Watershed Delineation On A Hexagonal Mesh Grid

Chang Liao[a], Teklu Tesfa[b], Zhuoran Duan[b], L. Ruby Leung[a]

[a]*Atmospheric Sciences and Global Change, Pacific Northwest National Laboratory, Richland, WA, USA*
[b]*Hydrology Group, Pacific Northwest National Laboratory, Richland, WA, USA*

Abstract

Spatial discretization is the cornerstone of all spatially-distributed numerical simulations including watershed hydrology. Traditional square grid spatial discretization has several limitations including inability to represent adjacency uniformly. In this study, we developed a watershed delineation model (HexWatershed) based on the hexagon grid spatial discretization. We applied this model to two different types of watershed in the US and we evaluated its performance against the traditional method. The comparisons show that the hexagon grid spatial discretization exhibits many advantages over the tradition method. We propose that spatially distributed hydrologic simulations should consider using a hexagon grid spatial discretization.

Keywords: Hydrology, Hexagon, Watershed delineation, Digital Global Grid System

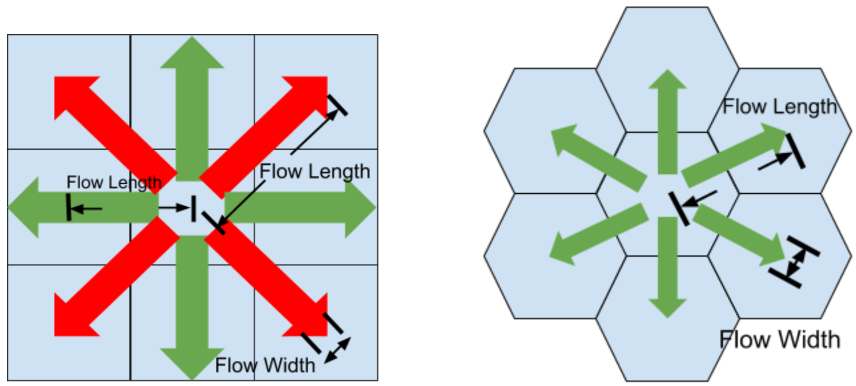
1. Introduction

Spatial discretization is the cornerstone of all spatially distributed numerical simulations including hydrologic simulations. In hydrologic modeling the study domain is commonly discretized using a Square Grid Spatial Discretization (SGSD). Few studies have investigated the performance of other spatial discretizations such as Hexagon Grid Spatial Discretization (HGSD) in hydrology [35, 33].

By definition, spatial discretization is the representation of the continuous real world with discrete information. In Geographic Information System (GIS), SGSD is the most widely used approach to represent spatial information. For example, a raster Digital Elevation Model (DEM) dateset is usually used to describe the surface elevation of a Region Of Interest (ROI) on the

Earth's surface [8]. Because SGSD data structure can be represented by a rectangular array /matrix directly, it is convenient for computation, analysis, visualization and storage. However, SGSD has several limitations.

First, SGSD cannot represent adjacency uniformly [4]. In a Cartesian coordinate system, each grid has two types of neighbors: direct and diagonal. The distances between the center of a grid and the center of its diagonal neighbors are further than that of the direct ones (Figure 1a). As a result, hydrologic models have to assign different weights, arbitrarily or empirically, to account for the differences in travel distances. These differences are also not always treated consistently. For example, in a coupled surface and subsurface hydrologic simulation, water flow in the diagonal direction may be considered in the surface hydrology component whereas it is ignored in the groundwater hydrology component [10, 12, 19, 15]. Consequently, it becomes one of the major model uncertainty sources (input data, model structure and parameters) in hydrologic modeling. In the remainder of the paper, the term "uncertainty" refers to model uncertainty caused by model structure unless otherwise specified. Because the diagonal neighbors are connected through the vertices instead of faces, they cannot represent stream width information correctly (Figure 1a). Similarly, flow width information of the surface runoff is misrepresented.



(a) (b)

Figure 1: Illustration of traditional D4/D8 neighbor definitions and the hexagonal D6 neighbor definitions. (a) the center square grid have 4 direct/face (green arrows) and 4 diagonal/vertex (red arrows) neighbors; (b) the center hexagon grid has 6 face neighbors (green arrows). The arrows also represent flow path with both length and width information. In D4/D8, diagonal flow length is longer than direct flow length, and flow width is out of boundary. In hexagonal D6, flow length is the same and flow width is within boundary.

Second, SGSD will create "island" effect due to the differences in D4 and D8 neighbor definitions, which causes problems for numerical simulations [3]. In this study, we define a single or group of grids that are connected through diagonal path at the edge of boundaries as an island. For this reason, watershed delineation results usually require tedious manual correction to eliminate these diagonal islands between subbasin boundaries [14]. Besides, because most groundwater flow models do not consider D8 neighbors, we cannot couple them with surface hydrology models directly [10, 19]. For example, we need to set the grid #1 as inactive in a coupled Groundwater and Surface Water Flow Model (GSFLOW) simulation (Figure 2) [18].

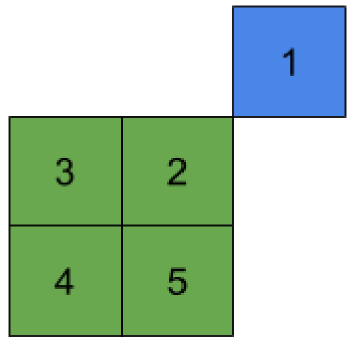


Figure 2: Illustration of "island" effect caused by D8 diagonal neighbor definition. The blue grid (#1) is a D8 diagonal neighbor of the green grid (#2). This blue grid can occur either within model domain or at the edge [14].

Third, SGSD cannot effectively represent a spherical topology, which will introduce significant spatial distortions (Figure 3). Hydrologic simulations at global scale that use longitude/latitude mesh grid will be undermined, especially when the ROI is also the most distorted areas. For this reason, longitude/latitude based river routing models (e.g., MOdel for Scale Adaptive River Transport (MOSART)) may contain larger uncertainty at high latitudes [24, 16]. Furthermore, as global scale oceanic models do not use SGSD method, it becomes cumbersome to couple land surface/hydrologic models with oceanic models.

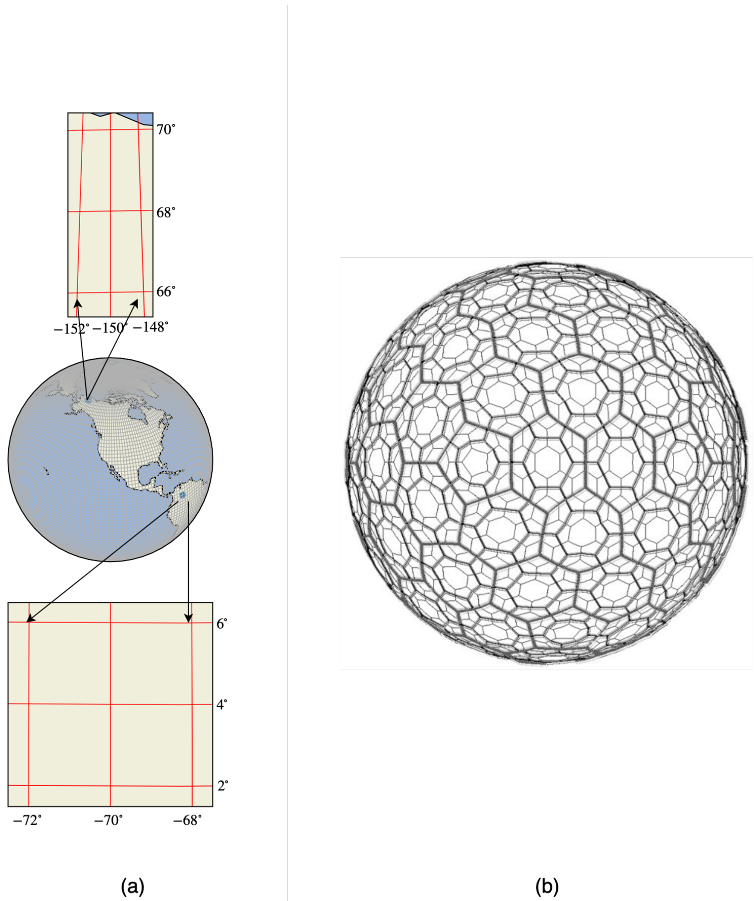


Figure 3: Illustration of the spatial distortion caused by longitude/latitude based SGSD method and a Discrete Global Grid System (DGGS) with uniform resolution. (a) is a longitude/latitude mesh at $2^\circ \times 2^\circ$ resolution. The ratio of distance in latitude to longitude increases with latitude. Top and bottom plots in (a) are zoom-in regions in Alaska, USA, and east Columbia. Their corresponding ratios are close to 2.5 and 1.0, respectively. (b) is a DGGS grid generated by DGGRID, which is made up by mostly hexagons [25][1]. All the hexagons have nearly the same resolution.

Other flow direction methods based on the SGSD also have similar limitations. For example, the D-infinity flow direction method can describe the flow direction in 360°and improve the partition of water flow in different directions [28]. However, these methods do not resolve the SGSD limitations

[1] There are 12 pentagons.

fundamentally and they are relatively less used in hydrologic simulations.

Triangular irregular networks (TIN) Spatial Discretization (TINSD) is also used in hydrologic models. One advantage of the TINSD method is that points of a TIN are distributed variably so it can provide high resolution near ROI whereas low resolution elsewhere [7]. However, this method is less popular because of the complex data structure. Because it has two types of neighbor connectivity (3 face neighbors and multiple vertice neighbors), it may also introduce uncertainty. For example, the Penn State Integrated Hydrologic Model (PIHM) only considers water flow through the face neighbors but ignores the vertice neighbors [22]. Hereafter, our discussion excludes TINSD unless otherwise specified. Watershed boundary-based spatial discretization (WBSD) is also used in large scale hydrologic simulations [30, 31]. However, this method essentially depends on the availability of watershed boundaries, which mostly come from SGSD based watershed delineation processes.

In contrast, the HGSD method can resolve these limitations:

1. In HGSD, each grid has only one type of neighbor with the same connectivity and distance (Figure 1b). As a result, we can route both surface and subsurface water flow consistently without using different weights, thus getting rid of the decadal old assumption on travel length. This will improve spatially distributed hydrologic models that rely on grid connectivity [17].
2. The "island" effect is automatically eliminated because all neighbors are connected through faces. No manual corrections are needed to resolve the diagonal traveling path issue [14].
3. It can provide continental to global coverage at consistent or variable spatial resolutions (Figure 3b) [26]. It can be used to couple land surface/hydrologic models with oceanic models using a unified mesh grid (e.g., the Voronoi tessellation of the Model for Prediction Across Scale (MPAS)) [5].

Additionally, it has other advantages:

1. It can be used for coupled surface (D6) and subsurface (9-point structured connectivity) hydrologic modeling to resolve the inconsistency in connectivity.
2. The conceptual model is more compatible with the flow width information because the flow path can be contained within the grid boundary (Figure 1b).

3. It can improve model performance as many studies show that numerical simulations based on hexagon grid perform better when compared with other mesh grids [4, 35, 34, 9].
4. Other flow direction methods (e.g., D-infinity) can also be implemented on HGSD with modifications to improve flow direction and partitions [28].

In recent decades, HGSD is widely used in Discrete Global Grid System (DGGS) [35, 23]. For example, the Icosahedral Snyder Equal Area (ISEA) tessellation method is used to generate a geodesic grid system within which most grids can be hexagons [26].

Despite all the advantages the HGSD method can provide, it is not widely used in numerical simulations. Partially it is because most existing datasets were generated in the SGSD format and numerical models were not designed to use the HGSD datasets. For example, nearly all the spatially distributed hydrologic models were developed based on square or unstructured grid discretization, and very few studies have used HGSD.

In this study, we made the first attempt to develop a watershed delineation model (HexWatershed) with a set of algorithms based on the HGSD method. In Section 2 we introduce the model algorithms. In Section 3 we apply the model to two different types of watersheds and analyze the model outputs. In Section 4 we evaluate the model performance against outputs from the traditional SGSD method. In Section 5 and 6 we discuss the limitations and future work.

2. Model Algorithm

2.1. Overview

Following the traditional watershed delineation algorithms, we developed a list of algorithms for the HGSD method. Because these algorithms are fundamentally similar in principle, we mainly focus on the differences that were introduced in the new model. Last, we describe the software requirements to run the HexWatershed model.

