

## Article

# Technical Note: A Drainage Density-Preserving River Network Delineation Algorithm

Heng Yang <sup>1</sup>, Shuanglong Chen <sup>2</sup>, and Hui Zheng <sup>3,\*</sup>

<sup>1</sup> Science and Technology Research Institute, China Three Gorges Corporation, Beijing 100038, China; yang\_heng2@ctg.com.cn

<sup>2</sup> Baihetan Hydropower Plant, China Three Gorges Corporation, Liangshan 615421, China; chen\_shuanglong@ctg.com.cn

<sup>3</sup> Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; hzheng\_iap@outlook.com

\* Correspondence: hzheng\_iap@outlook.com

**Abstract:** Accurate representation of drainage density in river network delineation is crucial for large-domain hydrodynamic simulations, but existing digital elevation model (DEM)-based methods fail to preserve observed drainage density across the entire network. This study introduces a fundamentally improved DEM-based river network delineation method that preserves observed drainage density by incorporating a novel concept of upstream accumulation length, which integrates flow direction and drainage density. A test case demonstrated the method's compatibility with widely used critical drainage area-based methods, producing identical results when using the same drainage density. The method was then applied to the Chinese Mainland using the MERIT-Hydro flow direction dataset and a drainage density dataset from the Third National Land Resources Survey of China. The resulting dataset shows superior performance in capturing drainage density variations, particularly in regions with complex topography, compared to existing datasets. The method is promising for delineating more accurate river networks by combining satellite-derived or surveyed drainage density with DEMs, thereby laying a valuable foundation for large-domain hydrodynamic simulations in vast regions of the globe that lack manually maintained river geometry datasets.

**Keywords:** river network delineation; digital elevation model; drainage density; river geomorphology

Received:

Revised:

Accepted:

Published:

**Citation:** Yang, H.; Chen, S.; Zheng, H.

Technical Note: A Drainage

Density-Preserving River Network

Delineation Algorithm. *Water* **2025**, *1*, 0.

<https://doi.org/>

**Copyright:** © 2025 by the authors.

Submitted to *Water* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hydrodynamic simulations on prescribed river network centerlines [1] are valuable for conducting hydrodynamic simulations at the scale of a country, continent, or the globe [2]. These so-called vector-based simulations capture the longitudinal water flow along rivers while significantly simplifying the dynamic changes in river width. This simplification is justified by the fact that the predominant flow direction of water in rivers is longitudinal. Compared with raster-based simulations, which model hydrodynamic processes on two-dimensional grids, vector-based simulations are computationally efficient. This efficiency makes vector-based simulations ideal for large-domain operational flood forecasting, such as the National Water Model of the United States [3]. They strike a good balance between the timeliness and computational demand of operational forecasts, especially when parameter calibration and data assimilation are involved.

The accuracy of the river network centerlines is essential for these vector-based simulations. In a few regions, such as the United States, high-accuracy river network datasets

are available for vector-based hydrodynamic simulations. These datasets are developed and maintained through extensive field surveys and manual delineation. However, in most regions of the world, such manually maintained datasets do not exist.

In the vast regions of the globe where manually maintained river network datasets are lacking, there are primarily two approaches to delineate river networks over large domains: digital elevation model (DEM)-based methods [4–8] and satellite or aerial imagery-based methods [9]. Imagery-based methods are well-suited for identifying the presence of rivers and often offer higher spatial resolution than DEM-based methods [9]. Imagery-derived river networks can serve as the basis for calculating drainage density. However, river networks extracted from imagery lack information on drainage directions and connectivity between river reaches, rendering them unsuitable for hydrodynamic simulations. On the other hand, DEM-based methods provide detailed information on flow direction and preserve the critical upstream-downstream relationships necessary for hydrodynamic simulations [4–6]. However, river networks delineated from DEMs are often not representative of the real network, leading to inaccuracies in various characteristics of the river network.

Among the various characteristics of river networks, drainage density is particularly significant [10]. Drainage density is defined as the total length of river segments per unit area [11,12] and serves as a key indicator of a region's hydrological characteristics [13]. It varies significantly under different hydroclimatic conditions, with higher values typically found in wetter regions and lower values in arid areas. In regions with high drainage density, water is drained more efficiently from the landscape, leading to more rapid hydrologic responses and potentially higher flood risks [14,15]. This characteristic also influences sediment transport, as more developed river networks can more effectively carry sediment [13].

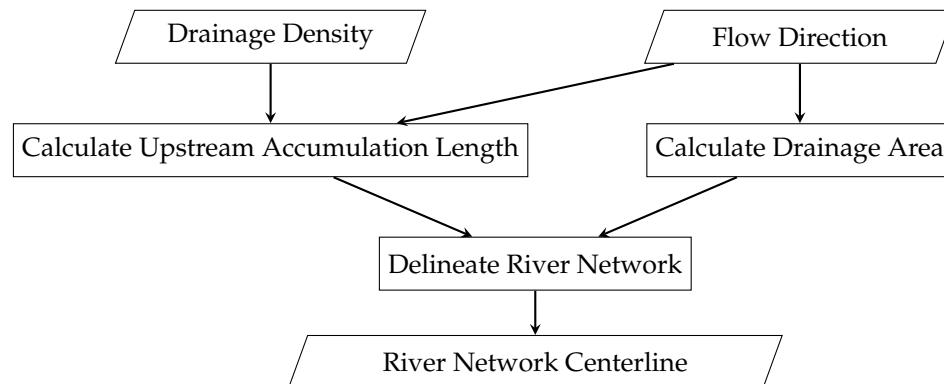
The difficulty in accurately representing drainage density is a major limitation for large-domain simulations, as drainage density naturally varies substantially across different regions. Several studies have attempted to address this limitation. Existing DEM-based methods utilize a parameter known as the critical drainage area to delineate river networks [4]. This parameter is defined as the minimum area required to sustain a river segment. The smaller the critical drainage area, the more river segments are delineated, and consequently, the higher the drainage density. Schneider et al. [16] developed a statistical model to estimate the critical drainage area using slope, lithology, and climate. This model was calibrated to match drainage densities from reference river networks in France and Australia and was subsequently applied globally to delineate river networks from DEMs. Another approach was proposed by Lin et al. [5], who introduced a method to trim river segments from networks delineated using a small critical drainage area threshold ( $1 \text{ km}^2$ ). The river with the smallest drainage area is trimmed first, and this process is repeated until the drainage density within a watershed matches a value derived from machine learning. However, all existing methods are performed in an ad hoc manner. Drainage density is not incorporated into the delineation process itself but is only used in the evaluation or post-processing step. As a result, none of these existing studies can effectively preserve the observed drainage density across the entire river network.

We take a fundamentally different approach to address the challenge of DEM-based delineation. A new river network delineation algorithm is proposed, using two-dimensional drainage density as a direct input. The algorithm is based on a novel concept of upstream accumulation length rather than the critical drainage area. Upstream accumulation length is determined by both flow direction and drainage density. Given a fixed flow direction, upstream accumulation length is mathematically equivalent to drainage density. Our proposed algorithm leverages the upstream accumulation length to preserve the observed drainage density across the entire river network.

