

# **SWOT River Database (SWORD)**

Product Description Document  
Release v17b  
March 2025

Elizabeth H. Altenau (ealtenau@unc.edu)  
University of North Carolina at Chapel Hill, Chapel Hill, NC

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

Michael T. Durand  
The Ohio State University, Columbus, OH

Elyssa Collins  
University of North Carolina at Chapel Hill, Chapel Hill, NC

## Table of Contents

<b>1. Versions</b>	<b>2</b>
<b>2. Overview</b>	<b>5</b>
<b>3. Data Sources</b>	<b>5</b>
<b>4. Data Formats</b>	<b>6</b>
4.1 NetCDF	6
4.2 Geopackage	21
4.3 Shapefiles	21
<b>5. Summary Figures and Statistics</b>	<b>27</b>
<b>6. References</b>	<b>31</b>

## 1. Versions

- Beta v0.1
  - Produced: September 2019
  - First beta version of SWORD
- Beta v0.2
  - Produced: December 2019
  - Added ghost nodes and reaches to the database (Type = 6).
- Beta v0.3
  - Produced: March 2020
  - Adjusted the “wth\_coef” values from 0.5 to 1.
- Beta v0.4
  - Produced: April 2020
  - Fixed formatting issues with multi-dimensional variables (/centerlines/rch\_id, /centerlines/node\_id, /nodes/cl\_ids, /reaches/cl\_ids/, reaches/rch\_id\_up, /reaches/rch\_id\_down) that occurred between v02 and v01.
- Beta v0.5
  - Produced: April 2020
  - Added discharge groups (reaches/MetroMan, reaches/BAM, reaches/HiVDI, reaches/MOMMA, reaches/SADS) and associated variables.
- Beta v0.6
  - Produced: June 2020
  - Improved topology algorithm.
  - Improved ghost node identification.
  - Added unconstrained/constrained discharge variables to netCDFs.
  - Updated GROD dataset used in reach definition.
  - Added “ext\_dist\_coef” attribute to dataset for improving errors caused by lakes-near-rivers in SWOT pixel cloud.
- Beta v0.7
  - Produced: September 2020
  - Added flow accumulation values to prior reach and node products.
  - Added low permeable dam and waterfalls to reach definition.
- Beta v0.8
  - Produced: November 2020
  - Improved local topology algorithm.
  - Improved distance-from-outlet algorithm.
- Beta v0.9 (Public v0)
  - Produced: January 2021
  - Adjusted reach definition to produce slightly longer reaches on average.
  - Added GROD and HydroFALLS ids to products.
  - Included SWOT pass and number of observations in shapefile products.
  - Added climatological ice flag values to netcdf.
- Release v10 (Public v1)
  - Produced: June 2021
  - Corrected missing GROD locations around the equator.
  - First round of manual edits to fix incorrect reach geometries.
  - Added “max\_width” to the netCDF attributes.
- Release v11
  - Produced: July 2021

- Included the Prior Lake Database (PLD) information in reach definition.
- Added “river\_name” and “sbQ\_rel” discharge parameter to netCDF.
- Improved centerline representations in coastal/estuary areas.
- Filled in width=1 node values with reach width values.
- Added “manual\_add” node attribute for identifying the nodes that were manually added to the original GRWL centerlines.
- Release v12 (Public v2)
  - Produced: October 2021
  - Added “meander\_length” and “sinuosity” attributes to the nodes.
  - Improved centerline representations in Peace-Athabasca Delta.
- Release v13
  - Produced: July 2022
  - Added SIC4DVar discharge group and associated variables.
  - Added “low\_slope\_flag” to reaches group.
  - Adjusted “ext\_dist\_coef” values for several reaches on the Yukon River.
- Release v14
  - Produced: November 2022
  - Improved topology in reaches covered by the 1-day CalVal SWOT orbit.
  - Custom reach definition for several Tier1 CalVal sites: Willamette River, Tanana River, Connecticut River, Peace Athabasca Delta, North Saskatchewan River, Sagavanirktok River, Waimak River.
  - Removed duplicated reaches around EU / AS border.
  - Added geopackage file format.
- Release v15
  - Produced: February 2023
  - Corrected topology in the world’s large river systems (i.e. Mississippi, Yukon, Congo, Amazon, etc.).
  - Centerline adjustments to several rivers:
    - Yukon, Kuskokwim, Slave, Mississippi, Atchafalaya, Ob, and Amur Rivers: Update requests by Bo Wang and Laurence Smith
    - Weser, Rhine, and Elbe Rivers: Update requests by Luciana Fenoglio
    - Centerline additions and connectivity improvements in large PLD lakes: Update requests by Jida Wang
    - Amazon River delta additions: Updates by Elizabeth Altenau
  - Added a tributary flag (“trib\_flag”) to the reach and node attributes. The tributary flag indicates whether a larger river identified in MERIT Hydro-Vector, but not in SWORD, is entering a reach or node.
- Release v15b
  - Produced: April 2023
  - Reduced default (max) “ext\_dist\_coef” value from 20 to 5.
  - Corrected “edit\_flag” values in custom calval basins.

- Release v16
  - Produced: August 2023
  - Continued topology updates in the high priority level 2 basins.
  - Reach ID corrections at the level 6 basin level. ~5% of Reach IDs changed between v15 and v17.
  - Removed reaches consisting of a single 30 m centerline point.
  - Updated `ext_dist_coef` values to be static for all nodes within reaches that have many water bodies near the river edge.
- Release v17
  - October 2024
  - Topological updates to ensure consistency.
  - Distance from outlet recalculation based on shortest paths between outlets and headwaters.
  - New variables for nodes and reaches: “`path_freq`”, “`path_order`”, “`path_segs`”, “`main_side`”, “`stream_order`”, “`end_reach`”, “`network`”.
  - Improved geometry for reach shapefiles.
  - Additional channels added for improved network connectivity.
  - New reach and node ids that better reflect the improved topology.
  - Corrected node lengths to match reach lengths when summed.
- Release v17b
  - March 2025
  - Type change for 1662 reaches and associated nodes globally.
  - Updates to reach and node lengths and distance from outlet to correct bug in node length calculation in select reaches (<2% globally).

