

Ni et al., 1998; Qian et al., 1987; Singh, 2000), because of its feasibility of determining the conflux order. TOPAZ, a popular program that extracts river network from the DEM data, automatically gives the Horton-Strahler order of each stream. There were deeper developments in the river network research due to the coming of the random river system models 20 years later. The researchers postulated that natural networks were random in the absence of strong geologic controls and proposed several stochastic models, such as the models that Shreve and Smart gave. They used morphometric parameters concerning topologic and length properties when doing researches on the configuration of river network.

With applications of GIS and RS techniques to water resource management, more and more researchers were involved in developing distributed hydrological model to simulate runoff and stream flow for watershed. GIS tools were used to facilitate to divide the study area into small sub-catchment units based on digital elevation model (DEM) information. It stimulated the development of classification and codification methods for stream networks in a large scale river basin. Pfafstetter (1989) proposed the Pfafstetter codification rules which were used by Yang et al. (2004) to develop a distributed hydrologic model. Luo improved the Pfafstetter codification method and applied it to the Yellow River basin (Luo et al., 2003; Luo et al., 2006). The results supplied important information to further develop a distributed hydrologic model for the whole Yellow River basin. Wang et al. (2005) developed the digital watershed model to simulate rain-runoff and soil erosion processes in a large scale river basin. Correspondingly, a new classification and codification method based on binary tree was proposed by Li et al. (2006) to facilitate the parallel computation of the developed digital watershed model for the Yellow River basin.

All application results indicated that classification and codification methods played an important role in both river networks delineating and distributed hydrologic modeling. Identification of the stream order could help in-depth understand the structure of the drainage system in a river basin. Moreover, various classification and codification methods could directly result in different computation efficiencies of a distributed hydrologic model, especially for a large scale river basin. Each classification and codification method has its own advantages and shortcomings so that it might be effective to one certain issue, but improper to other situations. Therefore, a comprehensive over review of existing classification and codification methods is desired to summarize their characteristics for supplying decision support for classification and codification method selection. In this paper, four classification and codification methods which were popularly used will be introduced and their advantages and disadvantages will be summarized. It will supply useful references to researchers who are engaging in hydrologic modeling.

2. Horton-Strahler Classification Method

2.1 Classification Rules

Delineating the geometrical morphology is the key concern of developing classification and codification methods of river networks. One of the most popular classification methods was the Horton-Strahler classification method which was modified by Strahler in 1952 (Knighton, 1984; Ni et al., 1998; Qian et al., 1987; Singh, 2000). He gave out the definition of drainage density and developed the classification method that was different from Gravelius' way. The classification rules are as follows (show in Figure 1):

- (a) the tributary which originates at a source and has no tributary injecting is designated as order 1;
- (b) the junction of two streams with same order u is designated as order $u+1$;
- (c) the junction of two streams with unequal order u and v , where $u < v$, is designated as order $\max(u, v)$.

The function of situation (b) and (c) can be described as:

$$n = \max(u, v) + \delta_{u,v}, \quad \text{where } \delta_{u,v} = \begin{cases} 1, & u = v \\ 0, & u \neq v \end{cases};$$

where n is the order of the stream formed by two streams with unequal order u and v .

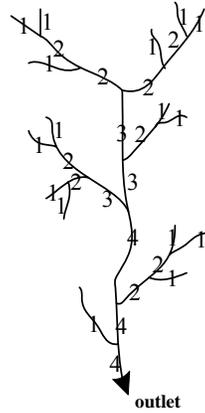


Figure 1 Channel segments ordered by Horton-Strahler

2.2 Geometric Laws

The characteristic factors of the water system can be described quantitatively at the view of geometry according to the stream order. The main morphological characteristic parameters of the water system are presented in Table 1.

