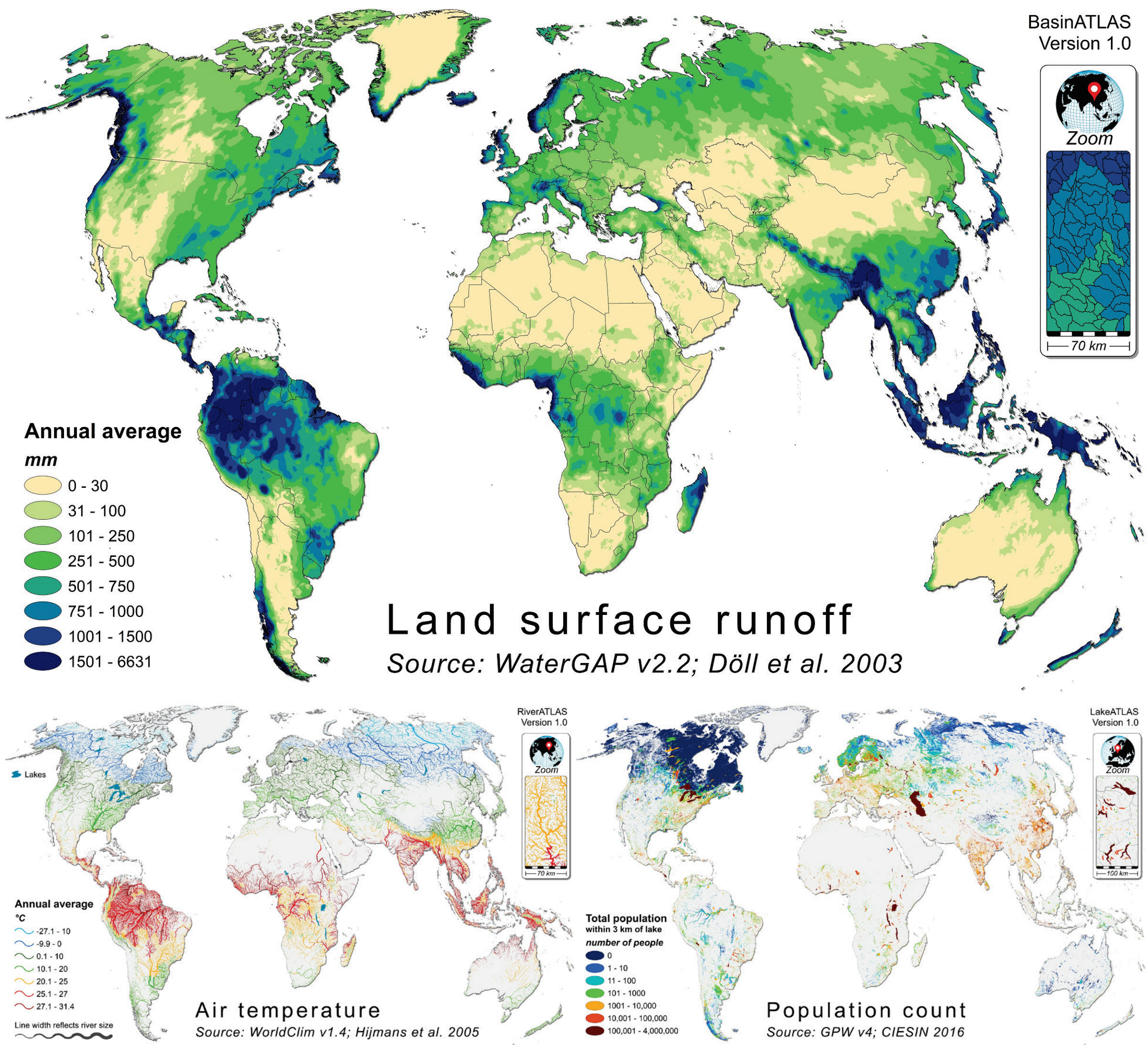
HydroATLAS

A global compendium of hydro-environmental sub-basin, river reach and lake characteristics at 15 arc-second resolution

Technical Documentation – version 1.0.1

prepared by Bernhard Lehner (bernhard.lehner@mcgill.ca)

June 2022



1. Background and introduction

The goal of HydroATLAS is to provide a broad user community with a standardized compendium of hydro-environmental attribute information for all catchments, rivers and lakes of the world at high spatial resolution. **This documentation describes the complete version 1.0 of HydroATLAS, including all three spatial units**.

Version 1.0 of HydroATLAS offers data for 56 hydro-environmental variables, partitioned into 281 individual attributes and organized in six categories: hydrology; physiography; climate; land cover & use; soils & geology; and anthropogenic influences (***Table 1*** *and* ***Appendix 4***).

HydroATLAS derives these hydro-environmental attributes by reformatting original data from well-established global digital maps. The attributes are then linked to (1) hierarchically nested sub-basin polygons at multiple scales; (2) individual river reach lines; and (3) lake polygons at a global scale (***Figure 1***). The sub-basin polygons and river reach lines are both extracted from the global HydroSHEDS database (Lehner et al. 2008) at 15 arc-second (~500 m) resolution, whereas the lake polygons are taken from the HydroLAKES database (Messager et al. 2016). The sub-basin, river reach and lake information are offered in three companion datasets: BasinATLAS, RiverATLAS and LakeATLAS. The standardized format of HydroATLAS ensures easy applicability while the inherent topological information supports basic network functionality such as identifying up- and downstream connections. HydroATLAS is fully compatible with other products of the overarching HydroSHEDS project enabling versatile hydro-ecological assessments. Updates of HydroATLAS are envisioned as new data become available.

The HydroATLAS documentation is organized in two parts: Part 1 (this document) provides an overview of the database and general explanations. Part 2 is provided in three alternative files: 'BasinATLAS_Catalog', 'RiverATLAS_Catalog', and 'LakeATLAS_Catalog'. Each catalog file first provides a summary table listing all hydro-environmental variables and their basic characteristics. This is followed by detailed information on each individual variable, including source data descriptions, units, conversion methodology, and citations, as they pertain to the sub-basins, river reaches and lakes, respectively. Each variable is presented on one standardized sheet which includes a map at global extent indicating the spatial distribution of values of the respective variable. The summary table and information sheets are hyperlinked within each catalog. Note that the variables and attributes are generally similar between the three catalogs, but some difference exist.

