Lake-TopoCat

A global lake drainage topology and catchment database

Version 1.0

Product description

December 10, 2022

Md Safat Sikder (msikder@ksu.edu), Jida Wang (jidawang@ksu.edu), Meng Ding
Kansas State University, Manhattan, KS

Yongwei Sheng
University of California, Los Angeles, CA

George H. Allen
Texas A&M University, College Station, TX

Dai Yamazaki
The University of Tokyo, Tokyo, Japan

Chunqiao Song
Chinese Academy of Sciences, Nanjing, China

Jean-François Crétaux
LEGOS, CNES, Toulouse, France

Tamlin M. Pavelsky
University of North Carolina, Chapel Hill, NC

Contents

1. Background and overview

Lakes store the maximum amount of liquid freshwater on Earth's surface and are an important part of the water cycle. The water mass balance in a lake and the states of its water qualities (e.g., sediment, nutrient, and organic matter concentrations), are affected by surface and subsurface inflows accumulated in the lake catchment. The water storage variation and limnological properties of one lake can also affect those of another via water and energy transfer through the connecting river networks. Therefore, understanding the spatial domain of each lake catchment and how lakes are topologically connected to each other through rivers has important hydrological and ecological significance. Surprisingly, there is no global lake topology and catchment database available to evaluate the hydrologic and ecologic connections between lakes. Here, we introduce the global Lake drainage Topology and Catchment database (Lake-TopoCat), which reveals detailed lake hydrography information with a careful consideration of possible multifurcation. The database was constructed using the HydroLAKES (*Messager et al.*, 2016) lake polygons, and with the help of MERIT Hydro (*Yamazaki et al.*, 2019). The database contains the outlet(s) and catchment(s) of each HydroLAKES lake, all potential drainage connections between the lakes, and a wide suite of attributes that depict detailed lake drainage relationships. This lake topology dataset may facilitate a variety of limnological applications including water quality diagnosis, agriculture and fisheries, lacustrine connectivity monitoring, and integrated lake-river modeling.

2. Input data and methods

The Lake-TopoCat was constructed using the HydroLAKES v1.0 lake and the 3-arc-second-resolution hydrography dataset MERIT Hydro v1.0.1. The HydroLAKES contains about 1.4 million lake polygons larger than 10 ha, which compiled from various sources. MERIT Hydro was developed based on the hydrologically-adjusted MERIT DEM (*Yamazaki et al.*, 2017), where drainage direction was calibrated by multi-source open water maps. We used MERIT Hydro to generate lake outlets, the associated catchments, and the interconnecting drainage paths between the lake outlets (or inland/coastal sinks). The drainage type of each HydroLAKES lake, such as isolated, inflow-headwater, headwater, flow-through, terminal, and coastal (see Section 3), was determined with assistance of MERIT Hydro-Vector (*Lin et al.*, 2021), a high-resolution river network dataset with spatially-variable drainage densities. The scheme of our algorithm is illustrated in Fig. 1.

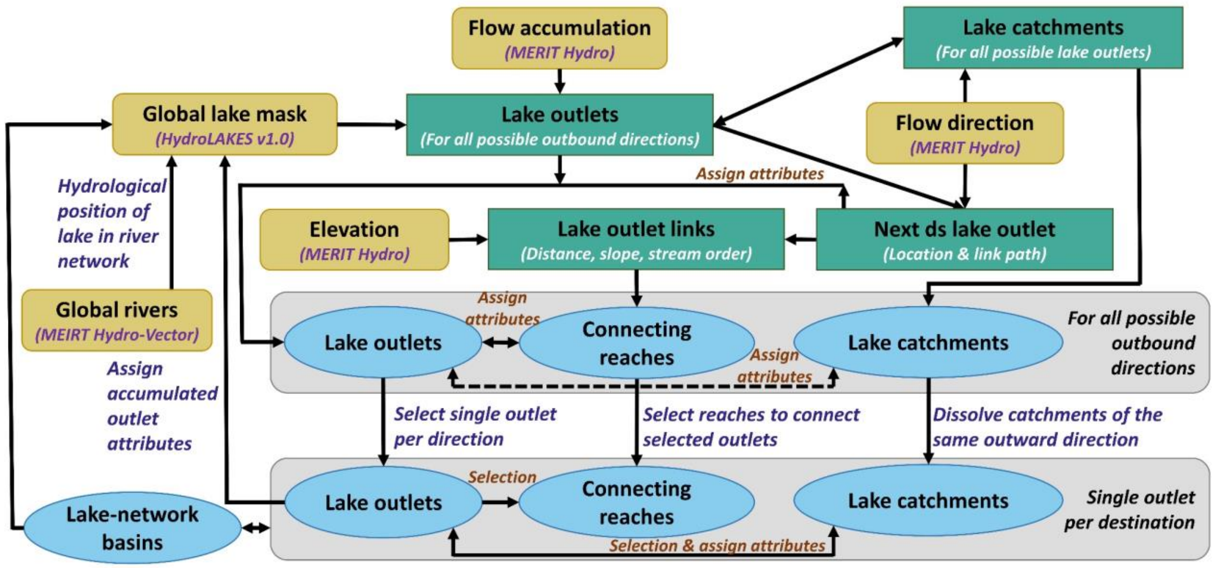

Figure 1. Schematic diagram of the algorithm used to construct the Lake-TopoCat.

