River Classification and River Network Structuration in River Auto-selection[1]

Lili Jiang[2], Fang Zhou, An Zhang, Qingwen Qi[3]

Institute of Geographic Sciences and Natural Resource Research, CAS, Beijing, China
State Key Laboratory of Resources and Environmental Information System, Beijing, China

Abstract. River auto-selection is an essential part of the automatic generalization of thematic maps. When accomplishing the selection, what is critical is to reasonably evaluate each river's importance in the global structure. A grading system is needed here to quantify the importance. In this system, the importance of river differs among different levels. Though several grading systems (Horton, Strahler, etc.) are already there, each with its strengths, still they cannot meet the requirements of river auto-selection. Meanwhile, as for river auto-selection, river classification is extremely crucial and indispensable. And, when taking river classification into consideration, it is mainly based on the following two points: (1) high-level river should be chosen to ensure the connectivity of its network, with the assistant of the grading system; (2) rivers in a same level can be selected by length, density and other indicators. Consequently, a basin-based grading system of river classification is proposed in this study. Consider from each river's importance in partial river network and selecting accordingly, this system can make the density of the selected rivers as consistent as the network before.

---

[1] Foundation item: The National Natural Science Fund Youth Project, project approval number: D41101447, " The drainage Auto Select Method Based on Basin Characteristics and Drainage network structure "

[2] Author introduction: LiLi Jiang (1972), female, doctor, research assistant, research direction: cartography and geographic information.

[3] Corresponding author: QingWen Qi (1963), male, Ph.D., professor, research direction: cartography, remote sensing, geographic information, etc.

Keywords: Cartographic generalization, River selection, Basin, River classification, River network structuration

1. Introduction

Normally, the cartographic generalization begins with the generalization of rivers. Since river selection being an important step and one of the key contents of rivers' generalization, the quality of its results will directly affect the quality of the whole map.

River selection refers to selecting the rivers which is relatively long, important and accordant to the theme of the map, or the rivers that can reflect the regional characteristics of map, and giving up one or a few river objects that is not so related with the contents of the map. There exits three ways available now for river auto-selection: (1) selecting by simple quantitative index; (2) selecting by network analysis and hierarchical structure model; (3) selecting by knowledge and intelligence. Studies and experiments using these three methods have noticed the role rivers' distribution characteristics play in the auto-selection, to various degrees. While lacking of consideration on rivers' distribution characteristics in different basins, these attempts destroyed the original relative differences of river distribution. Therefore they have failed to fundamentally solve the key problem of river auto-selection, and not any substantial progress has yet been made in it.

River is some sort of complex spatial object which is highly structured. Complex spatial relationship among rivers and river systems which developed in different hydrological conditions and terrain environment, even their distribution characteristics, cannot be easily acknowledged by some kind of simple analysis with GIS, as well as the uncertainty of river objects that constitute a river network. A river network is made up of different river systems. Each river system has its own source and scope of influence (basin). Rivers originated from one same source, may develop into a variety patterns of small networks, due to different geological and geomorphological conditions. These small river networks also have their own basin. For a long time till now, the quantitation of river networks has been delayed. A major reason to this is nothing else but the sheer complexity of these networks' spatial distribution.

Rivers commonly have a dendritic, feathery, or checkerboard-like shape in their spatial distribution structure. Since being different from other linear features in the characteristics of planar structure, river represents its specialty accordingly in the selection. On account of the order and connectivity character of rivers' topological relations, the erasable ones are confined to the external and some higher grade rivers.

The shape of the basin can reflect the morphological structure of the river network. The basin is also the comprehensive reflection of river levels, length, distribution density and other geometrical characteristics, making itself the key geographic characteristic factor to judge the importance of each branch. Thus, a prime task of river auto-selection is to acquire the basins of appropriate scale as the selection unit for the rivers on the relevant scale.

In order to study the selection in different types of river networks, this paper will divide a large-scale basin into small ones. Accordingly, the river network is cut into a few correlative individuals (a small set of rivers). Inside each small basin, the spatial distribution structure of the corresponding individual has a simple single shape. Each small basin can be regarded as the space distribution range of the individual. And river selection is undertaken inside each small basin, according to a grading system.

