

Article

# Selection Method of Dendritic River Networks Based on Hybrid Coding for Topographic Map Generalization

Chengming Li <sup>1,2,3</sup>, Wei Wu <sup>1</sup>, Pengda Wu <sup>2,3,\*</sup>, Yong Yin <sup>1,2,3</sup> and Zhaoxin Dai <sup>2,3</sup>

<sup>1</sup> College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China; licm@casm.ac.cn (C.L.); wuwei@casm.ac.cn (W.W.); yinyong@casm.ac.cn (Y.Y.)

<sup>2</sup> National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Xi'an 710072, China; daizx@casm.ac.cn

<sup>3</sup> Chinese Academy of Surveying and Mapping, Beijing 100830, China

\* Correspondence: wupd@casm.ac.cn; Tel.: +86-131-2121-5851

Received: 31 March 2020; Accepted: 4 May 2020; Published: 10 May 2020



**Abstract:** As the coding of a dendritic river system can be used to represent the stream order and spatial-structure of a river network, it is always used in river selection, which is a key step in topographic map generalization. There are two categories of conventional hydrological coding systems, one is the top-down approach, and the other is the bottom-up approach. However, the former does not accurately reflect the hierarchies of a dendritic river network, which is produced by catchment relationships, and it is not appropriate for the stream selection of river networks with uniform distributions of tributaries. The latter cannot directly indicate the subtree depth of a stream, and it is not favorable to stream selection of river systems that have topologically deep structures. Therefore, a selection method for dendritic river networks based on hybrid coding is proposed in this paper. First, the dendritic river network is coded through classical top-down Horton coding. Second, directed topology trees are constructed to organize the river network data, and stroke connections are calculated to code the river network in the bottom-up approach. Third, the river network is marked through hybrid usage of the top-down approach and bottom-up approach, and based on the spatial characteristics of the river network, the river network is classified into three kinds of subtrees: deep branch, shallow branch and modest branch. Then, appropriate coding is assigned automatically to different subtrees to achieve river selection. Finally, actual topographic map data of a river system in a region of Hubei Province are used to comparatively validate the hybrid coding system against two existing isolated coding systems. The experimental results demonstrate that the hybrid coding method is very effective for river network selection, not only in highlighting hierarchies formed by catchment relationships but also in the uniform distribution of tributaries.

**Keywords:** map generalization; dendritic river networks; river network selection; hybrid ordering; subtree classification

## 1. Introduction

The river network is the skeleton element in multi-scale topographic maps. Due to scale sensitivity, hydrographic data requires frequent generalization to maintain the main and typical characteristics of the river network [1]. Generally, dendritic river features are characterized by distinct hierarchical and fractal features [2,3], and coding is an effective approach for revealing the spatial structure of river features and the order of each stream. Therefore, the coding of dendritic river features has always been a focal point for map generalization studies [4,5].

Horton [6] proposed the classic Horton coding system based on the natural river property, in which water flows under the effects of gravity. The Horton coding system distinguishes the order of each stream and provides an intuitive reflection of the subtree depth of the stream, which makes it a powerful tool in quantitative analyses of the structural features of river networks. For example, Moharir et al. [7] and Harish et al. [8] combined Horton's stream ordering system with remote-sensing images to analyze the topographic features of watersheds. Most of all, since Horton's coding system can be used to distinguish the stream order, the coding system was used for stream selection in the generalization of river systems [9].

Strahler [10] further developed Horton's theories to establish the Strahler coding system, which is based on river segments. Strahler's coding system reflects the sizes and morphological features of substreams, and it is useful for highlighting the dendritic features of a river network. Stanislawski et al. [11] used the Strahler coding system as a criterion by which to identify the importance of a reach during stream generalization and selection. Since the reach-based coding system is a better fit for organizing river networks in computers, Shreve [12], Horsfield and Cumming [13] proposed their own hydrologic coding systems based on Strahler's coding system to reflect features such as the number of tributaries and differences in the density of a river network. The aforementioned coding systems are widely applied in hydrographic analyses and river generalizations. However, these systems generally use a top-down approach, wherein streams are ordered from the end of a river system (i.e., river source) towards the main stem (i.e., estuary). When these coding systems are used to process large dendritic river systems, the resulting stream codes are not very suitable for highlighting the hierarchies of a river system, which are formed by catchment relationships.

Gravelius [14] and Hack [15] proposed a classic definition for the stream order according to catchment relationships. Based on this definition, an automated bottom-up hydrologic coding system was developed for dendritic river systems [16,17]. In the method, river entities are treated as elements, and the stream hierarchy is accounted for in the coding of the stream order. Compared to Horton's stream coding system, this coding system cannot directly indicate the subtree depth of a stream, yet it does provide a better indication of the stream hierarchies produced by catchment relationships. In short, during stream selection and generalization, Horton's stream coding system is better suited for dendritic river systems that have topologically deep structures, whereas the Gravelius [14] and Hack [15] stream coding system is better suited for topologically shallow dendritic river systems with a uniform distribution of tributaries.

In the traditional river selection method for topographic map generalization, the aforementioned two coding systems were used individually, scholars either choose a top-down approach or a bottom-up approach. However, the spatial distribution structure of river networks is complex and diverse in actual data, it is difficult for a single code system to describe the importance of each stream during the selection operation. As a result, some river subtrees can't keep the original spatial distribution characteristics after automatic selection. Therefore, by combining the advantages of both of these methods, a selection method of dendritic river network based on hybrid coding is proposed in this paper.

The paper includes five sections. Section 2 introduces the existing top-down and bottom-up coding methods, and simultaneously, their shortcomings are analyzed. The proposed selection method for dendritic river networks based on hybrid coding is explained in detail in Section 3. Section 4 explains the experiments and then presents and analyzes the results, followed by the discussions and conclusions in Section 5.

## 2. Related Works

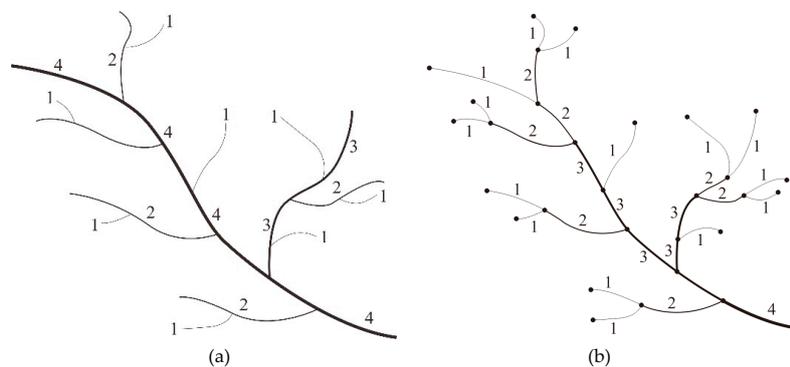
### 2.1. Current Stream Coding Methods

#### 2.1.1. Top-Down Coding System

There are two classical top-down coding systems. One is Horton's stream coding system, and the other is Strahler's stream coding system.

The objective of Horton's coding system is to categorize river entities in different watersheds that have similar shapes and structural features into the same order. To this end, branchless terminal tributaries in a dendritic river feature are assigned an order of one in the Horton stream system. Based on the direction of each stream, rivers that only contain first-order streams are defined as second-order streams, while rivers that contain first- and second-order streams are defined as third-order streams. In this way, the main stem of a river system, which has the greatest number of tributaries, is defined as the highest-order stream ( $n$ th-order), as shown in Figure 1a.