We partitioned the global landmass (excluding Antarctica) to 68 basins/regions to enable parallel computing (Fig. 2). These regions follow the boundaries of the 61 Pfafstetter level-2 basins as in MERIT Hydro (*Lin et al.*, 2021), and include another seven regions aggregated from islands and archipelagos across the global oceans. For convenience, our data products were also organized based on these 68 regions (subsets), with their subset IDs shown below (Fig. 2).

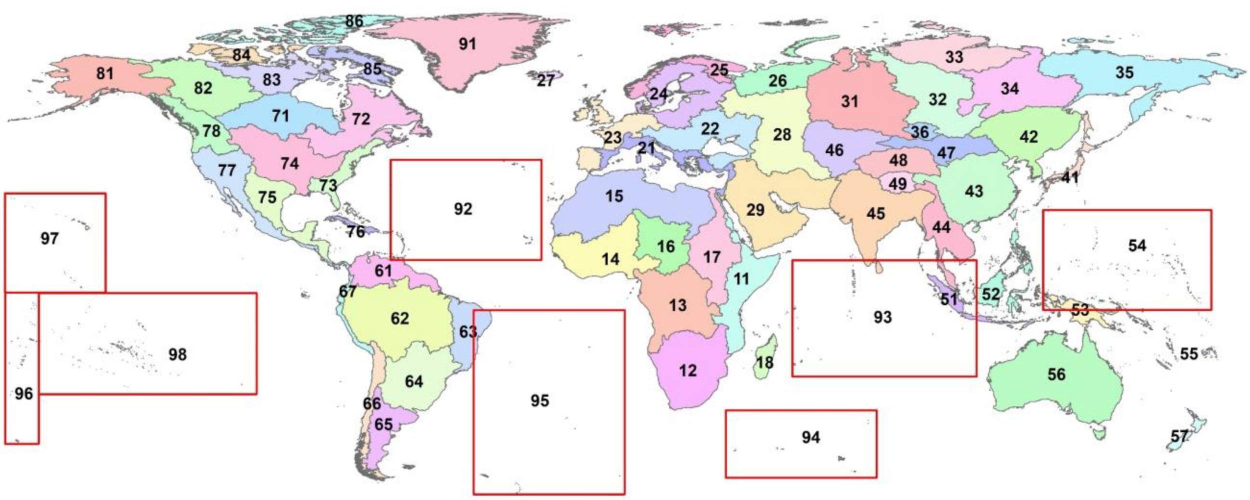

Figure 2. The 68 partitioning regions (Pfafstetter L-2 basins) in the world for Lake-TopoCat production and organization.

3. Product structure and description

Lake-TopoCat consists of five feature components, each with multiple attributes depicting lake drainage relationships. The five features are (1) lake boundaries (polygons), (2) lake outlets (points), (3) unit catchment boundaries (polygons) defining the drainage areas between cascading (i.e., immediately upstream and downstream) lake outlets, (4) inter-lake reaches (lines) defining the drainage networks that connect the lake outlets to the inland sinks or the ocean, and (5) lake-network basins (polygons) that define the entirety of the drainage area containing each inter-lake network (i.e., a complete basin from the headwater to an inland sink or the ocean for all basins containing lakes). The file names of these five components are:

- Lake boundaries: *Lakes_pfaf_xx*
- Lake outlets: *Outlets_pfaf_xx*
- Unit catchments: *Catchments_pfaf_xx*
- Inter-lake reaches: *Reaches_pfaf_xx*
- Lake-network basins: *Basins_pfaf_xx*

Note: "*pfaf_xx*" indicates the Pfafstetter Level-2 basin ID (Fig. 2)