2. Method

In the process, as river being the direct object of the auto-selection, a thorough understanding of its features and its overall structural characteristics is required for the design of a targeted method. The primary mission of river auto-selection is to pick out the rivers that can reflect the geographical characteristics of the mapping area and reject the ones in a critical level but can't. The shape of a basin mirrors the morphological structure of its river network, and it is also the comprehensive reflection of river levels, length, distribution density and other geometrical characteristics. For these reasons, this research proposes a hierarchical structure model for river classification based on small basins.

2.1. Existing River Classification Model

There are two existing river classification model now—one is based on the node–reach of the river, the other is based on the mainstream-tributary of the river.

- River Classification Based on Node-Reach

Considered from the meaning of a geographic entity, river network is a tree-structured one. And its data model can be defined as a four-level structure—river network, river, river reach and river node. River entity refers to a complete channel between the source and the estuary, or between a branch source and its estuary. As being an entity with complete geographic significance, river is the basic unit of a river network, and it also corresponds to one encoded entity in Horton coding system. Meanwhile, river reach is the

segmental arc between all kinds of nodes—river source, bifurcation, etc. Therefore, river reach is the basic unit of a river, and it also corresponds to one encoded entity in Strahler coding system. Additionally, river reach can be presented by initial node, terminal node and its flow direction.

Figure 1 visualizes the process of river classification based on Strahler coding system:

Figure 1. Strahler hierarchical coding.

Figure 1-1 represents a basic river network (encoded by Strahler grading system). Figure 1-2 shows the river network in which all (or external) source reaches in the first figure are deleted. Those deleted reaches are defined as Level 1. The new source reaches appeared in the second figure are defined as Level 2, deleting which we could get Figure 1-3, and in it there leaves Level 3 reaches of this river network.

Reach classification is mainly based on the flow direction and the topology network of a river. After the establishment of topological relationships between reaches, the data is no longer unrelated and disorganized, and the links between the reaches are established. The topology network helps to make further analysis of the river structure. It also builds the foundation to judge the mainstream and the tributaries.

- River Classification Based on Mainstream-Tributary

In the data organized with the node-reach pattern, a reach is not a complete entity of a river. The mainstream of a river cannot be completely specified.

Some other methods is needed to judge the relationship between the mainstream and the tributaries. River classification based on mainstream-tributary can fit better with our understanding of river's tree structure, with a better performance of each element's importance in a river network and the confluent relations among branches.

It is quite a complex issue to identify the mainstream and the tributaries automatically. The identification generally begins with the determination of the mainstream. But mainstream determination cannot rely on a single restriction, or on one local reach. The entire river network should be taken into consideration. In general, considering the morphology of the river, the mainstream is determined according to the following points:

1. Length Priority Principle to Identify the Mainstream

Generally speaking, the mainstream has the characteristic of being the longest reach in a river network (Richardson, 1993). Therefore, the length priority principle is helpful to find out the mainstream. According to river's tree structure, there's only one access between any two reaches. For this reason, the mainstream of a river is composed by longest reaches from the source reach to the terminal reach. And the longest access from all the source reaches to their related terminal reaches forms the mainstream.

2. 180°Approximation Principle to Identify the Mainstream

With the awareness of the flow direction of the entire river, it is easy to discover that at the join node, the mainstream of the river is likely to maintain its original flow trend. That is, there's a trend to approximate to 180° between any two reaches of the mainstream. Therefore, at the join nodes of two river reaches, the angle between the lower reach and the upper reach should be calculated in a clockwise (or counter-clockwise) —the reach with an angle more close to 180°proved itself to be the mainstream, and the others are the tributaries.

3. Using River Level to Identify the Mainstream (Horton)

In Horton grading system, each of the initial ditches flow (small tributaries), is designed as the Level 1 stream. Two Level 1 streams are intersected to produce the Level 2 stream. Afterwards, every two streams of a same level confluent and generate a higher level stream, and so forth. In this system, the highest level stream in a basin flows from the estuary back to the source. It traces upstream from the estuary, along the main stem and acquiring the minimum differentiation with its direction, concluding streams from the highest level until the lowest level.