2.2. Hexagon Grid Resolution

The hexagon grid resolution is defined using grid area instead of edge length so it is comparable with the traditional square grid [33]. For example, if the edge length of a hexagon is $10\,m$, its area is approximately $259.81m^2$, then its equivalent/effective square grid resolution is 16.11 m, which is the square root of its area (Equation 1 and 2).

$$A = \frac{3 \times \sqrt{3}}{2} L^2 \tag{1}$$

$$R_e = \sqrt{A} \tag{2}$$

where A is the hexagon area; L is the actual hexagon edge length; and R_e is the effective hexagon resolution.

Although applications of HexWatershed in this study use a constant resolution hexagonal mesh (Section 3), it also supports variable resolution hexagonal mesh. All algorithms within HexWatershed are designed and implemented independent of resolution. For example, the stream grid algorithm considers the total drainage area because area of each hexagon grid may be different (Section 2.9).

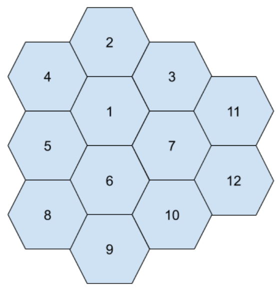
2.3. Grid Topology

In SGSD, grid is often referred by its array/matrix indices (i, j) and its neighbors can be referred by moving up/down the indices $(i \pm 1, j \pm 1)$. However, in HGSD, we cannot use array indices directly unless a dedicated hexagon grid index system is available [27]. As a result, an algorithm is required to obtain the topology. Depending on how the hexagon grid was generated, there are different ways to obtain the grid topology.

For the sake of generality, we assume that there is no prior grid topology stored within the hexagon grid. HexWatershed rebuilds the topology with three steps:

1. Assign a unique global ID for each hexagon;
2. For each hexagon, identify hexagons that share the same vertex/edge as neighbors; and
3. Save the global IDs of these neighbors into a look-up table for each hexagon.

The final look-up table of this algorithm is illustrated in Figure 4.



(a)

Hexagon ID	Neighbor ID					
1	2	3	4	5	6	7
2	1	3	4			
3	1	2	7	11		
...						

(b)

Figure 4: Illustration of hexagon topology. The indices are hexagon grid global IDs. (a) is a hexagon grid; and (b) is the look-up table which stores the hexagon topology. Each hexagon can have a maximal of 6 neighbors.

2.4. DEM Resampling

Similar to traditional raster datasets that use grid center to store information, HexWatershed uses the hexagon center to store elevation.

Theoratically, hexagonal DEM can be obtained by resampling from either high resolution traditional DEM, or high resolution hexagonal DEM. In this study, the former approach was used because most available DEM datasets are stored in the SGSD format. To further simplify this process, the nearest neighbor resampling method is used.

2.5. DEM Depression Filling

Similar to traditional DEM, hexagonal DEM could potentially have local depressions when generated. We developed an algorithm following the method proposed by Richard Barnes, which uses the priority-flood method to fill the depressions in any grid system [2]. Priority-flood is an efficient algorithm to fill DEM depressions by sequentially "flooding" the domain from the boundary inward to adjust elevations to assure that surface will drain [21]. To ensure an absolute drain, a minimal slope (0.01 in percentage) is added when applicable [32]. The step-by-step instructions are provided in Appendix B and illustrated in Figure 5.

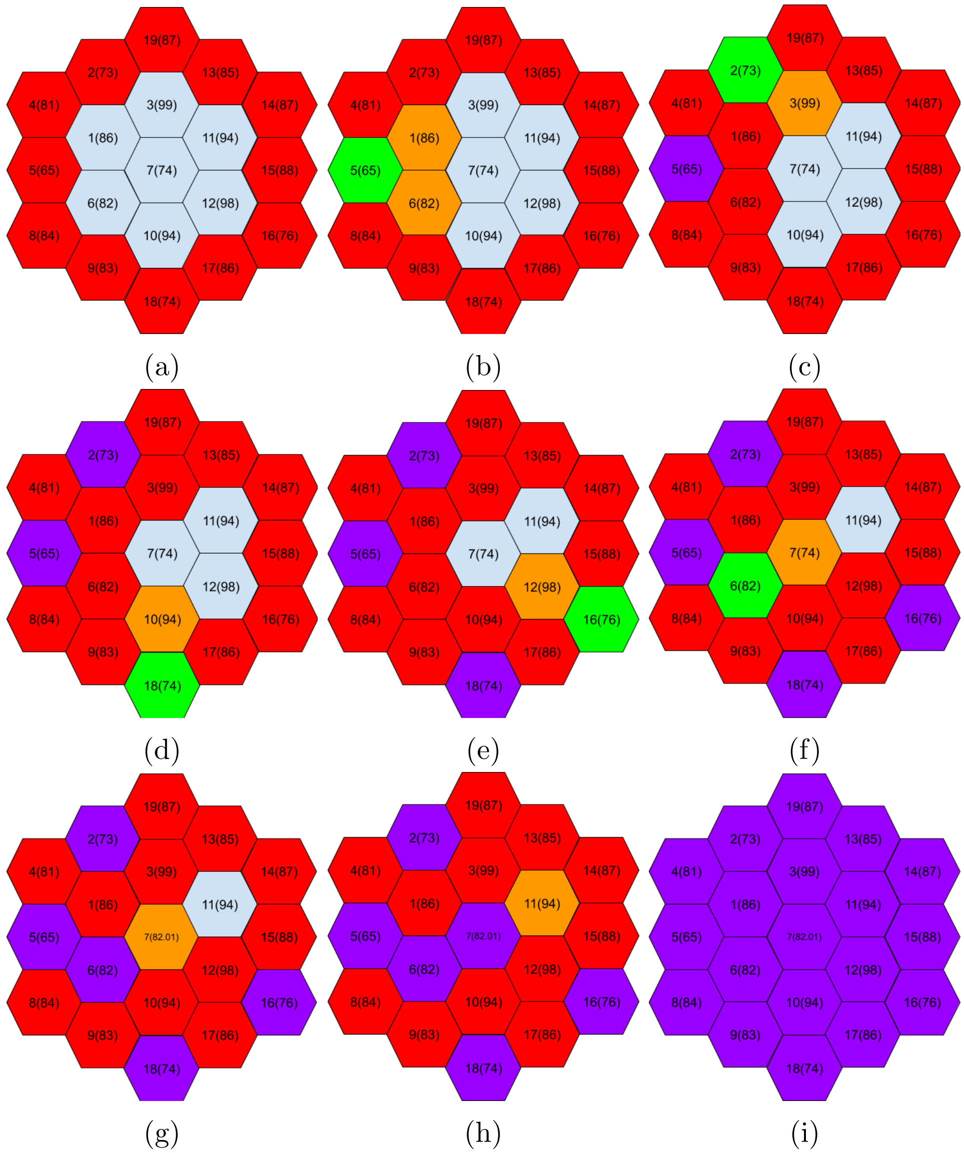


Figure 5: Illustration of the priority-flood depression filling for the HGSD method. Light blue grids represent the initial default state; red grids represent the boundary; green grids represent the to-be-removed grid from the queue; orange grids represent the to-be-added grids into the queue; and purple grids are finished grids. Numbers within each grid represent its global ID and elevation (in parentheses, unit: m), respectively. The algorithm gradually "floods" the domain using a boundary queue (red). If a to-be-added grid has equal or smaller elevation than a to-be-removed grid, its elevation is increased. For example, the elevation of grid #7 is increased from 74 to 82.01 in (g). The step-by-step instructions are provided in Appendix B.

2.6. Flow Direction

Flow direction is defined from the hexagon center to the center of neighbor hexagon which has the lowest elevation. In other words, flow direction is the flow path which has the steepest slope. The global ID of this downslope neighbor is stored in an attribute table. Unlike the traditional SGSD method that uses indices (1, 2, 4, 8, etc.) to represent flow direction, HexWatershed currently represents the flow direction using a flow routing map.

2.7. Flow Accumulation

We developed a flow accumulation algorithm based on the concept from ArcGIS flow accumulation [29]. In short, this algorithm scans all the hexagon grids and sums up the accumulations once all the accumulations of upslope hexagons are calculated. It runs recursively until accumulations of all hexagons are calculated (Figure C.17).

The flow accumulation algorithm also provides the option to consider variable resolution hexagonal mesh.

2.8. Watershed Boundary

The hexagon grid that has the highest flow accumulation is defined as the watershed outlet. This algorithm scans all the hexagon grids and identifies all the hexagons that contribute to this outlet using the flow routing map. Among them, those at the edge with less than 6 neighbors are used to define the watershed boundary.

2.9. Stream Grid

A hexagon grid is defined as a stream grid if its total drainage area exceeds the minimal drainage area threshold. For a constant resolution hexagonal mesh, each grid's drainage area is proportional to its accumulation value. For a variable resolution hexagonal mesh, each grid's drainage area is summarized from its upslope grids plus its own area.

In HexWatershed, a stream grid is also named a "stream reach", which makes up a stream segment (Section 2.11 and Figure 6).

2.10. Stream Confluence

Stream confluences are defined based on the flow routing map and stream grids. In short, if a hexagon grid is a stream grid and it has multiple upslope stream reaches, it is defined as a stream confluence. In rare scenarios, a stream confluence may have three or more upslope stream reaches.

2.11. Stream Networks

In HexWatershed, a stream segment is defined as the stream component between headwater/outlet and confluence or between two confluences.

To maintain the ascending order from upstream to downstream, we developed an algorithm to define the stream segments reversely from watershed outlet to headwaters.

Starting from the watershed outlet, the algorithm searches for stream confluence following the stream grids. Once a stream confluence is found, all of its upslope stream reaches are identified, and the algorithm continues to search recursively until all stream segments are identified. This algorithm works in the following steps:

1. Calculate the total number (N) of stream segments based on stream confluences information;
2. Set current outlet and current segment index as the watershed outlet and N, respectively;
3. Starting from the current outlet, search upstream and assign the current segment index to each stream reach;
4. If a confluence is found, set the current outlet and current segment index as this confluence and $N = N - 1$, respectively;
5. Repeat step 3 and loop through all the upstream segments of this confluence;
6. Stop until all confluences and segments are treated (N=1).

At the end of this algorithm, a stream segment is made up of a list of stream reaches which have the same segment index. The topology information of these stream reaches within each stream segment is also defined (Figure 6).

2.12. Stream Topology

Stream topology is defined based on the stream reaches information. If the first stream reach of a stream segment does not have a positive or valid upstream segment index, this stream reach is headwater (Reach #1 in Figure 6). Similarly, the segment index (Segment #6) of the downstream of the last stream reach (Reach #3) of a stream segment (Segment #5) is a downstream of Segment #5. Details of stream topology are explained in Figure 6.

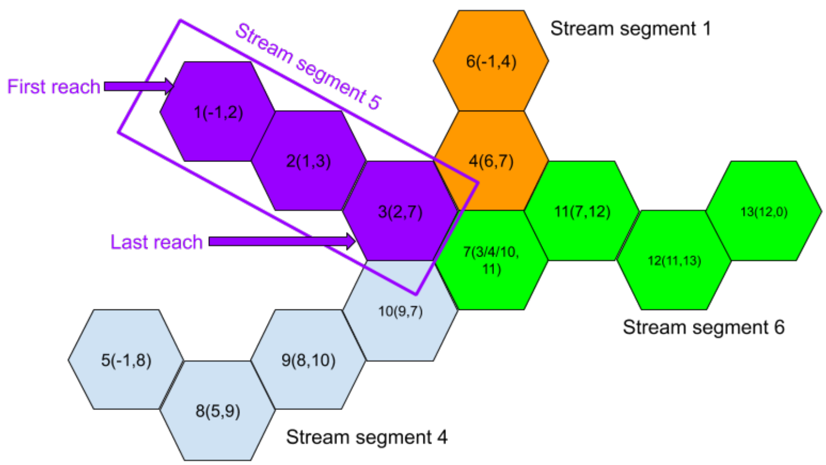


Figure 6: Illustration of the stream topology. Different colors represent 4 different stream segments, respectively. Each stream segment is made up by several stream reaches. For example, the purple stream segment #5 is represented by 3 stream reaches. Numbers with each stream reach represent global ID, upslope and downslope grids global IDs, respectively. Grid #1 and #3 are the first and last reaches of stream segment #5, respectively. Grid #1 does not have upslope stream reach grid and is a headwater. Grid #7 is a confluence and it receives inflow from grid #3, #4 and #7. Therefore stream segment #5 and #6 have a upstream-downstream topology relationship.