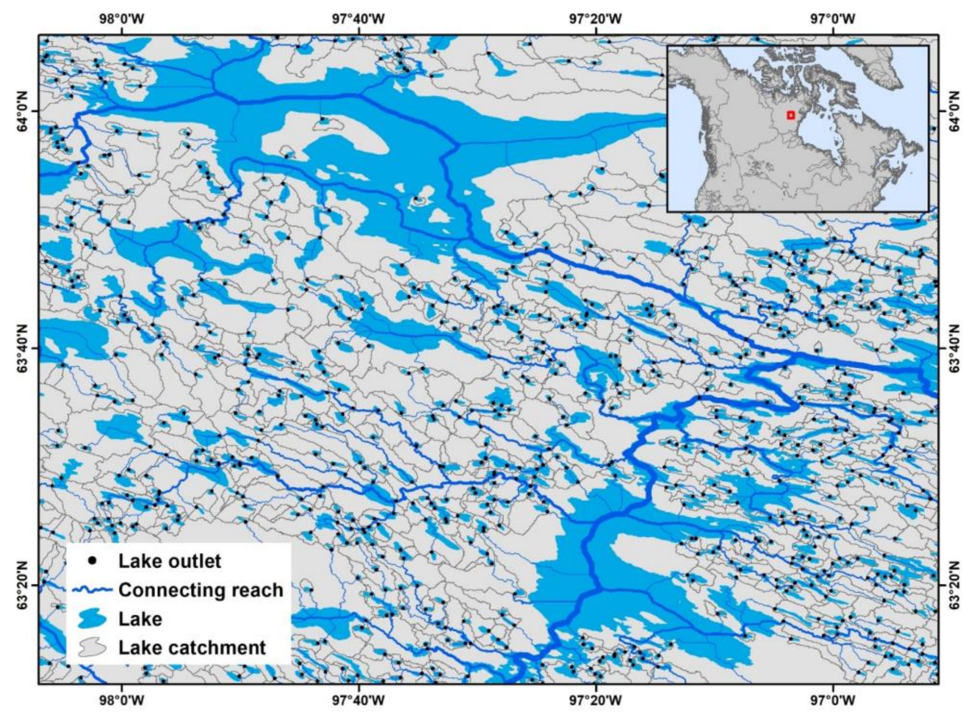

Figure 3. An example of Lake-TopoCat in the Canadian Shield. This example covers an area of ~6300 km^2 with ~850 lakes represented. Displayed features are lake outlets, inter-lake reaches (with width representing the abundance or count of the upstream lakes), lake unit catchments, and lake boundaries. All lakes in this region belong to the same lake-network basin shown in inset.

For user convenience, we also release the preliminary Lake-TopoCat lake outlets, unit catchments, and inter-lake reaches, with the affix "*_prelim*" in the file names (as explained in Section 3.2 and 3.3). An example of the five feature components is given for a small area in the Canadian Shield (Fig. 3). The attribute tables for each of the feature components are explained in Section 4.

3.1 Lake boundaries

Lake boundaries in Lake-TopoCat are geometrically the same as the HydroLAKES lake polygons, except that the former includes attributes informing lake drainage relations. About 99.95% of the ~14 million HydorLAKES polygons were located within or intersected the MERIT Hydro data boundary. These lakes were used to construct Lake-TopoCat (Fig. 4).

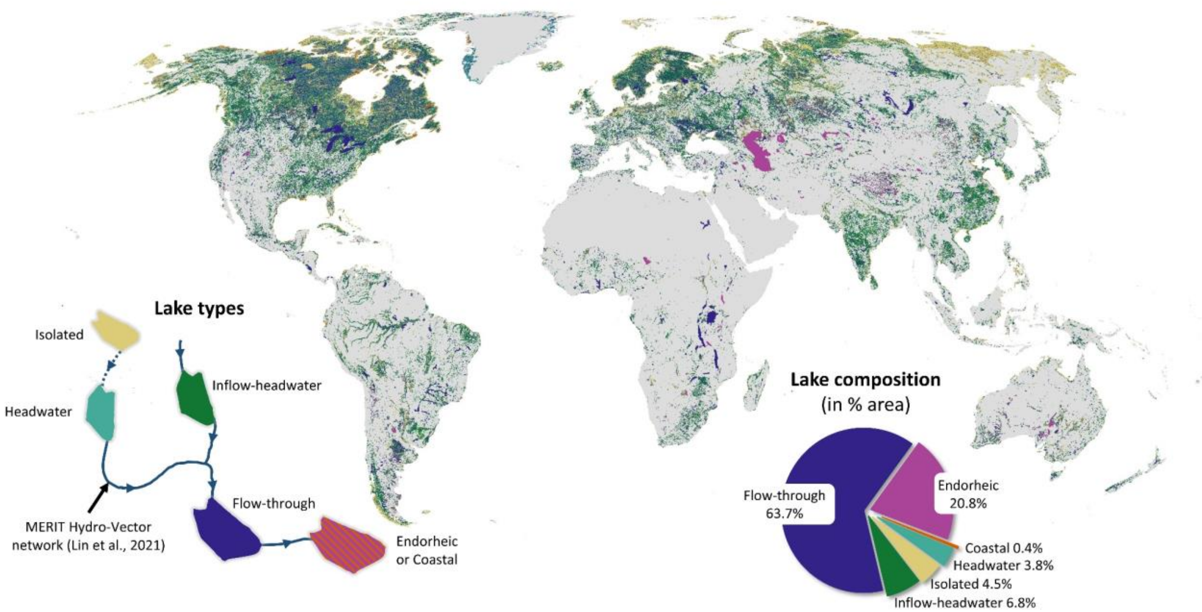

Figure 4. Global lakes and their drainage positions in Lake-TopoCat. Lake types ("Laktyp_mhv" attribute) configured based on the hydrologic position in global perennial and intermittent rivers provided in MERIT Hydro-Vector (*Lin et al.*, 2021). Pie chart in shows the global lake area composition by drainage type.

For consistency, we used the original HydroLAKES lake IDs to index lakes and their associated topological attributes in Lake-TopoCat (Table 1). The attributes associated with each lake boundary feature include the counts of directly connected upstream/downstream lakes, the ID lists of the directly connected downstream lakes, the total count of upstream lakes (from the headwater to this lake), and the total count of downstream lakes (from this lake to the drainage sink or ocean). The total area of the upstream catchments is also reported. To take into account lake bifurcation or multifurcation (i.e., a lake draining in two or more directions), we also reported the count of outlets for each HydroLAKES lake.