This technical note is organized as follows. Section 2 details the proposed algorithm. Section 3 sets up two test cases for the algorithm: one is a synthetic case, and the other is a real-world application. Section 4 presents the results of the two cases. Section 5 discusses the limitations of the algorithm. Finally, Section 6 concludes the paper.

## 2. The Drainage Density-Preserving River Network Delineation Algorithm

As shown in Figure 1, the algorithm consists of three main steps: (1) calculation of upstream accumulation length using flow direction and drainage density, (2) calculation of drainage area using flow direction, and (3) river network delineation using flow direction, drainage area, and upstream accumulation length. While extensive research has been conducted on the second step (i.e., the calculation of drainage area) [17], we will focus on the first and third steps in this section.



**Figure 1.** Flowchart illustrating the steps of the density-preserving river network delineation algorithm, including calculation of upstream accumulation length, drainage area, and final river network delineation.

### 2.1. Upstream Accumulation Length

In this section, we first introduce the definition of upstream accumulation length. Then, we show how to calculate the upstream accumulation length using drainage density. Third, we prove that, given the flow direction, the upstream accumulation length is mathematically equivalent to drainage area by establishing the formula for inverting drainage density from upstream accumulation length. Finally, we present a fast algorithm to calculate the upstream accumulation length with  $O(N)$  time complexity.

As illustrated in Figure 2, the upstream accumulation length for a grid cell (Point O) is defined as the total length of river segments (i.e., segments A, B, and C) that lie within the catchment of the grid cell. This can be expressed mathematically as:

$$L = \int_{l \in \text{Catchment}} dl, \quad (1)$$

where  $L$  represents the upstream accumulation length for the grid cell, Catchment denotes the catchment area of the grid cell, and  $l$  is the length of the river segments within the catchment.

Equation 1 can be rewritten using the definition of drainage density ( $D = \frac{dl}{dA}$ ) as follows:

$$L = \int_{A \in \text{Catchment}} \frac{dl}{dA} dA = \int_{A \in \text{Catchment}} D dA, \quad (2)$$

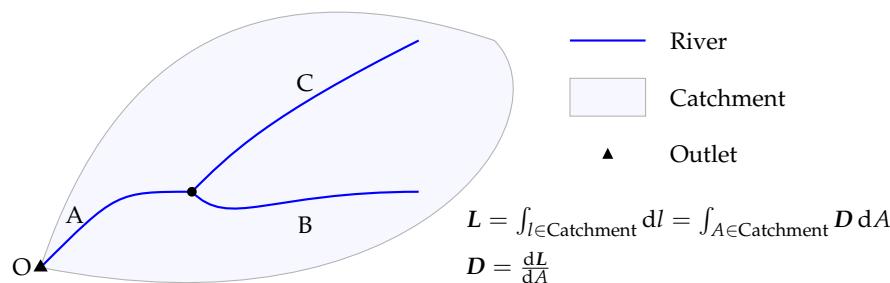
where  $dA$  denotes an infinitesimal area within the catchment and  $dl$  denotes the length of the river segments within the infinitesimal area. The equation states that the upstream accumulation length is equal to the integral of the drainage density over the area of the

catchment. In other words, if we have two-dimensional drainage density data and know the catchment of every grid cell over the area (which is feasible given flow direction), we can calculate a two-dimensional upstream accumulation length at every grid cell using Equation 2.

Furthermore, given flow direction, upstream accumulation length and drainage density have a strictly one-to-one relationship. We can not only precisely estimate upstream accumulation length from drainage density (Equation 2), but also vice versa. The formula for calculating drainage density from upstream accumulation length can be derived by differentiating Equation 2 as follows:

$$D = \frac{dL}{dA}. \quad (3)$$

This equation states that the drainage density of a grid cell is equal to the increase in upstream accumulation length divided by the area of the grid cell. The increase in upstream accumulation length can be estimated as the difference between the upstream accumulation length of a grid cell and the sum of the upstream accumulation length of the grid cell's donor cells. This one-to-one relationship between upstream accumulation length and drainage density is valid at any point of a two-dimensional domain.



**Figure 2.** Illustration of upstream accumulation length. Triangle symbols denote points in the study domain. The polygon represents the catchment of a point. Blue lines represent the river segments within the catchment. The upstream accumulation length ( $L$ ) is defined at the point, representing the total length of all the rivers within the catchment of the point. The symbol  $D$  represents drainage density.

Given flow direction and drainage density, the calculation of upstream accumulation length can be performed efficiently. Inspired by the work of Zhou et al. [17], we propose an algorithm to calculate the upstream accumulation length with  $O(N)$  time complexity. Algorithm 1 depicts the pseudocode. In the algorithm, we define three types of grid cells: (1) source cells, (2) intersection cells, and (3) interior cells. The number of donor cells for source cells is 0, for intersection cells is greater than 1, and for interior cells is 1. The algorithm starts from the source cells and iteratively calculates the upstream accumulation length downstream until an intersection cell is encountered. The number of donor cells for the encountered intersection cell is then decreased by 1. The algorithm continues until all source cells are processed. Similar to Zhou et al. [17], every grid cell is processed at most  $n$  times, where  $n$  is the maximum number of donor cells for a grid cell. Consequently, the time complexity of the algorithm is  $O(N)$ , where  $N$  is the number of grid cells in the study area.

**Input:** Two-dimensional flow direction ( $F$ ) and drainage density ( $D$ )  
**Output:** Two-dimensional upstream flow accumulation length ( $L$ ) for each grid cell

```

/* Count the number of upstream (donor) cells for each cell */          141
1 Let  $n$  be a two-dimensional array of the same size as  $F$ , filled by zeros;    142
2 foreach grid cell  $(i, j)$  in the study area do                                143
3   Check the flow direction of  $(i, j)$ , identify the downstream grid cell  $(i', j')$ ; 144
4    $n(i', j') = n(i', j') + 1$ ;                                                 145
5   /* Find the source cells */                                                 146
6 Let  $S$  be a stack of grid cell indices;                                         147
7 foreach grid cell  $(i, j)$  in the study area do                               148
8   if  $n(i, j) = 0$  then                                                 149
9     Push  $(i, j)$  to  $S$ ;                                                 150
10  /* Initialize  $L$  */                                                 151
11 Let  $L$  be a two-dimensional array of the same size as  $F$ ;                  152
12 foreach grid cell  $(i, j)$  in the study area do                            153
13    $L(i, j) = D(i, j) \times A(i, j)$ , where  $A(i, j)$  is the grid cell area; 154
14   /* Propagate length from sources downstream */                           155
15 while  $S$  is not empty do                                              156
16   Pop  $(i, j)$  from  $S$ ;                                                 157
17   while Check the flow direction  $F$ , if  $(i, j)$  has a downstream cell  $(i', j')$  do 158
18      $L(i', j') = L(i', j') + L(i, j)$ ;                                     159
19     if  $n(i', j') \geq 2$  then                                               160
20        $n(i', j') = n(i', j') - 1$ ;                                         161
21       Break;                                                               162
22     else                                                               163
23       Let  $(i, j) \leftarrow (i', j')$ ;                                         164
24   
```

**Algorithm 1:** Algorithm for calculating upstream accumulation length.

## 2.2. River Network Delineation with Upstream Accumulation Length

The key idea of our proposed algorithm is to transform drainage density into upstream accumulation length and to use the upstream accumulation length to delineate river networks. Since the upstream accumulation length and drainage density are mathematically equivalent, if the upstream accumulation length of every grid cell in the delineated river network matches the observed value, the drainage density of the delineated river network will also match the observed value at any point of a two-dimensional study area.