The development of BasinATLAS and RiverATLAS is fully described in Linke et al. (2019) and should be cited as:

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6: 283. https://doi.org/10.1038/s41597-019-0300-6

The development of LakeATLAS is fully described in Lehner et al. (2022) and should be cited as:

Lehner, B., Messager, M.L., Korver, M.C., Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. Scientific Data 9: 351. https://doi.org/10.1038/s41597-022-01425-z

If general references are made to the overall HydroATLAS product, citations to both Linke et al. (2019) and Lehner et al. (2022) should be used.

For more details on how to refer to explicit data or partial information as provided within he HydroATLAS data compendium, and for further acknowledgements, see section 4.4 below.

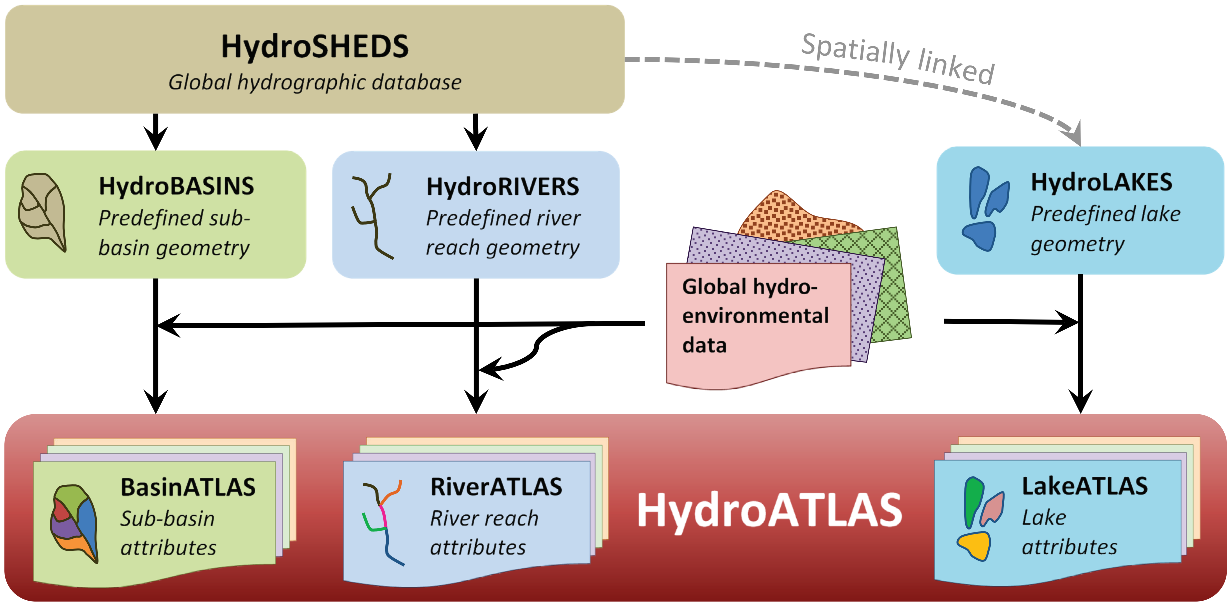


Figure 1: Conceptual design of HydroATLAS and relationship to underpinning HydroSHEDS database. HydroATLAS consists of three companion attribute datasets: BasinATLAS and RiverATLAS (fully described in Linke et al. 2019) as well as LakeATLAS (Lehner et al. 2022). BasinATLAS provides sub-basin characteristics for hierarchically nested watersheds at twelve spatial scales. RiverATLAS contains the same attributes yet calculated for river and stream reaches rather than sub-basins. The geospatial units for both databases, i.e., sub-basin polygons and river reach line segments, respectively, have been derived from the global hydrographic database HydroSHEDS (Lehner et al. 2008) at a spatial resolution of 15 arc-seconds (~500 m at the equator). LakeATLAS follows the same overall format and structure and contains the same hydro-environmental attributes yet calculated for lake polygons. Its geospatial units are provided by the HydroLAKES database (Messager et al. 2016) and are linked to the sub-basins and river reaches of HydroSHEDS via corresponding IDs.

Table 1: Categories of hydro-environmental variables offered in the HydroATLAS database.

Identifier	Category	Description
H	Hydrology & hydrography	Hydrological and hydrographic characteristics related to quantity, quality, location and extent of terrestrial water *Examples: natural runoff and discharge, groundwater table depth, lake cover*
P	Physiography	Topographic characteristics related to terrain, relief or landscape position *Examples: elevation, slope*
C	Climate	Climatic characteristics *Examples: mean temperature, climate moisture index, global aridity*
L	Land cover & land use	Land cover and land use characteristics including biogeographic regions *Examples: land cover classes, permafrost extent, freshwater ecoregions*
S	Soils & geology	Soil and geology related characteristics including substrate types and soil conditions *Examples: percentage clay in soil, soil water stress, lithography, soil erosion*
A	Anthropogenic influences	Anthropogenic characteristics including demographic and socioeconomic aspects *Examples: population density, human footprint, GDP per capita*

2. Methods and data characteristics

The methods used to create HydroATLAS are fully described in Linke et al. (2019) (for BasinATLAS and RiverATLAS) and Lehner et al. (2022) (for LakeATLAS), respectively.

All spatial units of BasinATLAS and RiverATLAS, i.e., either sub-basin polygons or river reach lines, were extracted from World Wildlife Fund's HydroSHEDS database (Lehner et al. 2008; Lehner and Grill 2013) at a grid resolution of 15 arc-seconds (approx. 500 m at the equator). For more information on HydroSHEDS please refer to the Technical Documentation at http://www.hydrosheds.org.

All spatial units of LakeATLAS, i.e., the polygons of all lakes in the world with at least 10 ha in surface area, were provided by the HydroLAKES database (Messager et al. 2016). For more information on HydroLAKES please refer to the Technical Documentation at https://www.hydrosheds.org/hydrolakes.

HydroATLAS consists of three complementary parts: BasinATLAS, RiverATLAS, and LakeATLAS. BasinATLAS provides hydro-environmental attributes for sub-basins (polygons). RiverATLAS provides hydro-environmental attributes for stream and river reaches (line segments). LakeATLAS provides hydro-environmental attributes for lakes (polygons).