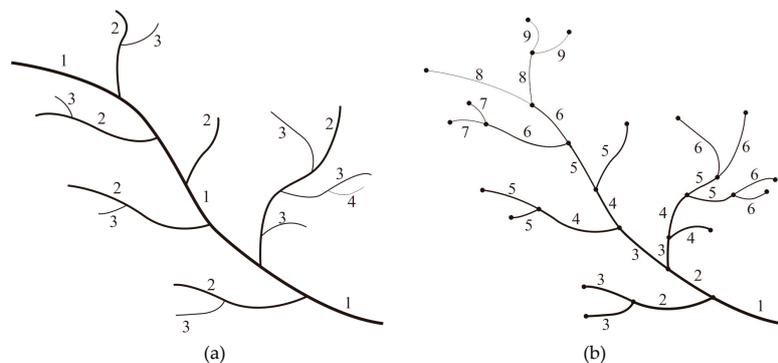
Strahler stream ordering is performed on a dendritic river features based on river segments. River segments that are connected to river source nodes are defined as first-order reaches, while reaches formed by the convergence of two or more first-order reaches have a stream order of 2. In this way, any reach formed by two or more  $n$ th-order reaches has a stream order of  $n + 1$ . All of the reaches are traversed until all of the streams have been coded. The Strahler codes of a river system are then obtained, as shown in Figure 1b.



**Figure 1.** Two classic stream coding systems for dendritic river features: (a) Horton coding based on river streams and (b) Strahler coding based on river segments.

### 2.1.2. Bottom-Up Coding System

Just like the top-down coding system, there are two kinds of bottom-up coding systems. One is Hack's stream coding system based on river streams, and the other is Marani's stream coding system based on river segments. A bottom-up coding system was proposed and implemented [14–17] through a classic definition of the stream order, in which the main stem is defined as a first-order stream, while the tributaries connected to the main stem are second-order streams. In this way, branchless tributaries at the end of a river system are defined as the highest-order streams, as shown in Figure 2a. This definition of the stream order clearly defines the hierarchies formed by catchment relationships in a river system. Marani et al. [18] proposed a coding system based on river segments, which is also known as the topological diameter coding method [16], as shown in Figure 2b.

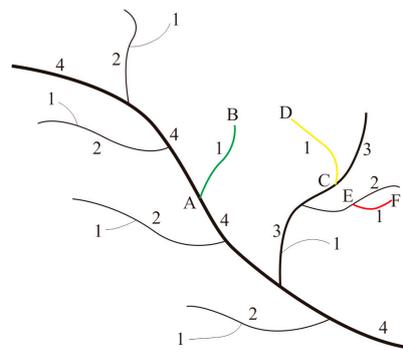


**Figure 2.** Bottom-up coding system for dendritic river features: (a) Hack coding based on river streams and (b) Marani coding based on river segments.

## 2.2. Inadequacies of the Current Methods

### 2.2.1. Inadequacies of the Top-Down Approach in River Selection

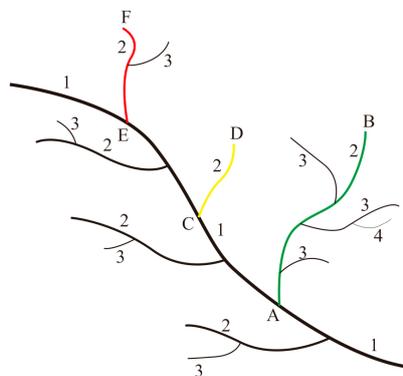
The top-down coding approach is often used in river generalization to reflect the importance of a river [1,11]. Compared to Strahler's coding system, Horton's coding system has more advantages in expressing the connectivity of a river because the unit in Horton's coding is river entities. Figure 1a shows that Horton's coding system (which is based on river entities) is an excellent method for distinguishing the stream order and describing the subtree depth of a stream. However, since Horton's coding system uses a top-down approach, tributary entities that have the same Horton order in large river systems may not actually have the same stream order. In Figure 3, the AB (green), CD (yellow) and EF (red) streams all have a Horton code of 1, despite having different stream orders in reality. Stream AB is connected to the main stem and is, therefore, a first-order stream; stream CD is connected to a first-order tributary and is, therefore, a second-order stream; and stream EF is connected to a second-order tributary and is, therefore, a third-order stream. Hence, river entities that have the same Horton stream order cannot be compared to each other, and the orders assigned by this system do not accurately reflect the stream hierarchies formed by the catchment relationships.



**Figure 3.** Inadequacies of top-down coding methods: three streams (AB, CD and EF) with different levels have the same code one.

### 2.2.2. Inadequacies of the Bottom-Up Approach in River Selection

We still focus on the coding approach based on river streams, which code the streams rely on in the catchment relationships of each river [15,16] are shown in Figure 4. However, since the coding system uses a top-down approach, tributary entities that have the same coding order in large river systems may not actually describe the subtree depth of a stream. In Figure 4, the AB (green), CD (yellow) and EF (red) streams all have a code of two, despite their deep structures being different in reality.



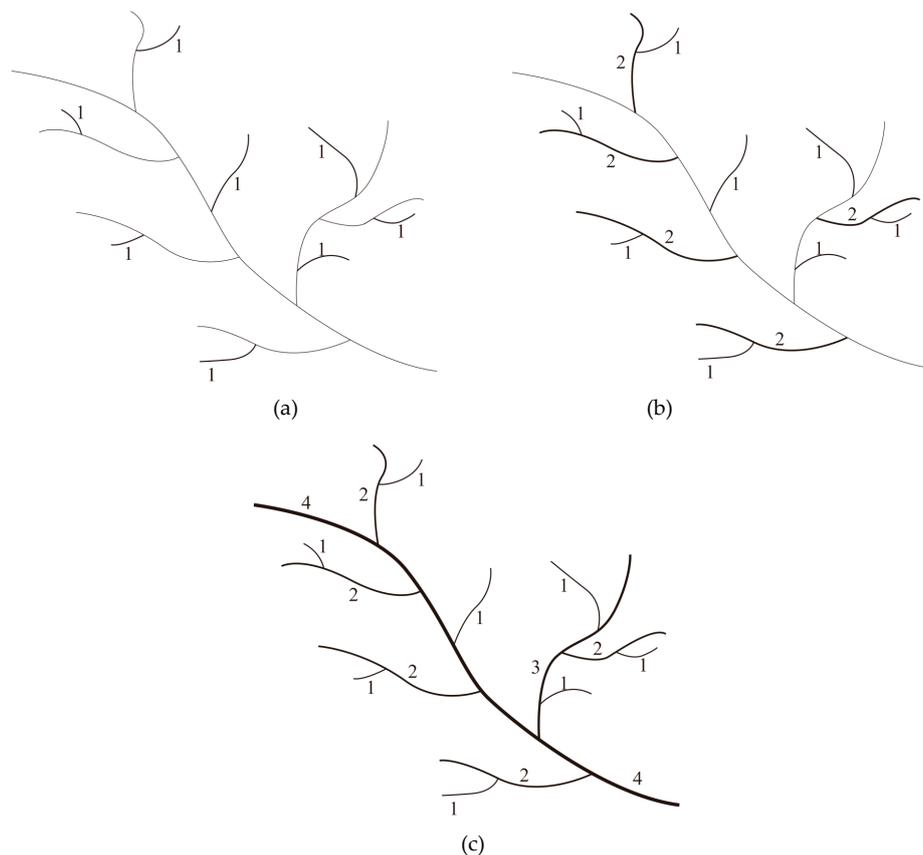
**Figure 4.** Inadequacies of bottom-up coding methods: three streams (AB, CD and EF) with different depths have the same code of two.

### 3. Selection Method for Dendritic River Networks Based on Hybrid Coding

A selection method for dendritic river networks based on hybrid coding is proposed, and it consists of three parts: (1) the dendritic river network is coded by classical top-down Horton coding; (2) directed topology trees are constructed to organize the river network data, while stroke connections are calculated to code the river network in the bottom-up approach; and (3) the river network is marked through hybrid usage of the top-down approach and bottom-up approach, after which a selection algorithm for dendritic river features based on hybrid coding is put forward.