## 2. Overview

The upcoming Surface Water and Ocean Topography (SWOT) satellite mission, planned to launch in 2022, will vastly expand observations of river water surface elevations (WSE), inundation extent, and slope [Biancamaria *et al.*, 2016]. For practical interpretation and application of SWOT measurements, a global prior database of river networks and reaches is required. The **SWOT River Database (SWORD)** was built to support the development of RiverObs, the central algorithm that will process SWOT pixel cloud data into vector products. A major purpose for SWORD is to provide fixed node locations, reach boundaries, and high-resolution reach centerlines in a way that facilitates the generation of SWOT vector products. SWORD provides high-resolution river nodes (200 m) and reaches (~10 km) in vector and netCDF formats with attached hydrologic variables (WSE, width, slope, etc.) as well as a consistent topological system for global rivers 30 m wide and greater.

## 3. Data Sources

SWORD is generated by combining multiple global hydrography databases into one congruent product. This section briefly describes the main data sources that are used in the development of SWORD. Table 1 provides a summary of data sets and the attributes they contribute to the final product. For detailed information regarding the development of SWORD see [Altenau *et al.*, 2021].

*Table 1: Summary of data sets used in the development of SWORD.*

Dataset	Attribute Contribution
Global River Widths from Landsat (GRWL) [Allen & Pavelsky, 2018]	Provides river centerline locations at 30 m resolution and associated width, water body type, and number of channels attributes.
MERIT Hydro [Yamazaki <i>et al.</i> , 2019]	Provides elevation and flow accumulation at 3 arc-second resolution (~90 m at the equator).
HydroBASINS [Lehner & Grill, 2013]	Provides Pfafstetter nested basin codes up to level 6.
Global River Obstruction Database (GROD) [Whittemore <i>et al.</i> , 2020]	Provides global locations of anthropogenic river obstructions along the GRWL river network.
Global Delta Maps [Tessler <i>et al.</i> , 2015]	Provides the spatial extent of 48 of the world's largest river deltas.

SWOT Orbits [ <a href="https://www.aviso.altimetry.fr/en/missions/future-missions/swot/orbit.html">https://www.aviso.altimetry.fr/en/missions/future-missions/swot/orbit.html</a> ]	Provides polygons containing SWOT track coverage for each pass throughout the 21-day cycle orbit.
HydroFALLS [ <a href="http://wp.geog.mcgill.ca/hydrolab/hydrofalls/">http://wp.geog.mcgill.ca/hydrolab/hydrofalls/</a> ]	Provides global locations of waterfalls and natural river obstructions.

## 4. Data Formats

SWORD data is provided in netCDF, geopackage, and shapefile formats. NetCDF and geopackage files are distributed at continental scales, while shapefiles are split into level 2 basins within each continent. HydroBASINS Pfafstetter codes have a total of 9 continental regions, however, for ease of distribution some HydroBASINS regions are grouped under a single continent in the SWORD database. In these cases, the original HydroBASINS Pfafstetter codes remain the same. For example, HydroBASINS Pfafstetter levels representing North America (7), the Arctic (8) and Greenland (9) are all grouped under the North American identifier ('na'). Additionally, HydroBASINS Pfafstetter levels representing Asia (4) and Siberia (3) are grouped under the Asia identifier ('as') in the SWORD database. File syntax denotes the regional information for each file and is described for each format below.

### 4.1 NetCDF

A comprehensive version of SWORD is available in netCDF format. Each netCDF file contains a set of global attributes and three groups of variables for the different spatial scales of the database (/centerlines, /nodes, /reaches). Table 3 provides descriptions of the groups and variables in the netCDF files. NetCDF file names are distributed at continental scales and are defined by a two-digit identifier (Table 2): [continent]\_sword\_v1.nc (*i.e.* na\_sword\_v17.nc).

*Table 2: Continent identifiers in the SWORD database.*

Identifier	Continent
na	North America
eu	Europe/Middle East
as	Asia
sa	South America
af	Africa
oc	Oceania

*Table 3: NetCDF variable and attribute descriptions. (\*These variables contain fill values).*

Group	Variable	Description	Units	Dimensions
global attributes	Name	2 letters identifying the continent (NA – North America, SA – South America, AF – Africa, EU – Europe/Middle East, AS – Asia, OC – Australia/Oceania).	none	N/A
	x_min, x_max, y_min, y_max	Bounding box of longitudes and latitudes included in a file. Note that files may have overlapping boxes.	decimal degrees	N/A
	production date	Date when the files were generated.	none	N/A
/centerlines	cl_id	high-resolution centerline point id	none	[number of points]
	x	longitude of the point ranging from 180°E to 180°W	decimal degrees	[number of points]
	y	latitude of the point ranging from 90°S to 90°N	decimal degrees	[number of points]
	reach_id	id of each reach the high-resolution centerline point is associated with. The format of the id is as follows: CBBBBBRRRRRT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream, <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 –	none	[4, number of points]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		unreliable topology, 6 – ghost reach)		
	node_id	id of each node the high-resolution centerline point is associated with. The format of the id is as follows: CBBBBBRRRRNNNT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream), <b>N</b> = Node id (assigned sequentially within a reach starting at the downstream end working upstream), <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost node)	none	[4, number of points]
/nodes	cl_ids	minimum and maximum high-resolution centerline point ids along each node.	none	[2, number of nodes]
	x	longitude of each node ranging from 180°E to 180°W	decimal degrees	[number of nodes]
	y	latitude of each node, ranging from 90°S to 90°N	decimal degrees	[number of nodes]
	node_id	id of each node. The format of the id is as follows: CBBBBBRRRRNNNT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin code up to	none	[number of nodes]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream), <b>N</b> = Node id (assigned sequentially within a reach starting at the downstream end working upstream), <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost node)		
	node_length	node length measured along the high-resolution centerline points	meters	[number of nodes]
	reach_id	id of the reach each node is associated with. The format of the id is as follows: CBBBBBRRRRT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream), <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach)	none	[number of nodes]
	wse	node average water surface elevation	meters	[number of nodes]
	wse_var	water surface elevation variance along the high-resolution centerline points used to calculate the average water surface elevation for each node	meters <sup>2</sup>	[number of nodes]
	width	node average width	meters	[number of nodes]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

