AUTOMATIC RECOGNITION OF DRAINAGE PATTERNS IN RIVER NETWORKS AND ITS APPLICATION TO GENERALIZATION

ZHANG Ling

Ph.D

The Hong Kong Polytechnic University

2014

The Hong Kong Polytechnic University

Department of Land Surveying and Geo-Informatics

Automatic Recognition of Drainage Patterns in River Networks and Its Application to Generalization

ZHANG Ling

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

August 2013

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________(Signed)

ZHANG Ling (Name of student)

***** Dedicated to my grandfather *****

ABSTRACT

In both GIS and terrain analysis, drainage systems are important components. Owing to local topography and subsurface geology, a drainage system achieves a particular drainage pattern based on the form and texture of its network of stream channels and tributaries. The drainage pattern can reflect the geographical characteristics of a river network to a certain extent, because it depends on the topography and geology of the land. Although research has been done on the description of drainage patterns in geography and hydrology, automatic drainage pattern recognition in river networks is not well developed. In addition, whether in cartography or GIS, hydrography is one of the most important feature classes to generalize to produce representations at various levels of detail. There are many methods for river network generalization, but few of them consider the drainage pattern in the first place, and the generalized results are always inspected by expert cartographers visually. Therefore, this research focuses on the drainage pattern and its application to map generalization.

First of all, this thesis introduces a new method for automatic classification of drainage systems in different patterns. The method applies to river networks and the terrain model is not required in the process. A set of geometric indicators describing each pattern are presented and the membership of a network is defined based on fuzzy logic. For each pattern, the fuzzy set membership is given by a defined IF-THEN rule composed of several indicators and logical operators. The method was implemented and experimental results are presented and discussed.

Second, this thesis proposes a method that evaluates the quality of a river network generalization by assessing if drainage patterns are preserved. This method provides a quantitative value that estimates the membership of a river network in different drainage patterns. Assessing the quality of a generalization is done by comparing and analyzing the value before and after the network generalization. This assessment method is tested with several river network generalization methods on different sets of networks.

Finally, this thesis proposes a solution to deal with multiple factors at same time during the river network generalization. The multi-objective optimization problem is settled by the genetic algorithm with consideration of the drainage pattern. According the characteristic of each drainage pattern, the factors, such as drainage pattern membership, stream order and tributary balance, are considered and built into objective functions. In the multi-objective model, different weights are used to aggregate all objective functions into a fitness function. Then, the generalization is implemented by a designed genetic algorithm.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor Dr. Eric Guilbert. It is such an honour to be his first Ph.D. student. I appreciate all his contributions of time, ideas, and funding to support my Ph.D. study. The insights and enthusiasm he has for his research influenced and encouraged me a lot, especially during tough times in the Ph.D. pursuit. Without his support and help, I would not be possible to finish this thesis.

Then, I want to express my thanks to my co-supervisor, Prof. Zhilin Li, who gave me the opportunity to carry out the research on map generalization and helped me to establish the direction taken by my research. I am also thankful for the excellent example he has provided as a successful researcher and professor on geo-informatics.

Many thanks should be given to all academic and general staffs in LSGI, especially, Prof. Xiaoli Ding, Prof. Wu Chen, Dr. Lilian Pu, Prof. Wenzhong Shi, Prof. Bruce King, Prof. Baki IZ and Dr. Bo Wu, for your suggestions and comments to my research at any of seminars or discussions; Ms. Vaness Yu, Ms. Ziki Cheung, Ms. Violet Tsang, Ms. Fiona Fong, Ms. Justin Yuen and Mr. Kenneth Yau, for your help in daily life and work. In addition, special thanks to Dr. Paul Fox for your work of the proofreading to the manuscript of my thesis.

I would also like to take this opportunity to thank my colleagues and friends for their help in both my study and life. I cannot list all your names, but several close friends should be mentioned. To Mi Jiang, Qianxiang Xu, Rui Xu, Jian Guo, Hui Sun, Peng Ti, Dr. Qi Zhou, Jingya Yan, Yuming Hong, Dr. Yaohua Hu, Gao Li, Shaojie Zhang, Dr. Ziwei Xu, Yazhong Xu, Jue Wang, Dr. Hao Wang, Kun Cheng, Ran Zhang, Dr. Yang Liu, Dr. Xiaofan Liu, and Xiaotian Sun etc., it is so nice and lucky to meet you in Hong Kong, thank you.

Last, but not least, I wish to express my deepest gratitude to my parents, Qiping Zhang and Huizhen Sui, and my girlfriend, Teng Fei, for their love, support and encouragement as always.

Ling Zhang
PolyU, Hong Kong
Aug, 2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AABB	Axis-Aligned Bounding Box
BGL	Boost Graph Library
CDTIN	Constrained DTIN
CLC	Coefficient of Line Correspondence
DEM	Digital Elevation Model
DTIN	Delaunay TIN
EA	Evolutionary Algorithm
GA	Genetic Algorithm
GIS	Geographic Information System
MBR	Minimum Bounding Rectangle
MF	Membership Function
NHD	National Hydrography Dataset
RRIIS	Russian River Interactive Information System
TIN	Triangulated Irregular Network
TOPAZ	Topographic Parameterization
WMS	Watershed Modeling System

Chapter 1 Introduction

1.1 Research background

A natural drainage system is the pattern formed by streams, rivers and lakes in a drainage basin. The drainage system is an important component in Geographic Information System (GIS) and in terrain analysis as it provides a morphological partition of the terrain. In a drainage system, a stream or a river is a natural watercourse, usually freshwater, flowing towards an ocean, a lake, or another river. Apart from a few cases where a river simply flows into the ground or dries up completely before reaching another body of water, rivers always connect together to form networks, achieving a particular drainage pattern. The drainage pattern is "*the arrangement in which a stream erodes the channels of its network of tributaries*" (Chernicoff & Whitney, 2006). It is different from the channel pattern which is used to "*describe the plan view of a reach of river as seen from an airplane*" (Leopold & Wolman, 1957). The river pattern describes the morphological structure of a river network at the river basin scale and is different from the channel pattern which describes the river morphology at the river channel scale.

There are several types of drainage patterns. They are commonly classified as dendritic, parallel, trellis, rectangular, radial, centripetal and reticulate patterns (Ritter, 2006). Dendritic patterns, also named tree-like patterns, can usually be found where there is no strong geological control (Charlton, 2008). Parallel, trellis and rectangular drainage patterns develop in areas with strong regional slopes but have their own specific characteristics. Streams radiating from a high central area form a pattern of radial drainage while streams forming a centripetal one gather in low-lying land. Reticulate drainage patterns are usually found on floodplains and deltas where rivers often interlace with each other (Fagan & Nanson, 2004).

In GIS, the drainage system can be digitized manually or extracted from the Digital Elevation Model (DEM) by computing the flow direction and accumulation on the terrain (Florinsky, 2009; Nardi et al., 2008; O'Callaghan & Mark, 1984; Ortega & Rueda, 2010; Tarboton, Bras, & Rodriguez-Iturbe, 1991; Tarboton, 1997; Vogt, Colombo, & Bertolo, 2003) and is represented as a river network where each tributary stream is defined by a polyline connected to its main stream. Although semantic information can be added at the river level, no semantic information is computed and stored at the network level. Inside a network, different patterns can be observed and related to other geographical factors. In a drainage basin, a number of factors such as topography, soil type, bedrock type, climate and vegetation cover influence input, output and transport of sediment and water (Charlton, 2008). These factors also influence the nature of the pattern of water bodies (Twidale, 2004). As a consequence, to a certain extent, a drainage pattern can reflect the geographical characteristics of a river network. In structural geology, drainage patterns not only offer clues to geological structure, but also help to decode regional geological chronology (Hills, 1972). Moreover, drainage patterns are useful in the search for minerals (e.g. Binks & Hooper, 1984; De Wit, 1999). At present, much research has been done on the description of drainage patterns in geography and hydrology (e.g. Howard, 1967; Lambert, 2007; Pidwirny, 2006; Twidale, 2004). However, automatic drainage pattern recognition in river networks is not well developed.

Automated map generalization is always an important issue and major challenge in the research of cartography and GIS. Regarded as the skeleton of terrain, the drainage system should be considered in research on automated map generalization in the first place. Further, as the most important component of the drainage system, generalization of rivers properly becomes a focal point. There are several reasons: (1) rivers are an important part of the land, and need to be represented in maps of any kind; (2) rivers are fundamental concepts used for various analyses in geo-science. For instance, geologists can get original slope and original structure from drainage patterns. As a set of line features, river networks are generalized from a large scale to a small scale by two main steps: selective omission and simplification of tributaries (Li, 2007). There are lots of existing methods for tributaries selective omission and simplification, but most research focuses on graph of river networks during its generalization and the generalized results are always inspected by expert cartographers visually. In fact, assessment of the quality of the generalized map is regarded as "*a forgotten consideration*" (Muller, Weibel, Lagrange, & Salge, 1995) and received "*little attention*" (Weibel & Dutton, 1999) in research. Until now scant work has been done on this aspect (Bard, 2004; Zhang, 2012). Considering map generalization from geographic level first (Ai, Liu, & Chen, 2006; Poorten & Jones, 2002), drainage patterns can be considered in the generalization process as patterns are important in generalization and should be explicitly measured and evaluated (Mackaness & Edwards, 2002).

1.2 Research objective and significance

The purposes of this research are to classify the drainage pattern of a river network automatically, and to apply this knowledge to river network generalization and evaluation. There are three objectives:

(1) The first objective is to propose and implement a new method for automatic drainage pattern recognition of a river network.
(2) The second one is to propose a quality evaluation method for river network generalization by assessing if the generalized river network preserves its original pattern.
(3) The last one is to develop a method for river tributary selection with consideration of different factors.

In GIS, such classification can be useful for terrain analysis as it can help provide a qualitative description of the terrain, or it can help with generalization as the process may be adapted to the type of network. At present, many researchers have started to pay attention to geographical features of river networks during the process of generalization (Ai et al., 2006; Buttenfield, Stanislawski, & Brewer, 2010; Stanislawski & Buttenfield, 2011; Stanislawski, 2009), which follows the idea that "*generalization is not a mere reduction of information – the challenge is one of preserving the geographic meaning*" (Bard & Ruas, 2005). Considering drainage pattern as a geographical factor in river network generalization helps to retain geographical features of the networks.

1.3 Structure of the dissertation

The thesis is divided into 6 chapters. Apart from the introduction of the first chapter, a literature review is presented in Chapter 2. Chapter 3 introduces the new method for automatic drainage pattern recognition. An application to cartographical generalization of the method is proposed in Chapter 4 and Chapter 5 presents a new method for tributary selection that makes use of the drainage information. Chapters are organized as follows:

(1) Chapter 2: the related work and definitions about drainage patterns classification and river network organization are first introduced. River network generalization and its qualitative evaluation are then reviewed in detail showing that geographical factors were not explicitly considered in the generalization process and the results evaluation.

(2) Chapter 3: following conclusions from the previous chapter, a novel method for drainage pattern classification from a river network is introduced. Classification relies on different geometric indicators such as the junction angle, the tributary sinuosity and the shape of a catchment. A network belongs to a pattern if its indicators fall into some sets of values. Providing crisp sets as threshold values or intervals is not reliable. Therefore fuzzy sets are defined and thus pattern classification depends on the degree of membership of the network for each pattern. Moreover, a hierarchy is given to organize the drainage tree based on drainage patterns.

(3) Chapter 4: an assessment method for networks generalized by selective omission of their tributaries is presented. The classification method of Chapter 3 is applied to check if the drainage pattern of a river network is preserved after generalization. The evaluation is based on the membership value obtained in the fuzzy logic process so that preservation

of the original meaning can be scored. In this chapter, several methods of river network generalization, such as the elimination based on stroke, length, and catchment area, are tested and evaluated.

(4) Chapter 5 introduces a generalization method that takes into account the drainage information extracted from the pattern classification. A multi-objective optimization problem is set up by mean of a genetic algorithm (GA) to generalize a river network with consideration of different factors. Each drainage pattern has its own characteristics, so river networks with different patterns should consider different factors during the generalization process. The factors, such as drainage pattern membership, stream order, and length, are considered and built into objective functions. In the multi-objective model, different weights are used to aggregate all objective functions into a fitness function. Then, a GA is designed to implement the generalization.

(5) Chapter 6: conclusions of this research are summarized in the chapter discussing the contributions and limitations of the method. In addition, some perspective works are also presented, firstly in the domain of river network and DTM analysis, and secondly extending the application to other dataset modeled by network structures.

Chapter 2 Review of river network classification and generalization methods

2.1 Introduction

A river is a natural watercourse, usually freshwater, flowing towards an ocean, a lake, a sea, or another river. Small rivers may also be called by several other names, including stream, creek, tributary and rill; there is no general rule that defines what can be called a river, although in some countries or communities a stream may be defined by its size. A river is part of the hydrological cycle and a vein of fluvial system. Schumm (1977) conceptualized the fluvial system to consist of three zones (see Figure 2.1): (1) production zone, (2) transfer zone, and (3) deposition zone. The reference is made to sediment process.

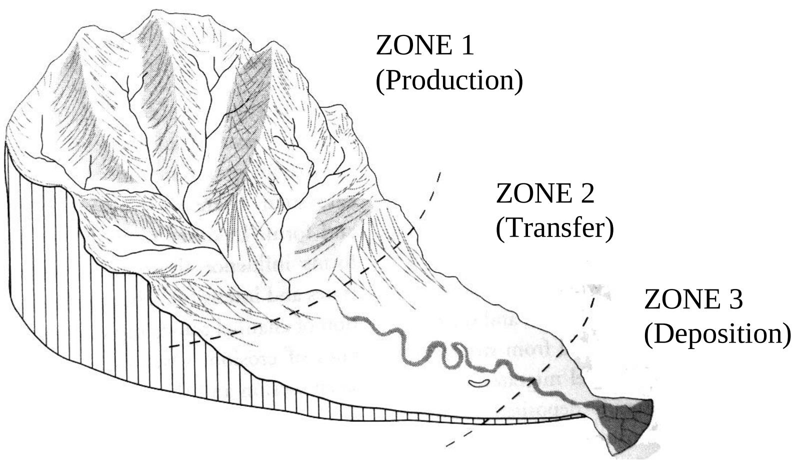


Figure 2.1 Idealized fluvial system (Schumm, 1977, diagram modified from Charlton, 2008)

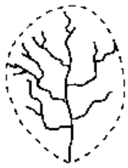
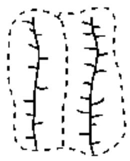
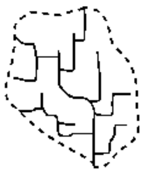
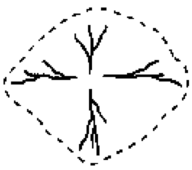
- Zone 1 is the upper portion of the system that is the watershed or drainage basin; this portion of the system functions as the sediment supply.
- Zone 2 is the middle portion of the system that is the river; this portion of the system functions as the sediment transfer zone.
- Zone 3 is the lower portion of the system and may be a delta, wetland, lake, or reservoir; this portion of the system functions as the area of deposition.

With the passage of time, a drainage system achieves a particular drainage pattern where its network of stream channels and tributaries is determined by local geological factors. In order to recognize the drainage pattern of a river network and apply it to generalization, the works of the classification of drainage patterns, river network organization and generalization are reviewed respectively in following sections.

2.2 Classification of drainage patterns and river networks

Drainage patterns are classified on the basis of their form and texture according to the terrain slope and structure. Their shape or pattern develops in response to the local topography and subsurface geology. There are 7 basic types of drainage network patterns as follows (see Table 2.1).

Table 2.1 Drainage network patterns (diagrams modified from Ritter, 2006)

Name	Schematic Diagram	Description
Dendritic		Dendritic pattern is the most common form of river system. In a dendritic river system, there are many contributing streams (analogous to the twigs of a tree), which join together and are the tributaries of a main river (Lambert, 2007).
Parallel		Parallel patterns form where there is a pronounced slope to the surface. Tributary streams tend to stretch out in a parallel-like fashion following the slope of the surface (Ritter, 2006).
Trellis		In a trellis pattern, as the river flows along a strike valley, smaller tributaries feed into it from the steep slopes on the sides of mountains. These tributaries enter the main river at approximately 90 degree angles, causing a trellis-like appearance of the river system (Ritter, 2006).
Rectangular		The rectangular pattern is found in regions that have undergone faulting. Movements of the surface due to faulting offset the direction of the stream. As a result, the tributary streams make sharp bends and enter the main stream at high angles (Ritter, 2006).
Radial		The radial pattern develops around a central peak or dome. This pattern is common to such conically shaped features as volcanoes. The tributary streams flow from the top downward to the bottom around a mountain (Ritter, 2006).
Centripetal		The centripetal pattern is just the opposite of the radial as streams flow toward a central depression. During wetter portions of the year, these streams feed ephemeral lakes, which evaporate away during dry periods (Ritter, 2006).
Reticulate		Reticulate drainage patterns usually occur on floodplains and deltas where rivers often interlace with each other forming a net (Fagan & Nanson, 2004).

According to the description of drainage patterns in Table 2.1, each drainage pattern has its own characteristics. Howard (1967) pointed out that dendritic patterns appear in horizontal sediments or uniformly resistant crystalline rocks with a gentle regional slope at present or at time of drainage inception. In Schumm, Dumont, & Holbrook (2002)'s book, parallel networks have moderate to steep slopes and appear in areas of parallel elongated landforms. Trellis patterns usually exist in dipping or folded sedimentary, volcanic, or low grade sedimentary rocks. A rectangular pattern is with joints and faults at right angles, in which streams and divides lack regional continuity.

Some experimental works have been done concerning morphological dependencies of river channel patterns, such as straight, meandering and braid patterns. Schumm and Khan (1972) determined an experimental relationship between slope and sinuosity for a fluvial channel, which can show threshold changes between pattern types. Here, sinuosity is the ratio of channel length to valley length. Results show that braided patterns appear on steep low-sinuosity channels. Schumm (1977) improved his model and pointed out that pattern adjustments, measured as sinuosity variations, are closely related to the type, size, and amount of sediment load. Although these works (e.g. Knighton, 1998; Lewin & Brewer, 2001) about morphological dependencies apply to river channel patterns rather than river networks, some of the above relationships will be considered in this thesis.

2.3 River network organization

A river network is composed of several connected river segments stored as line entities in GIS. The end points of the river segments are the nodes. There are three types of node: the junction node connecting river segments, the source node corresponding to river springs and the outlet towards where the flow goes. A river network is located in a catchment, also called drainage basin. The catchment controlled by a tributary flowing into a main stream is called a sub-catchment or sub-basin. All these features are illustrated in Figure 2.2. There are two kinds of organization for a river network. One is ordering scheme based on the hierarchical structure of river tributaries, and another is coding system also based on the hierarchical structure, but of drainage basin.

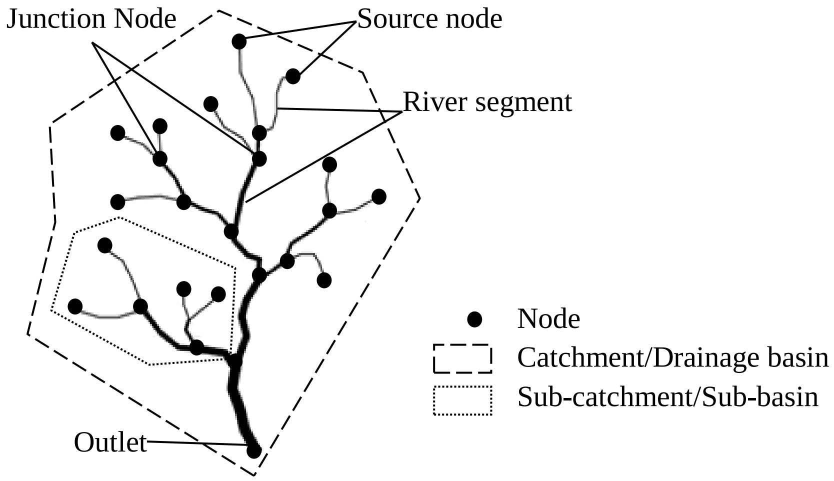


Figure 2.2 Features in a river network (modified from Li, 2007)

2.3.1 Extraction of drainage networks from DEM

In GIS, the drainage system can be extracted from the Digital Elevation Model (DEM) by computing the flow direction and accumulation on the terrain (Florinsky, 2009; O'Callaghan & Mark, 1984; Ortega & Rueda, 2010; Vogt et al., 2003). Verdin & Verdin (1999) presented a system, defined by topographic control of drainage and the topology of the resulting river networks, and implemented it with the North American portions of the GTOPO30 global DEM. Colombo et al. (2007) extracted river networks and catchment boundaries across the European continent from a medium resolution (250m) DEM. In the extraction process, flow directions determination from grid cells is the most important step. Much research on detecting the river flow direction of a river network with additional information from a digital elevation model has been done (e.g. Alves, 1993; Fairfield & Leymarie, 1991; O'Callaghan & Mark, 1984; Vogt et al., 2003).

Moreover, there are several computer tools for extraction of drainage networks. For example, the Watershed Modeling System (WMS) is a comprehensive graphical modeling environment for all phases of watershed hydrology and hydraulics; Topographic Parameterization (TOPAZ) is an automated digital landscape analysis tool for topographic evaluation, drainage identification, watershed segmentation and sub-catchment parameterization; in addition, ArcGIS also has Arc Hydro Tools supporting hydrologic and hydraulic analysis with GIS.

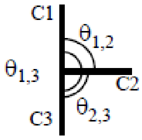
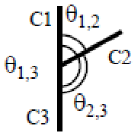
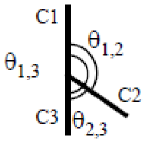
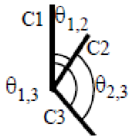
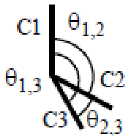
2.3.2 River network construction

In order to build a hierarchical structure of a drainage system automatically, inference of the flow direction and main stream in river networks is a necessary process. Usually, information about river networks only consists of the connectivity of channels, lacking any explicit information about the flow direction of the network (Paiva & Egenhofer, 2000). Although many researches of river networks assume that the flow direction is already known (Coffman & Turner, 1971; Smart, 1970), it is necessary to detect the river flow direction and main stream of a river network automatically.

Instead of using elevation information, a method based on the angles at which river segments connect in river networks has been put forward. The junction geometry describes the information among the related channels. The primary geometric concern is the information about the angles at which the channels flow together. Serres & Roy (1990) has found a set of inference rules (see Table 2.2) that match closely with dendritic river networks, and it is true for about 88% of the junctions in dendritic river networks empirically. Paiva & Egenhofer (2000) has presented an algorithm to find first the main branches of a network, from which it then infers the destination, based on the topology of channels and the angles at which river channels connect at junctions.

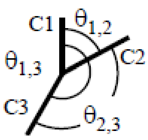
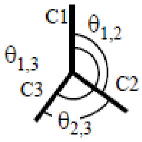
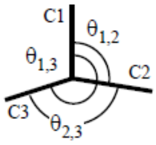
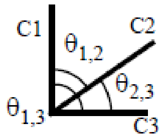
For the main stream detection, Rusak Mazur & Castner (1990) set two rules to determine the main stream: ① It has the same direction as the lower river (without consideration of any other geographic conditions); ② It has the longest length if several streams have a similar direction as the lower river. In (Paiva & Egenhofer, 2000)'s paper, "*the 180 °assumption*" is similar as the first rule above. The assumption presumes that the predominant continuation of the flow direction is along the main channel. The upstream channel that forms an angle closest to 180 °with the downstream channel is considered to be part of the main channel.