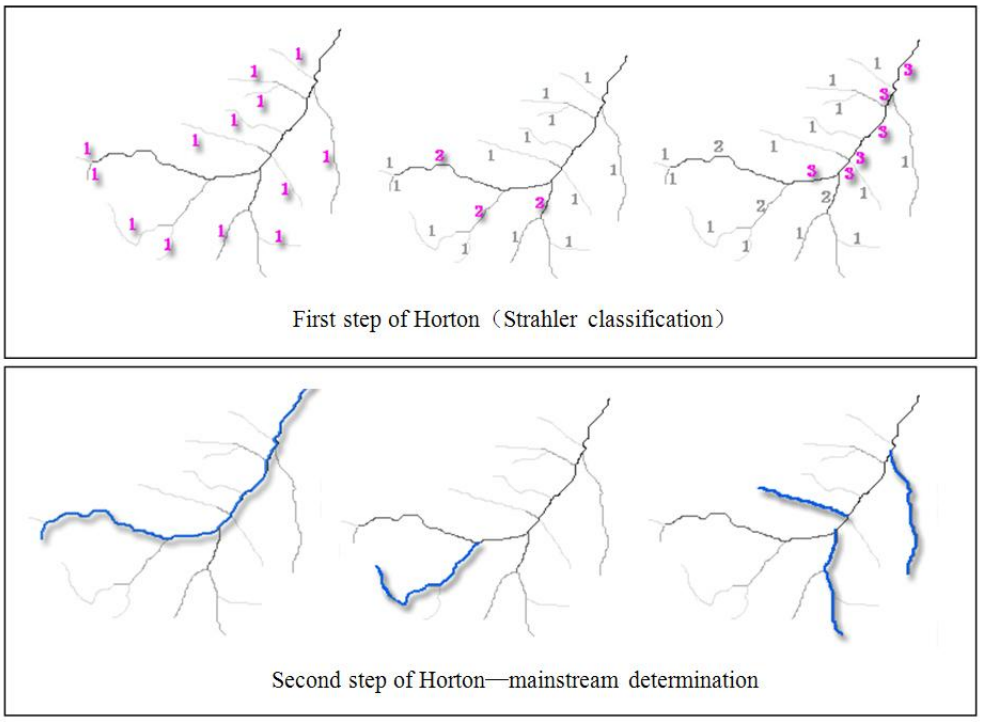

Figure 2. Horton hierarchical coding

2.2. Basin Division and River Classification

- Basin Division

In the integration of river network, basin is the key geographical feature factor to judge the importance of each river branch. It is the comprehensive reflection of the geometric characteristics—river level, length, distribution density, etc. As the unit of integrated selection, basin refers to the catchment area of a certain level. Therefore, to reasonably extract the catchment area, to build the hierarchical relationship between each other, and to establish a matchup between the hierarchical model of the catchment area and the river network auto-selecting unit, is fundamental for river auto-selection.

There are many mature approaches to extract the basin. The most effective one is the accumulated confluence threshold method. The threshold method is a simulation based on overland runoffs (have a great influence on the formation and development of valley). With different threshold value of accumulated confluence, we can get basins of different scales.

As well as the information about basins, information about elevation, flow direction, slope, aspect, etc. is also needed for river auto-selection. Pfafstetter coding method can be used here to establish the hierarchical structure of basins. Pfafstetter coding is built on top of the river network topology. To use Pfafstetter coding, the mainstream (the cumulated flow of which is required to be higher than other tributaries) should be identifyied firstly. Then identify its basins as 1, 3, 5, 7, and 9. Next, in accordance with the amount of the cumulated flow, select four branches along the mainstream and identify their basins as 2, 4, 6, 8. Next, encoding these 4 branches in the same way. After these steps, the hierarchy basin structure is established.