A stream segment always has only one downstream segment, but it may have multiple upstream segments unless it is headwater. Besides, both stream segment indices and stream reach indices within the same segment have ascending orders.

2.13. Stream Order

Stream order is defined following the classical stream order definition [29]. First the stream order of all the headwater stream segments are defined as 1. Then the stream orders of remaining stream segments are defined based on stream topology.

2.14. Subbasin Boundary

Similar to stream networks, subbasins are defined reversely. The algorithm works as follows:

1. Set the last stream segment and watershed outlet as the current stream segment index N and outlet, respectively;

2. Scan all the grids that contribute to the current outlet based on the flow routing map, set their subbasin indices to N;
3. Go to stream segment $N = N - 1$, repeat step 2;
4. Stop until all stream segments are treated (N=1).

2.15. Software Requirements and File I/O

HexWatershed was written in C++11 with OpenMP enabled. It can be applied to both regional and global scales. It is platform independent and parallel computing ready for high performance computing (HPC) [1].

To run the HexWatershed model, the minimal software requirements include:

1. GNU Compiler Collection (GCC) 4.9 and above; and
2. Geospatial Data Abstraction Library (GDAL 2.3).

The required model inputs include: (a) a high resolution traditional DEM raster file; and (b) a corresponding hexagonal mesh file, which can be generated by any mesh generator. In our study, we used the QGIS MMQGIS plugin. The input data must be prepared with the same spatial reference/map projection.

After a successful model simulation, HexWatershed produces a list of products including flow direction and stream networks. These products have the same spatial reference as the input data.

Because currently a standard file format for HGSD datasets is unavailable, all the model inputs and outputs are stored using the ESRI Shapefile format [6].

3. Application

3.1. Study Area

To test the performance of HexWatershed model, we applied the model to two different types of watersheds in the western US. Specifically, a mountainous area watershed and a flat area watershed are used to demonstrate the capability of HexWatershed model. Then we analyzed the model outputs.

The Tin Pan (TP) watershed is located near the northern border of New Mexico. This is a mountainous area watershed with relatively high average surface slope (Figure 7).

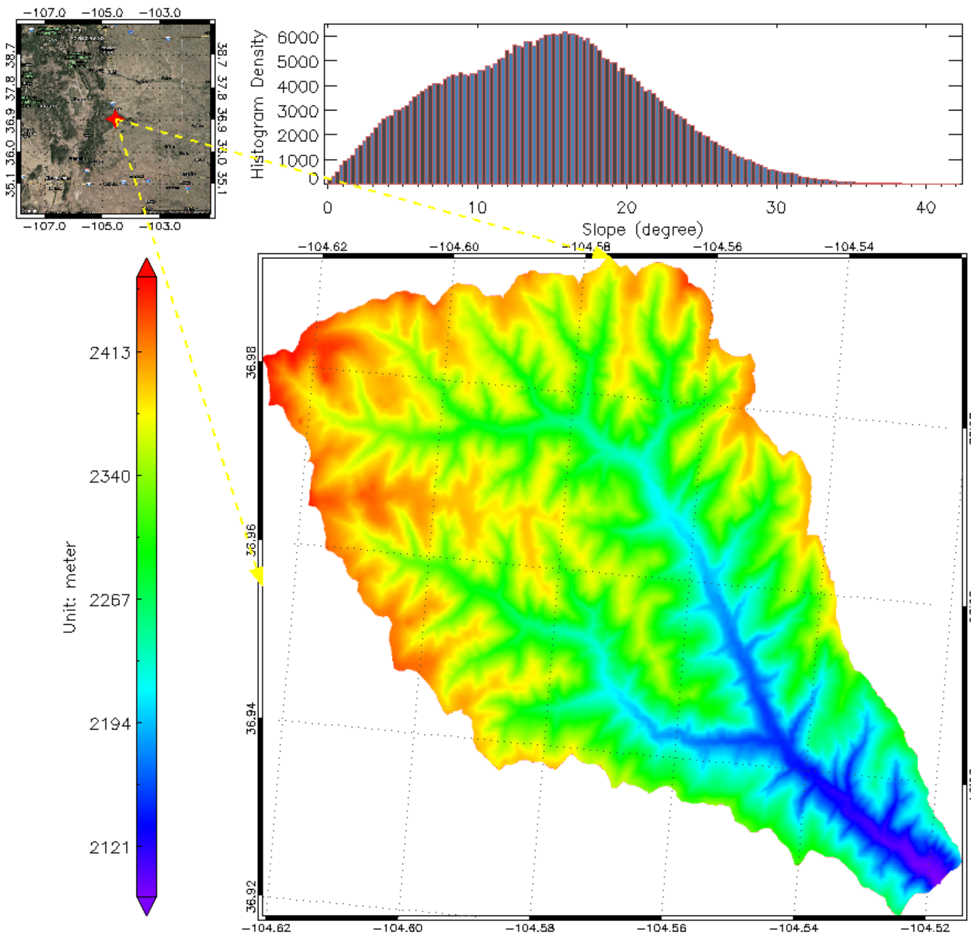


Figure 7: The spatial location, surface elevation and slope distribution of the Tin Pan watershed. Upper left is the location on Google Map; Upper right is the histogram of surface slope (degree); and bottom is the spatial distribution of surface elevation (m).

The Columbia Basin flat (CBF) watershed is located near the Columbia River, Washington. This is a flat area watershed with relatively low average surface slope (Figure 8).

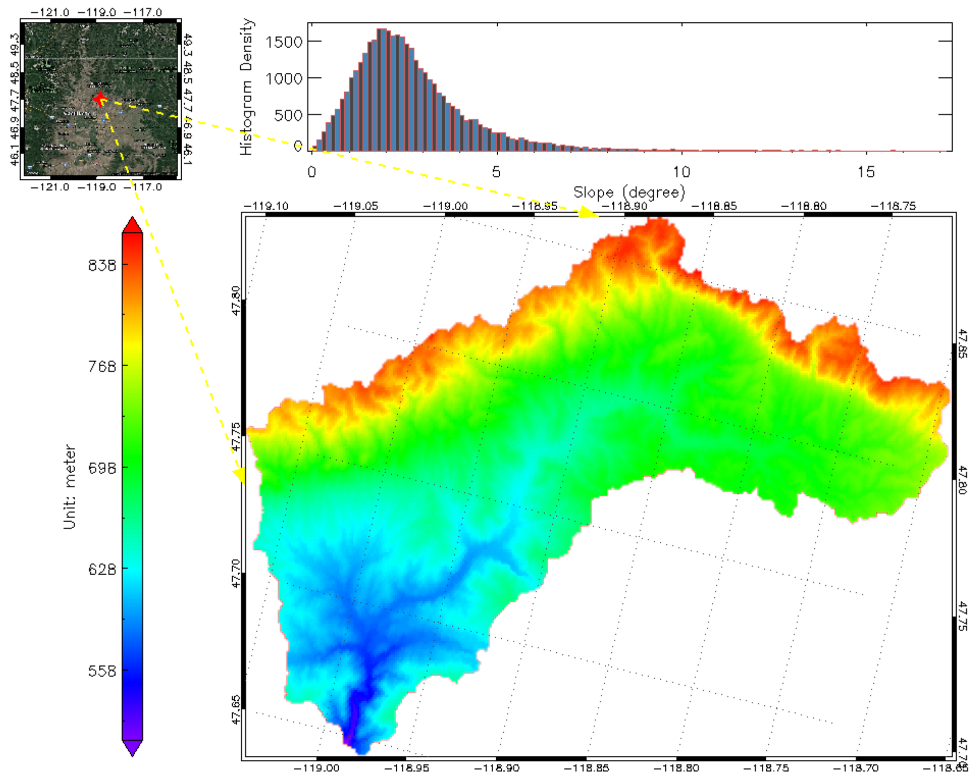


Figure 8: The spatial location, surface elevation and slope distribution of the CBF watershed. Upper left is the location on Google Map; Upper right is the histogram of surface slope (degree); and bottom is the spatial distribution of surface elevation (m).

Characteristics of the two watersheds are listed in Table 1.

Table 1: Characteristics of Tin Pan and Columbia Basin flat watersheds.

	TP	**CBF**
Location (lon, lat)	(-104.52, 36.94)	(-118.82, 47.74)
Elevation range (m)	2091 to 2457	510 to 859
Average slope (degree)	14.66	2.66
Total drainage (km^2)	42	308

3.2. Model Setup

First, we collected raster DEM for both Tin Pan (10 m resolution) and Columbia Basin flat (90 m resolution) watersheds. Then, we generated the hexagonal mesh files for Tin Pan (30 m resolution) and Columbia Basin flat (90 m resolution) watersheds. Last, we ran the HexWatershed model for both watersheds.

3.3. Results

Because of the unique structure, we mainly use visualization to present the model outputs. To provide a clear view of the data structure, we provide zoom-in views of the whole datasets. And the full views of these datasets are provided in Appendix D.

3.3.1. Tin Pan

Zoom-in views (upper left) of model outputs in the Tin Pan watershed are illustrated in Figure 9.

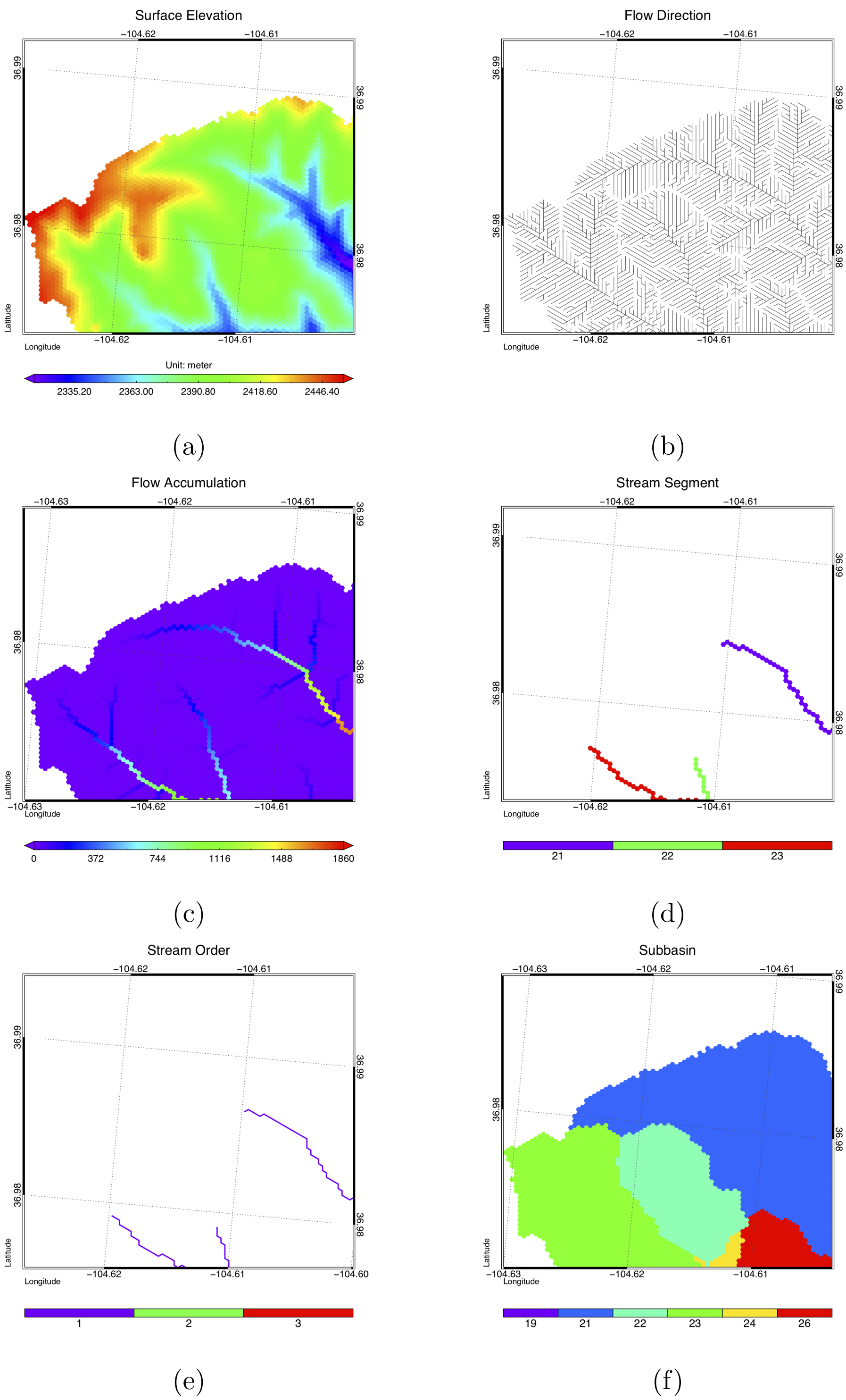


Figure 9: The zoom-in spatial distributions of model outputs in the Tin Pan watershed. (a) is the depression filled hexagonal DEM; (b) is the flow routing; (c) is the flow accumulation; (d) is the stream segments with indices; (e) is the stream order; and (f) is the subbasin with indices.