Table 2.2 Decision table for network junctions based solely on angles

	1	2	3	4	5
Case					
$\theta_{1,3}$	= 180 °	= 180 °	= 180 °	< 180 °	< 180 °
$\theta_{1,2}$	= 90 °	< 90 °	> 90 °	≤ 90°	≥ 90°
$\theta_{2,3}$	= 90 °	> 90 °	< 90 °	≥ 90°	≤ 90°
Downstream channel	C1 or C3	C3	C1	C3	C1

(*Continued*)

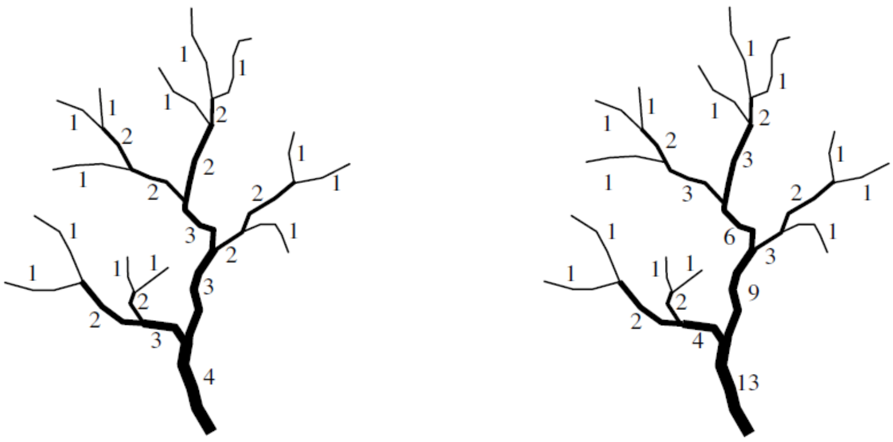
Table 2.2 (*Continued*)

	6	7	8	9
Case				
$\theta_{1,3}$	> 180 °	> 180 °	> 180 °	< 180 °
$\theta_{1,2}$	≤ 90°	≥ 90°	≥ 90°	< 90 °
$\theta_{2,3}$	≥ 90°	≤ 90°	≥ 90°	< 90 °
Downstream channel	C3	C1	C3 **if** $\theta_{1,2} < \theta_{2,3}$ C1 **if** $\theta_{2,3} < \theta_{1,2}$	C3 **if** $\theta_{1,2} < \theta_{2,3}$ C1 **if** $\theta_{2,3} < \theta_{1,2}$

2.3.3 Ordering scheme for river tributaries

Ordering schemes are built by assigning an order number to each tributary. Ordering starts by assigning order 1 to branchless tributaries. The order of a stream is always higher than the order of its tributaries so that the highest order is assigned to the segment connected to the outlet. In this procedure, the Horton-Strahler scheme based on (Horton, 1945) and modified by Strahler (1957), and the Shreve scheme (Shreve, 1966) have been considered the most relevant schemes for the multi-scale representation of river networks (Rusak Mazur & Castner, 1990).

The Horton-Strahler scheme assigns order 1 to all branchless tributaries, and higher order to those receiving tributaries following the river flow direction. This order scheme is illustrated in Figure 2.3A, and can be computed recursively (Gleyzer, Denisyuk, Rimmer, & Salingar, 2004). In the Shreve scheme, the order of a downstream tributary is the sum of the orders of the upper streams, as shown in Figure 2.3B.



A. Horton-Strahler order scheme B. Shreve order scheme

Figure 2.3 Horton-Strahler and Shreve order scheme (from Li, 2007)

2.3.4 Coding system for drainage basins

In order to support GIS-based hydrological analyses, much research on coding drainage networks has been done, which are applied not only to a river network but also to its associated drainage basin.

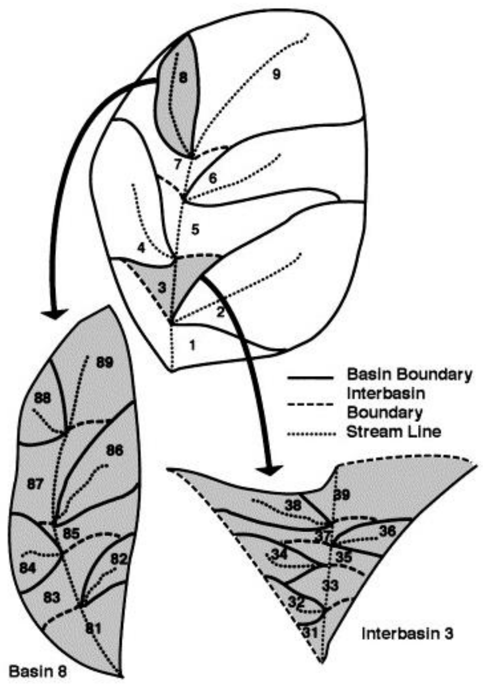


Figure 2.4 Prafstetter codification (from Verdin & Verdin, 1999)

Pfafstetter coding system, proposed by Otto Pfafstetter in 1989, is a subdivision and codification method for describing river basins based on the natural topology of the land surface (Verdin & Verdin, 1999). The system is built into a hierarchal structure from a whole basin to its sub-basins step by step recursively. For a basin, it can be divided into up to a maximum of 10 sub-basins, which are assigned a number from 0 to 9 based on their location and area (Furnans & Olivera, 2001). There is a sample of subdivisions of a basin by applying the Pfafstetter codification illustrated in Figure 2.4.

The advantage of the Pfafstetter coding system is that the code can be used not only to obtain a sub-basin directly but also to decide the topological relationship in the whole basin. Much research has been done to modify and apply Pfafstetter codification method (Fürst & Hörhan, 2009; Jia et al., 2006; Shrestha, Kazama, & Newham, 2008; Verdin & Verdin, 1999).

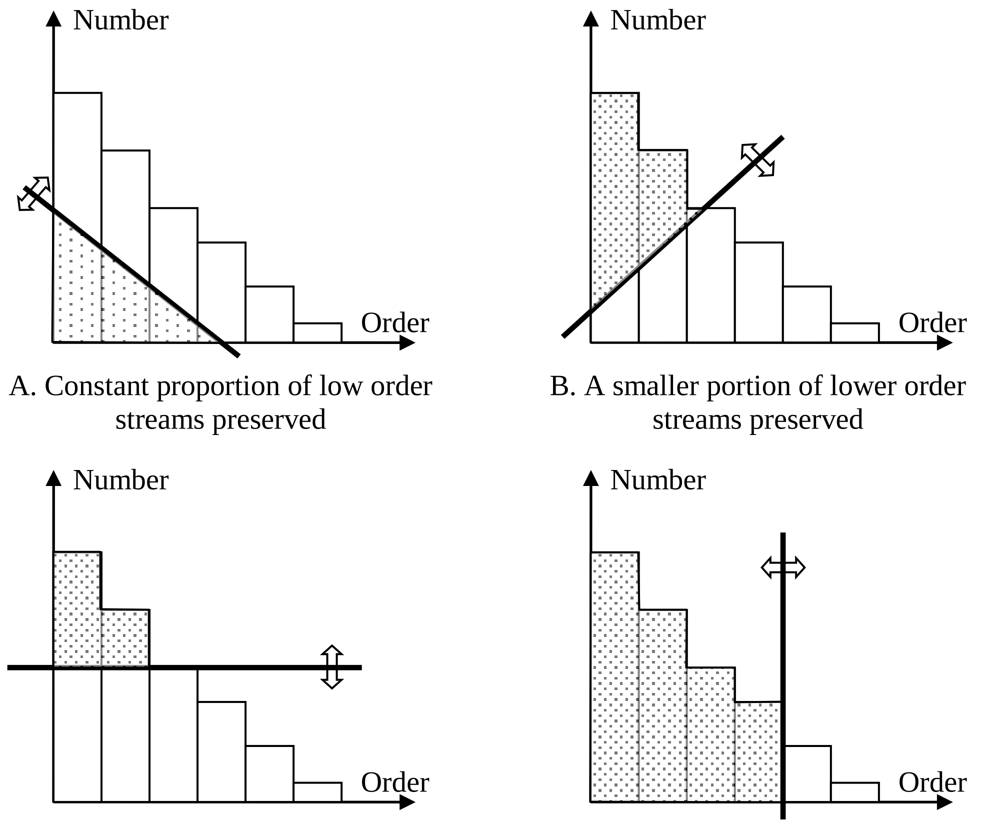
In order to overcome the weakness of the Pfafstetter method, which is not so effective applied in a large scale river network with high spatial resolutions, Li, Wang, & Chen (2010) proposed an efficient and effective codification method to support a complex hydrological model. This new method is already integrated with the established Digital Yellow River Model (G. Wang, Wu, & Li, 2007).

2.4 River network generalization

In general, there are two typical operations in a river network generalization, one is selective omission, and the other is scale-driven generalization (Li, 2007). In Li's book, "*The former eliminates the less important branches, and the latter makes the variation of the selected rivers simpler to suit representations at a smaller scale.*"

2.4.1 Tributary selection

Rusak Mazur & Castner (1990) have given four possible options for the elimination of river tributaries. They are shown in Figure 2.5, in which the x axis is the order of the tributaries and the y axis is the total number of tributaries at each order. The shadowed area indicates the portion to be eliminated. Richardson (1993) presents a method to selected rivers based on Horton order (Horton, 1945) and river length. Thomson & Brooks (2000) apply the Gestalt recognition principles in river network generalization judging the main channel and omitting less important channels. A mainstream is detected based on the strokes using their Horton order and their length. But, determining the main stream using the longest path on clipped river network causes big mistakes. Touya (2007) presents a method for river network selection that relies on the organization of river strokes in hierarchy. His work adds the management of river islands, irrigation zones and allows the building of strokes on a clipped area where some sources are not natural. But, it only focuses on the geometric graphs of river networks, and it does not select river network maintaining geomorphologic structures.



A. Constant proportion of low order streams preserved

B. A smaller portion of lower order streams preserved

C. Same number of low order streams preserved

D. All low order streams not preserved

Figure 2.5 Options for the elimination of river tributaries (from Li, 2007)

Since the distribution of river network is associated with the terrain surface, Wolf (1988) builds a weighted network data structure integrating the drainage, ridge, and peak and pit point. This data structure can determine the significance of a river. The river tree structures have various patterns leading to different generalization strategies. Wu (1997) investigates the characteristics of river tree and develops a method based on buffer spatial analysis to establish the river tree structure. Ai, Liu, & Chen (2006) present a method to focus on the decision of channel importance during the river network generalization applying the integrated hydro-graphic concept, namely watershed area to replace several geometric parameters of river feature. The parameters of density and the upstream drainage area are also used to prune the river network (Stanislawski, 2008, 2009). For man-made ditches, Sandro, Massimo, & Matteo (2011) present a typification method for generalization of groups of ditches, which are represented as a regular pattern of straight lines.

For applying drainage patterns to river tributary selection, little research focuses on this aspect. Touya (2007) and Jiang et al. (2009) both acknowledge the drainage pattern as an important factor in river network generalization, but no details about how to apply them are described. In order to preserve the main hydrographical properties, Jiang et al. (2009) obtained a simple representation of river networks by keeping the same drainage pattern after a selection operation but they did not go further to explain how patterns are preserved. In different drainage patterns, different factors should be considered during the river network generalization.

In order to consider different geographical factors, such as river length, river tributaries spacing, catchment area, and river network density, there is a multi-objective optimization (also known as multi-criteria or multi-attribute optimization) process in river tributary selection. Zhai et al. (2006) have built a structure river data model facing the river system’s spatial knowledge, and selected the river tributaries automatically based on genetic multi-objective optimization algorithm. In his model, the indicators, such as length, interval and importance of a river, have been taken into account while selecting the rivers.

2.4.2 Tributary simplification

After river tributary selection, several operations can be applied to the selected tributaries to simplify the complexity of line features. For a line, there are many simplification algorithms which have been divided into hierarchical method and non-hierarchical method. The purpose of line generalization is to remove non-relevant information on the line with the minimum point to keep emphasis on details of prime importance for visualization (Weibel & Dutton, 1999).

Line generalization is often perceived as a point-reduction or smoothing process. In point-reduction approaches, the main methods take geometric parameters or function of these parameters as criteria (see Table 2.3).

Table 2.3 An overview of line point-reduction methods

Algorithm	Foundation	Classification
AEG algorithm (Lang, 1969)	Perpendicular distance	Sequential algorithm
Li algorithm (Li, 1988)	Minima and maxima	Iterative algorithm
Douglas-Peucker algorithm (Douglas & Peucker, 1973)	Perpendicular distance	
Ansari-Delp algorithm (Ansari & Delp, 1991)		
Visvalingham-Whyatt algorithm (Visvalingam & Whyatt, 1993)	Area	
Rosenfeld-Johnstion algorithm (Rosenfeld & Johnston, 1973)	Cosine value	Algorithm with functions of geometric parameters as criteria
Teh-Chin algorithm (Teh & Chin, 1989)	Distance/Chord ratio	
Nakos-Mitropoulos algorithm (Nakos & Miropoulos, 2003)	Local length ratio	

In smoothing approaches, algorithms can be performed in the space domain or in the frequency domain. An overview of line smoothing methods is in Table 2.4.

Table 2.4 An overview of line smoothing methods

Technique/Method	Principle	Performing domain
B-spline (Guilbert & Lin, 2006; Saux, 1998)	Curve fitting	Space domain
Snake (Burghardt & Meier, 1997; Kass, Witkin, & Terzopoulos, 1988; Steiniger & Meier, 2004)		
Empirical mode decomposition (Li et al., 2004)	Component exclusion	Frequency domain
Fourier transforms (Brenner, 1969; Cooley & Tukey, 1965)	Frequency cutting	
Wavelet transforms (Balboa & López, 2000; Plazanet, Affholder, & Fritsch, 1995)		

These researches above are applied to a single line transformation, but the selected rivers are a set of line features representing a network. Gutman & Weaver (2012) applied wavelet transforms to river network generalization, and the connectivity issue of a river network for all scales has been settled. However, generalization of a set of lines as a whole should consider more issues such as spatial conflicts (Li, 2007), which are complex problems.

2.5 Generalized river network quality assessment

From the literatures, little research drew attention to the aspect of the assessment of generalization (Muller et al., 1995; Weibel & Dutton, 1999). Traditionally, the generalization is evaluated by visual assessment by the cartographic experts to grade the quality by questionnaires (Weibel, 1995). This method is based on knowledge of experience of experts, and it is rather subjective (Joao, 1998; Mackaness & Ruas, 2007). But, generalization is important on map features and analysis, and the methods for quantify generalization results should be developed (Joao, 1998). Bard (2004) proposed a general method to evaluate the cartographic generalization. For the quality assessment of river network generalization, especially in the operation of selective omission, the related work is few. The most relative one is that a Coefficient of Line Correspondence (CLC) is calculated to evaluate the generalized data by comparing with the existing data (Buttenfield et al., 2010; Stanislawski, 2009). CLC is given based on length only, which cannot assess the generalized river network comprehensively. To the river tributaries simplification, only some related studies focus on line features as single geometric primitive (Joao, 1998; Skopeliti & Tsoulos, 2001). In general, the methods to evaluate river network generalization quality are not well developed, and visual assessment is still often used.

2.6 Summary

From the literature review of previous work on drainage pattern classification, river network organization and generalization, the summaries are given as follows:

(1) At present, much research has been done on the definition, classification and description of drainage patterns in geography and hydrology. Many scholars work on predicting river channel patterns from in-channel characteristics, such as slope and discharge, but not drainage patterns. Although drainage pattern is recognized as an important element in GIS, its classification has not yet been considered. Therefore, this thesis studies the geometric and topologic characteristics of each type of drainage pattern to allow automatic river network classification.

(2) The current methods on river networks generalization have been well developed, and much work has been done on river network selective omission and selected tributaries simplification. But the methods of quality assessment for generalized river network are few. Recently, many researchers have paid more attention on geospatial patterns in cartographic generalization (Mackaness & Edwards, 2002; Zhang, 2012). This thesis will provide an evaluation method to check how much the generalized river network preserves the drainage pattern.

(3) For applications of drainage patterns, river network generalization is an important one. Much research on river network generalization focus on geometric properties only. Map generalization is the process of "*information abstraction*" rather than a "*data compression*" (Ai et al., 2006). Drainage patterns can reflect the geographical features of river networks to some extent. So, this thesis will apply drainage patterns into river networks generalization.

Chapter 3 A new approach for automatic drainage pattern recognition

3.1 Introduction

As reviewed in Section 2.2, seven types of drainage pattern were introduced. Each drainage pattern has its own geometric and topological characteristics. Among them, the first five patterns (the dendritic, parallel, trellis, rectangular and reticulate patterns) are characterized by the geometric organization of the river segments inside the patterns while radial and centripetal patterns depend on the spatial organization of a group of networks. This chapter focuses on the description of individual networks, and addresses the identification of the first five patterns based on geometric characteristics identified inside a network. A new approach is proposed to recognize the drainage pattern of a given river network automatically. A reticulate pattern is identified by graph theory (Bondy & Murty, 2008). For others, according to the geometric characteristics of drainage patterns, geometric indicators are defined. A classical usage of indicators to distinguish objects is to set threshold value based on the knowledge and experience of users. However many networks would be unclassified or polymorphic if the threshold values are too restrictive or too loose. Therefore, as fuzzy logic (Zadeh, 1965) can provide an approximate way rather than fixed or exact threshold values to reflect the inherent vagueness of drainage patterns, it is more appropriate to be applied to this work. Eight predicates will be extracted from defined indicators and set as membership functions. Combing the predicates with fuzzy operators, each pattern is determined by a rule. Then, the drainage pattern can be recognized by a fuzzy process. In addition, as a river network can be identified as a drainage pattern, inside sub-networks also can be classified as may be different patterns. A hierarchy of drainage tree is also provided to organize the river network based on drainage patterns in this chapter.

The next sections of this chapter are organized as follows. Section 3.2 gives a summary of the characteristics of drainage patterns based on the literature review. Sections 3.3 and 3.4 introduce the methodologies of the drainage pattern classification based on graph theory and fuzzy logic respectively. Then, an organization of drainage tree for a river network according to pattern is given in Section 3.5. In Section 3.6, the methods are tested on a river network and results are discussed. Finally, summaries and conclusions are presented.

3.2 Drainage pattern characteristics

Based on the description of different types of drainage patterns, each pattern has its own geographical characteristics, which can be reflected in some quantifiable variable related to some topological and geometrical aspects. Therefore, each pattern can be characterized by a combination of different variables. In Table 2.1 (p. 9), patterns are characterized by different geometric indicators measured on each segment of a network or describing the shape of the drainage while radial and centripetal patterns depend on the spatial organization of a group of networks. This work focuses on the description of individual patterns. A list of characteristics for each of them is proposed and shown in Table 3.1.

Table 3.1 Drainage pattern characteristic

Drainage pattern	Geometric and Topologic Characteristic
Dendritic	-Tributaries joining at acute angle
Parallel	- Parallel-like - Elongated catchment - Long straight tributaries - Tributaries joining at small acute angle
Trellis	- Short straight tributaries - Tributaries joining at almost right angle
Rectangular	- Tributary bends - Tributaries joining at almost right angle
Reticulate	- Tributaries cross together forming a cycle

Non-reticulate river networks are represented by a hierarchical graph and are characterized by geometric parameters related to the length and angle measured in the network. The reticulate pattern is a specific pattern because rivers intersect and cross together like a net. Due to that, a river network would form a cycle instead of a tree. Therefore, reticulate networks are identified first. They are taken out of the graph and replaced by nodes. The remaining part forms a hierarchical network with the outlet as the root which can be characterized by one of the four remaining patterns. Recognition and removal of reticulate networks is discussed in the next section while identification of other patterns based on geometric indicators is introduced in Section 3.4.

3.3 Reticulate pattern recognition based on graph

Graph theory has a long history in the study of graphs (Biggs, Lloyd, & Wilson, 1986). In mathematics and computer science, graphs are used to model relations between objects from a certain collection, where the objects are called vertices and relations that link some pairs of vertices are edges (Trudeau, 1994). Graphs are widely used in many areas, such as computer science, biology and GIS. In GIS, geometric networks are similar to graphs, and borrow many concepts from graph theory to perform spatial analysis on road networks or utility grids (e.g. Buyya, 2005; Porta, Crucitti, & Latora, 2006; Zhan & Noon, 1998).

In graph theory, a graph comprises a set *V* of vertices (or nodes) with a set *E* of edges (or links, lines), and is represented as $G = (V, E)$. A cut-edge (also known as a bridge) is an edge removal of which produces a graph with more components than the original (Bondy & Murty, 2008). Equivalently, an edge is a bridge if, and only if, it is not contained in any cycle. Figure 3.1 illustrates the cut-edges in an undirected graph, where the dashed line is cut-edge and the solid line is an edge contained in a cycle.

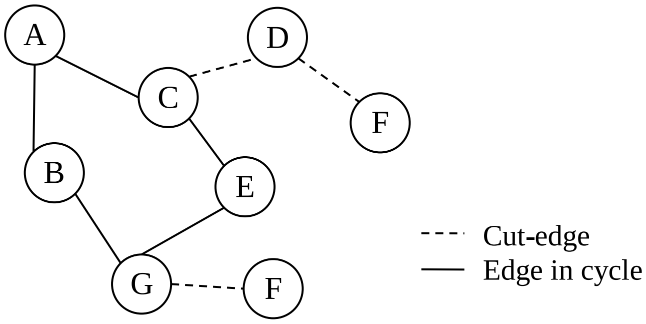


Figure 3.1 Cut-edges in an undirected graph

Considering the river network as a graph by setting river segments as edges and nodes in river network as nodes in graph, all cut-edges are found using a bridge-finding algorithm (Tarjan, 1974). Edges which are not identified as cut-edges are components of cycles and form reticulate patterns.

3.4 Recognition of other patterns based on fuzzy logic

In this section, some geometric quantitative indicators are defined to recognize dendritic, parallel, trellis and rectangular patterns. From the geometric characteristics of drainage patterns in Table 3.1, the most important variable is the angle formed by a tributary with its main stream at a junction node. The average junction angle of all angles in a catchment is one quantitative indicator. In order to distinguish rectangular pattern, the shape of a tributary is also needed. In this pattern, tributary streams make sharp bends almost at a right angle. The amount of bending of tributaries can be estimated by the sinuosity of the river segments. Another difference between parallel and trellis pattern is length: the tributaries in parallel pattern are long relative to trellis. The average length ratio of tributaries to the main stream is the third indicator. The fourth indicator is the catchment elongation used to identify parallel patterns in an elongated basin. The catchment elongation is characterized by the ratio of long edge to short edge of the Minimum Bounding Rectangle (MBR) of the catchment. If the catchment is elongated, this ratio is large.

3.4.1 Geometric indicators

(1) Junction angle

The angle at a junction is a useful parameter that can be used in flow direction and main stream inference (Paiva & Egenhofer, 2000; Serres & Roy, 1990). In general, a tributary joins into a main stream (Figure 3.2a), or two tributaries gather together forming a new stream (Figure 3.2b). In this situation, the angle is easy to obtain. However, it is further complicated when several tributaries (more than three river segments join at a junction) flow into a main stream at the same place (Figure 3.2c and d). In Figure 3.2, the arrow refers to the flow direction, and river segments in bolder line are with a higher stream order.

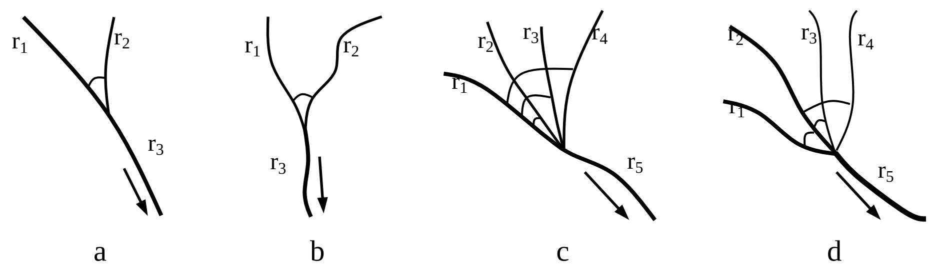


Figure 3.2 Different cases of river segments joining at a junction

In the case of three river segments joining at one junction, the angle is formed by the two upper river segments (e.g. river segments r_1 and r_2 in Figure 3.2a and b). In another case, where more than three river segments connect to a junction, the most important thing is to find the main stream in the upper river segments to measure the angles between the tributaries and the main stream. In order to get the main stream, the stream order is considered in the first place. The upper river segment with the highest order is the main stream. If there are two or more upper river segments with the same highest order, Rusak Mazur & Castner

(1990) set two rules to determine the main stream: ① It has the same direction as the lower river (without consideration of any other geographic conditions); ② It has the longest length if several streams have a similar direction as the lower river.

For example in Figure 3.2c, r_1 is the main stream because it has the highest order of all four upper river segments r_1, r_2, r_3 and r_4. So angles are formed by r_1 with r_2, r_3 and r_4 respectively. In Figure 3.2d, r_1 and r_2 have the same order, but r_2 is the main stream because it has the same direction from the junction with lower river r_5. Three angles are computed in the average, which are formed by r_2 with r_1, r_3 and r_4 respectively.

