Supporting Information for "Topological relationships-based flow direction modeling: mesh-independent river networks representation"

Chang Liao[1], Tian Zhou[1], Donghui Xu[1], Matt Cooper[1], Darren Engwirda[2],

Hong-Yi Li[3], L. Ruby Leung[1]

[1]Atmospheric Sciences and Global Change, Pacific Northwest National Laboratory, Richland, WA, USA

[2]T-3 Fluid Dynamics and Solid Mechanics Group, Los Alamos National Laboratory, Los Alamos, NM, USA

[3]University of Houston, Houston, TX, USA

Contents of this file

1. Text S1 to S4

2. Figures S1 to S13

3. Tables S1 to S2

Introduction

This supplementary information includes details of the following topics.

1. Spatial reference and computational geometry (Text S1)

2. River network simplification algorithm (Text S2, Figures S8-S10)

3. MPAS mesh and additional results (Text S3, Figure S11)

4. Comparison with the raster DEM-based (with stream burning) method (Text S4, Figure S13)

————

October 12, 2022, 7:35pm

5. Advantages of topological relationship-based method on the rectangle and hexagon meshes (Figures S1-S6)

6. Data model and definition (Figures S7 and S12)

7. Comparisons of existing river network representation methods at watershed and global scales (Table S1)

8. Major model configurations (Table S2)

Text S1.

Because different grid/mesh systems use different types of spatial references, computing and visualizing several metrics (e.g., distance and area) under the same framework is not straightforward. Therefore, we converted all meshes into the geographic coordinate system (GCS) spatial reference and implemented geometric algorithms based on the GCS framework. For example, the original structured square and hexagon meshes use the projected coordinate system (PCS) system. During the mesh generation process, all the coordinates were re-projected to the GCS system.

Geodesic distance

We assume the Earth surface is a perfect sphere. We define the great-circle length, the distance between two vertices on a curved surface, as the vertex distance.

Geodesic area

We define the spherical sector area, the area enclosed by more than two vertices on a curved surface, as the mesh cell area.

Geodesic angle

We define the geodesic angle, the vector angle in a three-dimensional sphere, as the river confluence branching angle.

Text S2.

Step 1 format conversion

This step converts the user provided vector-based river networks into the model framework, i.e., GCS-based flowline.

Step 2 connect disconnected flowlines

This step connects two flowlines separated due to digitization error.

October 12, 2022, 7:35pm

Step 3 define flowline vertices

A vertex classification algorithm was developed to classify all the vertices into head-water/outlet vertex, middle vertex, and confluence vertex depending on the associated number of flowlines. For example, in Figure S8a, vertices A, C, and F are defined as headwater/outlet vertices because they are associated with just one flowline. Vertices B and E are defined as middle vertices because B is connected to both flowline A->B and B->E and E is connected to both flowline B->E and E->F. Vertex D is initially defined as a headwater/outlet vertex because it is the end vertex of flowline C->D. However, after the split-by-vertex algorithm, flowline A->B splits into A->D and D->B, which redefines D as a confluence vertex because it is now connected to three flowlines.

Step 4 and 8 flowlines split and merge

Based on vertex classification, the flowlines are split or merged recursively until the network cannot be further simplified (Figure S8).

Step 5 recondition flowline

This step corrects flowline direction using the user-provided outlet location. The algorithm recursively searches from the river outlet. For example, if the starting-vertex of a flowline is the same as the outlet vertex, this flowline direction is then reversed.

Step 6 cycle removal

A cycling detection algorithm was developed to identify and simplify braided river flowlines. This algorithm scans the network and finds braided rivers. If a vertex has multiple downstream flowlines, it only keeps the one with the highest stream order as the dominant flowline (Figure S9).

Step 7 small river removal

October 12, 2022, 7:35pm

Small headwater flowlines can be optionally removed using a user-provided length threshold. This step may be run multiple times because the previous iteration may convert non-headwaters into headwaters.

Step 9 flowline indexing

After flowline simplification, the individual flowline is considered a stream segment (Flowline A->D, C->D, and D->F in Figure S8c). Both stream segment index and stream order (Strahler) are re-calculated(USGS, 2013; Liao et al., 2022). This information is used by the topological relationships reconstruction algorithm.

Text S3.

The Model for Prediction Across Scales (MPAS) mesh is an unstructured mesh widely used by the Earth system model community to provide a unified modeling framework for atmosphere, ocean, and other earth-system components(Ringler et al., 2013). It offers a gradual refinement of horizontal resolution to resolve several numerical difficulties in nested-domain modeling applications. It is based on Spherical Centroidal Voronoi Tessellations (SCVTs) and is composed of an irregular mesh of polygons (mostly hexagons). MPAS mesh can be generated by the JIGSAW mesh generator(Engwirda, 2017).

Text S4.

Because our method is an extension of our earlier studies(Liao et al., 2020, 2022), the results of this study are comparable to both large-scale flow direction datasets(Wu et al., 2012) and regional/watershed scale raster DEM-based (with stream burning) results depending on the model configurations. The full flow direction comparison will be provided in our part 2 study. Below we provide the comparison between our method and the raster DEM-based (with stream burning) method.

:

Because the raster DEM-based method is very sensitive to the DEM spatial resolution, we used the configuration with the highest resolution (5km) for comparison. Specifically, we compared the PCS rectangle mesh (square)-based results from Case 7 with the existing raster DEM-based method(Winchell et al., 2007). We first resampled the USGS National Elevation Dataset (NED) high resolution $30\,\mathrm{m}$ DEM to 5km, then ran the method using the resampled DEM and user-provided river networks.

Because the raster DEM-based method relies on flow accumulation, the final modeled flowlines are different from the user-provided flowlines. Our comparison only focuses on the flowlines produced by both methods.

