

Advances in hexagon mesh-based flow direction modeling

Chang Liao^{a,*}, Tian Zhou^a, Donghui Xu^a, Richard Barnes^{b,c,d}, Gautam Bisht^a, Hong-Yi Li^e, Zeli Tan^a, Teklu Tesfa^f, Zhuoran Duan^f, Darren Engwirda^g, L. Ruby Leung^a

^a*Atmospheric Sciences and Global Change, Pacific Northwest National Laboratory, Richland, WA, USA*

^b*Energy & Resources Group (ERG), University of California, Berkeley, CA, USA*

^c*Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA*

^d*Berkeley Institute for Data Science (BIDS), University of California, Berkeley, CA, USA*

^e*University of Houston, Houston, TX, USA*

^f*Hydrology Group, Pacific Northwest National Laboratory, Richland, WA, USA*

^g*T-3 Fluid Dynamics and Solid Mechanics Group, Los Alamos National Laboratory, Los Alamos, NM, USA*

Abstract

Watershed delineation and flow direction representation are the foundations of streamflow routing in spatially distributed hydrologic modeling. A recent study showed that hexagon-based watershed discretization has several advantages compared to the traditional Cartesian (latitude-longitude) discretization, such as uniform connectivity and compatibility with other Earth system model components based on unstructured mesh systems (e.g., oceanic models). Despite these advantages, hexagon-based discretization has not been widely adopted by the current generation of hydrologic models. One major reason is that there is no existing model that can delineate hexagon-based watersheds while maintaining accurate representations of flow direction across various spatial resolutions. In this study, we explored approaches such as spatial resampling and hybrid breaching-filling stream burning techniques to improve watershed delineation and flow direction representation using a newly developed hexagonal mesh watershed delineation model (HexWatershed). We applied these improvements to the Columbia River basin and performed 16 simulations with different configurations. The results show that (1) spatial resampling modulates flow direction around headwaters and provides an opportunity to extract subgrid information; and (2) stream burning corrects the flow directions in mountainous areas with complex terrain features.

Keywords: Watershed delineation, unstructured mesh, stream burning, flow routing

1. Introduction

Flow direction and watershed delineation are the foundations of runoff and streamflow routing in spatially distributed hydrologic modeling[1–5]. However, one must first perform spatial discretization to calculate flow directions.

To date, most existing flow direction methods have been developed based on Cartesian grid discretization[6–15]. They are thus subject to several limitations[16]: (1) the Cartesian grid cannot represent adjacency uniformly. In a Cartesian system, each rectangular cell has two types of neighbors, direct/face and diagonal/vertex neighbors. As a result, hydrologic models need to consider the differences in the travel path of these types of neighbors (Figure

*Corresponding author

Email address: chang.liao@pnnl.gov (Chang Liao)

1). (2) Because of the diagonal travel path in the D8 scheme[17], an “island” may be created and cause numerical
 10 issues in a coupled surface-subsurface hydrology model (Figure 1)[18, 19].

Island effect and its impact on surface-subsurface hydrologic model

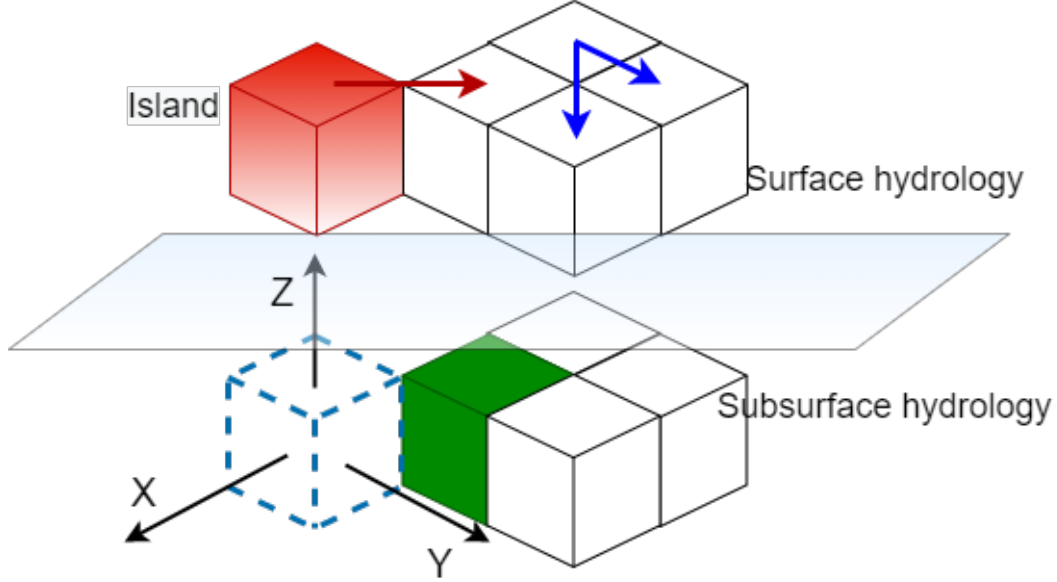


Figure 1: Illustration of the island effect and its impact on coupled surface-subsurface hydrologic models. The red cube/cell is an island in the surface hydrology model and is connected through a diagonal path (red arrow). The blue arrows illustrate the differences between direct and diagonal neighbors. The dashed cube underneath the red cube cannot be modeled in a Cartesian grid-based subsurface hydrologic model because only the top face is connected with the red cube. The green cube is often not considered a “neighbor” of the dashed cube in the subsurface hydrologic models[20, 21].

(3) the Cartesian grid cannot provide spherical coverage without significant spatial distortions, especially at high latitude regions. (4) Earth system model components, e.g., atmospheric and oceanic models, commonly use sophisticated grid structures (cubed-sphere, icosahedron, unstructured mesh) to address the convergence of meridians toward the poles. Because of that, Cartesian grid-based hydrologic models cannot be coupled to these components
 15 seamlessly and often require a spatial interpolation for flux exchange. In contrast, spatial interpolation will be significantly reduced or removed in a unified mesh framework. The use of fully unstructured model components (e.g. the Model for Prediction Across Scales (MPAS)[22, 23] and the Finite-volumeE Sea-ice/Ocean Model (FESOM)[24]) is an area of growing interest in coupled Earth system modeling.

Triangular irregular network (TIN)-based methods are also very popular due to the advantages of TIN mesh
 20 discretization. TIN can capture terrain variation with fewer samples on the Digital Elevation Model (DEM) than traditional mesh at higher resolution. TIN can also resolve the spherical coverage limitation. However, TIN based methods are still subject to other limitations: (1) In a TIN mesh, each triangle still has two types of neighbors: face neighbors and vertex neighbors (Section 6.1 Figure S1). Most TIN-based hydrologic models use face edge for flux exchange and represent the stream network using TIN edges[25–28]. (2) TIN is commonly adopted in regional
 25 hydrologic models without depression filling and stream burning. Although it can be extrapolated to a global

scale, it will not perform well without these capabilities because of the scale differences between mesh resolution and river networks[29]. (3) Although regional scale oceanic models use triangle mesh in the finite volume method (FVM) or finite element method (FEM), global scale oceanic models also use hexagonal-like C-grids to address issues including the horizontal divergence oscillations on unstructured triangular C-grids[30–32]. Therefore, the
30 different mesh systems must still be resolved.

Hexagonal mesh-based discretization can resolve the aforementioned limitations based on its geometrical properties: (1) it has only one type of neighbor, i.e, face neighbor. (2) Because of the uniform face connectivity, it directly eliminates the island effect. (3) It can cover spheres with flexibility similar to TIN. With some relaxation, a hexagonal mesh can be turned into a fully unstructured variable resolution mesh. (4) A hydrologic model based
35 on the hexagonal mesh can be seamlessly coupled with an oceanic model. As a result, it can be used to produce high-quality flow directions, especially for coupled land, river, and ocean modeling[16]. Despite these advantages, hexagonal mesh-based flow directions have not been widely adopted by current hydrologic models. One major reason is the limited access to models that can delineate watersheds on hexagonal mesh while maintaining accurate flow directions at various spatial resolutions. As the first step in this direction, we developed a hexagonal mesh-based
40 watershed delineation model (HexWatershed 1.0), which has shown great potential in improving stream network representations[16]. However, challenges still remain in applying it to the current generation of hydrologic models, such as accurately delineating watersheds and stream networks at relatively coarse spatial resolution.

The performance of watershed delineation tends to decrease for any type of mesh as spatial resolution decreases[16, 33, 34]. Fundamentally, poor representation of terrain elevation, DEM in coarse spatial resolutions, leads to poor
45 representation of the flow direction and incorrect stream networks due to the resolution mismatch. To date, there are several approaches to mitigate this issue in the Geographic Information System (GIS) and hydrology communities. These approaches can be used either individually or in conjunction with each other.

The first approach involves the use of advanced spatial resampling methods to improve DEM representation, studied extensively in the GIS literature[35–37]. The second approach uses advanced data structures to improve
50 DEM quality. For example, TIN-based DEMs have been widely used because they can follow terrain with variable spatial resolutions, providing a good balance between spatial resolution and performance[38]. The third approach is “stream burning”, a common practice in hydrology for adjusting surface elevation near stream channels so that water always converges into observed stream channels[39, 40]. Finally, some studies have attempted to define coarse spatial resolution flow direction using fine-scale hydrography datasets[41, 42].

55 In addition, flow direction determination issues can be exacerbated by processes in which “depression filling” algorithms are used to remove local depressions within a DEM which would otherwise trap water. This approach is exemplified by the Priority-Flood algorithm which is widely used in many hydrological applications[16, 43].