For a tributary joining a main stream at junction P_1, supposing points P_2 and P_3 are the "from" nodes of the upper stream and the tributary, the junction angle $\angle P_2P_1P_3$ can be computed by the law of cosines:

$$\angle P_2P_1P_3 = \arccos\left(\frac{a^2+b^2-c^2}{2ab}\right), \tag{3.1}$$

where a is the distance between P_1 and P_2, b is the distance between P_1 and P_3, and c is the distance between P_2 and P_3.

The first parameter is the junction angle between the tributaries and the main stream. The parameter is given by the average value α of angles measured at all junctions. The dendritic pattern only requires that junction angles are acute, which can be translated by $\alpha < 90°$. Parallel patterns are characterized by angles more acute than in dendritic patterns, therefore $\alpha << 90°$. For trellis and rectangular patterns, tributaries join at a right angle and $\alpha \approx 90°$.

(2) Sinuosity

Schumm (1977) set the sinuosity variable of a stream as the ratio of the channel length to the valley length to quantify how much a river or stream meanders. In GIS, a stream is stored as a polyline. Then, the sinuosity can be approximately calculated as polyline length divided by length between end points (Figure 3.3).

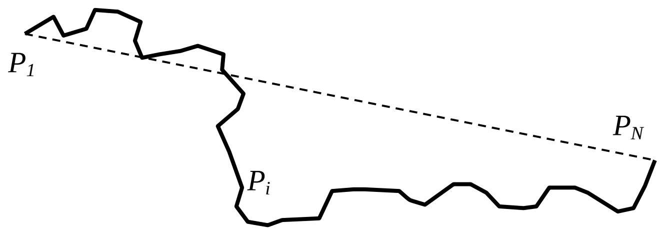


Figure 3.3 River segment represented as a polyline

Supposing a river segment is composed of N points P_i with P_1 and P_N the end points, the sinuosity ratio SI is

$$SI = \sum_{i=1}^{N-1} Dis(P_i, P_{i+1}) \Big/ Dis(P_1, P_N) \quad (3.2)$$

where *Dis*() is the distance between two points.

A perfect straight stream would have a sinuosity ratio of 1; while the higher this ratio is above 1, the more the stream meanders. If the sinuosity ratio is equal to or is greater than 1.5, the stream is considered to be meandering (Ritter, 2006). In both trellis and rectangular patterns, tributaries connect to the main stream at right angles. However, in trellis, tributaries are straight, while in rectangular pattern, most tributaries have sharp bends. A tributary is considered to have sharp bends if it has a high sinuosity. The indicator that is chosen is not the overall sinuosity of the network as a rectangular drainage can contain straight and sinuous streams which may yield a relatively low sinuosity value. Instead, the number of bended tributaries is considered. A parallel drainage or a trellis shall have very few bended tributaries in comparison to a rectangular drainage. A second indicator, the percentage of bended tributaries β is used. This parameter is calculated as the number of bended tributaries divided by the total number of tributaries, where a bended tributary has the sinuosity ratio ≥ 1.5. A rectangular pattern should yield a high value of β while, in trellis and parallel, β should tend towards 0.

(3) Length ratio

Long tributaries in a parallel pattern and short tributaries in a trellis pattern are relative concepts in geography. The river absolute length cannot be used to distinguish different drainage patterns directly. This work takes the length ratio between the tributaries and the main stream as an indicator. Here, the main stream is not only a river segment straight connected to the tributary, it is composed of several segments connected together with the same direction and same order. This is illustrated in Figure 3.4, where the arrow refers to the flow direction, and the river segments in dashed boxes are main streams.

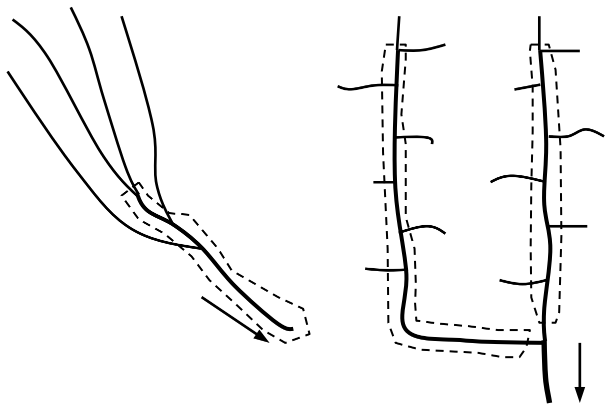


Figure 3.4 Main streams calculated in length ratio

The parameter average length ratio γ is used to distinguish parallel and trellis patterns. In parallel patterns, tributaries have long length so $\gamma > 1$; otherwise, $\gamma \ll 1$ indicates that most tributaries are shorter than the main stream, as expected in a trellis.

(4) Catchment elongation

The exact location of the catchment area is usually computed from the DEM which is not available. Approximations can be obtained from the river network such as the convex hull, the axis-aligned bounding box (AABB) or the oriented minimum bounding rectangle (MBR) (Figure 3.5). In this work, the objective is to estimate whether the catchment is elongated or not. The MBR of the river network is considered as it follows the orientation of the network. The breadth of the river network is given by the length of the MBR side that forms the largest angle with the main stream. The length of the other side which is roughly aligned with the main stream corresponds to the depth. The elongation is defined by the ratio between its depth and breadth. For example, in Figure 3.5, the depth is less than the breadth. The catchment area is not elongated, so that the drainage cannot be considered as parallel. Parallel and trellis patterns form in elongated catchments and are therefore characterized by a high elongation δ.

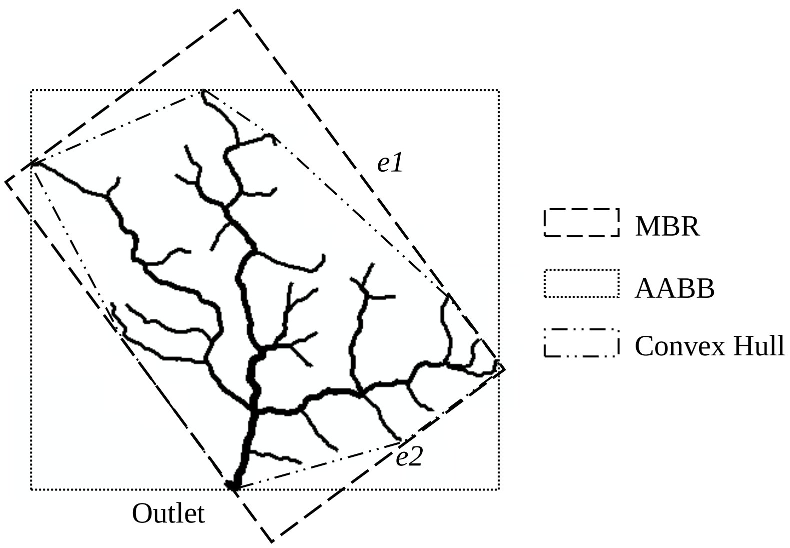


Figure 3.5 MBR of a river network. The edge *e1* has a bigger angle with mainstream, *ratio* = *e2*/*e1* < 1, it is not an elongated river basin.

Geometric characteristics of different patterns presented in Table 3.1 are only defined qualitatively. In order to identify patterns based on these characteristics, statistical measures are obtained from the network and compared with threshold values. The different indicators are summarized in Table 3.2. They are expressed by qualitative predicates and are translated into geometric indicators. These indicators can be directly implemented and measured on a river network. Values associated with each pattern are vague as they represent qualitative properties, and classification into one pattern depends on several of these values.

Table 3.2 List of indicators

Drainage Pattern	Average Junction Angle (α)	Bended Tributaries Percentage (β)	Average Length Ratio (γ)	Catchment Elongation (δ)
Dendritic	Acute $\alpha < 90°$	-	-	Broad $\delta < 1$ or $\delta \approx 1$
Parallel	Very acute $\alpha << 90°$	Not bended $\beta \rightarrow 0$	Long $\gamma \approx 1$ or $\gamma > 1$	Elongated $\delta >> 1$
Trellis	Right angle $\alpha \approx 90°$	Not bended $\beta \rightarrow 0$	Short $\gamma << 1$	Elongated $\delta >> 1$
Rectangular	Right angle $\alpha \approx 90°$	Bended $\beta \rightarrow 100\%$	-	-

Setting crisp threshold values defining the acuteness of an angle or the breadth of a catchment is an empirical task which relies on the user's judgment and expertise and which does not reflect the inherent vagueness of drainage patterns. Furthermore, they do not provide a robust enough classification. Too restrictive threshold values will leave many networks unclassified while too loose values will end up in networks that may belong to different patterns. Therefore, assertion of each predicate is not defined by crisp sets of values but by fuzzy sets and the membership to a set is based on fuzzy logic (Zadeh, 1965).

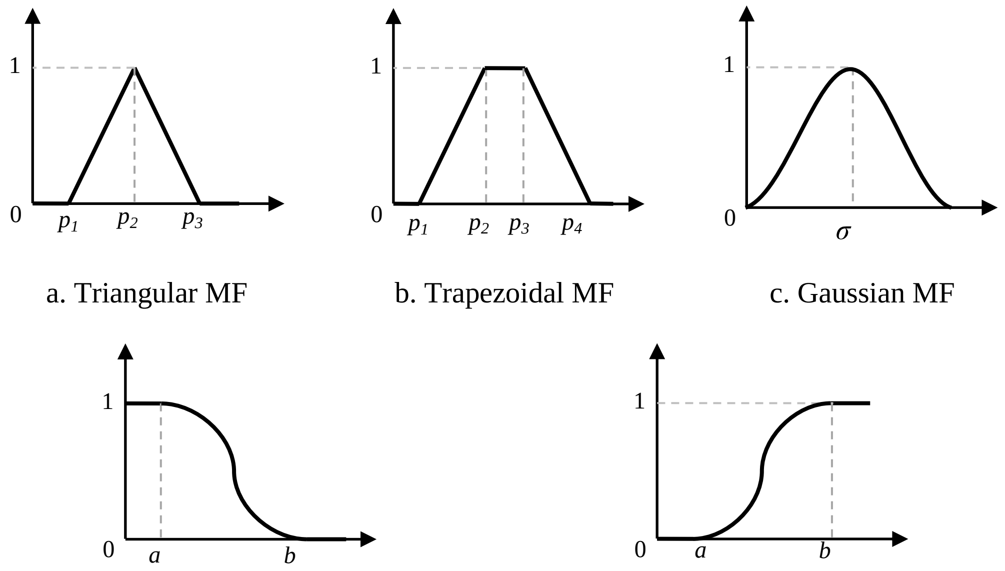
3.4.2 Fuzzy logic process

3.4.2.1 Fuzzy logic fundaments

Zadeh (1965) introduced fuzzy logic in the proposal of fuzzy set theory. Fuzzy logic has been applied to many fields, such as control theory (H. O. Wang, Tanaka, & Griffin, 1996; L.-X. Wang, 1993; Ying, Siler, & Buckley, 1990) and artificial intelligence (Fukuda & Shibata, 1992; Liao & Tsao, 2004; Yen & Langari, 1998). Fuzzy logic is composed of fuzzy sets, fuzzy operators and fuzzy rules.

(1) *Fuzzy set*. A fuzzy set is a set whose membership is not defined by a binary value (an element belongs or not to a set) but by a value between 0 and 1 corresponding to different grades of membership. A membership function (MF) associated with a given fuzzy set maps an input value to its appropriate

membership value. There are five common MFs in use as shown in Figure 3.6. Fuzzy set theory allows approximated reasoning on values which are imprecise or incomplete.



a. Triangular MF　　b. Trapezoidal MF　　c. Gaussian MF

d. Z curve MF　　e. S curve MF

Figure 3.6 Five common MFs

The Gaussian, Z and S curve MFs will be used in this work, and they are described mathematically as follows.

The Gaussian function is given as

$$g(x;\mu,\,\sigma)=e^{\frac{-(x-\sigma)^2}{2\mu^2}}\,, \tag{3.3}$$

where x is input, σ is center and μ controls the width of the curve.

The Z curve MF is a spline-based function of input x,

$$z(x;a,b)=\begin{cases} 1, & x\le a \\ 1-2\left(\dfrac{x-a}{b-a}\right)^2, & a<x\le\dfrac{a+b}{2} \\ 2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2}<x<b \\ 0, & x\ge b \end{cases}, \tag{3.4}$$

where *a* and *b* are the extremes of the sloped portion of the curve, and $a < b$.

The S curve MF is a mirror-image function of Z curve,

$$s(x;a,b) = \begin{cases} 0, & x \le a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a < x \le \dfrac{a+b}{2} \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2} < x < b \\ 1, & x \ge b \end{cases}, \tag{3.5}$$

where *a* and *b* are the extremes of the sloped portion of the curve, and $a < b$.

(2) *Fuzzy operator*. In fuzzy logic, the truth of any statement is a matter of degree between 0 and 1. Zadeh (1965) suggested the minimum, maximum and complement methods for AND, OR and NOT operators respectively. For two fuzzy set values A and B within the range (0, 1), fuzzy logic operations are (Figure 3.7):

$$A\ AND\ B = min(A, B) \tag{3.6}$$

$$A\ OR\ B = max(A, B) \tag{3.7}$$

$$NOT\ A = 1 - A \tag{3.8}$$

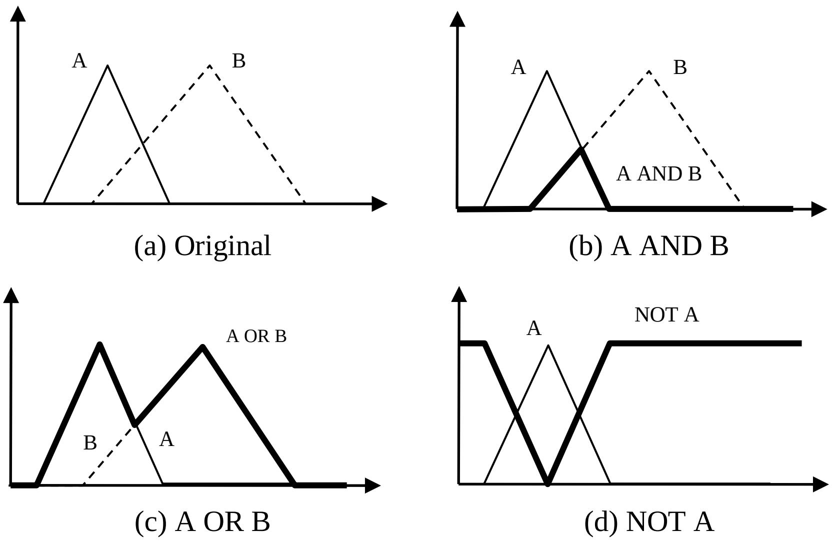


Figure 3.7 AND, OR and NOT operations

(3) *Fuzzy rule*. Fuzzy rules, also called IF-THEN rules, are used to represent the conditional statements with fuzzy sets and fuzzy operators. A single fuzzy IF-THEN rule is like

IF (*a is X*) AND (*b is Y*) THEN (*c is Z*),

where *a* and *b* are input variables, *c* is output, and *X*, *Y* and *Z* are defined by fuzzy sets. Here, it should be noted that there is no ELSE part in a fuzzy rule.

Usually, there are several fuzzy rules for a fuzzy logic application. In the process, all rules should be evaluated, and the outputs can be aggregated to get a result, which is also a fuzzy set sometimes. Therefore, especially for a fuzzy control system, defuzzification is typically needed, which is a process of producing a quantifiable result from a fuzzy set (Leekwijck & Kerre, 1999).

3.4.2.2 Fuzzy logic applied in drainage pattern recognition

In this research, fuzzy set theory is used to perform classification with predicates that cannot be asserted as true or false in all cases but require gradual assessment. A total of eight predicates are extracted from indicators defined in Table 3.2. They are:

- α IS acute
- α IS very acute
- α IS right
- β IS bended
- γ IS long
- γ IS short
- δ IS broad
- δ IS elongated

The degree of membership to each predicate is asserted by a MF. A MF is a curve that defines how each element in fuzzy set is mapped to a membership degree between 0 and 1. Membership degrees of the first two predicates are defined by Z curves, i.e. asymmetrical polynomial curves open to the left (Figure 3.8) of the form $z(\alpha;\ a,\ b)$ where α is the junction angle and *a* and *b* locate the extremes of the sloped portion of the curve. The degree of membership is 1 if $\alpha < a$ and 0 if $\alpha > b$. If $a < \alpha < b$, the degree is decreasing. Obviously, a very acute angle should be smaller than an acute angle so that $a < a'$ and $b < b'$ in Figure 3.8. MFs can be non-zero for a same α value. That means that an angle may be considered as very acute, acute and right at different degrees. A Gaussian distribution curve $g(\alpha;\ \sigma,\ m)$ is used to define the degree of membership to the third predicate (Figure 3.8). The value m is the average angle on which the function is centered and is equal to 90°. Parameter σ controls the width of the curve; the larger it is, the broader the curve.

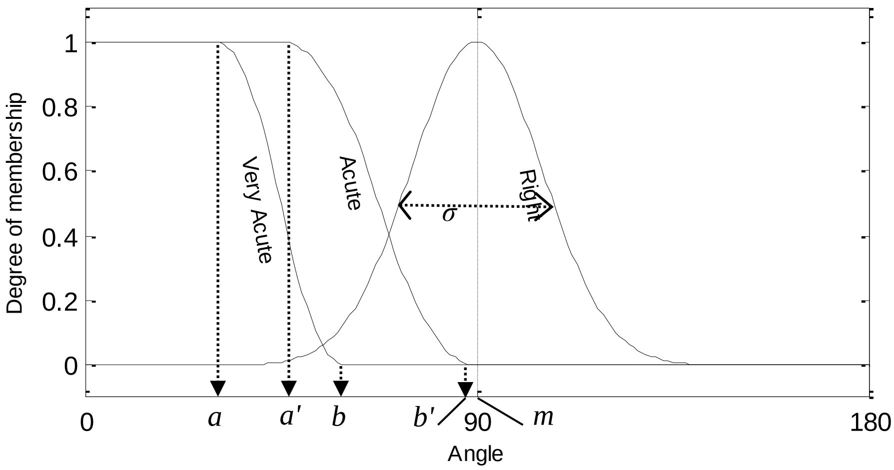
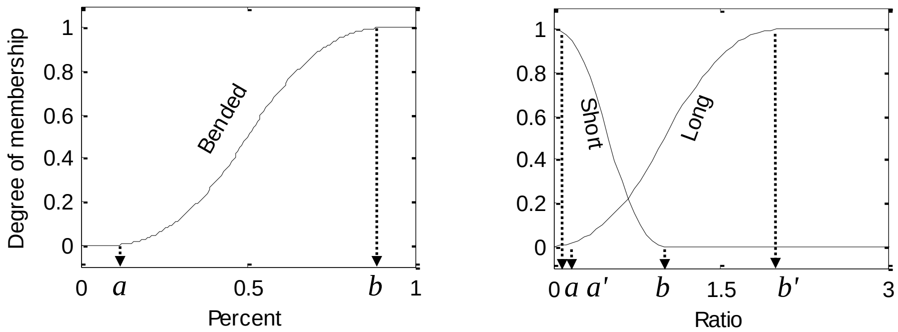
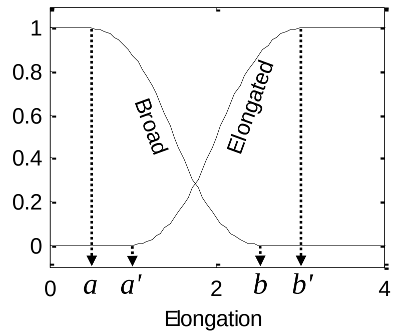


Figure 3.8 MFs for very acute, acute and right angle, input is the junction angle α



(a) MF for bended tributaries, input is bended tributaries percentage β

(b) MF for short tributary, input is average length ratio γ



(c) MF for elongated catchment, input is catchment elongation δ

Figure 3.9 MFs for bended tributaries, short tributary and elongated catchment

The degree of membership to the bend, long tributaries and elongated catchment predicates are estimated by S curves, i.e. asymmetrical polynomial curves open to the right, of the forms $s(\beta; a, b)$, $s(\gamma; a', b')$ and $s(\delta; a, b)$ (Figure 3.9a-c). The smaller the input value, the smaller the degree of membership. Finally, the degree of membership to the short tributaries and broad catchment predicates are characterized by a Z curve where a small ratio has a high degree (Figure 3.9b and c).

Combining the predicates in more complex rules characterizing each drainage pattern is done by using fuzzy Boolean operators AND, OR and NOT. Each pattern of Table 3.2 is defined by the following IF-THEN rules based on fuzzy logic operations:

(1) IF (α IS acute) AND (δ IS broad) THEN pattern IS dendritic
(2) IF (α IS very acute) AND NOT (β IS bended) AND (γ IS long) AND (δ IS elongated) THEN pattern IS parallel
(3) IF (α IS right) AND NOT (β IS bended) AND (γ IS short) AND (δ IS elongated) THEN pattern IS trellis
(4) IF (α IS right) AND (β IS bended) THEN pattern IS rectangular

In fuzzy logic, there is no ELSE rule and all the rules should be evaluated. Therefore, each network is given a degree of membership for each pattern. To get a crisp decision, the maximum-method is used to defuzzify the set of singletons and the pattern with the maximum degree of membership is chosen. The process is illustrated in Figure 3.10 and goes through the following steps:

Step 1: input the indicators α, β, γ and δ of a river network;

Step 2: evaluate all rules according to fuzzy inputs by applying logic operations, and obtain the outputs of all rules;

Step 3: defuzzify the results and get the final output as the pattern with maximum degree.

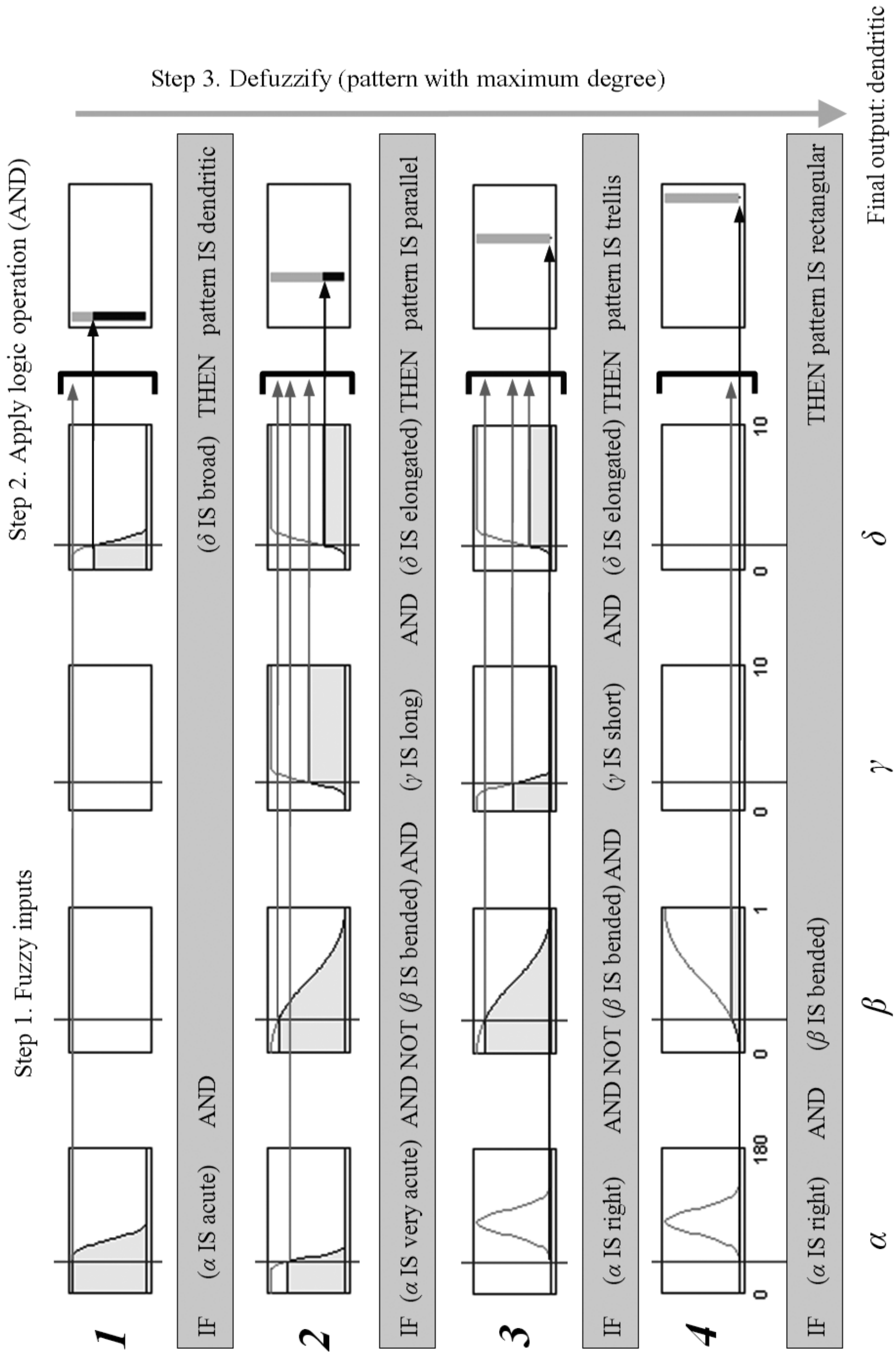


Figure 3.10 The fuzzy logic process for drainage pattern recognition

3.5 Drainage tree construction

The previous section analyzed the drainage pattern of a given river network by comparing the shape of a main stream with its tributaries. As the graph formed by the river network has a tree structure, sub-trees formed by tributaries can also be characterized with a drainage pattern, forming a drainage tree describing the network at different levels of detail. Two kinds of sub-networks are analyzed: reticulate networks and sub-trees formed by considering tributaries as main rivers with their own tributaries at a higher level of detail.

