HydroSHEDS

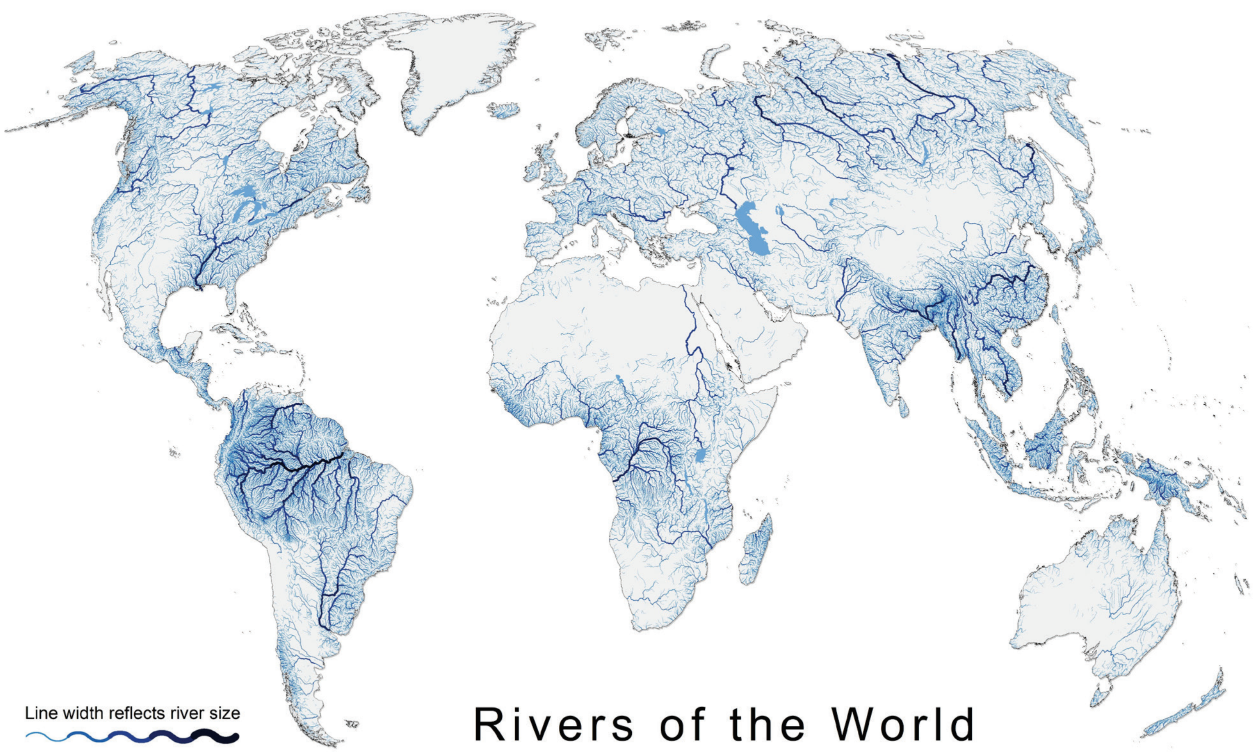
A global hydrographic database of integrated river network and sub-basin delineations at high spatial resolution

Technical Documentation version 2.0.0

Data version 2.0

prepared by Bernhard Lehner on behalf of Confluvio Inc. (bernhard.lehner@confluvio.com)

July 2026



1. Background and introduction

HydroSHEDS provides hydrographic information in a consistent and comprehensive format for regional and global-scale applications. HydroSHEDS offers a suite of geo-referenced datasets in raster and vector format, including stream networks, watershed boundaries, drainage directions, and ancillary data layers such as flow accumulations, flow distances, and river topology information. The overarching goal of HydroSHEDS is to offer application-ready baseline data in support of regional and global watershed analyses, hydrological modeling, and freshwater conservation planning at a quality, resolution and extent that have previously been unachievable.

Since its introduction in 2008, version 1 of the HydroSHEDS database (Lehner et al. 2008) has transformed large-scale hydro-ecological research and applications worldwide by offering standardized spatial units for hydrological assessments. At its core, HydroSHEDS provides digital hydrographic information that can be applied in Geographic Information Software (GIS) or hydrological models to delineate river networks and catchment boundaries at multiple scales, from local to global. Its various data layers form the basis for applications in a wide range of disciplines including environmental, conservation, socioeconomic, human health, and sustainability studies. As version 1 of HydroSHEDS was derived from SRTM elevation data, coverage above 60° northern latitude (i.e., largely the Arctic) is missing for the 3 arc-second product and is of low quality for coarser products. To overcome this limitation, and to improve overall quality of the resulting products, version 2 of HydroSHEDS is currently extracted from a more recent digital elevation model covering the entire global land surface including northern latitudes: the TanDEM-X dataset (TerraSAR-X add-on for Digital Elevation Measurement), created in partnership between the German Aerospace Agency (DLR) and Airbus.