#### 3.1. Top-Down Coding for Dendritic River Features

The first step of hybrid coding is to code the stream in the dendritic river network by top-down Horton coding. Terminal tributaries connected to river sources are assigned an order of 1. Based on the direction of each stream, streams that only contain first-order streams are defined as second-order streams and are assigned an order of 2. In this way, any stream containing  $n$ th-order streams has an order of  $n + 1$ . The main stem of a river system, which has the greatest number of tributaries, is defined as the highest-order stream ( $n$ th-order), as depicted in Figure 5. The detailed process of Horton coding has been described in existing research [6,9,19]; thus, it is not covered here.



**Figure 5.** Procedures of the automated coding method for dendritic river features: (a) coding the terminal tributaries, (b) coding of second-order streams and (c) final results of stream coding.

#### 3.2. Bottom-Up Coding for Dendritic River Features

In fact, rivers are usually stored as a collection of reach elements in original map data. Regarding selection, it is important that a stream be processed as a whole. To code a stream in a dendritic river network by the bottom-up approach, directed topology trees are constructed to organize the river network data, and stroke connections are calculated from estuaries to track the main stream in this paper [17].

### 3.2.1. Directed Topology Trees and Stroke Connections

Topological structural diagrams of dendritic river features that contain flow directions are also known as DTTs [20], which are collections of nodes and arcs that record information such as the degree, out-degree and in-degree of each node [21]. The direction of each arc (edge) is defined as the direction from the source node to the terminal node. In addition, DTTs also contain semantic (e.g., name and type) and geometric (e.g., length and width) stream information.

The recognition of main stems is an important step in the coding of river systems. Some scholars have proposed that river reaches could be connected to form river entities using stream lengths and basin areas [16]. However, these methods do not account for the flow direction characteristics of a stream. Thomson et al. [22,23] proposed that the principle of connectivity in Gestalt psychology could be used to construct strokes for river systems based on the semantics, geometries, topologies and directions of their streams. Excellent results have been obtained using this approach. In our method, the stroke connections of a dendritic river feature are iteratively calculated using DTTs based on the principles of semantic consistency, directional consistency and length prioritization (i.e., longer is better). The main stem is then identified on this basis. The procedures for the calculation of stroke connections and main stem identification are described below, with Figure 6 being used as an illustrative example:

Step 1: In most cases, the downstream reaches of a dendritic river feature only have one estuary. Therefore, we have chosen the estuary as the origin (Point O) from which the stroke connections are traced. The arc associated with the estuary is the tracing arc (Arc OP), and the other node of this arc (Point P) is the tracing node.

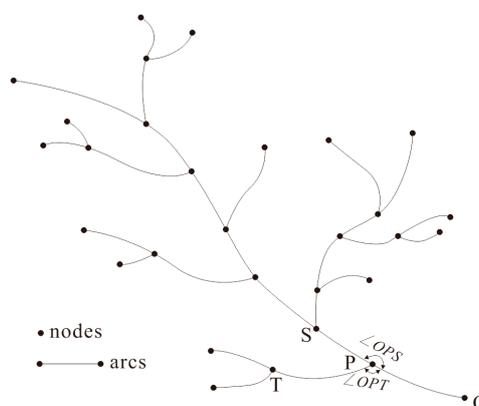
Step 2: The arcs associated with P are incorporated into the set of candidate arcs for stroke connection R {PS, PT}. The angles of these arcs with respect to OP (i.e.,  $\{\angle OPS, \angle OPT\}$ , respectively) are also calculated.

Step 3: The arc in set R that forms a stroke with OP is selected according to the principles of semantic consistency, directional consistency and length prioritization. In this case, S is selected to form OS, and S is now the tracing node.

Step 4: Stroke connections are continuously calculated and traced according to the methods of Steps 2 and 3 until the stream has been traced to its source. The resulting stroke is then the main stem of the estuary.

Step 5: All of the stroke connections to the arc associated with the river source are then calculated to obtain the tributaries of the main stem.

Step 6: All of the stroke connections associated with the aforementioned tributaries are calculated until all of the arcs in the DTT have been calculated. The calculation of stroke connections terminates at this point.



**Figure 6.** The construction of stroke connections: starting point O, tracing node P in the first stage, tracing node S and T in the second stage.

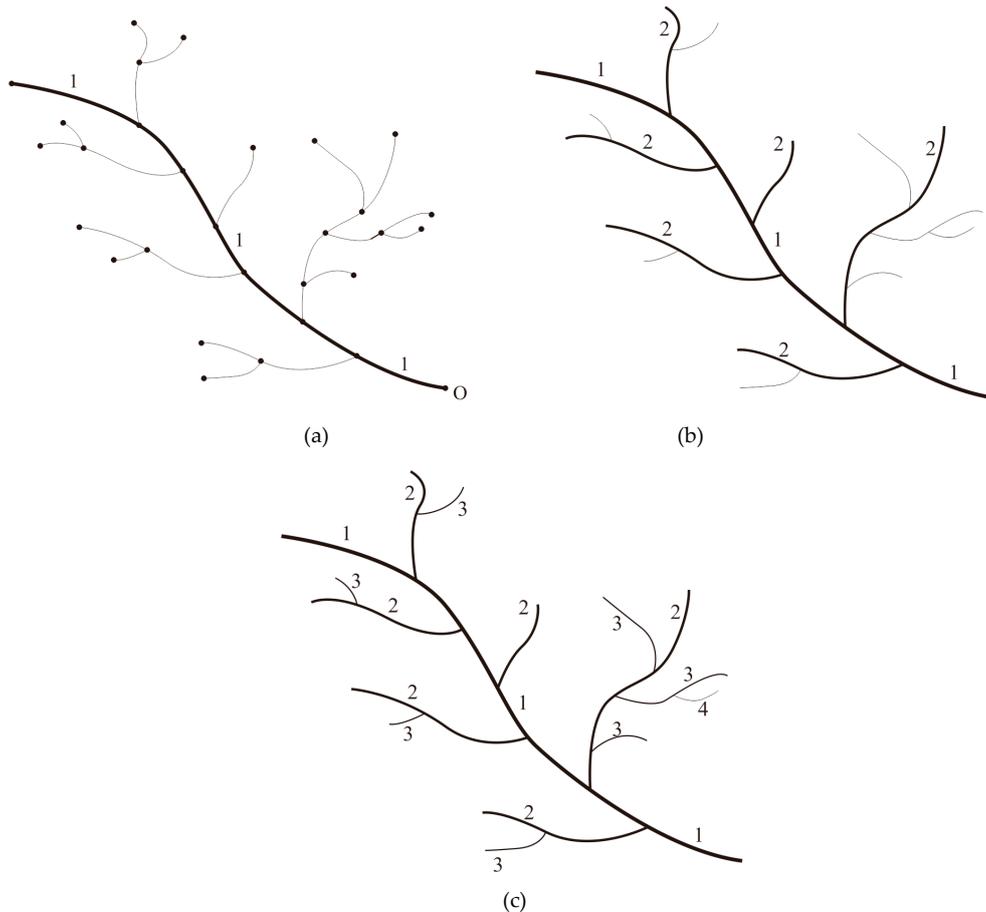
### 3.2.2. Procedures of Automated Bottom-Up Coding

The procedures of the automated coding method for dendritic river features are described below, with the river system shown in Figure 7 being used as an illustrative example:

Step 1: Stroke connections are calculated from the estuary, O, until the main stem of the river system has been calculated. The main stem is assigned a code (order) of 1, as shown in Figure 7a.

Step 2: All of the stroke connections associated with the arc of the main stem are calculated to obtain the tributaries of the main stem. These tributaries are assigned a code of 2, as shown in Figure 7b.

Step 3: All of the stroke connections to the tributary arcs are calculated and coded. The final results of the stream ordering process are shown in Figure 7c.