	width_var	width variance along the high-resolution centerline points used to calculate the average width for each node	meters <sup>2</sup>	[number of nodes]
	n_chan_max	maximum number of channels for each node	none	[number of nodes]
	n_chan_mod	mode of the number of channels for each node	none	[number of nodes]
	obstr_type	Type of obstruction for each node based on GROD and HydroFALLS databases. Obstr_type values: 0 - No Dam, 1 - Dam, 2 - Lock, 3 - Low Permeable Dam, 4 - Waterfall.	none	[number of nodes]
	grod_id	The unique GROD ID for each node with obstr_type values 1-3.	none	[number of nodes]
	hfalls_id	The unique HydroFALLS ID for each node with obstr_type value 4.	none	[number of nodes]
	dist_out	distance from the river outlet for each node	meters	[number of nodes]
	facc	maximum flow accumulation value for each node	km <sup>2</sup>	[number of nodes]
	lakeflag	GRWL water body identifier for each node: 0 – river, 1 – lake/reservoir, 2 – canal , 3 – tidally influenced river.	none	[number of nodes]
	max_width	maximum width value across the channel for each node that includes island and bar areas.	meters	[number of nodes]
	wth_coef	coefficient that is multiplied by the width variable to inform the RiverObs search window for pixel cloud points.	none	[number of nodes]

	ext_dist_coef	coefficient that informs the maximum RiverObs search window for pixel cloud points.	none	[number of nodes]
	meander_length	length of the meander that a node belongs to, measured from beginning of the meander to its end in meters. For nodes longer than one meander, the meander length will represent the average length of all meanders belonging to the node.	meters	[number of nodes]
	sinuosity	the total reach length the node belongs to divided by the Euclidean distance between the reach end points.	none	[number of nodes]
	river_name	all river names associated with a node. If there are multiple names for a node they are listed in alphabetical order and separated by a semicolon.	none	[number of nodes]
	manual_add	binary flag indicating whether the node was manually added to the public GRWL centerlines. These nodes were originally given a width = 1, but have since been updated to have the reach width values.	none	[number of nodes]
	edit_flag	numerical flag indicating the type of update applied to SWORD nodes from the previous version. Flag descriptions:  1 - reach type change; 2 - node order change; 3 - reach neighbor change; 41 - flow accumulation update; 42 - elevation update;	none	[number of nodes]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		<p>43 - width update;  44 - slope update;  45 - river name update;  5 - reach id change;  6 - reach boundary change  7 - reach/node addition</p> <p>Nodes where multiple updates have been applied will have each flag number separated by commas, i.e: "41,2".</p>		
	trib_flag	binary flag indicating if a large tributary not represented in SWORD is entering a node. 0 - no tributary, 1 - tributary.	none	[number of nodes]
	path_freq	the number of times a node is traveled to get to any given headwater from the primary outlet.	none	[number of nodes]
	path_order	unique values representing continuous paths from the river outlet to the headwaters. Values are unique within level two Pfafstetter basins. The lowest value is always the longest path from outlet to farthest headwater point in a connected river network. Higher path values branch off from the longest path value to other headwater points.	none	[number of nodes]
	path_segs	unique values indicating continuous river segments between river junctions. Values are unique within level two Pfafstetter basins.	none	[number of nodes]
	main_side	value indicating whether a node is on the main network (0), side network (1), or is a	none	[number of nodes]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		secondary outlet on the main network (2).		
	stream_order	stream order based on the log scale of the path frequency. stream order is calculated for the main network only (see “main_side” description). stream order is not included for side channels which are given a no data value of -9999.	none	[number of nodes]
	end_reach	value indicating whether a reach is a headwater (1), outlet (2), or junction (3) reach. A value of 0 means it is a normal main stem river reach.	none	[number of nodes]
	network	unique value for each connected river network. Values are unique within level two Pfafstetter basins.	none	[number of nodes]
/reaches	cl_ids	minimum and maximum high-resolution centerline point ids along each reach.	none	[2, number of reaches]
	x	longitude of the reach center ranging from 180°E to 180°W	decimal degrees	[number of reaches]
	y	latitude of the reach center ranging from 90°S to 90°N	decimal degrees	[number of reaches]
	x_max, x_min, y_max, y_min	Bounding box of longitudes and latitudes for a reach. Note that reaches may have overlapping boxes.	decimal degrees	[number of reaches]
	reach_id	id of each reach. The format of the id is as follows: CBBBBBRRRRRT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> =	none	[number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream, <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach)		
	reach_length	reach length measured along the high-resolution centerline points	meters	[number of reaches]
	wse	reach average water surface elevation	meters	[number of reaches]
	wse_var	water surface elevation variance along the high-resolution centerline points used to calculate the average water surface elevation for each reach	meters <sup>2</sup>	[number of reaches]
	width	reach average width	meters	[number of reaches]
	width_var	width variance along the high-resolution centerline points used to calculate the average width for each reach	meters <sup>2</sup>	[number of reaches]
	max_width	maximum width value across the channel for each reach that includes island and bar areas.	meters	[number of reaches]
	n_nodes	number of nodes associated with each reach	none	[number of reaches]
	n_chan_max	maximum number of channels for each reach	none	[number of reaches]
	n_chan_mod	mode of the number of channels for each reach	none	[number of reaches]