In practice, the foregoing approaches are often used together to best represent watershed and stream networks in Cartesian grid systems[44, 45]. However, the quality of the final outputs depends on many factors including the
60 order in which the operations are conducted[46].

To date, less attention has been paid to hexagon mesh-based watershed delineation models and their corresponding stream network representations, especially when used at coarse spatial resolutions.

In this study, we introduce HexWatershed 2.0, which incorporates the above-mentioned techniques to improve

upon HexWatershed 1.0. We then compare the results from HexWatershed 1.0, 2.0, and a Cartesian grid-based
 65 flow direction model, i.e., the Dominant River Tracing (DRT) model[41] in the Columbia River basin. Last, we
 discuss the limitations of HexWatershed 2.0 and the next steps to continue improving model performance.

2. Methods

2.1. HexWatershed 1.0

The stream burning feature is not available in HexWatershed 1.0 and the nearest-neighbor resampling method
 70 was used to assign the elevation of each hexagon from the high spatial resolution DEM (yellow dot in Figure 2).
 In detail, the model obtains elevation in the following steps: (1) obtain the center location of each mesh cell; (2)
 calculate the image coordinates (row and column index) from the DEM data based on the location; (3) read the
 elevation value using the index. The nearest-neighbor resampling method is the easiest and most straightforward
 method but may introduce large bias in areas with large elevation variations.

2.2. New features in HexWatershed 2.0

2.2.1. Zonal mean spatial resampling

In HexWatershed 2.0 we provided an additional resampling method, the zonal mean resampling method, to
 improve elevation representation. The zonal mean elevation of a mesh cell is estimated with high-resolution DEM
 (see an example in Figure 2). To guarantee there are enough high-resolution DEM values within each hexagon,
 80 the ratio of resolutions between the target hexagon mesh and DEM should be larger than 3. The definition of hexagon
 mesh resolution is provided in Section3.2.

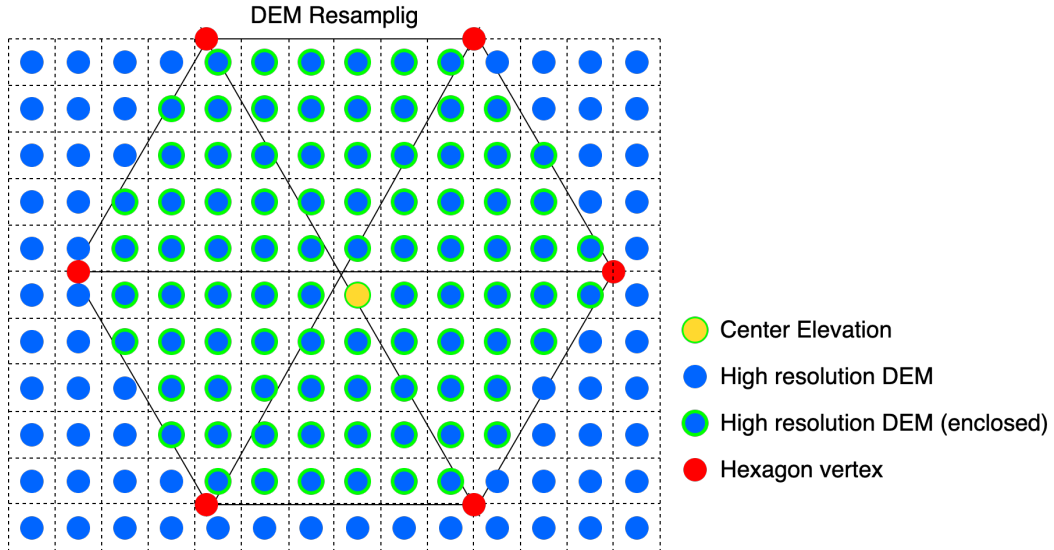


Figure 2: Illustration of DEM resampling using the nearest and zonal mean methods. Red dots represent the vertices of a hexagon cell;
 blue dots are high spatial resolution DEM raster, and dots with green outlines are those enclosed within the hexagon. In the nearest
 resampling method, the closest cell (yellow dot with green outline) to the hexagon center is used as hexagon elevation. In the zonal
 mean method, the average elevation of all enclosed grid cells is used as the hexagon elevation.

This process was implemented using the Geospatial Data Abstraction Library (GDAL) to identify all the high-resolution DEM cells inside a given coarse-resolution cell[47]. The zonal mean resampling approach also provides the capability to calculate the elevation profile and variations within each polygon mesh cell, important information
85 required by Earth system models (ESMs)[2].

2.2.2. Stream burning

We developed a stream burning algorithm based on several earlier studies[40, 48]. Stream burning, also often referred to as “DEM reconditioning”, is a technique to manipulate flow direction by lowering the elevations within and near the stream channels so that water always flows into prescribed channels.

90 A longstanding challenge in watershed delineation models is that depression filling and stream burning may interfere with each other because both algorithms modify the surface elevations[45]. To address this in HexWatershed 2.0, we developed a hybrid algorithm that combines depression filling and stream burning. Specifically, stream burning was implemented within the priority flood algorithm seamlessly in a way that elevation will be modified only once. Priority-flood is an efficient algorithm that fills DEM depressions by sequentially flooding the
95 domain from a boundary inward to adjust elevations to assure surface drainage[43]. To further simplify the process, the stream burning algorithm only modifies surface elevations in the immediate riparian zone, which is a single mesh cell buffer zone on both sides of the stream channels. The supplementary information provides more details (Section 6.3).

After obtaining all the necessary information (DEM, hexagon mesh with stream tag, etc.), the modified stream-
100 burning-priority-flood algorithm operates via the following steps:

1. Define the initial priority-flood queue as the domain boundary.
2. Find the node with the lowest mesh cell elevation in the priority-flood queue.
3. Check whether this node is a stream mesh cell or not. If it is a stream mesh cell, find its upstream mesh cells and modify stream elevation if the upstream elevation is lower or equal, then go to step 3a; if it is not a
105 stream mesh cell, process as a normal priority-flood node, then go to step 3b.
 - (a) Adjust node elevations in stream mesh cell buffer zones.
 - (b) Adjust node elevations if they are lower than the center node.
4. Mark a mesh cell as processed if its elevation is visited.
5. Push out processed nodes and push their neighbors into the priority-flood queue.
- 110 6. Go to step 2 unless the priority-flood queue is empty.

The priority-flood algorithm[16] and additional stream burning algorithm are explained and illustrated in our earlier study as well as Figures 3 and 4.

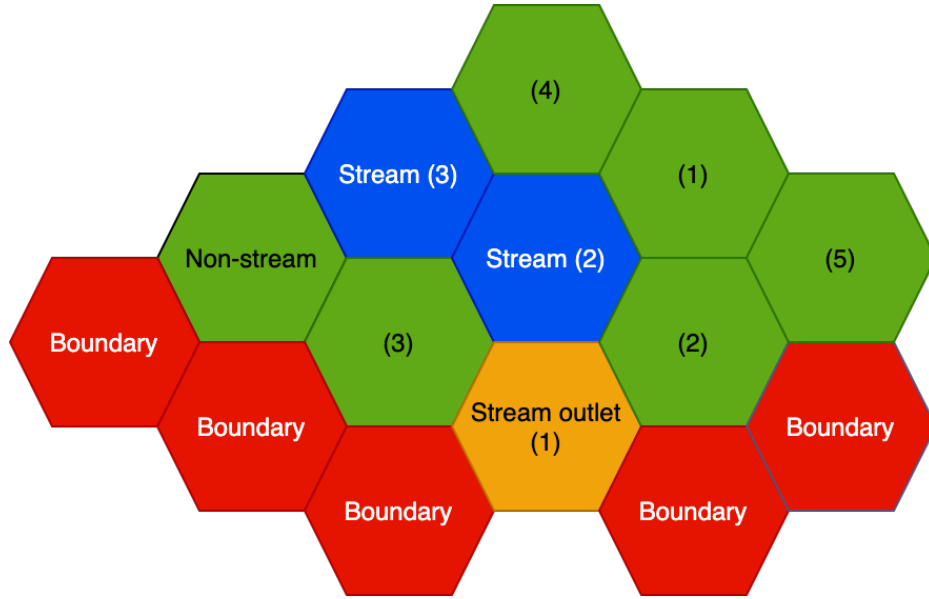


Figure 3: Illustration of the stream-burning-priority-flood algorithm on a 2D plane. Red cells represent the priority queue at the boundary; the orange cell is the user-defined watershed outlet; blue cells are stream cells; and green cells are normal non-stream cells. Indices with cells are cell IDs. When the stream outlet (orange #1) is identified, its upstream stream cells (blue #2 and #3) are processed first. Then the rest of the normal cells (green #2 and #3, etc.) are processed. Breaching or filling operates depending on the elevation differences between stream cells orange #1 and blue #2.