These results show that HexWatershed is able to produce all the traditional watershed delineation characteristics.

3.3.2. Columbia Basin Flat

Zoom-in views (lower left) of model outputs in the Columbia Basin flat watershed are illustrated in Figure 10.

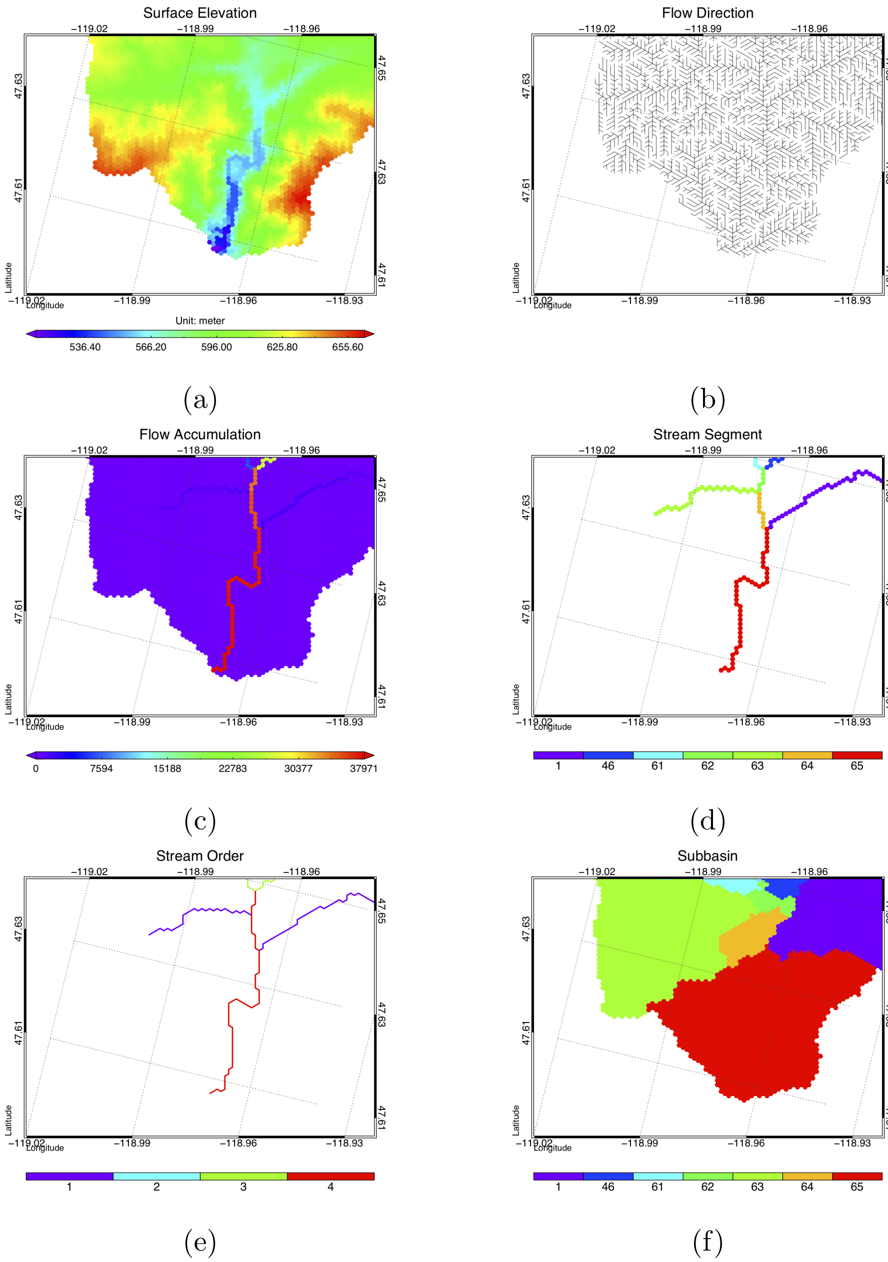


Figure 10: The zoom-in spatial distributions of model outputs in the Columbia Basin flat watershed. (a) is the depression filled hexagonal DEM; (b) is the flow routing; (c) is the flow accumulation; (d) is the stream segments with indices; (e) is the stream order; and (f) is the subbasin with indices.

Simulation results from the Columbia basin flat watershed demonstrate that HexWatershed is robust in basins with flat terrain.

4. Comparisons

To evaluate the performance of the HexWatershed model, we compared the model outputs against outputs from the traditional SGSD method.

To produce the traditional watershed delineation characteristics, we used the ArcSWAT watershed delineation tool. ArcSWAT is an ArcGIS extension for the Soil & Water Assessment Tool (SWAT), which is widely used in basin scale hydrologic simulations [20, 36].

Although HexWatershed is able to produce all the watershed delineation characteristics, we only compared characteristics that are commonly used in hydrologic simulations. Because HexWatershed is robust in different watersheds, we mainly provide comparison in the Tin Pan watershed and show only selected results from the Columbia Basin flat watershed. To evaluate the sensitivity of HexWatershed to mesh resolution, we ran an addition simulation at Tin Pan watershed using a $100\,m$ resolution hexagonal mesh.

4.1. Hexagonal DEM

The depression filled hexagonal DEM (Figure 9a) has the same spatial pattern as the traditional DEM (Figure 7), and it fits the land surface reasonably well.

4.2. Flow Direction

Because HexWatershed does not use indices (1, 2, 4, etc.) to represent flow direction, we cannot compare its flow direction against ArcSWAT flow direction output directly.

4.3. Flow Accumulation

The spatial patterns of flow accumulation from ArcSWAT and HexWatershed are similar. However, their spatial distributions are different. For example, ArcSWAT produces more grids with accumulation value at 1, 2 and 3 whereas HexWatershed produces more between 4 and 10 (Figure 11). Because of this, HexWatershed produces less spatial variability in flow accumulation.

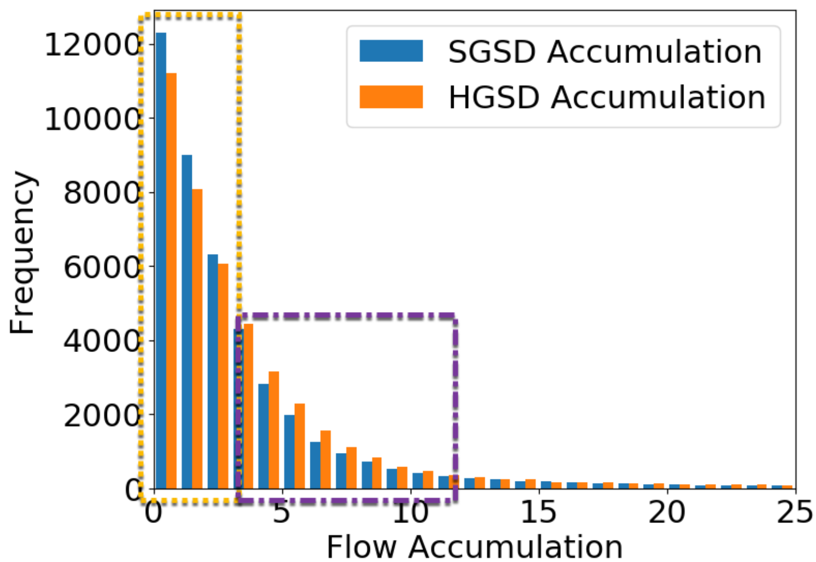


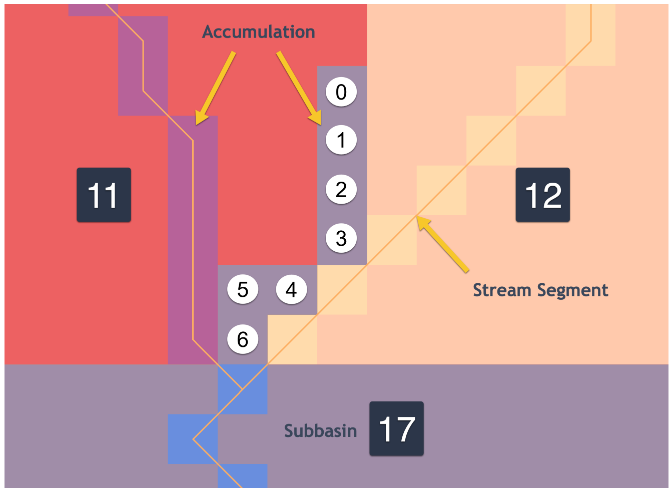
Figure 11: Histograms of flow accumulation from ArcSWAT (SGSD) and HexWatershed (HGSD), respectively. Yellow rectangle features where ArcSWAT produces more grids with accumulation at 1, 2 and 3. Purple rectangle features where HexWatershed produces more grids with accumulation between 4 and 10.

4.4. Subbasin Boundary

ArcSWAT produces diagonal travel path at the subbasin interfaces. For example, grids with circles are defined within Subbasin #17 when grids with accumulation 0, 1, 2, and 3 should be in either Subbasin #11 or #12 (Figure 12a).

Meanwhile, HexWatershed is able to eliminate the diagonal travel path, and the corresponding hexagon grids are clearly defined within Subbasin #2 [2] (Figure 12b).

[2]Because ArcSWAT and HexWatershed use different index systems, subbasin indices are different.



(a)

(b)

Figure 12: Comparison of flow accumulation and subbasin boundary between model outputs from ArcSWAT and HexWatershed at the same location. (a) are the outputs from ArcSWAT. Square grids with indices represent subbasin. Circles with indices represent flow accumulation. Line features represent stream networks. Because circles from 0 to 6 are connected through diagonal flow path, they are defined in Subbasin #17. (b) are model outputs from HexWatershed. Each hexagon is labelled with its accumulation. Yellow lines between hexagon grids are flow paths.

4.5. Stream Networks

Stream networks produced by HexWatershed are very close to the ArcSWAT produced stream networks. To further compare the differences, we calculated the enclosed area of differences between modelled stream networks and the National Hydrography Dataset (NHD) flowline datasets. We treat the NHD flowline as the "true" flow path. In theory, the smaller the total enclosed area is, the closer the stream networks are to the NHD flowline. To test the robustness, we compared model outputs at different spatial resolutions. (Figure 13 and Table 2).

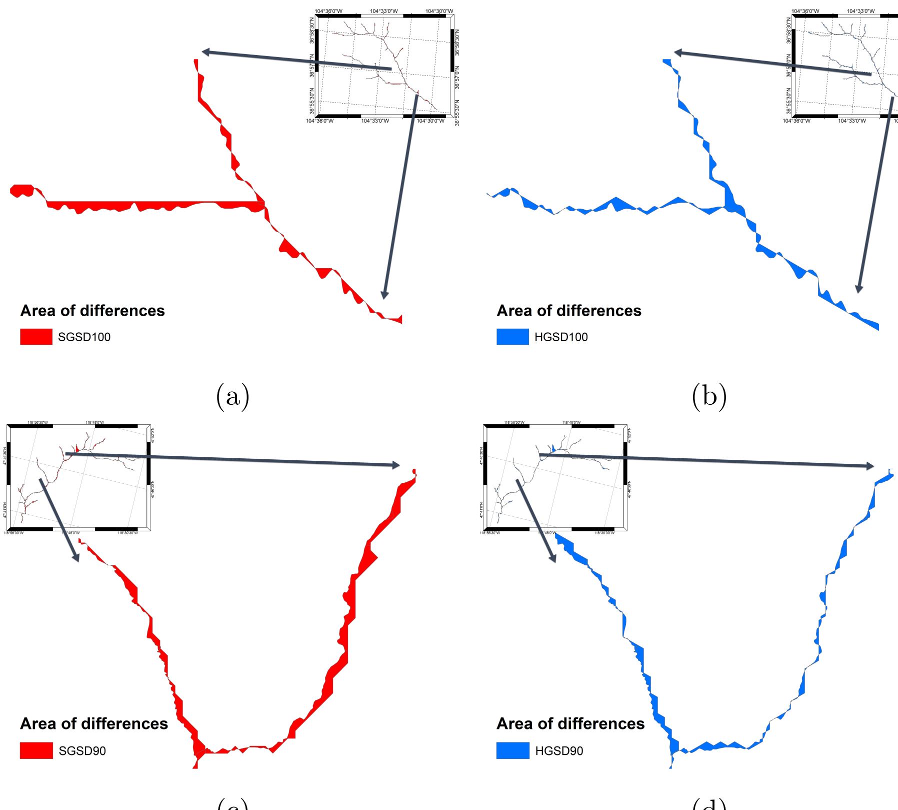


Figure 13: Comparisons of stream networks through enclosed area of differences. These polygons were generated by connecting the NHD flowline with the modelled stream networks. (a) and (b) are areas of differences for ArcSWAT and HexWatershed at 100m resolution in the Tin Pan watershed, respectively. (c) and (d) are areas of differences for ArcSWAT and HexWatershed at 90m resolution in the Columbia Basin flat watershed, respectively.