At $5\,\mathrm{km}$ spatial resolution, the spatial patterns from both methods are very close (Figure S13). However, there are two major differences. First, some mesh cells that contain user-provided flowlines are excluded in the raster DEM-based method because their flow accumulations are smaller than the accumulation threshold. Therefore the DEM-based method cannot capture the river meander in many locations and produces a much shorter total length of modeled flowlines. In contrast, the topology-based method mostly preserves these mesh cells and generates a much longer total length (Figure S13). Second, the DEM-based method often produces shifted river confluence locations. In contrast, the topology-based method can capture river confluences in most cases. However, the topology-based method can also produce incorrect confluences due to the cycle removal algorithm.

Figure S1.

Figure S2.

Figure S3.

Figure S4.

Figure S5.

Figure S6.

Figure S7.

Figure S8.

Figure S9.

Figure S10.

Figure S11.

Figure S12.

Figure S13.

Table S1

Table S2

Table S1. Comparisons of existing river network representation methods at watershed and global scales.

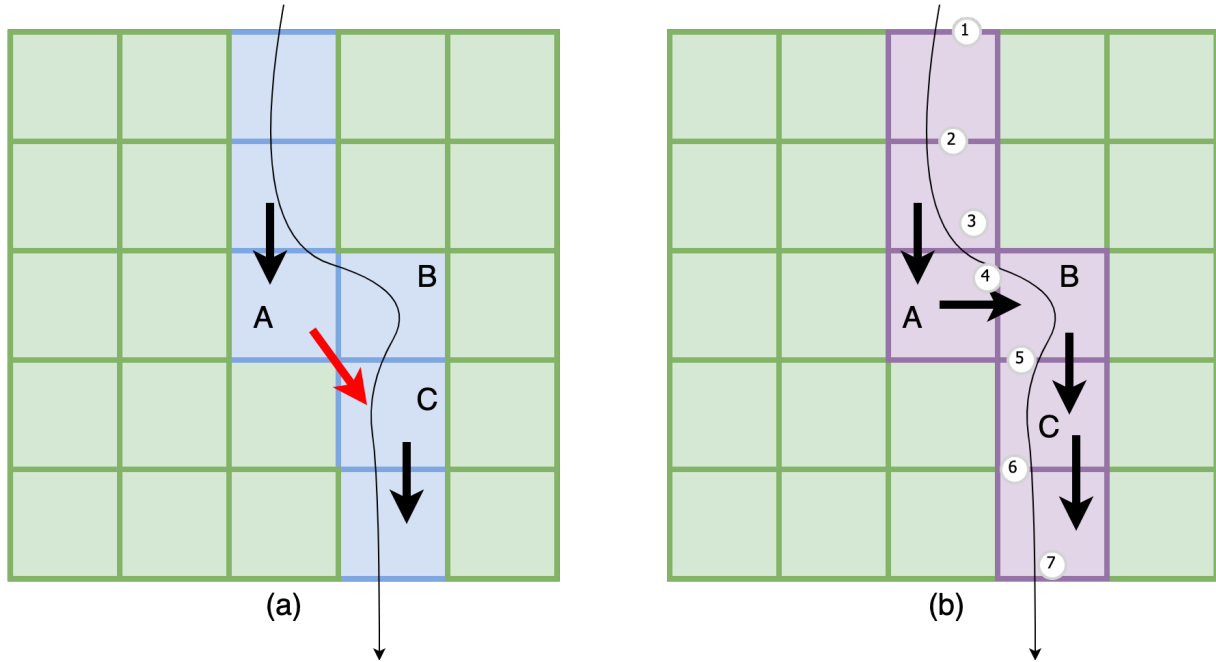|	Watershed scale		Global scale
	Vector	**Raster DEM (stream burning)**	**e.g., COTAT/NTM**
Inputs	Vector flowline	Raster DEM + vector flowline	Vector flowline + Cartesian mesh
Recommended resolution	High + moderate	High	Coarse
Flat area	Not sensitive	Less sensitive	Not sensitive
Support unstructured mesh	Not applicable	No	No
Small river removal	Optional	Optional	Automate
Braided river removal	Automate	Require pre-processing	Automate
Coverage	Only river	Entire domain	Entire domain
Impact on slope	Not applicable	High	Low
Impact on mesh generation	Not applicable	Not applicable	Not applicable