However, achieving this target exactly is computationally demanding. We can slightly relax the target and restate the objective of this study as follows: match the upstream accumulation length at every catchment outlet with the observed value. This relaxation significantly reduces the computational burden. Meanwhile, if the target is achieved, the average drainage density within every catchment of the delineated river network will also match the catchment-averaged observed value. This characteristic represents a substantial improvement over the existing methods discussed in the Introduction.

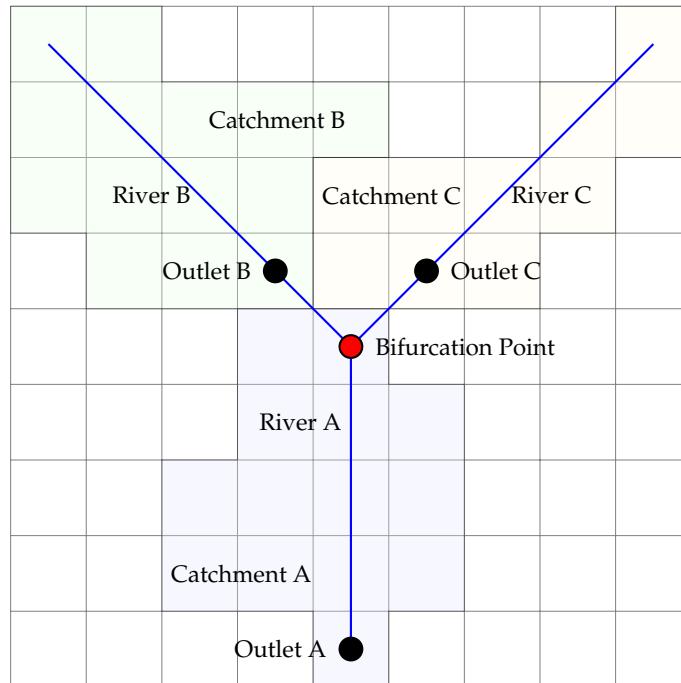
The delineation with the relaxed objective can be efficiently implemented using a divide-and-conquer strategy. Using Figure 3 as an illustration, the algorithm starts from the outlets of the whole domain (Outlet A in Figure 3) and iteratively grows the centerlines of the river network (blue lines) by adding grid cells. If a bifurcation is encountered, the algorithm splits the river network into two or more parts (three parts in Figure 3). A

bifurcation point (the red point) is defined as a grid cell that lies on the existing centerline and has two or more donor cells that flow from existing centerlines into it. The donor cells of the bifurcation point (Outlets B and C) are then identified using flow direction. These donor cells serve as the outlets of the upstream catchments. The newly identified outlets split the river network into parts, each with a river segment (River A, B, and C) and a corresponding catchment (Catchment A, B, and C).  
161  
162  
163  
164  
165  
166

The algorithm calculates the expected length for each river segment based on the upstream accumulation length. If a reach has no upstreams (Rivers B and C), the expected length of the reach is the upstream accumulation length at its outlet (Outlets B and C). If a reach has upstreams (River A), the expected length is the upstream accumulation length at its outlet (Outlet A) minus the sum of the upstream accumulation lengths of all its upstreams at their outlets (Outlets B and C). The algorithm grows each river segment by adding grid cells and repeats the growing process until the length of every reach is greater than or equal to the expected length. When the iteration stops, the upstream accumulation length at every catchment outlet is equal to the observed value.  
167  
168  
169  
170  
171  
172  
173  
174  
175

In Figure 3, if the upstream accumulation length of the delineated network is the same as the observed value at Outlets B and C, the drainage density within Catchments B and C must match the observed drainage density. Since the upstream accumulation length at Outlet A is equal to the sum of the lengths of Rivers A, B, and C, if the expected length of River A is also achieved, the upstream accumulation length at Outlet A is equal to the observed value. Consequently, the drainage density within the combined catchments A, B, and C must also match the observed drainage density. Given that the drainage density within Catchments B and C is already equal to the observed value, the drainage density within Catchment A must also match the observed value. In this case, the drainage density within every catchment of the delineated river network is equal to the observed value.  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185

The last piece of the algorithm is how to add grid cells to grow the river segments. There are certainly many ways to do this. We would like to keep our algorithm compatible with existing critical drainage area-based river network delineation algorithms. In critical drainage area-based algorithms, grid cells on the delineated rivers always have a larger drainage area value than those not delineated as rivers. Based on this observation, we grow the rivers by adding one grid cell at a time. The added grid cell is the one with the largest drainage area value among the neighbors of the delineated centerlines that flow into the centerlines. In this way, we ensure that the grid cells on the delineated centerlines always have a larger drainage area value than those not delineated as centerlines. This is the same idea as in critical drainage area-based algorithms.  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195



**Figure 3.** Divide-and-conquer strategy for river network delineation using upstream accumulation length. The algorithm grows rivers to expected lengths independently within the three catchments. Blue lines denote the rivers. Grid cells filled with different colors denote the catchments of different rivers. Black dots denote the outlets of each catchment.

Algorithm 2 shows pseudocode for the density-preserving river network delineation algorithm. The algorithm utilizes two queues: the Open queue, which is a temporary store of partially delineated rivers, and the Lines queue, which stores the finalized rivers. The algorithm grows each river in the Open queue. If the length of a river is greater than or equal to the expected length, the river is finalized and moved to the Lines queue. If a bifurcating condition is encountered, the algorithm divides the river into parts, as illustrated in Figure 3, and puts the parts back into the Open queue.

The algorithm is efficient. The divide-and-conquer strategy ensures that the river-growing process focuses on a specific reach and its neighboring grid cells at a time, without having to consider the entire study domain. This locality is especially important when dealing with large domains such as at the scale of a country, a continent, or even the globe.