Statistics of areas of differences due to changing spatial resolutions are provided in Table 2. These statistics show that HexWatershed performs much better at coarse resolutions.

5. Discussion

Based on model outputs and comparisons, HexWatershed can provide equivalent and potentially even better performance than the traditional method. Because of the close relationship between watershed delineation and surface hydrology, most hydrologic processes will be affected. Because we didn't present results on a sphere, our discussion will focus on watershed scale only.

First, HexWatershed has successfully eliminated the "island" effect and tedious manual corrections are no longer needed. Because both subbasin boundary and watershed boundary are improved, stream discharge in hydrologic simulations will be improved (Figure 14).

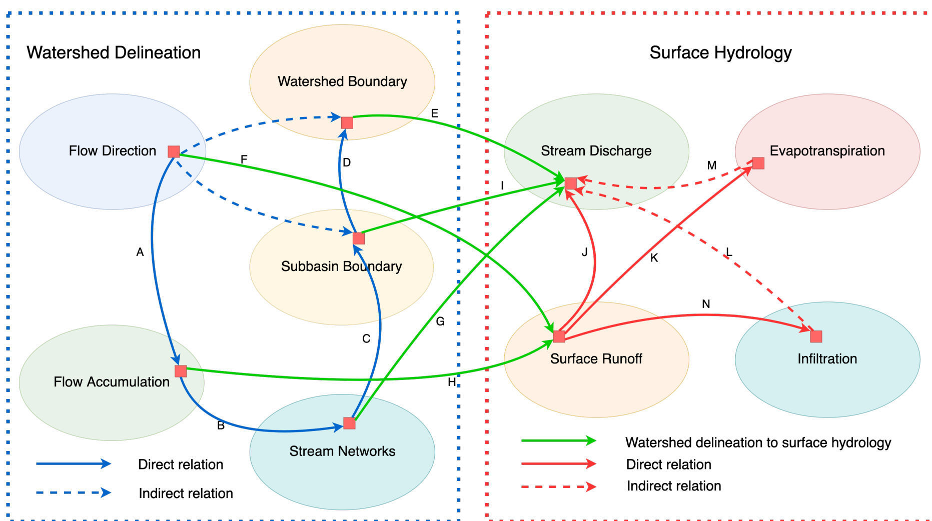


Figure 14: The relationship between watershed delineation and surface hydrology. Circles are major watershed delineation characteristics (left) and hydrologic processes (right). Blue lines show the relationships between watershed characteristics, red lines show the relationships between hydrologic processes, and green lines show the relationships between watershed characteristics and hydrologic processes. The arrows of the lines indicate the direction of influence.

Second, although we didn't compare flow direction directly, which essentially determines flow accumulation, and subsequently stream networks,

Table 2: Statistics of areas of differences of stream networks for both 30m and 100m spatial resolutions (m^2).

Watershed	**Method with resolution**	**Mean**	**Maximum**	**Sum**	**Standard Deviation**
Tin Pan	SGSD30	554.85	8387.66	225270.04	891.68
	HGSD30	821.74	14328.59	296649.35	1457.44
	SGSD100	3027.82	53887.28	593453.52	4961.07
	HGSD100	2779.15	17059.82	589181.20	2959.10
CBF	SGSD90	25024.08	422748.00	4679503.03	50613.98
	HGSD90	10855.88	504569	3528163.97	33022.46

subbasin boundary, it is potentially improved because the latter ones are improved. For example, due to the uniform connectivity, flow accumulation is smoother in spatial transitions (Figure 9c and 11). Consequently, surface runoff, evapotranspiration, infiltration and stream discharge will be improved (Figure 14).

Last, comparisons of stream networks suggest that spatial resolution has an impact on model performance. Our analysis shows that HexWatershed performs much better at coarse resolutions.

Taken together, HexWatershed should be applied to hydrologic models to improve hydrologic simulations.

6. Limitations

Based on our analysis, we have identified a few limitations in this study:

1. Currently a standard file format to manage HGSD based datasets is unavailable. In this study, we rely on ESRI Shapefile for storage and visualization. However, due to the limitations of Shapefile, we can use NetCDF or HDF to improve performance. For example, NetCDF is the file format currently used by MPAS mesh grid.
2. The HexWatershed model relies on the accuracy of DEM resampling. There are challenges in converting SGSD based raster DEM to HGSD based DEM. Currently we use the nearest resampling method because it will not introduce new values into the system and it's computationally efficient. Other advanced resampling methods should be used in the future [13].
3. Our model currently only considers the steepest slope as the single flow direction. In some scenarios, multiple flow directions should be considered. And we can implement the D-infinity algorithm on the HGSD method [12].
4. We didn't implement the stream "burn-in"capability in the current version, which could further improve the performance under certain circumstances [11].
5. Currently there is not a tool that can be used to convert existing SGSD based datasets to the HGSD based format. But similar function is already available in many coupled Earth system models. In the future, we plan to provide a tool to resolve this limitation.

6. Currently we only ran HexWatershed model at $30\,m$, $90\,m$ and $100\,m$ spatial resolutions. More simulations at different resolutions are needed to evaluate the model sensitivity to spatial resolution.

7. Conclusion

We have developed a watershed delineation model (HexWatershed) using the hexagon grid spatial discretization method. We have applied this model to two different types of watersheds in the western US featuring steep and flat terrain. Model outputs have shown that HexWatershed can reproduce all the watershed delineation characteristics.

Comparisons between outputs from HexWatershed and the traditional square grid spatial discretization method have shown that the HGSD method has multiple advantages including removal of "island" effect and improvement of flow direction and all watershed characteristics such as subbasin boundary that depend on the flow direction because of the consistent connectivity. Analysis also suggests that spatially distributed hydrologic simulations which rely on connectivity/routing can be improved if the HGSD method is used.

Our model can be applied to continental or global scale to improve large scale hydrologic simulations.

8. Acknowledgement

The research described in this paper was primarily funded by the Laboratory Directed Research and Development (LDRD) Program Quickstarter project 206583 at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy. LRL was supported by U.S. Department of Energy Office of Science Biological and Environmental Research through the Earth and Environmental System Modeling program as part of the Energy Exascale Earth System Model (E3SM) 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 Github HexWatershed. Some datasets in the Tin Pan watershed were produced from an effort of hydrology-based design of geomorphic evapotranspiration covers for reclamation of mine land. 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.

9. References

[1] Richard Barnes. Parallel non-divergent flow accumulation for trillion cell digital elevation models on desktops or clusters. *Environmental Modelling & Software*, 92:202–212, 2017.

[2] Richard Barnes, Clarence Lehman, and David Mulla. Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, 62:117–127, 2014.

[3] Colin P D Birch, Sander P Oom, and Jonathan A Beecham. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological modelling*, 206(3-4):347–359, 2007.

[4] Luis de Sousa, Fernanda Nery, Ricardo Sousa, and Joao Matos. Assessing the accuracy of hexagonal versus square tilled grids in preserving DEM surface flow directions. In *7th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, pages 191–200, 2006.

[5] Qiang Du, Max D Gunzburger, and Lili Ju. Voronoi-based finite volume methods, optimal Voronoi meshes, and PDEs on the sphere. *Computer methods in applied mechanics and engineering*, 192(35-36):3933–3957, 2003.

[6] ESRI and White PaperdJuly. ESRI shapefile technical description. *Comput. Stat*, 16:370–371, 1998.

[7] Antonio Francipane, Valeriy Y Ivanov, Leonardo V Noto, Erkan Istanbulluoglu, Elisa Arnone, and Rafael L Bras. tRIBS-Erosion: A parsimonious physically-based model for studying catchment hydro-geomorphic response. *Catena*, 92:216–231, 2012.

[8] Dean Gesch, Michael Oimoen, Susan Greenlee, Charles Nelson, Michael Steuck, and Dean Tyler. The national elevation dataset. *Photogrammetric engineering and remote sensing*, 68(1):5–32, 2002.

[9] Brian Hamilton and Alberto Torin. Finite Difference Schemes on Hexagonal Grids for Thin Linear Plates with Finite Volume Boundaries. In *DAFx*, pages 5–12. Citeseer, 2014.

[10] Arlen 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.

[11] Ferdi Hellweger and 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)*, 1997.

[12] Wesley R Henson, Rose L Medina, C Justin Mayers, Richard G Niswonger, and R Steven Regan. *CRT–Cascade Routing Tool to define and visualize flow paths for grid-based watershed models.* US Department of the Interior, US Geological Survey, 2013.

[13] Innchyn Her and Chi-Tseng Yuan. Resampling on a pseudohexagonal grid. *CVGIP: Graphical Models and Image Processing*, 56(4):336–347, 1994.

[14] Craig M Johnston, Thomas G Dewald, Timothy R Bondelid, Bruce B Worstell, Lucinda D McKay, Alan Rea, Richard B Moore, and Jonathan L Goodall. Evaluation of catchment delineation methods for the medium-resolution national hydrography dataset. Technical report, 2009.

[15] Christian D Langevin, Joseph D Hughes, Edward R Banta, Richard G Niswonger, Sorab Panday, and Alden M Provost. Documentation for the MODFLOW 6 groundwater flow model. Technical report, 2017.

[16] Hongyi Li, Mark S Wigmosta, Huan Wu, Maoyi Huang, Yinghai Ke, Andr M Coleman, and L Ruby Leung. A physically based runoff routing model for land surface and earth system models. *Journal of Hydrometeorology*, 14(3):808–828, 2013.

[17] Chang Liao, Qianlai Zhuang, L Ruby Leung, and Laodong Guo. Quantifying Dissolved Organic Carbon Dynamics Using a ThreeDimensional Terrestrial Ecosystem Model at High SpatialTemporal Resolutions. *Journal of Advances in Modeling Earth Systems.*

[18] Steven L Markstrom, Richard G Niswonger, R Steven Regan, David E Prudic, and Paul M Barlow. *GSFLOW, Coupled Ground-Water and Surface-Water Flow Model Based on the Integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005)*. US Department of the Interior, US Geological Survey, 2008.

[19] Reed M Maxwell, Stefan J Kollet, Steven G Smith, Carol S Woodward, Robert D Falgout, Ian M Ferguson, Chuck Baldwin, William J Bosl, Richard Hornung, and Steven Ashby. ParFlow users manual. *International Ground Water Modeling Center Report GWMI*, 1(2009):129, 2009.

[20] S Neitsch, J Arnold, J Kiniry, J R Williams, and K W King. SWAT2009 Theoretical Documentation. *Texas Water Ressources Institute Technical Report*, (406), 2005.

[21] Olivier Planchon and Frdric Darboux. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, 46(2-3):159–176, 2002.

[22] Yizhong Qu. *An integrated hydrologic model for multi-process simulation using semi-discrete finite volume approach*. The Pennsylvania State University, 2005.

[23] David A Randall, Todd D Ringler, Ross P Heikes, Phil Jones, and John Baumgardner. Climate modeling with spherical geodesic grids. *Computing in Science & Engineering*, 4(5):32–41, 2002.

[24] H Renssen and J M Knoop. A global river routing network for use in hydrological modeling. *Journal of Hydrology*, 230(3-4):230–243, 2000.