With the help of the derived topology and MERIT Hydro-Vector, we were able to sort out where each of the HydroLAKES lakes is located in terms of drainage position. We reported such information in a drainage type attribute ("Laktyp_mhv" in Table 1), which informs if a lake is isolated (off the drainage network), at the headwater, in the drainage pathway (flow-through), or at the drainage terminal (either endorheic or coastal; also see Section 3.2). The schematic diagram in Fig. 4 illustrates how these drainage types were defined. The number of identified isolated, headwater, inflow-headwater, flow-through, endorheic, and coastal lakes are 31.5%, 15.0%, 20.8%, 32.5%, 0.1%, and 0.1% of the total lakes, respectively. The pie chart in Fig. 4 shows the global lake area composition according to the lake drainage types. Flow-through lakes cover the largest proportion of the global lake surface area (~64%) despite ~32% of the global lake number. We noted that flow-through lakes are usually larger in size than the upstream lake types (i.e., isolated, headwater, and inflow headwater), so having the largest coverage in area but relatively low frequency was expected. The endorheic lakes cover a significant part of the global lake surface area (~21%), mainly due to the Caspian Sea.

3.2 Lake outlets

Lake outlets in Lake-TopoCat represent the locations of lake drainage pour points. Although the majority of lakes have one outlet only, lake bifurcation or even multifurcation does exist (see the example in Fig. 5). To provide a more realistic representation of global lake topology, we leveraged the hydrography information in MERIT Hydro and the accurate lake boundaries in HydroLAKES to allow for lake multifurcation in the Lake-TopoCat product. Here a multifurcation lake is defined as a HydroLAKES water body that shows divergent drainage destinations (lakes in our case). In brief, we designed the algorithm to account for multifurcation as follows (Fig. 1). First, all possible pour points of a lake were identified using the hydrography data (i.e., preliminary outlets). These pour points were next grouped by the downstream outlets or sinks. In other words, we considered several pour points to be in the same group if they drain to the same downstream outlet. Third, the pour point with the maximum flow accumulation in each group was selected to represent the outlet associated with this drainage direction (i.e., final outlets). The rationale of this design was to captures all unique drainage destinations from each lake and meanwhile to ignore intermediate (trivial) drainage divergence between the lake and each of its unique destinations. Since multifurcation was allowed, the count of outlets in Lake-TopoCat is larger than the count of lakes. We identified 1,459,201 outlets for 1,426,967 lakes, where 29,190 lakes (~2% of the global lakes) show bi/multifurcation (Fig. 6).

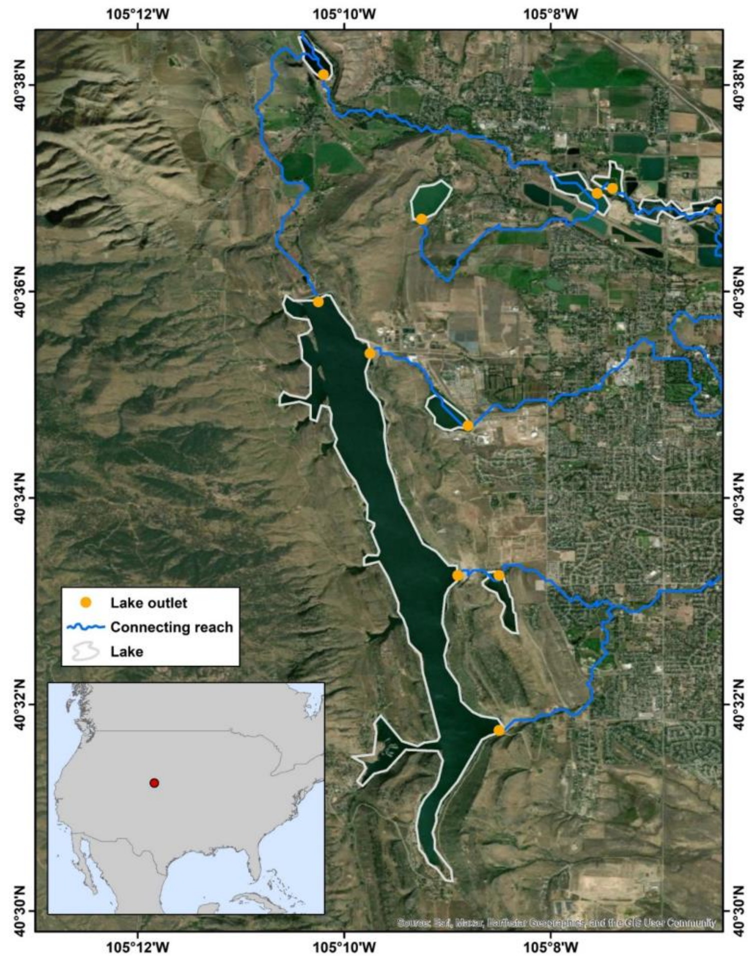

Figure 5. An example of a multifurcation lake (Horsetooth Reservoir in Colorado, US) with four verified drainage outlets (map background from Esri, Maxar, Earthstar Geographics, and the GIS User Community).