Basin and sub-basin delineations have been pre-processed as a derivative of HydroSHEDS at 15 arc-second resolution and are available as a stand-alone product termed HydroBASINS (for details see https://www.hydrosheds.org/hydrobasins). The HydroBASINS dataset offers a suite of 12 layers, each containing nested sub-basins that were subdivided and coded using the topological concept of the Pfafstetter system, which provides a methodology for the breakdown of sub-basins at different scales in a hierarchical and systematic manner (***Figure 2a***). It should be noted, however, that at the lowest Pfafstetter levels (i.e., 1-3) multiple river basins may be lumped into larger regions, and for coastal sub-basins (at any level) multiple smaller rivers may be lumped into one sub-basin—in these cases, the association of some particular attributes (such as river discharge) is ambiguous, and the assigned attribute value may refer to only one river within the sub-basin unit.

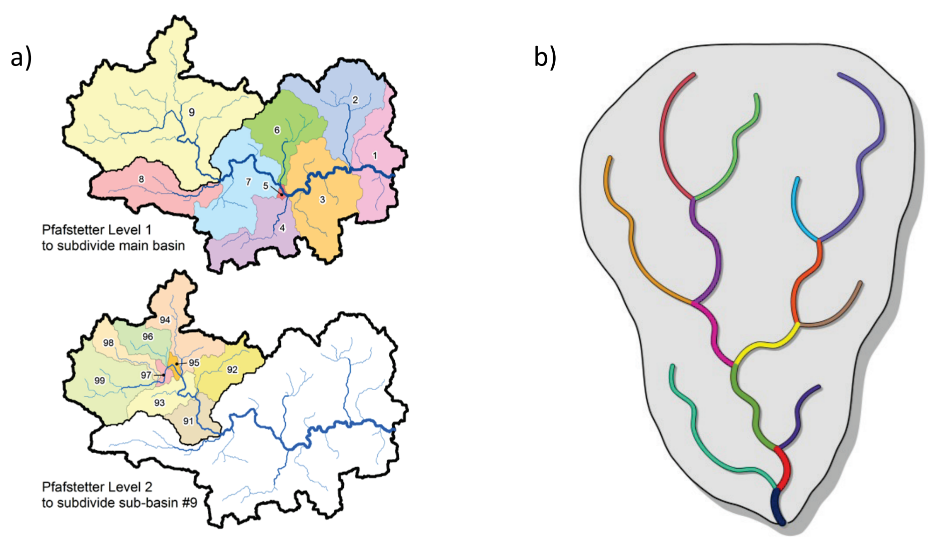


Figure 2: Overview of Pfafstetter sub-basin coding scheme used in BasinATLAS (a); and river reach concept used in RiverATLAS (b). Sub-basins are nested within 12 hierarchical levels. A river reach is defined as a stretch between two tributaries, or between the start/end of the network and a tributary.

Also, a global river network delineation has been extracted from HydroSHEDS at 15 arc-second resolution and is available as a stand-alone product termed HydroRIVERS (for details see https://www.hydrosheds.org/hydrorivers). For this network, rivers have been defined to start at all pixels where the accumulated upstream catchment area exceeds 10 km^2, or where the long-term average natural discharge exceeds 0.1 cubic meters per second, resulting in a line network consisting of individual stream and river reaches (***Figure 2b***).

For the lake surface polygons as provided by HydroLAKES, which comprise all lakes with a surface area of at least 10 ha (for details see https://www.hydrosheds.org/hydrolakes), an additional spatial unit has been created to accommodate the needs of LakeATLAS. This unit represents the direct vicinity around a lake within a buffer of 3 km from the lake's shoreline (without the lake surface itself). For all data processing steps, the lakes and buffers were converted to grid format at the standard 15 arc-second resolution of HydroATLAS.

It should be noted that the quality of HydroSHEDS data is significantly inferior for regions above 60 degrees northern latitude as there were no high-resolution elevation data (SRTM) available at the time of creating HydroSHEDS and thus a coarser source of elevation data has been inserted (HYDRO1k). The quality of HydroLAKES is also affected by its underpinning source data, providing the most accurate information for Canada (which contains 62% of global lakes) and the most inferior accuracy for Siberia above 60 degrees northern latitude (see Messager et al. 2016 for details).

3. Data format and distribution

a) *Data format and projection*

HydroATLAS is publicly available for download at http://www.hydrosheds.org/hydroatlas and from the *figshare* data repositories at https://doi.org/10.6084/m9.figshare.9890531 (BasinATLAS and RiverATLAS) and https://doi.org/10.6084/m9.figshare.19312001 (LakeATLAS). All map data layers, including attribute tables, are provided in ESRI© Geodatabase and Shapefile formats. The data are projected in a Geographic Coordinate System using the World Geodetic System 1984 (GCS_WGS_1984). The attribute table can also be accessed as a stand-alone file in dBASE format which is included in the Shapefile format. HydroATLAS data are available electronically in compressed zip file format. To use the data files, the zip files must first be decompressed. Each zip file includes a copy of the HydroATLAS Technical Documentation.

b) Layer name syntax and spatial coverage

HydroATLAS data layers are provided in two spatial extents:

- primarily as a seamless, fully global coverage;
- but for some datasets also (or only) as regional tiles (see Figure 2 for definition of regions).

The layer names follow the syntax:

- ***BasinATLAS_v10_levXX*** (for BasinATLAS layers with global coverage), where *XX* indicates the Pfafstetter level (1-12);
- ***RiverATLAS_v10*** (for RiverATLAS layer with global coverage); or
- ***RiverATLAS_v10_YY*** (for RiverATLAS layers in regional tiles), where *YY* indicates the region;
- ***LakeATLAS_v10_pol*** (for LakeATLAS layer of lake polygons with global coverage);
- ***LakeATLAS_v10_pol_YY*** (for LakeATLAS layers in regional tiles), where *YY* indicates the region;
- ***LakeATLAS_v10_pnt*** (for LakeATLAS layer of lake outlet points with global coverage);
- ***LakeATLAS_v10_pnt_YY*** (for LakeATLAS layers in regional tiles), where *YY* indicates the region.