[25] K Sahr. DGGRID version 6.2 b: User documentation for discrete global grid software, 2015.

[26] Kevin Sahr. Hexagonal discrete global grid systems for geospatial computing. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, 22:363–376, 2011.

[27] Kevin Sahr. Central Place Indexing: Hierarchical Linear Indexing Systems for Mixed-Aperture Hexagonal Discrete Global Grid Systems. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 54(1):16–29, 2019.

[28] David G Tarboton. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water resources research*, 33(2):309–319, 1997.

[29] David G Tarboton, Rafael L Bras, and Ignacio RodriguezIturbe. On the extraction of channel networks from digital elevation data. *Hydrological processes*, 5(1):81–100, 1991.

[30] Teklu K Tesfa, H-Y Li, L R Leung, Maoyi Huang, Yinghai Ke, Yu Sun, and Ying Liu. A subbasin-based framework to represent land surface processes in an Earth system model. *Geoscientific Model Development*, 7(3):947–963, 2014.

[31] Teklu K Tesfa, L Ruby Leung, Maoyi Huang, HongYi Li, Nathalie Voisin, and Mark S Wigmosta. Scalability of gridand subbasinbased land surface modeling approaches for hydrologic simulations. *Journal of Geophysical Research: Atmospheres*, 119(6):3166–3184, 2014.

[32] Lei Wang and Hongxing Liu. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science*, 20(2):193–213, 2006.

[33] Lu Wang and Tinghua Ai. The comparison of drainage network extraction between square and hexagonal grid-based dem. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(4), 2018.

[34] Hilary Weller, John Thuburn, and Colin J Cotter. Computational modes and grid imprinting on five quasi-uniform spherical C grids. *Monthly Weather Review*, 140(8):2734–2755, 2012.

[35] Hilary Weller, Henry G Weller, and Aim Fournier. Voronoi, Delaunay, and block-structured mesh refinement for solution of the shallow-water equations on the sphere. *Monthly Weather Review*, 137(12):4208–4224, 2009.

[36] M Winchell, R Srinivasan, M Di Luzio, and J Arnold. ArcSWAT interface for SWAT 2005. *Users Guide, Blackland Research Center, Texas Agricultural Experiment Station, Temple*, 2007.

Appendix A. Model Structure

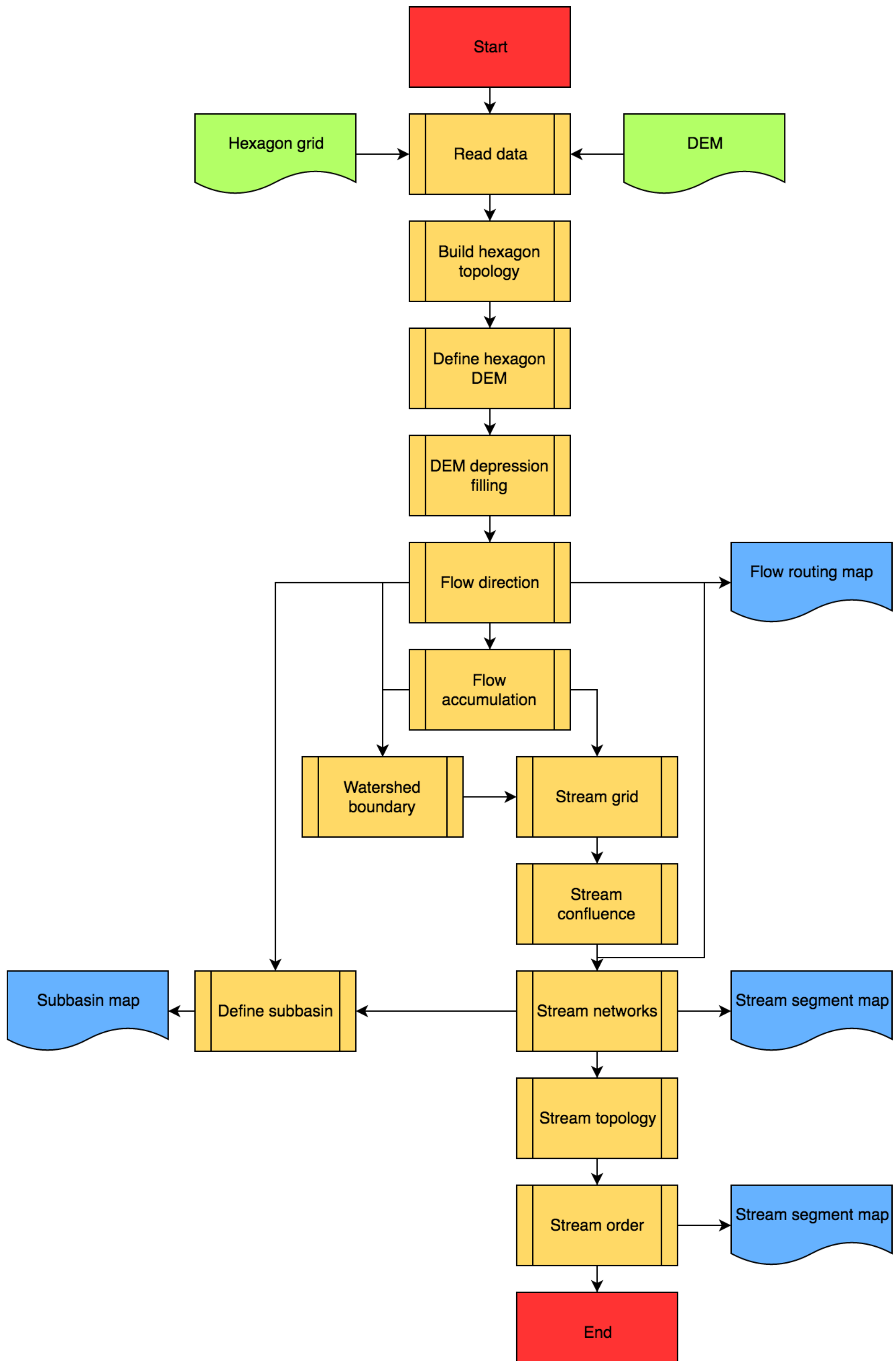


Figure A.15: The work flow of the HexWatershed model. The red tiles are start and end. The yellow tiles are processing steps. The green and blue tiles are major model inputs and outputs, respectively.

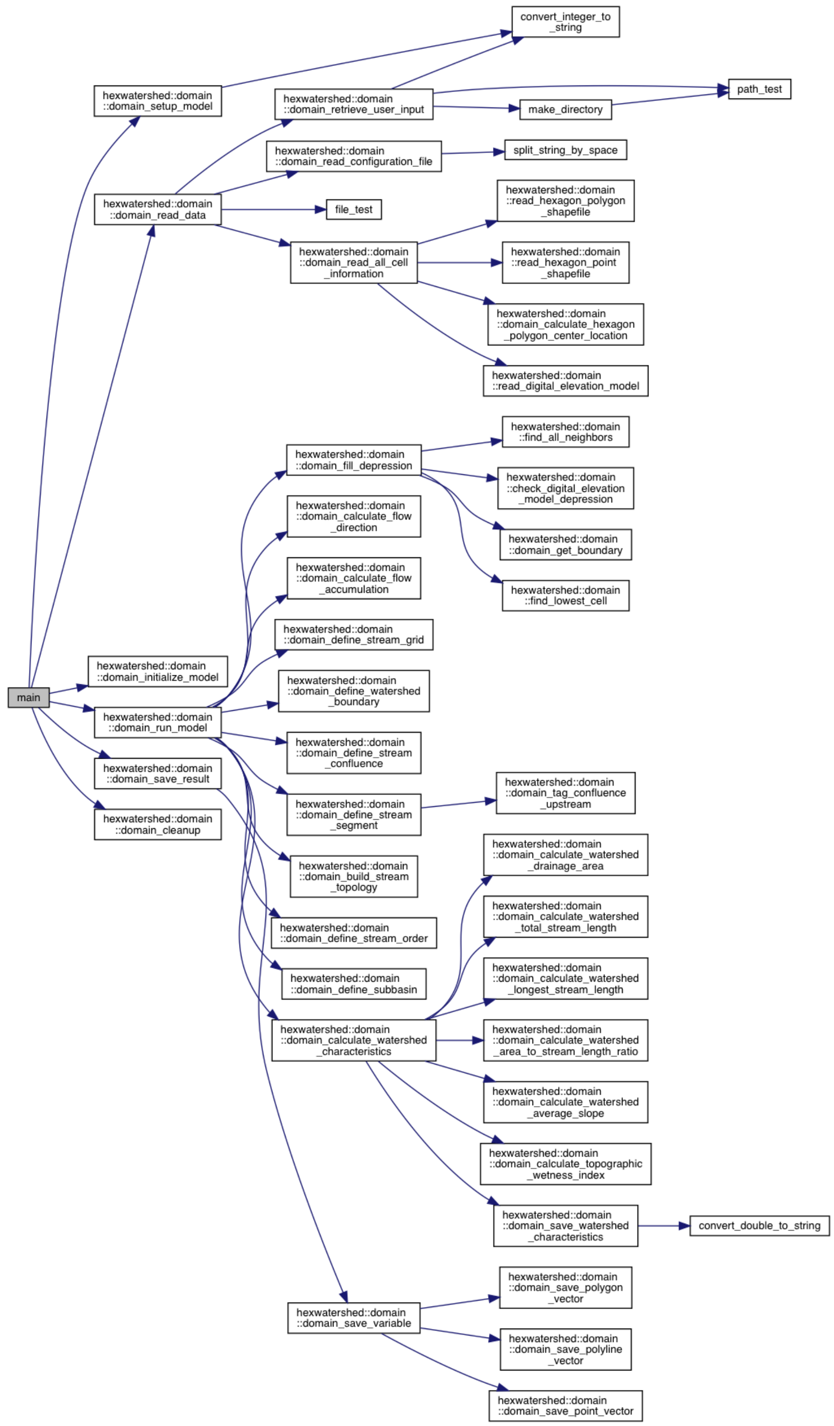


Figure A.16: The structure of HexWatershed model generated by Doxygen.

Appendix B. Depression Filling

1. Find the boundary of the grid system and push them into a queue Q;
2. Find the grid A which has the lowest elevation in Q;
3. Find all the untreated neighbors of grid A and put them into array B;
4. If any member of B has a lower elevation than A, increase its elevation to higher than A's;
5. Push B into Q and remove A from Q;
6. If there are still untreated grids, go to step 2.

Appendix C. Flow Accumulation

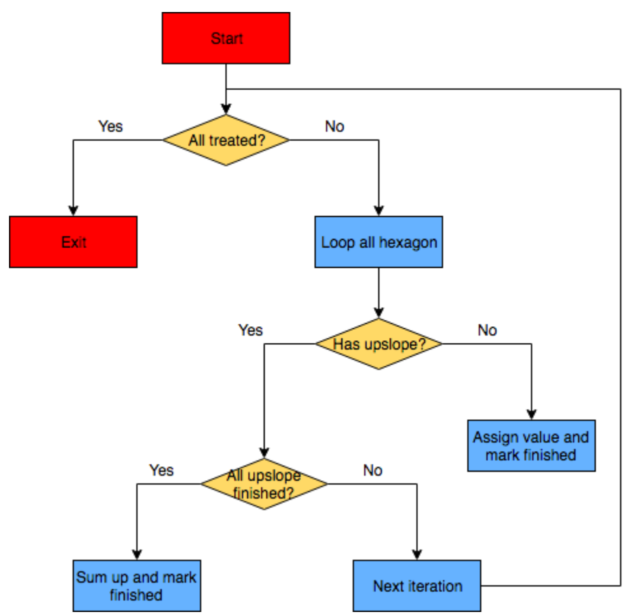


Figure C.17: Illustration of the flow accumulation algorithm. In a hexagon grid system, this algorithm loops through grids using the global IDs and calculates accumulation once its upslope accumulations are finished.