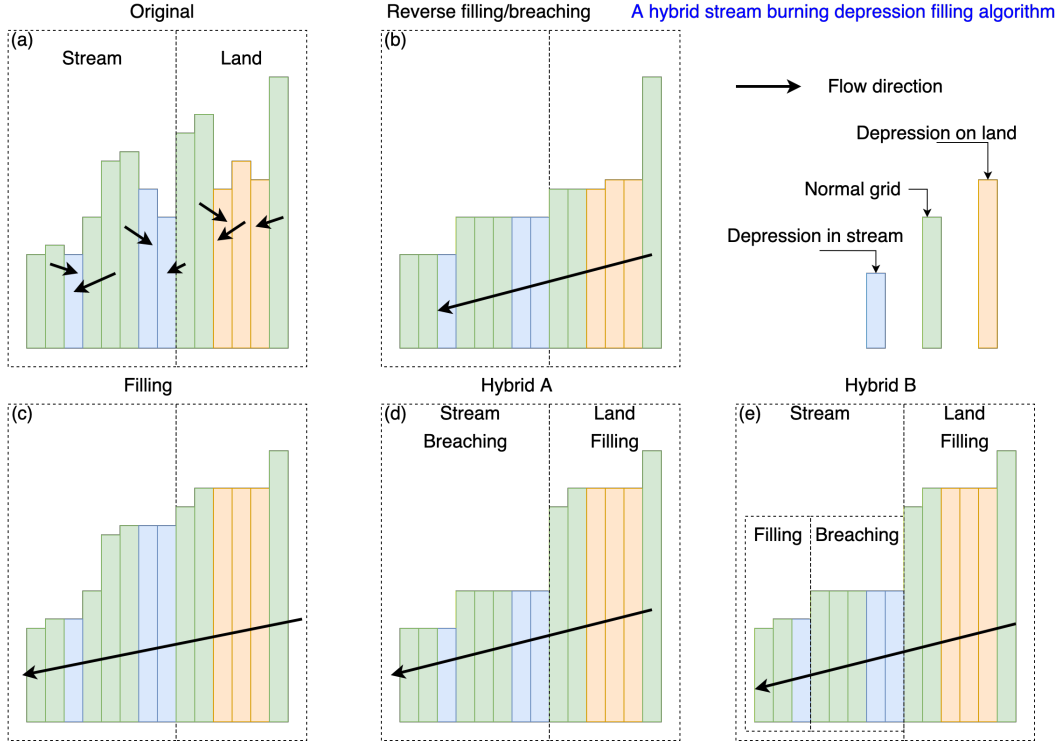


Figure 4: Illustration of the stream-burning-priority-flood algorithm on a 2D transect along the stream. Blue, green, and yellow rectangles represent elevations of stream and land mesh cells. Black arrows are flow directions. (a) is the raw elevation profile with 3 depression regions (2 in stream and 1 on land). (b) is the depression removal approach by lowering elevation from the peak elevation, similar to the breaching approach. (c) is the priority-flood approach starting from the lowest elevation. (d) is a hybrid approach with breaching in stream and filling on land. (e) is an improved hybrid approach that includes both breaching and filling in stream. Only option d is implemented in HexWatershed 2.0. Option e will be available in HexWatershed 3.0 because explicit stream topology (upstream-downstream relationship) is required (Section 5.5).

Unlike the original priority flood algorithm in HexWatershed 1.0, which only increases the elevation upstream when needed, the modified stream-burning-priority-flood algorithm does both breaching and filling when modifying stream and land elevations (Figures 4d and 5). This strategy was developed to minimize modification to the land mesh cell surface elevations[40]. A user-provided threshold parameter is employed to control the breaching and filling options. In traditional stream burning, the stream mesh cell elevation is decreased by a threshold value such as 100 m, serving as a large drop to enforce flow direction. In HexWatershed 2.0, we only modify the outlet elevation and use a threshold value like 10 m to adjust stream mesh cell elevation from outlet to headwater recursively. In order to generate a “smooth” gradient, we enforce the max drop between neighboring upstream and downstream mesh cells. For example, if the elevation differences between upstream and downstream exceed 10 m, the upstream elevation is lowered.

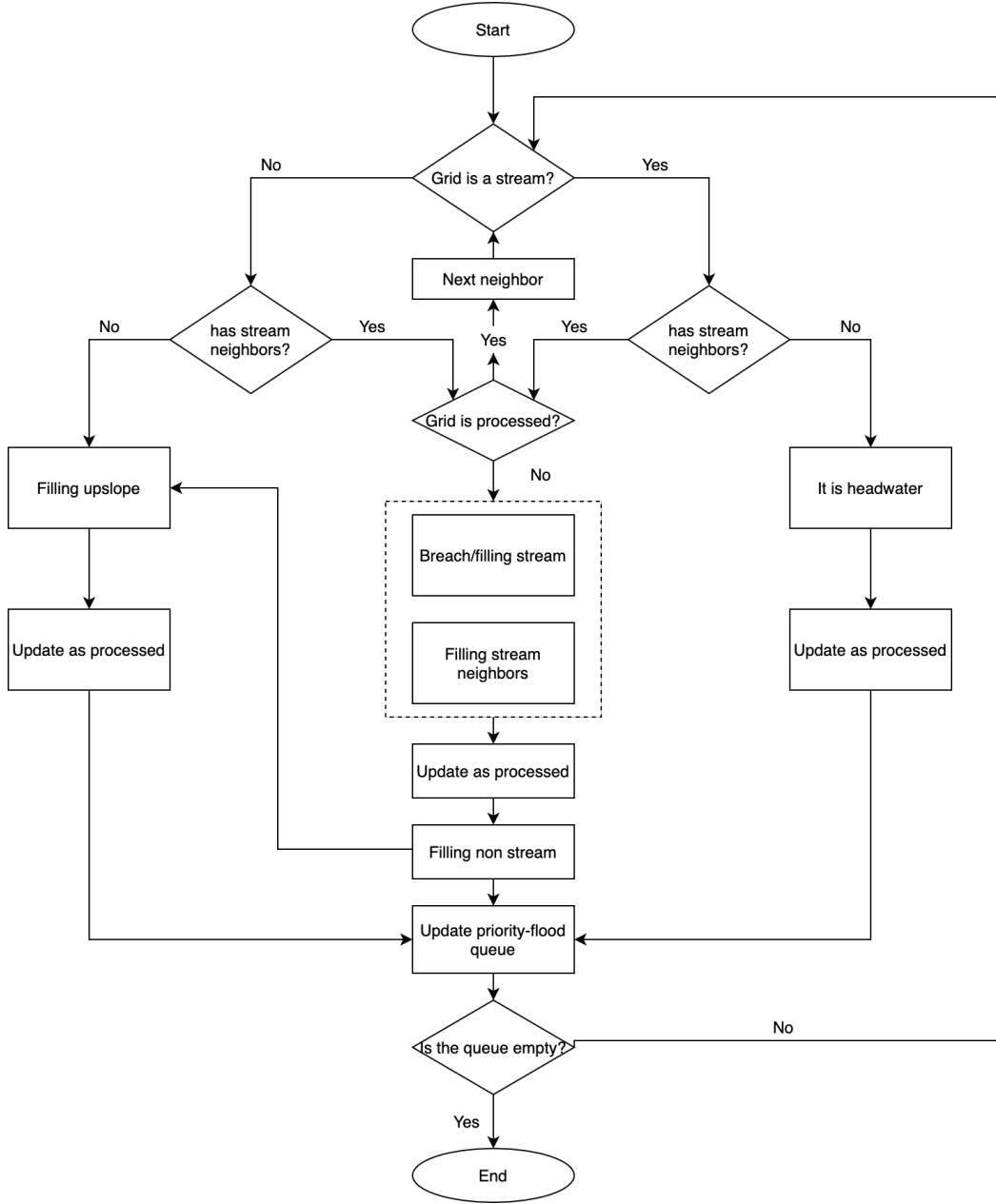


Figure 5: The workflow of the hybrid stream-burning-priority-flood depression filling algorithm.

The stream-burning algorithm can be turned on or off through newly added model configurations (Section 6.2 Table S1). Addition capabilities were also added to support advanced data management and visualization (Section 6.4 and 6.5).

HexWatershed 2.0 also supports multiple flow directions. As a result, the total number of upslope and downslope mesh cells is a constant of 6. The downslope flow direction with the steepest slope is defined as the dominant flow direction by default.

Both stream segment indices and order (Strahler) were calculated using the method introduced in our earlier study[16, 49]. The stream segment is defined as an individual stream channel connecting either headwater/outlet

with its immediate confluence or between two confluences. Each segment is assigned a unique ID ascending from headwater to outlet.

3. Application

3.1. Study area

135 To test the performance of HexWatershed 2.0, we applied the model to the Columbia River Basin (CRB). CRB is the drainage basin of the Columbia River in the Pacific Northwest region of North America. The drainage area of the CRB is approximately $6.7 \times 10^5 \text{ km}^2$ based on the Watershed Boundary Dataset (WBD)[50]. Its surface elevation ranges from -40 m to more than 4000 m above North American Vertical Datum of 1988 (NAVD 88). It contains both relatively high and low surface slopes in different areas (Figure 6). The largest tributary of the
140 Columbia River, the Snake River, flows from the southeast towards the west and joins the Columbia River at the center of the basin.