Table 1 Main parameters of the water system

Parameter	Symbol	Remark
Stream number of order u	N_u	
Drainage density	$D_u = \frac{\sum L_u}{A_u}$	$\sum L_u$: the total length of the streams ordered u A_u : the drainage area of the stream ordered u
Bifurcation ratio	$R_b = \frac{N_u}{N_{u+1}}$	
Stream length ratio	$R_L = \frac{\overline{L}_u}{\overline{L}_{u-1}}$	\overline{L}_u : the average length of streams ordered u
Drainage area ratio	$R_A = \frac{\overline{A}_u}{\overline{A}_{u-1}}$	\overline{A}_u : the average drainage area of streams ordered u

Several geometric laws relevant to these parameters were proposed by Horton and Strahler as follows:

(1) Law of stream numbers

$$N_u = R_b^{(\Omega-u)} \quad (1)$$

where N_u is the stream number of order u, R_b is the bifurcation ratio, Ω is the highest order of the river network. Equation (1) indicates that the stream order increases while the stream number decreases.

(2) Law of stream lengths

$$\overline{L}_u = \overline{L}_1 R_L^{(u-1)} \quad (2)$$

where \overline{L}_u is the average length of streams ordered u;

(3) Law of drainage areas

$$\overline{A}_u = \overline{A}_1 R_A^{(u-1)} \quad (3)$$

where \overline{A}_u is the average drainage area of streams ordered u.

2.3 Virtues and Shortcomings

The method proposed by Horton and Strahler could provide quantitative results (Knighton, 1984), that making this method became an important foundation of river network research. The method facilitated the development of river geomorphology. The Horton-Strahler classification method has been widely used for analyzing watershed structures to help develop various hydrological models. However, limitations exist as follows:

- (1) it cannot distinguish the mainstream from two streams with the same order;
- (2) if a lower-order stream enters into a high-order stream, it will only result in changing of both drainage area and flow discharge, but not the order of the high-order stream;
- (3) the law of stream numbers is actually the inevitable result of the random development of water under the condition of gravity. R_b , R_L and R_A are profoundly indifferent to network structure, and these ratios are insensitive to marked changes in network structure (David et al., 1996; Krichner et al., 1993; Qian et al., 1987).

3. Shreve-Smart Classification Method

Shreve and Smart (Krichner et al., 1993; Knighton, 1984; Qian et al., 1987) thought that the evolvement of a natural stream network was a random process with certain statistical laws. They proposed a classification method for stream networks, which is called Shreve-Smart classification method. Based on the proposed method, they developed two models for tackling the randomness and regularity characteristics of stream network structures through using the broader-based probabilistic-topological methods (David et al., 1996; Krichner et al., 1993; Qian et al., 1987).

3.1 Morphometric Parameters

Three types of river network nodes were defined as outlet, source and junction in the Shreve-Smart classification method (Knighton, 1984; Ni et al., 1998; Qian et al., 1987; Singh, 2000). Outlet was defined as the most downstream node in the river network. All those origin nodes of a stream network were defined as sources of the network. The node where two streams converged was defined as the junction.

The basic unit of a stream network is link, defined as an unbroken section of channel between successive nodes (sources, junctions or outlet). A link which connects a source with the nearest downstream junction was defined as an exterior link. It corresponds to a stream whose Horton-Strahler order is 1. A link which connects two neighborhood junctions or the outlet with the nearest junction was defined as an interior link. The link magnitude of an exterior link was set as 1. The junction within two adjacent links with magnitude M_1 and M_2 respectively, would have the link magnitude of $M_1 + M_2$ (as shown in Figure 2). With this definition, the whole stream networks could then be classified. The number of the streams with link magnitude M is denoted by N_M .

A stream network with N_1 sources contains N_1 exterior links and $N_1 - 1$ interior links, and the magnitude of a link which connects the outlet and the adjacent junction is N_1 .

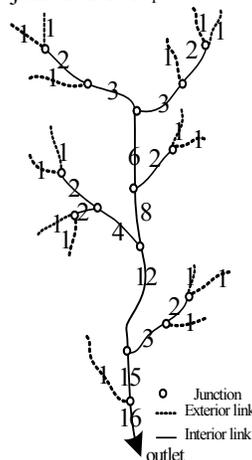


Figure 2 Channel segments ordered by Shreve-Smart

3.2 Random Models for determining a stream network structure

Several rules were found through the statistic and two random models were proposed by Shreve and Smart by using the morphometric parameters given above.