Figure S1. Illustration of flow direction on a rectangle mesh (a) without and (b) with the topological relationship at river meanders. Green mesh cells are land cells. Light blue and purple mesh cells (A, B, and C) are river cells. The black lines are the stream segments. Red arrow (A->C) and black arrows (A->B and B->C) are incorrect and correct flow directions, respectively. Circles with indices (1, 2, ..., 7) are vertices where a segment intersects with cell edges.
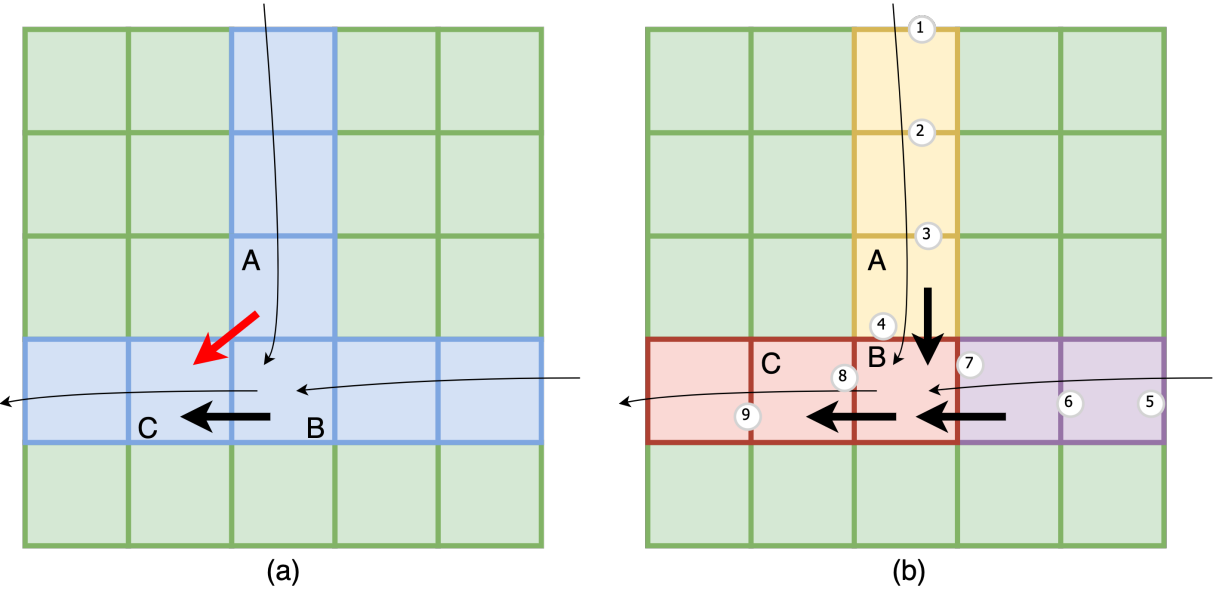
Figure S2. Illustration of flow direction on a rectangle mesh (a) without and (b) with the topological relationship at river confluence. Green cells are land. Light blue, yellow, red, and purple cells (A, B, and C) are river cells. Black lines are the stream segments. Red arrow (A->C) and black arrows (A->B and B->C) are incorrect and correct flow directions, respectively. Circles with indices (1, 2, ..., 9) are vertices where segments intersect with cell edges or each other.
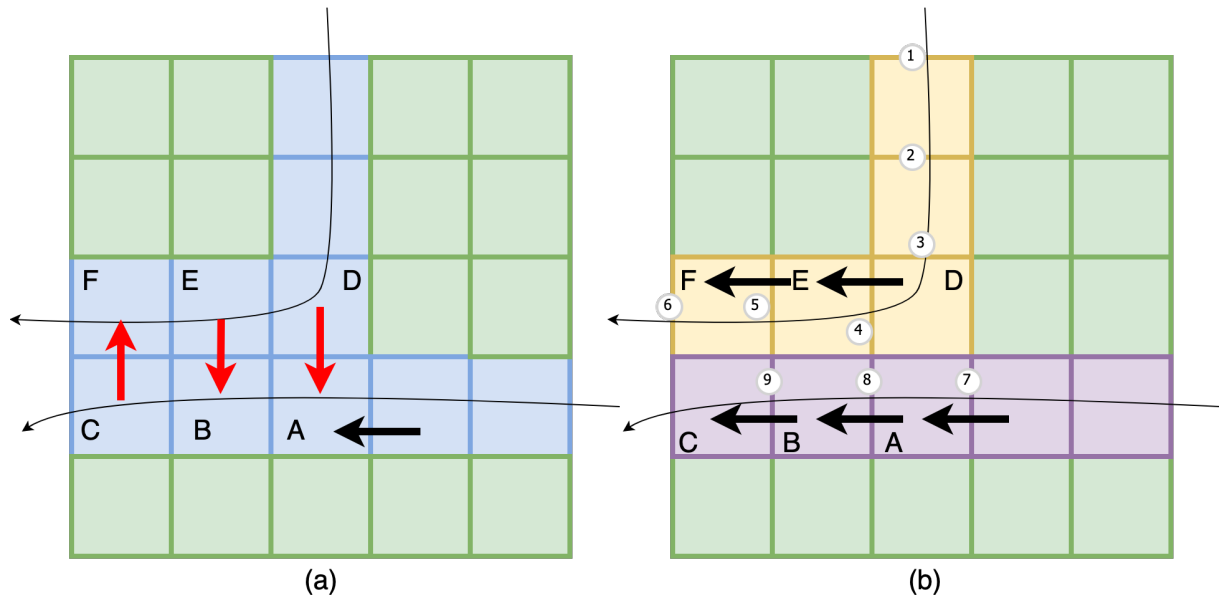
Figure S3. Illustration of flow direction on a rectangle mesh (a) without and (b) with the topological relationship when two rivers run in parallel. Green mesh cells are land cells. Light blue, yellow, and purple mesh cells (A, B,..., and F) are river cells. Black