196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206

**Input:** Two-dimensional flow direction ( $F$ ), drainage area ( $A$ ), and upstream accumulation length ( $L$ )

**Output:**  $Lines$  is a list of rivers, each represented by a list of grid cell indicies

```

/* Initialization of the delineation */
```

- 1 Let  $Open$  be an empty queue of lists of grid cell indicies;
- 2 **foreach** outlet  $(i, j)$  in the study area **do**
- 3 **if**  $A(i, j) > 0$  **then**
- 4  $Open.push(\text{single-element list } [(i, j),]);$
- 5 /\* Delineation of river centerlines \*/
- 6 **while**  $Open$  is not empty **do**
- 7  $L = Open.pop();$
- 8 **if**  $L$  has no upstreams **then**
- 9 Let the expected length  $l \leftarrow L(i, j)$ , where  $(i, j)$  is the last grid cell of  $L$ ;
- 10 **else**
- 11 Let the expected length  $l \leftarrow L(i, j) - \sum L(i', j')$ , where  $(i, j)$  is the last grid cell of  $L$ ,  $(i', j')$  is the last grid cell of the upstream of  $L$ ;
- 12 **loop**
- 13 Let  $N$  be a set of grid cells. Each of  $N$ : (1) is the neighbor of  $L$ , (2) flow into  $L$ , and (3) have not been included in any line in  $Lines$  or in  $Open$ ;
- 14 **if** Length of  $L \geq l$  **or**  $N$  is empty **then**
- 15  $Lines.append(L);$
- 16 **break**;
- 17 Find the grid cell  $(i_2, j_2)$  in  $N$  that has the largest  $A$ ;
- 18 Find  $(i', j')$  that is the grid cell that  $(i_2, j_2)$  flows to,  $(i_1, j_1)$  is the grid cell in  $L$  that also flows to  $(i', j')$ ;
- 19 **if**  $(i', j')$  is the last grid cell of  $L$  **then**
- 20  $L.append((i_2, j_2));$
- 21 **else**
- 22 /\* Bifurcation \*/
- 23 Split  $L$  into two sub-list:  $L_1$  is the list before  $(i', j')$  (inclusive) and  $L_2$  is the one after  $(i', j')$  (exclusive);
- 24  $Open.push(L_1);$
- 25  $Open.push(L_2);$
- 26  $Open.push(\text{one-element list } [(i_1, j_1),]);$
- 27 **break**;

**Algorithm 2:** Density-preserving river network delineation algorithm.

### 2.3. Avoidance of Short River Segments

The algorithm shown in Algorithm 2 strictly preserves the observed drainage density for every catchment of the delineated river network. However, the algorithm often produces a large number of short river segments. These short river segments are not only difficult to represent in hydrodynamic simulations but also have a negligible impact on the overall drainage density. Algorithm 3 shows an improved algorithm tailored for hydrodynamic simulations. Two thresholds are introduced in the algorithm: one for the minimum length of river segments ( $l_{cr}$ , line 24 in Algorithm 3) and the other for the minimum drainage area ( $A_{cr}$ , line 17 in Algorithm 3). The former is used to avoid overly short river segments on mainstreams, while the latter is used to avoid overly short river segments on head waters.

**Input:** Two-dimensional flow direction ( $F$ ), drainage area ( $A$ ), upstream accumulation length ( $L$ ), critical drainage area ( $A_{cr}$ ), and critical length ( $l_{cr}$ )

**Output:**  $Lines$  is a list of centerlines, each represented by a list of grid cell indicies

```

/* Initialization of the delineation */
1 Let  $Open$  be an empty queue of lists of grid cell indicies;
2 foreach outlet  $(i, j)$  in the study area do
3   if  $A(i, j) > 0$  then
4      $Open.push(\text{single-element list } [(i, j),]);$ 
      /* Delineation of river centerlines */
5 while  $Open$  is not empty do
6    $L = Open.pop();$ 
7   if  $L$  has no upstreams then
8     Let the expected length  $l \leftarrow L(i, j)$ , where  $(i, j)$  is the last grid cell of  $L$ ;
9   else
10    Let the expected length  $l \leftarrow L(i, j) - \sum L(i', j')$ , where  $(i, j)$  is the last grid
       cell of  $L$ ,  $(i', j')$  is the last grid cell of the upstream of  $L$ ;
11  loop
12  Let  $N$  be a set of grid cells. Each of  $N$ : (1) is the neighbor of  $L$ , (2) flow into
        $L$ , and (3) have not been included in any line in  $Lines$  or in  $Open$ ;
13  if Length of  $L \geq l$  or  $N$  is empty then
14     $Lines.append(L);$ 
15    break;
16  Find the grid cell  $(i_2, j_2)$  in  $N$  that has the largest  $A$ ;
17  if  $A(i_2, j_2) \leq A_{cr}$  then
18     $Lines.append(L);$ 
19    break;
20  Find  $(i', j')$  that is the grid cell that  $(i_2, j_2)$  flows to,  $(i_1, j_1)$  is the grid cell in
        $L$  that also flows to  $(i', j')$ ;
21  if  $(i', j')$  is the last grid cell of  $L$  then
22     $L.append((i_2, j_2));$ 
23  else
    /* Bifurcation */
24  if Length of  $L < l_{cr}$  then
25     $Lines.append(L);$ 
26    break;
27  Split  $L$  into two sub-list:  $L_1$  is the list before  $(i', j')$  (inclusive) and  $L_2$  is
       the one after  $(i', j')$  (exclusive);
28   $Open.push(L_1);$ 
29   $Open.push(L_2);$ 
30   $Open.push(\text{one-element list } [(i_1, j_1),]);$ 
31  break;

```

**Algorithm 3:** Density-preserving river network delineation algorithm with thresholds for river segment length and drainage area.

### 3. Experimental Setup

We test the proposed algorithm using two cases: a synthetic case and a real-world case. The synthetic case is designed to demonstrate the steps of the algorithm and to show

its compatibility with the critical drainage area-based algorithm. The real-world case is designed to demonstrate the computational efficiency and effectiveness of the proposed algorithm in delineating river networks over large domains.

The synthetic case is a  $4 \times 4$  grid with a grid size of  $1 \times 1$ . Figure 4b show the flow direction and drainage area, respectively. A river network is first delineated using the critical drainage area-based algorithm with a threshold of 4, as shown in Figure 4c. This river network is used as the synthetic truth. Drainage density is calculated from this network. We then delineate the river network using the proposed algorithm to check whether the proposed algorithm can produce the same river network shown in Figure 4c.

The real-world case involves delineating the river network over the Chinese Mainland. The MERIT-Hydro flow direction dataset [6] is used in this case. MERIT-Hydro is a global 3-arc-second (approximately 90 m) resolution dataset derived from the MERIT DEM [18]. This dataset has been widely used in hydrodynamic simulations [19]. Lin et al. [5] have shown that the river network delineated from the MERIT-Hydro flow direction dataset has high accuracy in terms of centerline position. The drainage density dataset used in this case is produced from the Third National Land Resources Survey of China [20]. The dataset covers the entire Chinese Mainland at a spatial resolution of 1 km. We bilinearly interpolated the 1-km drainage density data to the 90-m MERIT-Hydro grid. To ensure the delineated river network is continuous, we slightly extended the boundary of the study area outside the Chinese Mainland. In these extended areas, there is no surveyed drainage density data. The drainage density is estimated using a river network delineated from the MERIT-Hydro dataset and the critical drainage area-based method with a threshold of 1  $\text{km}^2$ .