- (1) Infinite topologically random model

Shreve developed the infinite topologically random model based on the assumption that a natural river network was topologically random in the absence of strong geologic controls (Knighton, 1984; Ni et al., 1998; Qian et al., 1987). The networks with the same number of sources had similar topologic complexity (i.e. the same number of links, junctions and the first order streams), but with distinguishable topologically structures. The number of topologically distinct channel networks (TDCN), for a river network with N_1 sources, was denoted as $N(N_1)$. The number of TDCN is calculated as follows:

$$N(N_1) = \frac{(2N_1 - 2)!}{N_1!(N_1 - 1)!} \quad (4)$$

Thus the numbers of TDCN for sources one to six are 1, 1, 2, 5, 14 and 42 respectively according to the Equation (4), i.e when $N_1=1,2,3,4,5,6$, $N(N_1)=1,1,2,5,14$ respectively accordingly to the formula (4).

A river network with N_1 sources may be composed of $n_1(n_1=N_1), n_2, \dots, n_{G-1}, n_G$ ($n_G=1$) streams whose Horton-Horton-Strahlers order are $1, 2, \dots, G-1, G$ respectively. The number of TDCN for this network is:

$$N(n_1, n_2, \dots, n_{G-1}, 1) = \prod_{g=1}^{G-1} 2^{(n_g - 2n_{g+1})} \binom{n_g - 2}{n_g - 2n_{g+1}} \quad (5)$$

Values of formula (5) are listed in Table 2 for 12 sources network which has 10 kinds of topological structures.

Table 2 TDCN of a network with 12 sources

$(n_1, n_2, \dots, n_{G-1}, 1)$	$N(n_1, n_2, \dots, n_{G-1}, 1)$	Probability $\frac{N(n_1, n_2, \dots, n_{G-1}, 1)}{N(N_1)}$
(12, 6, 3, 1)	2	0.00004
(12, 6, 2, 1)	24	0.00041
(12, 6, 1)	16	0.00027
(12, 5, 2, 1)	1080	0.01837
(12, 5, 1)	1440	0.02450
(12, 4, 2, 1)	3360	0.05716
(12, 4, 1)	13440	0.22862
(12, 3, 1)	26880	0.45725
(12, 3, 1)	11520	0.19596
(12, 2, 1)	1024	0.01742
(12, 1)		
Total	58786	1.00000

It can be seen from the Table 2 that the number of TDCN, with 12 streams of Horton-Strahler order 1, 3 streams of Horton-Strahler order 2 and 1 stream of Horton-Strahler order 3, is the largest one. The most possible natural stream networks should have the largest TDCN number (Krichner et al., 1993; Knighton, 1984; Qian et al., 1987). Thus this type of network is the most probable natural structure in the absence of strong geologic controls.

Shreve indicated that the most probable stream structure resulted from the random model is consistent with that from Horton's law of stream number (Shreve, 1966; Knighton, 1984). Horton's law of stream number is not a law in the strict physical sense but merely a statistical relationship describing the most probable state of network composition. Shreve calculated the network with N_1 (smaller than 100) sources and found that the most probable network had a bifurcation ratio $R_b \approx 4$, which is the modal value of R_b (Krichner, 1993). There was no significant difference between Horton's law of stream number and the random model although the latter one assumed regularity existing in the network structure.

(2) Random link length model

Smart derived a random link length model to investigate length properties of drainage networks (Knighton, 1984; Ni et al., 1998; Qian et al., 1987; Singh, 2000). The average length of the streams with Horton-Strahler order

g was defined as \overline{L}_g and:

$$\overline{L}_g = \overline{l}_i \prod_{a=2}^g (N_{a-1} - 1) / (2N_a - 1), \quad g \geq 2 \quad (6)$$

where \overline{l}_i is the average length of interior links, N_a is the number of the streams with Horton-Strahler order a . This model explains the behavior of ordered stream lengths and reveals that, as with the Horton's law of stream numbers, Horton's law of stream lengths is largely a result of applying the ordering scheme to a topologically random network. Equation (6) was thought to be the substitute of the Horton's law of stream lengths (Knighton, 1984).

The development of the random link length model established theoretical foundation of the geometrical rules (Jin, 1993) which include the stream number, stream length and drainage area laws. The models could be used to quantitatively describe a stream network. All the laws proposed by Horton could be further quantified with probability.