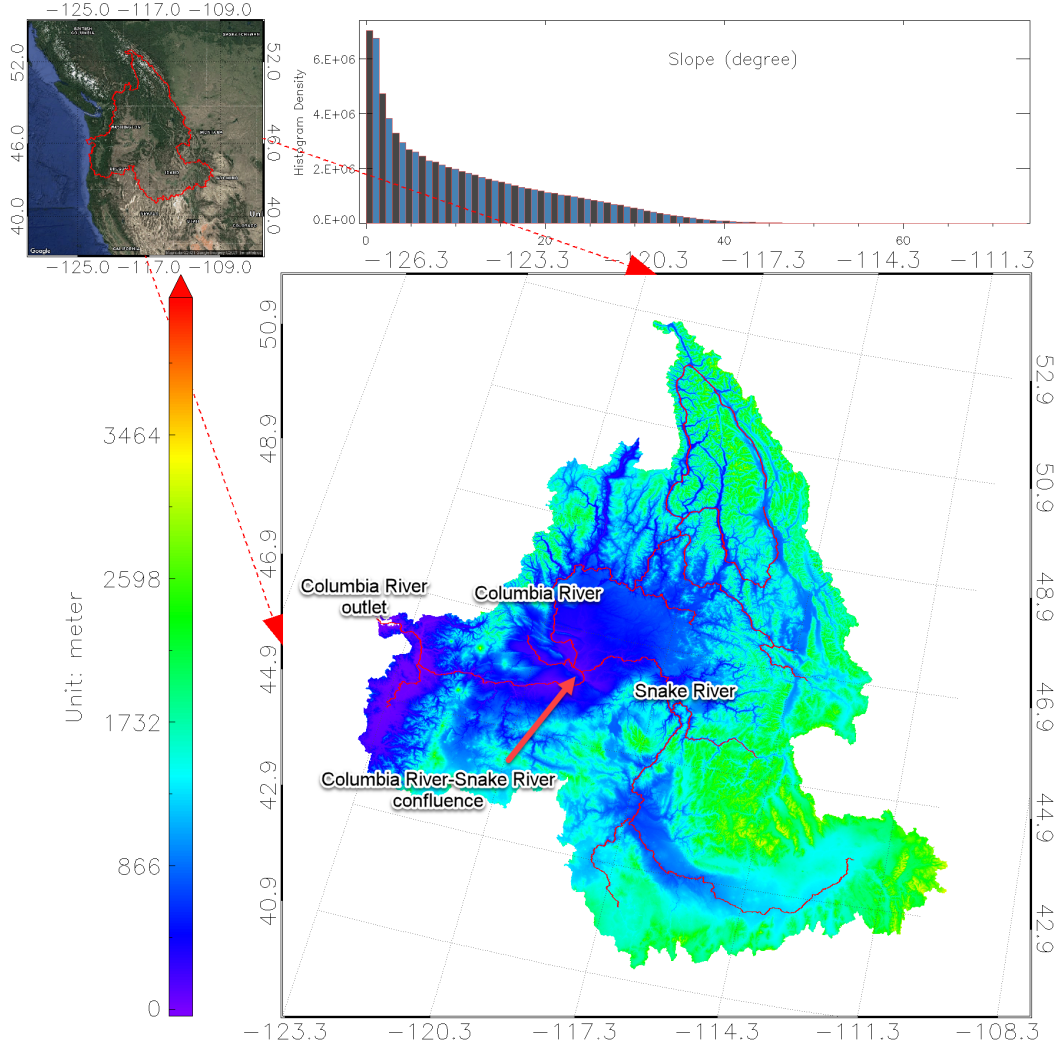


Figure 6: The location, surface elevation, and slope distribution of the Columbia River Basin. The upper left red polygon is the WBD watershed boundary on Google Map[50]; the upper right is the histogram of surface slope (degree); and the bottom is the topographic map (m). In the topographic map, the red lines are major rivers including the Columbia River (Drainage $6.7 \times 10^5 \text{ km}^2$) and Snake River (Drainage $2.8 \times 10^5 \text{ km}^2$).

3.2. Data

We collected raster DEM from the United States National Elevation Dataset (NED) and Canada Natural Resources (90 m resolution, 15000×16600)[51, 52]. We also obtained the stream networks and watershed boundary from the United States National Hydrography Dataset (NHD) and Watershed Boundary Dataset (WBD)[49].

The hexagon mesh can be generated using the QGIS-MMQGIS tool[53] or a Python script in the following steps: (1) retrieve the spatial reference (map projection) and extent of the DEM; (2) set the lower-left corner as the origin; (3) calculate the number of rows and columns based on the desired resolution and spatial extent; (4) calculate the vertex coordinates for each hexagon; (5) export the hexagons using the same spatial reference. To maintain consistency between different meshes, we define the hexagon mesh resolution (R_{hex}) using the equivalent

resolution, which is the root square of its area (Equation 1).

$$R_{hex} = \sqrt{\frac{3 \times \sqrt{3}}{2} L^2} \quad (1)$$

$$(2)$$

where L is hexagon mesh cell edge length (km).

Additionally, we “tagged” high order streams (Strahler stream order higher than 7) into the mesh using the QGIS intersection tool.

Detailed instructions for input data preparation are provided in the supplementary information (Section 6.6).

150 All input information is specified within the model configuration file (Table S1).

3.3. Model setup

Updated model configurations, including parameters for the stream burning feature, are listed in Table S1. To evaluate the sensitivity of HexWatershed 2.0 to spatial resolution and new features, we ran the model with different configurations with case indices used for illustrations (Table 1). These resolutions were selected to match up with
155 a flow direction dataset introduced in Section 4[54].

Table 1: Simulation configurations with case indices. The illustrations and analyses all use the same layout and indices.

Resolution	Nearest		Zonal mean	
	Stream burning off	Stream burning on	Stream burning off	Stream burning on
5 km	1	2	3	4
10 km	5	6	7	8
20 km	9	10	11	12
40 km	13	14	15	16

3.4. Results and analysis

Similar to version 1.0, HexWatershed 2.0 calculates most watershed delineation variables (flow direction, flow accumulation, river network, etc.). Although the new algorithms affect numerous results tied to these variables, we only present flow direction and stream networks as they are often used in hydrologic models. Graphical results
160 with the same index layout are presented in Table 1.

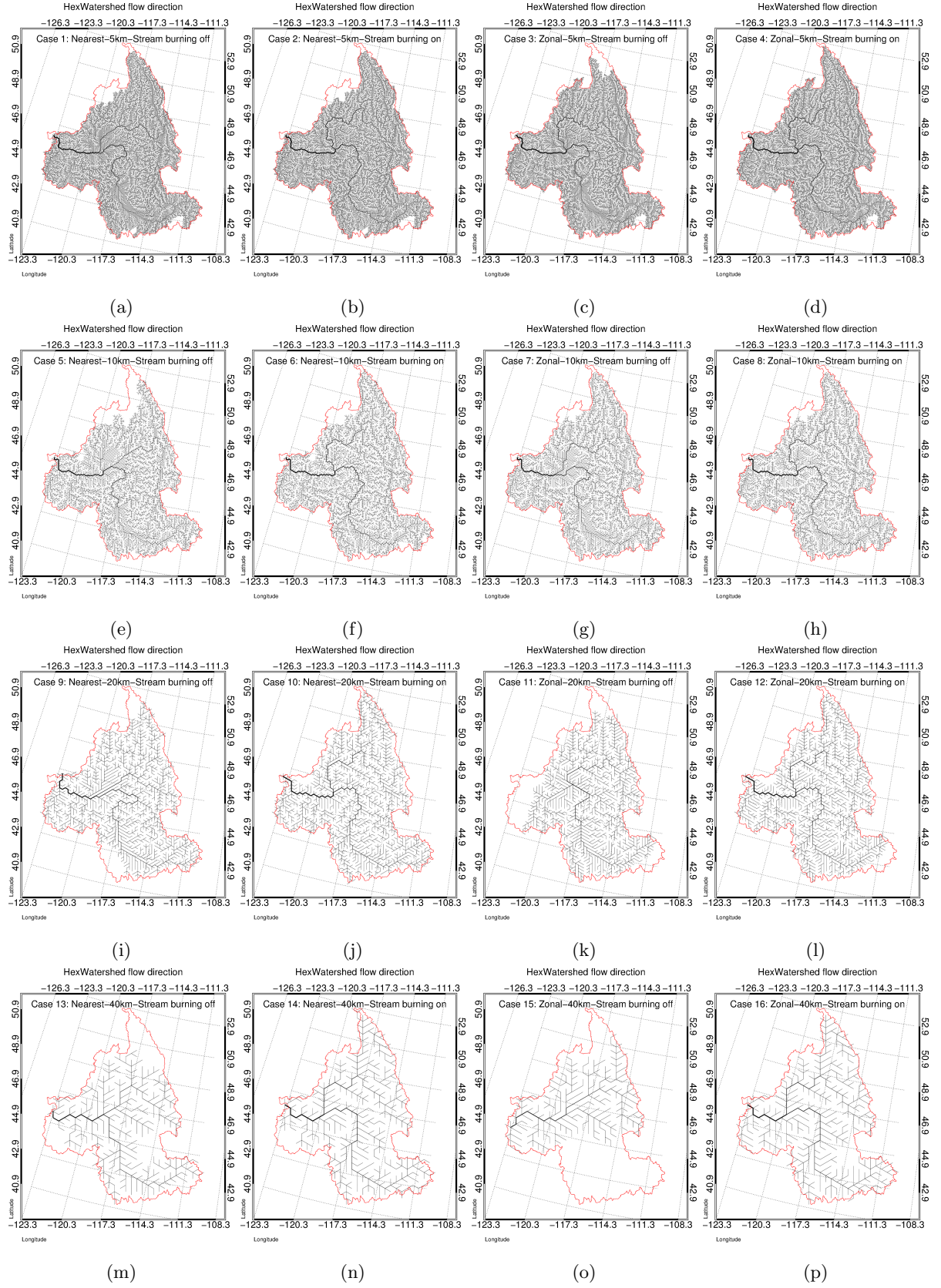


Figure 7: The spatial distributions of modeled flow direction in CRB from 16 cases in Table 1. The layout of figures and case indices is the same. Panels a to p correspond to Cases 1 to 16. Flow accumulation/upstream drainage area is scaled as the line thickness. The red line outlines the watershed boundary from the WBD dataset.

Additional information is provided in the supplementary information (Section 6.7 and 6.8). To maintain consistency, we only present final results that belong to the delineated watershed. For example, if the modeled flow direction is not within the delineated watershed, it is excluded even though it lies within the WBD based watershed boundary. Case 3 is such an example. NHD flowlines are used to evaluate and compare both flow direction and stream networks, but they are only plotted in Figure 8 to provide better visualizations.