## 4. Results

### 4.1. The Synthetic Case

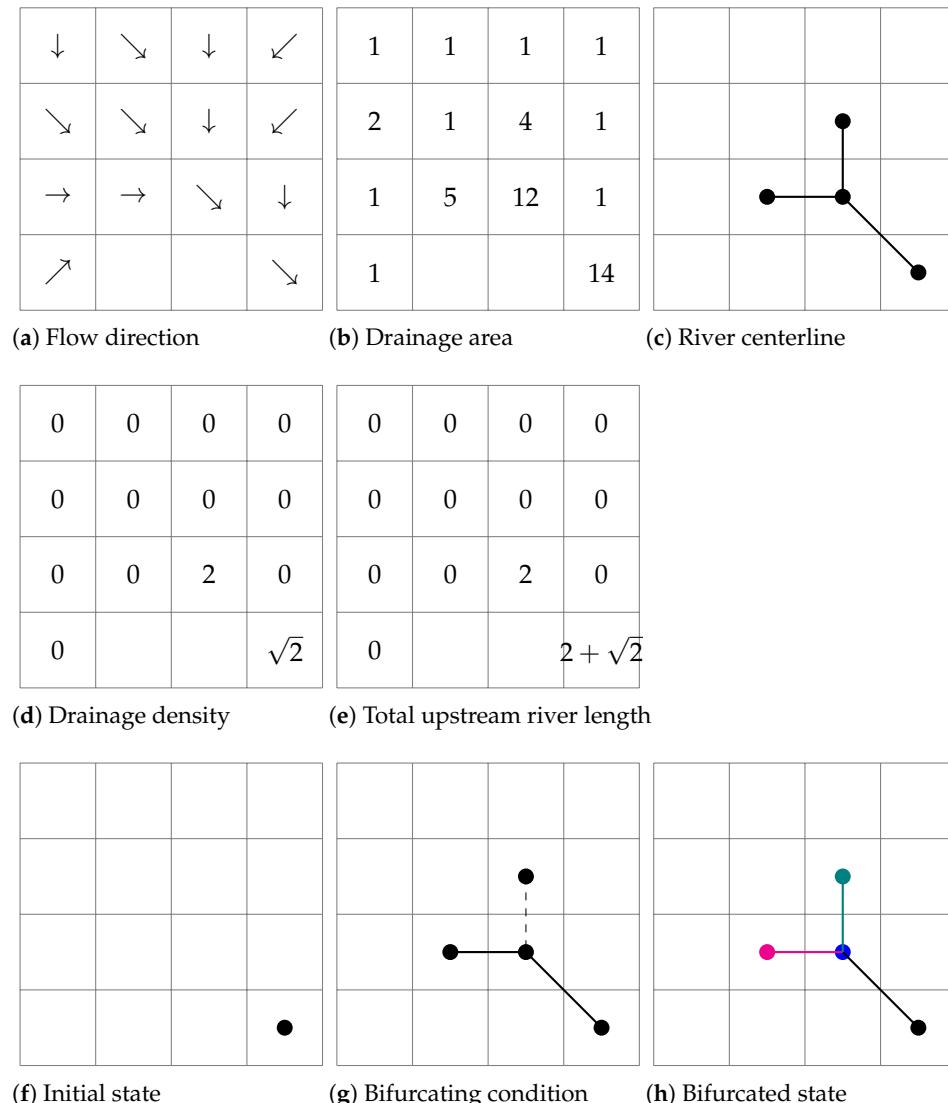
Figure 4d shows the drainage density calculated from Figure 4c. Note that the river length at each grid cell is calculated as the total length of the river segments connecting the grid cell center to the centers of its upstream cells. This calculation differs slightly from survey data or imagery-based data. If the grid cell is small (i.e., a typical grid cell is smaller than 100 m  $\times$  100 m), the difference is negligible. Figure 4e shows the total upstream river length calculated according to the algorithm described in Algorithm 1.

The last row of Figure 4 reveals the delineation process. Figure 4f shows the state after initialization. The *Open* queue contains the grid cells at the outlets of the whole domain. After initialization, a centerline is grown from the outlet by adding the grid cell with the largest drainage area value in the neighborhood of the already delineated centerlines. This process is repeated until a bifurcating condition is encountered. Figure 4g shows the bifurcating condition, where the solid lines denote the already delineated rivers and the dashed line connects the grid cell with the largest drainage area value.

As the grid cell with the largest drainage area value (denoted by the dot connected by the dashed line) does not flow to the head of the river, the river is split into two parts, denoted by the magenta and black lines in Figure 4h. The parts and the grid cell with the largest drainage area (i.e., the green line) are then added to the *Open* queue for further delineation. In this simple case, the lengths of all three centerlines are already greater than or equal to the expected river length at this stage, as described in Algorithm 2, so the algorithm terminates.

In this synthetic case, the river network delineated using our proposed algorithm (Figure 4h) is identical to the one delineated using the critical drainage area-based method (Figure 4c). This demonstrates that our proposed algorithm produces the same results as critical drainage area-based methods when using the same drainage density. The only

difference is that our algorithm uses two-dimensional drainage density (Figure 4d) as an input, while the critical drainage area-based method does not have a priori information about the drainage density.

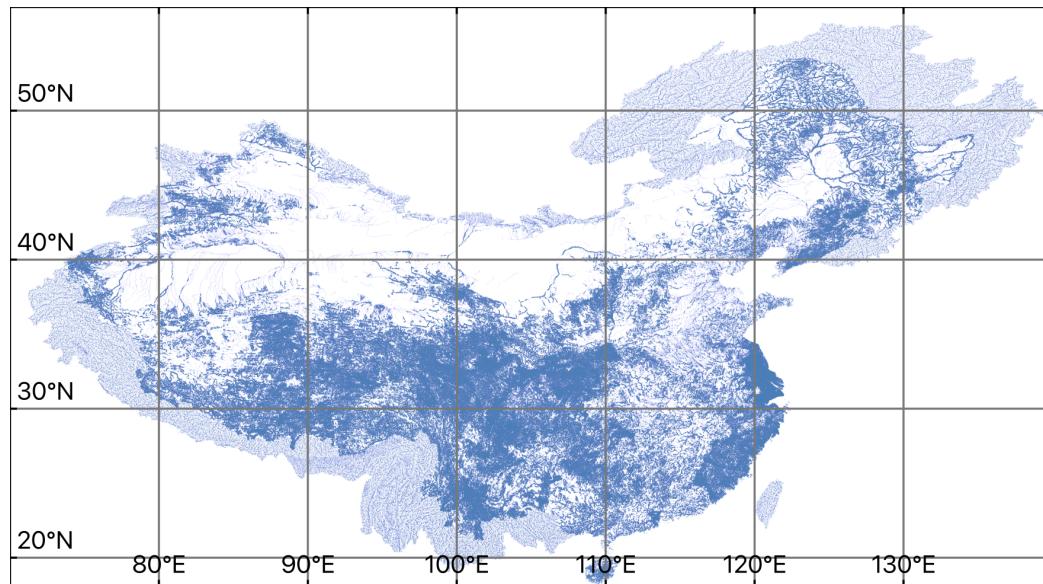


**Figure 4.** An illustration of the density-preserving river network delineation algorithm. Panel (c) serves as the synthetic truth of the river network.