	obstr_type	Type of obstruction for each reach based on GROD and HydroFALLS databases. Obstr_type values: 0 - No Dam, 1 - Dam, 2 - Lock, 3 - Low Permeable Dam, 4 - Waterfall.	none	[number of reaches]
	grod_id	The unique GROD ID for each reach with obstr_type values 1-3.	none	[number of reaches]
	hfalls_id	The unique HydroFALLS ID for each reach with obstr_type value 4.	none	[number of reaches]
	slope	reach average slope calculated along the high-resolution centerline points	m/km	[number of reaches]
	lakeflag	GRWL water body identifier for each reach: 0 – river, 1 – lake/reservoir, 2 – canal , 3 – tidally influenced river.	none	[number of reaches]
	dist_out	distance from the river outlet for each reach	meters	[number of reaches]
	facc	maximum flow accumulation value for each reach	km <sup>2</sup>	[number of reaches]
	n_rch_up	number of upstream reaches for each reach	none	[number of reaches]
	n_rch_down	number of downstream reaches for each reach	none	[number of reaches]
	rch_id_up	reach ids of the upstream reaches	none	[4, number of reaches]
	rch_id_dn	reach ids of the downstream reaches	none	[4, number of reaches]
	ice_flag	Ice flag values for each SWOT reach are modeled river ice	none	[366, number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		<p>conditions based on an empirical river ice model [Yang <i>et al.</i>, 2019], that takes surface air temperature (SAT) data from ERA5 Land (9 km resolution) as model input.</p> <p>Values include 0 – ice free, 1 – mixed, 2 – ice cover.</p>		
	swot_obs	The maximum number of SWOT passes to intersect each reach during the 21 day orbit cycle.	none	[number of reaches]
	swot_orbits	A list of the SWOT orbit tracks that intersect each reach during the 21 day orbit cycle.	none	[75, number of reaches]
	river_name	all river names associated with a reach. If there are multiple names for a reach they are listed in alphabetical order and separated by a semicolon.	none	[number of reaches]
	low_slope_flag	binary flag where a value of 1 indicates the reach slope is too low for effective discharge estimation.	none	[number of reaches]
	edit_flag	<p>numerical flag indicating the type of update applied to a SWORD reach from the previous version. Flag descriptions:</p> <p>1 - reach type change;                  2 - node order change;                  3 - reach neighbor change;                  41 - flow accumulation update;                  42 - elevation update;                  43 - width update;                  44 - slope update;</p>	none	[number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

		<p>45 - river name update;                      5 - reach id change;                      6 - reach boundary change                      7 - reach/node addition</p> <p>Reaches where multiple updates have been applied will have each flag number separated by commas, i.e: "41,2".</p>		
	trib_flag	binary flag indicating if a large tributary not represented in SWORD is entering a reach. 0 - no tributary, 1 - tributary.	none	[number of reaches]
	path_freq	the number of times a reach is traveled to get to any given headwater from the primary outlet.	none	[number of reaches]
	path_order	unique values representing continuous paths from the river outlet to the headwaters. Values are unique within level two Pfafstetter basins. The lowest value is always the longest path from outlet to farthest headwater point in a connected river network. Higher path values branch off from the longest path value to other headwater points.	none	[number of reaches]
	path_segs	unique values indicating continuous river segments between river junctions. Values are unique within level two Pfafstetter basins.	none	[number of reaches]
	main_side	value indicating whether a node is on the main network (0), side network (1), or is a secondary outlet on the main network (2).	none	[number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

	strm_order	stream order based on the log scale of the path frequency. stream order is not included for side channels which are given a no data value of -9999.	none	[number of reaches]
	end_rch	value indicating whether a reach is a headwater (1), outlet (2), or junction (3) reach. A value of 0 means it is a normal main stem river reach.	none	[number of reaches]
	network	unique value for each connected river network. Values are unique within level two Pfafstetter basins.	none	[number of reaches]
/reaches/area_fits				
	*h_break	the sub-domain boundary values for water surface elevation, calculated by the fitting of curves to height-width data	meters	[4, number of reaches]
	*w_break	the sub-domain width boundary values for water surface elevation	meters	[4, number of reaches]
	*h_variance	height variance for the calculation of the w,h covariance matrix	meters <sup>2</sup>	[number of reaches]
	*w_variance	width variance for the calculation of the w,h covariance matrix	meters <sup>2</sup>	[number of reaches]
	*hw_covariance	covariance between height and width the calculation of the w,h covariance matrix	meters <sup>2</sup>	[number of reaches]
	*fit_coeffs	fit coefficients for the computation of A prime	meters <sup>2</sup>	[2, 3, number of reaches]
	*med_flow_area	the cross-sectional area at median flow	meters <sup>2</sup>	[number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

	*h_err_stdev	height error standard deviation	meters	[number of reaches]
	*w_err_stdev	width error standard deviation	meters	[number of reaches]
	*h_w_nobs	number of observations for w,h covariance matrix	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/MetroMan				
	*Abar	wetted cross-sectional area during median discharge	meters <sup>2</sup>	[number of reaches]
	*ninf	friction relationship coefficient	none	[number of reaches]
	*p	friction relationship exponent	none	[number of reaches]
	*Abar_stdev	standard deviation of Abar	meters <sup>2</sup>	[number of reaches]
	*ninf_stdev	standard deviation of ninf	none	[number of reaches]
	*p_stdev	standard deviation of p	none	[number of reaches]
	*ninf_p_cor	correlation between ninf and p	none	[number of reaches]
	*ninf_Abar_cor	correlation between Abar and ninf	none	[number of reaches]
	*p_Abar_cor	correlation between Abar and p	none	[number of reaches]
	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/BAM				
	*Abar	wetted cross-sectional area during median discharge	meters <sup>2</sup>	[number of reaches]
	*n	manning coefficient	none	[number of reaches]

	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/HiVDI				
	*Abar	wetted cross-sectional area during median discharge	meters^2	[number of reaches]
	*alpha	reciprocal friction relationship coefficient	none	[number of reaches]
	*beta	negative friction relationship exponent	none	[number of reaches]
	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/MOMMA				
	*B	the elevation of zero flow	meters	[number of reaches]
	*H	bankfull water surface elevation	meters	[number of reaches]
	*Save	averaged slope	m/km	[number of reaches]
	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/SADS				
	*Abar	wetted cross-sectional area during median discharge	meters^2	[number of reaches]

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

	*n	manning coefficient	none	[number of reaches]
	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]
/reaches/discharge_models/[unconstrained][constrained]/SIC4DVar				
	*Abar	wetted cross-sectional area during median discharge	meters^2	[number of reaches]
	*n	manning coefficient	none	[number of reaches]
	*sbQ_rel	relative uncertainty of timeseries mean discharge at each reach, specified as a standard deviation	none	[number of reaches]

#### 4.2 Geopackage

SWORD geopackage files are split into two files for nodes and reaches per continental region, where nodes are represented as 200 m spaced points and reaches are represented as polylines. All geopackage files are in geographic (latitude/longitude) projection, referenced to datum WGS84. Attributes included in the node and reach files for both geopackage and shapefile formats are listed in Tables 4 and 5. Geopackage file names are distributed at continental scales and are defined by a two-digit identifier (Table 2): [continent]\_sword\_[nodes/reaches]\_v1.gpkg (*i.e.* *na\_sword\_nodes\_v17.gpkg*; *na\_sword\_reaches\_v17.gpkg*).

