

Topology Consistency Check

Testing Filename: sword_v17c_NA.gpkg

Total Reaches Checked: **38696**

1. Check whether 'n_rch_up' or 'n_rch_down' is consistent with the actual number of upstream/downstream reaches

--- Reach Type Statistics ---
Type 1 [River]: 22155
Type 3 [Lake on river]: 8680
Type 4 [Dam or waterfall]: 2130
Type 5 [Unreliable topology]: 1953
Type 6 [Ghost reach]: 3778

--- Topology Error Type Statistics ---
Type 6 [Ghost reach]: 16
Type 3 [Lake on river]: 11
Type 5 [Unreliable topology]: 3
Type 1 [River]: 2

--- Starting Upstream Topology Check ---

Found 16 inconsistencies in upstream topology!

	reach_id	n_rch_up	actual_n_up
495	71181901496	1	0
645	71185700176	1	0
918	71212001186	1	0
1033	71216000356	1	0
1519	71224200036	1	0
7498	72408300406	1	0
8441	72545000736	1	0
13216	73281000066	1	0
13218	73281000086	1	0
13448	73283000366	1	0
13453	73283000416	1	0
15709	74262300956	1	0
22699	77121000176	1	0
30947	82213000116	1	0
34888	83185001366	1	0
34890	83185001386	1	0

All ghost reaches (16/16)

--- Starting Downstream Topology Check ---

Found 16 inconsistencies in downstream topology!

	reach_id	n_rch_down	actual_n_down
481	71181901353	2	1
642	71185700143	2	1
881	71212000813	2	1
1006	71216000083	2	1
1521	71224300023	2	1
7471	72408300133	2	1
8379	72545000113	2	1
13197	73270900155	2	1
13223	73281000135	2	1
13434	73283000223	2	1
13441	73283000291	2	1
15662	74262300481	2	1
22695	77121000133	2	1
30943	82213000075	2	1
34773	83185000213	2	1
34774	83185000223	2	1

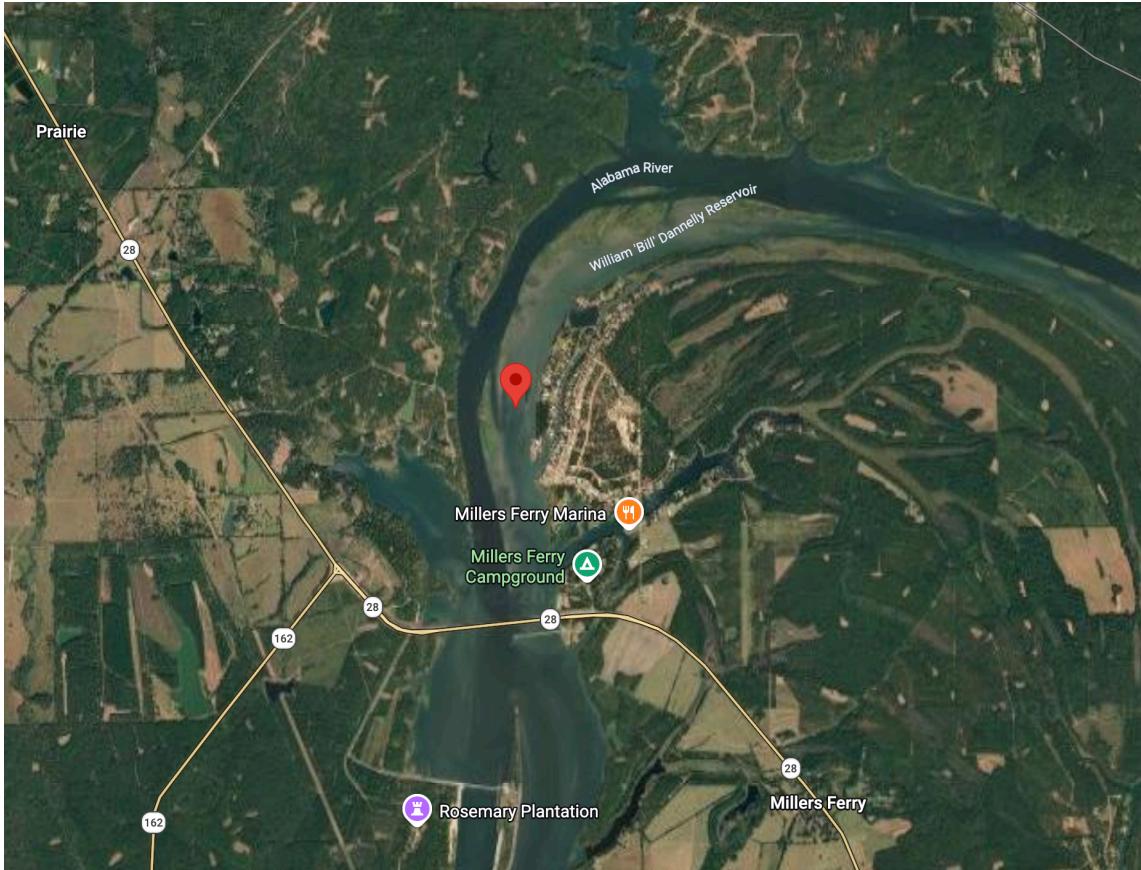
Most lake reaches (11/16)