Appendix D. Model Results

Appendix D.1. Tin Pan

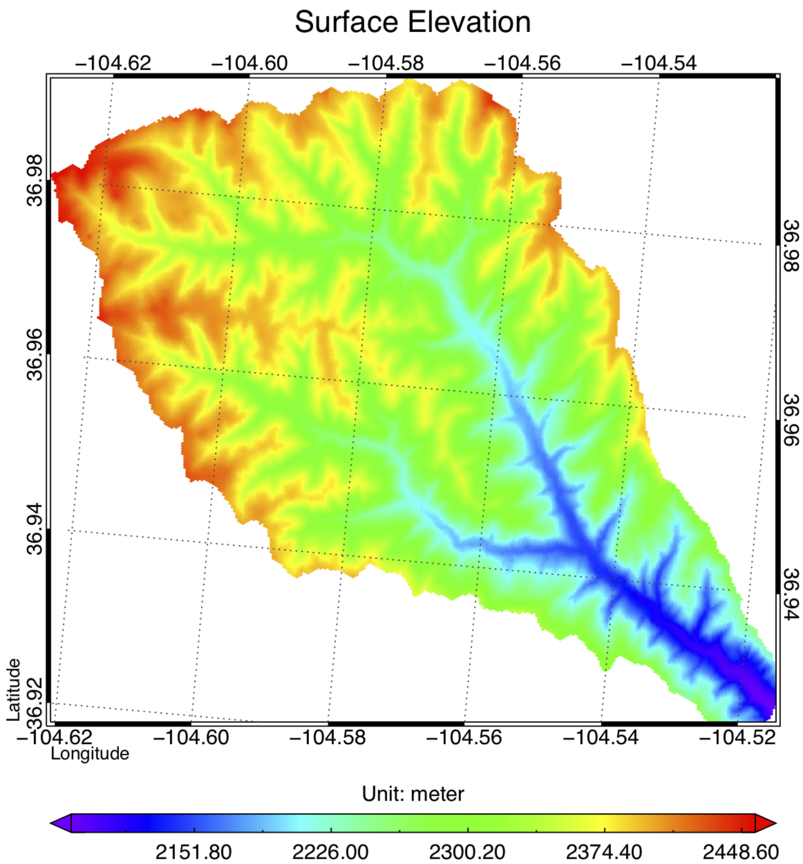


Figure D.18: The digital elevation model using the HGSD method (m).

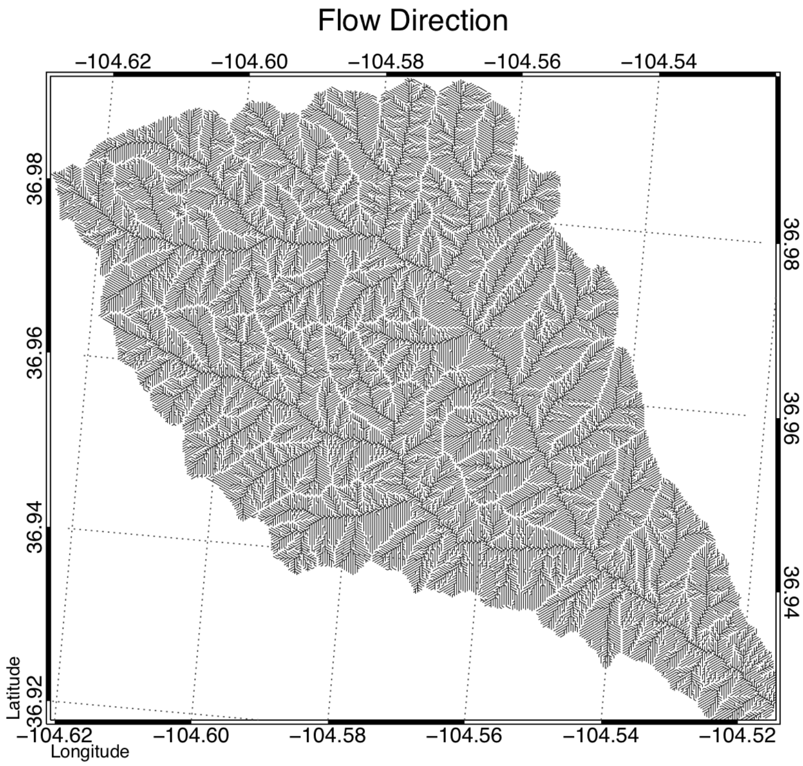


Figure D.19: The spatial distribution of simulated flow direction.

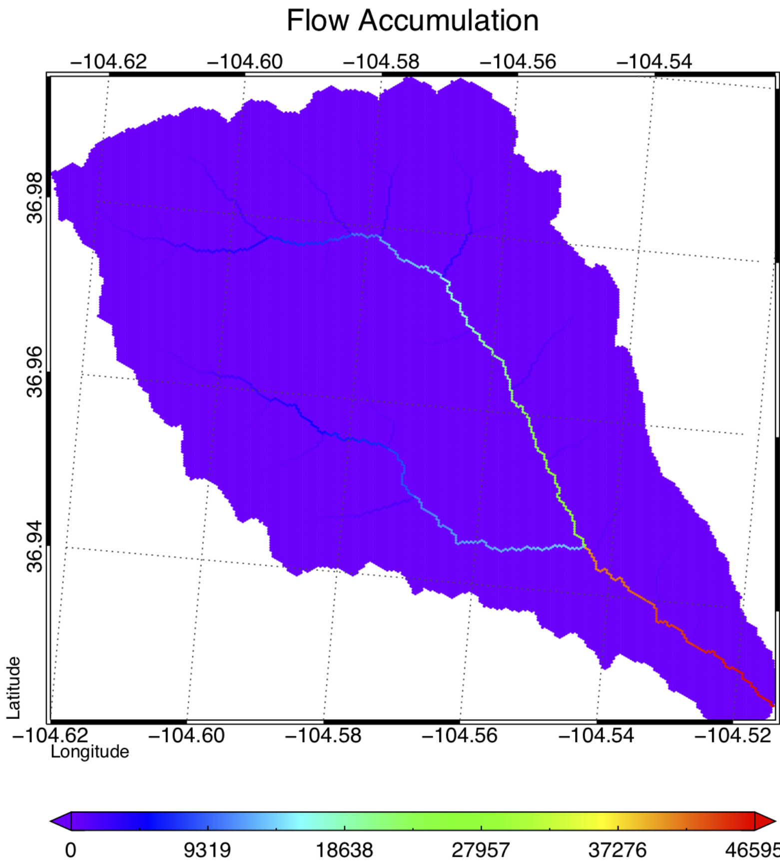


Figure D.20: The spatial distribution of simulated flow accumulation.

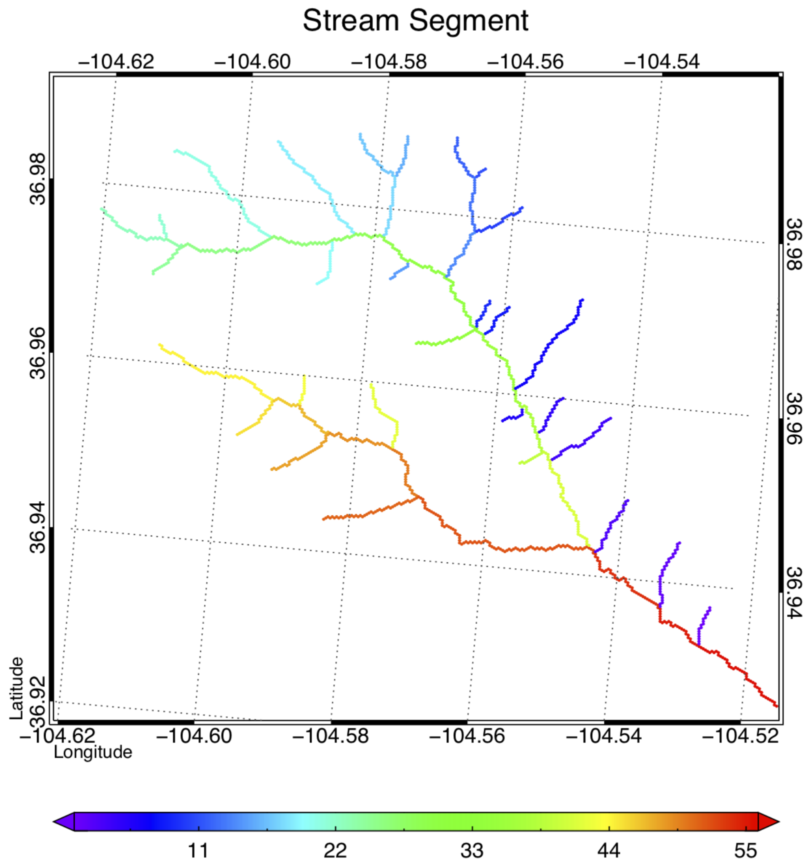


Figure D.21: The spatial distribution of simulated stream networks.

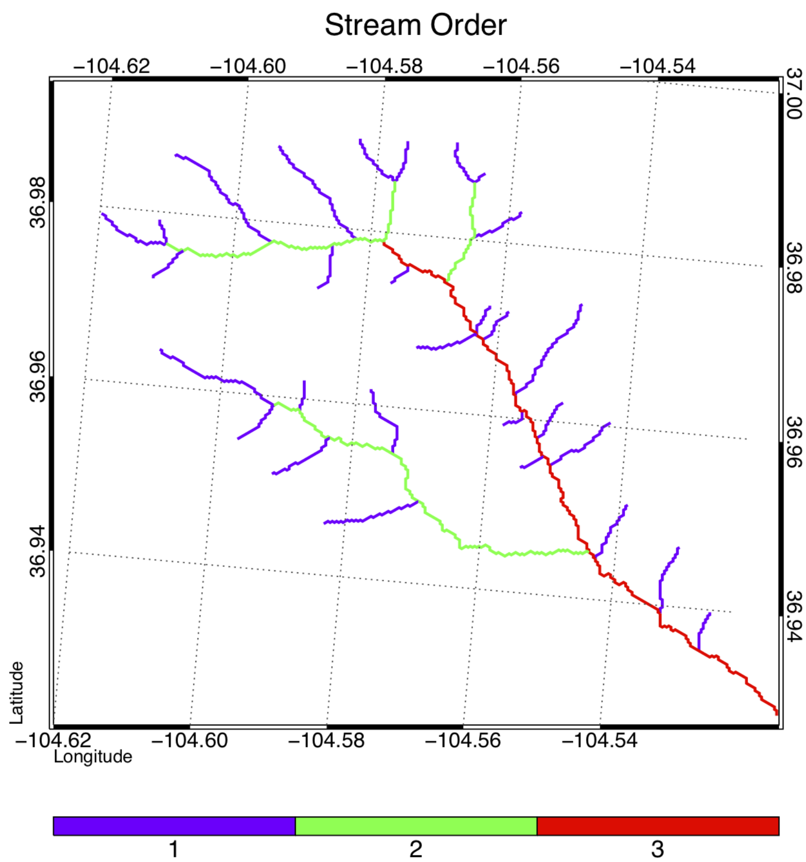


Figure D.22: The spatial distribution of simulated stream order.

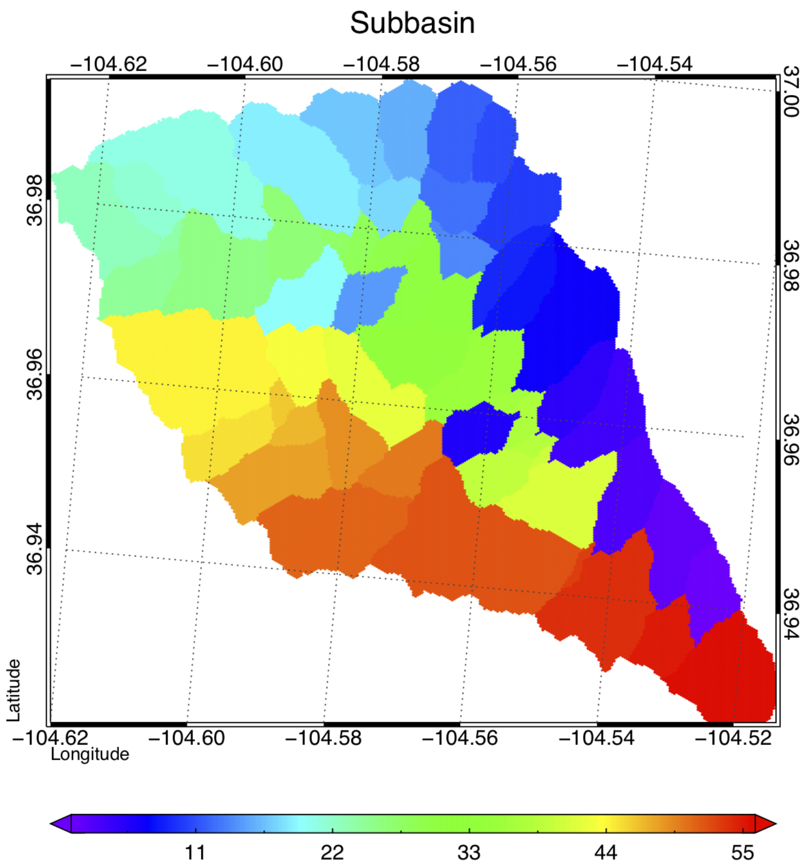


Figure D.23: The spatial distribution of simulated subbasin boundary. The colored polygons represent hexagons in the same subbasin.