4. Pfafstetter Codification Method

4.1 Pfafstetter Codification Rules

Otto Pfafstetter developed a numbering scheme, called Pfafstetter codification method, for labeling drainage networks in a river basin (Britton, 2002; Luo, 2003; Luo, 2006). Pfafstetter codification method could be used to divide the whole drainage basin into small sub-catchments according to the topology of the drainage network and the size of the surface drainage area. The rules of the Pfafstetter codification are as follows and shown in Fig.3:

- (1) the mainstream is identified with the biggest drainage area;
- (2) find four bigger tributaries with Horton-Strahler order 1 in the mainstream drainage area and code them with 2, 4, 6 and 8 from downstream to upstream respectively;
- (3) the mainstream is divided into five sections by the junctions where the tributary flows into the mainstream, and the five sections are coded with 1,3,5,7 and 9 from downstream to upstream respectively, until now the level 1 codification is finished;
- (4) repeat steps 1), 2), and 3) to a tributary of a stream who has four or more than four small tributaries. The codification of the tributary is called the level 2 codification. The tributary is coded with two digitals, the left one is the code of the level 1 codification and the right one is the code of the level 2 codification;
- (5) the whole coding of the river network will be finished order by order analogically by using the steps above.

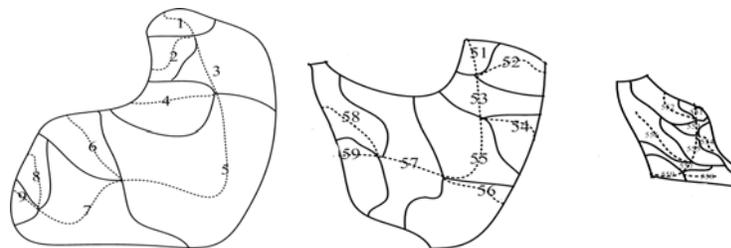


Figure 3 Pfafstetter coding method

The code of the sub-basin corresponds to its stream code. Thus the whole drainage basin can be divided into small sub-basins with coded number. Luo et al. (2006) improved the rules of Pfafstetter so that it could treat streams which had less than four tributaries. However there are no digital 8 in the code of the tributary and no digital 9 in the code of the mainstream when the step 2) and 3) are still used in the improved method.

4.2 Virtues and Shortcomings

The Pfafstetter codification method was widely used in America. North America was divided into 5020 sub-basins with an average surface area of 3640 km² (Britton, 2002). In China, Pfafstetter codification method was used for facilitating to develop the distributed hydrological model by Yang (2004). Luo et al. (2003) developed a program for automatic implementing the Pfafstetter codification method and applied it to the Wei River basin. With the Pfafstetter codification method, the whole Yellow River basin were coded and divided into 8255 sub-basins at

level 6 (Luo et al., 2006).

The codification method allows using decimal digit so that it is consistent with people's habit. Moreover, it is convenient for developing computing program to realize automatic codification process because of the regularity of coding rules. The codification results also have topological information implied in the code.

In real practices, it is sometimes difficult to compare areas among sub-catchments, which will result in difficulties in processes of watershed division. When making the drainage area of the sub-basins equal, there will be large codification level difference in the code of the sub-basins, thus the Pfafstetter codification method do not take account the uniformity of the drainage area. The Pfafstetter codification method can not identify the topological relationship directly if a stream network is with dendritic characteristics. It is easy to cause the lack of the stream code because of the limitation of the digit.

5. A Binary Tree Codification Method

5.1 Binary Tree Codification Rules

A binary tree codification method, which was thought to be more appropriate to the parallel computation for a distributed hydrologic model to simulate rain-runoff processes in a large scale river basin, was proposed by Li et al. (2006) to characterize topological relationships of a stream networks effectively based on the following assumptions:

- (1) there is only one tributary flowing into the mainstream at each node;
- (2) there is only one stream within two adjacent nodes.