Topology Consistency Check

Results are summarized in metadata_count_errors.csv.

Total Reaches Checked: **38696**

1. Basic Topology Number Check



Example reach (ID: 73283000223, Alabama River)

'n_rch_down=2' while the actual number of downstream reaches is 1, or another downstream ID is missing.

Type 3 [Lake on river]: 11
Type 5 [Unreliable topology]: 3
Type 1 [River]: 2

--- Found 16 / 16 reaches ---

reach_id	river_name	x	y
71181901353	NODATA	-100.910771	55.674191
71185700143	NODATA	-107.417665	55.955389
71212000813	NODATA	-99.171310	55.774000
71216000083	NODATA	-98.097504	54.441994
71224300023	NODATA	-95.016672	50.125007
72408300133	NODATA	-71.863511	48.607369
72545000113	NODATA	-76.286628	45.465798
73270900155	Bear Creek	-87.889636	30.971446
73281000135	Mobile River	-87.964745	30.857002
73283000223	Alabama River	-87.396492	32.131770
73283000291	Alabama River	-87.251699	32.107290
74262300481	Tennessee River	-85.821818	34.808517
77121000133	NODATA	-101.880340	18.450504
82213000075	Mackenzie	-134.143305	67.654577
83185000213	NODATA	-98.764202	65.999987
83185000223	NODATA	-98.879349	66.048299

Topology Consistency Check

Total Reaches Checked: **38696**

In a valid river network:

- IF Reach A lists Reach B as its downstream (rch_id_dn),
THEN Reach B MUST list Reach A as its upstream (rch_id_up).
- If Reach A points to B, but B does not acknowledge A,
THEN it is a Reciprocity Error.

2. Reciprocity Check

```
def check_reciprocity(row):  
    current_id = row['reach_id']  
    downs = [str(int(row[c])) for c in dn_cols if pd.notnull(row[c])  
and row[c] != 0]  
  
    for d_id in downs:  
        if d_id in up_map:  
            if current_id not in up_map[d_id]:  
                return True  
        else:  
            pass  
    return False
```

Number of reciprocity error: 0

The Topology is Correct:

The specific IDs are filled in correctly, and they successfully "shake hands" (A points to B, and B acknowledges A). The physical network is intact.

The Index is Wrong (Page 1):

The n_rch_up field is just a summary label, and it is likely outdated or was calculated incorrectly in a previous step.

Example: The "label" on the box says "Contains 2 items" (n_rch_up = 2), but when you actually look inside the box, there is only 1 item (rch_id_up_1 has data, rch_id_up_2 is empty).

Topology Consistency Check

Results are summarized in
distance_errors.csv.

Total Reaches Checked: **38696**

3. Distance Monotonicity Check

According to the ‘SWORD Product Description Document’, there is a need to check the distance from the river outlet for every reach. As the river drives towards the ocean (moves downstream), the distance must get smaller.

$$D_{current} \geq D_{downstream}$$

6. Notes for SWORD Users

The SWORD database is still under development and therefore contains some artifacts remaining from the automated algorithms. Below is a list of known limitations with the database:

1. *Topology Inconsistencies*: Currently, there are limited manual adjustments made to SWORD which results in the database containing some topological errors primarily due to artifacts in the flow accumulation and elevation values that occur during the merging process between the GRWL and MERIT Hydro databases. These topological errors typically show up as incorrectly ordered reaches or misidentified neighboring reaches. We apply a filter to automatically reduce these errors as much as possible, however, the remaining artifacts translate into incorrect topology definitions. We estimate that < 2% of reaches have topological inconsistencies. Additionally, the “dist_out” attribute, which defines the **distance** from the river outlet for every reach and node, is highly sensitive to topology. Therefore, large discrepancies in the “dist_out” variable can result from minor errors in the topological structure. We applied a filter to reduce these errors in the “dist_out” attribute, but users are advised to use “dist_out” as an estimate and not a precise value. In future SWORD versions released closer to the launch of SWOT in 2022, we plan to include manual adjustments where automatic methods fail, which will address the remaining topological inconsistencies. Additionally, we plan to use future SWOT elevations to improve errors in SWORD after launch.

Most unreliable reaches (11/16)

Number of wrong distance to outlet: 16					
	reach_id	type_code	type_desc	dist_out	
	1043	7121700073	3	Lake	7.056785e+05
	7822	72409000575	5	Unreliable	8.301096e+04
	19620	74300400105	5	Unreliable	1.173397e+05
	26319	81110900015	5	Unreliable	2.859791e+04
	27633	81190900085	5	Unreliable	7.141915e+04
	27810	81210500091	1	River	3.287429e+05
	27880	81210600171	1	River	6.669385e+05
	30557	82100800025	5	Unreliable	1.089118e+05
	30590	82100800355	5	Unreliable	1.827105e+05
	30614	82100900205	5	Unreliable	6.480299e+04
	30615	82100900215	5	Unreliable	6.953526e+04
	30643	82100900495	5	Unreliable	7.655340e+04
	30707	82211000255	5	Unreliable	1.482142e+05
	30940	82213000045	5	Unreliable	2.716486e+05
	32546	82282000031	1	River	2.474659e+06
	32813	82285000341	1	River	2.527936e+06

Topology Consistency Check

Total Reaches Checked: **38696**

Summarization:

1. 16 reaches show a number mismatch where 'n_rch_up=1', but the actual upstream count is 0. All identify as ghost reaches.
2. 16 reaches show a number mismatch where 'n_rch_down=1', but the actual downstream count is 0. Most identify as lake reaches.
3. Zero reciprocity errors were found across the full dataset of 38,696 North American reaches.
4. 16 reaches violate distance monotonicity (i.e., the downstream 'dist_out' value exceeds the current reach's value).
5. There is no overlap between these error categories (number errors vs. distance monotonicity errors).

For more topology consistency check/correction, refer to:

https://swotst.aviso.altimetry.fr/fileadmin/user_upload/SWOTST2024/20240620/hydro/Thurs_PM_Altenau_Elizabeth_SWORLD.pdf

Mass Conservation Check

The original logic of the Drainage Area Fix used in the Integrator is:

$$A_{downstream} = \sum A_{upstream} + \delta A$$
$$\delta A \geq 0$$

Jake raised a solution for bifurcations:

Bifurcation children (parent has 2+ downstream children): `corrected = corrected_parent * (width_child / sum_sibling_widths)` . Instead of every child getting 100% of the parent's facc (the UPA cloning error), each child gets its share proportional to channel width. If a child has no width data, children split equally as a fallback.

Therefore, the mass conservation check should be extended to:

$$\sum A_{downstream} = \sum A_{upstream} + \delta A$$
$$\delta A \geq 0$$

The mass conservation check is conducted on the dataset with a correct topology (38664/38696). The number errors group is removed, but the distance monotonicity errors group is kept.

Mass Conservation Check

The mass conservation check is conducted on the dataset with a correct topology (38664/38696). The number errors group is removed, but the distance monotonicity errors group is kept.

Error Type	What is broken?	Can water flow?	Action
Topology (32)	The Connection (ID)	No (ID is missing)	REMOVE (Data is unusable)
Distance (16)	The Label (Attribute)	Yes (Connection exists)	KEEP (Don't break the chain)

```
Starting Mass Conservation Check on 38664 reaches...
Initial violations found (Current < Sum_Upstream): 0
-----
Perfect! All reaches satisfy Mass Conservation (Out >= In).
```

Conclusion:

1. Jake's new method for fixing the drainage area is correct on all NA reaches with correct topology.
2. The topology issue needs to be reported and resolved.