3.4.1. Flow direction

The spatial distributions of flow direction for all 16 cases are illustrated in Figure 7. The results show that:

1. In general, model performance on flow direction decreases as spatial resolution decreases. For example, at high resolution (5 km and 10 km) (Case 1-4 and 5-8), the spatial coverage of modeled flow directions is close to the watershed boundary level. In contrast, at coarse resolution (40 km) (Case 13-16), the spatial coverage of modeled flow directions is much smaller due to the missing portions (e.g., upper right in Case 13 and lower right in Case 15). This is because the model is more likely to produce incorrect flow direction at some critical locations at coarse resolution, which results in these missing portions. There is no guarantee that the model will produce better flow direction even at finer spatial resolutions. For example, a portion in the upper right was captured in Case 7 (10 km), but not in Case 3 (5 km).
2. When stream burning is disabled (Columns 1 and 3 in Figure 7): Zonal mean resampling may produce a larger bias in spatial coverage than the nearest resample method (e.g., Case 15). Neither resampling method can reproduce the flow direction along the stream networks.
3. When stream burning is enabled (Columns 2 and 4 in Figure 7): The stream burning algorithm leads to the same flow direction near stream channels even though the corresponding elevations are different. Resampling method-induced differences in flow direction are only visible in areas far away from the stream channels (e.g., Case 14 and 16).
4. Regardless of the resampling method, the stream burning algorithm significantly improves the flow direction near the stream channels, especially at coarse resolutions (e.g., Columns 2 and 4 compared with Columns 1 and 3).

3.4.2. Stream networks

Similar to the flow direction, the spatial distributions of stream segments of all 16 cases are illustrated in Figure

8. The results show that:

1. Similar to flow direction, model performance on stream segments decreases as resolution decreases. For example, at 5 km resolution, modeled stream segments from Case 2 and 4 almost overlap with the NHD flowlines. In contrast, at 40 km resolution, modeled stream segments cannot preserve the spatial details.
2. When stream burning is disabled (Columns 1 and 3 in Figure 8), all cases failed to reproduce the NHD stream networks in flat areas (e.g., Case 1 and 3).
3. When stream burning is enabled (Columns 2 and 4 in Figure 8), modeled stream segments are close to NHD flowlines. The major differences are near river confluences, such as the Snake River-Columbia River confluence

(e.g., Case 14 and 16). The impacts of the resample method on stream segment are near low order streams such as headwaters (e.g., Case 10 and 12).

4. Even with the same accumulation threshold, the total number of stream segments differ due to differences in flow accumulation (e.g., Case 9-12).

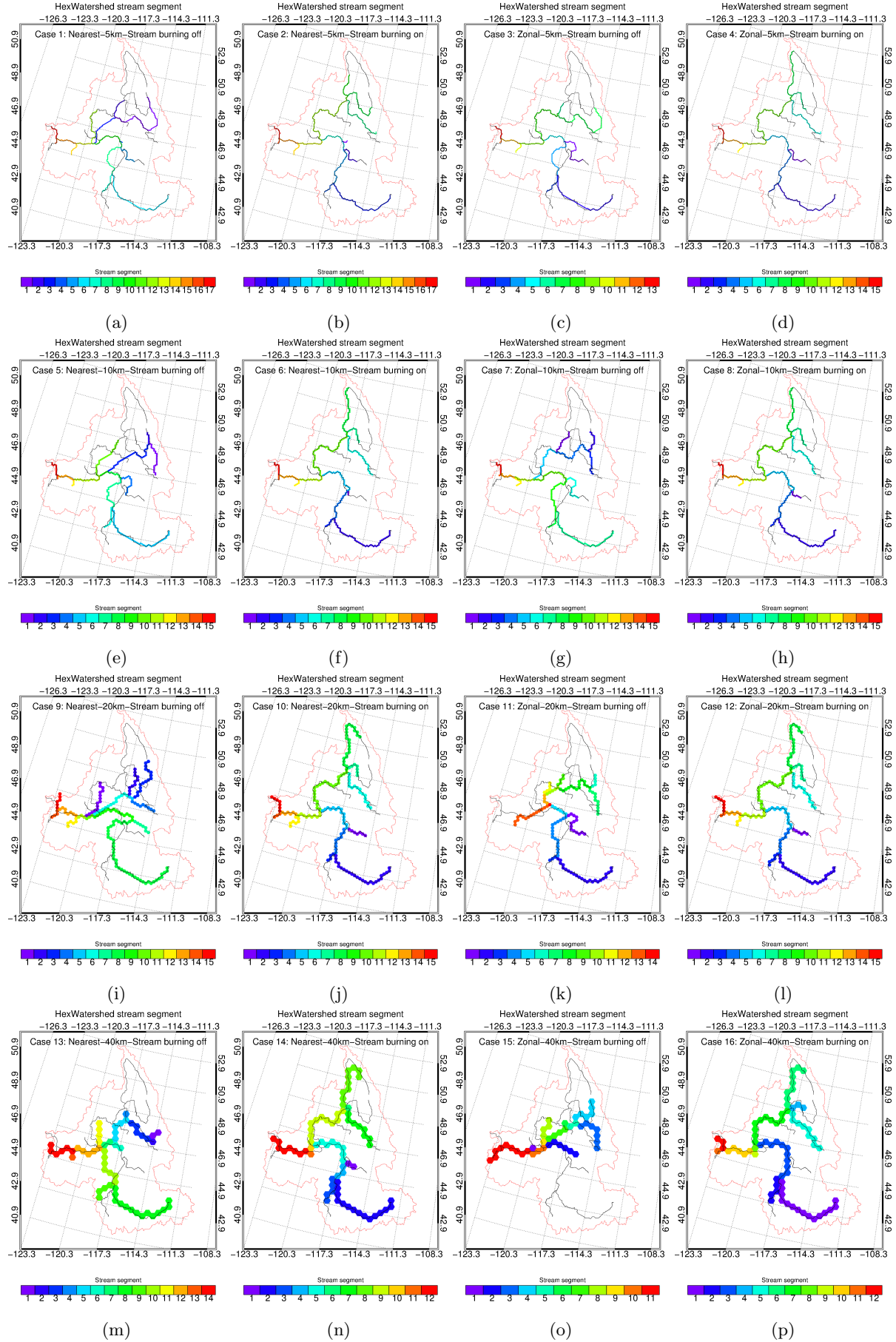


Figure 8: The spatial distributions of modeled stream segments in the CRB from 16 cases in Table 1. For better visualization, stream segments are represented by colored hexagonal polygons. Panels a to p correspond to Cases 1 to 16. The black line features are NHD flowlines. Colorbar indices represent the stream segment indices. The red line outlines the watershed boundary from the WBD dataset.

Similar to our earlier study[16], we calculated the enclosed area of differences between the modeled stream segments and NHD flowlines. We then used the data to evaluate the closeness of the modeled stream networks to the actual river networks. In this method, we used area to represent line feature discrepancy and closeness. If two or more line features intersect with each other, the intersected segments can be used to create enclosed polygons. In general, the smaller the total polygon area is, the closer these line features are. An illustration of this method is provided in Section 6.10 Figure S8. The comparison shows that at 40 km resolution, when stream burning is enabled, the enclosed area of difference is much smaller ($2.8 \times 10^4 \text{ km}^2$) than when stream burning is disabled ($7.1 \times 10^4 \text{ km}^2$) (Figure 9).

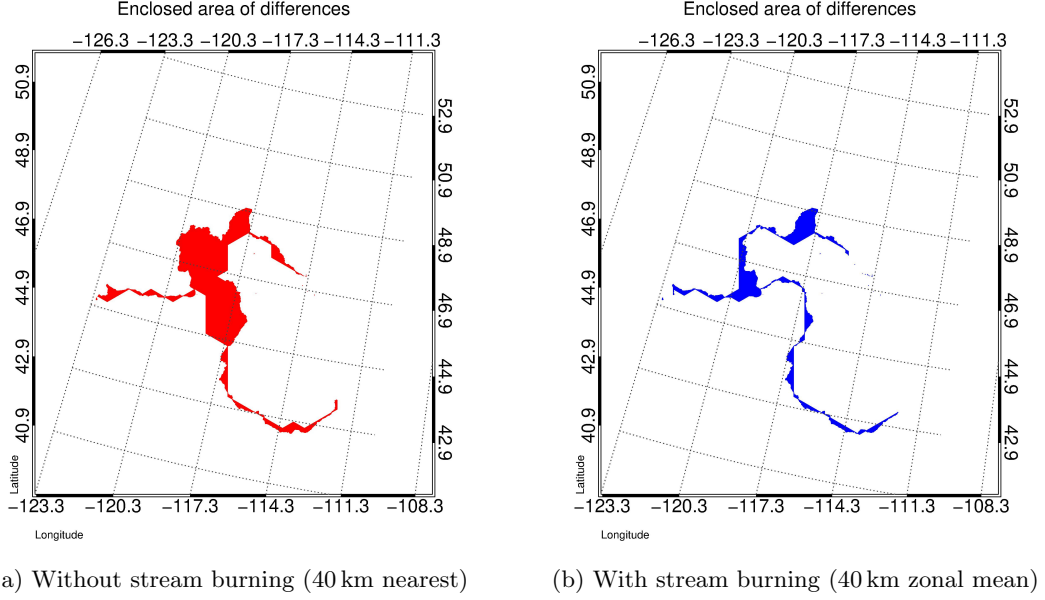


Figure 9: Differences between modeled stream segment and NHD flowlines through the enclosed area of differences at 40 km resolution with and without stream burning (Case 13 and 16 in Table 1). The red/blue polygons were generated by connecting the NHD flowlines with the modeled stream networks.