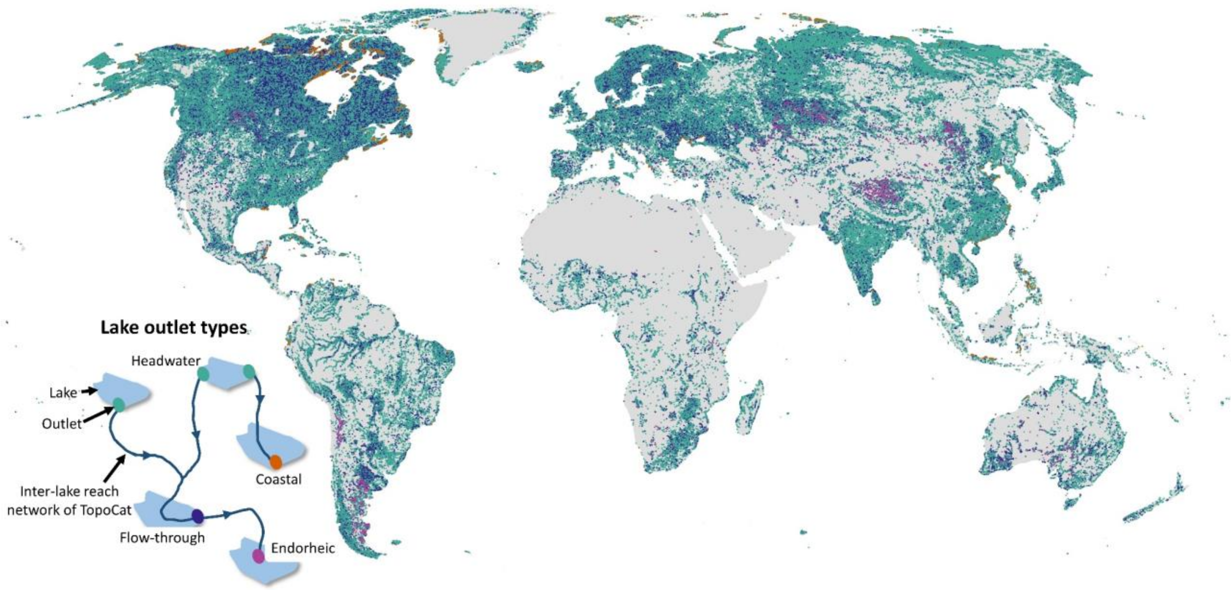

Figure 6. Global map of lake outlets and drainage types in Lake-TopoCat. Outlet types are based on drainage positions in the inter-lake reach network. The schematic diagram shows the outlet type definitions.

The attributes of each lake outlet feature (Table 1) include the unique ID of the outlet (which is different from the HydroLAKES lake ID), the associated HydroLAKES lake ID, and topological relationship with other outlets and lakes, e.g., the IDs of the directly connected downstream outlets/lakes. The attributes also contain important information on the associated lake catchment and drainage paths, which was derived with the assistance of the attributes of the catchment and reach features (Sections 3.3 and 3.4). Examples include the area of the local catchment (i.e., the drainage catchment from the outlet to the directly connected upstream outlet(s)), the area of the entre upstream catchment (i.e., the drainage catchment from the outlet to the headwater), the count of directly connected upstream lakes, the maximum, minimum, and mean drainage distances to the directly connected upstream outlets, the count of lakes in the entire upstream catchment, and the count of lakes in entire downstream region (to the inland sink or ocean). Drainage types of the outlets were assigned based on their hydrological positions in the inter-lake reach network (Fig. 6). The drainage distance from each outlet to the next downstream outlet and the accumulative drainage distance to the sink/ocean were also calculated using the inter-lake reach network. From Lake-TopoCat it was found that ~82% of the global lakes have a drainage proximity to other lakes of less than 10 km, and the percentage increases to ~95% if the proximity increases to 100 km.

3.3 Unit catchments

This feature file provides drainage catchments associated with each of the lake outlets. Note that the catchments are not based on one-lake to one-catchment relationship, but one-outlet to one-catchment relationship. This was to offer the catchment boundaries at the finest possible spatial detail, and when needed, users can easily dissolve the catchments using the lake ID attribute to form full catchments of each of the lakes. More specifically, unit catchments for all possible pour points for a lake were first delineated independently (i.e., preliminary unit catchments). The catchments for the pour points draining to the same downstream outlet/sink were then lumped to one "local or outlet catchment", and this catchment was assigned to the outlet with the maximum accumulation flow (i.e., final unit catchments; refer to outlet selection in Section 3.2). Therefore, the count of local catchments equal to the count of lake outlets, and bifurcation or multifurcation lakes have multiple local catchments. In total, the delineated catchments in Lake-TopoCat cover about 77.5 million km^2, which is about 57% of the Earth's land mass excluding the Antarctic (Fig. 7).