The development of version 2 of HydroSHEDS follows the same basic design principles and specifications as version 1. DLR has preprocessed the original 0.4 arc-second (12 m) TanDEM-X data to create a hydrologically pre-conditioned version at 1 arc-second (30 m) resolution. In this step, corrections with high-resolution vegetation and settlement maps are applied to reduce distortions caused by vegetation cover and in built-up areas. Following this preprocessing, refined hydrological optimization and correction algorithms are used to derive the drainage pathways, including improved 'stream-burning' techniques that incorporate recent data products such as high- resolution terrestrial open water masks, advanced identification of natural sinks and depressions, and improved tracing of drainage pathways as centerlines in global lake and river maps.

HydroSHEDS version 2 has been developed in partnership between World Wildlife Fund (WWF), McGill University, the German Aerospace Center (DLR), and Confluvio Inc. Major funding for this project was provided by a multitude of sponsors (see https://www.hydrosheds.org/sponsors for more details).

This documentation describes partial product releases under version 2 of HydroSHEDS and will be updated as new regions and data layers become available.

For more information, please visit https://www.hydrosheds.org.

Legacy products of version 1 are also available at the HydroSHEDS website.

The development and characteristics of HydroSHEDS v2.0 are not yet fully described in a scientific publication. In the interim, citations and acknowledgements of the database should be made as follows:

Lehner, B., Roth, A., Huber, M., Anand, M., Thieme, M. (2022). A sharper look at the world's rivers and catchments. Eos, 103, https://doi.org/10.1029/2022EO220167.

2. Methods

Note that HydroSHEDS v2 is currently under development and thus only an abbreviated version of the methods is available. This documentation will be updated as new data layers and full methodological descriptions become available.

Figure 1 shows a simplified flowchart depicting the main steps carried out to generate the core layers of HydroSHEDS v2.0. The workflow can be separated into (a) preconditioning processes, (b) conditioning of the DEM using automated processing, and (c) refinement of the conditioned DEM using manual corrections. From the final conditioned DEM, a flow direction grid is extracted which then serves as the source layer for follow-on products.

The main steps during preconditioning include the customized resampling of original TanDEM-X data from 0.4 arc-second resolution to the target resolution of 1 arc-seconds (approximately 30 m at the equator), data tiling and mosaicking, sink identification, vegetation and urban corrections, and water mask creation. The workflow uses a tiled grid approach to optimize computational efficiency, yet tiles are mosaicked at the end to provide seamless products. Each tile undergoes a series of automated hydrologic conditioning steps, applying DEM surface refinement techniques such as corrections for vegetation and urban distortions, removal of outliers, water mask smoothing, flow line generation through the center of water bodies, customized filling of anomalous depressions (pits and mines), hydrological filtering, and others. Then, a semi-automated stream 'imprint' technique (see details below) is applied with the goal of improving the resulting river network through customized stream burning and barrier removal techniques. Finally, an extensive manual correction process is performed which includes the generation of local correction lines based on high-resolution remote sensing imagery and ancillary datasets as well as the validation and adjustment of identified natural depressions (see details below).

Once all corrections are applied to the original DEM, initial flow directions are derived using the conditioned DEM in ESRI's "Derive Continuous Flow" tool; these initial flow directions are then combined with a customized DEM carving algorithm (developed by Confluvio Inc.) to produce a final conditioned DEM surface that matches, as closely as possible, the drainage directions as created by the "Derive" tool. From this elevation surface, the final flow directions are calculated using ESRI's geodesic D8 algorithm. Standard GIS tools from the ESRI Spatial Analyst Hydrology toolbox are used to generate additional data layers, including flow accumulations, river lines, and sub-basin polygons.