lines are the stream channels. Red arrows (D->A, B->E, and C->F) and black arrows (A->B, B->C, D->E, and E->F) are incorrect and correct flow directions, respectively. Circles with indices (1, 2, ..., 9) are vertices where river channels intersect with mesh cell edges.
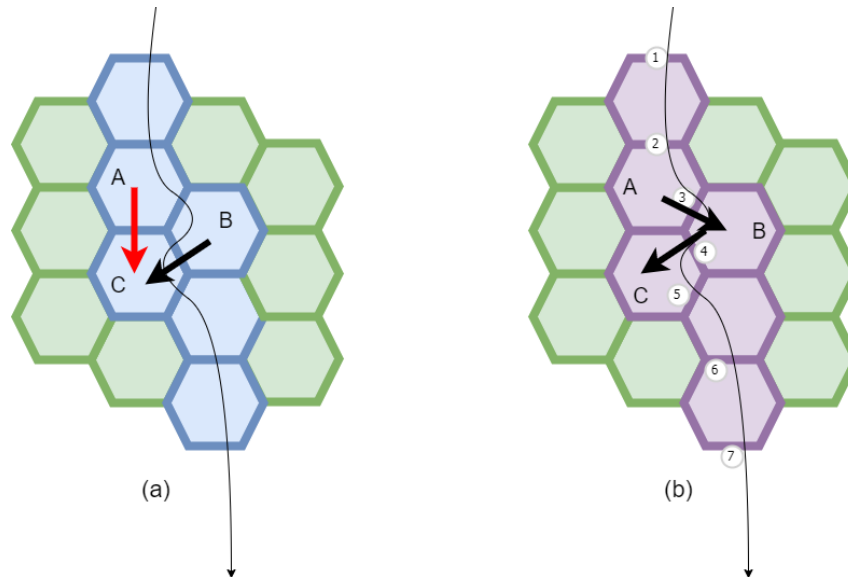
Figure S4. Illustration of flow direction on a hexagon mesh (a) without and (b) with the topological relationship when river meanders. Green mesh cells are land cells. Light blue and purple mesh cells (A, B, and C) are river cells. Black lines are the stream segments. Red and black arrows are incorrect (A->C) and correct (A->B and B->C) flow directions, respectively. Circles with indices (1, 2, ..., 7) are vertices where segment intersects with cell edges.
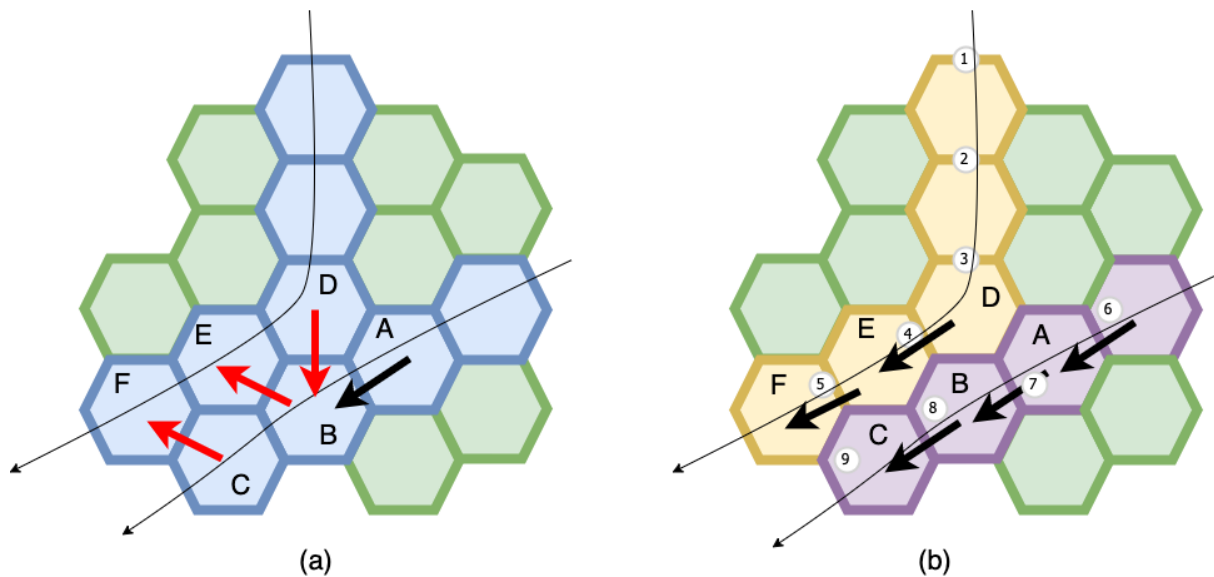
Figure S5. Illustration of flow direction on a hexagon mesh (a) without and (b) with the topological relationship at a river confluence. Green cells are land. Light blue, yellow, red, and purple cells (A, B, and C) are river cells. Black lines are the stream segments. Red and black arrows are incorrect (A->B) and correct (A->C and C->B) flow directions, respectively. Circles with indices (1, 2, ..., 9) are vertices where segments intersect with cell edges or each other.