#### 4.3 Shapefiles

SWORD shapefiles consist of four main files (.dbf, .prj, .shp, .shx). There are separate shapefiles for nodes and reaches, where nodes are represented as 200 m spaced points and reaches are represented as polylines. All shapefiles are in geographic (latitude/longitude) projection, referenced to datum WGS84. Attributes included in the node and reach files for both geopackage and shapefile formats are listed in Tables 4 and 5. Shapefiles are split into HydroBASINS Pfafstetter level 2 basins (hbXX) within each continent (Table 2) with a naming convention as follows: [continent]\_sword\_[nodes/reaches]\_hb[XX]\_v1.shp (*i.e.* *na\_sword\_nodes\_hb74\_v17.shp*; *na\_sword\_reaches\_hb74\_v17.shp*).

*Table 4: Node shapefile and geopackage attribute descriptions..*

<b>Attribute</b>	<b>Description</b>	<b>Units</b>
x	longitude of each node ranging from 180°E to 180°W	decimal degrees
y	latitude of each node, ranging from 90°S to 90°N	decimal degrees
node_id	id of each node. The format of the id is as follows: CBBBBBRRRRNNNT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin code up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream), <b>N</b> = Node id (assigned sequentially within a reach starting at the downstream end working upstream), <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost node)	none
node_length	node length measured along the high-resolution centerline points	meters
reach_id	id of the reach each node is associated with. The format of the id is as follows: CBBBBBRRRRRT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream), <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach)	none
wse	node average water surface elevation	meters
wse_var	water surface elevation variance along the high-resolution centerline points used to calculate the average water surface elevation for each node	meters^2
width	node average width	meters
width_var	width variance along the high-resolution centerline points used to calculate the average width for each node	meters^2
n_chan_max	maximum number of channels for each node	none
n_chan_mod	mode of the number of channels for each node	none
obstr_type	Type of obstruction for each node based on GROD and HydroFALLS databases. Obstr_type values: 0 - No Dam, 1 - Dam, 2 - Lock, 3 - Low Permeable Dam, 4 - Waterfall.	none
grod_id	The unique GROD ID for each node with obstr_type values 1-3.	none
hfalls_id	The unique HydroFALLS ID for each node with obstr_type value 4.	none

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

dist_out	distance from the river outlet for each node	meters
type	Node type identifier: 1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach.	none
face	Maximum flow accumulation value for each node.	kilometers <sup>2</sup>
lakeflag	GRWL water body identifier for each node: 0 – river, 1 – lake/reservoir, 2 – canal, 3 – tidally influenced river.	none
max_width	maximum width value across the channel for each node that includes island and bar areas.	meters
river_name	all river names associated with a node. If there are multiple names for a node they are listed in alphabetical order and separated by a semicolon.	none
sinuosity	the total reach length the node belongs to divided by the Euclidean distance between the reach end points.	none
meand_len	length of the meander that a node belongs to, measured from beginning of the meander to its end in meters. For nodes longer than one meander, the meander length will represent the average length of all meanders belonging to the node.	meters
manual_add	binary flag indicating whether the nodes was manually added to the public GRWL centerlines. These nodes were originally given a width = 1, but have since been updated to have the reach width values.	none
trib_flag	binary flag indicating if a large tributary not represented in SWORD is entering a node. 0 - no tributary, 1 - tributary.	none
path_freq	the number of times a node is traveled to get to any given headwater from the primary outlet.	none
path_order	unique values representing continuous paths from the river outlet to the headwaters. Values are unique within level two Pfafstetter basins. The lowest value is always the longest path from outlet to farthest headwater point in a connected river network. Higher path values branch off from the longest path value to other headwater points.	none
path_segs	unique values indicating continuous river segments between river junctions. Values are unique within level two Pfafstetter basins.	none
main_side	value indicating whether a node is on the main network (0), side network (1), or is a secondary outlet on the main network (2).	none

strm_order	stream order based on the log scale of the path frequency. stream order is not included for side channels which are given a no data value of -9999.	none
end_rch	value indicating whether a reach is a headwater (1), outlet (2), or junction (3) reach. A value of 0 means it is a normal main stem river reach.	none
network	unique value for each connected river network. Values are unique within level two Pfafstetter basins.	none

*Table 5: Reach shapefile and geopackage attribute descriptions.*

Attribute	Description	Units
x	longitude of the reach center ranging from 180°E to 180°W	decimal degrees
y	latitude of the reach center ranging from 90°S to 90°N	decimal degrees
reach_id	id of each reach. The format of the id is as follows: CBBBBBRRRRRT where <b>C</b> = Continent (the first number of the Pfafstetter basin code), <b>B</b> = Remaining Pfafstetter basin codes up to level 6, <b>R</b> = Reach id (assigned sequentially within a level 6 basin starting at the downstream end working upstream, <b>T</b> = Type (1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach)	none
reach_length	reach length measured along the high-resolution centerline points	meters
wse	reach average water surface elevation	meters
wse_var	water surface elevation variance along the high-resolution centerline points used to calculate the average water surface elevation for each reach	meters <sup>2</sup>
width	reach average width	meters
width_var	width variance along the high-resolution centerline points used to calculate the average width for each reach	meters <sup>2</sup>
n_nodes	number of nodes associated with each reach	none
n_chan_max	maximum number of channels for each reach	none
n_chan_mod	mode of the number of channels for each reach	none
obstr_type	type of obstruction for each reach based on GROD and HydroFALLS databases. Obstr_type values: 0 - No Dam, 1 - Dam, 2 - Lock, 3 - Low Permeable Dam, 4 - Waterfall.	none