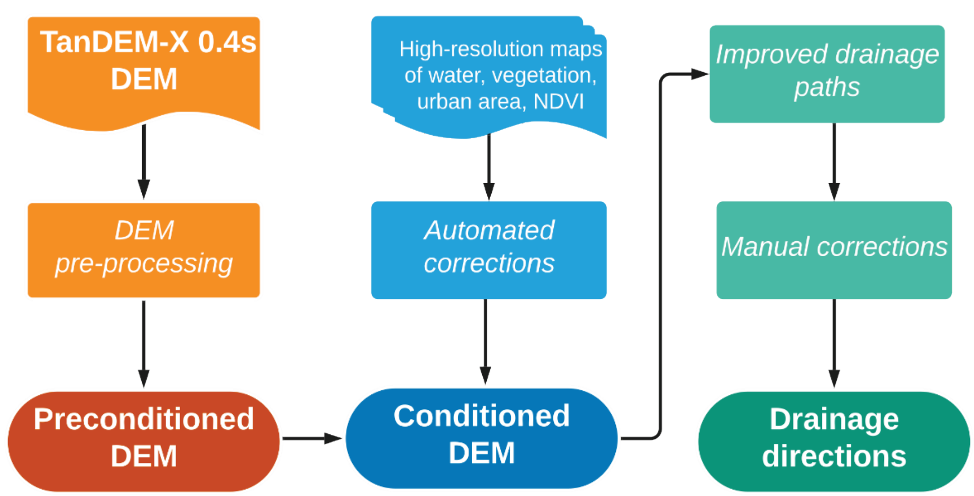


Figure 1: Overview of main processing steps for HydroSHEDS v2.

Manual corrections and stream 'imprinting'

As part of the DEM conditioning process, corrections are applied based on generating initial river lines followed by a thorough review of the results. If discrepancies are detected between the DEM-derived river lines and background satellite imagery or other auxiliary river maps, the spatial analyst can choose to add correction lines to redirect the river to its correct path. This process, often called river burning, is applied in an iterative approach, supported by a series of customized GIS tools, and repeated until flow paths are deemed sufficiently accurate.

In addition to these individual manual corrections, preselected OpenStreetMap (OSM) lines are incorporated through a complementary process termed 'imprinting' to expand the coverage of corrections and therefore improve the quality of the final river network. This process involves a custom-developed subtle stream burning technique which nudges the streams (especially smaller ones) to match OSM lines while not interfering with manual corrections or with established center lines within water surfaces. Results are visually inspected and the imprint layer is adjusted if necessary.

Sink identification and verification

A two-step process is applied to identify and subsequently verify natural depressions that form inland sinks (i.e., surface depressions in which water discontinues to flow, often forming lakes without outflow). During sink identification, a customized tool is used to create an initial set of candidate sinks. After checking depression nesting and hierarchy as well as the resulting river network patterns against satellite imagery and related literature, 'real' sinks are confirmed while spurious or non-natural depressions are discarded. Each sink is manually verified by a senior hydrologist, and further adjustments are made if necessary. For example, sinks can be removed if they represent seasonal depressions that fill and spill during high-flow events; or clusters of numerous small and/or nested sinks can be simplified into one 'representative' sink.

3. Data format and distribution

Version 2 of the HydroSHEDS database is publicly available for download from the HydroSHEDS website at https://www.hydrosheds.org/products. Currently, only a subset of select map layers and regions (North and South America) are available (see Figure 2).