Drainage patterns form a hierarchical structure following the river network structure however one river stream may not follow only one pattern but can go through different patterns along its course. Therefore, after recognition, adjacent patterns of the same type are merged in the drainage tree (Figure 3.12). The whole process includes 4 steps in sequence:

(1) Identify reticulate patterns in the river network (Section 3.3);
(2) Identify all sub-networks forming the drainage tree;
(3) Characterize drainage patterns in the sub-tree (Section 3.4);
(4) Merge adjacent patterns of the same type.

Following Section 3.5.1 describes the process building the drainage tree (step 2). Section 3.5.2 presents the merging process (step 4).

3.5.1 Drainage tree construction based on patterns

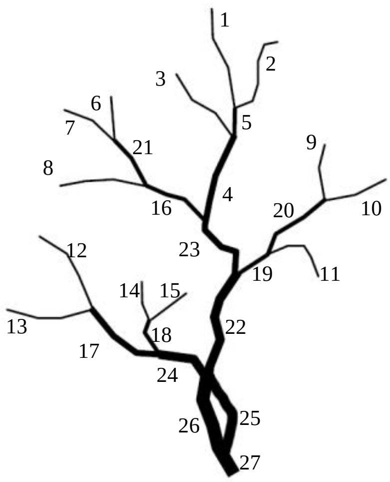
Construction of the drainage tree is done in two steps. First, all reticulate networks are identified as cycles and removed from the river network as described in Section 3.3. Second, the main stream of each sub-network is identified. Levels of representation are defined by the segment order following the Horton-Strahler scheme (Strahler, 1957). Each branchless segment is assigned an order 1. A segment is assigned an order equal to the highest order of its tributaries or to the highest order plus one if there are several tributaries of this order (Figure 3.11a). This order can be computed recursively (Gleyzer et al., 2004).

In the process, a pattern is defined for each sub-network therefore the drainage tree is structured following the sub-networks. In the river network, streams are defined by adjacent segments sharing the same Horton-Strahler order. A sub-network is defined by one river stream and all its tributaries down to order 1 segments. The set of all river streams forms a tree structure where the river stream of the highest order, connected to the outlet and representing the whole network, is the root. Reticulate networks which were extracted from the river network form sub-networks and are put back into the tree structure. A reticulate network is a sub-network connected to the closest river stream of highest order.

The data structures of a stream and a river segment is illustrated in Appendix A, where the data structure of a junction is also provided. The features of junctions and river segments correspond to the features described in Figure 2.2. In a river segment, the “from” and “to” junctions show the direction of river flow.

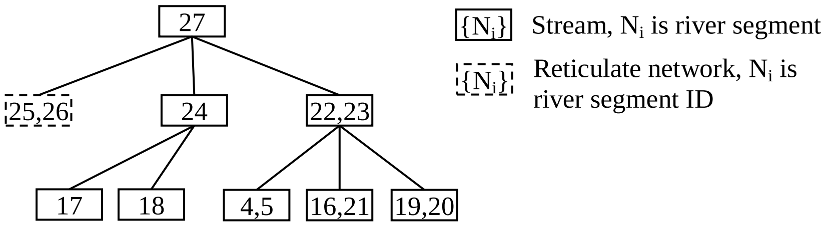
The drainage tree is built by starting from the outlet and, for each node of the tree, by adding the river streams or reticulate networks below. The algorithm ***Build(st, rs)*** building the hierarchy of sub-networks is described in Appendix B.1. This algorithm starts with a new stream as root and a river segment containing outlet node, and the river segment has been add to the stream.

An example of drainage tree obtained from the river network of Figure 3.11a is shown in Figure 3.11b. Each sub-network is identified by the segments forming the main stream or by the list of segments forming a cycle. The stream defined by segment 27 contains the outlet and represents the whole network noted (27). Segments 25 and 26 form a reticulate network (25, 26) and, with sub-networks (22, 23) and (24), are located under the root. That means that they form three sub-networks at the level below the main river network. A drainage pattern can be defined for each sub-network.



ID	Order	ID	Order
1	1	15	1
2	1	16	2
3	1	17	2
4	2	18	2
5	2	19	2
6	1	20	2
7	1	21	2
8	1	22	3
9	1	23	3
10	1	24	3
11	1	25	4
12	1	26	4
13	1	27	5
14	1		

a. Initial river data. The marking beside the river is river segment ID. The table is river segment ID and its Horton-Strahler order.



b. Result of sub-networks hierarchy.

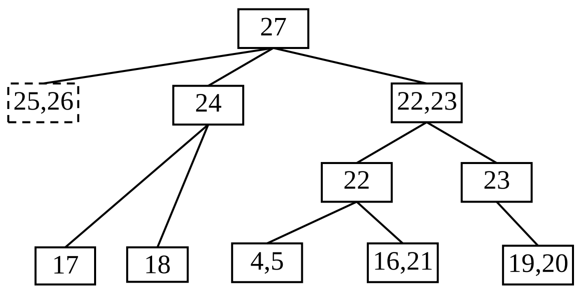
Figure 3.11 The drainage tree

3.5.2 Merging drainages along a river stream

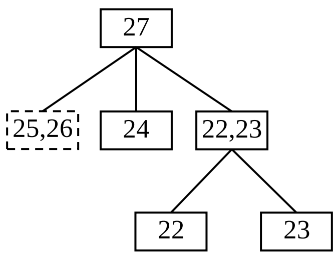
The process yields a drainage tree where all existing sub-networks are characterized. However, a river stream can go through different types of terrain where its tributaries follow different patterns. Therefore, a river stream can be split in sections forming different drainage patterns. The algorithm starts from the root of the drainage pattern tree and moves down to the leaves. For each river stream, if two adjacent sub-networks are of the same pattern, they can be merged into a larger drainage. Two drainages are adjacent if they connect on the same node of a river segment or on two adjacent nodes on the same side of the segment. The algorithm is illustrated in Appendix B.2. The function of ***Merge(st)*** is used for merging the new hierarchy of river network according to drainage patterns, which is also a recursive function.

Sub-networks can also be merged to remove information seen as redundant: if all the sub-networks sharing the same parent in the tree have the same drainage pattern as their parent, the sub-networks can be removed as their pattern information is already defined at the parent level. The ***Remove(st)*** function used for removing information is shown in Appendix B.3.

Taking the river network in Figure 3.11 as an example, sub-networks (24), (17) and (18) are supposed dendritic, sub-networks (4,5) and (16,21) are parallel, and sub-network (19,20) is a trellis (Figure 3.12a). Networks (17) and (18) have the same pattern as their parent (24) therefore they hold redundant information and can be removed. Networks (4, 5) and (16, 21) share the same pattern and are both connected to segment 22 while trellis (19, 20) is connected to segment 23. Therefore, the stream (22, 23) goes through two drainage systems: first a trellis and second a parallel drainage. Therefore, network (22, 23) can be split into one parallel network (22) and one trellis (23). The resulting drainage tree is shown in Figure 3.12b.



a. Drainage tree after splitting 22, 23 from stream (23, 22)



b. Drainage tree after removal of sub-networks with same pattern

Figure 3.12 Merged hierarchy of sub-network from Figure 3.11

3.6 Experiments and results

The method was implemented in C++ with the Boost Graph Library (BGL[1]) for cut-edge finding and indicator computation, and in MATLAB with Fuzzy Logic Toolbox[2] for drainage pattern recognition. Software Graphviz[3], which is a

---

[1] http://www.boost.org/doc/libs/1_48_0/libs/graph/doc/index.html
[2] http://www.mathworks.com/products/fuzzy-logic/
[3] http://www.graphviz.org/

package of open source tools initiated by AT&T Labs Research[4] for rendering graphs in DOT[5] language scripts, was applied to display the result of the drainage tree.

The experimental data set is the Russian river (Figure 3.13), California at the scale of 1:24,000 stored in a Shapefile from the Russian River Interactive Information System (RRIIS[6]). Original data set is composed of 5699 river segments. The bolder the line, the greater the Horton-Strahler order. The highest order in the network is equal to 6. In Figure 3.13, R_1 to R_9 are the regions of selected sub-networks for case studies.

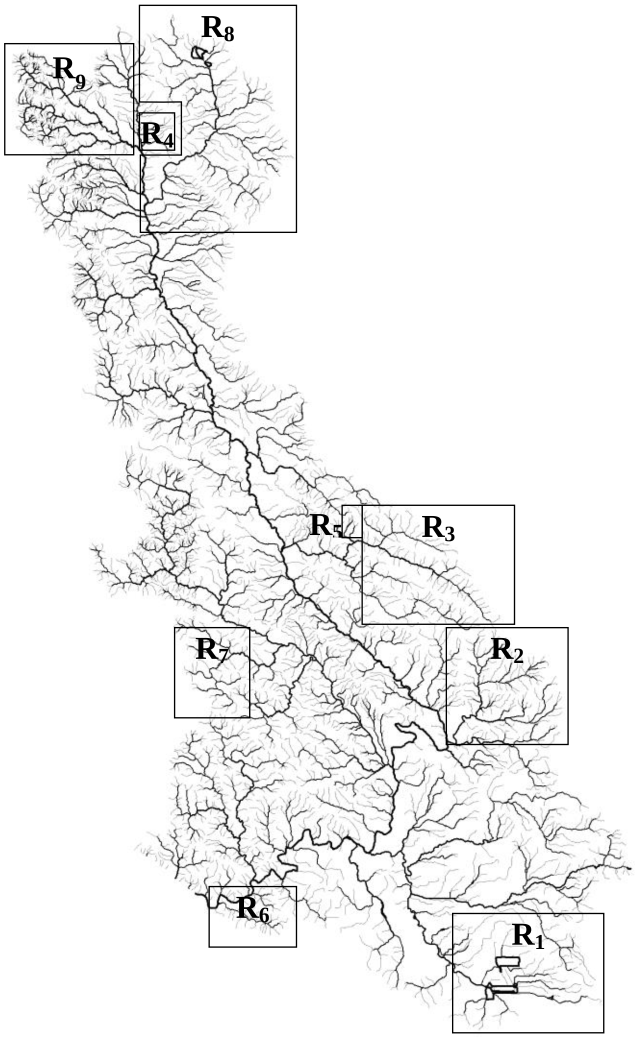
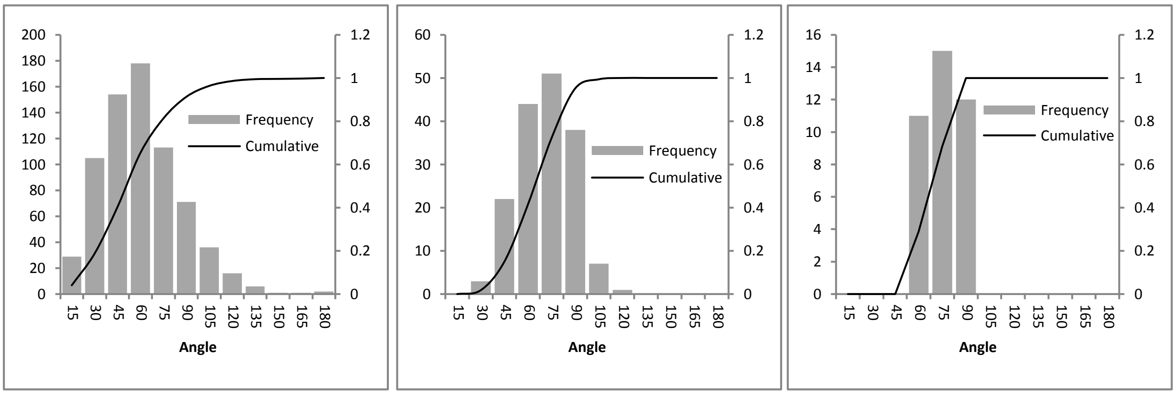


Figure 3.13 Russian river provided by RRIIS

[4] http://www.att.com/labs
[5] http://www.graphviz.org/content/dot-language
[6] http://www.rrwatershed.org/

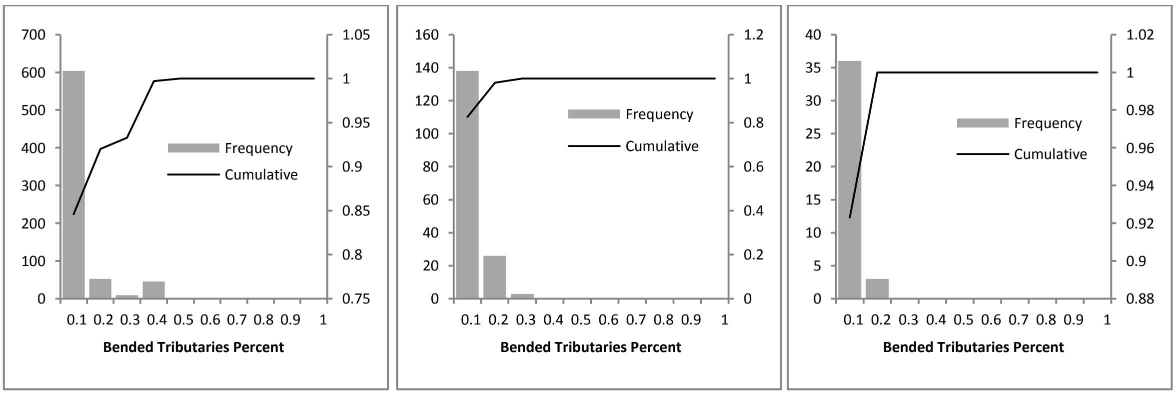
3.6.1 MF parameter settings for Russian river

Drainage pattern classification depends on the definition of MFs. Parameters to be set are the values of a and b for the Z and S curves and the value of σ and m for the Gaussian distribution. These values are mostly defined based on sample assessments and expert advice. As shown in Figure 3.14, the distribution of indicator values α, β, γ, δ among all networks is continuous. Therefore, setting a crisp threshold value is difficult as a small change in one value can have a significant impact on the classification. Working with fuzzy sets appears to be more appropriate as MFs can have very large supports or overlapping supports. For example an angle can be considered acute and right or very acute and acute at the same time so that classification is based on a compromise between different degrees of membership. Furthermore, reasoning on fuzzy sets has the benefit of being more robust as a small variation of the MFs would have a limited impact on the classification.



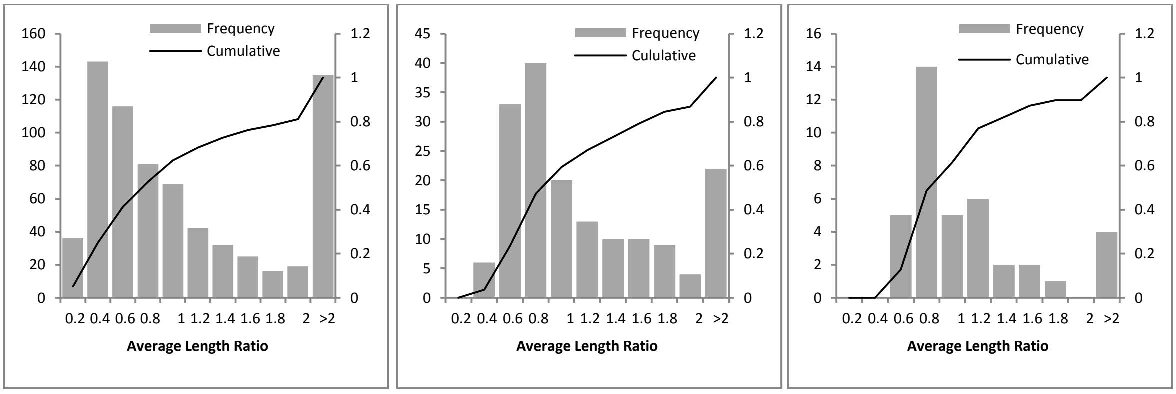
(a) Order 2 (b) Order 3 (c) Order 4

A. Frequency of average junction angle (α) in different levels



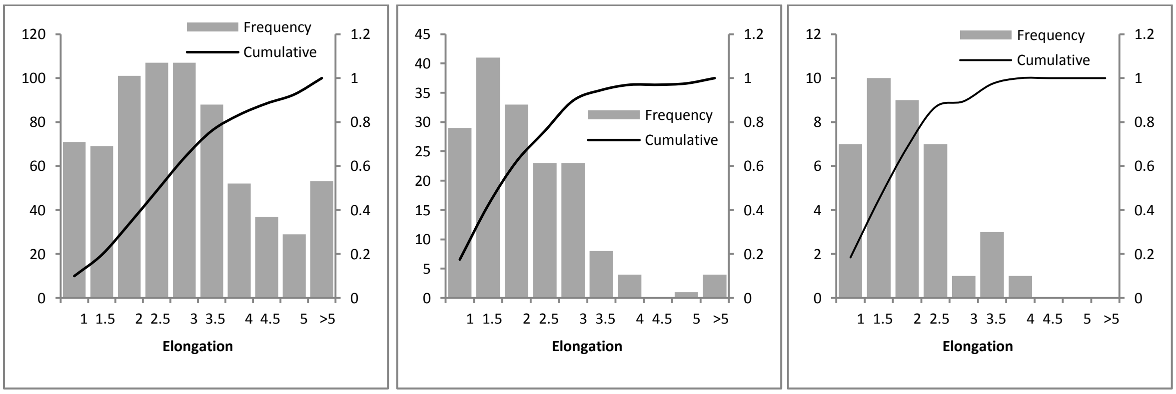
(a) Order 2 (b) Order 3 (c) Order 4

B. Frequency of bended tributaries percentage (β) in different levels



(a) Order 2 (b) Order 3 (c) Order 4

C. Frequency of average length ratio (γ) in different levels



(a) Order 2 (b) Order 3 (c) Order 4

D. Frequency of catchment elongation (δ) in different levels

Figure 3.14 Frequency of indicators in different levels

MFs used in this experiment are given in Table 3.3. For case I, the threshold values 30° and 60° are used to establish very acute MF. If the angle is smaller than 30°, it is definitely a very acute angle while if it is greater than 60°, it cannot be very acute. Between 30° and 60°, the greater the angle is, the smaller the membership. Similarly, angles are considered acute under 45°, and their degree of membership decreases when the angle increases. The standard deviation for right angles was set to 10°. The closer to 90°, the higher the membership value. MFs of bended tributaries and elongated catchment are both S curves. The closer to 1 β is, the more tributaries bend.

Table 3.3 Specific settings of MFs

Case	I	II	III
Very acute angle	*z(α;30°,60°)*	*z(α;35°,65°)*	*z(α;25°,55°)*
Acute angle	*z(α;45°,90°)*	*z(α;50°,95°)*	*z(α;40°,85°)*
Right angle	*g(α;10°,90°)*	*g(α;12.5°,90°)*	*g(α;7.5°,90°)*
Bended tributaries	*s(β;0,1)*	*s(β;0,0.9)*	*s(β;0.1,1)*
Long tributary	*s(γ;0,1)*	*s(γ;0,0.8)*	*s(γ;0.1,1.1)*
Short tributary	*z(γ;0,1)*	*z(γ;0.2,1.2)*	*z(γ;0,0.9)*
Broad catchment	*z(δ;1,3)*	*z(δ;1.25,3.25)*	*z(δ;0.8,2.8)*
Elongated catchment	*s(δ;1,3)*	*s(δ;0.75,2.75)*	*s(δ;1.2,3.2)*

Based on analyses of case studies, dendritic drainages had an elongation centered on 1 while trellis and parallel drainages had a much higher elongation (often greater than 2). Therefore, the elongated MF was set to *s(δ;1,3)* so that a network with an elongation up to 2 may still be considered as square. For the broad catchment, the MF is set opposite to elongated to *z(δ;1,3)*. During the tests, short tributaries appeared to be a less relevant indicator than the elongation and the angle to characterize the networks. According to the rules given in Table 3.3, the MF for short tributaries should have a large support and thus is set to *z(γ;0,1)*, and oppositely the MF for long tributaries is set to *s(γ;0,1)*.

In Table 3.3, two other cases are also presented. MFs in case II have a larger support than in case I, while in case III the support is smaller providing a stricter classification. As an example, an angle of 32° is definitely very acute in case II, but not in cases I and III. The degree of membership in case I would be higher than in case III though. These three cases were all tested in the experiment.

Sensitivity to parameter values was assessed by fixing all the parameters but one to values defined for case I. The free parameter was tested with values presented in cases II and III. Table 3.4 shows the results of sensitivity analysis and results are compared with case I classification. It appears that the model is mostly sensitive to α and δ. MFs of parameters α and δ for very acute angle and broad catchment have a strong influence on the classification results. On the opposite, the value of β has limited influence as few streams are bent.

Table 3.4 Number of drainage patterns with a parameter changing

	Test	Dendritic	Parallel	Trellis	Rectangular	Unrecognized	Changed
	Case I	405	339	130	18	28	-
α	VAA → $z(\alpha;35°,65°)$	382	370	122	18	28	31
	VAA → $z(\alpha;25°,55°)$	424	306	142	18	30	33
	AA → $z(\alpha;50°,95°)$	424	339	119	15	23	19
	AA → $z(\alpha;40°,85°)$	378	339	140	28	35	27
	RA → $g(\alpha;12.5°,90°)$	390	335	148	19	28	19
	RA → $g(\alpha;7.5°,90°)$	412	342	120	18	28	10
β	BT → $s(\beta;0,0.9)$	403	339	129	21	28	3
	BT → $s(\beta;0.1,1)$	407	339	131	9	34	9
γ	LT → $s(\gamma;0,0.8)$	401	343	130	18	28	4
	LT → $s(\gamma;0.1,1.1)$	409	335	130	18	28	4
	ST → $z(\gamma;0.2,1.2)$	401	339	136	17	27	6
	ST → $z(\gamma;0,0.9)$	406	341	123	19	31	7
δ	BC → $z(\delta;1.25,3.25)$	438	316	123	18	25	33
	BC → $z(\delta;0.8,2.8)$	364	361	143	20	32	41
	EC → $s(\delta;0.75,2.75)$	393	343	143	16	25	17
	EC → $s(\delta;1.2,3.2)$	420	331	116	21	32	22

(VAA, AA, RA, BT, LT, ST, BC and EC are short for very acute angle, acute angle, right angle, bended tributaries, long tributary, short tributary, broad catchment and elongated catchment respectively. "→" presents that the parameter apply another setting.)

3.6.2 Case studies in Russian river

3.6.2.1 Reticulate pattern in river networks

The cut-edge finding algorithm is used to identify all river segment parts of a reticulate pattern. Figure 3.15 shows a part of the network from region R_1 of Figure 3.13 with its reticulate patterns.