#### 4.2. The Real-World Case

In the real-world case, we applied the improved algorithm shown in Algorithm 3 to delineate the river network over the Chinese Mainland. The delineation was performed on the 90-m MERIT Hydro grids using a  $0.5 \text{ km}^2$  threshold for the minimum drainage area and a 0.5 km threshold for the minimum river segment length.

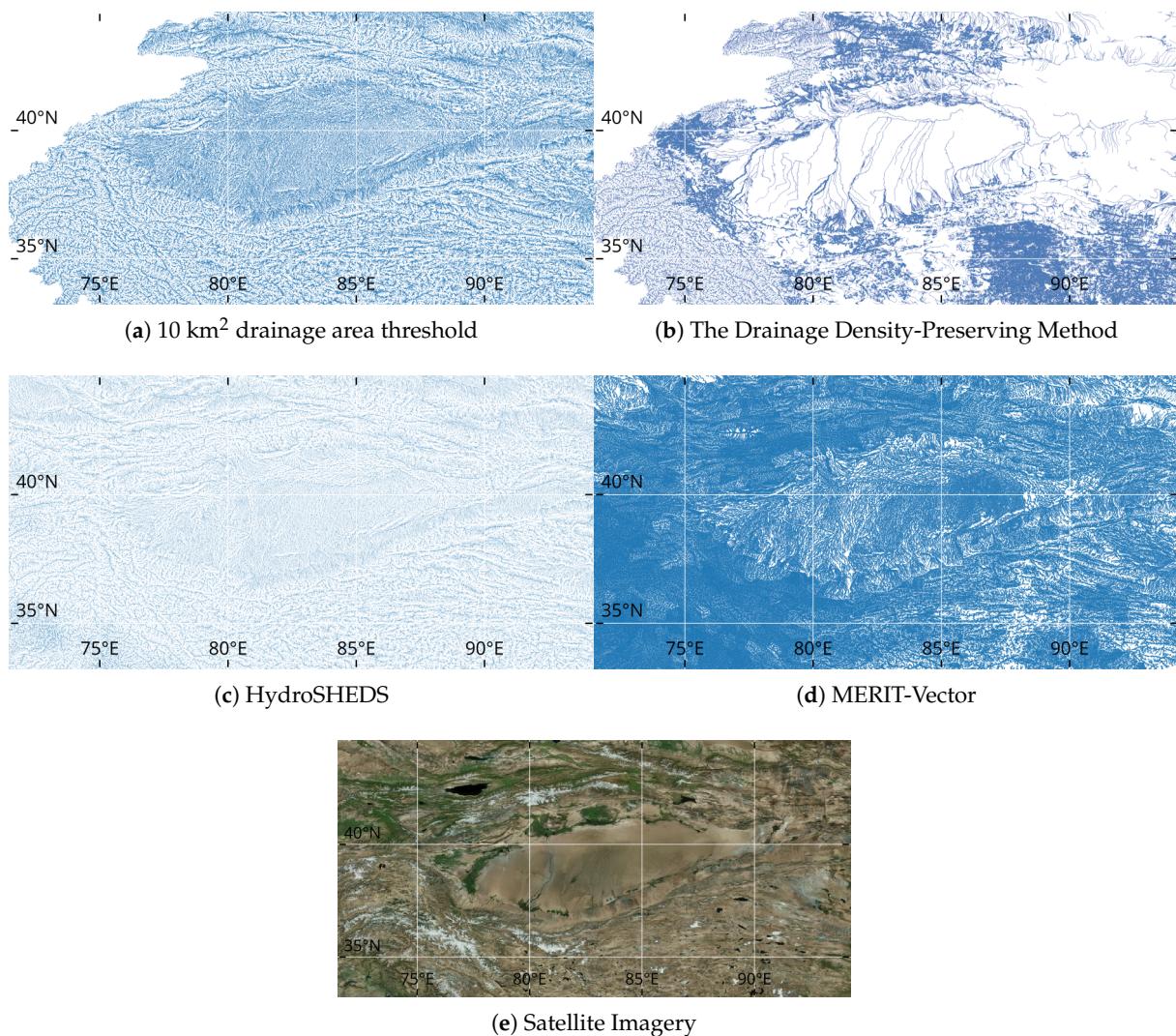
The delineation process was efficient, completing in less than 10 minutes using a single thread of an Apple M3 Max processor. This represents a substantially reduced computational burden compared to previous studies [5]. The efficiency is attributed to the divide-and-conquer strategy and the use of two-dimensional drainage density information as a priori input, rather than using it ad hoc in a trial-and-error process. The delineated river network consists of 2,314,986 river segments, which are available for download from <https://doi.org/10.57760/sciedb.23487> (accessed on 2024-04-10).



**Figure 5.** A Chinese Mainland river network dataset delineated using the drainage density-preserving method.

Besides showing the efficiency, we also demonstrate the effectiveness of the proposed algorithm in capturing the variation of drainage density. Figure 6 compares the river network delineation results in South Xinjiang, China. South Xinjiang is characterized by a variety of landscapes, including mountains and deserts. The drainage density varies significantly across the region, making it an ideal testbed for intercomparison. Figure 6a shows the river network delineated using the critical drainage area-based method with a threshold of  $10 \text{ km}^2$ . The river network appears nearly uniform across the region, failing to capture the variation in drainage density. Compared with satellite imagery, the river network is overly dense in the desert areas. Figure 6b shows the river network delineated using the drainage density-preserving method. The result is encouraging, not only representing the contrast in drainage density between the foothills and deserts, but also resolving rivers that flow across the desert. Figure 6c shows the hydrological data and maps based on the Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS) river network dataset, while Figure 6d shows the MERIT-Vector river network dataset [5]. MERIT-Vector has a much denser network than HydroSHEDS and shows more variation in drainage density. However, they both produce overly dense river networks in the desert and show smaller variations between foothills and desert than Figure 6b. The comparison demonstrates that our proposed algorithm can effectively capture the variation in drainage density across the entire river network, a feature that existing publicly available datasets have not performed as well in.

In this study, we have not evaluated the position accuracy of the delineated river network centerline. Since we are using the same flow direction dataset as Lin et al. [5], we expect that the river network delineated in this study will also achieve a similar level of centerline position accuracy as reported by Lin et al. [5].



**Figure 6.** Comparison of the river network delineation results in South Xinjiang. (a) delineation with critical drainage area of 10 km<sup>2</sup>, (b) delineation with the drainage density-preserving method, (c) the HydroSHED river network dataset, (d) the MERIT-Vector river network dataset, and (e) Satellite imagery.