3.4.3. Drainage area

We compared modeled drainage area at Columbia River and Snake River outlets (Figure 10). The analysis shows that: As spatial resolution decreases, variations in modeled drainage area increase. When stream burning is disabled, both nearest (blue) and zonal mean (purple) resampling methods produce a large bias in the drainage area (30.0% increase or decrease). When stream burning is enabled (green and orange), the differences between modeled drainage area and the WBD dataset are less than 8.0% (increase or decrease) and the impacts from resampling methods are also less than 5.0%.

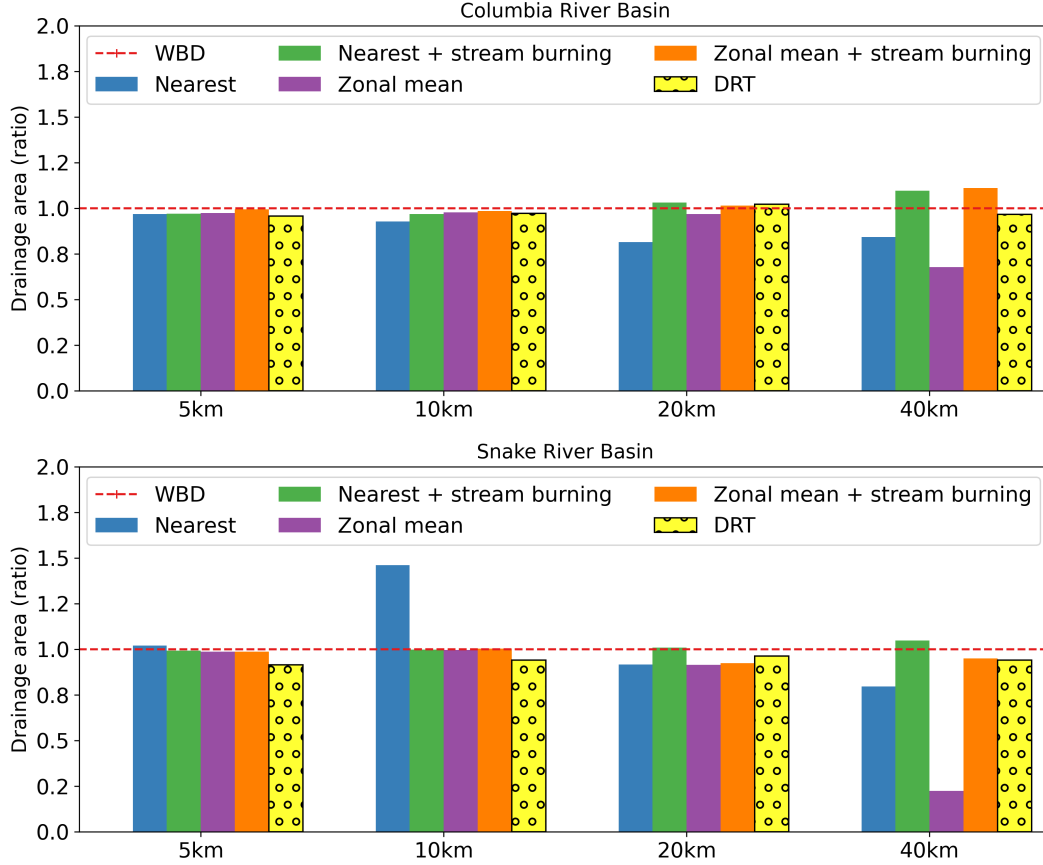


Figure 10: Comparison of the total drainage area of the Columbia River and Snake River from different model configurations. The X-axis is the spatial resolution and Y-axis is the ratio between modeled and observed (from Watershed Boundary Dataset (WBD)) drainage area. The dashed red line is the WBD drainage area. Colored (blue, green, purple, and orange) bars without texture represent the drainage area from 16 different model configurations in Table 1. Colored (yellow) bars with texture represent drainage area modeled by the Dominant River Tracing (DRT) model. Without stream burning, HexWatershed is more likely to produce an incorrect flow direction at some critical point, resulting in underestimated drainage area (see purple bar).

4. Comparison with DRT

To evaluate the performance of HexWatershed 2.0, we compared the model outputs to another flow direction model, the Dominant River Tracing (DRT) model[41]. DRT is a multi-scale flow direction model and has been used to generate flow direction maps for large-scale river routing models including the Model for Scale Adaptive River Transport (MOSART) and Variable Infiltration Capacity (VIC)[2, 41].

To address the stream networks representation issue raised at coarse resolutions, DRT uses both fine resolution river networks and flow accumulation instead of coarse resolution DEM to define flow directions. Like most large-scale flow direction models, DRT was developed based on the Cartesian latitude-longitude grids and defines flow direction (north, northeast, east, southeast, south, southwest, west, and northwest) using the D8 flow direction scheme[17, 41]. This introduces large biases in the high latitudes due to spatial distortion[55]. Currently, DRT has been used to generate flow direction maps at several resolutions (e.g., 1/2, 1/4, 1/8, and 1/16 degrees).

We mainly focus on the comparison at coarse resolutions because both models generally perform well at fine

resolutions (Section 6.9). Although both HexWatershed and DRT produce many products, we only focus on flow direction and drainage area given their widespread use by spatially-distributed hydrologic models.

Because of mesh differences, we compared HexWatershed with DRT using an “equivalent” resolution approach (Table 2). Specifically, because the actual resolution of latitude-longitude grids varies with latitude, the “equivalent” resolution (R_{drt}) is defined using the average latitude (Equation 3).

$$R_{drt} = R_{Earth} \times \sin\left(\frac{90 - Lat}{180} \times \pi\right) \times 2\pi \times \frac{R_{Lon}}{360} \quad (3)$$

where R_{Earth} is the average radius of Earth (km), Lat is the average latitude of CRB (degree), and R_{Lon} is the
230 MOSART resolution (degree).

Strictly speaking, any geometry, especially at coarse resolutions, on Earth’s surface is distorted because Earth is an imperfect sphere[56]. Therefore, the aforementioned area and resolution are only approximations.

Table 2: HexWatershed and its corresponding equivalent Dominant River Tracing (DRT) resolutions in the Columbia River Basin (-115°W, 45°N).

HexWatershed	DRT
5 km	1/16 degree
10 km	1/8 degree
20 km	1/4 degree
40 km	1/2 degree

The comparisons show that the spatial patterns of flow direction from both models are similar when stream burning is enabled (Figure 11 and 7).

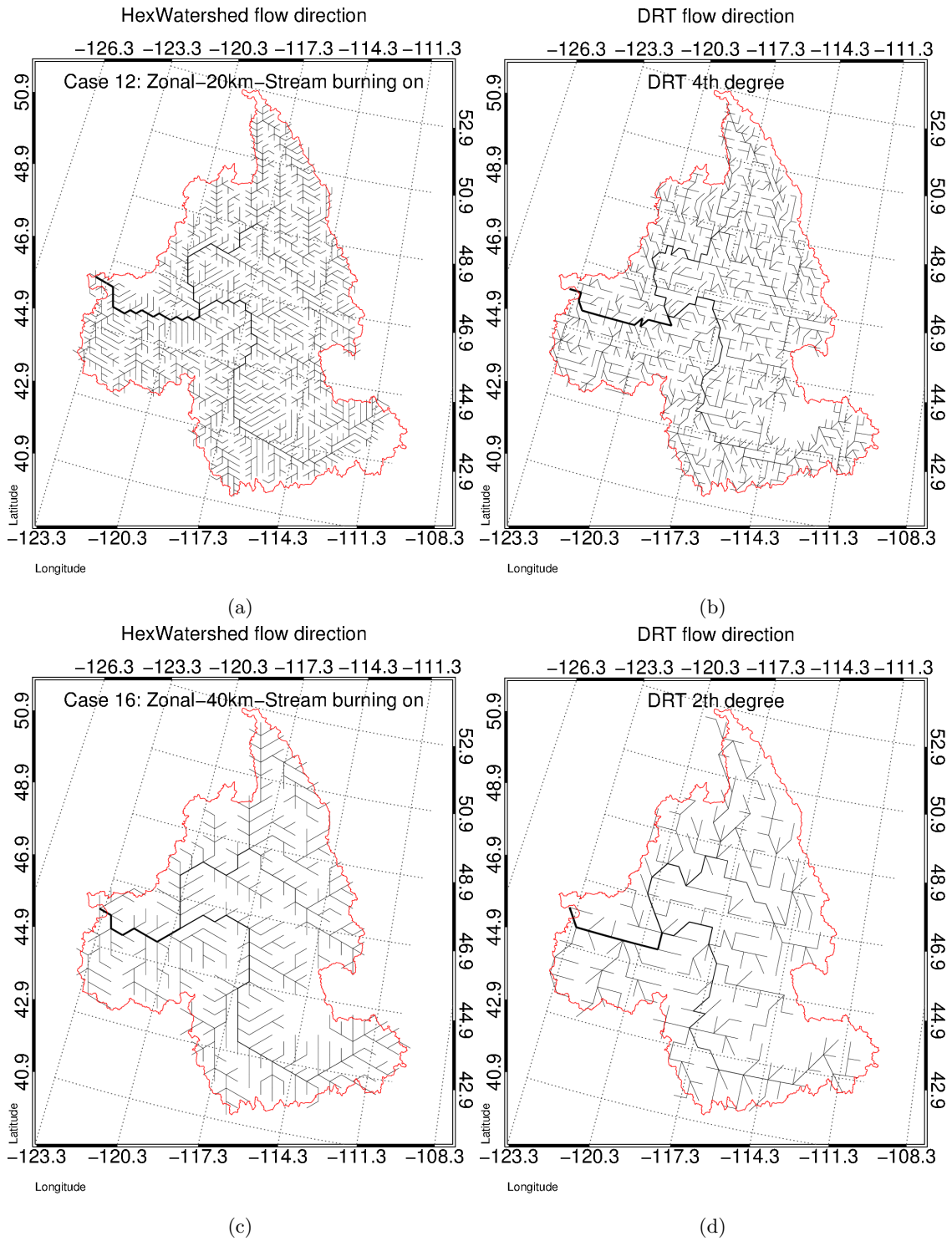


Figure 11: The spatial distributions of flow direction in the Columbia River Basin calculated by the HexWatershed 2.0 and Dominant River Tracing (DRT) models at multiple spatial resolutions: (a) 20 km (HexWatershed), (b) 1/4 degree (DRT), (c) 40 km (HexWatershed), and (d) 1/2 degree (DRT). Flow accumulation/upstream drainage area is scaled as line thickness. The red line outlines the watershed boundary from the WBD dataset.