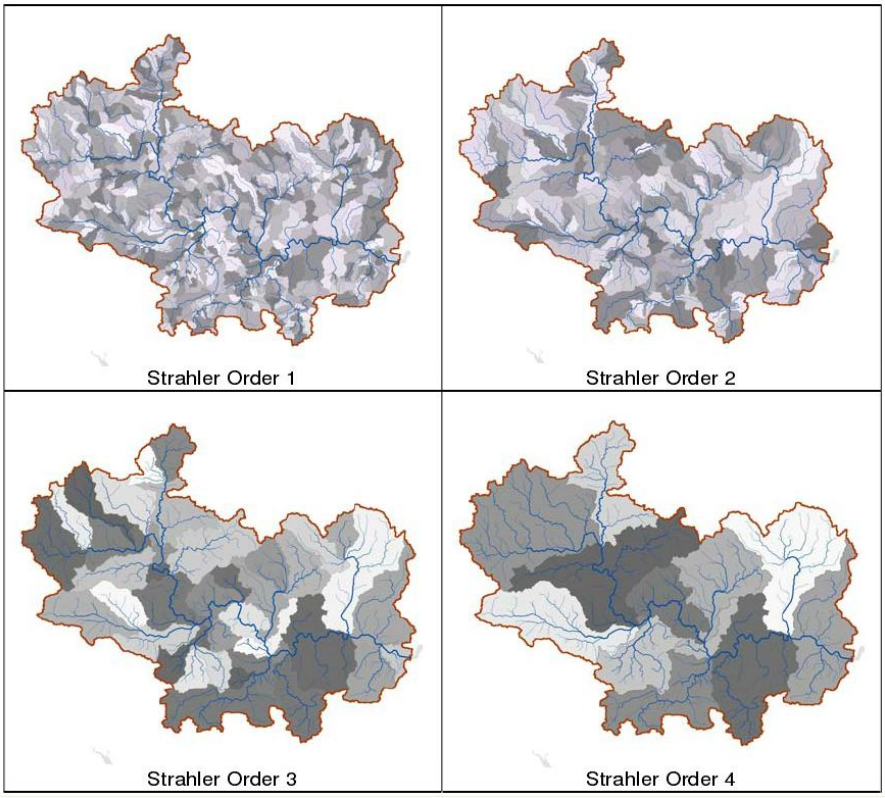

Figure 3. Basin hierarchical chart

Figure 3 depicts the hierarchical basin structure after Pfafstetter coding. From this picture we can clearly identify the hierarchical relationship

between basins of different levels and the relationship between the basins and the rivers inside them.

As for river selection at different measuring scales, required basin units are at different levels. This paper focuses on the river auto-selection of a scale from 1:250,000 to 1:1,000,000, hence a suitable threshold is needed here to extract basins, which will be the units of river auto-selection under this scale.

- Establishment of a Hierarchical River Structure Model

The distribution characteristics of rivers mainly lay on the distribution relationship between the mainstream and the tributaries. Being the distribution axis of the basin, the mainstream is the core to control the whole basin. With this axis as the center (not necessarily to the axis of symmetry), other rivers dispersed to the two sides of the mainstream and form a network. Therefore, the mainstream of a basin is the mandatory entity to be selected. And when selecting a tributary, we should consider if it is "the one" that reflects the distribution characteristics of the river, if so, select, otherwise delete. Intersection angle between mainstream and tributaries is also a factor to reflect the distribution characteristics of river. And only after all the rivers are graded, can we calculate the angle between the mainstream and the tributaries. Therefore, river classification is very important for the study of the distribution of river network. Other relevant distribution factors like river shape, river length and its changes of rivers at the same level are also needed for river selection.

River classification is the basis of hierarchical structure of river network inside a basin. Through the analysis of two common river grading systems mentioned above, it appears that, for river selection, classification based on mainstream-tributaries and based on river reaches could be combined together to automatically build the hierarchical structure of river network. The river reaches are classified using the Strahler grading system, and the mainstream is recognized relying on the hierarchical distribution relationship between mainstream and tributaries (Honton principle for mainstream identifying), together with the length priority principle. The specific process is shown in Figure 4:

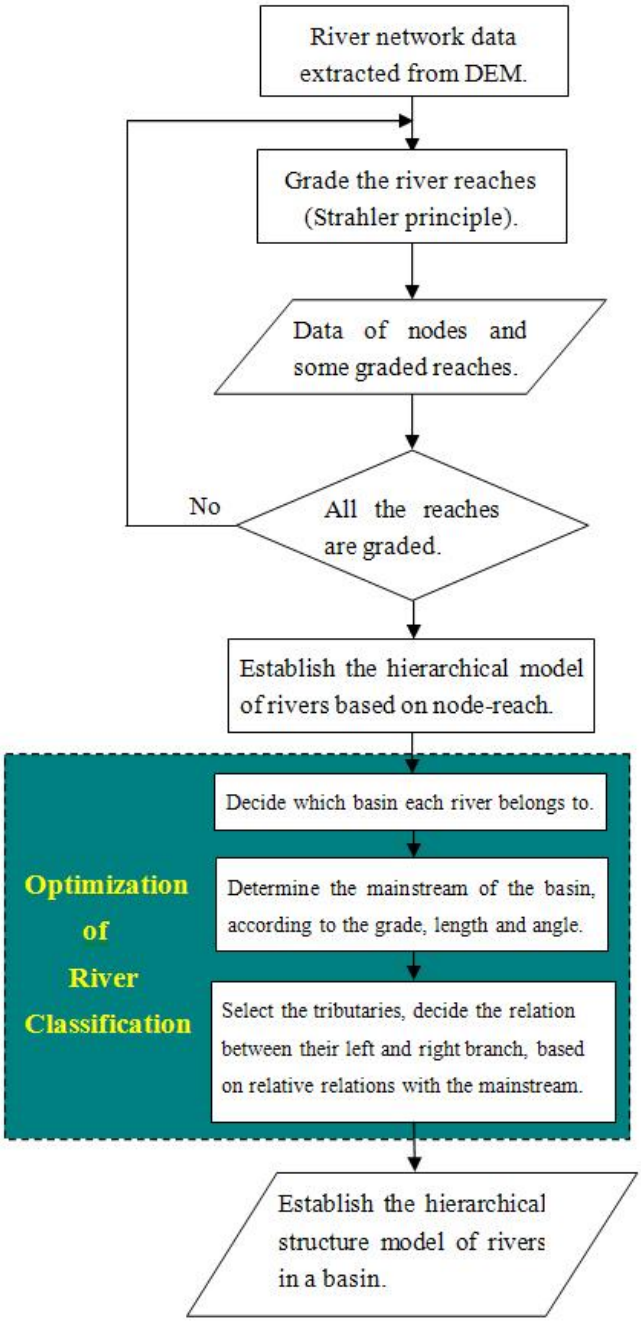

Figure 4. Optimized process of river classification based on small basins

1. Establish the information database of the nodes and reaches in the river network, and obtain the in and out degree and its flow direction
2. Obtain the source node of all the rivers, record the direction of river reaches, and set the grade of the lower reach as 1.
3. According to the principle of Strahler grading, give the correspondent grades to all of the river reaches and calculate the length of them.
4. Divide the basin and determine which basin each river belongs to.
5. Determine the mainstream of the basin: start from the river reaches of the highest grade, find out the connected reaches in turn; on the basis of the mainstream characteristics (highest grade, longest and most straight), reversely search out the mainstream of the river network, encode the mainstream and count the amount of the rivers (need to add two fields in the river network database: CLASS and RID, and set 1 to the values of mainstream's CLASS and RID), to identify the river grade and the river itself (not the reaches). That is to say, the river reaches with the same field value form a complete river branch.
6. Take the river reaches joint to the mainstream as the origin of tributaries; reversely search out its tributaries of the first level (set 2 to the value of its CLASS). Till all of the reaches are appointed to the river tributaries, the rivers have finally been connected and the hierarchical structure of river network been established. Meanwhile the CLASS code of the first level tributaries is valued (CLASS=1), and continue to set value for RID. Then search for the tributaries of the second level (the nodes connected with the first level tributaries are the origins, set 2 to the value of CLASS, and so on, until all the river reaches are checked, the coding of river grade is completed.)
7. When traversing the river branches, decide whether the left or right branch the tributary belongs to, based on the right-hand rule. Set up the ORIENT field, and set -1 when it's a left branch, set 1 when it's a right branch. Then, the hierarchical structure of the spatial river network data organized by the pattern of node-reach would have been built up through the method above. Relying on the hierarchical structure, the classification of river network based on mainstream—tributaries can be completed. Save the information of nodes and reaches corresponding to mainstream and tributaries of different levels, to accomplish the reestablishment of the hierarchical river structure.

In the grading system raised in this paper, river reach is the actual storage unit of a river in the database. But the database also has a field（RID）to note the river entity (RID values of the same river entity are the same). When it comes to the some processing towards river entity, a set of them can be produced at any time, according to the field of RID. The advantage is obvious—include both the advantage of data organization by reach-node

and by mainstream-tributaries, reduce the storage and fulfill the demand of river selection.