Figure S6. Illustration of flow direction on a hexagon mesh (a) without and (b) with the topological relationship when two rivers run in parallel. Green mesh cells are land cells. Light blue, yellow, and purple mesh cells (A, B,..., and F) are river cells. Black lines are the stream channels. Red and black arrows are incorrect (D->B, B->E, and C->F) and correct (A->B, B->C, D->E, and E->F) flow directions, respectively. Circles with indices (1, 2, ..., 9) are vertices where river channels intersect with mesh cell edges.
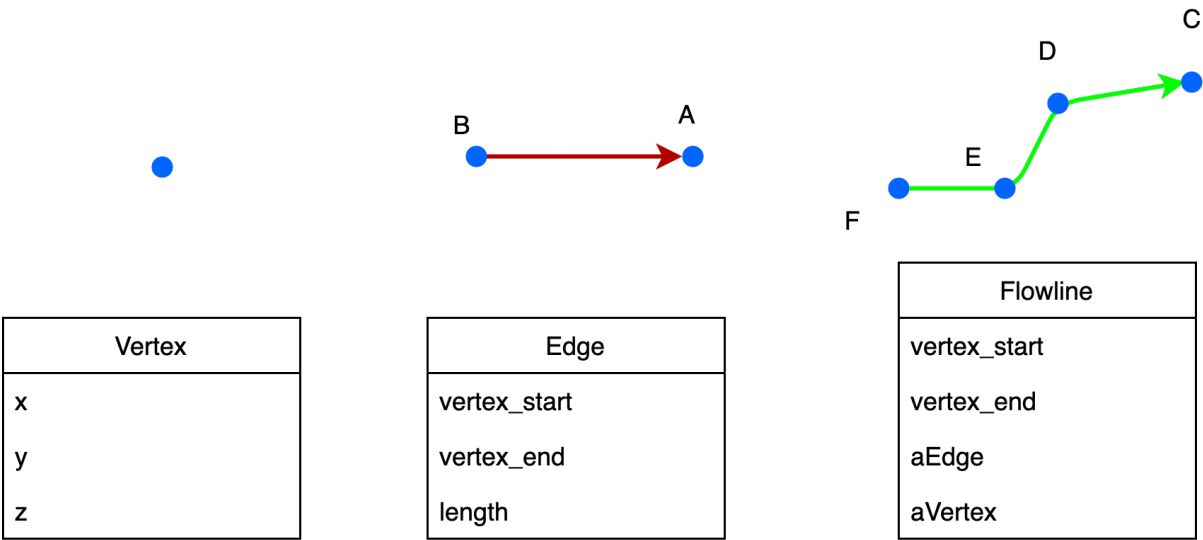
Figure S7. Illustration of the three basic elements (with attributes) used to represent a river channel. A vertex is a point with location attributes (x, y, and z). A directed edge is a link connecting two vertices (start and end). It also has the length attribute. A flowline is a list of connected edges (an edge is also a single-edge flowline). It has both start and end vertices. Its length is the sum of all member edges. For notation, a single letter/number will be used for a vertex (e.g., vertex A), two letters/numbers will be used for a directed edge (e.g., B->A), and multiple letters/numbers will be used for a directed flowline (e.g., F->E->D->C).
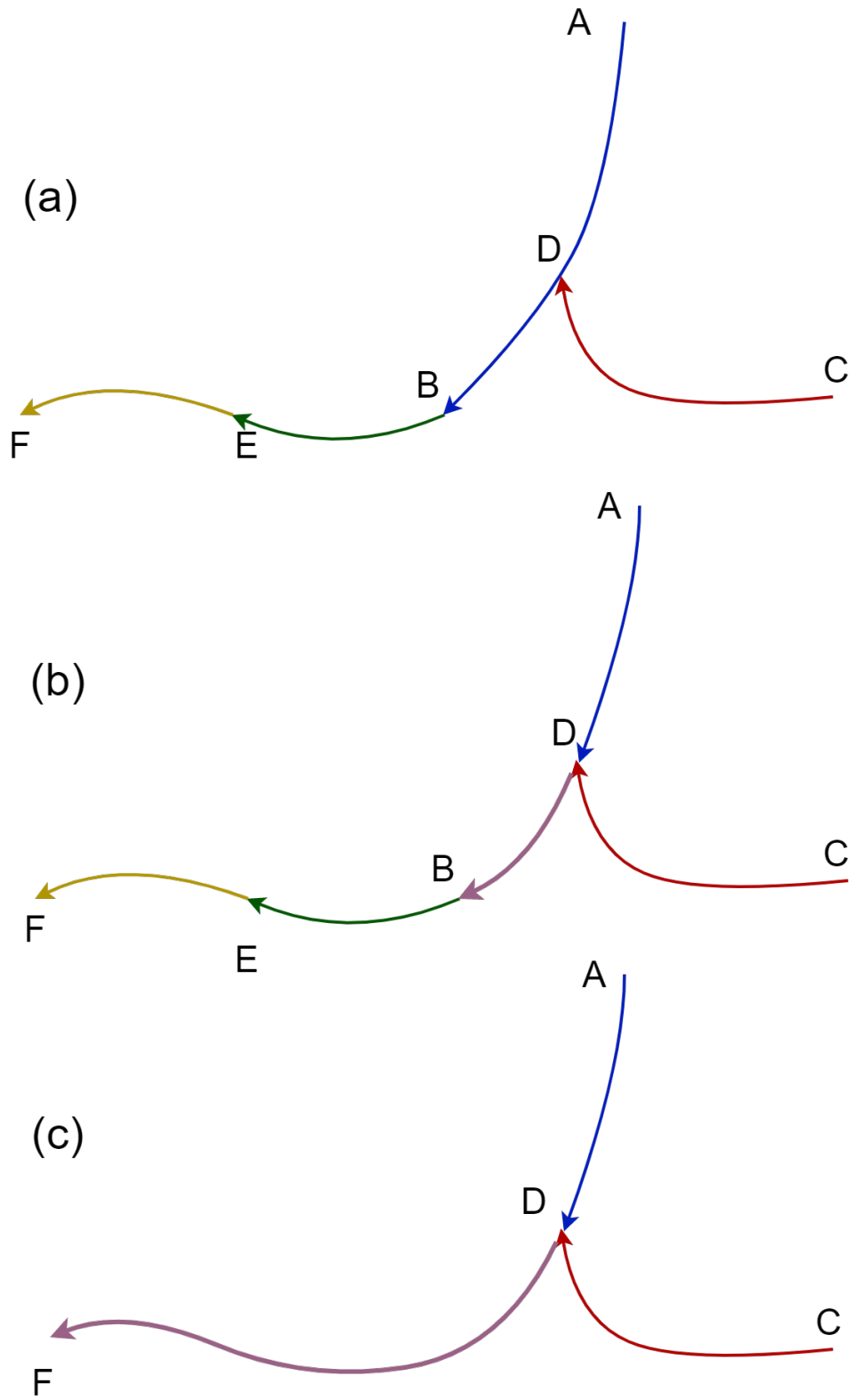
Figure S8. Illustration of flowline split and merge algorithms. The original river networks have 4 flowlines (colored lines A->B, B->E, E->F, and C->D). First, the split-by-vertex algorithm splits A->B using all vertices to obtain 5 flowlines (A->D, C->D, D->B, B->E, and E->F). Second, vertex D is defined as a confluence vertex. Third, the algorithm merges flowlines D->B, B->E, and E->F because vertex B and E are middle vertices. In the end, only 3 flowlines, A->D, C->D, and D->F, are defined.
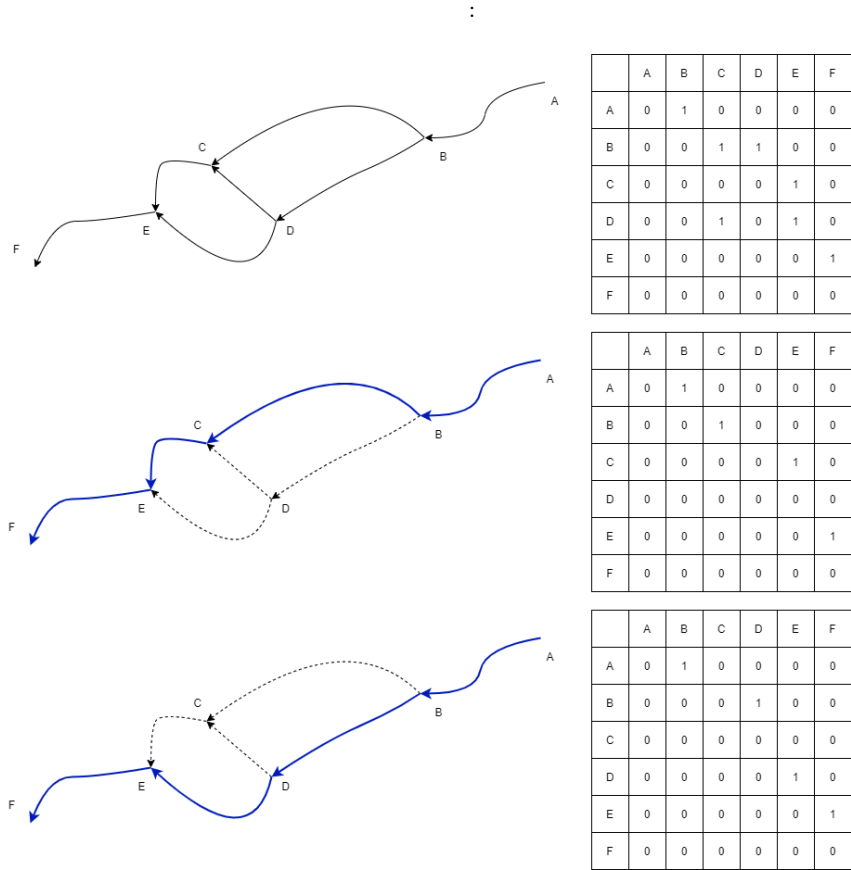
	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	1	1	0	0
C	0	0	0	0	1	0
D	0	0	1	0	1	0
E	0	0	0	0	0	1
F	0	0	0	0	0	0

