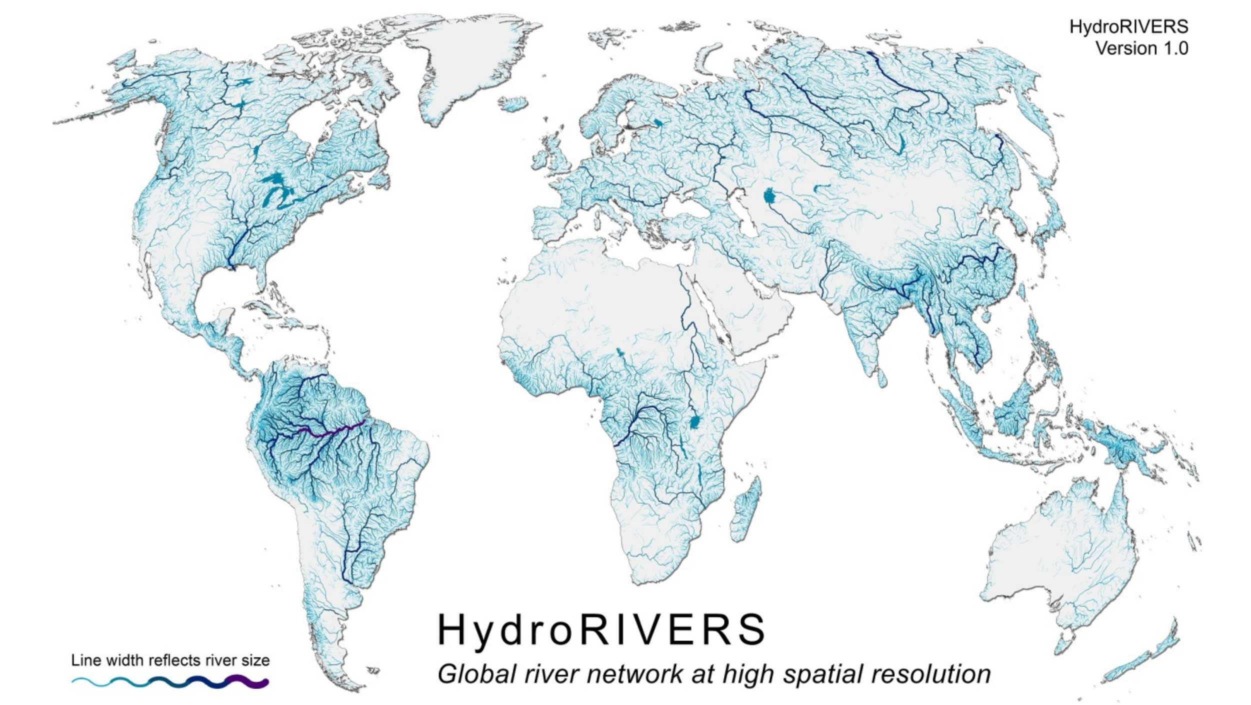
HydroRIVERS

Global river network delineation
derived from HydroSHEDS data at 15 arc-second resolution

Technical Documentation Version 1.0

prepared by Bernhard Lehner
(bernhard.lehner@mcgill.ca)
October, 2019



1. Overview and background

This documentation accompanies the HydroRIVERS line layers that depict a river network delineation at global scale. The goal of this product is to provide a seamless global coverage of consistently sized river reaches, supported by geometric information that allows for basic analysis of river network topology such as up- and downstream connectivity and distances.

River reaches provide important geospatial units for many applications, but at a global scale there is a lack of high-quality mapping sources. The HydroSHEDS database (Hydrological data and maps based on SHuttle Elevation Derivatives; Lehner et al. 2008; for more information see https://www.hydrosheds.org) provides hydrographic data layers that allow for the derivation of river networks for any given location based on the near-global, high-resolution SRTM digital elevation model. Using this hydrographic information at 15 arc-second spatial resolution, rivers were delineated in a consistent manner. The resulting line layers are termed HydroRIVERS and represent a subset of the HydroSHEDS database.

The HydroRIVERS product has been developed on behalf of World Wildlife Fund US (WWF), Washington DC, USA, in collaboration with McGill University, Montreal, Canada. Major funding for the underpinning HydroSHEDS project was provided to WWF by Sealed Air Corporation.

2. Methods and data characteristics

2.1 Extraction of line network

The HydroRIVERS data layers were derived from World Wildlife Fund's HydroSHEDS data (Lehner et al. 2008; Lehner and Grill 2013) based on a grid resolution of 15 arc-seconds (approximately 500 m at the equator). For more information on the HydroSHEDS database please refer to its Technical Documentation at https://www.hydrosheds.org.

A global river network delineation has been extracted from HydroSHEDS and is available as a stand-alone product termed HydroRIVERS. The extraction process used the flow direction grid of HydroSHEDS at 15 arc-second resolution, as well as an auxiliary grid of upstream catchment areas (corrected for latitudinal distortions in pixel sizes due to the given geographic projection) and a grid of long-term average discharge estimates (see section 2.2 below for details).