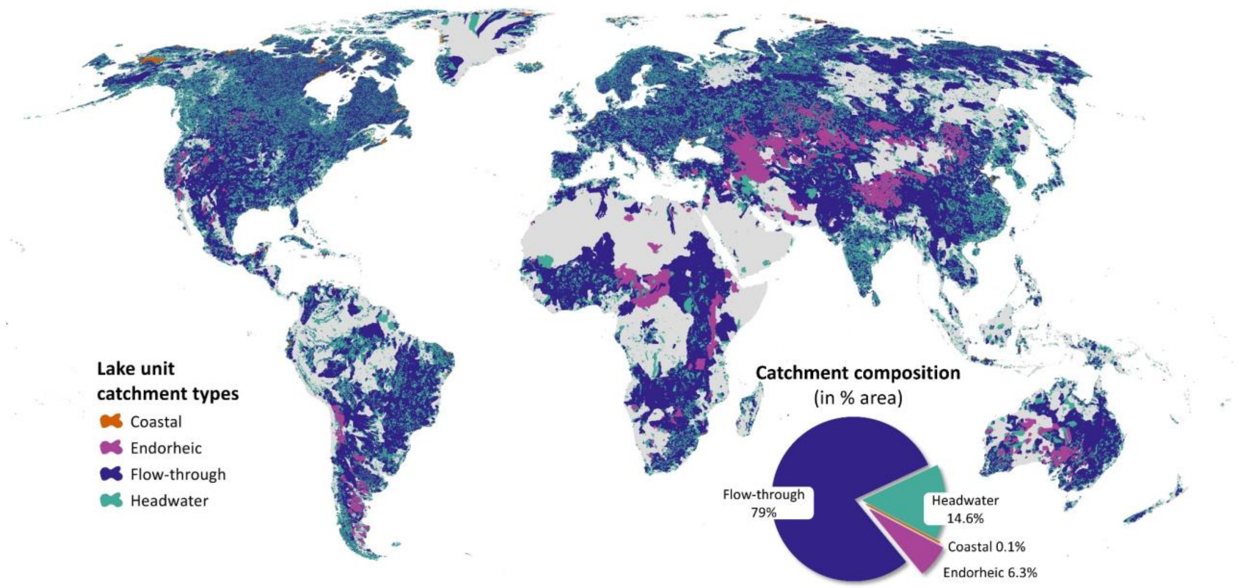

Figure 7. Global map of lake unit catchments in Lake-TopoCat. The catchment boundaries define the drainage areas between cascading lake outlets (i.e., one catchment for each lake outlet), and when there is no lake further upstream, the catchment defines the drainage area from the headwater to the lake outlet. The pie chart shows the composition of global lake catchment area by outlet drainage type.

Attributes of each lake catchment feature include the IDs of the associated outlet and lake, the IDs of the directly connected downstream lake and outlet, and the area of this catchment. The pie chart in Fig. 7 summarizes the composition of global lake catchments in terms of outlet drainage types. The catchments associated with flow-through outlets covers the largest proportion (~79%) in area. About 80% of Lake-TopoCat lakes have lake-area to catchment-area ratios ranging from 2 to 70.

3.4 Inter-lake reaches

This feature file consists of detailed reach segments with drainage topology, and depict how a lake drains to another and eventually reach the ocean of an inland sink. About 3 million connecting reaches were generated among ~1.4 million outlets (Fig. 8). The total length of these inter-lake connecting reaches is ~10 million km. Main attributes of each connecting reach include a unique reach ID, the IDs of the directly connected upstream outlet and lake, the count of all upstream lakes (from the headwater), the IDs of the directly connected downstream outlet and lake, the geodesic distance and slope of the reach, and the Strahler stream order of the reach.

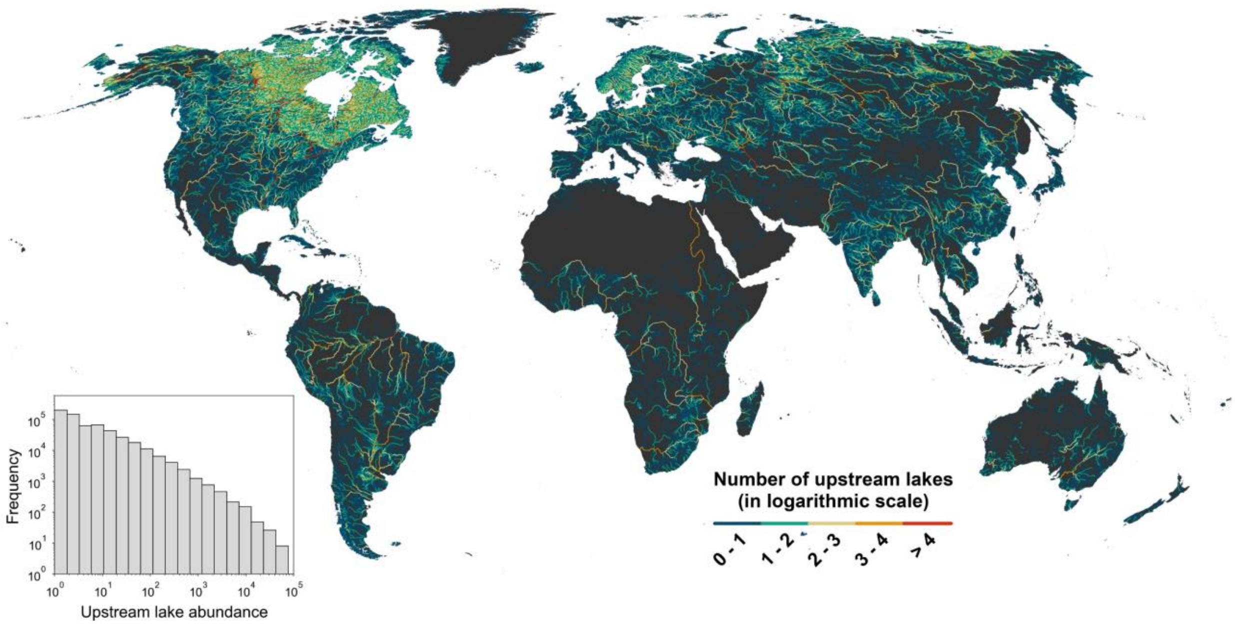

Figure 8. Global map of inter-lake reaches in Lake-TopoCat. Reach colors illustrate the accumulative lake abundance or count upstream to each inter-lake reach. The histogram on the lower-left corner shows the frequency of upstream lake abundance.