3. Data Processing

The experimental area for this study is the Yangtze River Basin and the Yellow River Basin. 30 meters resolution DEM data, 1:250,000 and 1:1,000,000 standard river system data are used in this study. The research topic of this paper is river selection, with a concern of how to make the morphological structure of the river network as consistent as that before generalized. This paper focuses on the single line river. Because in the classification of the patterns of river network, this simplex linear river data is much more effective in the calculation and statistical analysis of the characterization factors and selection indicators, and it can avoid the complexity of characteristics calculation brought by multiple types of features, to improve the efficiency and accuracy of the computation of factors. Therefore, experimental data should be preprocessed to guarantee its organization in a standard way, and to meet the demand of the analysis on classifying characterization factors.

- Result of Basin Distribution

First, 1:250,000 standard river system data should be processed as follows: extracting single line rivers from double line rivers, as well as the rivers which connect the river network to the nearby lakes. The processed rivers must keep the connectivity with the network. Then, the basins come into being according to the flow amount extracted from DEM. And finally, river classification and river selection can be completed inside these basins.

The experiment shows that the accumulated flow threshold of 1:250,000 river data is about 10% of the total flow amount. Therefore, in the auto-selection from 1:250,000 river data to 1:1,000,000 river data, as an example of the experiment, 10% is determined to be used as the threshold to extract the basin units. Figure 5 below displays the result of the extracted units.

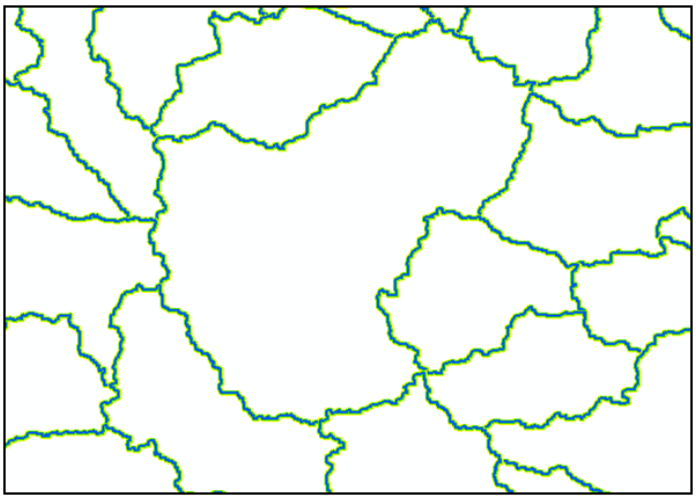

Figure 5. Result of basin distribution

- Result of River Classification

Result of river classification based on Strahler grading system:

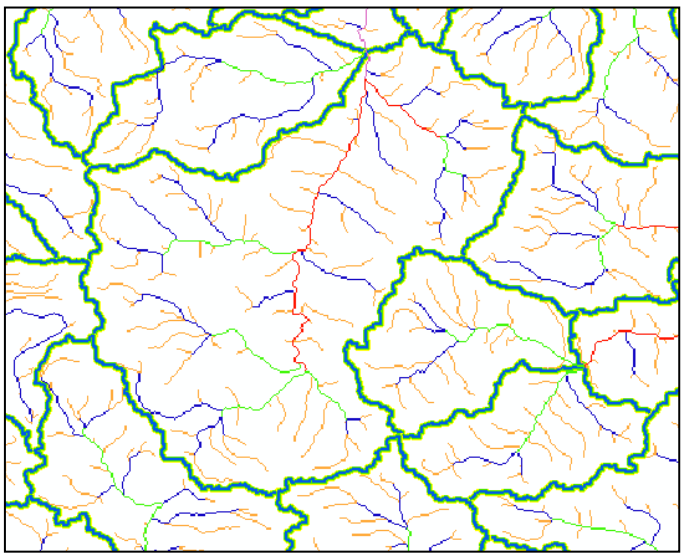

Figure 6. Result of Strahler

- Result of River Classification Based on Small Basins

Rivers belongs to different basins. Inside each basin, the rivers are classified by the optimization method above-proposed in this paper (see Figure 7).