The regional extents (see ***Figure 3***) are defined by a two-digit identifier:

Identifier	*Region*
af	Africa
ar	North American Arctic
as	Central and South-East Asia
au	Australia and Oceania
eu	Europe and Middle East
gr	Greenland
na	North America and Caribbean
sa	South America
si	Siberia

Note that the Shapefile format is limited to a maximum file size of 2 GB. Therefore, the RiverATLAS data in Shapefile format are only provided in regional tiles (with further subdivisions into north and south parts where needed); and the LakeATLAS data in Shapefile format are provided for two regions: the western hemisphere (west = ar, gr, na, sa) and the eastern hemisphere (east = af, as, au, eu, si). Currently, all other data layers are provided in full global coverage, but more versatile regional breakdowns and data packages may be offered in future iterations.

Figure 3: Spatial extent of regional tiles of HydroATLAS layers.

c) *Available columns and column name syntax*

The attribute tables of HydroATLAS contain the pre-existing columns of HydroBASINS, HydroRIVERS, and HydroLAKES, respectively (see their respective Technical Documentations at http://www.hydrosheds.org for details; as well as ***Appendices 1-3*** in this document for a list of columns). The hydro-environmental attributes are then appended in a series of additional columns. This section provides information on the column name syntax used for the identification of each sub-basin, river reach or lake attribute provided in the HydroATLAS database. All existing attributes and their associated column names are summarized in ***Appendix 4*** and at the beginning of the BasinATLAS, RiverATLAS and LakeATLAS catalogs.

Each hydro-environmental attribute column name has 10 digits (for example '*dis_m3_pyr*') and its syntax is as follows:

<Layer name key>_<Unit key>_<Spatial key>< Dimension key>

Layer name key:

Three digits that describe the name of the attribute. The layer name key is unique to the attribute it represents. *Example: 'dis' for discharge.*

Unit key:

Two digits that describe the units of the attribute value. See ***Table 2*** for possible keys.

Spatial extent key:

One digit that describes the spatial extent of the attribute. See ***Table 3*** for possible keys.

Dimension key:

Two digits that describe the dimension of the attribute in terms of its aggregation level or other type of spatio-temporal association. The dimension key can refer to a temporal dimension, a statistical aggregation, or a class or year association. See ***Table 4*** for possible keys.

Table 2: Unit keys. Note that some values are stored in factors of the given units (to efficiently store them as integers without losing precision), e.g., temperature is stored in tenths of degrees; these factors are listed in the respective data sheet of each variable in the HydroATLAS catalogs.

Key	Unit of values
cl	Classes
cm	Centimeters
ct	Count (e.g., number of people)
dc	Degrees Celsius (°C)
dg	Degrees
dk	Decimeters per kilometer
ha	Hectares
id	ID number
ix	Index value
kh	Kilogram per hectare per year (kg/ha/yr)
m3	Cubic meters per second (m^3/s)
mc	Million cubic meters (mcm)
mk	Meters per square kilometer (m/km^2)
mm	Millimeters
mt	Meters *or* Meters above sea level (m.a.s.l.)
pc	Percent *or* Percent cover
pk	Per square kilometer (e.g., people per square kilometer)
tc	Thousand cubic meters
th	Metric tonnes per hectare
ud	US dollars

Table 3: Spatial extent keys. Note that all attributes represent average values within the spatial unit unless stated otherwise in the attribute's catalog sheet.

Key	Spatial representation
c	In reach catchment (i.e., the local catchment that drains directly into the reach)
l	Inside lake polygon
p	At pour point of sub-basin, river reach, or lake
r	Along reach segment
s	In sub-basin
u	In total watershed upstream of sub-basin, river reach, or lake; extracted at their respective pour point location
v	Within a 3-km vicinity buffer around lake polygon (excluding the lake polygon)

Table 4: Dimension keys.

Key	Temporal or statistical aggregation or other association
01-12	Calendar month (January to December) for monthly data
00-99	Class number (e.g., for spatial extent calculations of individual classes)
00-99	Other numbers may be used & explained as needed (e.g., to represent a specific year)
av	Average
g1-g9	Class groupings (individual groups are defined in HydroATLAS catalogs)
lt	Long-term maximum
mj	Spatial majority (dominant value)
mn	Minimum *or* Annual minimum
mx	Maximum *or* Annual maximum
se	Spatial extent (%)
su	Sum
va	Value
yr	Annual average

4. License, disclaimer and acknowledgement

4.1 License agreement

HydroATLAS forms a Collective Database, i.e., a collection of information from independent datasets, and as a whole is licensed under a Creative Commons Attribution 4.0 International License (CC-BY 4.0; http://creativecommons.org/licenses/by/4.0/). However, the individual parts (content) of this Collective Database are still governed by their own licenses. In version 1.0 of HydroATLAS, all attribute columns are licensed under either a Creative Commons Attribution 4.0 International License (CC-BY 4.0) or an Open Data Commons Open Database License (ODbL 1.0; https://opendatacommons.org/licenses/odbl/1-0/index.html), both permitting reuse of the data for any purpose including commercial. In cases where original licenses differ from CC-BY 4.0 or ODbL 1.0, special permission was obtained from the original author(s) to release their works in the format of HydroATLAS under a CC-BY 4.0 or ODbL 1.0 license. Note that the licenses of the underpinning source datasets in their original format are not affected or altered by these licenses. Detailed information regarding the specific license that applies to each attribute column is provided in the respective data sheet of the BasinATLAS, RiverATLAS and LakeATLAS catalogs.

By downloading and using the data the user agrees to the terms and conditions of these licenses.

4.2 Disclaimer of warranty

The HydroATLAS database and any related materials contained therein are provided "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose, noninterference, system integration, or noninfringement. The entire risk of use of the data shall be with the user. The user expressly acknowledges that the data may contain some nonconformities, defects, or errors. The authors do not warrant that the data will meet the user's needs or expectations, that the use of the data will be uninterrupted, or that all nonconformities, defects, or errors can or will be corrected. The authors are not inviting reliance on these data, and the user should always verify actual data.

4.3 Limitation of liability

In no event shall the authors be liable for costs of procurement of substitute goods or services, lost profits, lost sales or business expenditures, investments, or commitments in connection with any business, loss of any goodwill, or for any direct, indirect, special, incidental, exemplary, or consequential damages arising out of the use of the HydroATLAS database and any related materials, however caused, on any theory of liability, and whether or not the authors have been advised of the possibility of such damage. These limitations shall apply notwithstanding any failure of essential purpose of any exclusive remedy.

4.4 Data citations and acknowledgements

When using an attribute contained in HydroATLAS, citations and acknowledgements should be made to both the original data source and the respective HydroATLAS compendium. For example, the following templates illustrate a reference to precipitation data sourced from HydroATLAS.

