding contributing catchments, using the HydroBASINS dataset. This work supports downstream hydrological analyses by delineating lake-level drainage areas and optionally identifying full or isolated upstream networks.

# 2. Datasets

Two primary datasets were used:  
- HydroBASINS (North America): Level 8 polygon delineations of sub-basins, derived from HydroSHEDS (Lehner & Grill, 2013). Key attributes include HYBAS\_ID and NEXT\_DOWN.  
- CCILakes (V2.0): A global vector layer of lakes, including Lake\_ID as a unique identifier (Klein et al., 2017).

# 3. Methodology

## 3.1 Assignment of Catchments to Lakes

To associate each lake with its intersecting or nearest catchment(s), we performed the following:  
  
1. Preprocessing: Catchment geometries were cleaned to retain only valid Polygon or MultiPolygon features. Lakes were spatially clipped to the HydroBASINS extent using gpd.clip() to eliminate irrelevant global lakes.  
2. Spatial Matching:  
- A spatial intersection (gpd.sjoin) was used to assign each lake to catchments it intersects.  
- For lakes with no direct intersection, a nearest-neighbor assignment (gpd.sjoin\_nearest) was applied using lake centroids.  
- The original lake geometries were restored following the nearest join.  
3. Data Cleaning and Export:  
- Duplicated associations were removed.  
- Outputs were saved using Fiona, supporting large floating-point Lake\_ID values.

## 3.2 Upstream Catchment Delineation

To trace upstream catchments contributing to each lake:  
  
1. Graph Construction: A reverse-directed graph was built from the NEXT\_DOWN attribute to define the upstream flow network:  
 ReverseGraph[j] = {i | NEXT\_DOWN(i) = j}  
2. Upstream Traversal: For each Lake\_ID, catchments were used as starting nodes. Breadth-first traversal was applied recursively.  
3. Conflict-Aware Tracing: An alternative method halted traversal at catchments linked to other lakes, yielding non-overlapping upstream zones.  
4. Aggregation: Catchments were dissolved into single geometries per lake for further analysis.

# 4. Output

Two key outputs were produced:  
- lake\_catchments\_na\_cleaned.shp: links each lake to its assigned catchments.  
- upstream\_stop\_at\_other\_lakes.gpkg: dissolved upstream catchments per lake, truncated at other lakes.

# 5. Remarks

This pipeline ensures hydrologically consistent assignment of upstream catchments to lakes. The inclusion of both full-network and truncated upstream tracing enables flexible downstream analyses, such as nutrient budgeting, flow modeling, or catchment-based classification.

# References

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world’s large river systems. Hydrological Processes, 27(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Klein, I., Griessinger, N., & Döll, P. (2017). Development and validation of a global dataset of lakes, reservoirs and wetlands. Journal of Hydrology, 555, 365–380. https://doi.org/10.1016/j.jhydrol.2017.10.032