For version 1.0 of HydroRIVERS, streams 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, or both. Streams smaller than these thresholds were not extracted as they are increasingly unreliable in their spatial representation due to the inherent uncertainties in the underpinning global geometric and hydrologic data. All identified stream pixels at 15 arc-second resolution were then converted into vector format to produce a line network consisting of individual segments, i.e. river reaches (Figure 1). The global coverage of HydroRIVERS encompasses a total of 8,477,883 individual river reaches with an average length of 4.23 km, representing a total of 35.85 million km of rivers globally.

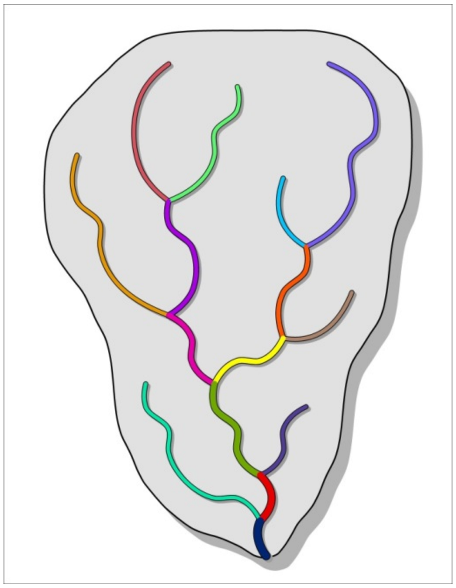


Figure 1: *Overview of the river reach concept used in the HydroRIVERS dataset. Every river reach, depicted by a line segment in a different color, is defined as a stretch of river between two tributaries, or between the start/end of the network and a tributary.*

It should be noted that the quality of HydroSHEDS data—and thus also of HydroRIVERS—is significantly inferior for regions above 60 degrees northern latitude as there is no underlying SRTM elevation data available and thus a coarser scale DEM has been inserted (HYDRO1k provided by USGS (2000), see http://gcmd.nasa.gov/records/GCMD_HYDRO1k.html).

2.2 Discharge estimates

For the delineation of the river network presented in HydroRIVERS, a discharge threshold was used to define the initiation points of all headwater streams. For this purpose, estimates of long-term (1971-2000) average discharge were derived through a geospatial downscaling procedure (Lehner and Grill 2013) from the 0.5 degree resolution runoff and discharge layers of the global WaterGAP model (Döll et al. 2003; version 2.2 as of 2014), a well-documented and validated integrated water balance model. After downscaling, the global total river flow into all oceans matched the original flow as modeled in WaterGAP within an error margin of 0.13%, indicating no significant distortion of large-scale totals due to the downscaling process. In addition, a validation of the downscaled discharge estimates against observations at 3,003 global gauging stations, provided by the Global Runoff Data Center, Koblenz, Germany, representing river sizes from 0.004 to 180,000 m^3/s, confirmed good overall correlations for long-term average discharges ($R^2 = 0.99$ with 0.2% positive bias and a symmetric mean absolute percentage error sMAPE of 35%, improving to 13% for rivers $\geq$100 m^3/s). Despite these overall good correlations, significant uncertainties were observed in certain regions, in particular areas that are dominated by snow or glacier runoff, large wetland complexes, or arid and semi-arid conditions.

2.3 Relationship to the HydroBASINS product and associated features

A second stand-alone product derived from the HydroSHEDS database is the delineation and nested breakdown of sub-basins at different spatial scales provided by the HydroBASINS dataset, which is also available at https://www.hydrosheds.org. Details and specifications of HydroBASINS are described in its respective Technical Documentation. While each river reach of HydroRIVERS resides in exactly one sub-basin of HydroBASINS (one-to-one relationship), each sub-basin can contain none, one, or multiple river reaches (one-to-many relationship). Each river reach can be joined or linked to its corresponding sub-basin in which it resides—at the highest spatial resolution of Pfafstetter level 12—via the HydroBASINS ID that is provided as an attribute column in the HydroRIVERS dataset (column *HYBAS_ID*, see section 3.2 below). Using the additional information gained by joining the attribute information of the corresponding sub-basin, in particular its Pfafstetter code, additional spatial relationships can be analyzed or linked, such as subsequent associations to other (coarser) Pfafstetter levels (1-11).

As a particular characteristic, HydroBASINS provides additional information on endorheic basins (i.e., inland sinks that are not connected to the sea). Many large river basins contain small endorheic sub-basins inside them or adjacent to their watershed divide. For example, the Nile Basin contains many small endorheic basins in the dry middle region which are not connected to the main river via surface water flows yet are typically considered to be part of the overall basin (as they may be hydrologically connected via groundwater or during flood events). In order to allow for these endorheic basins to be lumped with the larger river basin, and to enable topological queries in which the endorheic discontinuities can be eliminated (traversed) to create contiguous regions, some endorheic sinks have been assigned a 'virtual' connection to an appropriate downstream sub-basin. These virtual connections can be identified in the attribute table of HydroBASINS (see its respective Technical Documentation). By linking river reaches to sub-basins, these virtual connections can be assessed for rivers as well.