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	1	0	0	0
C	0	0	0	0	1	0
D	0	0	0	0	0	0
E	0	0	0	0	0	1
F	0	0	0	0	0	0

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	0	1	0	0
C	0	0	0	0	0	0
D	0	0	0	0	1	0
E	0	0	0	0	0	1
F	0	0	0	0	0	0

Figure S9. Illustration of braided river simplification using the cycling detection algorithm. On the left, black lines are original river flowlines, blue flowlines are 2 out of 3 candidate simplified flowlines (A->B->C->E->F or A->B->D->E->F). Dashed flowlines are removed flowlines. On the right, each matrix is the corresponding network matrix with 0 and 1 representing no flow and flow, respectively. In a directed network, a braided river occurs when a vertex is the starting vertex of more than one flowline. Correspondingly, its network matrix has multiple 1s on the same row.
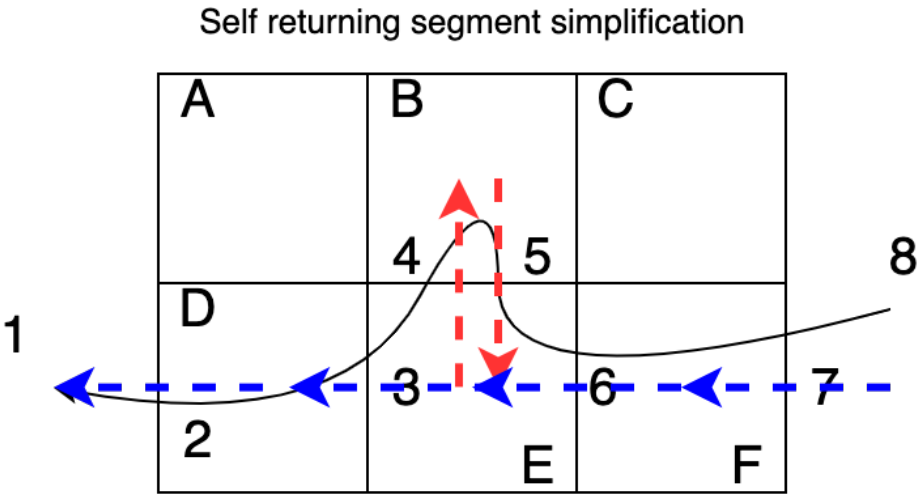
Self returning segment simplification