If the data are provided by BasinATLAS or RiverATLAS, please cite as:

> "*Precipitation data from the WorldClim v1.4 database (Hijmans et al. 2005) have been used in the spatial format as provided by BasinATLAS/RiverATLAS v1.0 (Linke et al. 2019).*"

If the data are provided by LakeATLAS, please cite as:

> "*Precipitation data from the WorldClim v1.4 database (Hijmans et al. 2005) have been used in the spatial format as provided by LakeATLAS v1.0 (Lehner et al. 2022).*"

Information regarding the reference(s) for each hydro-environmental attribute is provided on the individual attribute sheets in the BasinATLAS, RiverATLAS, and LakeATLAS catalogs. In addition, every data source may have individual requests for acknowledgements, and users of HydroATLAS are asked to honor those requests when using the respective attributes.

General citations and acknowledgements of HydroATLAS should be made as follows:

> *Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6: 283. https://doi.org/10.1038/s41597-019-0300-6*

We kindly ask users to cite both source data and HydroATLAS in any published material produced using the data. If possible, online links to the HydroATLAS website should be provided (http://www.hydrosheds.org/hydroatlas).

5. 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

Lehner, B., Verdin, K., Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU 89(10): 93-94. https://doi.org/10.1029/2008EO100001

Lehner, B., Messager, M.L., Korver, M.C., Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. Scientific Data Sci Data 9: 351. https://doi.org/10.1038/s41597-022-01425-z

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6: 283. https://doi.org/10.1038/s41597-019-0300-6

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

6. Appendices

Appendix 1: *Pre-existing attributes of HydroBASINS included in BasinATLAS version 1.0*

BasinATLAS includes the following attribute columns of HydroBASINS (for more details see Technical Documentation of HydroBASINS at https://www.hydrosheds.org/hydrobasins):

Note that in the shapefile format the field 'FID' and in the geodatabase format the fields 'OBJECTID', 'Shape_Length' and Shape_Area' are added by default by the ArcGIS software—these fields are not officially part of BasinATLAS.

Column	***Description***
Hybas_id	Unique basin identifier. The code consists of 10 digits: • First 1 digit represents the region: 1 = Africa; 2 = Europe; 3 = Siberia; 4 = Asia; 5 = Australia; 6 = South America; 7 = North America; 8 = Arctic (North America); 9 = Greenland • Next 2 digits define the Pfafstetter level (01-12). The value '00' is used for the 'Level 0' layer that contains all original sub-basins and all Pfafstetter codes (at all levels); 'Level 0' only exists in the standard format of HydroBASINS (without lakes). • Next 6 digits represent a unique identifier within the HydroSHEDS network; values larger than 900,000 represent lakes and only occur in the customized format (with lakes) • Last 1 digit indicates the side of a sub-basin in relation to the river network (0 = noSide; 1 = Left; 2 = Right). Sides are only defined for the customized format (with lakes).
Next_down	Hybas_id of the next downstream polygon. This field can be used for navigation (up- and downstream) within the river network. The value '0' indicates a polygon with no downstream connection. Note that small endorheic sinks may have a 'virtual' connection to an appropriate downstream polygon to allow for topological queries in larger river basins where discontinuities should be eliminated (e.g., the larger Nile Basin contains smaller endorheic basins that are virtually connected to the larger basin). Virtual connections can be identified as they carry a value of '2' in the 'Endo' field AND a value larger than '0' in the 'Next_down' field. Users can thus decide whether or not to terminate the routing at endorheic sinks.
Next_sink	Hybas_id of the next downstream sink. This field indicates either the ID of the next downstream endorheic sink polygon (if there is one) or the most downstream polygon of the river basin (if there is no endorheic sink in between). This field can be used to identify the entire, fully connected watershed that a polygon belongs to.
Main_bas	Hybas_id of the most downstream sink, i.e., the outlet of the main river basin. This field indicates the ID of the most downstream polygon of the river basin and can be used to identify the entire river basin that a polygon belongs to, including all associated endorheic basins. Note: small endorheic parts are typically lumped (via virtual connections) with their corresponding larger basin, while large endorheic watersheds can form their own basins.
Dist_sink	Distance from polygon outlet to the next downstream sink along the river network, in kilometers. This distance is measured to the next downstream endorheic sink (if there is one) or (if there is none) to the most downstream sink (i.e., the ocean).

Dist_main	Distance from polygon outlet to the most downstream sink, i.e. the outlet of the main river basin along the river network, in kilometers. The most downstream sink or outlet is that of the larger basin (to which smaller endorheic sub-basins may be virtually connected), i.e. either the outlet at the ocean, or the final sink of a large endorheic watershed which forms its own basin. Note that when small endorheic basins are lumped with a larger river basin, the virtual linkages are not measured as true distances but are calculated as direct (zero distance) connections.
Sub_area	Area of the individual polygon (i.e., sub-basin), in square kilometers.
Up_area	Total upstream area, in square kilometers, calculated from the headwaters to the polygon location (including the polygon). The upstream area only comprises the directly connected watershed area, i.e., it does not include endorheic regions that may be part of the larger basin through virtual connections.
Pfaf_id	The Pfafstetter code. For more details see Technical Documentation of HydroBASINS. The Pfafstetter code uses as many digits as the level it represents. This field can be used to cluster or subdivide sub-basins into nested regions. This field is only available for levels 1-12 (i.e., not for the 'Level 0' layer of the standard format).
Side	Indicates the side of a sub-basin in relation to the river network: L = Left; R = Right; M = Merged (direction defined looking downstream). This index enables a distinction between the two sides along lake shorelines (see text for more explanation). Polygons have only been split into left and right parts where lakes exist. This field is only available in the customized format (with lakes).
Lake	Indicator for lake types: 0 = no Lake; 1 = Lake; 2 = Reservoir; 3 = Lagoon. This field is only available in the customized format (with lakes).
Endo	Indicator for endorheic (inland) basins without surface flow connection to the ocean: 0 = not part of an endorheic basin; 1 = part of an endorheic basin; 2 = sink (i.e., most downstream polygon) of an endorheic basin.
Coast	Indicator for lumped coastal basins: 0 = no; 1 = yes. Coastal basins represent conglomerates of small coastal watersheds that drain into the ocean between larger river basins.
Order	Indicator of river order (classical ordering system): order 1 represents the main stem river from sink to source; order 2 represents all tributaries that flow into a 1^{st} order river; order 3 represents all tributaries that flow into a 2^{nd} order river; etc.; order 0 is used for conglomerates of small coastal watersheds.
Sort	Indicator showing the record number (sequence) in which the original polygons are stored in the shapefile (i.e., counting upwards from 1 in the original shapefile). The original polygons are sorted from downstream to upstream. This field can be used to sort the polygons back to their original sequence or to perform topological searches.
Pfaf_1 to Pfaf_12	Pfafstetter codes for all levels (1 to 12). For general description see literature (e.g., Verdin and Verdin 1999). The Pfafstetter code uses as many digits as the level it represents. These fields can be used to create sub-basins at all Pfafstetter levels by dissolving the polygons accordingly. These fields are only available for the 'Level 0' layer of the standard format (without lakes).