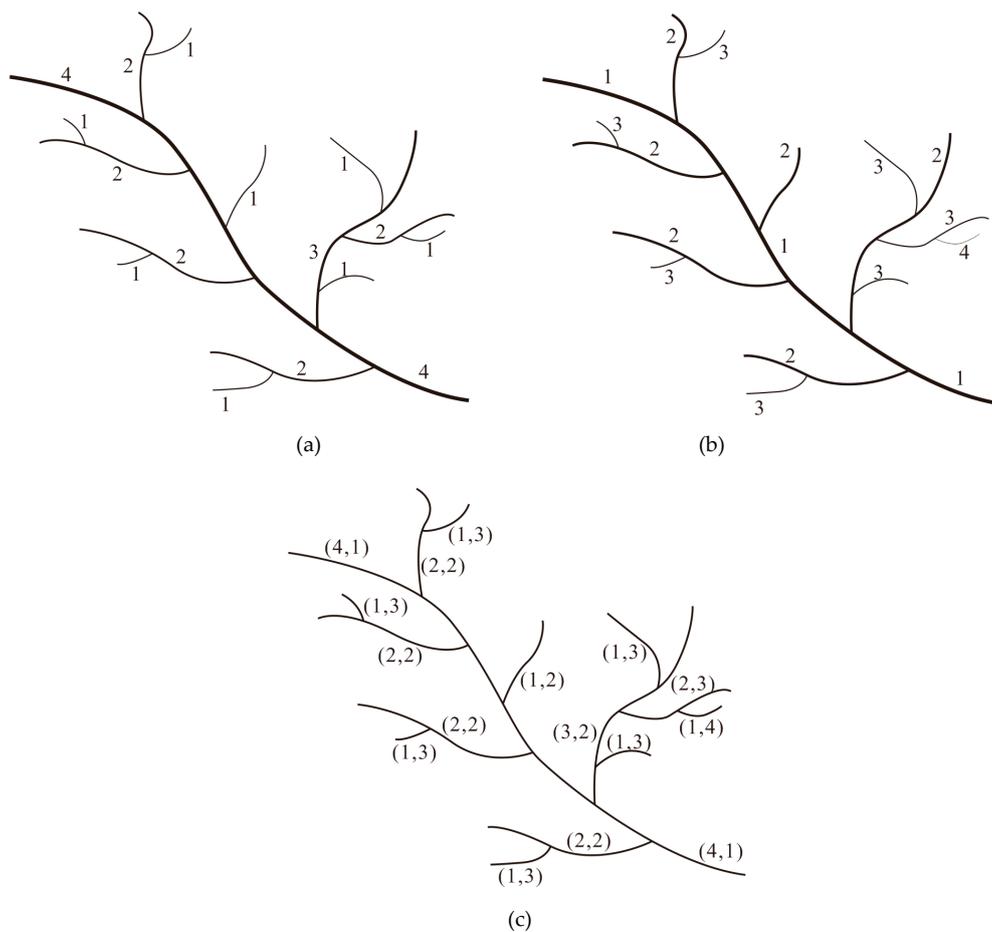


**Figure 7.** Procedures of the automated coding method for dendritic river features: (a) coding the main stem, (b) coding of tributaries and (c) final results of stream coding.

### 3.3. Selection Algorithm for Dendritic River Features Based on Hybrid Coding

#### 3.3.1. Hybrid Coding

By combining the advantages of both coding methods, hybrid coding is proposed to describe the spatial characteristic of river networks. The basic idea is that any stream in a river network is marked through hybrid usage of the top-down approach and bottom-up approach, and the coding structure is described as (top-down coding, bottom-up coding), as shown in Figure 8.



**Figure 8.** Three different coding methods: (a) top-down coding method, (b) bottom-up coding method and (c) hybrid coding method.

As can be seen from Figure 8, the deep structure and hierarchical relationship can be described at the same time by hybrid coding.

### 3.3.2. Determination of the Dichotomy Threshold of Each Stream

According to the analysis in Section 2.1, the information described by the two existing coding methods is different, which affects the judgment of the river network spatial structure. To make full use of the advantages of the two codes, the subtrees connected to the main stream of the river network are classified into different patterns.

First, the depth information of each stream (DS) is identified based on Horton's coding, and the depth threshold (DT) is calculated on the basis of the variance of the depth information distribution to classify the streams into two categories. If  $DS > DT$ , the stream is a deep stream, which means that the stream is very important for maintaining the overall depth characteristics of the river network, and if  $DS \leq DT$ , the stream is a shallow stream, which means that its importance in maintaining the overall depth characteristics is lesser.

The depth threshold (DT) is calculated by the Otsu's method [24], which is an effective algorithm for image segmentation and has been widely employed in various field applications. Let the depth distribution of a given river network be represented by  $L$  levels  $[1, 2, \dots, L]$ . The number of streams at level  $i$  is denoted by  $n_i$ , and the total number of streams is represented by  $N = n_1 + n_2 + \dots + n_L$ . Then, the probability distribution of each depth level is  $p_i = n_i / N$ .

Now, suppose that we dichotomize the depth into two classes,  $C_0$  and  $C_1$ , by a threshold at level  $k$ .  $C_0$  denotes the depths of levels  $[1, \dots, k]$ , and  $C_1$  denotes the depths of levels  $[k + 1, \dots, L]$ . Then the between-class variance is given by [24]:

$$\sigma_B^2 = \omega_0(u_0 - u_T)^2 + \omega_1(u_1 - u_T)^2 = \omega_0\omega_1(u_1 - u_0)^2 = \frac{[u_T\omega(k) - \mu(k)]^2}{\omega(k)(1 - \omega(k))} \quad (1)$$

where  $\omega_0$ ,  $\omega_1$ ,  $u_0$ , and  $u_1$  are the probabilities of class occurrence and the class mean levels, respectively,

$$\omega(k) = \omega_0, u_T = \sum_{i=1}^L ip_i, \text{ and } \mu(k) = \sum_{i=1}^k ip_i.$$

The optimal threshold  $k^*$  is [24]:

$$\sigma_B^2(k^*) = \max_{1 \leq k < L} \sigma_B^2(k) \quad (2)$$

### 3.3.3. Classification of Dendritic River Networks

After the dichotomy of all streams is determined, each subtree connected to the main stream is classified into one of three patterns based on the spatial distribution and connection relationship of the two kinds of streams, as follows: (1) deep branch, which consists of deep streams and the shallow streams contained by these deep streams; (2) shallow branch, which consists of shallow streams or only of a deep stream and the shallow streams contained by this stream; and (3) modest branch, which consists of shallow streams without depth differences or among which the difference is 1. During the generalization process, the deep branches are coded using the top-down approach, the shallow branches are coded via the bottom-up approach, and the modest branches are coded by one of the aforementioned two approaches.

The classification is carried out based on the hybrid coding of each stream. The subtrees connected to the main stream can be extracted through bottom-up coding, and these streams are coded 2. Top-down coding is used for comparison with DT to classify the streams into the deep or shallow category, and subtrees are classified into different patterns based on the aforementioned classification rules.

### 3.3.4. Selection Algorithm

River selections are performed using the classic selection method proposed by Stanislawski et al. [11]. The detailed procedure is as follows:

Step 1: The number of streams that should be retained is determined using the ‘Radical Law’ model [25,26],  $n_f = n_a \sqrt{M_a/M_f}$ , where  $n_f$  is the number of objects that can be shown at the target scale,  $n_a$  is the number of objects shown on the original map, and  $M_a$  and  $M_f$  are the scale denominators of the original and target maps, respectively.

Step 2: The river network is hybrid coded to describe the depth and hierarchy information of each stream, and then the depth threshold (DT) is calculated by Otsu’s method.

Step 3: The streams are classified into deep or shallow types, and the subtrees of the main stream are classified into deep, shallow and modest patterns based on the spatial characteristic and connection relationships of the streams.