Appendix D.2. Columbia Basin Flat

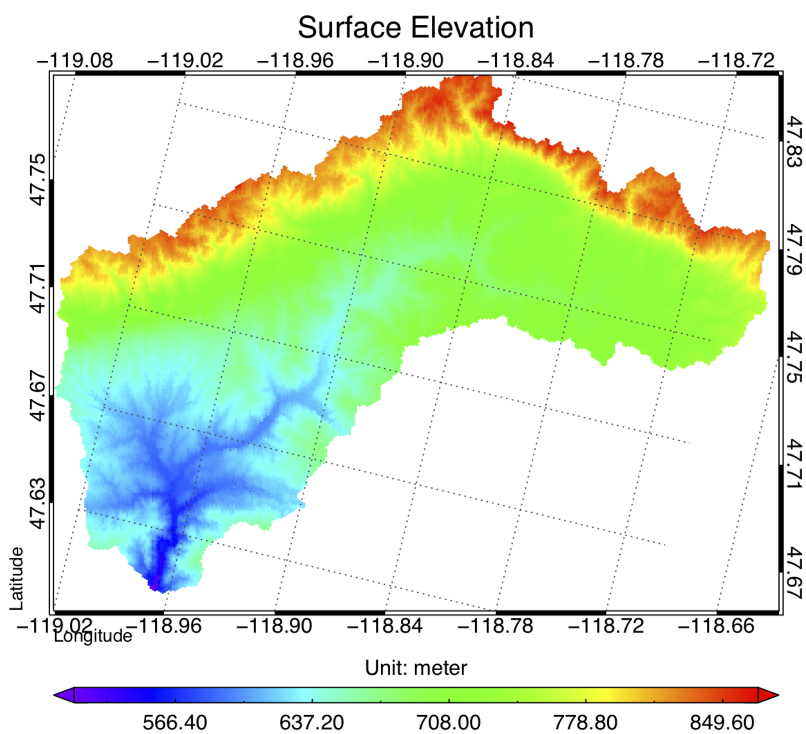


Figure D.24: The digital elevation model using the HGSD method (m).

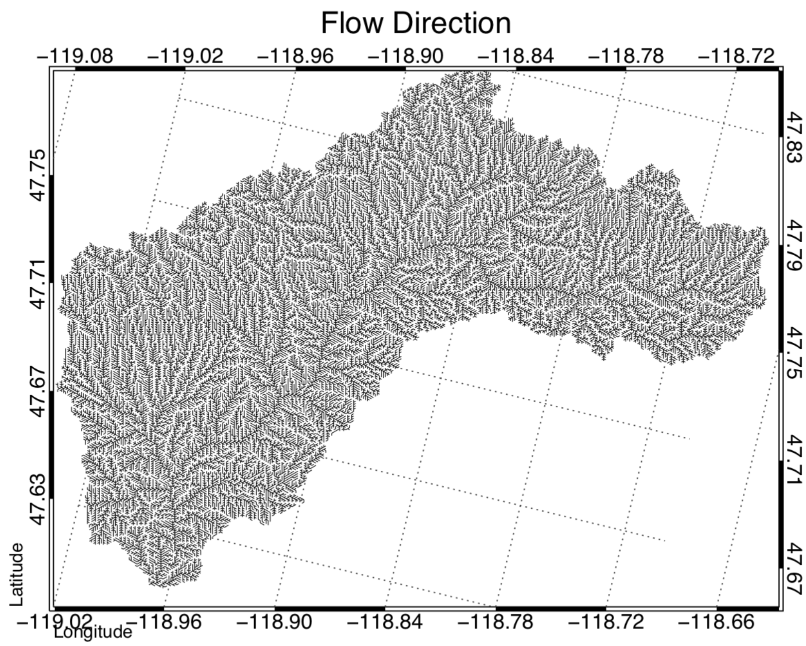


Figure D.25: The spatial distribution of simulated flow direction.

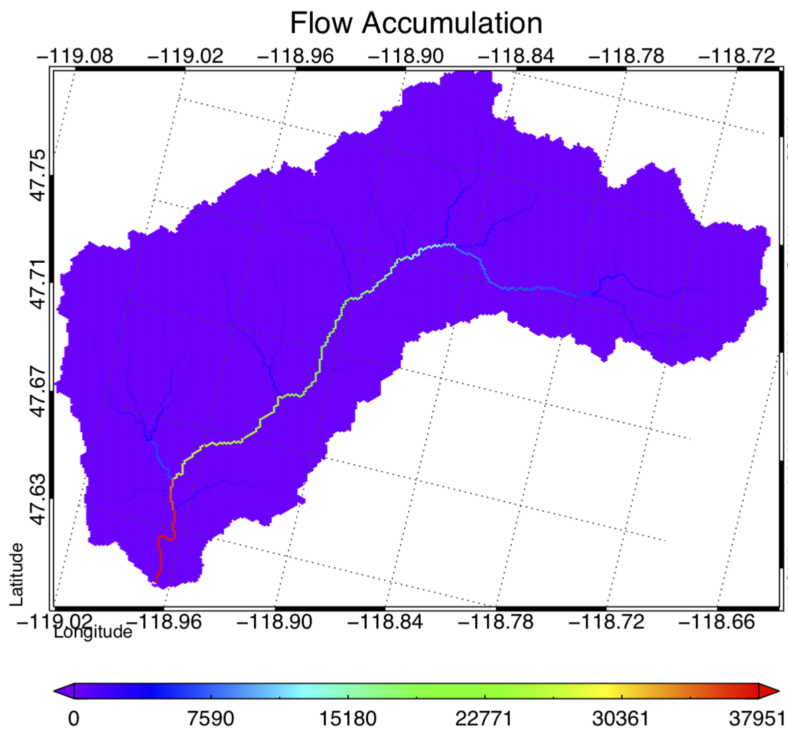


Figure D.26: The spatial distribution of simulated flow accumulation.

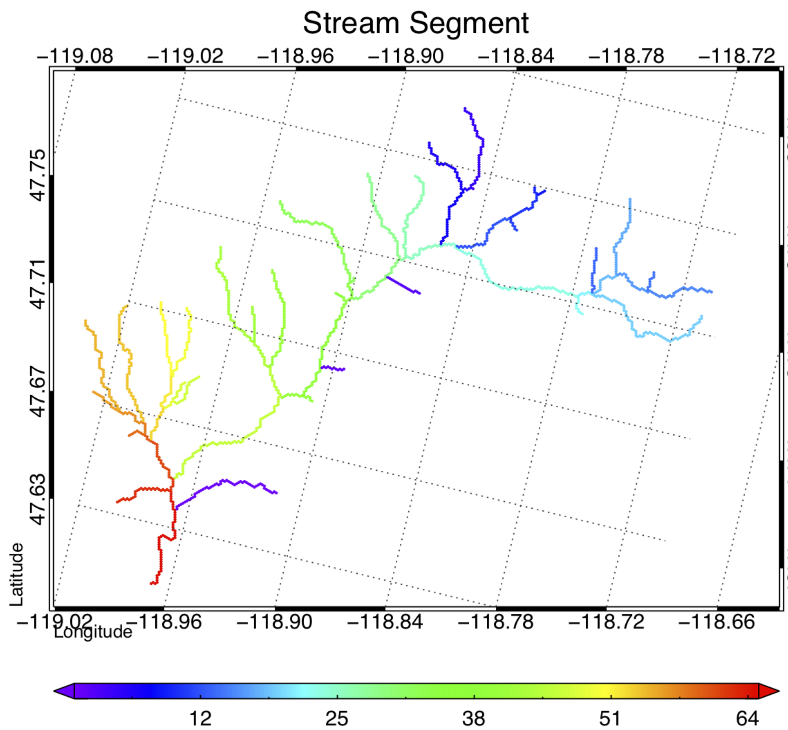


Figure D.27: The spatial distribution of simulated stream networks.

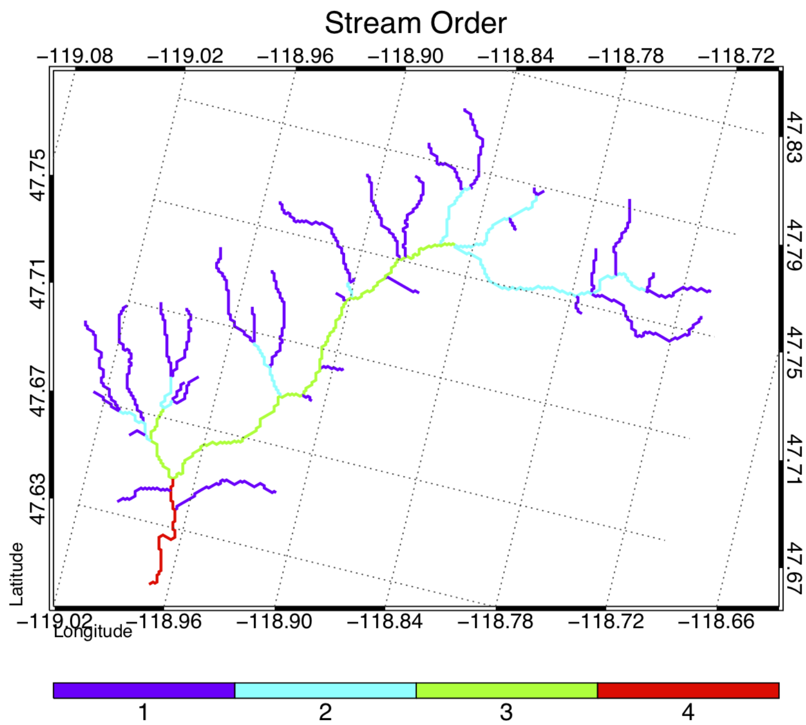


Figure D.28: The spatial distribution of simulated stream order.

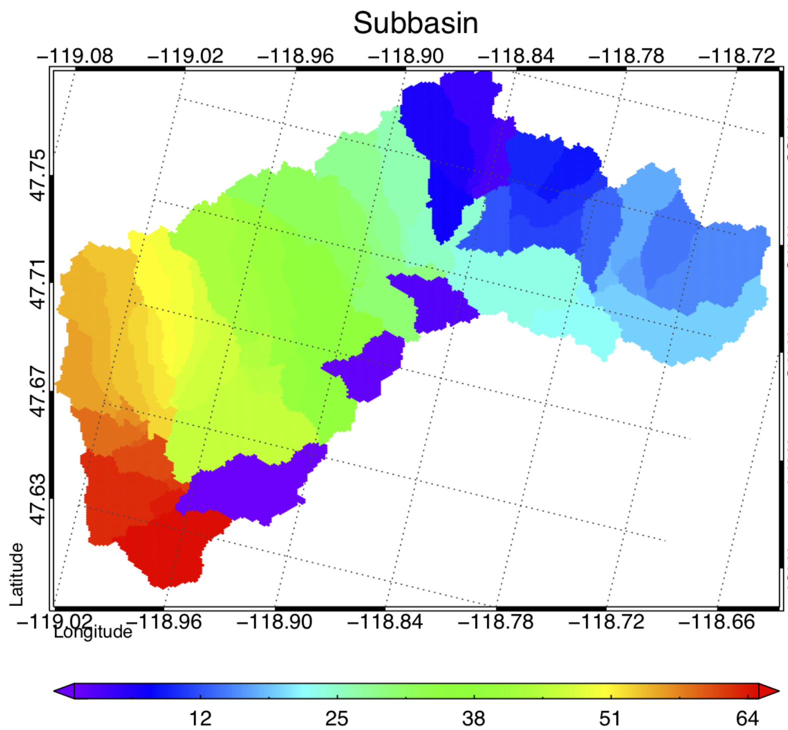


Figure D.29: The spatial distribution of simulated subbasin boundary. The colored polygons represent hexagons in the same subbasin.