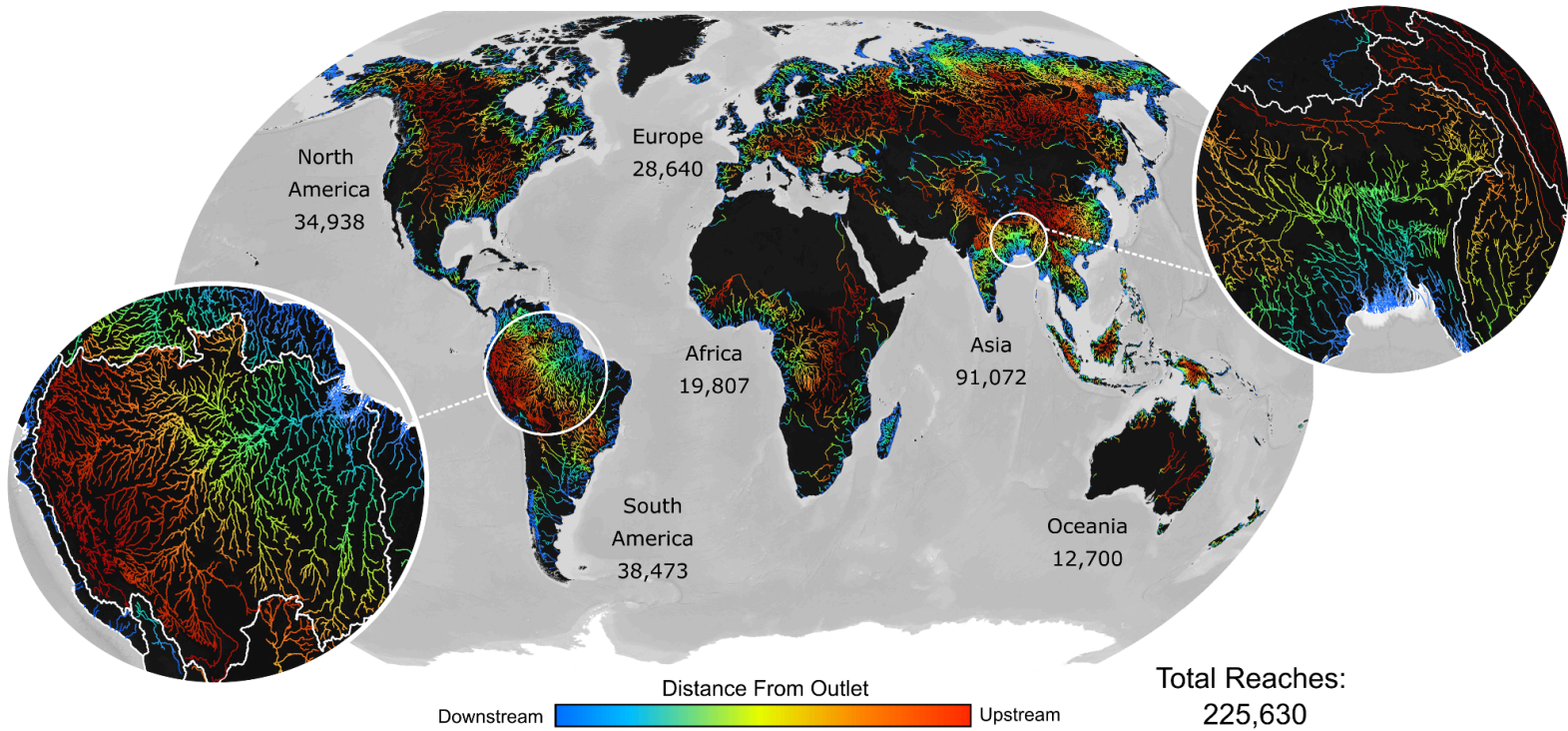
If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

grod_id	the unique GROD ID for each reach with obstr_type values 1-3.	none
hfalls_id	the unique HydroFALLS ID for each reach with obstr_type value 4.	none
slope	reach average slope calculated along the high-resolution centerline points	m/km
dist_out	distance from the river outlet for each reach	meters
n_rch_up	number of upstream reaches for each reach	none
n_rch_down	number of downstream reaches for each reach	none
rch_id_up	reach ids of the upstream reaches	none
rch_id_dn	reach ids of the downstream reaches	none
lakeflag	GRWL water body identifier for each reach: 0 – river, 1 – lake/reservoir, 2 – canal, 3 – tidally influenced river.	none
max_width	maximum width value across the channel for each reach that includes island and bar areas.	meters
type	Reach type identifier: 1 – river, 3 – lake on river, 4 – dam or waterfall, 5 – unreliable topology, 6 – ghost reach.	none
facc	Maximum flow accumulation value for each reach.	kilometers <sup>2</sup>
swot_obs	The maximum number of SWOT passes to intersect each reach during the 21 day orbit cycle.	none
swot_orbits	A list of the SWOT orbit tracks that intersect each reach during the 21 day orbit cycle.	none
river_name	all river names associated with a reach. If there are multiple names for a reach they are listed in alphabetical order and separated by a semicolon.	none
trib_flag	binary flag indicating if a large tributary not represented in SWORD is entering a reach. 0 - no tributary, 1 - tributary.	none
path_freq	the number of times a reach is traveled to get to any given headwater from the primary outlet.	none
path_order	unique values representing continuous paths from the river outlet to the headwaters. Values are unique within level two Pfafstetter basins. The lowest value is always the longest path from outlet to farthest headwater point in a connected river network. Higher path values branch off from the longest path value to other headwater points.	none