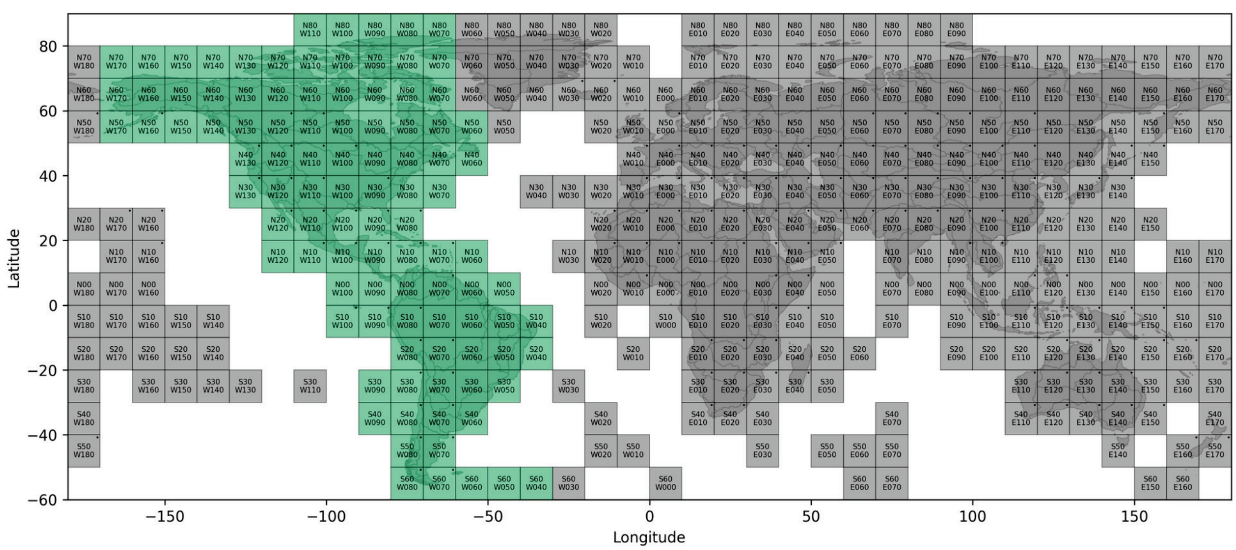


Figure 2: Current extent of available HydroSHEDS v2 data (indicated in green).

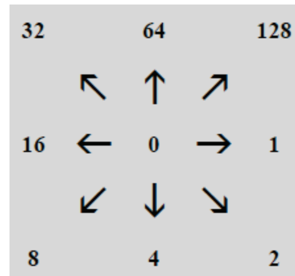
All data layers are projected in a Geographic Coordinate System using the World Geodetic System 1984 (GCS_WGS_1984). Gridded data are available in GeoTIFF format either as 10x10 degree tiles (see Figure 2) or as seamless continental grids. Vector layers are provided by continent in both ESRI geodatabase format and in platform-independent GeoPackage format.

Raster datasets

1) Flow Directions (DIR)

Currently available datasets:

- **north-america_DIR_1s_v2r0.tif**
- **south-america_DIR_1s_v2r0.tif**
- **tiled grids for North and South America**



The drainage direction maps distributed as part of HydroSHEDS v2 define the direction of surface flow from each cell in the conditioned DEM to its steepest down-slope neighbor. Values of flow direction vary from 1 to 128 (see graph on right) and follow the convention adopted by ESRI's D8 single flow direction implementation. Final pixels at the coast and natural sinks are flagged by a flow direction value of 0. The maps are delivered as an unsigned 8-bit raster in a spatial resolution of 1 arc-seconds (approximately 30 m at the equator).

2) Flow Accumulation (ACC = accumulation in number of cells; ACA = accumulation in upstream area)

Currently available datasets:

- **north-america_ACC_1s_v2r0.tif**
- **south-america_ACC_1s_v2r0.tif**
- **north-america_ACA_1s_v2r0.tif**
- **south-america_ACA_1s_v2r0.tif**
- **tiled grids for North and South America**

The flow accumulation maps define the extent of upstream area (in number of cells or in catchment area) draining into each cell. The drainage direction layer is used to define which cells flow into the target cell. The number of accumulated cells is a proxy of the upstream catchment area. However, since the cell size of the HydroSHEDS dataset depends on latitude, the cell accumulation value cannot directly be translated into drainage areas in square kilometres. Therefore, a second flow accumulation map reflecting true catchment areas (in square kilometres) – termed ACA – takes the latitudinal distortions of cells into account. Note that in both instances, the cell itself is added to the flow accumulation value, i.e., the first upstream accumulation value is ‘1’ (in ACC) or the cell’s own area (in ACA) rather than zero. The maps are delivered in a spatial resolution of 1 arc-seconds (approximately 30 m at the equator). To ensure flow accumulation values are correct for exceptionally large upstream cell counts, the resulting accumulation grids are provided as a 64-bit raster.

Vector datasets

3) Stream Network (RIV_STREAMS) and associated Reach Nodes (RIV_NODES)

Currently available datasets:

- north-america_RIV_STREAMS_1s_v2r0
- south-america_RIV_STREAMS_1s_v2r0
- north-america_RIV_NODES_1s_v2r0
- south-america_RIV_NODES_1s_v2r0

A flow accumulation threshold of 1 square kilometer was used to define the initiation of streamlines. The resulting stream extent raster was then converted to a vector river network using ESRI's "Stream to Feature" tool. Some additional information on select stream attributes was generated (see Table 1).

A second vector (point) layer (RIV_NODES) complements the river network with corresponding node features that mark the start and end of each line segment (see Table 2).

Table 1: Attributes of the streams network layer. Note that the list of attributes will change during the development of the global HydroSHEDS v2 database and may not fully reflect the current data version.