## 5. Discussion

The proposed algorithm uses two-dimensional drainage density as a priori information to delineate river networks, and the delineated river network can preserve the observed drainage density at every delineated catchment. This represents a significant improvement over existing methods, which often fail to capture the variation of drainage density across large domains. The capability of capturing the variation of drainage density is not limited to the spatial dimension but could also be extended to the temporal dimension. Given a time-varying drainage density dataset, the proposed algorithm can be used to delineate river networks that preserve the observed drainage density at every time step. Although we do not expect the algorithm to capture fragmented rivers, it is still able to represent the extension and retreat of headwaters. Such studies would be useful for understanding the changes in non-perennial rivers and the impacts of climate change on river networks [21–24].

However, the proposed algorithm is still subject to several limitations that are common to all DEM-based methods. First, the resolution of the input DEM significantly affects the delineation results [25]. A higher resolution DEM provides more accurate flow direction

310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325

and drainage area information, would lead to a more accurate delineation of river networks. However, DEM-based methods discussed in this study have difficulties when the DEM resolution is too fine. As shown by Bernard et al. [26], Costabile et al. [27], these methods, including the one proposed in this study, are not able to capture the morphology of rivers wider than the resolution. Special treatment of the river surface [28] is necessary to adjust the DEM. Second, the delineated river network is affected by the accuracy of the input DEM and drainage density data. Dense vegetation in the tropics and steep terrain at the edge of mountains can introduce errors in the DEMs, which can propagate to the delineated river network. In this study, we used the MERIT-Hydro dataset, which is produced by correcting multiple errors in DEMs, including the impacts of vegetation. However, the impacts of errors in drainage density data are still unknown. We leave the investigation of drainage density's impact for future research to keep this paper focused and concise. Third, the delineated river network represents mere line skeletons without width information. For use in large-domain hydrodynamic simulations, the cross-section of the rivers is parameterized using simple geometry and calibratable geometric parameters [29]. This approach often faces difficulties in representing floodplains accurately. Two-dimensional hydraulic simulations on DEMs show potential to address these limitations. However, the computational cost hinders the application of two-dimensional hydraulic simulations on large domains, like the Chinese Mainland presented in this study.

## 6. Conclusions

This study introduces a novel river network delineation method that preserves observed drainage density across the entire river network. The method is based on a novel concept named upstream accumulation length, which is calculated using flow direction and drainage density. The delineation algorithm is designed to be compatible with critical drainage area-based methods, allowing for the preservation of observed drainage density in every single catchment of the river network.

The proposed algorithm was applied to the Chinese Mainland, utilizing the MERIT-Hydro flow direction dataset and a kilometer-resolution drainage density dataset produced from the Third National Land Resources Survey of China. The resulting river network dataset is segmented into nearly uniform lengths of approximately one kilometer, providing a critical foundation for the development of national-scale flood forecasting systems in the Chinese Mainland.

The proposed algorithm was tested on a synthetic case and a real-world case. The results from the synthetic case demonstrated that the proposed algorithm can produce the same river network as critical drainage area-based methods when using the same drainage density. The real-world case that to delineate the river network over Chinese Mainland showed that the proposed algorithm is both computationally efficient and effective in capturing the variation of drainage density. 2,314,986 river segments were delineated in less than 10 minutes on a single thread of an Apple M3 Max processor. The delineated river network was shown to be superior in capturing the variation of drainage density across the entire river network compared to existing publicly available datasets such as HydroSHEDS and MERIT-Vector.

The results are encouraging, since the method opens the possibility of combining aerial satellite imagery and DEMs to delineate river networks accurately on a global scale. This is particularly important, as only a few countries (e.g., the United States of America, the United Kingdom, France, and Australia) have manually maintained river network datasets. The vast majority of the world, including China, lacks such datasets. The proposed algorithm can be applied to delineate river networks in these regions, providing a valuable resource for hydrodynamic simulations and flood forecasting systems.

**Author Contributions:** Conceptualization, H.Z.; methodology, H.Z.; software, H.Z.; validation, S.C. and H.Y.; formal analysis, S.C. and H.Y.; investigation, S.C. and H.Y.; resources, S.C., H.Y. and H.Z.; data curation, S.C. and H.Y.; writing—original draft preparation, S.C. and H.Y.; writing—review and editing, S.C., H.Y., and H.Z.; visualization, S.C. and H.Y.; supervision, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is funded by China Three Gorges Corporation (contract Z532302035), the National Key Research and Development Program of China (2023YFF0805501), and the Natural Science Foundation of China (grants 42075165 and 42275178).

**Data Availability Statement:** The MERIT-Hydro dataset was obtained from [https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT\\_Hydro/](https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/) (accessed on 2024-11-10). The 1 km Drainage Density Dataset in China (2022) was obtained from <https://doi.org/10.57760/sciencedb.16001> (accessed on 2024-11-10). The MERIT-Vector river network dataset was obtained from <https://doi.org/10.6084/m9.figshare.c.5052635.v1> (accessed on 2025-01-10). The HydroSHED river network dataset was obtained from <https://www.hydrosheds.org> (accessed on 2025-01-10). The delineated river network dataset in this study can be found on <https://doi.org/10.57760/sciencedb.23487> (accessed on 2024-04-10).