The global distribution of our delineated inter-lake reach network shown in Fig. 8, where different colors illustrate the total count of lakes upstream of each reach. For example, the abundance of upstream lakes per connecting reach in the Canadian Shield, the most lake-rich region, is several orders higher than the rest of the world. The histogram on Fig. 8 suggests that the abundance of total upstream lakes follows a quasi-Pareto distribution.

3.5 Lake-network basins

For user convenience, we provided another feature file "Lake-network basins". Each lake-network basin, in principle, defines the entire drainage area of an inter-lake reach network, which stretches from the headwater to an inland sink or the ocean. If a multifurcation lake infringes more than one reach networks that drain to different termini, these networks were considered related to each other through this shared lake, and their drainage areas were merged into one lake-network basin.

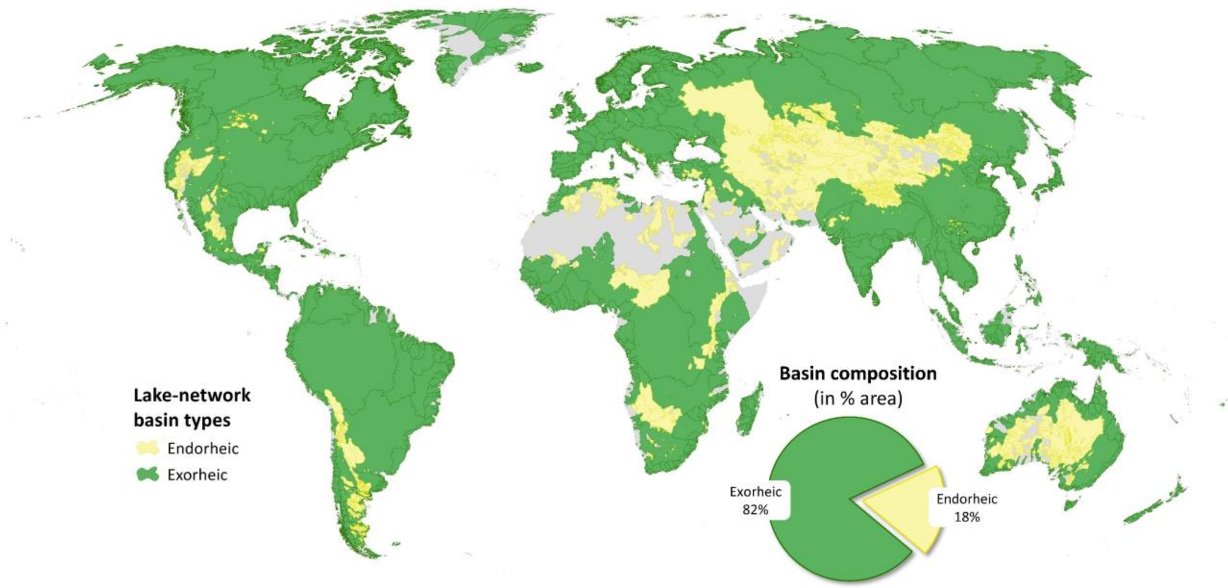

Figure 9. Global map of lake-network basins in Lake-TopoCat. A lake-network basin is defined as the entire drainage area from the headwater to the ocean or the inland sink, which contains a hydrologically independent inter-lake network. Gray indicates basins having no lake presence. The pie chart illustrates the composition of global lake basin areas by their drainage type (endorheic and exorheic).

A total of 47,340 lake-network basins were identified as shown in Fig. 9. Among them, endorheic basins ("Basin_type" in Table 1) account for 5.1% by count and 18% by area of all lake-network basins. These endorheic basins cover ~15.4% of global surface excluding Antarctica. Overall, the spatial pattern of the lake-network basins is consistent with the river basins. Importantly, these lake-network basins define how the millions of global lakes are partitioned by their drainage dependence, which can potentially help users gauge or optimize parallel processing (e.g., for global lake modeling and coupled lake-river modeling). For this reason, we also assigned the ID of the affiliated lake-network basin ("Basin_id") to each of the other Lake-TopoCat features.

4. Attributes

Table 1. Attributes in each of the five feature components of Lake-TopoCat. Here, dd, and masl are abbreviations of decimal degrees, and meters above sea level, respectively.