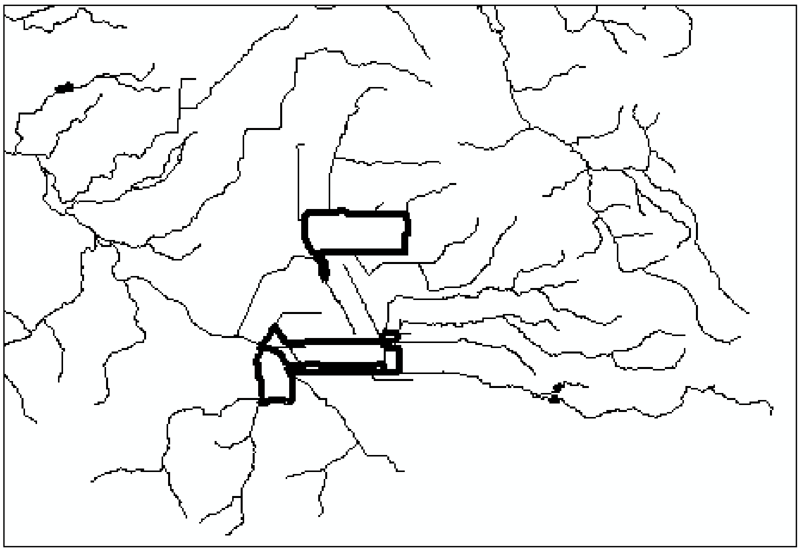
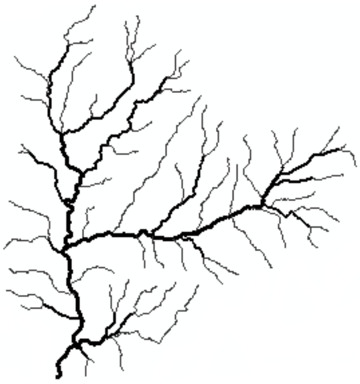


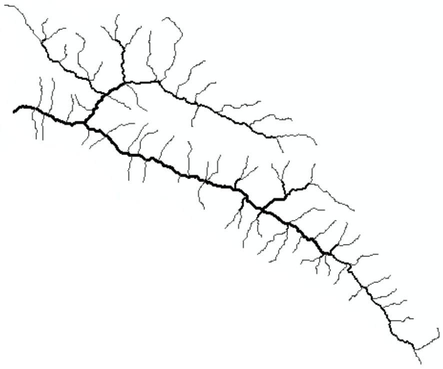
Figure 3.15 Reticulate networks in bold

3.6.2.2 Dendritic, parallel, trellis and rectangular pattern in river networks

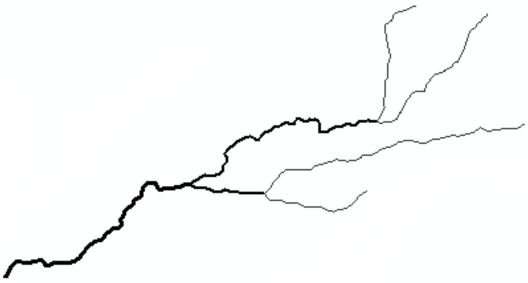
Selected sub-catchments are shown in Figure 3.16. Locations of sub-catchments in the whole river basin can be seen in Figure 3.13. Sub-catchments (a), (b), (c), (d), (e) and (f) have a highest Horton-Strahler order of 4, 4, 3, 3, 3 and 3 respectively.



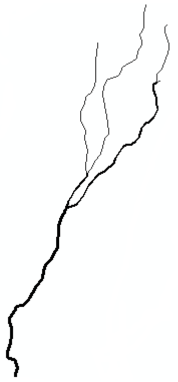
(a) Region R_2



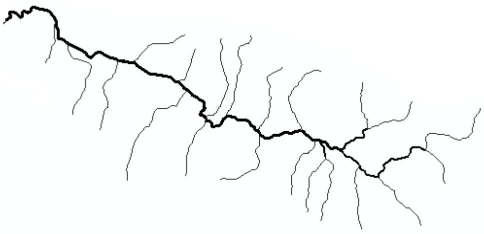
(b) Region R_3



(c) Region R_4



(d) Region R_5



(e) Region R_6

(f) Region R_7

Figure 3.16 Sub-catchments of Russian river basin, (a), (b), (c), (d), (e) and (f) corresponding to regions R_2, R_3, R_4, R_5, R_6 and R_7 in Figure 3.13 respectively

Results from several sub-catchments from the Russian river basin are provided for discussion with values of indicators as well as membership degrees for all four rules in Table 3.5.

Table 3.5 Information of sub-catchments in Figure 3.16

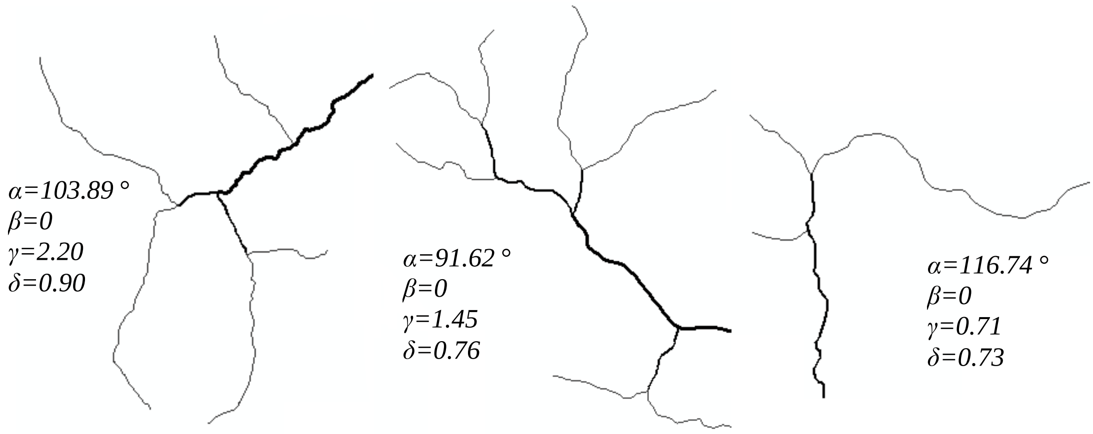
	Indicator values	Case	MF degree				Output
			rule 1 (p_1/p_2)*	rule 2 ($p_1/p_2/p_3/p_4$)*	rule 3 ($p_1/p_2/p_3/p_4$)*	rule 4 (p_1/p_2)*	
(a)	α=51.64 ° β=3.54% γ=0.74 δ=0.87	I	0.957 (0.957/1)	0 (0.155/0.998/ 0.865/0)	0 (0.001/0.998/ 0.135/0)	0.001 (0.001/0.003)	dendritic
		II	0.997 (0.997/1)	0.007 (0.397/0.997/ 0.989/0.007)	0.007 (0.009/0.997/ 0.423/0.007)	0.003 (0.009/0.003)	dendritic
		III	0.866 (0.866/0.998)	0 (0.025/1/ 0.063/0)	0 (0/1/ 0.063/0)	0 (0/0)	dendritic
(b)	α=81.14 ° β=1.49% γ=0.74 δ=3.17	I	0 (0.078/0)	0 (0/0.999/ 0.865/1)	0.135 (0.675/0.999/ 0.135/1)	0 (0.675/0)	trellis
		II	0.003 (0.189/0.003)	0 (0/0.999/ 0.989/1)	0.423 (0.778/0.999/ 0.423/1)	0.001 (0.778/0.001)	trellis
		III	0 (0.015/0)	0 (0/1/ 0.741/0.999)	0.063 (0.498/1/ 0.063/0.999)	0 (0.498/0)	trellis
(c)	α=21.52 ° β=0 γ=1.28 δ=3.53	I	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
		II	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
		III	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
(d)	α=22.88 ° β=0 γ=1.55 δ=5.13	I	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
		II	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
		III	0 (1/0)	1 (1/1/1/1)	0 (0/1/0/1)	0 (0/0)	parallel
(e)	α=85.43 ° β=9.76% γ=0.62 δ=2.87	I	0.008 (0.021/0.008)	0 (0/0.981/ 0.711/0.992)	0.289 (0.901/0.981/ 0.289/0.992)	0.019 (0.901/0.019)	trellis
		II	0.072 (0.091/0.072)	0 (0/0.977/ 0.898/1)	0.647 (0.935/0.977/ 0.647/1)	0.024 (0.935/0.024)	trellis
		III	0 (0/0)	0 (0/1/ 0.539/0.946)	0.194 (0.831/1/ 0.194/0.946)	0 (0.831/0)	trellis
(f)	α=94.93 ° β=4.08% γ=0.33 δ=0.87	I	0 (0/1)	0 (0/0.997/ 0.218/0)	0 (0.886/0.997/ 0.782/0)	0.003 (0.886/0.003)	rectangular
		II	0 (0/1)	0 (0/0.996/ 0.340/0.007)	0.007 (0.925/0.996/ 0.966/0.007)	0.004 (0.925/0.004)	trellis
		III	0 (0/0.998)	0 (0/1/ 0.106/0)	0 (0.806/1/ 0.731/0)	0 (0.806/0)	-

* The content in the parenthesis is the degree of all predicates of each rule.

Based on indicators, network (a) has acute junction angles and is broad. The MF value of the first rule is the highest in all cases (0.957/0.997/0.866), meaning that (a) is dendritic. From Table 3.5, (b) and (e) have a highest membership value for the third rule and other membership values are very small, so they are classified as trellis pattern. Networks (c) and (d) are definitely parallel, because they both have very acute junction angles (21.52° and 22.88°), elongated catchments (3.53 and 5.13 both bigger than 3) and long tributaries (1.28 and 1.55). Network (f) has average angle greater than 90°, so it is neither dendritic nor parallel. It is identified as rectangular in case I and trellis in case II, but cannot be recognized in case III. The membership value from rule 4, however, is only 0.003, which might be too small to consider (f) as rectangular. In case II, the membership value from rule 3 is 0.007, which is not large enough to consider (f) as trellis.

The quantitative indicators and fuzzy logic method can characterize the drainage patterns of the river network. From the experiment, it is verified that a small variation of MFs has a limited impact on the classification result, but the MFs with a more tolerant setting can support more ambiguous situations. In addition, the maximum method for defuzzifying the fuzzy outputs may show some limitations in some cases. For example, sub-network (f) in Figure 3.16 has been output as a rectangular pattern as the final result in case I, but the degree of membership of the result is only 0.003 which is too farfetched to classify (f) as rectangular. In such a case, a solution may be not to use the maximum method but to obtain the final result by giving a threshold of the degree of membership.

A few networks remained unclassified such as network (f) in case III of Figure 3.16, for which all membership degrees of all rules are 0. Some unclassified networks in case I are illustrated in Figure 3.17. Overall, unclassified networks are broad and have tributaries joining at obtuse angles. They also have few tributaries so that indicators computed as an average value from tributaries may not be objective.



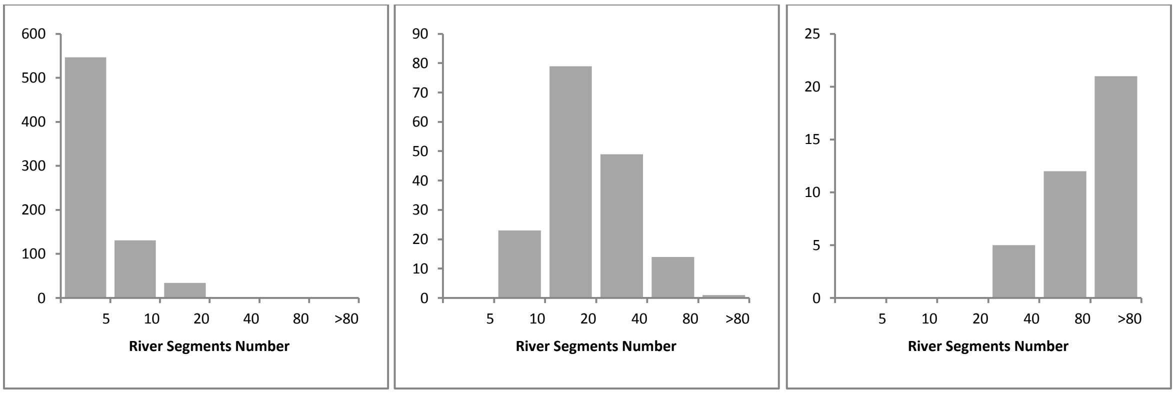
(a) (b)* (c)

Figure 3.17 Some unclassified sub-networks in case I (*dendritic in case II)

In Figure 3.17, sub-networks (a) and (b) have a junction angle (103.89°, 91.62°) close to a right angle, but catchments are not elongated (δ both smaller than 1). Moreover, neither of them have bended tributaries, and their tributaries are not short. So, (a) and (b) cannot be classified. The junction angle of (c) is bigger than 90°, so it cannot be dendritic and parallel. Also, (c) cannot be identified as trellis because its catchment is not elongated (δ = 0.73), neither can it be rectangular because there are no bended tributaries (β = 0). However, (b) can be classified as dendritic in case II due to the wide support on MF of acute angle, and the membership value of rule 1 is 0.011.

3.6.3 Drainage pattern recognition results and discussion

In the previous section, some sub-networks from the Russian river are selected to show the results of case study. In this section, the method is applied to the whole network which can be decomposed in different sub-networks at different orders. In the experiment, different catchment units lead to different results. According to the Horton-Strahler order of its main stream, a catchment unit can belong to different orders from 2 to 4. Order 1 catchments were not considered as they correspond to single stream networks. On one hand, the smaller the order, the more catchment units. On the other hand, low order networks have fewer tributaries which may make average values less significant. Figure 3.18 shows the frequency of river segment numbers in a sub-catchment at different orders.



(a) Order 2 (b) Order 3 (c) Order 4

Figure 3.18 Frequency of river segment numbers in a catchment at different orders

Frequency distributions vary for different stream orders. From Figure 3.18, most of the river sub-catchments in order 2 are composed of fewer than 5 river segments, and no sub-catchment in order 4 has fewer than 20 river segments. The number of river segments in a sub-catchment would influence the indicators such as average angle and catchment elongation. The percentage of bended tributaries indicator is not related with the river segments number because it depends more on the shape of each single river segment. Table 3.6 shows the number of drainages at each order for each pattern by the MFs settings in Table 3.3. Reticulate networks are identified in a preliminary step.

Table 3.6 Number of drainage patterns

	Case	Dendritic	Parallel	Trellis	Rectangular	Unrecognized	Total	Reticulate
Order 2	I	253	320	107	9	25	714	12
	II	253	329	111	5	16		
	III	247	300	114	10	43		
Order 3	I	119	19	18	8	3	167	
	II	119	21	19	7	1		
	III	114	18	19	4	12		
Order 4	I	33	0	5	1	0	39	
	II	31	1	6	1	0		
	III	32	0	5	0	2		

In Table 3.6, although cases I, II and III have different sets of values of MFs, there is little change in the number of drainages recognized for each pattern at a given order. This shows that the classification obtained with fuzzy logic is robust.

Case II recognizes the largest number of networks while case III recognizes the smallest number. The result is expected as MFs in case II have larger support. Therefore, changing the threshold values of MFs can help to avoid unclassified networks to some extent. The proportion of dendritic drainages increases with the order while the proportion of parallel drainages decreases. This variation may be linked to some geomorphological properties of the terrain but may also be due to the fact that MFs "Acute" and "Very Acute" partially overlap and for some networks, both MFs are equal to 1. In that case, the catchment elongation becomes the main indicator. A river network at a higher order tends to form a more complex network with a larger number of rivers spreading in various directions and eventually to exhibit a broader catchment, hence a larger proportion of dendritic patterns at order 4. Indeed, in some cases, networks at order 4 can represent very large systems where a main stream goes through different types of terrain and follows different patterns. The number of trellis remains stable in proportion because drainages identified as trellis tend to form less complex networks with a smaller number of sub-networks and so remain elongated. In general, all three MF settings provide reliable results. Drainages that change from one pattern to another have a low membership value in any case. They mainly belong to order 2 where the number of streams can be small, making the process more sensitive to indicator variations. In some cases, drainages that remain unclassified do not belong to any pattern because they do not satisfy any criterion.

Statistics of average value of indicators for classified drainages is shown in Table 3.7. It can be noted that the junction angle of parallel drainages is close to 30°, indicating that most of the parallel drainages had a junction angle far below the limit. Junction angle of trellis is around 85° meaning that many streams do not join at right angles. Therefore, the MF needs to be set with a rather large support. Trellis has a significantly smaller value γ than for other patterns for which it is twice as long as the main stream. This mostly relates to the way main streams are defined. Finally, it can be noted that the catchment elongation of parallel and trellis patterns is much larger than dendritic, which makes sense, but also that parallel drainages are on average more elongated than trellis.

Table 3.7 Statistics for each classified drainage

	Case	Average(α)	Average(β)	Average(γ)	Average(δ)
Dendritic	I	58.97 °	4.27%	2.09	1.57
	II	59.50 °	4.26%	2.08	1.56
	III	57.90 °	4.31%	1.99	1.57
Parallel	I	34.82 °	3.48%	2.49	3.84
	II	35.45 °	3.48%	2.49	3.80
	III	33.52 °	3.62%	2.61	3.85
Trellis	I	85.50 °	3.29%	0.43	2.65
	II	87.09 °	3.74%	0.44	2.53
	III	81.47 °	3.14%	0.41	2.83
Rectangular	I	94.01 °	18.81%	2.64	1.76
	II	95.46 °	17.52%	3.31	1.81
	III	97.30 °	26.07%	1.91	1.24

3.6.4 Drainage pattern hierarchy

The whole Russian river is too large to show details of the drainage tree. Here, a sub-network of the Russian river is selected to illustrate the results in the process of the drainage tree construction. The selected river network corresponds to region R_8 in Figure 3.13, which is illustrated in Figure 3.19.

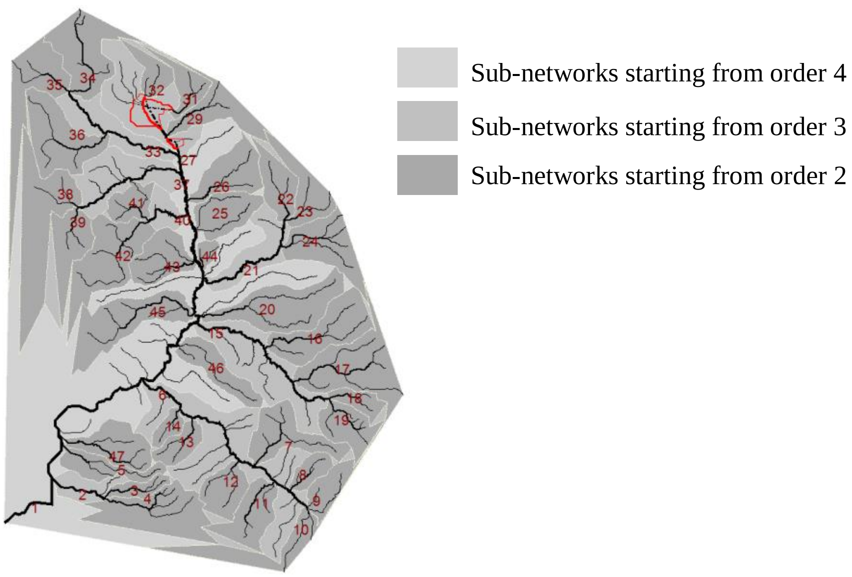


Figure 3.19 Selected river network from the Russian river

In Figure 3.19, there are two reticulate parts in the river network, shown as red pieces. Inside the river network, each sub-network is located in a drainage basin extracted by the method of the hierarchical watershed partitioning (Ai et al., 2006). Each sub-network is assigned with a number. All sub-networks are identified and formed as a drainage tree for the selected experimental data, which is shown in Figure 3.20.

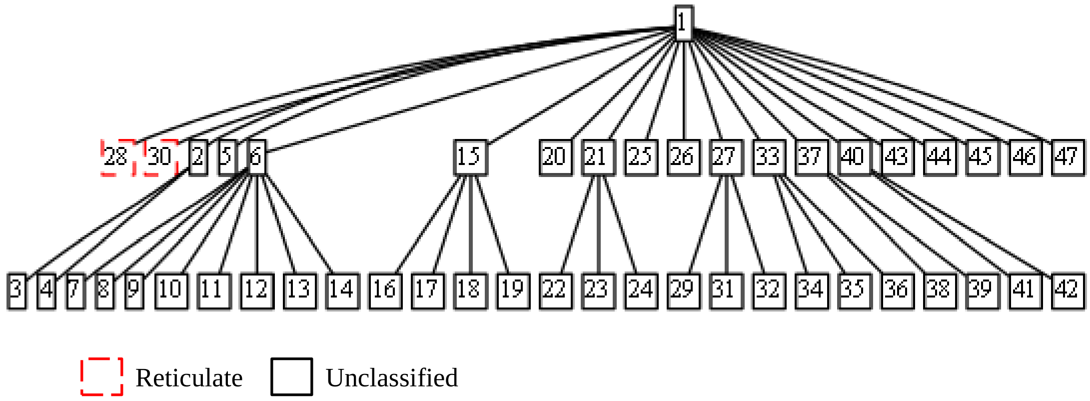


Figure 3.20 Drainage tree of all sub-networks

In the hierarchy graph, the root node is arranged at the top, and leaves are at the bottom. A node, which indicates a sub-network, is represented as a number that corresponds to the numbers of the sub-networks in Figure 3.19. The sub-networks classified as reticulate are located in the nodes with red dashed box. These features of the hierarchy are followed not only in this sub-network hierarchy but also in the drainage tree in the following sections.

In the sub-network hierarchy of the selected river network (Figure 3.20), there are 46 sub-networks. The whole river network is noted as (1), and two reticulate networks are (28) and (30) respectively. The DOT script result of the hierarchy in Figure 3.20 is detailed in Appendix C.1. For characterizing drainage patterns in the sub-tree, the MFs adopt the setting of case I in Table 3.3, and the result is shown in Figure 3.21. In the result, red dashed boxes also indicate reticulate networks. In addition, boxes filled with sky blue, orange, yellow and tomato colors represent dendritic, parallel, trellis and rectangular networks respectively.

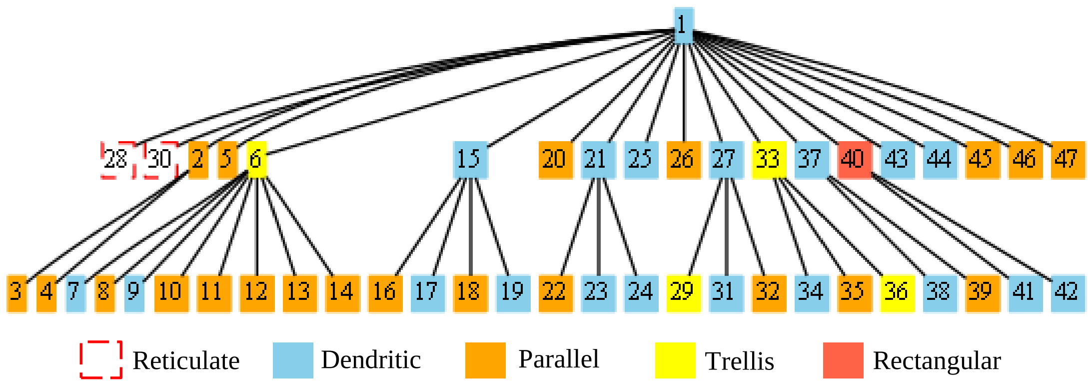


Figure 3.21 Drainage tree after pattern classification (Appendix C.2)

In Figure 3.21, the selected river network is recognized as dendritic. Inside the network, 18 sub-networks are classified as dendritic and 21 sub-networks are parallel. Besides, there are 4 trellis networks and 1 rectangular network. Figure 3.22 shows the drainage pattern classifications in the selected region divided by watershed and rendered with color.

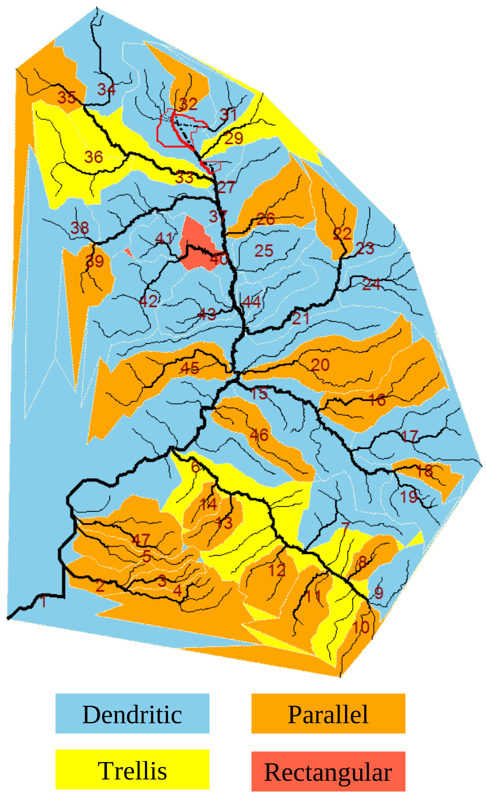


Figure 3.22 Sub-catchments with drainage patterns

From Figure 3.22, we can see that some networks are adjacent. For example, sub-networks (2), (5) and (47) should be merged due to their locations on the right side of the main stream. So, some nodes of the drainage tree should be split and merged. After this process, the drainage tree is illustrated in Figure 3.23. The DOT script is given in Appendix C.3.