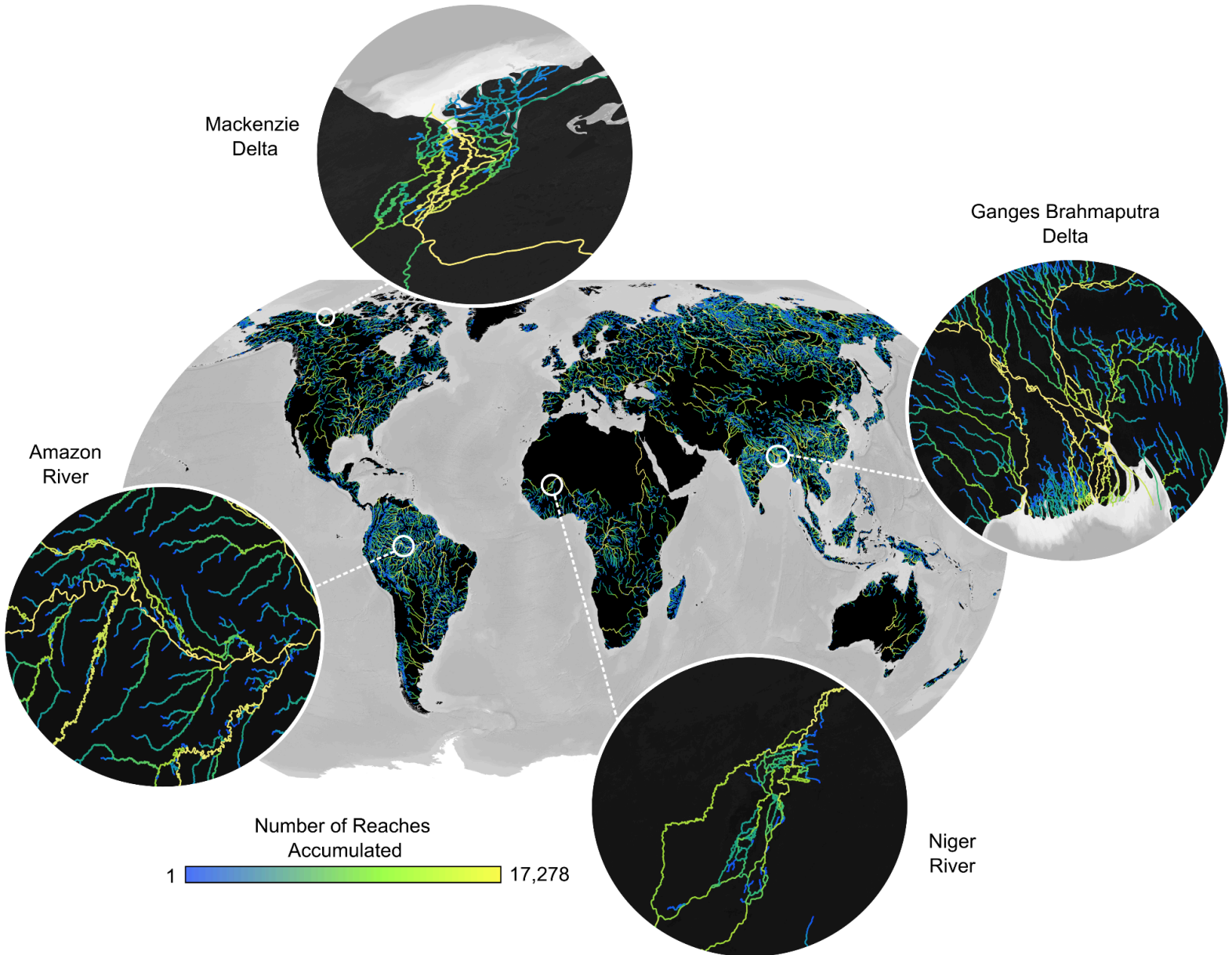
If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

path_segs	unique values indicating continuous river segments between river junctions. Values are unique within level two Pfafstetter basins.	none
main_side	value indicating whether a node is on the main network (0), side network (1), or is a secondary outlet on the main network (2).	none
strm_order	stream order based on the log scale of the path frequency. stream order is not included for side channels which are given a no data value of -9999.	none
end_rch	value indicating whether a reach is a headwater (1), outlet (2), or junction (3) reach. A value of 0 means it is a normal main stem river reach.	none
network	unique value for each connected river network. Values are unique within level two Pfafstetter basins.	none

## 5. Summary Figures and Statistics



*Figure 1: SWORD reach numbers per continent (not including ghost reaches). Colors display the updated distance from outlet based on shortest paths between outlets and headwaters.*



*Figure 2: Modified lumped routing results of accumulated reaches based on updated topology.*