Appendix 2: ***Pre-existing attributes of HydroRIVERS included in RiverATLAS version 1.0***

RiverATLAS includes the following attribute columns of HydroRIVERS (for more details see Technical Documentation of HydroRIVERS at https://www.hydrosheds.org/hydrorivers):

Note that in the shapefile format the field 'FID' and in the geodatabase format the fields 'OBJECTID' and 'Shape_Length' are added by default by the ArcGIS software—these fields are not officially part of RiverATLAS.

Column	***Description***
HYRIV_ID	Unique identifier for each river reach. The code consists of 8 digits: • The first digit represents the region: 1 = Africa; 2 = Europe; 3 = Siberia; 4 = Asia; 5 = Australia; 6 = South America; 7 = North America; 8 = Arctic; 9 = Greenland • The other 7 digits represent a unique identifier within the river network
NEXT_DOWN	HYRIV_ID of the next downstream line segment. This field can be used for navigation (up- and downstream) within the river network. The value '0' indicates a line with no downstream connection, i.e., the last river reach draining into the ocean or into an inland sink. Note that endorheic rivers are identified in the 'ENDORHEIC' field.
MAIN_RIV	HYRIV_ID of the most downstream reach of the connected river basin. This field indicates the ID of the most downstream reach of the river basin and can be used to identify the entire river network that belongs to this basin (by querying for that ID). Note: if small endorheic river networks are nested within a larger surrounding river basin, users may want to include these as part of the larger basin, despite a missing fluvial connection. These topologic relationships can be analyzed by joining the sub-basin table of HydroBASINS (via column 'HYBAS_L12' below) which offers some additional information about 'virtual flow connections' (see Technical Documentation of HydroBASINS for more details).
LENGTH_KM	Length of the river reach segment, in kilometers.
DIST_DN_KM	Distance from the reach outlet, i.e., the most downstream pixel of the reach, to the final downstream location along the river network, in kilometers. This downstream location is either the pour point into the ocean or an endorheic sink.
DIST_UP_KM	Distance from the reach outlet, i.e., the most downstream pixel of the reach, to the most upstream location along the river network, in kilometers. The most upstream location is the furthest upstream point from this reach on the watershed divide.
CATCH_SKM	Area of the catchment that contributes directly to the individual reach, in square kilometers. The catchment only relates to the reach itself, while the contributing area of all upstream reaches is not included (see next column).
UPLAND_SKM	Total upstream area, in square kilometers, calculated from the headwaters to the pour point (i.e., the most downstream pixel) of the reach. The upstream area only comprises the directly connected watershed area, i.e., it does not include endorheic regions that may be nested within the larger basin.
ENDORHEIC	Indicator for endorheic (inland) basins without surface flow connection to the ocean: 0 = not part of an endorheic basin; 1 = part of an endorheic basin.
DIS_AV_CMS	Average long-term discharge estimate for river reach, in cubic meters per second. See Technical Documentation of HydroRIVERS for more information.

ORD_STRA	Indicator of river order following the Strahler ordering system: order 1 represents headwater streams; when two 1^{st} order streams meet, they form a 2^{nd} order river; when two 2^{nd} order rivers meet, they form a 3^{rd} order river; etc.
ORD_CLAS	Indicator of river order following the classical ordering system: order 1 represents the main stem river from sink to source; order 2 represents all tributaries that flow into a 1^{st} order river; order 3 represents all tributaries that flow into a 2^{nd} order river; etc. This ordering system can be used to identify 'backbone' rivers, i.e., the main stem of a river from source to sink.
ORD_FLOW	Indicator of river order using river flow to distinguish logarithmic size classes: order 1 represents river reaches with a long-term average discharge ≥ 100,000 m^3/s; order 2 represents river reaches with a long-term average discharge ≥ 10,000 m^3/s and < 100,000 m^3/s; … order 9 represents river reaches with a long-term average discharge ≥ 0.001 m^3/s and < 0.01 m^3/s; and order 10 represents river reaches with a long-term average discharge < 0.001 m^3/s (i.e., 0 in the provided data due to rounding to 3 digits).
HYBAS_L12	HYBAS_ID of the corresponding HydroBASINS sub-basin in which the river reach resides. This ID refers to HydroBASINS at Pfafstetter level 12 (without lakes).

Appendix 3: *Pre-existing attributes of HydroLAKES included in LakeATLAS version 1.0*

LakeATLAS includes the following attribute columns of HydroLAKES (for more details see Technical Documentation of HydroLAKES at https://www.hydrosheds.org/hydrolakes):

Note that in the shapefile format the field 'FID' and in the geodatabase format the field 'OBJECTID' as well as (for polygon layers) the fields 'Shape_Length' and Shape_Area' are added by default by the ArcGIS software—these fields are not officially part of LakeATLAS.