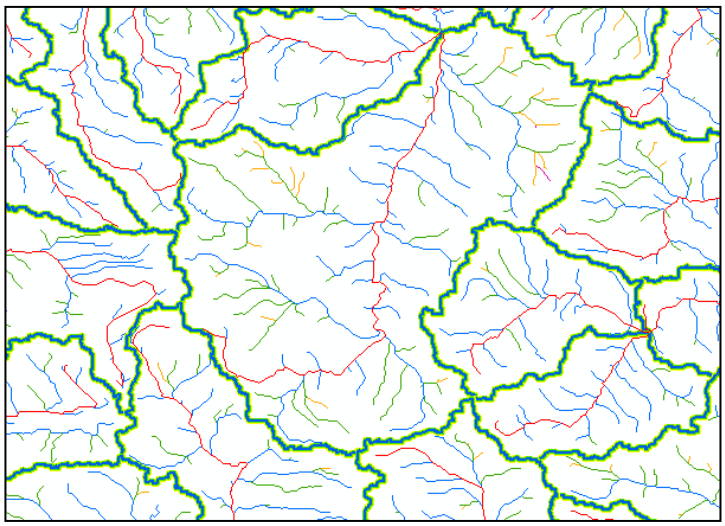

Figure 7. Result of river classification based on small basins

4. Findings

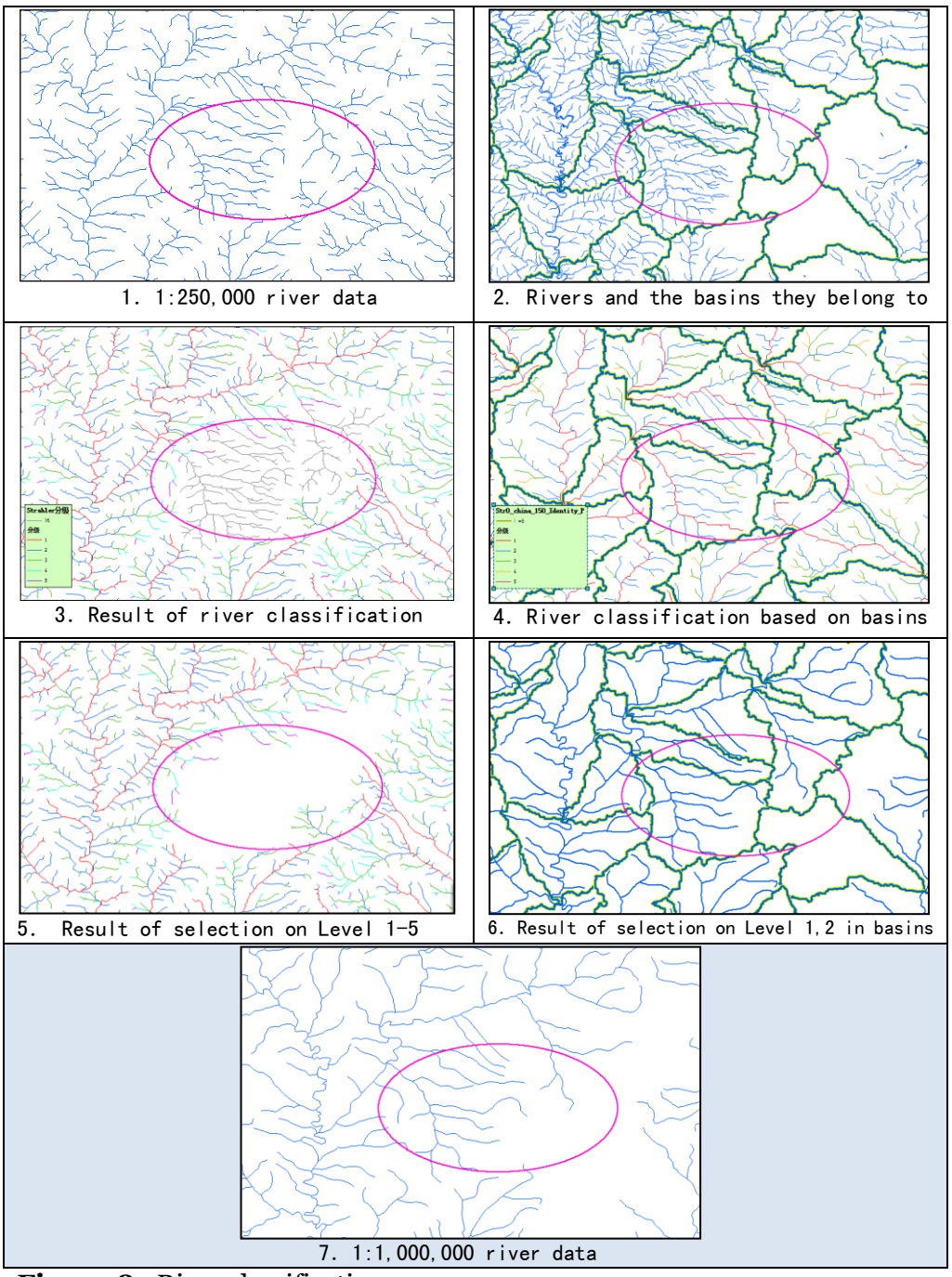

Figure 8. River classification

Figure 8-1 presents 1:250,000 standard river system data in the experimental area. Figure 8-2 presents the same river data and the basin each river belongs to. Figure 8-3 displays the river classification result based on the whole Yangtze/Yellow River Basin. Figure 8-4 displays the river classification result based on each small basin. Figure 8-5 shows the river selection result on the grade of one to five, based on the whole Yangtze/Yellow River Basin. Figure 8-6 shows the river selection result of the first and second grade, based on each small basin. Figure 8-7 presents 1:1,000,000 standard river system data in the experimental area.

In accordance with the traditional method (based on the a few large-scale basins), the rivers inside the red circle (figures above) are all in a relatively low grade and should be completely abandoned as long as the selection grade is greater than or equal to five. While compared with the 1:1,000,000 standard river system data, these rivers should be kept to achieve the consistency. It follows that river selection based on the whole experimental region does not apply to the river auto-selection.

Obviously, in the situation of considering grade only, river classification based on each small basin does much better at river auto-selection, and is more reflective of the regional distribution. In comparison, river classification based on the whole experimental region may lead to some instances that the rivers in a basin are entirely eliminated.

The reason for the instance above is that, on the basis of the existing methods for river classification, as for some complex river network, especially some with more branches, the selection inevitably results in some unpleasant conditions that rivers in some partial regions are completely deleted. Therefore, it is a strong proof that the method existed in river classification cannot meet the requirements of river auto-selection and need to be optimized. While comparatively speaking, in this study, the river classification experiment based on each small basin, get a remarkable result in solving the problem.

5. Discussion and conclusion

Some conclusions can be obtained as follows:

- The shape of the basin can reflect the morphological structure of the river network. The basin is also the comprehensive reflection of river levels, length, distribution density and other geometrical characteristics, making itself the key geographic characteristic factor to judge the importance of each branch. Thus, basin is significant to river auto-selection.

- The major concern here in river selection is the importance of river. So, river classification based on each small basin is, in effect, to order the regional importance of rivers. Since the higher the grade, the higher the importance, the river is more likely to be selected.
- River selection based on classification in basins is remarkable in keeping the connectivity of river network and maintaining the consistency of its density.

Acknowledgements

Prof. Qi Qingwen reviewed this paper and raised valuable comments which greatly improved the quality of this paper. Many engineers and students took part in this field survey, providing generous expertise help. They are: Mr. Sun Yafu, from Nanjing Institute of Geography and Limnology under CAS also joined this field survey.

References

Guillaume, Touya A Road Network Selection Process Based on Data Enrichment and Structure Detection. 10th ICA Workshop on Generalisation and Multiple Representation. Moscow

Horton RE (1945) Erosion Development of Streams and Drainage Basins: Hydrophysical Approach to Quantitative Morphology. Bulletin of the Geological Society of America 56(3): 275-370

Lili Jiang, Qingwen Qi, Zhong Zhang (2009) The abstraction method research of river network based on catchments' characters deriving digital elevation data. spie

Lili Jiang, Qingwen Qi, Zhong Zhang (2008) The generalization method research of river network based on morph structure and catchments' character knowledge. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXVII. Part B4. Beijing 2008.597-602

LiuYaoLin, Martin.Molenaar, Menno-Jan.KRAAK (2002) Automated Organization of Hierarchical Catchments in River Network Based Constrained Delaunay Triangulation. FIG XXII International Congress

Qingnian Zhang (2004) Road Network Generalization Based on Connection Analysis. The 11th International Symposium on Spatial Data Handling, 343-353.

Tinghua Ai (2007) The drainage network extraction from contour lines for contour line generalization. ISPRS Journal of Photogrammetry and Remote Sensing 62: 93-103.

TOPFER.F The Principles of Selection: A Means of Cartographic Generalization. The Cartographic Journal 3(1):10-16.

William Machaness, Geoffrery Edwards (2002) The importance of modeling pattern and structure in automated map generalization Multi-scale representations of spatial data.

http://www.esri.com