Figure S10. Illustration of the self-returning stream segment simplification in a 2-row by 3-column Cartesian mesh. Letters on the top left corner from A to F represent the mesh cells. Numbers from 1 to 8 represent the start, end, and intersected vertices. Black line is the actual flowline. Because it enters and exits the edge between grid B and E twice, the grid topological relationships F->E->B->E->D (red dashed arrows in the loop) demonstrate cycling. The algorithm detects and removes cycling and the final topological relationships are F->E->D (blue dashed arrows).
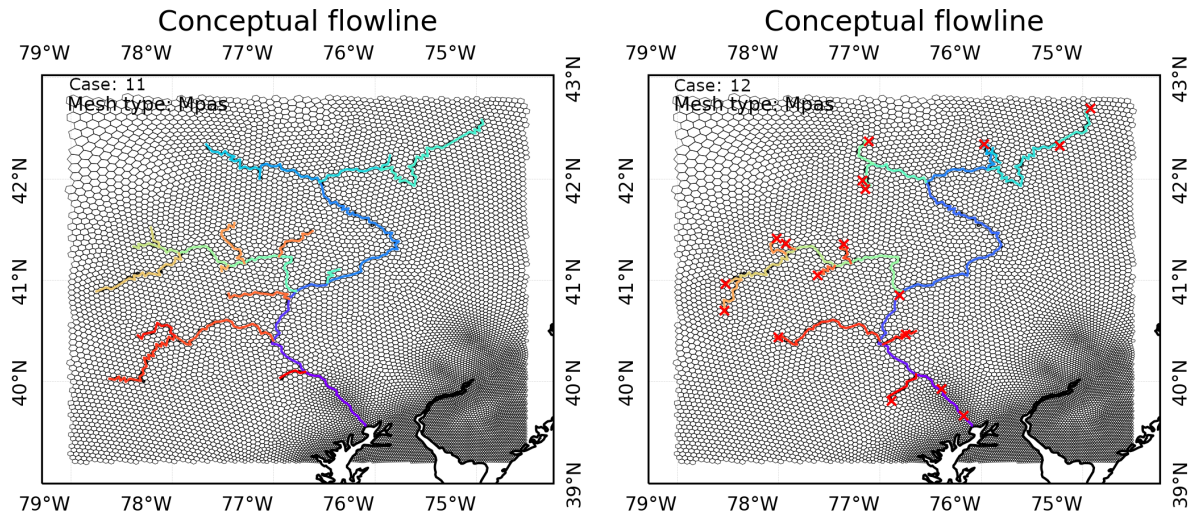
Figure S11. Modeled flowline-guided MPAS mesh (clipped to the study domain) and conceptual flowlines in the SRB from Cases 11 and 12 (Table 1). Black line features are the simplified flowlines. Colored line features are the modeled conceptual flowlines. (a) is modeled conceptual flowlines with a stream order 6 threshold. (b) is modeled conceptual flowlines with a stream order 7 with dam burnt in. Red crosses are dam locations. All the flowlines downstream of each dam are included.
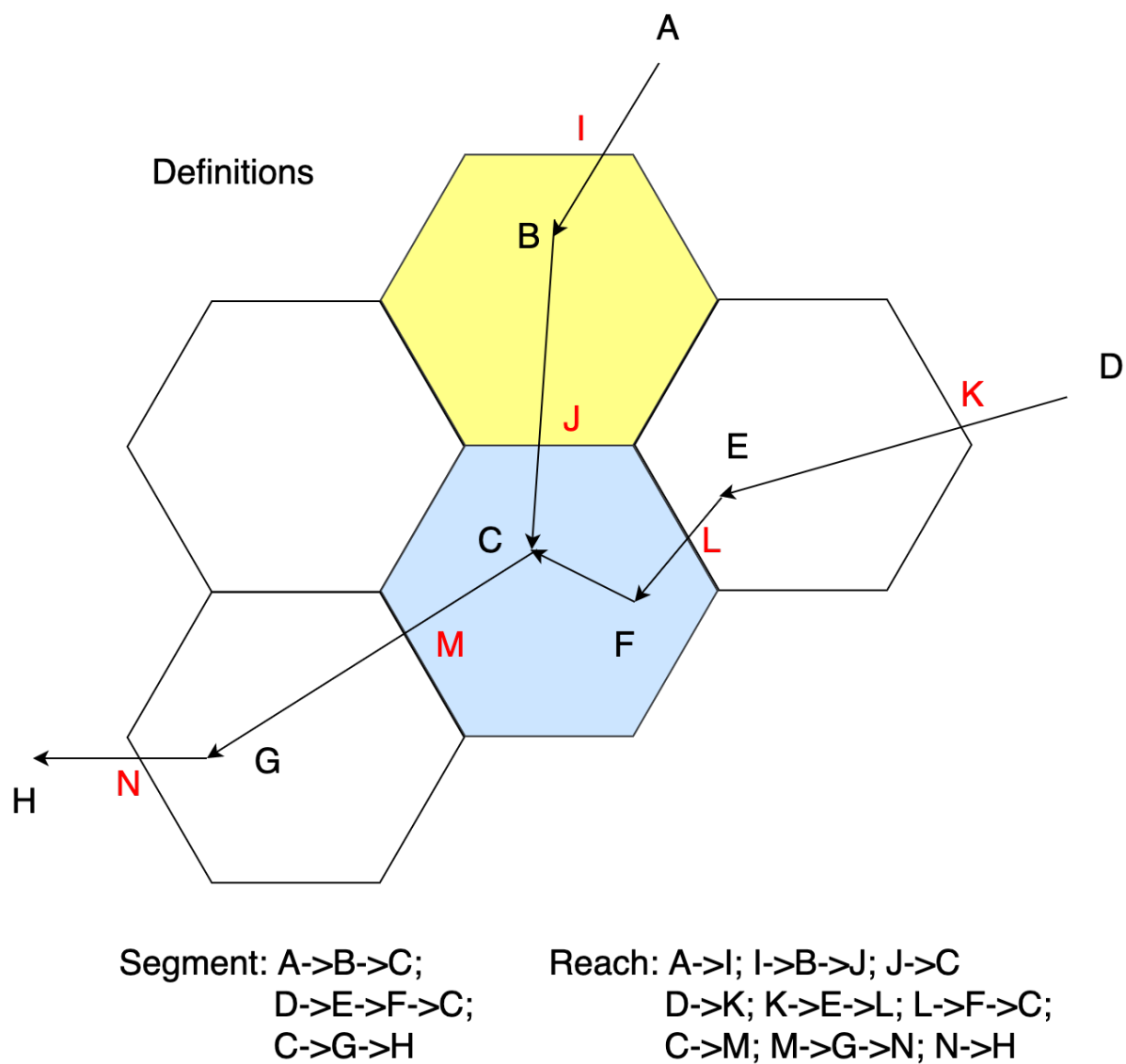
:

Segment: A->B->C;
         D->E->F->C;
         C->G->H

Reach: A->I; I->B->J; J->C
       D->K; K->E->L; L->F->C;
       C->M; M->G->N; N->H

Figure S12. Definitions of stream segment and reach with an intersection in a hexagon mesh. Initially, there are 3 stream segments (A->B->C, D->E->F->C, and C->G->H). After the intersection, each segment is broken into several reaches. For example, segment A->B->C is broken into reach A->I, I->B->J, and J->C. The yellow cell has only one reach and the light blue cell has 3 reaches.
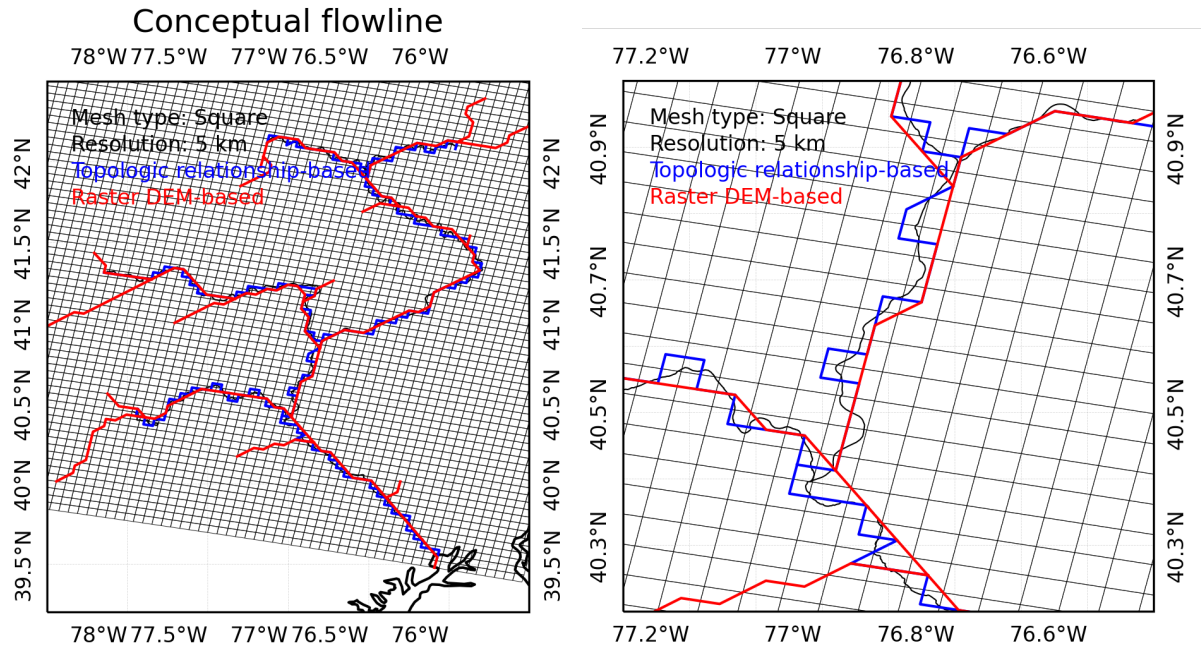
Figure S13. Comparisons of modeled flowlines from our model and the raster DEM-based method at a 5 km spatial resolution. The black, blue, and red lines are the simplified, square mesh topological relationship-based, and raster DEM-based flowlines, respectively. (b) is a zoomed-in view of (a) near the confluence.

:

Table S2. Major model configurations.

Parameter	Usage	Default
iFlag_removesmallriver	Option to remove small river	1
iFlag_shortcut	Option to enable shortcut simplification	1
dLongitude_outlet_degree	Longitude of outlet in degree	0.0
dLatitude_outlet_degree	Latitude of outlet in degree	0.0
dResolution_meter	Mesh resolution	10000 m
dThreshold_smallriver	The threshold to remove small river	Same as resolution
dThreshold_shortcut	The threshold to ignore cell index	5000 m
sMesh_type	Mesh type	hexagon

References

Engwirda, D. (2017). JIGSAW-GEO (1.0): locally orthogonal staggered unstructured grid generation for general circulation modelling on the sphere. *Geoscientific Model Development*, *10*(6), 2117–2140.

Liao, C., Tesfa, T., Duan, Z., & Leung, L. R. (2020). Watershed delineation on a hexagonal mesh grid. *Environmental Modelling & Software*, *128*, 104702. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1364815219308278` doi: 10.1016/j.envsoft.2020.104702

Liao, C., Zhou, T., Xu, D., Barnes, R., Bisht, G., Li, H.-Y., . . . Engwirda, D. (2022). Advances in hexagon mesh-based flow direction modeling. *Advances in Water Resources*, 104099. doi: 10.1016/j.advwatres.2021.104099

Ringler, T., Petersen, M., Higdon, R. L., Jacobsen, D., Jones, P. W., & Maltrud, M. (2013). A multi-resolution approach to global ocean modeling. *Ocean Modelling*, *69*, 211–232. doi: 10.1016/j.ocemod.2013.04.010

USGS. (2013). *National Hydrography Geodatabase: The National Map viewer available on the World Wide Web, accessed [June 02, 2015], at url [http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd].* Retrieved from `http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd`

Winchell, M., Srinivasan, R., Di Luzio, M., & Arnold, J. (2007). ArcSWAT interface for SWAT 2005. *User's Guide, Blackland Research Center, Texas Agricultural Experiment Station, Temple*.

Wu, H., Kimball, J. S., Li, H., Huang, M., Leung, L. R., & Adler, R. F. (2012). A new global river network database for macroscale hydrologic modeling. *Water resources research*, *48*(9).