Column	***Description***
Hylak_id	Unique lake identifier. Values range from 1 to 1,427,688.
Lake_name	Name of lake or reservoir. This field is currently only populated for lakes with an area of at least 500 km^2; for large reservoirs where a name was available in the GRanD database; and for smaller lakes where a name was available in the GLWD database.
Country	Country that the lake (or reservoir) is located in. Note that for the creation of this attribute, international or transboundary lakes were assigned to the country in which the lake's corresponding pour point is located; assignments may thus be arbitrary for pour points that fall on country boundaries.
Continent	Continent that the lake (or reservoir) is located in. Geographic continent: Africa, Asia, Europe, North America, South America, or Oceania (Australia and Pacific Islands)
Poly_src	Source of original lake polygon: CanVec; SWBD; MODIS; NHD; ECRINS; GLWD; GRanD; or Other More information on these data sources can be found in the Technical Documentation of HydroLAKES.
Lake_type	Indicator for lake type: 1: Lake 2: Reservoir 3: Lake control (i.e., natural lake with regulation structure) Note that the default value for all water bodies is 1, and only those water bodies explicitly identified as other types (mostly based on information from the GRanD database) have other values; hence the type 'Lake' also includes all unidentified (typically small) human-made reservoirs and regulated lakes.
Grand_id	ID of the corresponding reservoir in the GRanD database, or value 0 for no corresponding GRanD record. This field can be used to join additional attributes from the GRanD database.
Lake_area	Lake surface area (i.e., polygon area), in square kilometers.
Shore_len	Length of shoreline (i.e., polygon outline), in kilometers.

Shore_dev	Shoreline development, measured as the ratio between shoreline length and the circumference of a circle with the same area. A lake with the shape of a perfect circle has a shoreline development of 1, while higher values indicate increasing shoreline complexity.
Vol_total	Total lake or reservoir volume, in million cubic meters (1 mcm = 0.001 km^3). For most polygons, this value represents the total lake volume as estimated using the geostatistical modeling approach by Messager et al. (2016). However, where either a reported lake volume (for lakes ≥ 500 km^2) or a reported reservoir volume (from GRanD database) existed, the total volume represents this reported value. In cases of regulated lakes, the total volume represents the larger value between reported reservoir and modeled or reported lake volume. Column 'Vol_src' provides additional information regarding these distinctions.
Vol_res	Reported reservoir volume, or storage volume of added lake regulation, in million cubic meters (1 mcm = 0.001 km^3). 0: no reservoir volume
Vol_src	1: 'Vol_total' is the reported total lake volume from literature 2: 'Vol_total' is the reported total reservoir volume from GRanD or literature 3: 'Vol_total' is the estimated total lake volume using the geostatistical modeling approach by Messager et al. (2016)
Depth_avg	Average lake depth, in meters. Average lake depth is defined as the ratio between total lake volume ('Vol_total') and lake area ('Lake_area').
Dis_avg	Average long-term discharge flowing through the lake, in cubic meters per second. This value is derived from modeled runoff and discharge estimates provided by the global hydrological model WaterGAP, downscaled to the 15 arc-second resolution of HydroSHEDS (see Technical Documentation of HydroLAKES for more details) and is extracted at the location of the lake pour point. Note that these model estimates contain considerable uncertainty, in particular for very low flows. -9999: no data as lake pour point is not on HydroSHEDS landmask
Res_time	Average residence time of the lake water, in days. The average residence time is calculated as the ratio between total lake volume ('Vol_total') and average long-term discharge ('Dis_avg'). Values below 0.1 are rounded up to 0.1 as shorter residence times seem implausible (and likely indicate model errors). -1: cannot be calculated as 'Dis_avg' is 0 -9999: no data as lake pour point is not on HydroSHEDS landmask
Elevation	Elevation of lake surface, in meters above sea level. This value was primarily derived from the EarthEnv-DEM90 digital elevation model at 90 m pixel resolution by calculating the majority pixel elevation found within the lake boundaries. To remove some artefacts inherent in this DEM for northern latitudes, all lake values that showed negative elevation for the area north of 60°N were substituted with results using the coarser GTOPO30 DEM of USGS at 1 km pixel resolution, which ensures land surfaces ≥0 in this region. Note that due to the remaining uncertainties in the EarthEnv-DEM90 some small negative values occur along the global ocean coastline south of 60°N which may or may not be correct.

Slope_100	Average slope within a 100 meter buffer around the lake polygon, in degrees. This value is derived from the EarthEnv-DEM90 digital elevation model at 90 m pixel resolution. Slopes for each pixel were computed with latitudinal corrections for the distortion in the XY spacing of geographic coordinates by approximating the geodesic distance between cell centers. For 12 lakes located above the northern limit of the EarthEnv-DEM90 digital elevation model (83°N), slopes were computed from the GTOPO30 DEM of USGS at 1 km pixel resolution. -1: slope values were not calculated for the largest lakes (polygon area ≥ 500 km^2)
Wshd_area	Area of the watershed associated with the lake, in square kilometers. The watershed area is calculated by deriving and measuring the upstream contribution area to the lake pour point using the HydroSHEDS drainage network map at 15 arc-second resolution. -9999: no data as lake pour point is not on HydroSHEDS landmask
Pour_long	Longitude of the lake pour point, in decimal degrees.
Pour_lat	Latitude of the lake pour point, in decimal degrees.
HYBAS_L12	HYBAS_ID of the corresponding HydroBASINS sub-basin in which the river reach resides. This ID refers to HydroBASINS at Pfafstetter level 12 (without lakes).
HYRIV_RCH	HYRIV_ID of the corresponding HydroRIVERS stream reach into which the lake drains at its pour point location. A HYRIV_RCH value of zero indicates that the lake's pour point is not located directly on a river reach that is depicted in HydroRIVERS but lies offstream.
HYRIV_CAT	HYRIV_ID of the corresponding HydroRIVERS reach catchment (i.e., the catchment that directly drains into the stream reach) in which the lake's pour point resides. Lakes with pour points directly located on a stream reach have the same HYRIV_RCH and HYRIV_CAT value. A HYRIV_CAT value of zero indicates that the lake's pour point is not located within the catchment of a river reach that is depicted in HydroRIVERS; this can be the case for small catchments that drain directly into the ocean or into an inland (endorheic) sink.