Both HexWatershed and DRT produced similar drainage areas at the basin outlet and Snake River-Columbia River confluence (Figure 10).

5. Discussion

Simulation results show common patterns which were also observed in traditional watershed delineation processes (Figures 7 and 8)[33, 57]. This occurs because the mesh only affects the connectivity of the underlying data. The overall spatial patterns and the influences of algorithms on data are the same. Because of the differences in meshes, there are also some unique features and challenges.

5.1. Spatial resampling

Spatial resampling is important because it is the first step in watershed delineation and can affect several remaining steps. For flow direction, its impacts mainly manifest near the headwaters. It also affects the average slope between mesh cells. Compared with the nearest resampling, zonal mean resampling produces smoother average elevation and, consequently, milder slopes. However, there is no evidence that the zonal mean resampling method produces better flow direction. Advanced spatial resampling methods such as zonal mean provide an opportunity for more detailed geostatistical analysis (elevation variations, etc.).

5.2. Stream burning

The stream burning algorithm has a dominant impact on the delineated stream networks and overall model performance (Figure 8 and 10). When this feature is enabled, it directly defines the flow direction near stream channels and produces final stream segments that are most consistent with provided flowlines. The impact of spatial resampling on flow direction is thus negligible near stream channels because of the extensive manipulation by stream burning over the average elevation in the mesh.

However, the performance of stream burning is also subject to several factors, including the quality of user-provided flowline datasets. In our study, we only used high-order stream channels to simplify the simulation process. Low order streams can be included to further improve the model performance.

When stream burning is disabled, flow direction and stream networks are defined based on the DEM resampling results. While the delineated stream networks in our study deviate from the NHD flowlines, especially in the eastern Washington area, they actually represent the natural flow path formed by ice age glacial outburst floods well (Figure 8)[58, 59].

5.3. Spatial resolution

The spatial resolutions of input datasets play an important role in flow direction modeling. First, the spatial resolution of the raster DEM affects the performance of DEM resampling and the remaining steps. Ideally, the spatial resolution should always be finer than the desired hexagonal mesh resolution.

In our earlier study[16], we used much finer resolution hexagon meshes (30 m, 90 m and 100 m), causing the model to perform reasonably well even without stream burning. However, this is not applicable to this study. In all 16 simulation cases, the modeled flow direction contains large biases in many areas if stream burning is disabled because the hexagon resolutions are much higher (more than 5 km). This is consistent with other traditional watershed delineation methods. This also highlights the importance of stream burning, especially at coarse spatial

resolutions. Besides resolution, the quality of the DEM dataset is an important factor as it affects the resampling operation. In our study, we mainly use the NED dataset. Other DEM datasets might be used[60].

Although user-provided vector flowlines do not have “resolution”, their level of detail can be viewed as resolution and should be considered together with the hexagon mesh resolution. When the resolution of the hexagon mesh is coarse, only high order streams are important because low order streams may be mapped onto the same hexagon mesh cells along with the high order ones. When the resolution of the hexagon mesh is high, then datasets with low order streams (e.g., high resolution NHD flowlines) should be used.

5.4. Hybrid breaching-filling algorithm

The newly developed hybrid breaching-filling algorithm performs reasonably well (Figure 7 and 8) because it combines the strengths of both depression filling and stream burning. This enables the model to produce the correct flow direction while minimizing its modification to surface elevations. However, the modifications to stream channel elevations may be extensive, which can be problematic for river geometry (e.g., river bed slope) definition. In this study, the threshold parameters used to choose filling or breaching and the adjustment to the stream elevation are critical for the stream elevation calculation. More tests or other advanced methods are needed to improve or even replace this parameter.

5.5. Limitations

There are a few limitations to the method/work presented in this study:

1. The zonal mean resampling method may not perform well when hexagon mesh resolution is close to the source DEM resolution. Also, the zonal mean resampling method may not perform well if the hexagon mesh is partially covered by DEM at the boundary. To address these issues, an area-weighted zonal mean method should be used, especially at coarse resolutions[61].
2. The current stream burning method does not make use of stream order information. As a result, when a stream mesh cell is also a stream confluence, the breaching-filling algorithm does not prioritize the stream channel with higher order. This limitation can be addressed by adding the stream order information into the mesh.
3. The current breaching-filling algorithm modifies stream mesh cell elevations extensively, which may cause issues if stream channel geometry such as riverbed slope is desired. To address this, an improved approach (e.g, automated parameter adjustment) should be used to minimize the modification to the stream mesh cell elevations[40, 62]. To achieve this, stream topology information is required to allow the breaching algorithm to operate on the desired downstream mesh cell near the confluences.
4. Currently, lakes, regardless of size, are considered “local depressions” because they often have lower elevations than their surroundings. As a result, lakes are filled during depression filling, which results in incorrect flow directions. This becomes an issue in areas where the lake plays an important role in hydrology. A “Fill-Spill-Merge” approach should be used to consider lakes, especially endorheic lakes[42, 63, 64].

305 6. Conclusions

Based on our analysis and comparisons, we conclude that the hexagon mesh-based flow direction method could resolve several limitations in existing Cartesian grid-based methods. Through the addition of zonal mean resampling and stream burning capabilities, the HexWatershed 2.0 model significantly improves watershed delineation model performance compared to HexWatershed 1.0, especially at coarse resolutions. HexWatershed 2.0 improves the
310 flow direction both near and away from stream networks through stream burning and zonal mean resampling techniques. Because of that, the model is robust even in flat areas at coarse resolutions. With these new capabilities, HexWatershed can be applied to regional and large-scale hydrologic models as well as coupled Earth system models for land-river-ocean simulations.

Acknowledgements

315 This work was supported by the Earth System Model Development program areas of the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research as part of the multi-program, collaborative Integrated Coastal Modeling (ICoM) project. The data used for model simulations are listed in the tables and all input data and model outputs are archived on the computers at PNNL and will be available by contacting Chang Liao (chang.liao@pnnl.gov). The HexWatershed program can be accessed through the Github website
320 (<https://github.com/pnnl/hexwatershed>). A portion of this research was performed using PNNL Research Computing at Pacific Northwest National Laboratory. PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

Conflict of Interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any fi-
325 nancial interest (such as honoraria; educational grants; participation in speakers bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

- 330 [1] S. Neitsch, J. Arnold, J. Kiniry, J. R. Williams, K. W. King, SWAT2009 Theoretical Documentation, Texas Water Resources Institute Technical Report (406).
- [2] H. Li, M. S. Wigmosta, H. Wu, M. Huang, Y. Ke, A. M. Coleman, L. R. Leung, A physically based runoff routing model for land surface and earth system models, *Journal of Hydrometeorology* 14 (3) (2013) 808–828.
- [3] S. L. Markstrom, R. S. Regan, L. E. Hay, R. J. Viger, R. M. Webb, R. A. Payn, J. H. LaFontaine, PRMS-IV,
335 the precipitation-runoff modeling system, version 4, Tech. rep. (2015).

- [4] C. Liao, Q. Zhuang, Quantifying the role of snowmelt in stream discharge in an Alaskan watershed: An analysis using a spatially distributed surface hydrology model, *Journal of Geophysical Research: Earth Surface*.
- [5] C. Liao, Q. Zhuang, L. R. Leung, L. Guo, Quantifying Dissolved Organic Carbon Dynamics Using a ThreeDimensional Terrestrial Ecosystem Model at High SpatialTemporal Resolutions, *Journal of Advances in Modeling Earth Systems*.
- [6] T. G. Freeman, Calculating catchment area with divergent flow based on a regular grid, *Computers & Geosciences* 17 (3) (1991) 413–422.
- [7] P. Quinn, K. Beven, P. Chevallier, O. Planchon, The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrological processes* 5 (1) (1991) 59–79.
- [8] P. Holmgren, Multiple flow direction algorithms for runoff modelling in grid based elevation models: an empirical evaluation, *Hydrological processes* 8 (4) (1994) 327–334.
- [9] D. G. Tarboton, A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water resources research* 33 (2) (1997) 309–319.
- [10] P. Pilesjö, Q. Zhou, L. Harrie, Estimating flow distribution over digital elevation models using a form-based algorithm, *Geographic Information Sciences* 4 (1-2) (1998) 44–51.
- [11] S. Orlandini, G. Moretti, M. Franchini, B. Aldighieri, B. Testa, Pathbased methods for the determination of nondispersive drainage directions in gridbased digital elevation models, *Water resources research* 39 (6).
- [12] J. Seibert, B. L. McGlynn, A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models, *Water resources research* 43 (4).
- [13] S. Orlandini, G. Moretti, Determination of surface flow paths from gridded elevation data, *Water Resources Research* 45 (3).
- [14] S. D. Peckham, Mathematical surfaces for which specific and total contributing area can be computed: Testing contributing area algorithms, *Proceedings of Geomorphometry*.
- [15] J. P. Wilson, G. AGGETT, D. Yongxin, Water in the landscape: a review of contemporary flow routing algorithms, in: *Advances in digital terrain analysis*, Springer, 2008, pp. 213–236.
- [16] C. Liao, T. Tesfa, Z. Duan, L. R. Leung, Watershed delineation on a hexagonal mesh grid, *Environmental Modelling & Software* 128 (2020) 104702. doi:10.1016/j.envsoft.2020.104702.
URL <http://www.sciencedirect.com/science/article/pii/S1364815219308278>
- [17] ESRI, Flow Direction (2020).
URL <https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/flow-direction.htm>