Step 4: The deep branches are coded using the top-down approach, the shallow branches are coded via the bottom-up approach, and the modest branches are classified by one of the aforementioned two approaches to prepare the selection.

Step 5: The selection numbers are allocated unequally to each subtree based on the number of streams,  $n_i = n_f \times N_i/N_t$ , where  $n_i$  is the number of streams shown at subtree  $i$ ,  $n_f$  is the total number of streams shown,  $N_i$  is the number of streams at subtree  $i$ , and  $N_t$  is the total number of streams across the whole river network.

Step 6: The selective principle is to preserve main streams and remove the tributaries. Specifically, for a subtree, selection starts from the main stream of the subtree and extends to the tributaries. Streams in the same order are selected according to their length information, and the longer streams are preserved.

Step 7: Because of the special structure of the modest branches with depth code 1, they are placed together and selected as a whole.

Step 8: Steps 6 and 7 are repeated until a sufficient number of streams (as defined by the square root model) have been selected, and then the selection algorithm ends.

## 4. Experiments and Analysis

### 4.1. Experimental Data and Environment

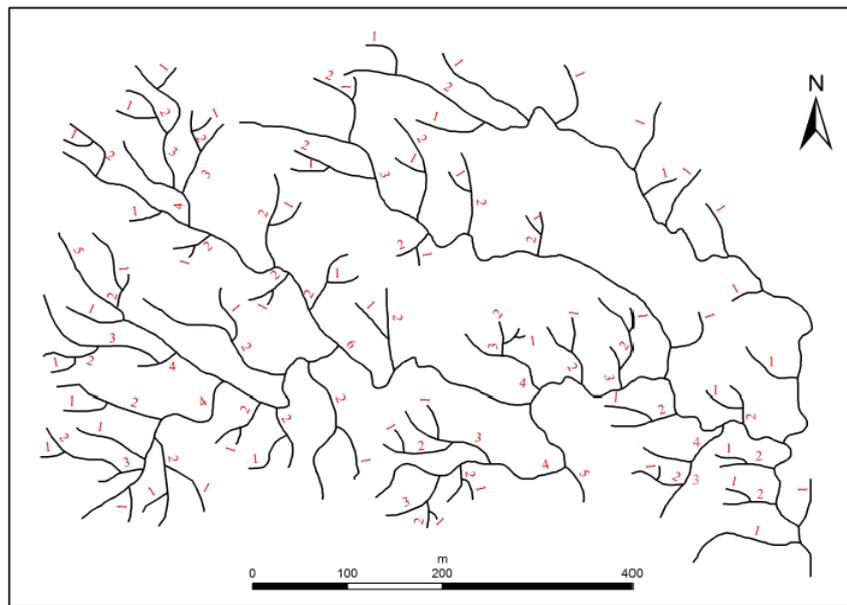
The hybrid coding method was embedded into the WJ-III mapping workstation developed by the Chinese Academy of Surveying and Mapping. The new stream coding method was experimentally validated via a comparison with the classic top-down coding method and bottom-up method. The experimental data correspond to part of a dendritic river feature in a city of Hubei Province in China. The original scale of these data is 1:100,000, and there is a total of 108 rivers in this dataset. The storage space of the experimental data is 55 KB. The hardware environment for this experiment runs on a Microsoft Windows 7 64-bit operating system equipped with an Intel Core i7-3770 running at 3.2 GHz, with 16 GB of RAM and a 1024 GB solid-state drive.

The experiment included two parts: first, coding the rivers in the same experimental area using three different coding methods to analyze their differences and characteristics; second, comparing the selection effect under three different coding methods via classifying of river subtrees in three typical areas.

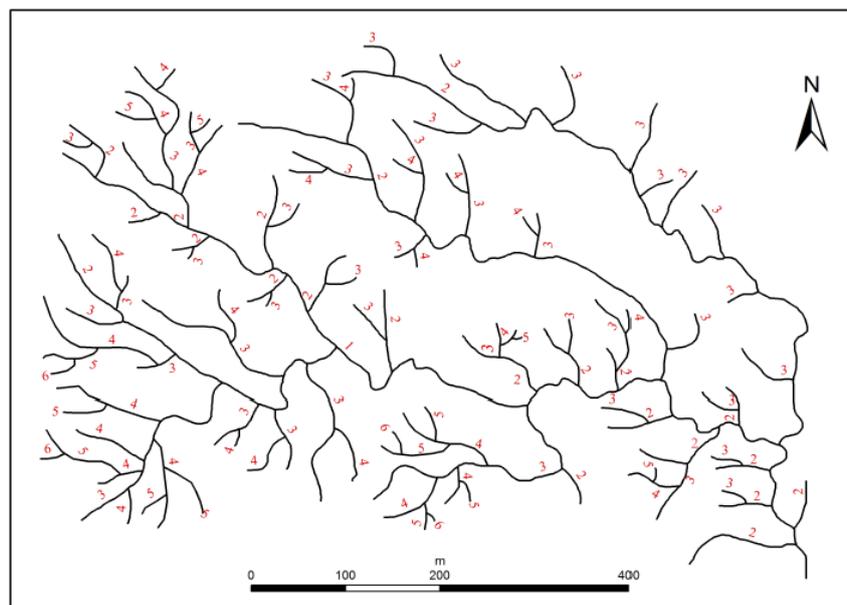
The time consumed by the selection method based on river coding mainly includes three parts: data preparation time, river coding time and river selection time. For the same experimental area, the total time to complete the above three operations for each classical coding system is 8 s, and the total time based on the proposed method is 10 s.

### 4.2. Comparative Analysis of Each Coding Method

Figure 9 illustrates the coding results of the experimental river network obtained using three different coding methods, including (a) the top-down Horton's coding method, (b) the bottom-up coding method and (c) the hybrid coding method. As depicted in Figure 9, the ranges of the stream orders given by these three methods are identical, i.e., from 1 to 6. However, the spatial structural characteristics that they describe are different. (1) Horton coding coded the river network from the source to the estuary, so the depth information was recorded. For example, a stream that has a Horton stream order of 3 must contain a bifurcation of first-order and second-order tributaries, and the subtree depth of this stream is 3. (2) The bottom-up approach coded the river network from the estuary to the source, so the hierarchy formed by catchment relationships was recorded. For example, a stream that has a code of 2 must connect to the main stream, which is coded by 1. (3) The hybrid coding method combined the advantages of the aforementioned two coding methods, and the depth and hierarchy information of a stream can be read at the same time.

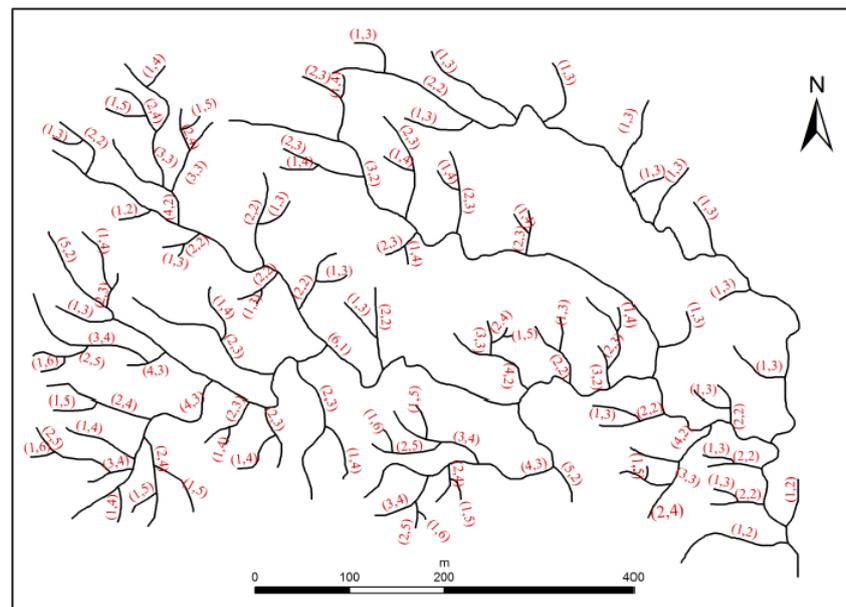


(a)



(b)

Figure 9. Cont.