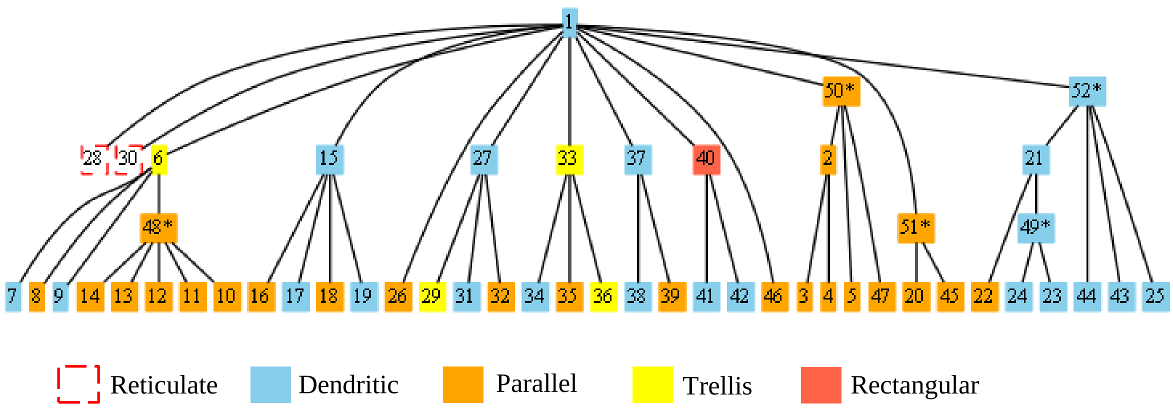


Figure 3.23 Drainage tree after merging and splitting

In Figure 3.23, nodes with an asterisk (*) are new networks split from the main stream. The sub-networks (2), (5) and (47) have been merged, because all of them are parallel and located near each other. In network (6), along its main stream from outlet to source, (10), (11), (12), (13) and (14) are all on the right, so the main stream should be split for their merging. Although (7) and (9) are also on the same side, they cannot be merged because of the interruption of network (8). Networks (20) and (45) are placed on opposite sides of the main stream, but they also are merged because they both connect to the same river segment.

The last process is to remove the redundant information in the drainage tree. Inside the network (49*), two sub-networks (23) and (24) are removed because both of them are identified as dendritic as their parent (49*). Networks (10), (11), (12), (13) and (14) are merged and noted as network (48*), under which all sub-networks can be removed. Similarly, the sub-networks under networks (50*) and (51*) also should be removed. Although network (52*) and its direct sub-networks are identified as dendritic, it cannot be simplified. Because there is a sub-network, (22), under (21) which is parallel. The final result of the drainage tree is provided in Figure 3.24. The detailed DOT script is shown in Appendix C.4.

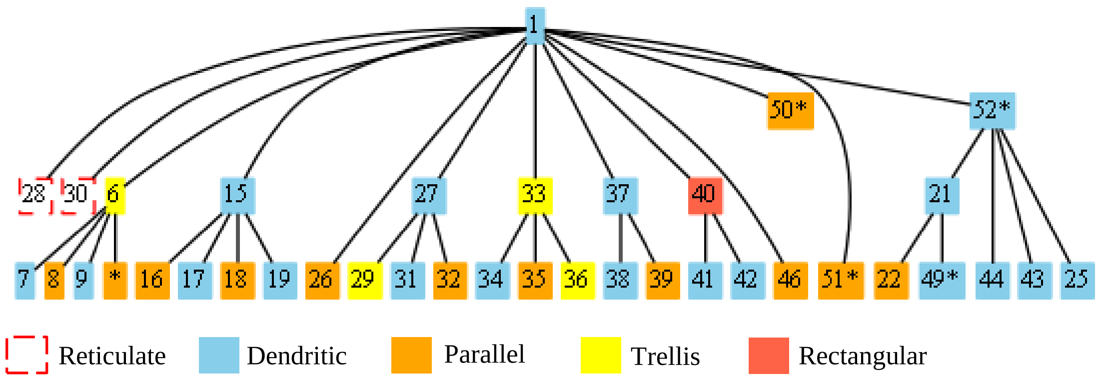


Figure 3.24 Final result of drainage tree

The selected river network is a typical dendritic drainage, where most of the tributaries flow into a larger one with an angle less than 90 degrees and the catchment is broad. The river network is in the upper course of the Russian river, which is a headwater region that collects and funnels water to the main stream. From the drainage tree in Figure 3.24, most of the sub-networks are classified as dendritic and parallel patterns. It is reasonable because the dendritic river network has many contributing streams that are used for collecting water. Parallel networks are formed where there is a pronounced slope, and they can be found in the upper course. As the upper course is steep, V-shaped valleys are formed by the prevailing downward erosion, and it is one of the landform with pronounced slopes.

However, there are still some other patterns such as reticulate, trellis and rectangular patterns. In general, in the upper course, these patterns do not appear except for human intervention. The area is located in Figure 3.25, the Potter Valley, a census-designated place in Mendocino County, California. Man-made irrigation canals or ditches destroy the nature of the river network at the reticulate region to a certain extent.

Figure 3.25 Image of the selected experimental area (from Google Map)

3.6.5 Multiple representation of drainage trees at different levels

A drainage tree is used to organize a river network according to the drainage patterns. In a drainage tree, the sub-networks with the same drainage pattern have been clustered by the adjacent locations. Although some sub-networks in a drainage tree have been merged, in a large river network, the drainage tree might still be complex. For example, the drainage tree in Figure 3.24 looks complicated, because at the bottom, the sub-networks start from order 2. As discussed in the previous chapter, many catchments in order 2 have less than 5 tributaries, and usually only 3 tributaries. In a large river network, the sub-networks are not necessarily divided too small. In this experiment, an application of drainage trees to the problem of multi-levels representation based on the order is given. A sub-network is selected from the Russian river for the case study. Figure 3.26 shows the selected river network for this experiment is in the region of R_9 in the Russian river. The highest order of the network is 6.

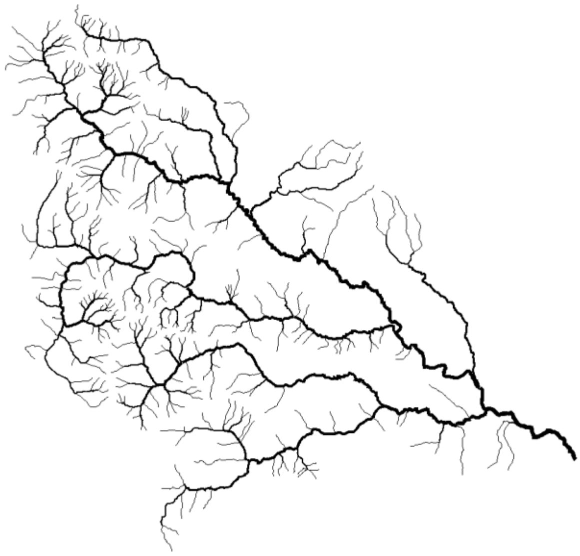


Figure 3.26 Sub-network corresponding to the region R_9 in Figure 3.13

The drainage tree of the selected river network is illustrated in Figure 3.27.

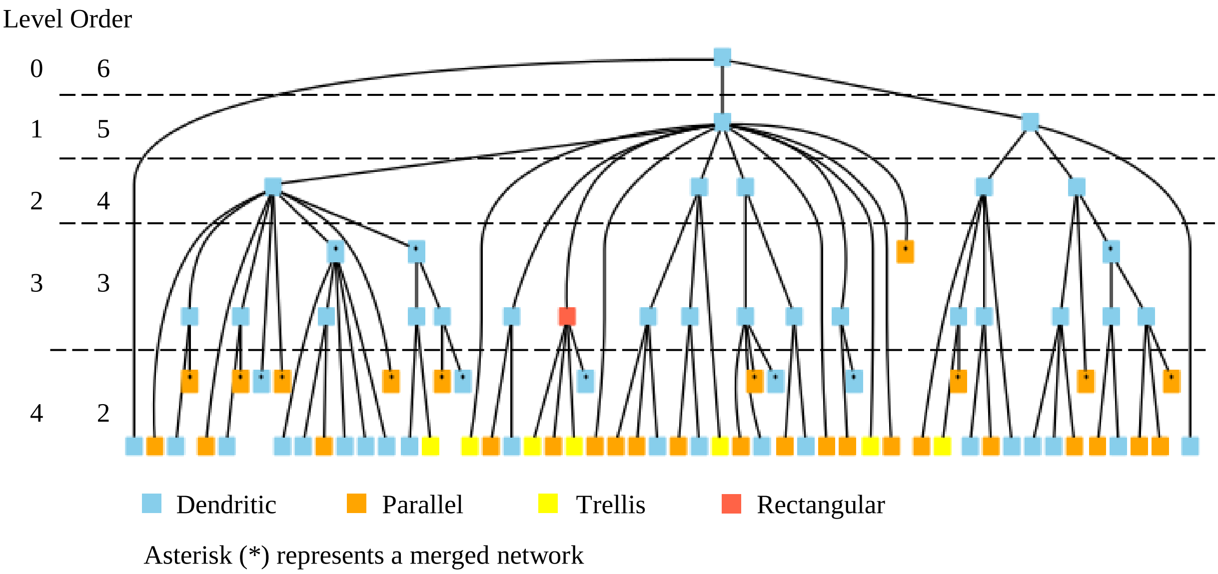


Figure 3.27 Result of drainage tree (Appendix B.5)

In the drainage tree, 5 levels are generated from order 6 to 2. There are only 2 sub-networks in level 1 less than last case but more sub-networks in level 2 to 4 especially in level 4. The drainage tree is more complicated. This network is identified as dendritic, and from level 1 to 3, most networks are classified as dendritic too. This is a typical dendritic network. However, in level 4, sub-networks beginning at order 2 are classified as various patterns, and parallel networks are in majority. There is only one sub-network which is rectangular in level 3, but not because the landform has undergone faulting. It is a coincidence that the network has some similar features of the rectangular pattern such as bended tributaries and the average angle close to 90 degree. Here, the rectangular sub-network is regarded as an exception.

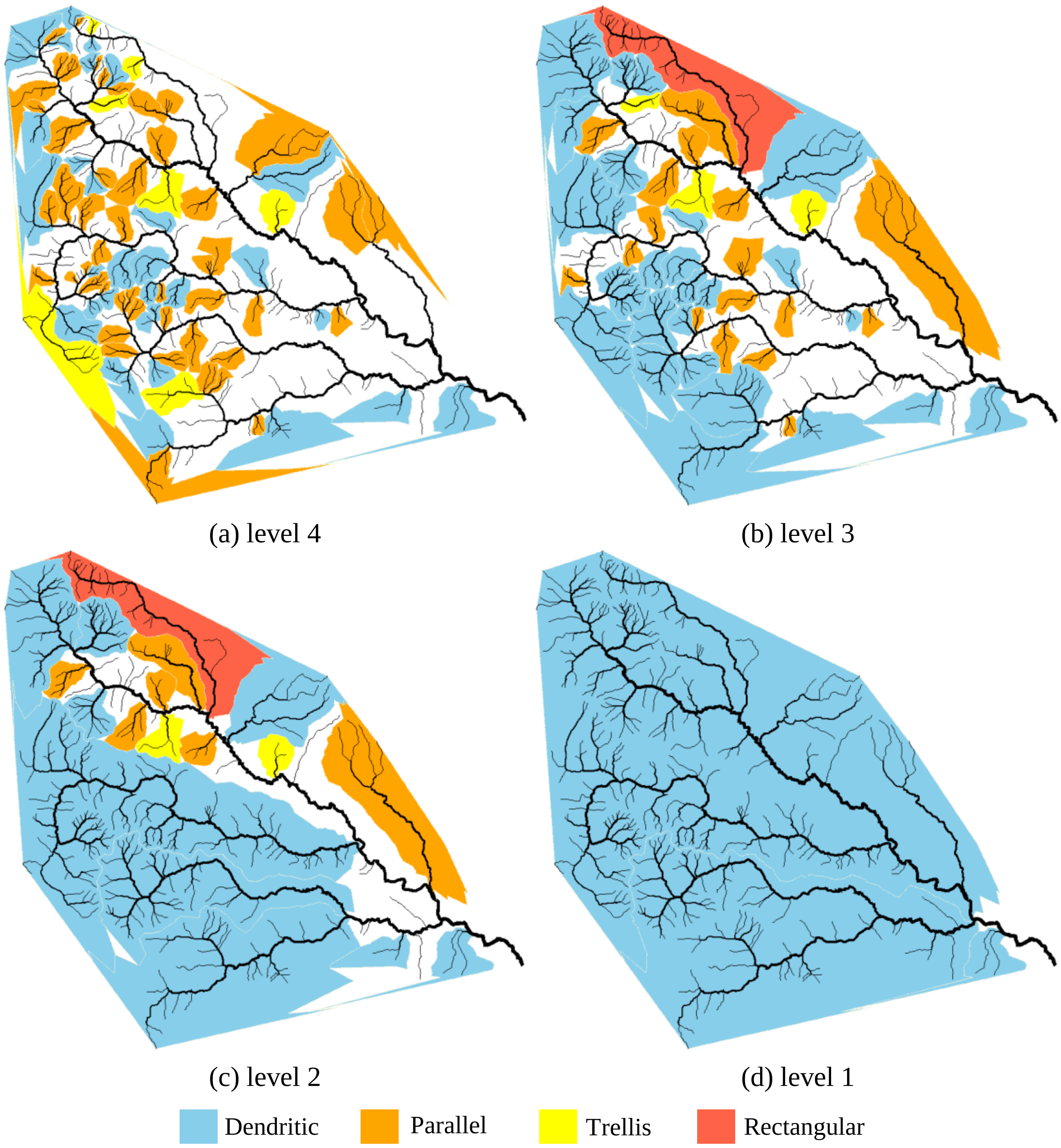


Figure 3.28 The selected network and representations at multiple levels for case 2

The multi-representations of the drainage tree at different levels based on the order are shown in Figure 3.28. In level 3 and 4, the results are similar with the last case. Adjacent sub-networks are usually classified as the same pattern. Inside dendritic networks in level 3, many sub-networks are identified as the parallel pattern. This is because sub-networks starting from order 2 do not have enough tributaries. The fewer tributaries, the easier to form a long and narrow drainage basin. This is the key factor to distinguish a dendritic network from a parallel one, especially if both of them have an acute average angle. In level 1, all sub-networks are classified as dendritic although there are many parallel sub-networks at the leaves of the drainage tree.

From the case study of multi-representations of the drainage tree at different levels, several conclusions can be given as follows: (1) Most river networks starting from order 3 and above are classified as dendritic; parallel sub-networks usually appear at the lower level especially in order 2. If river networks are extracted starting from 3 or above, this matches closely with the result that about 88% of river networks are dendritic (Serres & Roy, 1990). (2) In general, adjacent river networks are classified in the same pattern due to the similar landform, and river networks sharing a drainage divide are also classified as the same pattern. (3) The reticulate and rectangular river networks rarely appear in the upper river course unless there are particular reasons. For reticulate pattern, the man-made channels can lead to the crossing of rivers.

3.7 Summary

The drainage pattern is an important geographic factor for a river basin. This study proposed a method for automatic recognition of drainage patterns in river networks. The method recognizes drainage patterns from a river network defined by a directed graph. The terrain model is not required for the classification. Five types of pattern are classified: dendritic, parallel, trellis, rectangular and reticulate patterns. The method is based on geometric indicators, such as the junction angle, sinuosity, and catchment elongation, to classify the patterns automatically. Different patterns in a river network were identified separately and correspond to more or less complex networks with different Horton-Strahler orders, and were organized into a hierarchical structure representing levels of description of the drainage pattern. The method was finally applied on a case study, the Russian river network, and the resulting classification was discussed.

The advantage of this work is that proposed geometric indicators are easy to obtain and calculate. They can easily be implemented in a GIS and applied to a river network defined in a Shapefile or extracted from DEMs. However, in this last case, the quality of the classification may depend on the quality of the extracted network from the DEM (Grimaldi et al., 2007; Nardi et al., 2008).

As rules defining each pattern are vague and depend on a combination of indicators, classification made use of fuzzy logic to improve robustness of the result. Due to the tree-like characteristic of a river network, the hierarchical structure for drainage patterns is built based on a recursive method which can be stated shortly and clearly and implemented as shown in experiments. Such classification and organization can be useful for terrain analysis as it can help provide a qualitative description of the terrain or for generalization as river selection can be adapted to the type of network.

Validation of the results is based on assessments done on case studies. Some networks still remained unclassified either because they could belong to several different patterns, or to none. They are usually networks at the low order where there are not enough tributaries. Classification depends on the MF definitions. Different definitions were tested and yielded consistent results. It appears that among the different indicators, the junction angle and the elongation are the most significant: parallel drainages are mainly characterized by very acute junction angles and elongated catchment, trellis are also elongated but with orthogonal junctions. Some networks were not classified because they do not belong to any of the defined patterns. However, distinguishing drainages which are misclassified from those which cannot be classified may require assistance from hydrologic experts to provide a finer tuning of membership function supports.

In the experimentation, it appeared that the proportion of drainages in each pattern varied with the order. One explanation may be that drainages at the higher order tend to be broader and therefore more likely classified as dendritic. Further investigation is required in this direction, as large drainages may also be formed by clusters of drainages of different types as rivers go through different types of terrain. This evolution can be tracked by clustering adjacent drainages showing similar patterns into larger drainages that do not necessarily form a hierarchical structure as in this work.

The influence of scale shall also be studied. Results were discussed on a large scale model. However, the scale may affect the number of tributaries represented in the network and the computation of indicators. As the drainage system is often extracted from the terrain model, the accuracy of drainage pattern classification at different orders may be related to the resolution of the terrain model.

In terms of future research, in the short term, the first aspect for further work is the addition of other parameters for further pattern descriptions. On top of geometric indicators computed in a single network, other topologic indicators expressing relationships between networks can be considered. This would allow the study of the structure of the river network according to stream order and location inside the network. Drainage patterns such as radial and centripetal patterns have not been addressed in this work. Their identification requires the characterization of spatial relationships between networks rather than geometric indicators.

In the longer term, the drainage pattern can be considered for applications in terrain analysis and cartography. In cartography, drainage patterns provide information about the network structure and can be used in river tributary selection for map generalization (Touya, 2007). It may also be used in map updating to check the existence of inconsistencies between terrain elements and the streams and to correct the conflicts (Chen et al., 2007). The drainage pattern can be used to analyze and correlate the drainage patterns with the catchment areas extracted from a terrain model. As the drainage pattern is related to the morphology of a terrain, it can be used to enrich the terrain model and characterize morphologic features. Other physiologic and morphologic information may be overlaid on the river network to perform statistical analysis and improve the classification by adding other factors.

Chapter 4 Evaluation of generalization methods in preserving the drainage pattern

4.1 Introduction

In the previous chapter, the drainage patterns are recognized automatically for a river network in a fuzzy logic process. The drainage pattern of a river network can be applied to assess a generalized river network by checking whether its pattern has changed or not. In the drainage pattern recognition, for a river network, all rules established for all drainage patterns are estimated, and the pattern is determined by the highest membership value of all rules. For quality assessment, the membership values of the original and generalized river network are compared. As reticulate pattern is identified by bridge finding algorithm in graph theory, it is not considered yet here. Four drainage patterns are addressed in this study: dendritic, parallel, trellis and rectangular patterns.

There are two steps in river network generalization: selective omission and selected tributaries simplification (Li, 2007). In general, "*feature selection is normally the first step to any generalization project, independent of the generalization model being employed*" (Wilmer & Brewer, 2010). The simplification step has limited influence to the drainage pattern of a generalized river network. This research focuses on selective omission for river network generalization. In this chapter, in Section 4.2, the tributary selection problem is presented first, and several existing methods are introduced. The evaluation method is proposed in Section 4.3. Section 4.4 provides a design of experiments. In Section 4.5, case studies are performed on the Russian river, and in Section

4.6, the evaluation method is applied to a large amount of networks in the Russian river basin. The conclusion is given in Section 4.7.

4.2 Tributary selection techniques

Two questions are raised for selective omission of tributaries in a river network, they are:

(1) How many river tributaries are selected?
(2) Which river tributaries are selected?

The first question is answered by applying the "*Radical Law*" to determine how many river tributaries are selected in a specific map scale. For the second question, several tributary selection methods are introduced including selection by order and length and by watershed partitioning.

4.2.1 Tributary selection amount

The problem addressed in this section is to decide how many river tributaries should be removed (or retained) from a large map scale to a smaller one. In map generalization, a classical principle of selection (Topfer & Pillewizer, 1966), which is the so-called "*Radical Law*", was discovered by F. Topfer in 1961. The method is given as follows:

$$n_f = n_a \sqrt{\frac{M_a}{M_f}}, \tag{4.1}$$

where n_f is the number of objects shown at the smaller scale M_f, and n_a is the number of objects shown at the larger scale M_a.

This method is a basic principle, however, it may be not applicable everywhere. In Topfer & Pillewizer's (1966) work, as equation (4.1) is not so useful to small scale maps, a modified equation is also provided as

$$n_f = n_a C_b C_z \sqrt{\frac{M_a}{M_f}}, \tag{4.2}$$

where C_b is the "*Constant of Symbolic Exaggeration*" and C_z is the "*Constant of Symbolic Form*".

For a specific situation, new factors should be taken into consideration in hydrologic map generalization. In Wilmer & Brewer's (2010) work, a factor called "*Constant of Flowlines*" (C_f) is added to the basic equation. The modified equation is

$$n_f = n_a C_f \sqrt{\frac{M_a}{M_f}} \quad (4.3)$$

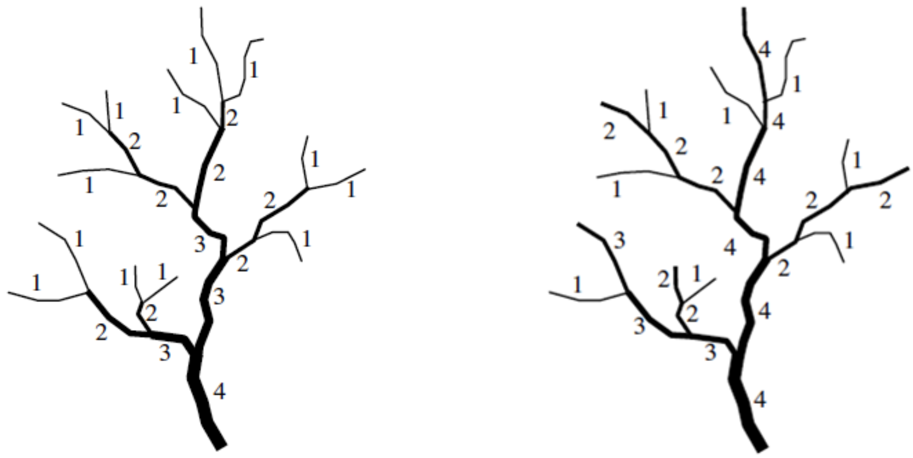
To apply the "*Radical Law*" in NHD, the constant C_f in the equation has three possible values: 1, 1.7 and 0.6. The value 1 is for large scale (24K) to medium scale (100K), 1.7 is for local scale (5K) to other scale, and 0.6 for comparisons to small scale (2M). The constant of 1 are used in the experiment.

4.2.2 Tributary selection modeling

Tributary selection methods are reviewed in Section 2.4.1. In this section, some of them are applied to obtain generalized river networks, which are used to be assessed by the evaluation method. The selection methods based on stroke and watershed partitioning are introduced in detail.

4.2.2.1 By stroke and length

Thomson & Brooks (2000) proposed a "*stroke*" concept and applied it to generalization and analysis for geographic networks such as road and river networks. In his work, for river networks, the Horton stream ordering after upstream routine is used to build the strokes of a river network. The Horton-Strahler order scheme is first performed, and then an upstream routine is applied to determine the main stream. Here, in a river network, the main stream is referred to as a "*stroke*". Figure 4.1 shows the strokes of a river network.