**Conflicts of Interest:** Authors Heng Yang and Shuanglong Chen were employed by the China Three Gorges Corporation. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. David, C.H.; Maidment, D.R.; Niu, G.Y.; Yang, Z.L.; Habets, F.; Eijkhout, V. River network routing on the NHDPlus dataset. *Journal of Hydrometeorology* **2011**, *12*, 913–934. <https://doi.org/10.1175/2011JHM1345.1>.
2. David, C.H.; Famiglietti, J.S.; Yang, Z.L.; Habets, F.; Maidment, D.R. A decade of RAPID—Reflections on the development of an open source geoscience code. *Earth and Space Science* **2016**, *3*, 226–244. <https://doi.org/10.1002/2015ea000142>.
3. Cosgrove, B.; Gochis, D.; Flowers, T.; Dugger, A.; Ogden, F.; Graziano, T.; Clark, E.; Cabell, R.; Casiday, N.; Cui, Z.; et al. NOAA’s National Water Model: Advancing operational hydrology through continental-scale modeling. *Journal of the American Water Resources Association* **2024**, *60*, 247–272. <https://doi.org/10.1111/1752-1688.13184>.
4. Tarboton, D.G.; Bras, R.L.; Rodriguez-Iturbe, I. On the extraction of channel networks from digital elevation data. *Hydrological Processes* **1991**, *5*, 81–100. <https://doi.org/10.1002/hyp.3360050107>.
5. Lin, P.; Pan, M.; Wood, E.F.; Yamazaki, D.; Allen, G.H. A new vector-based global river network dataset accounting for variable drainage density. *Scientific Data* **2021**, *8*, 28. <https://doi.org/10.1038/s41597-021-00819-9>.
6. Yamazaki, D.; Ikeshima, D.; Sosa, J.; Bates, P.D.; Allen, G.H.; Pavelsky, T.M. MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research* **2019**, *55*, 5053–5073. <https://doi.org/10.1029/2019wr024873>.
7. Amatulli, G.; Garcia Marquez, J.; Sethi, T.; Kiesel, J.; Grigoropoulou, A.; Üblacker, M.M.; Shen, L.Q.; Domisch, S. Hydrography90m: a new high-resolution global hydrographic dataset. *Earth System Science Data* **2022**, *14*, 4525–4550. <https://doi.org/10.5194/essd-14-4525-2022>.
8. Munier, S.; Decharme, B. River network and hydro-geomorphological parameters at 1/12° resolution for global hydrological and climate studies. *Earth System Science Data* **2022**, *14*, 2239–2258. <https://doi.org/10.5194/essd-14-2239-2022>.
9. Wang, Z.; Liu, J.; Li, J.; Meng, Y.; Pokhrel, Y.; Zhang, H. Basin-scale high-resolution extraction of drainage networks using 10-m Sentinel-2 imagery. *Remote Sensing of Environment* **2021**, *255*, 112281. <https://doi.org/10.1016/j.rse.2020.112281>.
10. Mutzner, R.; Bertuzzo, E.; Tarolli, P.; Weijs, S.V.; Nicotina, L.; Ceola, S.; Tomasic, N.; Rodriguez-Iturbe, I.; Parlange, M.B.; Rinaldo, A. Geomorphic signatures on Brutsaert base flow recession analysis. *Water Resources Research* **2013**, *49*, 5462–5472. <https://doi.org/10.1002/wrcr.20417>.
11. Horton, R.E. Drainage-basin characteristics. *Eos, Transactions American Geophysical Union* **1932**, *13*, 350–361. <https://doi.org/10.1029/TR013i001p00350>.
12. Horton, R.E. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *GSA Bulletin* **1945**, *56*, 275–370. [https://doi.org/10.1130/0016-7606\(1945\)56\[275:EDOSAT\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2).
13. Tarboton, D.G.; Bras, R.L.; Rodriguez-Iturbe, I. A physical basis for drainage density. *Geomorphology* **1992**, *5*, 59–76. [https://doi.org/10.1016/0169-555X\(92\)90058-V](https://doi.org/10.1016/0169-555X(92)90058-V).
14. Singh, S.K.; Griffiths, G.A. Prediction of streamflow recession curves in gauged and ungauged basins. *Water Resources Research* **2021**, *57*, e2021WR030618. <https://doi.org/10.1029/2021WR030618>.

15. Pallard, B.; Castellarin, A.; Montanari, A. A look at the links between drainage density and flood statistics. *Hydrology and Earth System Sciences* **2009**, *13*, 1019–1029. <https://doi.org/10.5194/hess-13-1019-2009>. 426  
427
16. Schneider, A.; Jost, A.; Coulon, C.; Silvestre, M.; Théry, S.; Ducharne, A. Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophysical Research Letters* **2017**, *44*, 2773–2781. <https://doi.org/10.1002/2016GL071844>. 428  
429  
430
17. Zhou, G.; Wei, H.; Fu, S. A fast and simple algorithm for calculating flow accumulation matrices from raster digital elevation. *Frontiers of Earth Science* **2019**, *13*, 317–326. <https://doi.org/10.1007/s11707-018-0725-9>. 431  
432
18. Yamazaki, D.; Ikeshima, D.; Tawatari, R.; Yamaguchi, T.; O'Loughlin, F.; Neal, J.C.; Sampson, C.C.; Kanae, S.; Bates, P.D. A high-accuracy map of global terrain elevations. *Geophysical Research Letters* **2017**, *44*, 5844–5853. <https://doi.org/10.1002/2017GL072874>. 433  
434  
435
19. Yang, Y.; Pan, M.; Lin, P.; Beck, H.E.; Zeng, Z.; Yamazaki, D.; David, C.H.; Lu, H.; Yang, K.; Hong, Y.; et al. Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bulletin of the American Meteorological Society* **2021**, *102*, E2086–E2105. <https://doi.org/10.1175/BAMS-D-20-0057.1>. 436  
437  
438
20. Zhang, D.X.; Peng, J. A dataset of 1km Grid Drainage Density in China (2022), 2024. <https://doi.org/10.57760/SCIENCEDEB.16001>. 439  
440
21. Durighetto, N.; Botter, G. On the relation between active network length and catchment discharge. *Geophysical Research Letters* **2022**, *49*, e2022GL099500. <https://doi.org/10.1029/2022GL099500>. 441  
442
22. Zanetti, F.; Botter, G.; Camporese, M. Stream network dynamics of non-perennial rivers: insights from integrated surface-subsurface hydrological modeling of two virtual catchments. *Water Resources Research* **2024**, *60*, e2023WR035631. <https://doi.org/10.1029/2023WR035631>. 443  
444  
445
23. Price, A.N.; Jones, C.N.; Hammond, J.C.; Zimmer, M.A.; Zipper, S.C. The drying regimes of non-perennial rivers and streams. *Geophysical Research Letters* **2021**, *48*, e2021GL093298. <https://doi.org/10.1029/2021gl093298>. 446  
447
24. Messager, M.L.; Lehner, B.; Cockburn, C.; Lamouroux, N.; Pella, H.; Snelder, T.; Tockner, K.; Trautmann, T.; Watt, C.; Datry, T. Global prevalence of non-perennial rivers and streams. *Nature* **2021**, *594*, 391–397. <https://doi.org/10.1038/s41586-021-03565-5>. 448  
449
25. McMaster, K.J. Effects of digital elevation model resolution on derived stream network positions. *Water Resources Research* **2002**, *38*, 13–1–13–8. <https://doi.org/10.1029/2000WR000150>. 450  
451
26. Bernard, T.G.; Davy, P.; Lague, D. Hydro-geomorphic metrics for high resolution fluvial landscape analysis. *Journal of Geophysical Research: Earth Surface* **2022**, *127*, e2021JF006535. <https://doi.org/10.1029/2021JF006535>. 452  
453
27. Costabile, P.; Costanzo, C.; Lombardo, M.; Shavers, E.; Stanislawska, L.V. Unravelling spatial heterogeneity of inundation pattern domains for 2D analysis of fluvial landscapes and drainage networks. *Journal of Hydrology* **2024**, *632*, 130728. <https://doi.org/10.1016/j.jhydrol.2024.130728>. 454  
455  
456
28. Barnes, R.; Lehman, C.; Mulla, D. Priority-Flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences* **2014**, *62*, 117–127. <https://doi.org/10.1016/j.cageo.2013.04.024>. 457  
458
29. Read, L.K.; Yates, D.N.; McCreight, J.M.; Rafieeinabas, A.; Sampson, K.; Gochis, D.J. Development and evaluation of the channel routing model and parameters within the National Water Model. *JAWRA Journal of the American Water Resources Association* **2023**, *59*, 1051–1066. <https://doi.org/10.1111/1752-1688.13134>. 459  
460  
461

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.