(c)

**Figure 9.** Comparison of the stream coding systems: (a) top-down coding method, (b) bottom-up coding method and (c) hybrid coding method.

To further analyze the characteristics and differences of different coding methods, the number of streams corresponding to each stream order is shown in Table 1.

**Table 1.** Comparison analysis between different coding methods.

Stream Coding Method	Number and Percentage of Streams for Each Stream Order							
	Order	1	2	3	4	5	6	Total
Top-down coding	Code	6	5	4	3	2	1	/
	Number of streams	1	2	6	10	35	54	108
	Percentage (%)	0.93	1.85	5.56	9.26	32.40	50.00	100
Bottom-up coding	Code	1	2	3	4	5	6	/
	Number of streams	1	22	42	26	13	4	108
	Percentage (%)	0.93	20.37	38.89	24.07	12.04	3.70	100
Cross-check analysis	Number of streams with inverted code	1	2	3	4	4	4	18
	Percentage (%)	0.93	1.84	2.78	3.70	3.70	3.70	16.67
	Number of streams with non-inverted code	0	0	3	6	31	50	90
	Percentage (%)	0	0	2.78	5.56	28.70	46.29	83.33

Streams with inverted codes refer to the stream whose codes by two different coding systems were added together, and the result is 7. Otherwise, the stream is with non-inverted codes.

As depicted in Table 1, although the coding ranges of these two coding methods are identical, there are significant differences in the number of streams corresponding to each stream order. In the Horton method, the number of streams decreases with the stream order. Rivers with codes 1 and 2 at the boundary account for 82.4%, which means that the importance of these streams can be distinguished accurately by the coding. The distribution of streams corresponding to our coding method exhibits an n-type pattern, i.e., lower- and higher-order streams are fewer in number, whereas medium-order streams are greater in number, which makes it easy to judge the importance of the streams around the outside of river network but difficult to judge the streams in the middle.

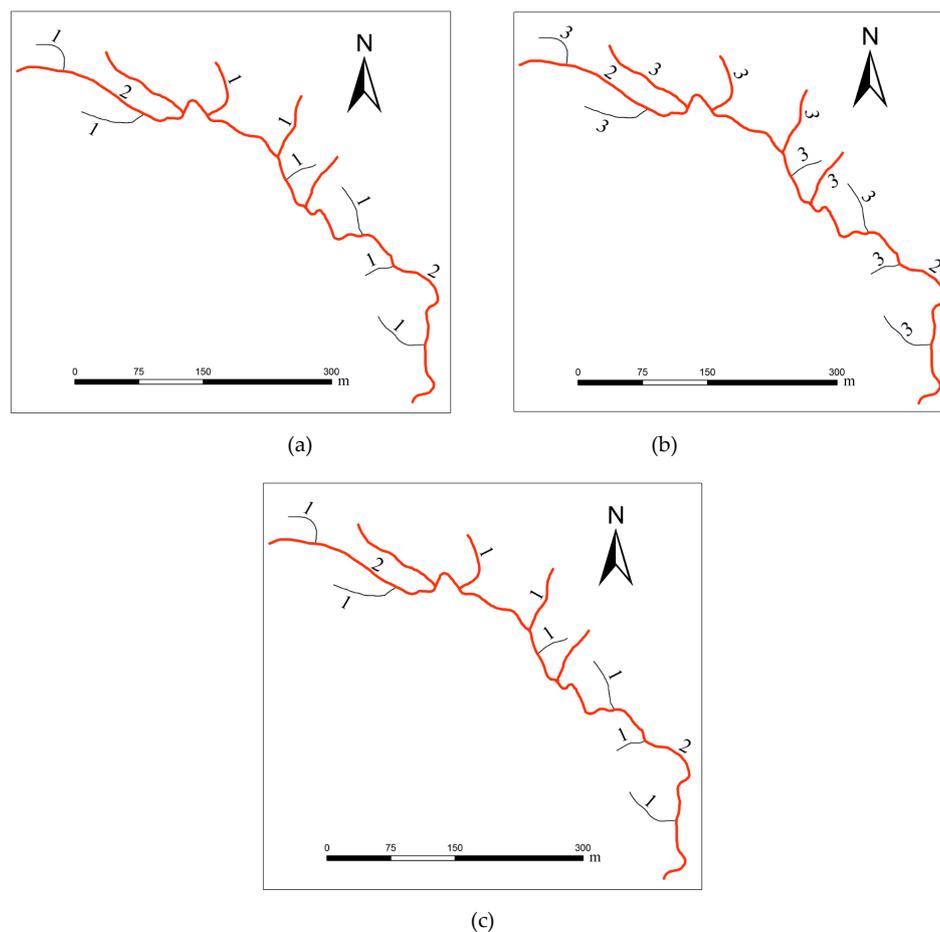
In addition, a cross-check analysis was carried out in this paper. Taking Horton's coding as the reference code, the numbers of streams with inverted codes and non-inverted codes in the bottom-up

coding were recorded. As depicted in Table 1, only 16.67% of the rivers are streams with inverted codes, which means that the comprehensive utilization of the two methods can't be achieved through simple algebraic transformation but requires an in-depth analysis of the functions of the two methods in expressing different spatial structures.

#### 4.3. Analyzing the Efficacy of Each Coding System in River Selection and Generalization

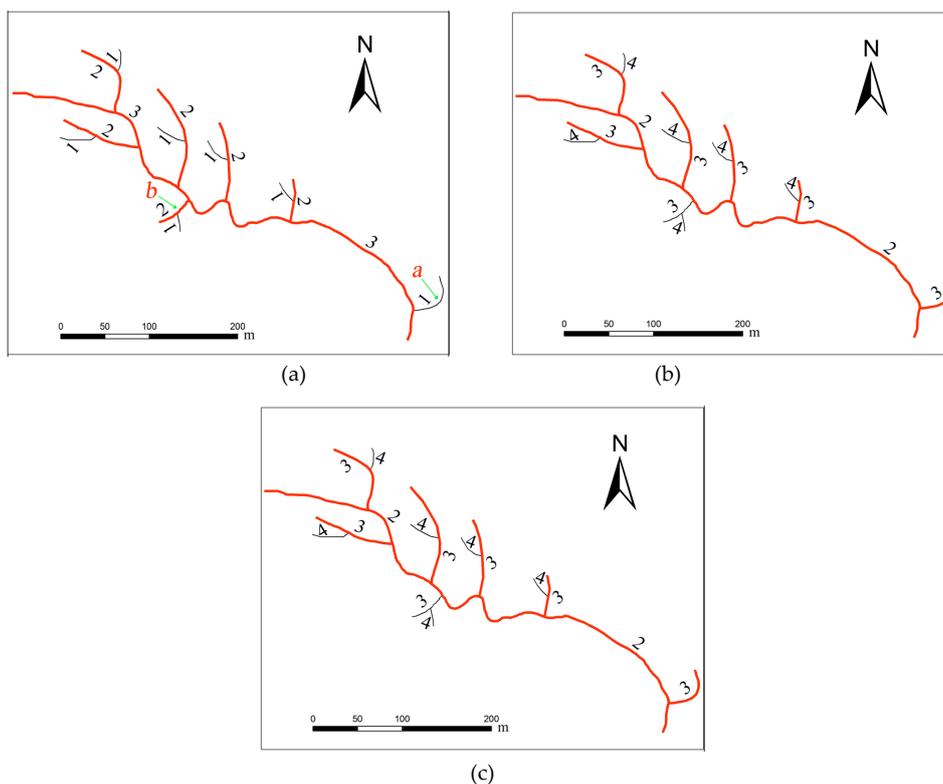
The target scale of river network generalization is 1:500,000. Therefore, 52 streams can be shown at the target scale based on the square root model. In addition, according to Otsu's method [24], for each depth level, the between-class variance was calculated. The results were as follows.  $\sigma_B^2(1) = 206.21$ ,  $\sigma_B^2(2) = 222.25$ ,  $\sigma_B^2(3) = 112.53$ ,  $\sigma_B^2(4) = 41.84$ ,  $\sigma_B^2(5) = 14.59$ , and  $\sigma_B^2(2)$  was the max value, so the depth threshold (DT) was set to 2. On the basis of the DT, 6 deep branches, 12 shallow branches and 4 modest branches were obtained.