A. Horton-Strahler order scheme B. After upstream routine ("*stroke*")

Figure 4.1 "*Strokes*" of a river network (from Li, 2007)

Tributary selection based on order can be done in four possible ways (Rusak Mazur & Castner, 1990) as listed in Figure 2.5 (p. 17). The easy way is to eliminate all low order tributaries and preserve high order tributaries in the first place (Figure 2.5D). The shortage is that all tributaries in an order will be removed in a step. Sometimes, in a specific scale, some tributaries should be preserved in an order. So the length is a factor taken into consideration.

Similarly, taking a stroke as an entity, there are two steps in the generalization process: (1) remove the low order stroke first; (2) remove the shorter strokes if they are in the same order.

4.2.2.2 By watershed partitioning

Ai et al. (2006) proposed a method by constructing a hierarchy of different level watersheds. It focuses on the channel importance during the river network generalization replacing several geometric parameters of river feature by the watershed area. The watershed area is not obtained from the DEM but constructed on spatial competition by triangulations of the network. Obviously, tributaries should not be crossed over in the construction; they are constrained edges in the TIN. The selection method is to eliminate tributaries according to the catchment area. The tributary with smaller catchment area will be removed first. An example of the hierarchical watershed partitioning is shown in Figure 4.2.

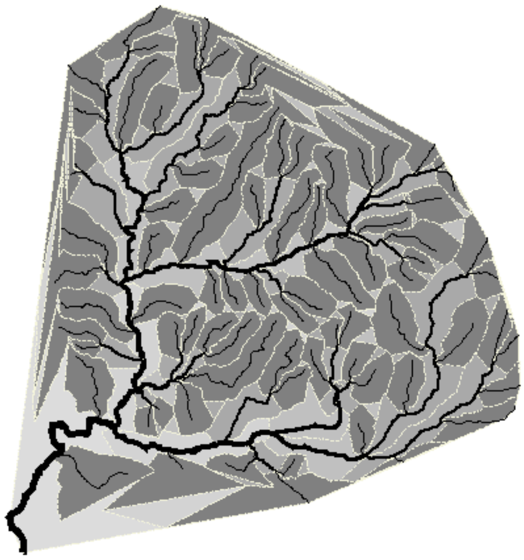


Figure 4.2 Hierarchical partitioning of river catchments

Consequently, the catchment area is the area of the watershed polygon. For a simple polygon with n vertices (x_i, y_i) ($1 \le i \le n$), the first and last vertices are the same, i.e. $x_n = x_1$, and $y_n = y_1$. The area is given by the Surveyor's formula (Braden, 1986):

$$A = \frac{1}{2}\left|\sum_{i=1}^{n-1}\left(x_i y_{i+1} - x_{i+1} y_i\right)\right|, \tag{4.4}$$

where A is the area of the polygon. If the vertices are stored sequentially in the counterclockwise direction, the absolute value sign in the formula can be omitted.

4.3 Evaluation method for each drainage pattern

The indicators used for drainage pattern recognition are introduced in Section 3.4.1; they are average junction angle (α), bended tributaries percentage (β), average length ratio (γ) and catchment elongation (δ). These indicators are also applied in the evaluation. The degree of a rule is used to assess the generalized river network quantitatively. In Section 3.4.2, the common MFs are introduced. Here, the evaluation methods for each pattern are described mathematically.

4.3.1 Dendritic pattern

In the fuzzy logic process, the rule defined for the dendritic pattern is

IF (α IS acute) AND (δ IS broad) THEN pattern IS dendritic.

Therefore, the degree of a dendritic network can be calculated as "(α IS acute) AND (δ IS broad)", which can be represented in the following formula.

$$f(\alpha,\delta) = \min(z(\alpha;a,b), z(\delta;a',b')), \tag{4.5}$$

where α and δ are inputs, $z(\alpha;a,b)$ and $z(\delta;a',b')$ are defined MFs for an acute angle and a broad catchment respectively.

4.3.2 Parallel pattern

The rule for parallel pattern is given by

IF (α IS very acute) AND NOT (β IS bended) AND (γ IS long) AND (δ IS elongated) THEN pattern IS parallel.

The degree of a parallel network is given as

$$f(\alpha,\beta,\gamma,\delta) = \min\left(z(\alpha;a,b), 1-s(\beta;a',b'), s(\gamma;a'',b''), s(\delta;a''',b''')\right), \tag{4.6}$$

where α, β, γ and δ are inputs, $z(\alpha;a,b)$, $s(\beta;a',b')$, $s(\gamma;a'',b'')$, and $s(\delta;a''',b''')$ are MFs for a very acute angle, bended tributaries, a long tributary and an elongated catchment respectively.

4.3.3 Trellis pattern

For the trellis pattern, the rule is defined as

IF (α IS right) AND NOT (β IS bended) AND (γ IS short) AND (δ IS elongated) THEN pattern IS trellis.

Then, the degree of a trellis network can be calculated as

$$f(\alpha,\beta,\gamma,\delta) = \min\left(g(\alpha;a,b), 1-s(\beta;a',b'), z(\gamma;a'',b''), s(\delta;a''',b''')\right), \tag{4.7}$$

where α, β, γ and δ are inputs, $g(\alpha;a,b)$, $s(\beta;a',b')$, $z(\gamma;a'',b'')$, and $s(\delta;a''',b''')$ are MFs for a right angle, bended tributaries, a short tributary and an elongated catchment respectively.

4.3.4 Rectangular pattern

The rule for the rectangular pattern is set as

IF (α IS right) AND (β IS bended) THEN pattern IS rectangular.

Therefore, the degree of a rectangular network is formulated as

$$f(\alpha,\beta) = \min(g(\alpha;a,b), s(\beta;a',b')) , \tag{4.8}$$

where α and β are inputs, $g(\alpha;a,b)$ and $s(\beta;a',b')$ are MFs for a right angle, bended tributaries respectively.

4.4 Experiment design

4.4.1 Testing data

Russian river datasets are tested in the experiment. Two different scales are used: 1:24,000-scale (1:24K) and 1:100,000-scale (1:100K). The large scale data is the one tested in Chapter 3. Small scale data is provided by the National Hydrography Dataset (NHD[7]) of the USA. From the history of the establishment of the NHD, the medium resolution data is built first, and then a conflation tool is used to help generate the 1:24K hydrological data from the medium resolution data. The medium resolution in the NHD data is at 1:100K. Therefore, the 1:100K data is not generalized from 1:24K, it is built manually. The Horton-Strahler order scheme was then computed. The testing data are illustrated in Figure 4.3.

[7] http://nhd.usgs.gov/data.html

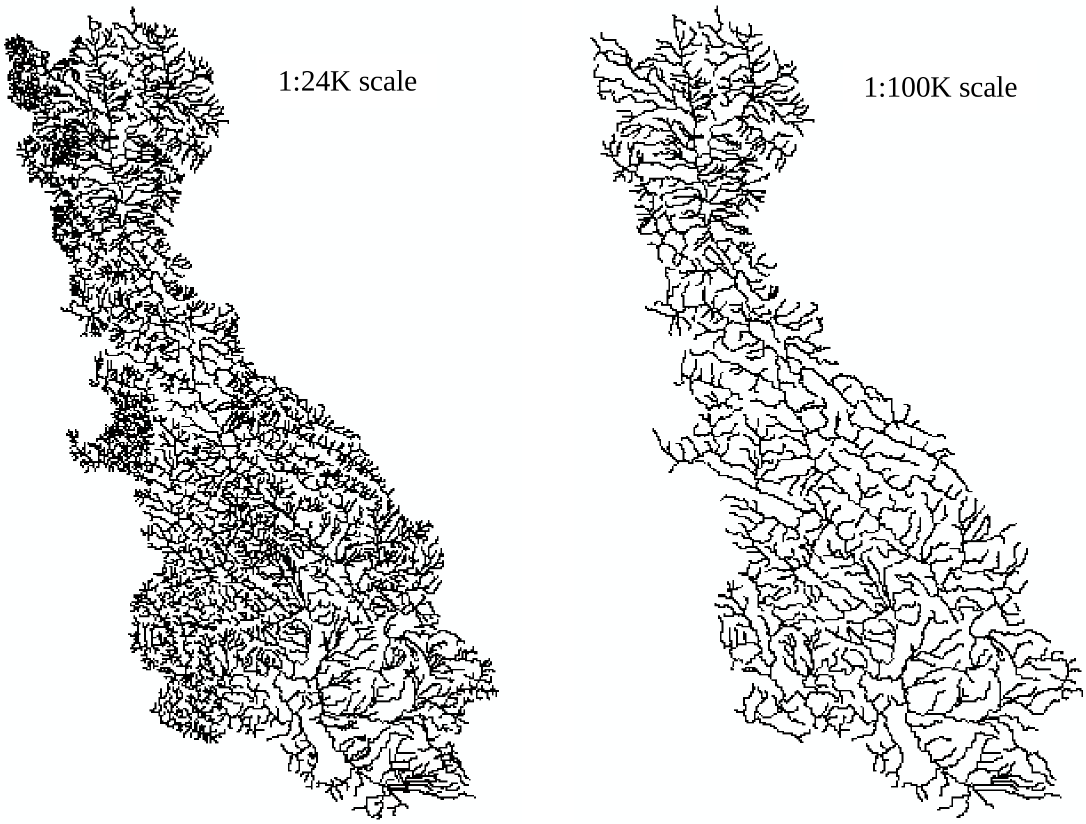


Figure 4.3 Experiment datasets

In the experiment, in order to assess a river network maintains the same drainage pattern after generalization, three generalization methods are tested (Table 4.1). The first two are automatic, and the last one is the manually generalized data. The detailed methods are introduced in Section 4.2.2.

Table 4.1 Testing on three generalization methods

No.	Approaches	Methods
I	Hierarchy	Stroke + Length
II		Watershed partitioning (Catchment)
III	Manual	

4.4.2 MF parameter settings for testing

From the four defined rules for predicting the drainage pattern of river networks, eight predicates are applied into MFs. For the testing, they are set as in Table 4.2. This setting is the same as case I of MFs in the previous chapter.

Table 4.2 MF parameter settings for testing

Predicate	MF	Setting
α IS acute	$z(x;a,b)$	$z(\alpha;45°,90°)$
α IS very acute		$z(\alpha;30°,60°)$
γ IS short		$z(\gamma;0,1)$
δ IS broad		$z(\delta;1,3)$
α IS right	$g(x;a,b)$	$g(\alpha;10°,90°)$
β IS bended	$s(x;a,b)$	$s(\beta;0,1)$
γ IS long		$s(\gamma;0,1)$
δ IS elongated		$s(\delta;1,3)$

4.5 Case studies in Russian river

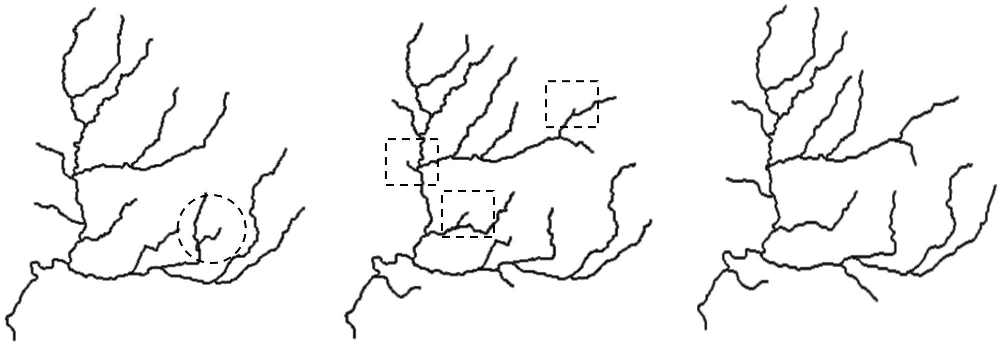
4.5.1 Case 1: a dendritic river network

Figure 4.4(a) shows the tested river network for this case, which is selected from the Russian river corresponding to the region R_2 in Figure 3.13. The bolder the river tributary, the greater the Horton-Strahler order. It is a typical dendritic network with membership value of 0.933. The river network with Horton-Strahler order after upstream routine is illustrated in Figure 4.4(b), which is used to select tributaries by stroke and length.

(a) Horton-Strahler order (b) After upstream routine

Figure 4.4 Tested river network for dendritic case

Generalized river networks by the three methods are illustrated in Figure 4.5. The manual generalized river network at 100K scale from the NHD is shown in Figure 4.5(III). It is generalized overly and does not follow the selection principle of "*Radical Law*". Tributaries are eliminated by stroke and catchment according to the amount of the manual one, so that they can be compared at the same level. River networks generalized by stroke and catchment are shown in Figure 4.5 (I) and (II) respectively.



(I) Stroke + Length (II) Catchment (III) Manual

Figure 4.5 Generalized networks by three methods for dendritic case

In Figure 4.5, all generalized networks are good with visual assessment. However, the manual one is better than others in some details. For example, the tributary in the dashed circle in network (I) is short with a twist that should be eliminated. It is preserved in network (I) because its order is greater than other longer tributaries. There are some short tributaries maintained in network (II) generalized by catchment which are shown in dashed boxes. Network (I) is better than (II), and (III) is the best one.

Table 4.3 Assessment result of generalized networks in Figure 4.5

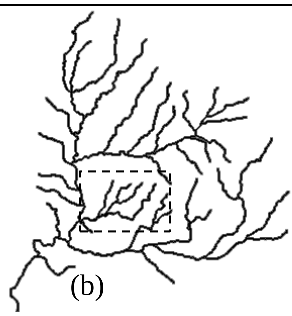
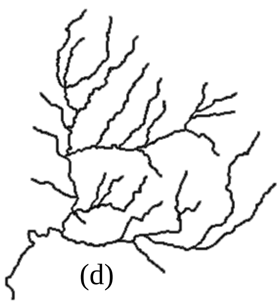
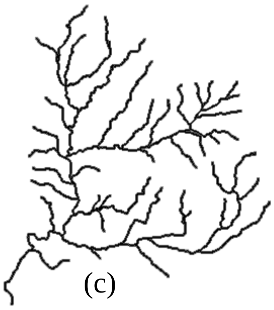
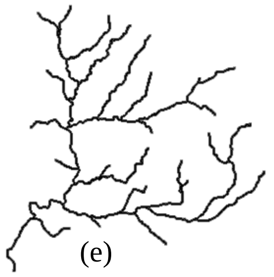
Method		Indicator				Membership Value			
		α	*β*	*γ*	*δ*	D	P	T	R
Stroke + Length	(I)	59.19 °	4.00%	1.10	1.20	0.801	0.002	0	0.003
Catchment	(II)	61.52 °	8.57%	0.58	1.16	0.730	0	0.013	0.015
Manual	(III)	56.52 °	10.34%	0.64	0.99	**0.869**	0	0	0.004

Table 4.3 shows the assessment result of generalized river networks by the three methods. From the table, the membership value of manual network is 0.869,

which is the greatest among all generalized networks. Membership values of network (I) and (II) are 0.801 and 0.730 respectively. Network (I) is better than (II) from the membership, and that is also confirmed by visual assessment.

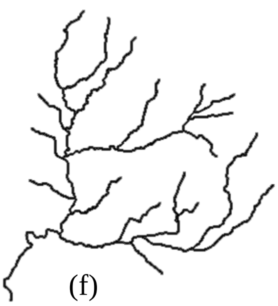
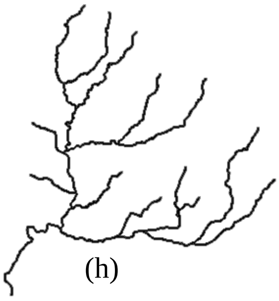
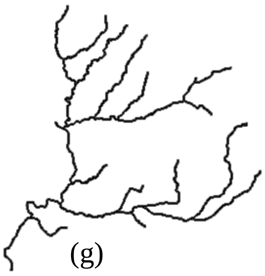
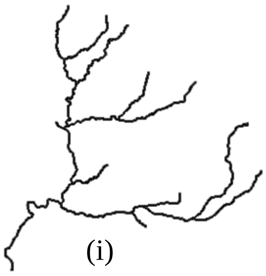
Table 4.4 shows generalized networks by stroke and catchment at different scales. In this case study, 1:100K, 1:250K, 1:500K, 1:1M and 1:5M scales are tested. In the table, the river network (a) is the original data at 24K-scale used for comparison. All generalized results are good, but in general, the stroke and length method provides better results by visual checking. At 1:100K scale, network (c) has more short tributaries due to the shortage of the method, and (b) has better details in the dashed box than (c). At 1:250K, 1:500K and 1:1M scales, networks (d), (f) and (h) look more balanced than (e), (g) and (i) respectively. The first method eliminates tributaries based on strokes that keep the tributaries straighter and longer than the second method. It can be verified visually from the results of (e), (g) and (i) compared to (d), (f) and (h) respectively. At 1:5M scale, network (j) has better shape than (k) as the skeleton of the original network is maintained well in network (j). Obviously, network (j) is better than (k) at this scale.

Table 4.4 Generalized river network for dendritic case at different scales

1:24K	Method	Scale	
		1:100K	1:250K
 (a)	Stroke + Length	 (b)	 (d)
	Catchment	 (c)	 (e)

(*Continued*)

Table 4.4 (*Continued*)

Method	Scale		
	1:500K	1:1M	1:5M
Stroke + Length	 (f)	 (h)	 (j)
Catchment	 (g)	 (i)	 (k)

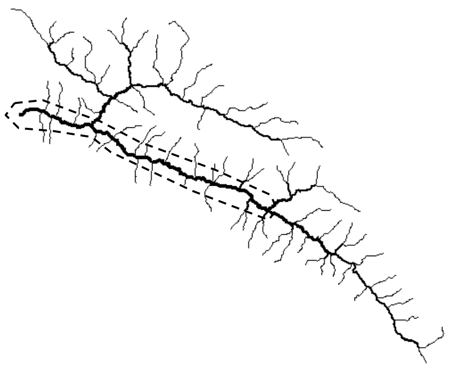
The assessment result for the generalized river networks by different methods at different scales during the generalization process is listed in Table 4.5, and it shows the same findings with the visual assessment. At 1:100K, 1:250K, 1:500K, 1:1M and 1:5M scales, the membership values of the generalized networks by stroke are 0.869, 0.884, 0.762, 0.801 and 0.561 respectively, and they are greater than the values by catchment at each scale. At 1:100K scale, the difference of the memberships between the two methods is very small, which is also confirmed visually that river network (b) and (c) in Table 4.4 are both acceptable. From the membership value of network (k), it changed the pattern from dendritic to rectangular. So network (j) is much better than (k) at 1:5M scale, which also corresponded to the visual assessment. Overall, the stroke method brings better results than the catchment method in this case study from the membership values.

Table 4.5 Assessment result of dendritic case at different scales

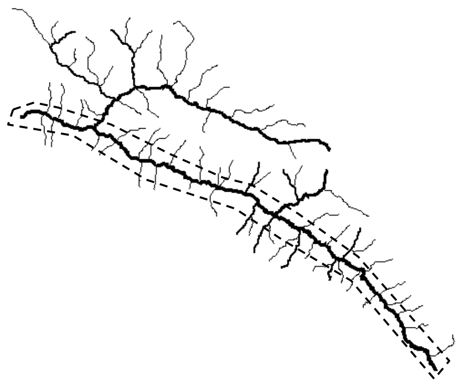
Scale	Method		Indicator				Membership Value			
			α	β	γ	δ	D	P	T	R
1:24K		(a)	53.24°	3.68%	0.69	1.14	0.933	0.010	0.001	0.001
1:100K	I	(b)	55.53°	2.53%	0.68	1.15	**0.869**	0.011	0.004	0.001
	II	(c)	55.85°	3.61%	0.86	1.21	0.861	0.022	0.004	0.001
1:250K	I	(d)	57.55°	4.08%	0.90	1.20	**0.844**	0.013	0.005	0.003
	II	(e)	59.26°	6.38%	0.62	1.10	0.799	0.001	0.006	0.008
1:500K	I	(f)	60.51°	2.86%	0.88	1.20	**0.762**	0	0.013	0.002
	II	(g)	63.76°	6.45%	0.48	1.16	0.653	0	0.014	0.008
1:1M	I	(h)	59.19°	4.00%	1.10	1.20	**0.801**	0.002	0	0.003
	II	(i)	65.16°	4.76%	0.63	1.16	0.599	0	0.014	0.005
1:5M	I	(j)	66.08°	11.11%	1.29	1.23	**0.561**	0	0	0.025
	II	(k)	76.83°	42.86%	0.63	1.18	0.171	0	0.017	0.367

4.5.2 Case 2: a trellis river network

The selected experimental data for this case is a trellis river network in the region of R_3 of the Russian river. It is shown in Figure 4.6(a), and it is arranged as trellis. Figure 4.6(b) illustrates the Horton order after upstream routine.



(a) Horton-Strahler order

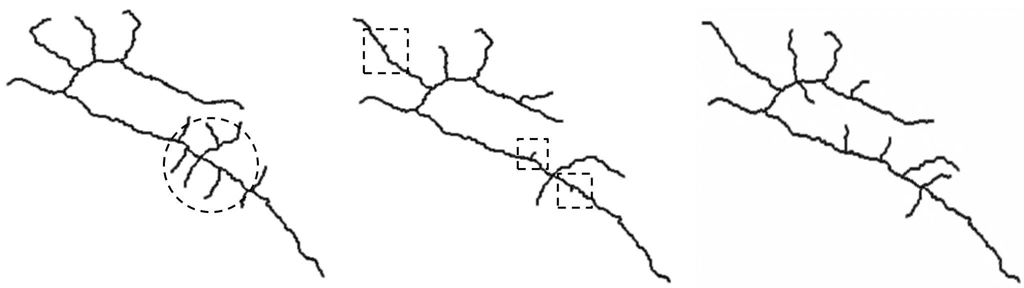


(b) After upstream routine

Figure 4.6 Tested river network for trellis case

In the automatic drainage pattern recognition, the Horton-Strahler order is used for classification. Here, as the river network is already classified as a trellis, the order after upstream routine is used to evaluate generalized results. This is in order to obtain the value of length ratio indicator based on the same main streams. Because the method of stroke and length builds strokes first that is according to the Horton-Strahler after upstream routine. The length ratio values will be higher if other methods do not follow the upstream routine as main streams are shorter. An example of the difference is shown by the dashed polygons in Figure 4.6, where the main stream is obtained owing to different order schemes.

Figure 4.7 shows generalized results by the three methods. In the figure, (III) shows the trellis river network from NHD at 100K scale. It also did not meet the requirement of "*Radical Law*" as too many tributaries are eliminated at this scale in comparison with 1:24K scale network. Networks (I) and (II) are generalized according to the amount left by the manual one. By checking visually, network (III) is well distributed as it is more balance than other results, and tributaries do not gather together as tributaries in the dashed circle in network (I). Network (I) is better than (II) because some short tributaries are preserved by the catchment method such as tributaries in the dashed boxes. Network (III) is still the best result among all generalized networks.



(I) Stroke + Length (II) Catchment (III) Manual

Figure 4.7 Generalized networks by three methods for trellis case

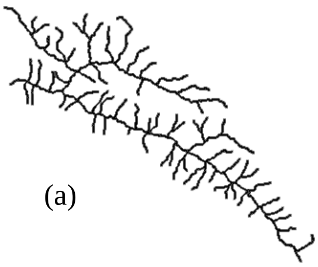
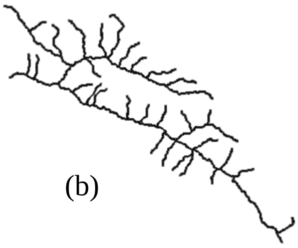
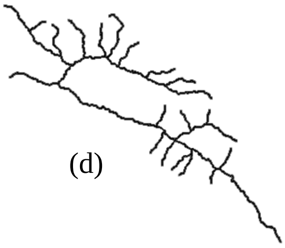
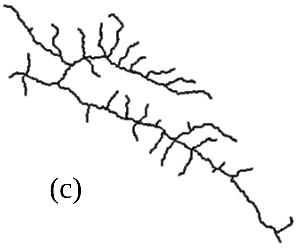
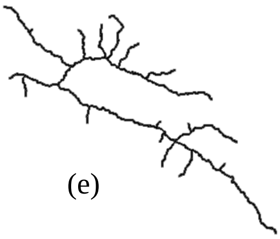
The evaluation result is shown in Table 4.6 which corresponds to the outcome by visual assessment. The manual network obtains a maximum membership value of all generalized networks. Its membership is 0.842, which is greater than both 0.684 of network (I) and 0.396 of network (II). From membership values, network (I) generalized by stroke and length is better than (II) by catchment, which is also confirmed by visual checking.