Therefore, the whole stream network looks like a binary tree. With the binary tree codification method, each stream is labeled with a binary string as shown in Figure 4. The detailed codification processes are as follows:

- (1) firstly, assigning "0" to the outlet stream of the main stream;
- (2) searching its upstream along the main stream. The first tributary is labeled through adding "1" to the right side of the code for the main stream reach which is right downstream of this tributary as shown in Figure 4; and the right upstream main reach is labeled through adding "0" into the code of the downstream reach;
- (3) repeating step (2) till all streams are coded.

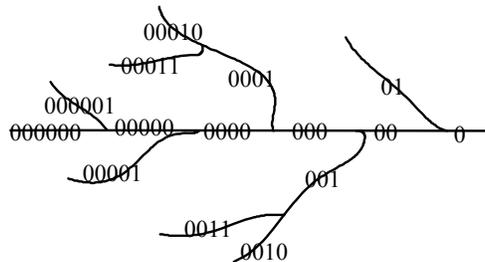


Figure 4 Binary string codification of a drainage network

After the whole stream is coded, each node is represented in the form of (L_n, V_n) as shown in Figure 5, where L_n is the length of the binary string and V_n is the decimal value of the binary string. L_n can reflect the distance from the stream to the outlet. The larger the L_n is, the longer the distance from the stream to the outlet is. V_n can reflect the distance from the stream to the mainstream from the view of topology. In the corresponding binary tree, L_n represents the layer of node in the binary tree; and V_n represents the code of the node in the binary tree.

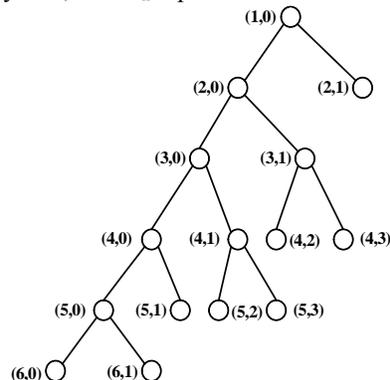


Figure 5 The binary tree representation of a drainage network

There are remarkable characteristics that the numerical component of the mainstream is 0 and the farther the tributary from the mainstream is, the bigger the numerical component of the tributary is. The numerical component increases in the terms of 2^x . However when a tributary develops to a certain degree, the numerical component will be too big to denote and store, even the component will overflow. Thus the bigger size tributary is needed to be separated and coded itself by the same way when the component overflows. The mainstream of the separated tributary is also assigned with the numerical component 0. If the tributary of the separated tributary overflows again, the smaller separation will be done repeatedly. The above steps are done circularly till all the fingertip tributaries are coded.

5.2 Virtues and Shortcomings

The binary tree codification method can be used to help index any section in a stream network directly through searching it from the fingertip tributaries to the outlet. Furthermore, the topological relationship of a stream network is represented by two parameters (L_n and V_n), which can be directly used to facilitate spatially parallel simulation of hydrologic processes.

Since the binary tree codification method directly reflects the binary tree character of stream networks, the relations among streams coincides with evolvement features of streams in a river basin. By using the binary tree codes, that are L_n and V_n , a drainage network stored in a database table can be efficiently delineated in binary tree data structure in memory. And the binary tree data structure facilitates a parallel computation algorithm, which disassembles the drainage network into branches in controllable size and sequence from upstream to downstream. The unification of the runoff and sediment process and routing computation is supported by the binary tree codification method.

The binary tree method is useful for the topological location and calculation; however the bigger tributary inevitably needs to be separated from the mainstream for the reason of code overflow. There must be something extra to be done to denote the inflow junction of the separated tributaries. It brings a little more complexity to do conflux calculation.

6. Other Codifications

Other codification methods are proposed for stream networks with different geography characteristics. The German method of catchment coding (Britton, 2002) has been established by Landerarbeitsgemeinschaft Wasser(LAWA). The coding system is hierarchical and purely numeric. The system is very similar to the Pfafstetter system, except that the sub-basin is coded from the headwaters to the river mouth, which is in the opposite direction with the Pfafstetter system. Thus, like the Pfafstetter system, it has a limited number of basin code, and it is not able to provide a code of equal format in all levels and areas.