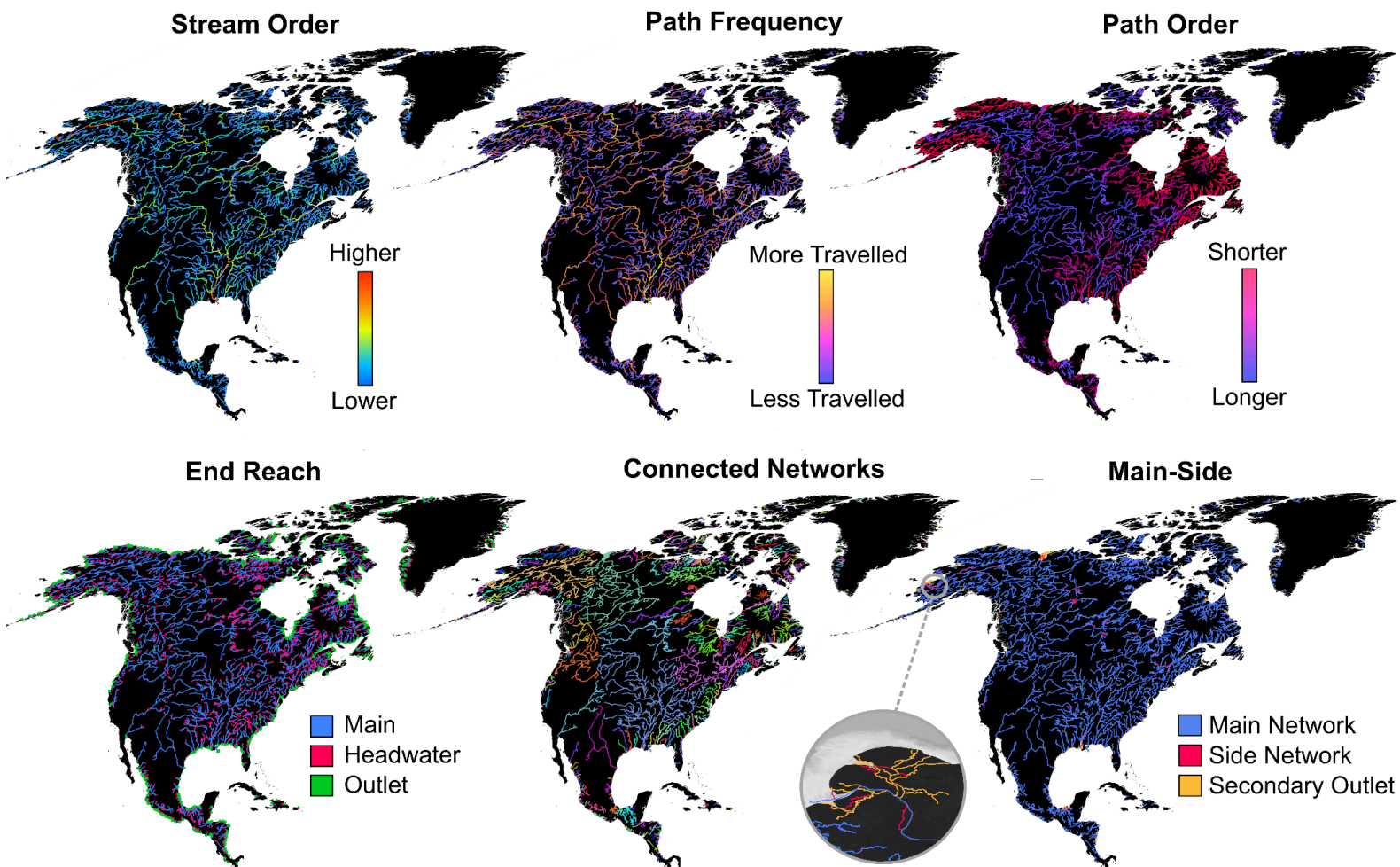
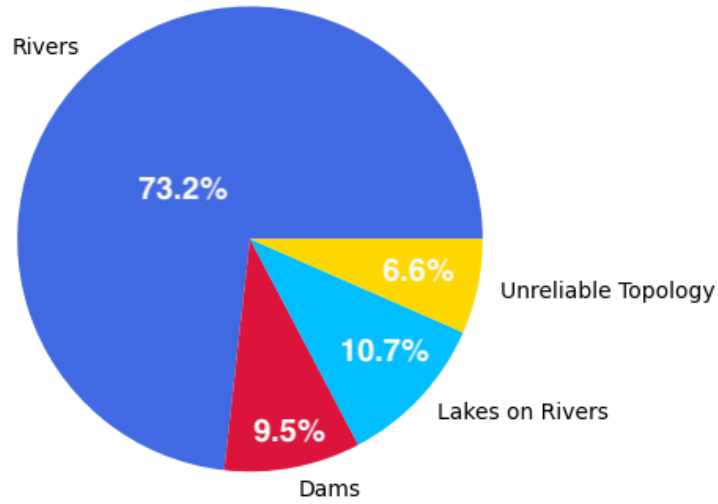


Figure 3: New SWORD variables starting in version 17. Variable descriptions in tables 3-5.



**Figure 4:** Pie chart of global reach type identifiers (1 – river, 3 – lake on river, 4 – dam (or waterfall), 5 – unreliable topology). Percentages do not include ghost reaches.

**Table 6:** Reach lengths per continent (excluding ghost reaches). Numbers in parentheses are calculated for reaches with river type identifiers only.

Reach Length (L)	NA	SA	AS	EU	AF	OC	Global
L < 5 km	23.9% (14.2%)	13.1% (7.4%)	27.4% (14.7%)	37.1% (19.5%)	12.5% (7.1%)	13.6% (7.4%)	23.6% (12.6%)
5 km ≤ L < 10 km	17.8% (13.5%)	16.5% (14.8%)	15.5% (15.9%)	15.7% (17.0%)	15.6% (13.9%)	20.9% (16.4%)	16.3% (15.3%)
10 km ≤ L ≤ 20 km	58.2% (72.3%)	70.4% (77.8%)	57.8% (69.4%)	47.1% (63.4%)	71.9% (79.0%)	65.5% (76.2%)	60.0% (72.1%)
L > 20 km	0.02% (0.03%)	0.02% (0.02%)	0.02% (0.03%)	0.01% (0.02%)	0.01% (0.01%)	0% (0%)	0.02% (0.02%)
Mean	9.7 km (11.1 km)	11.0 km (11.9 km)	9.3 km (11.0 km)	8.1 km (10.4 km)	11.2 km (12.1 km)	10.7 km (11.8 km)	9.8 km (11.3 km)
Median	10.3 km (10.9 km)	10.5 km (10.7 km)	10.3 km (10.8 km)	9.5 km (10.6 km)	10.6 km (10.8 km)	10.4 km (10.7 km)	10.3 km (10.8 km)

If you use the SWORD Database in your work, please cite: Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. <https://doi.org/10.1029/2021WR030054>

## 6. References

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, 361(6402), 585-588.

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products". *Water Resources Research*.

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. In *Remote Sensing and Water Resources* (pp. 117-147). Springer, Cham.

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](http://www.hydrosheds.org).

Tessler, Z. D., Vörösmarty, C. J., Grossberg, M., Gladkova, I., Aizenman, H., Syvitski, J. P. M., & Foufoula-Georgiou, E. (2015). Profiling risk and sustainability in coastal deltas of the world. *Science*, 349(6248), 638-643.

Whittemore, A., Ross, M. R., Dolan, W., Langhorst, T., Yang, X., Pawar, S., Jorissen, M., Lawton, E., Januchowski-Hartley, S., & Pavelsky, T. (2020). A Participatory Science Approach to Expanding Instream Infrastructure Inventories. *Earth's Future*, 8(11), e2020EF001558.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G., & Pavelsky, T. (2019). MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets. *Water Resources Research*. <https://doi.org/10.1029/2019WR024873>.

Yang, X., Pavelsky, T. M., Allen, G. H. (2019). The past and future of global river ice. *Nature*.

SWOT Orbits: <https://www.aviso.altimetry.fr/en/missions/future-missions/swot/orbit.html>

HydroFALLS: <http://wp.geog.mcgill.ca/hydrolab/hydrofalls/>