3. Data format and distribution

3.1 File name syntax

HydroRIVERS data layers are provided in two spatial extents:

- as a seamless, fully global coverage; and
- as regional tiles (see Figure 2 for definition of regions).

File names follow the syntax:

- ***HydroRIVERS_v10*** (for global coverage); or
- ***HydroRIVERS_v10_XX*** (for regional tiles), where *XX* indicates the region.

The regional extents 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

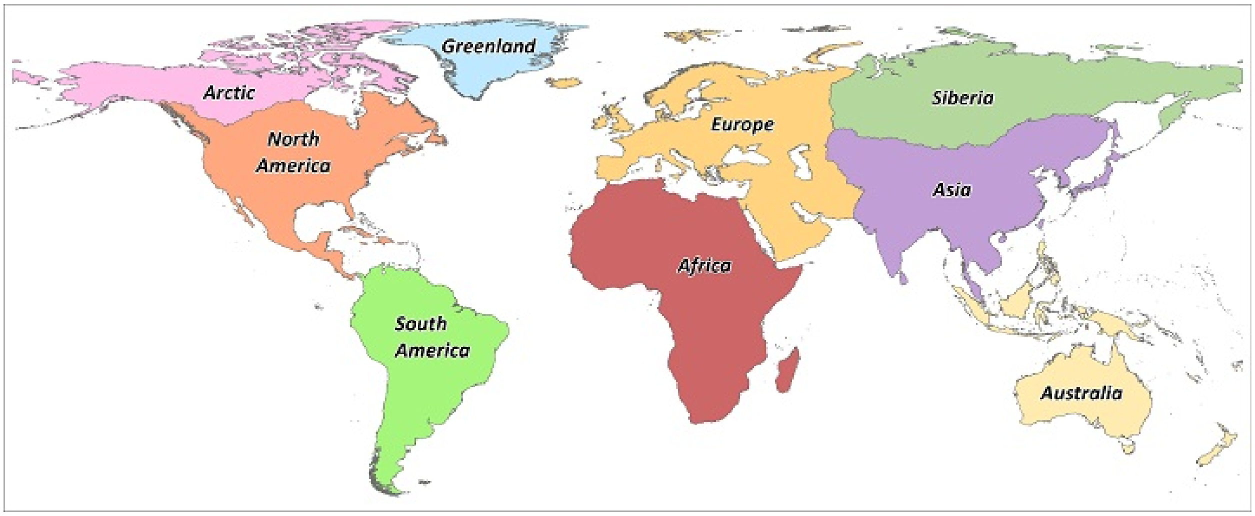


Figure 2: *Spatial extent of regional tiles of HydroRIVERS data files.*

3.2 Attribute table

Each HydroRIVERS data layer contains an attribute table with the following column structure and information:

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 section 2.2 for more information.
ORD_STRA	Indicator of river order following the Strahler ordering system: order 1 represents headwater streams; when two 1st order streams meet, they form a 2nd order river; when two 2nd order rivers meet, they form a 3rd 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 1st order river; order 3 represents all tributaries that flow into a 2nd 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).

3.3 Vector data format and projection

The line network datasets of HydroRIVERS are distributed both in ESRI® 'geodatabase' and 'shapefile' formats. Each HydroRIVERS shapefile consists of five main files (.dbf, .sbn, .sbx, .shp, .shx), and projection information is provided in an ASCII text file (.prj). All HydroRIVERS data layers are provided in geographic (latitude/longitude) projection, referenced to datum WGS84. In ESRI® software this projection is defined by the geographic coordinate system GCS_WGS_1984 and datum D_WGS_1984.

3.4 Data distribution

HydroRIVERS data are available electronically in compressed zip file format from https://www.hydrosheds.org. To use the data files, the zip files must first be decompressed. Each zip file includes a copy of the HydroRIVERS Technical Documentation.

4. Disclaimer and acknowledgement

4.1 License agreement

4.2 Acknowledgement and citation

We kindly ask users to cite HydroRIVERS in any published material produced using the data. If possible, online links to the HydroSHEDS hosting website (https://www.hydrosheds.org) should be provided.

Citations and acknowledgements of the HydroRIVERS data should be made as follows:

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. Data is available at www.hydrosheds.org.

5. References

Döll, P., Kaspar, F., Lehner, B. (2003): A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology, 270, 105-134.

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.

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

USGS – U.S. Geological Survey (2000): HYDRO1k Elevation Derivative Database. USGS EROS Data Center, Sioux Falls, SD.