***Appendix 4:** Attributes included in version 1.0 of HydroATLAS (for details see HydroATLAS catalogs)*

HydroATLAS Attributes (version 1.0)						
ID	**Category**	**Variable**	**Source Data**	**Citation**	**Column(s)**	**Count**
H01	Hydrology	Natural Discharge	WaterGAP v2.2	Döll et al. 2003	dis_m3_---	x3
H02	Hydrology	Land Surface Runoff	WaterGAP v2.2	Döll et al. 2003	run_mm_---	x1
H03	Hydrology	Inundation Extent	GIEMS-D15	Fluet-Chouinard et al. 2015	inu_pc_---	x6
H04	Hydrology	Limnicity (Percent Lake Area)	HydroLAKES	Messager et al. 2016	lka_pc_---	x2
H05	Hydrology	Lake Volume	HydroLAKES	Messager et al. 2016	lkv_mc_---	x1
H06	Hydrology	Reservoir Volume	GRanD v1.1	Lehner et al. 2011	rev_mc_---	x1
H07	Hydrology	Degree of Regulation	HydroSHEDS & GRanD	Lehner et al. 2011	dor_pc_---	x1
H08	Hydrology	River Area	HydroSHEDS & WaterGAP	Lehner & Grill 2013	ria_ha_---	x2
H09	Hydrology	River Volume	HydroSHEDS & WaterGAP	Lehner & Grill 2013	riv_tc_---	x2
H10	Hydrology	Groundwater Table Depth	Global Groundwater Map	Fan et al. 2013	gwt_cm_---	x1
P01	Physiography	Elevation	EarthEnv-DEM90	Robinson et al. 2014	ele_mt_---	x4
P02	Physiography	Terrain Slope	EarthEnv-DEM90	Robinson et al. 2014	slp_dg_---	x2
P03	Physiography	Stream Gradient	EarthEnv-DEM90	Robinson et al. 2014	sgr_dk_---	x1
C01	Climate	Climate Zones	GEnS	Metzger et al. 2013	clz_cl_---	x1
C02	Climate	Climate Strata	GEnS	Metzger et al. 2013	cls_cl_---	x1
C03	Climate	Air Temperature	WorldClim v1.4	Hijmans et al. 2005	tmp_dc_---	x16
C04	Climate	Precipitation	WorldClim v1.4	Hijmans et al. 2005	pre_mm_---	x14
C05	Climate	Potential Evapotranspiration	Global-PET v1	Zomer et al. 2008	pet_mm_---	x14
C06	Climate	Actual Evapotranspiration	Global Soil-Water Balance	Trabucco & Zomer 2010	aet_mm_---	x14
C07	Climate	Global Aridity Index	Global Aridity Index v1	Zomer et al. 2008	ari_ix_---	x2
C08	Climate	Climate Moisture Index	WorldClim & Global-PET	Hijmans et al. 2005	cmi_ix_---	x14
C09	Climate	Snow Cover Extent	MODIS/Aqua	Hall & Riggs 2016	snw_pc_---	x15
L01	Landcover	Land Cover Classes	GLC2000	Bartholomé & Belward 2005	glc_cl_---	x1
L02	Landcover	Land Cover Extent	GLC2000	Bartholomé & Belward 2005	glc_pc_---	x44
L03	Landcover	Potential Natural Vegetation Classes	EarthStat	Ramankutty & Foley 1999	pnv_cl_---	x1
L04	Landcover	Potential Natural Vegetation Extent	EarthStat	Ramankutty & Foley 1999	pnv_pc_---	x30
L05	Landcover	Wetland Classes	GLWD	Lehner & Döll 2004	wet_cl_---	x1
L06	Landcover	Wetland Extent	GLWD	Lehner & Döll 2004	wet_pc_---	x22
L07	Landcover	Forest Cover Extent	GLC2000	Bartholomé & Belward 2005	for_pc_---	x2
L08	Landcover	Cropland Extent	EarthStat	Ramankutty et al. 2008	crp_pc_---	x2
L09	Landcover	Pasture Extent	EarthStat	Ramankutty et al. 2008	pst_pc_---	x2
L10	Landcover	Irrigated Area Extent (Equipped)	HID v1.0	Siebert et al. 2015	ire_pc_---	x2
L11	Landcover	Glacier Extent	GLIMS	GLIMS & NSIDC 2012	gla_pc_---	x2
L12	Landcover	Permafrost Extent	PZI	Gruber 2012	prm_pc_---	x2
L13	Landcover	Protected Area Extent	WDPA	IUCN & UNEP-WCMC 2014	pac_pc_---	x2
L14	Landcover	Terrestrial Biomes	TEOW	Dinerstein et al. 2017	tbi_cl_---	x1
L15	Landcover	Terrestrial Ecoregions	TEOW	Dinerstein et al. 2017	tec_cl_---	x1
L16	Landcover	Freshwater Major Habitat Types	FEOW	Abell et al. 2008	fmh_cl_---	x1
L17	Landcover	Freshwater Ecoregions	FEOW	Abell et al. 2008	fec_cl_---	x1
S01	Soils & Geology	Clay Fraction in Soil	SoilGrids1km	Hengl et al. 2014	cly_pc_---	x2
S02	Soils & Geology	Silt Fraction in Soil	SoilGrids1km	Hengl et al. 2014	slt_pc_---	x2
S03	Soils & Geology	Sand Fraction in Soil	SoilGrids1km	Hengl et al. 2014	snd_pc_---	x2
S04	Soils & Geology	Organic Carbon Content in Soil	SoilGrids1km	Hengl et al. 2014	soc_th_---	x2
S05	Soils & Geology	Soil Water Content	Global Soil-Water Balance	Trabucco & Zomer 2010	swc_pc_---	x14
S06	Soils & Geology	Lithological Classes	GLiM	Hartmann & Moosdorf 2012	lit_cl_---	x1
S07	Soils & Geology	Karst Area Extent	Rock Outcrops v3.0	Williams & Ford 2006	kar_pc_---	x2
S08	Soils & Geology	Soil Erosion	GloSEM v1.2	Borrelli et al. 2017	ero_kh_---	x2
A01	Anthropogenic	Population Count	GPW v4	CIESIN 2016	pop_ct_---	x2
A02	Anthropogenic	Population Density	GPW v4	CIESIN 2016	ppd_pk_---	x2
A03	Anthropogenic	Urban Extent	GHS S-MOD v1.0 (2016)	Pesaresi & Freire 2016	urb_pc_---	x2
A04	Anthropogenic	Nighttime Lights	Nighttime Lights v4	Doll 2008	nli_ix_---	x2
A05	Anthropogenic	Road Density	GRIP v4	Meijer et al. 2018	rdd_mk_---	x2
A06	Anthropogenic	Human Footprint	Human Footprint v2	Venter et al. 2016	hft_ix_---	x4
A07	Anthropogenic	Global Administrative Areas	GADM v2.0	University of Berkeley 2012	gad_id_---	x1
A08	Anthropogenic	Gross Domestic Product	GDP PPP v2	Kummu et al. 2018	gdp_ud_---	x3
A09	Anthropogenic	Human Development Index	HDI v2	Kummu et al. 2018	hdi_ix_---	x1