For each pattern, the subtrees with the largest numbers of streams are selected as typical areas for visual analysis of the selection results, as shown in Figures 10–12. Figure 10 shows a modest branch, in which all of the tributaries connected to the main stem of the subtree have the same topological depth. Figure 11 shows a shallow branch with uniformly distributed tributaries, in which the tributaries of the main stem are relatively shallow in terms of topological depth, and their depths do not vary significantly. Figure 12 shows a topologically deep branch, in which the tributaries connected to the main stem have very large differences in topological depth.



**Figure 10.** Comparison of the selections in a modest branch: (a) selection result based on the top-down coding method, (b) selection result based on the bottom-up coding method and (c) selection result based on the hybrid coding method (the selection results are shown as red lines, and the original rivers are shown as black lines).

There are 11 streams in the modest branch shown in Figure 10. Based on the square root model, 5 streams need to be preserved, while 6 streams have to be removed. The spatial distribution characteristics of streams show that the hierarchies and depths of each tributary are the same, except for the main stream in this branch; hence, their codes have no difference, no matter which coding method is used. During the generalization process, Horton's coding was selected as the coding method in this paper. The selection results for each method are shown in Figure 10a–c, and the three ordering methods resulted in the same selection results in this scenario.

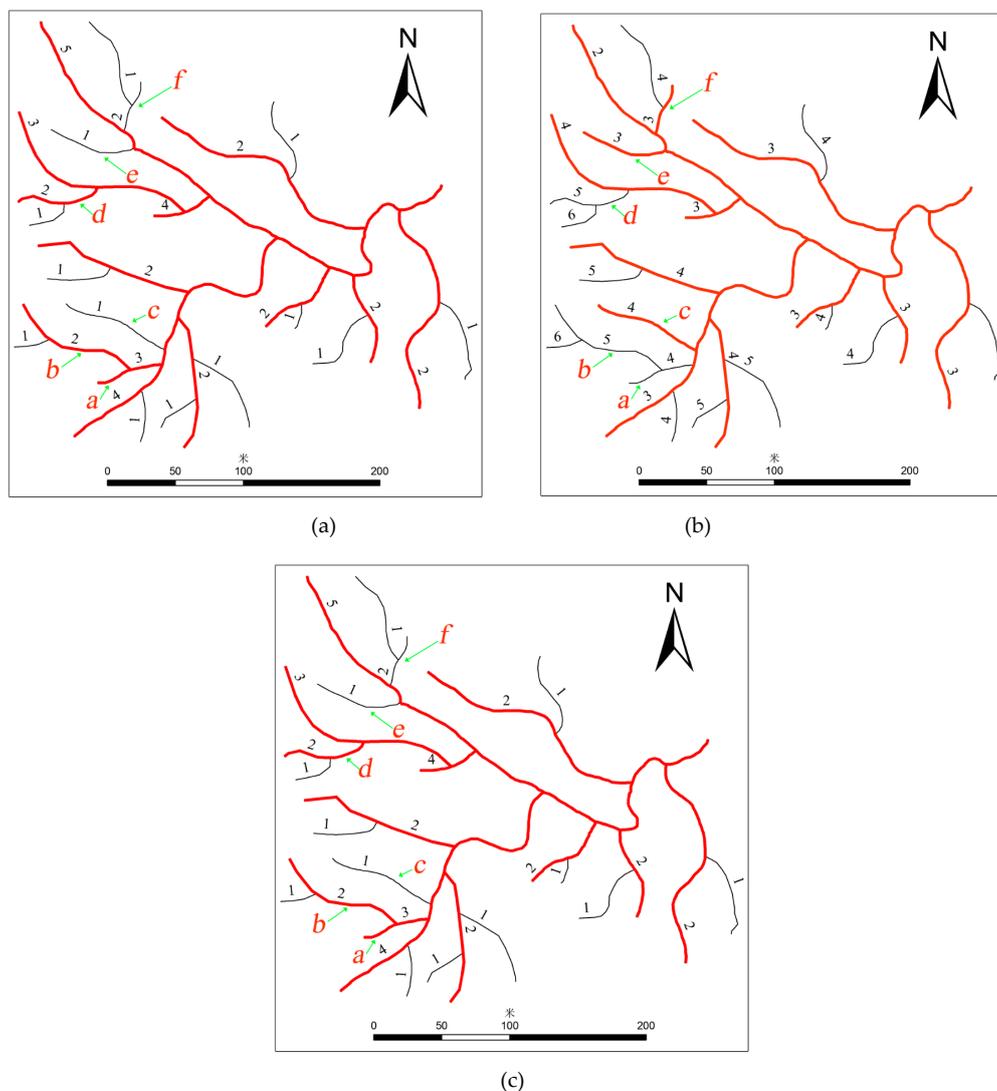


**Figure 11.** Comparison of the selections in a shallow branch: (a) selection result based on the top-down coding method, (b) selection result based on bottom-up coding method and (c) selections based on the hybrid coding method (the selection results are shown as red lines, and the original rivers are shown as black lines).

Figure 11 shows that there are 13 streams in the shallow branch. According to the square root model, 6 streams are retained, while 7 streams need to be removed. These streams are evenly uniformly distributed around the main stream of this branch, but the hierarchies and depths of each tributary have slight differences. For this pattern, bottom-up coding was selected as the coding method in this paper. The selection results for each method are shown in Figure 11a–c. The results obtained using Horton's coding method retain stream b and remove stream a. Conversely, the results obtained using bottom-up coding remove stream b and retain stream a. As can be seen from Figure 11, although stream b has a depth of two, the depth structure in this branch is as not important as the hierarchy structure, which can better reflect the main spatial structure of this branch, as shown in stream a. Hence, the bottom-up coding method and the proposed method can efficiently preserve the spatial structures of the original map.

There are 27 streams in the deep branch shown in Figure 12. Based on the square root model, 13 streams have to be retained, while 14 streams have to be removed. These rivers form a deeply nested structure, and the hierarchies and depths of each tributary have obvious differences. For this pattern, top-down coding was selected as the coding method in this paper. The selection results for each method are shown in Figure 12a–c. The results obtained using Horton's coding method retain

streams a, b and d and remove streams c, e and f. Conversely, the results obtained using bottom-up coding remove streams a, b and d and retain streams c, e and f. As can be seen from Figure 12, streams a, b and d with deeper depth structures are more important than are streams c, e and f in reflecting the main spatial structure of this branch. Removing streams a and b leads to the obvious loss of the tree structure. Hence, the top-down coding method and the proposed method can efficiently preserve the spatial structure of the original map.



**Figure 12.** Comparison of the selections in a deep branch: (a) selection result based on the top-down coding method, (b) selection result based on bottom-up coding method and (c) selection result based on the hybrid coding method (the selection results are shown as red lines, and the original rivers are shown as black lines).