Field	Description
STRM_ID	Unique identifier for each stream reach.
STRM_DN	STRM_ID of the next downstream reach. The value '-1' indicates a line with no downstream connection, i.e., the last stream reach draining into the ocean or into an inland sink.
STRM_UP	STRM_ID(s) of the next upstream segment(s). The IDs of all next upstream segments are collated into a string using an underscore between. No value (empty) indicates a line with no upstream reaches (headwater).
NODE_ID_UP	Identifier for the upstream node of this stream reach (see associated node feature class).
NODE_ID_DN	Identifier for the downstream node of this stream reach (see associated node feature class).
LENGTH_KM	Geodesic length of the stream reach segment in kilometers.
UP_CELLS	Total number of upstream cells contributing to this stream reach (measured at the most downstream pixel of the reach before it enters a confluence or the ocean).
UPLAND_SKM	Total upstream geodesic watershed area in square kilometers (measured at the most downstream pixel of the reach before it enters a confluence or the ocean).
MAIN_BAS	Unique ID to distinguish the main hydrological basin to which the stream reach belongs (i.e., a unique ID that identifies the entire basin from headwaters to ocean outlet).

Table 2: Attributes of the node layer associated with the stream network layer. Note that the list of attributes will change during the development of the global HydroSHEDS v2 database and may not fully reflect the current data version.

Field	Description
NODE_ID	Unique identifier for each node point.
NODE_ID_DOWN	NODE_ID of the next downstream node.
NODE_ID_UP	NODE_ID(s) of the next upstream node(s). The IDs of all next upstream nodes are collated into a string using an underscore between.

4) Sub-basins (BAS – beta version only)

Currently available datasets:

- **north-america_BAS_1s_v2r0_BETA**
- **south-america_BAS_1s_v2r0_BETA**

A beta version of a sub-basin layer has been produced following the stream network delineation. This layer represents watershed boundaries and sub-basin delineations that relate directly to the stream network. I.e., there is one sub-basin associated with each river reach which defines the incremental area that directly drains into the reach. Any basin too small to form a river reach (i.e., small coastal basins ≤ 1 km^2 that drain into the ocean in between larger river basins) have been aggregated into relatively equal sizes of about 1 km^2 using an automated algorithm to form lumped coastal drainage units. For these coastal basins, no corresponding river reach exists. Additional information on available attributes, including connectivity information, is shown in Table 3; however, no full set of attributes is available in the beta version.

Table 3: Attributes of the sub-basins data layer. Note that the list of attributes will change during the development of the global HydroSHEDS v2 database and may not fully reflect the current data version.

Field	Description
HYBAS_ID	Unique identifier for each catchment (corresponds to STRM_ID, where possible).
UPSTREAM_ACC	Total number of upstream cells contributing to this sub-basin (measured at the most downstream pixel of the sub-basin before it enters the next confluence or the ocean).
UPLAND_SKM	Total upstream geodesic watershed area contributing to this sub-basin in square kilometers (measured at the most downstream pixel of the sub-basin before it enters the next confluence or the ocean).
STRM_ID	Unique ID to distinguish the stream reach that drains the sub-basin (corresponding to the stream network feature class).

4. License, disclaimer and acknowledgement

4.1 License agreement

All products of version 2 of HydroSHEDS are distributed under a Creative Commons Attribution 4.0 International License. By downloading and using the data the user agrees to the terms and conditions of this license. A copy of the license is available at http://creativecommons.org/licenses/by/4.0/. Notwithstanding this free license, we ask users to refrain from redistributing the data in whole in its original format on other websites without the explicit written permission from the authors. HydroSHEDS v2 is available at https://www.hydrosheds.org/products.

4.2 Disclaimer of warranty

All products of the HydroSHEDS v2 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 any product of the HydroSHEDS v2 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 Citations and acknowledgements

The development and characteristics of HydroSHEDS v2 are not yet fully described in a scientific publication. In the interim, citations and acknowledgements of the database should be made as follows:

Lehner, B., Roth, A., Huber, M., Anand, M., Thieme, M. (2022). A sharper look at the world's rivers and catchments. Eos, 103, https://doi.org/10.1029/2022EO220167.

We kindly ask users to cite the HydroSHEDS v2 database in any published material produced using the data. If possible, links to the HydroSHEDS website should be provided (https://www.hydrosheds.org).

5. References

Lehner, B., Verdin, K., Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. Eos, 89(10): 93–94. https://doi.org/10.1029/2008eo100001