Feature	Attribute	Description	Unit
Lakes	Hylak_id	HydroLAKES ID for this lake	N/A
	Lake_area	Lake water surface area	km^2
	Outlet_n	Count of outlets of this lake (>1 indicating multifurcation)	N/A
	D_hylak_id	List of IDs of the next (i.e., directly connected) downstream lakes	N/A
	D_lake_n	Count of the next downstream lakes	N/A
	D_lak_ntot	Count of all downstream lakes to the sink or ocean, excluding this lake	N/A
	U_lake_n	Count of the next upstream lakes	N/A
	U_lak_ntot	Count of all upstream lakes from the headwater, excluding this lake	N/A
	Cat_a_lake	Area of the total upstream catchment(s) (accumulative from the headwater) for this lake. In case of multifurcation, this value aggregates the areas of the upstream catchments for all outlets of this lake.	km^2
	Lake_type	Drainage type of this lake in relation to the inter-lake reach network: headwater, flow-through, terminal, or coastal.	N/A
	Lake_order	Strahler order of the lake in the inter-lake reach network	N/A
	Laktyp_mhv	Drainage type of this lake in relation to rivers in MERIT Hydro-Vector: isolated, headwater, inflow-headwater, flow-through, terminal, or coastal.	N/A
	Basin_id	ID of the lake-network basin (see below) this lake belongs to	N/A
Outlets	Outlet_id	ID of this lake outlet, which is different from Hylak_ID as there can be more than one outlet for a multifurcation lake.	N/A
	Hylak_id	ID of the associated HydroLAKES lake	N/A
	Outlet_lat	Latitude at this outlet	dd
	Outlet_lon	Longitude at this outlet	dd
	Outlet_elv	Elevation at this outlet (based on hydrologically adjusted MERIT DEM, as for other elevation values)	masl
	D_out_id	ID of the next downstream outlet	N/A
	D_hylak_id	ID of the next downstream lake	N/A
	D_lak_ntot	Count of all downstream lakes to the sink or ocean, excluding this lake	N/A
	D_reach_id	ID of the connecting downstream reach	N/A
	D_slope	Hydraulic gradient (slope) from this outlet to the next downstream outlet or sink	N/A
	D_distance	Drainage distance from this outlet to the next downstream outlet or sink	m
	D_dst_sink	Total drainage distance to the sink or ocean	m
	U_lake_n	Count of the next upstream lakes	N/A
	U_dist_avg	Mean distance from the next upstream outlets	m
	U_dist_min	Minimum distance from the next upstream outlets	m
	U_dist_max	Maximum distance from the next upstream outlets	m
	U_lak_ntot	Count of all upstream lakes from the headwater, excluding this lake	N/A
	Cat_area	Area of the associated unit catchment	km^2
	Cat_a_tot	Area of the entire upstream drainage basin, i.e., from the headwater to this outlet	km^2

	Out_type	Drainage type of this lake outlet in relation to the inter-lake reach network: headwater, flow-through, terminal, or coastal.	N/A
	Out_order	Strahler order of the outlet in the inter-lake reach network	N/A
	Basin_id	ID of the lake-network basin this outlet belongs to	N/A
Unit catchments	Outlet_id	ID of the associated outlet, also used as the ID of the unit catchment.	N/A
	Hylak_id	ID of the associated lake	N/A
	D_out_id	ID of the next downstream outlet	N/A
	D_hylak_id	ID of the next downstream lake	N/A
	Cat_area	Area of this unit catchment	km^2
	Cat_type	Drainage type of this unit catchment in relation to the inter-lake reach network (same as in "Outlet_type").	N/A
	Basin_id	ID of the lake-network basin this unit catchment belongs to	N/A
Inter-lake reaches	Reach_id	ID of this inter-lake reach	N/A
	D_reach_id	ID of the next downstream reach	N/A
	D_out_id	ID of the connected downstream outlet	N/A
	D_hylak_id	ID of the connected downstream lake	N/A
	U_out_id	ID of the connected upstream outlet	N/A
	U_hylak_id	ID of the connected upstream lake	N/A
	U_lak_ntot	Count of the total upstream lakes from the headwater	N/A
	Start_elv	Elevation at the starting node of this reach	masl
	End_elv	Elevation at the ending node of this reach	masl
	Rch_length	Reach geodesic length	m
	Rch_slope	Hydraulic gradient (slope) of this reach	N/A
	Rchint_mhv	Proportion of this inter-lake reach (in % length) overlapped by river channels in MERIT Hydro-Vector, implying the likelihood of this reach being more perennial or ephemeral	N/A
	Rch_order	Strahler order of this reach	N/A
	Basin_id	ID of the lake-network basin this reach belongs to	N/A
Lake-network basins	Basin_id	Lake-network basin ID	N/A
	Basin_type	Type of this lake-network basin: endorheic (draining to a terminal lake or an inland sink) or exorheic (draining to the ocean)	N/A
	Basin_area	Area of this lake-network basin	km^2

5. Data availability and usage

The Lake-TopoCat v1.0 database is available in both shapefile and geodatabase formats through Zenodo (https://doi.org/10.5281/zenodo.7420810) under the Creative Commons Attribution 4.0 International license.

6. Disclaimer

Authors of this dataset claim no responsibility or liability for any consequences related to the use, citation, or dissemination of Lake-TopoCat v1.0.

7. References

Lin, P., Pan, M., Wood, E. F., and Allen, G. H.: A new vector-based global river network dataset accounting for variable drainage density, *Scientific Data*, 8, 28, https://doi.org/10.1038/s41597-021-00819-9, 2021.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nature Communications, 7, 13603, https://doi.org/10.1038/ncomms13603, 2016.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: a high-resolution global hydrography map based on latest topography dataset, *Water Resources Research*, 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, *Geophysical Research Letters*, 44, 5844–5853, https://doi.org/10.1002/2017GL072874, 2017.