The Norwegian Register of Catchment Areas (REGINE) coding system (Britton, 2002) is developed and applied to delineate the whole river systems in Norway. The water system area on the highest hierarchy level are subdivided into two parts, a main catchment with one single outlet, and a remaining area with diffuse drainage systems or minor watercourses along a lake or the coastline. Furthermore, the main catchment and the remaining area are labeled with Z and O respectively. The main catchment areas are then further divided into 24 sub-areas at most, labeled with the letters A to Z excluding O and Z. The area with diffuse drainage systems is subdivided into 9 sub-areas through using the numbers 1 to 9. Thus the code is economical in that it uses few digits. However, and coastal areas in the system may not be clearly indicated. The different steps of subdivision, one with the addition of a new digit and one with letters, make the system somewhat unclear and difficult to handle.

The Finish coding system (Britton, 2002) is a structured hierarchical approach and successfully applied to many real practices. However, the codification method using 18 characters to code stream networks results in significant complexes. It is difficult to code a new tributary which joins into the stream networks since the coded network has to be re-coded again.

ERICA coding method (Britton, 2002) is an extension of the Pfafstetter method. It could deal with 99 sub-catchments and 49 tributary catchments at most, thus making it easy to maintain a relationship between catchment size and levels of coding. Sea and marine borders are handled. The ERICA method using two digits to code 49 tributaries not only increases the length of the code but also decreases the discernment capacity.

7. Conclusion

In this paper several most popularly used methods on classification and codification of stream networks has been reviewed. The rules and principles of these methods have been briefly introduced. Advantages and disadvantages of all methods have been summarized. The results can supply decision support for classification and codification methods selection for water resources management and distributed hydrologic modeling. The review will supply useful references to researchers who are engaging in hydrologic modeling.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 50221903) and the Key Technologies Research and Development Program of China (Grant No. 2005BA901A11).

References

- Britton P. (2002). Review of Existing River Coding Systems for River Basin Management and Reporting (WFD GIS Working Group: European Coding Systems Task Group), October 2002, http://193.178.1.168/River_Coding_Review.htm
- David G. (1996). Tarboton. Fractal river networks, Horton's laws and Tokunaga cyclicity. *Journal of Hydrology*, 187, 105-117
- Jia, Y. W., Wang, H., et al. (2005). *The theory and practice of the distributed hydrologic models*. Beijing: China Water Resources and Hydropower Press
- Jin, C. X. (1993). Empirical relations and theoretical explanation to the landforms of drainage areas. *Journal of Shanxi Institute of Mechanical Engineering*, 9, 45—51
- Kirchner J. W. (1993). Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks. *Geology*, 21, 591—594
- Knighton D. (1984). *Fluvial Forms and Processes*. Britain: Edward Arnold
- Li, T. J., Liu, J. H., Wang, G. Q. (2006). Drainage network codification method for digital watershed model[J], *Advances in Water Science*, 17, 658—664
- Lu, Z. C. (1987). *Watershed geomorphology system*. Dalian: Dalian Science Press
- Luo, X. Y., Jia, Y. W., Wang, J. H., et al. (2003). System for topologically coding river basin and its application, *Advances in Water Science*, 14, 89—93
- Luo, X. Y., Jia, Y. W., Wang, J. H., et al. (2006). Method for delineation and codification of a large basin based on DEM and surveyed river network, *Advances in Water Science*, 17, 259—264
- Ni, J. R., Ma, A. N. (1998). *River dynamical geomorphology*. Beijing: Peking University Press
- Qian, N., Zhang, R., Zhou, Z. D. (1987). *Fluvial processes*. Beijing: Science Press
- Shreve, R. L. (1966). Statistical law of stream numbers, *Journal of Geology*, 74, 17-37
- Singh V. P. (2000). *Drainage basin simulation*. Zhengzhou: Yellow River Conservancy Press
- Wang, G. Q., Liu, J. H. (2005). Beijing: Science Press
- Xiong, L. H., Guo, S. L., et al. (2004). *The distributed hydrologic models*. Beijing: China Water Resources and Hydropower Press
- Yang, D. W., Li, C., Ni, G. H., et al. (2004). Application of a distributed hydrological model to the Yellow River basin. *Acta Geographica Sinica*, 59, 143—154

Li Zhang is presently a Ph. D. candidate, studying on the river network analysis.