- [18] Y. Fan, G. MiguezMacho, C. P. Weaver, R. Walko, A. Robock, Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, *Journal of Geophysical Research: Atmospheres* 112 (D10).
- 370 [19] M. Camporese, C. Paniconi, M. Putti, S. Orlandini, Surfacesubsurface flow modeling with pathbased runoff routing, boundary conditionbased coupling, and assimilation of multisource observation data, *Water Resources Research* 46 (2).
- [20] A. W. Harbaugh, MODFLOW-2005, the US Geological Survey modular ground-water model: The ground-water flow process, US Department of the Interior, US Geological Survey Reston, VA, USA, 2005.
- 375 [21] C. D. Langevin, J. D. Hughes, E. R. Banta, R. G. Niswonger, S. Panday, A. M. Provost, Documentation for the MODFLOW 6 groundwater flow model, Tech. rep. (2017).
- [22] T. Ringler, M. Petersen, R. L. Higdon, D. Jacobsen, P. W. Jones, M. Maltrud, A multi-resolution approach to global ocean modeling, *Ocean Modelling* 69 (2013) 211–232.
- [23] W. C. Skamarock, J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, T. D. Ringler, A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering, *Monthly Weather Review* 140 (9) (2012) 3090–3105.
- 380 [24] S. Danilov, D. Sidorenko, Q. Wang, T. Jung, The finite-volume sea iceocean model (fesom2), *Geoscientific Model Development* 10 (2) (2017) 765–789.
- [25] V. Y. Ivanov, E. R. Vivoni, R. L. Bras, D. Entekhabi, Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resources Research* 40 (11).
- 385 [26] V. Y. Ivanov, E. R. Vivoni, R. L. Bras, D. Entekhabi, Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach, *Journal of Hydrology* 298 (1–4) (2004) 80–111.
- [27] C. Paniconi, M. Putti, Physically based modeling in catchment hydrology at 50: Survey and outlook, *Water Resources Research* 51 (9) (2015) 7090–7129.
- 390 [28] C. B. Marsh, J. W. Pomeroy, H. S. Wheeler, The Canadian Hydrological Model (CHM) v1. 0: a multi-scale, multi-extent, variable-complexity hydrological model design and overview, *Geoscientific Model Development* 13 (1) (2020) 225–247.
- [29] H. N. Davies, V. A. Bell, Assessment of methods for extracting low-resolution river networks from high-resolution digital data, *Hydrological Sciences Journal* 54 (1) (2009) 17–28.
- 395 [30] J. Thuburn, T. D. Ringler, W. C. Skamarock, J. B. Klemp, Numerical representation of geostrophic modes on arbitrarily structured C-grids, *Journal of Computational Physics* 228 (22) (2009) 8321–8335.
- [31] H. Weller, J. Thuburn, C. J. Cotter, Computational modes and grid imprinting on five quasi-uniform spherical C grids, *Monthly Weather Review* 140 (8) (2012) 2734–2755.

- [32] P. J. Wolfram, O. B. Fringer, Mitigating horizontal divergence checker-board oscillations on unstructured triangular C-grids for nonlinear hydrostatic and nonhydrostatic flows, *Ocean Modelling* 69 (2013) 64–78.
- [33] T. Goulden, C. Hopkinson, R. Jamieson, S. Sterling, Sensitivity of watershed attributes to spatial resolution and interpolation method of LiDAR DEMs in three distinct landscapes, *Water Resources Research* 50 (3) (2014) 1908–1927.
- [34] W. Buakhao, A. Kangrang, DEM resolution impact on the estimation of the physical characteristics of watersheds by using SWAT, *Advances in Civil Engineering* 2016.
- [35] D. Weber, E. Englund, Evaluation and comparison of spatial interpolators, *Mathematical Geology* 24 (4) (1992) 381–391.
- [36] P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, *Geographical information systems: principles, techniques, applications, and management*, John Wiley & Sons, 1999.
- [37] P. V. Arun, A terrain-based hybrid approach towards DEM interpolation, *Annals of GIS* 19 (4) (2013) 245–252.
- [38] A. Francipane, V. Y. Ivanov, L. V. Noto, E. Istanbuluoglu, E. Arnone, R. L. Bras, tRIBS-Erosion: A parsimonious physically-based model for studying catchment hydro-geomorphic response, *Catena* 92 (2012) 216–231.
- [39] F. Hellweger, D. Maidment, AGREE-DEM surface reconditioning system, Online at <http://www.ce.utexas.edu/prof/maidment/gishydro/ferdi/research/agree/agree.html> (last accessed March 3, 2005).
- [40] J. B. Lindsay, The practice of DEM stream burning revisited, *Earth Surface Processes and Landforms* 41 (5) (2016) 658–668.
- [41] H. Wu, J. S. Kimball, N. Mantua, J. Stanford, Automated upscaling of river networks for macroscale hydrological modeling, *Water Resources Research* 47 (3).
- [42] D. Yamazaki, T. Oki, S. Kanae, Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map, *Hydrology and Earth System Sciences* 13 (11) (2009) 2241–2251. doi:10.5194/hess-13-2241-2009.
- [43] R. Barnes, C. Lehman, D. Mulla, Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models, *Computers & Geosciences* 62 (2014) 117–127. doi:10.1016/j.cageo.2013.04.024.
- [44] M. Winchell, R. Srinivasan, M. Di Luzio, J. Arnold, *ArcSWAT interface for SWAT 2005, Users Guide*, Blackland Research Center, Texas Agricultural Experiment Station, Temple.
- [45] W. Lidberg, M. Nilsson, T. Lundmark, A. M. Ågren, Evaluating preprocessing methods of digital elevation models for hydrological modelling, *Hydrological Processes* 31 (26) (2017) 4660–4668.
- [46] Esri Water Resources Team, *Arc Hydro Tools - Tutorial*, Tech. rep. (2011).

- [47] GDAL/OGR contributors, Geospatial Data Abstraction software Library (2019).
URL <https://gdal.org>
- [48] J. B. Lindsay, Efficient hybrid breachingfilling sink removal methods for flow path enforcement in digital elevation models, *Hydrological Processes* 30 (6) (2016) 846–857.
- 435 [49] USGS, 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>] (2013).
URL <http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd>
- [50] USDA-NRCS, EPA, USGS, Watershed Boundary Dataset, Tech. rep. (2017).
- [51] D. Gesch, M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, D. Tyler, The national elevation dataset, Pho-
440 togrammetric engineering and remote sensing 68 (1) (2002) 5–32.
- [52] N. R. C. (NRC), Canadian digital elevation model (CDEM). Government of Canada (2015).
URL <https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333>.
- [53] M. Minn, MMQGIS for QGIS Geographic Information System, Open Source Geospatial Foundation Project. [(accessed on 17 May 2018)].
- 445 [54] H. Wu, J. S. Kimball, H. Li, M. Huang, L. R. Leung, R. F. Adler, A new global river network database for macroscale hydrologic modeling, *Water resources research* 48 (9).
- [55] A. Sood, V. Smakhtin, Global hydrological models: a review, *Hydrological Sciences Journal* 60 (4) (2015) 549–565.
- [56] C. M. R. Fowler, C. M. R. Fowler, M. Fowler, *The solid earth: an introduction to global geophysics*, Cambridge
450 University Press, 1990.
- [57] J. N. Callow, K. P. Van Niel, G. S. Boggs, How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis?, *Journal of hydrology* 332 (1-2) (2007) 30–39.
- [58] J. H. Bretz, H. T. U. SMITH, G. E. Neff, Channeled Scabland of Washington: New data and interpretations, *Geological Society of America Bulletin* 67 (8) (1956) 957–1049.
- 455 [59] N. R. W. G. Survey, Washington’s Ice Age Floods, Tech. rep. (2018).
URL <https://wadnr.maps.arcgis.com/apps/Cascade/index.html?appid=84ea4016ce124bd9a546c5cbc58f9e29>
- [60] D. Yamazaki, D. Ikeshima, J. Sosa, P. D. Bates, G. H. Allen, T. M. Pavelsky, MERIT Hydro: A highresolution global hydrography map based on latest topography dataset, *Water Resources Research* 55 (6) (2019) 5053–5073.
- 460 [61] P. A. Ullrich, M. A. Taylor, Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part I, *Monthly Weather Review* 143 (6) (2015) 2419–2440.

- [62] J. B. Lindsay, W. Yang, D. D. Hornby, Drainage network analysis and structuring of topologically noisy vector stream data, *ISPRS International Journal of Geo-Information* 8 (9) (2019) 422.
- [63] R. Barnes, K. L. Callaghan, A. D. Wickert, Computing water flow through complex landscapes, Part 3: Fill-Spill-Merge: Flow routing in depression hierarchies, *Earth Surface Dynamics Discussions* (2020) 1–22.
- [64] K. Liu, C. Song, L. Ke, L. Jiang, R. Ma, Automatic watershed delineation in the Tibetan endorheic basin: A lake-oriented approach based on digital elevation models, *Geomorphology* (2020) 107127.
- [65] B. Lorensen, VTK Example (2020).
URL <https://lorensen.github.io/VTKExamples/site/>