## 5. Conclusions

River coding plays an important role in river network selection for topographic map generalization. Two existing classical coding methods show different characteristics—the top-down coding method is not appropriate for the stream selection of river networks with uniform distributions of tributaries, and the bottom-up coding method is not favorable for the stream selection of river systems that have topologically deep structures. The aforementioned two coding systems were used individually on the traditional river selection method, which leads to that the selection results of some river subtrees can't keep the original spatial distribution characteristics. Therefore, a selection method for dendritic

river networks based on hybrid coding is proposed in this paper, which combines the advantages of the above two methods and can preserve the spatial structural characteristics of river networks well by selecting the appropriate coding mode according to the spatial structure of the river subtrees. The following conclusions were drawn from the experimental validation using actual data:

(1) Only 16.67% of the rivers are inverted-coded, which means that the comprehensive utilization of the two methods can't be achieved through simple algebraic transformation. The hybrid coding method proposed in this paper is effective, and the depth and hierarchy information of a stream can be read at the same time.

(2) The spatial distribution structure of river networks is determined based on the hybrid coding. First, streams were classified into deep or shallow streams based on the top-down coding approach, which reflects the depth information; second, river subtrees were identified based on the bottom-up coding approach, which reflects the hierarchy information; finally, the classified into three kinds of patterns based on the distribution characteristics and connection relationship of streams, which is deep branch, shallow branch and modest branch.

(3) For different patterns, the appropriate coding mode was selected automatically in this paper, which ensured that the selection results of different spatial structures was reasonable. For the modest branch, one of two coding approaches was used arbitrarily, because the selection result was the same. For the shallow branch and deep branch, the bottom-up coding method and the top-down coding method were selected, respectively.

The method proposed in this paper can be directly used in the river map generalization and also be helpful in the determination of river catchment regions. Future research will focus on (1) the generality of the proposed method in river selection of multi-scale and large-spatial range river network datasets; (2) exploring widely the applications of the proposed method in fields, such as hydrographic analyses, flood, retention activities, etc. At present, the proposed method embedded in WJ-III map station has been used in five provinces in China to perform multi-scale representation of river networks data in the National Geographic Census, such as Guizhou, Gansu, Guangdong, etc. The research results in river selection of multi-scale and large-spatial range river network datasets need further summarization.

**Author Contributions:** Chengming Li conceived the original idea for the study; all co-authors conceived and designed the methodology; Wei Wu, Pengda Wu, Yong Yin and Zhaoxin Dai conducted processing and analysis of the data; Chengming Li and Pengda Wu drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a project supported by the National Natural Science Foundation of China, (Grant No. 41871375), and Basal Research Fund of CASM, (Grant Nos. AR 2008/ 2009). We are grateful to the National Geomatics Center of China for providing data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Battenfield, B.P.; Stanislawski, L.V.; Anderson-Tarver, C.; Gleason, M.J. Alternate methods for automatic selection of primary paths through braided hydrographic networks. In Proceedings of the 16th ICA Workshop on Generalization & Multiple Representations, Dresden, Germany, 23–24 August 2013.
2. Tarboton, D.G.; Bras, R.L.; Rodriguez-Iturbe, I. The fractal nature of river networks. *Water Resour. Res.* **1988**, *24*, 1317–1322. [[CrossRef](#)]
3. He, Z. *Principle and Method of Map Data Processing Model*; Wuhan University Press: Wuhan, China, 2004. (In Chinese)
4. Zhang, L.; Wang, G.Q.; Dai, B.X.; Li, T.J. Classification and codification methods of stream network in a river basin, a review. *Environ. Inform. Arch.* **2007**, *5*, 364–372.
5. Gülgén, F. A stream ordering approach based on network analysis operations. *Geocarto Int.* **2017**, *32*, 322–333. [[CrossRef](#)]
6. Horton, R.E. Erosional development of streams and their drainage basins; hydro physical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **1945**, *56*, 275–370. [[CrossRef](#)]

7. Moharir, K.N.; Pande, C.B. Analysis of morphometric parameters using Remote-sensing and GIS techniques in the lonar nala in Akola district Maharashtra India. *Int. J. Tech. Res. Eng.* **2014**, *1*, 1034–1040.
8. Harish, N.; Kumar, P.S.; Raja, M.S.; Lokesh, V.; Reddy, M.G.S.; Shalisad, S.; Sazid, S. Remote sensing and GIS in the morphometric analysis of macro-watersheds for hydrological Scenario assessment and characterization—A study on Penna river sub-basin. SPSR Nellore district, India. *Int. Res. J. Eng. Technol.* **2016**, *3*, 1579–1583.
9. Sen, A.; Gokgoz, T. An experimental approach for selection/elimination in stream network generalization using support vector machines. *Geocarto Int.* **2015**, *30*, 311–329. [[CrossRef](#)]
10. Strahler, A.N. Quantitative analysis of watershed geomorphology. *Eos Trans. Am. Geophys. Union* **1957**, *38*, 913–920. [[CrossRef](#)]
11. Stanislawski, L.V.; Savino, S. Pruning of hydrographic networks: A comparison of two approaches. In Proceedings of the 14th ICA Workshop on Generalization and Multiple Representation, Jointly Organized with ISPRS Commission II/2 Working Group on Multiscale Representation of Spatial Data, Paris, France, 3–8 July 2011.
12. Shreve, R.L. Statistical law of stream numbers. *J. Geol.* **1966**, *74*, 17–37. [[CrossRef](#)]
13. Horsfield, K.; Cumming, G. Morphology of the bronchial tree in man. *J. Appl. Physiol.* **1968**, *24*, 373–383. [[CrossRef](#)] [[PubMed](#)]
14. Gravelius, H. *Flusskunde*; Goschen'sche Verlagshandlung: Berlin, Germany, 1914.
15. Hack, J. Studies of longitudinal stream profiles in Virginia and Maryland. *U.S. Geol. Surv. Prof. Pap.* **1957**, *294B*, 45–95.
16. Jasiewicz, J.L.; Metz, M. A new GRASS GIS toolkit for Hortonian analysis of drainage networks. *Comput. Geosci.* **2011**, *37*, 1162–1173. [[CrossRef](#)]
17. Dai, Z.X.; Li, C.M.; Wu, P.D. An automated bottom up coding system for dendritic river system. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-4/W16*, 169–175.
18. Marani, A.; Rigon, R.; Rinaldo, A. A note on fractal channel network. *Water Resour. Res.* **1991**, *27*, 3041–3049. [[CrossRef](#)]
19. Whitaker, S.; Stanislawski, L.; Hamann, M. Automated stream leveling for the high-resolution national hydrography dataset. In Proceedings of the 22nd Annual ESRI International User Conference, San Diego, CA, USA, 8–12 July 2002; pp. 8–12.
20. Wu, J.; Deng, M.; Liu, H. An Integrated Model to Represent Topological Relation and Directional Relation Between Directed Line Objects. *Geomat. Inf. Sci. Wuhan Univ.* **2013**, *11*, 1358–1363. (In Chinese)
21. Zhang, Y.; Li, L.; Jin, Y.; Zhu, H.H. Structured Design of Dendritic River Networks Based on Graph. *Geomat. Inf. Sci. Wuhan Univ.* **2004**, *29*, 537–539. (In Chinese)
22. Thomson, R.C.; Brooks, R. Efficient generalization and abstraction of network data using perceptual grouping. In Proceedings of the 5th International Conference on Geo-Computation, Chatham, UK, 23–25 August 2000; pp. 23–25.
23. Thomson, R.C.; Brooks, R. Exploiting perceptual grouping for map analysis, understanding and generalization: The Case of Road and River Networks. In *International Workshop on Graphics Recognition*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 148–157.
24. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
25. Töpfer, F. *Study of the Application of the Principle of Square Root in Cartographic Generalization*; Surveying and Mapping Press: Beijing, China, 1963.
26. Töpfer, F.; Pillewizer, W. The principles of selection, a means of cartographic generalization. *Cartogr. J.* **1966**, *3*, 10–16. [[CrossRef](#)]