Table 4.6 Assessment result of generalized networks in Figure 4.7

Method		Indicator				Membership Value			
		α	β	γ	δ	D	P	T	R
Stroke + Length	(I)	98.73°	8.33%	0.21	3.03	0	0	0.684	0.014
Catchment	(II)	103.61°	5.00%	0.29	3.29	0	0	0.396	0.005
Manual	(III)	86.67°	4.35%	0.28	3.65	0	0	**0.842**	0.004

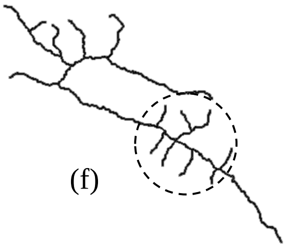
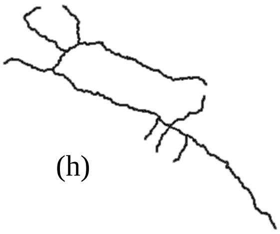
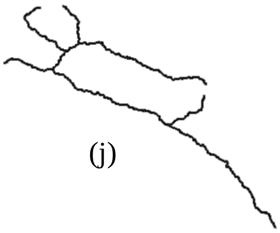
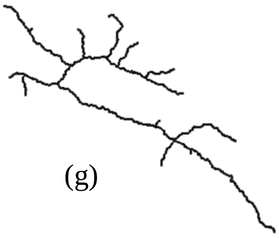
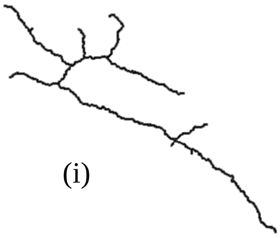
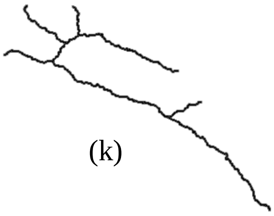
During the generalization process, the trellis river network is handled to generalize from 1:24K scale to 1:100K, 1:250K, 1:500K, 1:1M and 1:2M scales in this case. The results of this case study are listed in Table 4.7, where network (a) is the original trellis river network at 1:24K scale.

Table 4.7 Generalized river network for trellis case at different scales

1:24K	Method	Scale	
		1:100K	1:250K
 (a)	Stroke + Length	 (b)	 (d)
	Catchment	 (c)	 (e)

(*Continued*)

Table 4.7 (*Continued*)

Method	Scale		
	1:500K	1:1M	1:2M
Stroke + Length	 (f)	 (h)	 (j)
Catchment	 (g)	 (i)	 (k)

In Table 4.7, visually, at 1:100K scale, network (c) is better than (b) as (c) looks more balanced, but (b) is still an acceptable result. Network (e) preserves more short tributaries and (d) has more long ones due to stroke establishment. From the aspect of length, network (d) is better than (e), because short tributaries should be removed after generalization especially by manual network. The catchment of a tributary receives all catchment of its upper stream, so the catchment area would be large even if its length is short. That is why short tributaries are preserved in networks (e), (g) and (i). Networks (d), (f) and (h) are more satisfied than (e), (g) and (i) respectively. For the generalized results at 1:2M scale, they are both trellis pattern as the tributaries are short and straight and all junction angle are large. But the tributaries are too few to discuss pattern issue. As a result, most of the generalized networks are better by stroke and length than by catchment at each scale except at 1:100K scale.

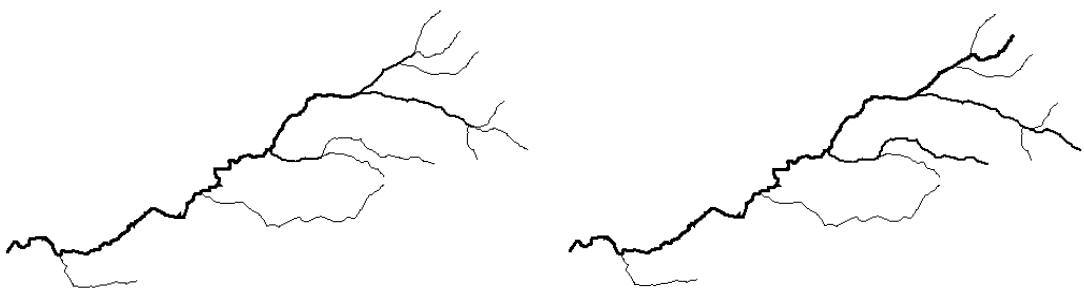
Table 4.8 Assessment result of trellis case at different scales

Scale	Method		Indicator				Membership Value			
			α	*β*	*γ*	*δ*	D	P	T	R
24K		(a)	81.14 °	1.49%	0.20	3.17	0	0	0.675	0
100K	I	(b)	84.72 °	1.56%	0.20	3.35	0	0	0.870	0.001
	II	(c)	88.25 °	1.67%	0.14	3.17	0	0	**0.961**	0.001
250K	I	(d)	84.28 °	2.50%	0.27	3.35	0	0	**0.849**	0.001
	II	(e)	95.83 °	2.94%	0.17	3.09	0	0	0.844	0.002
500K	I	(f)	96.61 °	3.57%	0.23	3.35	0	0	**0.896**	0.003
	II	(g)	100.63 °	4.55%	0.27	3.09	0	0	0.568	0.004
1M	I	(h)	94.13 °	5.00%	0.22	3.65	0	0	**0.907**	0.005
	II	(i)	112.24 °	6.25%	0.31	3.29	0	0	0.843	0.008
2M	I	(j)	98.80 °	8.33%	0.25	3.65	0	0	**0.679**	0.014
	II	(k)	99.87 °	0	0.29	4.13	0	0	0.615	0

Table 4.8 shows the assessment result of the trellis river network during the generalization process. In the table, from the assessment, all generalized networks are preserved as the trellis pattern. At 1:100K scale the membership of (c) is 0.961, which is greater than 0.870 of network (b), and is confirmed by visual assessment. The membership values of networks (d) and (e) are almost the same at 0.849 and 0.844 respectively. But network (d) is better than (e), which also corresponds to the results visually. For other scales, the method of stroke and length brings higher membership values than by catchment as they are 0.896 > 0.568, 0.907 > 0.843 and 0.679 > 0.615 at 1:500K, 1:1M and 1:2M scales respectively.

4.5.3 Case 3: a parallel river network

The river network tested in this experiment is a parallel river network. It is illustrated in Figure 4.8(a), and the river network built after upstream routine with Horton-Strahler order is shown in Figure 4.8(b).



(a) Horton-Strahler order (b) After upstream routine

Figure 4.8 Tested river network for parallel case

Figure 4.9 illustrates generalized networks by the three methods. In the figure, network (III) is the manual river network at 1:100K scale from the NHD. It does not correspond to the selection principle in that it has more tributaries than the network generalized after "*Radical Law*". So, networks (I) and (II) are generalized to the same level of tributary amount of network (III). From the figure, network (II) is not as good as the others because tributaries in the dashed box are short and twist. Networks (I) and (III) are the same results after generalization; however, tributaries in the dashed circles of network (III) are smoother than (I). Although they are the same selection results, the membership values would be a little different. In general, all generalized networks are acceptable by preserving the parallel pattern.

(I) Stroke + Length (II) Catchment (III) Manual

Figure 4.9 Generalized networks by three methods for parallel case

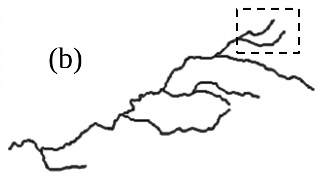
The assessment result for generalized networks by three methods is shown in Table 4.9. Although network (III) from the NHD is nearly the same as (I) simplified by stroke and length, their membership values are a little different: (III) is 0.926 and (I) is 0.930. The smoothed tributaries influenced the calculation of indicators α and β. However, memberships of networks (I) and (III) show the result is better than (II). The result is confirmed by visual assessment.

Table 4.9 Assessment result of generalized networks in Figure 4.9

Method		Indicator				Membership Value			
		α	β	γ	δ	D	P	T	R
Stroke + Length	(I)	34.02°	15.38%	0.81	3.86	0	**0.930**	0	0
Catchment	(II)	41.15°	15.38%	0.57	3.44	0	0.632	0	0
Manual	(III)	27.63°	0%	0.81	4.84	0	0.926	0	0

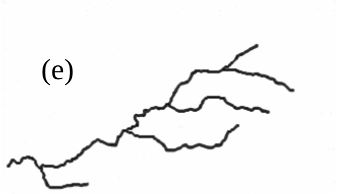
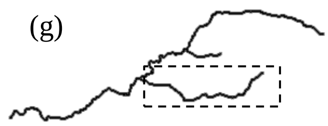
There are not so many tributaries in the parallel network, so only three map scales are involved in the generalization process: 1:50K, 1:100K and 1:250K scales. The scales smaller than 250K are not necessary to be tested as there are only 5 river segments left at 1:250K scale. Table 4.10 shows the generalized results for this case in different scales. Network (a) is the original parallel river network, and its membership value is 0.792.

Table 4.10 Generalized river network for parallel case at different scales

1:24K	Method	Scale
		1:50K
(a) 	Stroke + Length	(b) 
	Catchment	(c) 

(*Continued*)

Table 4.10 (*Continued*)

Method	Scale	
	1:100K	1:250K
Stroke + Length		
Catchment		

In Table 4.10, visually, all generalized river networks maintained the parallel pattern. At 1:50K scale, network (b) and (c) are both acceptable, but (b) is better than (c) because the junction angle in the dashed box of (b) is smaller, which is in keeping with characteristics of the parallel pattern, and tributary in the dashed circle of network (c) is not as straight as the same place in (b). At 1:100K scale, junction angles in both network (d) and (e) are the same, under which circumstance the network with longer and straight tributaries is more satisfactory. So, (d) looks like a parallel network more than (e). By visual checking, network (g) looks more parallel than (f) because (g) preserves a longer tributary in the dashed box of (g) although its order is only 1.

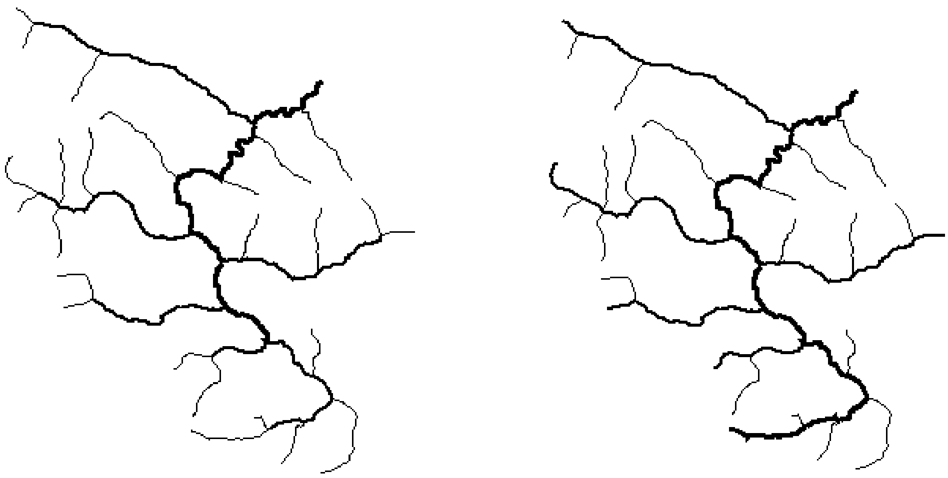
Table 4.11 Assessment result of parallel case at different scales

Scale	Method		Indicator				Membership Value			
			α	β	γ	δ	D	P	T	R
1:24K		(a)	39.67 °	10.53%	0.69	3.14	0	0.792	0	0
1:50K	I	(b)	34.02 °	15.38%	0.81	3.86	0	**0.930**	0	0
	II	(c)	41.15 °	15.38%	0.57	3.44	0	0.632	0	0
1:100K	I	(d)	41.80 °	22.22%	0.34	3.86	0	**0.225**	0	0
	II	(e)	41.80 °	22.22%	0.29	3.71	0	0.165	0	0
1:250K	I	(f)	47.58 °	20.00%	0.34	4.09	0	0.235	0	0
	II	(g)	35.27 °	20.00%	0.47	3.71	0	**0.439**	0	0

The assessment result of the parallel network is shown in Table 4.11. From the table, membership values of the network generalized by stroke and length are 0.930 and 0.225 at 1:50K and 1:100K scales, respectively. They are greater than the generalized network by catchment at both scales, which is confirmed by visual checking. At 1:250K scale, the membership value of network (g) is greater than (f) (0.439 > 0.235), which reflects (g) looks more parallel than (f). This result is confirmed by visual assessment.

4.5.4 Case 4: a rectangular river network

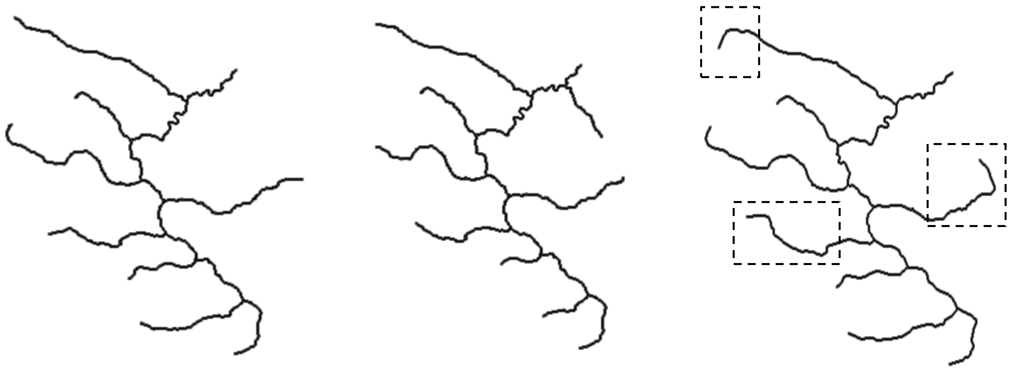
This testing network for the rectangular case is selected from the Russian river corresponding to the region R_2. Figure 4.10(a) illustrates the river network with Horton-Strahler order, and (b) shows it after upstream routine with the order.



(a) Horton-Strahler order (b) After upstream routine

Figure 4.10 Tested river network for rectangular case

Figure 4.11 shows generalized river networks by the three methods. Network (III) is the manual one from NHD at 1:100K scale. All river networks are generalized to the same level on tributary amount. Networks (I) and (II) are generalized by stroke and length and by catchment respectively. Visually, all networks are acceptable by preserving the rectangular pattern, and they are almost the same. Network (III) generalized manually has more bended tributaries than others (see dashed boxes). The manual one is better than the others.



(I) Stroke + Length (II) Catchment (III) Manual

Figure 4.11 Generalized networks by three methods for rectangular case

The assessment result in Table 4.12 shows the same conclusion with visual assessment. The membership of network (III) is 0.037, which is slightly greater than the others. Membership values of networks (I) and (II) are 0.030 and 0.028, respectively, which are almost the same. The result corresponds to the visual assessment.

Table 4.12 Assessment result of generalized networks in Figure 4.11

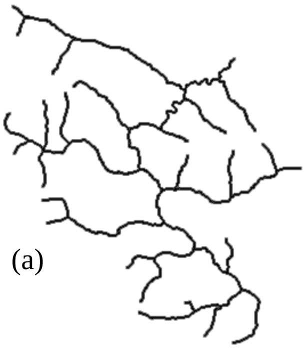
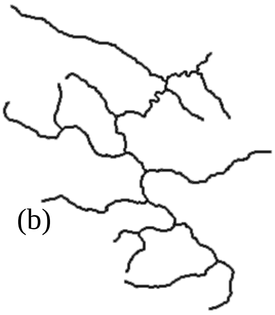
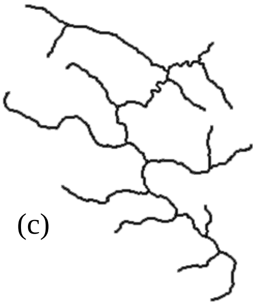
Method		Indicator				Membership Value			
		α	β	γ	δ	D	P	T	R
Stroke + Length	(I)	116.54ˁ	13.33%	0.32	0.87	0	0	0	0.030
Catchment	(II)	115.33ˁ	11.76%	0.25	1.12	0	0	0.007	0.028
Manual	(III)	115.74ˁ	20.00%	0.35	0.80	0	0	0	**0.037**

In this case, 1:100K, 1:250K and 1:1M scales take part in the experiment. Table 4.13 shows the generalized results by stroke and catchment from 1:24K scale to 1:1M scale. Network (a) is the original data. Networks (b), (d) and (f) are given by stroke and length at 1:100K, 1:250K and 1:1M scales respectively, and catchment method provides (c), (e) and (g) at each scale.

In Table 4.13, all generalized river networks look rectangular more than the original one (a) because (a) has more straight tributaries. At 1:100K scale, network (b) preserves more bended tributaries than (c), from which aspect, (b) is better than (c). With network (d) and (e) it is hard to tell which is better, but tributaries in (d) are longer than (e). In fact, they are both acceptable visually. At

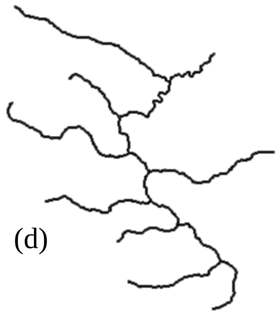
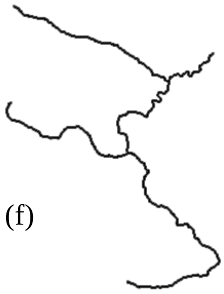
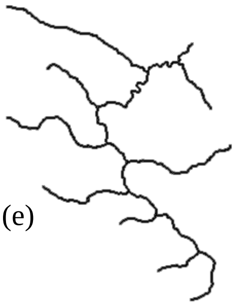
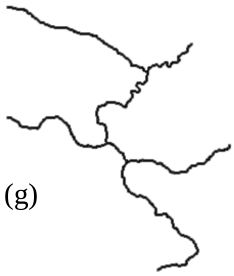
1:1M scale, there are few tributaries left in networks (f) and (g), and they are both good.

Table 4.13 Generalized river network for rectangular case at different scales

1:24K	Method	Scale 1:100K
 (a)	Stroke + Length	 (b)
	Catchment	 (c)

(*Continued*)

Table 4.13 (*Continued*)

Method	Scale	
	1:250K	1:1M
Stroke + Length	 (d)	 (f)
Catchment	 (e)	 (g)

The assessment result of this case is shown in Table 4.14. From the table, we can see the membership value of the original network (a) is 0.003 for the rectangular pattern only. As membership values of other patterns are 0, it can only be classified as rectangular. However, generalized networks are better than the original one. At 1:100K scale, the membership of (b) is 0.034, which is greater than 0.013 of (c). From the assessment, at 1:250K scale, membership values of networks (d) and (e) are 0.030 and 0.028 respectively. They are so close that it is difficult to evaluate visually. Networks (f) and (g) have the same membership value of the rectangular pattern and also corresponded with visual assessment.

Table 4.14 Assessment result of rectangular case at different scales

Scale	Method		Indicator				Membership Value			
			α	β	γ	δ	D	P	T	R
1:24K		(a)	94.93°	4.08%	0.33	0.87	0	0	0	0.003
1:100K	I	(b)	105.93°	13.04%	0.61	0.87	0	0	0	**0.034**
	II	(c)	106.89°	8.00%	0.47	0.97	0	0	0	0.013
1:250K	I	(d)	116.54°	13.33%	0.32	0.87	0	0	0	**0.030**
	II	(e)	115.33°	11.76%	0.25	1.12	0	0	0.007	0.028
1:1M	I	(f)	120.58°	28.57%	0.80	0.61	0	0	0	0.009
	II	(g)	120.58°	28.57%	0.66	0.93	0	0	0	0.009

4.6 Evaluation results in Russian river

The evaluation method is applied to the Russian river to assess generalized river networks by the three methods: stroke and length, catchment and manual work. The data process is as follows:

(1) According to the river data from NHD at 1:100K scale, eliminate tributaries in the Russian river to get the generalized river network by manual work; rebuild the network by combining river segments and reassign the Horton-Strahler order, then establish the drainage tree for the network.

(2) According to the river segment IDs, get the corresponding sub-networks from the Russian river at 1:24K scale.

(3) Generalize the sub-networks by stroke and catchment method to the same level amount of river segments with manual generalized networks.

(4) Assess each generalized river network by the evaluation method.

Table 4.15 lists the number of preserved or changed drainage patterns after river network generalization.

Table 4.15 Number of drainage patterns after generalization

	Manual			Catchment			Stroke + Length		
	Order 2	Order 3	Order 4	Order 2	Order 3	Order 4	Order 2	Order 3	Order 4
Dendritic	15	29	13	14	34	17	13	29	15
Parallel	14	4	0	17	6	0	16	5	0
Trellis	2	6	2	3	7	4	3	6	3
Rectangular	0	2	1	0	3	0	0	3	1
Unclassified	2	0	0	2	1	0	2	0	0
D→P	16	0	0	15	2	0	19	0	0
D→T	15	2	1	13	2	1	9	5	1
D→R	9	4	1	5	5	1	9	3	1
D→U	4	1	0	1	0	0	6	0	0
P→D	2	1	0	1	0	0	2	0	0
P→T	3	0	0	0	0	0	1	0	0
P→R	0	0	0	0	0	0	0	0	0
P→U	0	0	0	0	0	0	0	0	0
T→D	0	0	0	0	0	0	0	0	0
T→P	5	0	0	2	0	0	3	0	0
T→R	1	0	0	0	0	0	0	0	0
T→U	1	1	0	2	0	0	3	0	0
R→D	0	0	0	0	0	0	0	0	0
R→P	1	0	0	0	0	0	0	0	0
R→T	3	0	0	3	1	0	3	0	0
R→U	0	0	0	0	0	0	0	0	0
U→D	1	0	0	0	0	0	1	0	0
U→P	0	0	0	1	0	0	0	0	0
U→T	1	0	0	0	0	0	1	0	0
U→R	1	0	0	1	0	0	1	0	0
Total	164			164			164		

* “D” – Dendritic, “P” – Parallel, “T” – Trellis, “R” – Rectangular, “U” – Unclassified; “→” means that one pattern changes to another.

In Table 4.15, the first five rows give the number of preserved patterns, and following rows are the changed numbers of each pattern in detail. There are 164 river networks at different orders that are extracted and evaluated. From the table, many of the generalized river networks are preserved drainage patterns by the three methods. There are 90, 108, and 96 generalized river networks that preserve their patterns by manual work, catchment and stroke respectively. Although patterns of many networks are changed after generalization, it happens in order 2. In manual work, 74 generalized networks alter patterns, but 85% (63/74) of them are in order 2. Similarly, 79% (44/56) and 85% (58/68) of changed patterns by catchment and stroke respectively are in order 2. The possible reason is that indicators from a river network are statistic values, which rely on the amount of river segments. If there are few river segments in a river network, the indicators would be not so robust to reflect the pattern of the river network. From the result of the previous chapter in Figure 3.18 (p. 53), most of the river networks in order 2 have less than 5 river segments. Therefore, if a river network in order 2 is generalized from the network in order 3 or a higher order, two situations would arise: one is that the pattern does change after generalization, and another is that the evaluation method is not available due to insufficient river segments. In addition, from the table, most patterns change from dendritic to parallel, trellis and rectangular.

Table 4.16 shows average membership values of all generalized river networks where their patterns are preserved. From the table, the average membership value of river networks generalized by manual work is 0.59, which is slightly greater than by catchment (0.52) and by stroke and length (0.57). It indicates that, from the aspect of drainage patterns, river networks generalized by manual work are better than by catchment and stroke, which corresponds to the result from case studies. The average value given by the stroke and length method is close to the manual generalized river networks. Here the stroke is established based on the Horton-Strahler order after upstream routine, which has been considered as the one that "*most closely approximates the generalisation decisions made by a human cartographer*" (Thomson & Brooks, 2000).

Table 4.16 Average membership value of preserved patterns

Method	Stroke + Length	Catchment	Manual
Average membership value	0.57	0.52	0.59

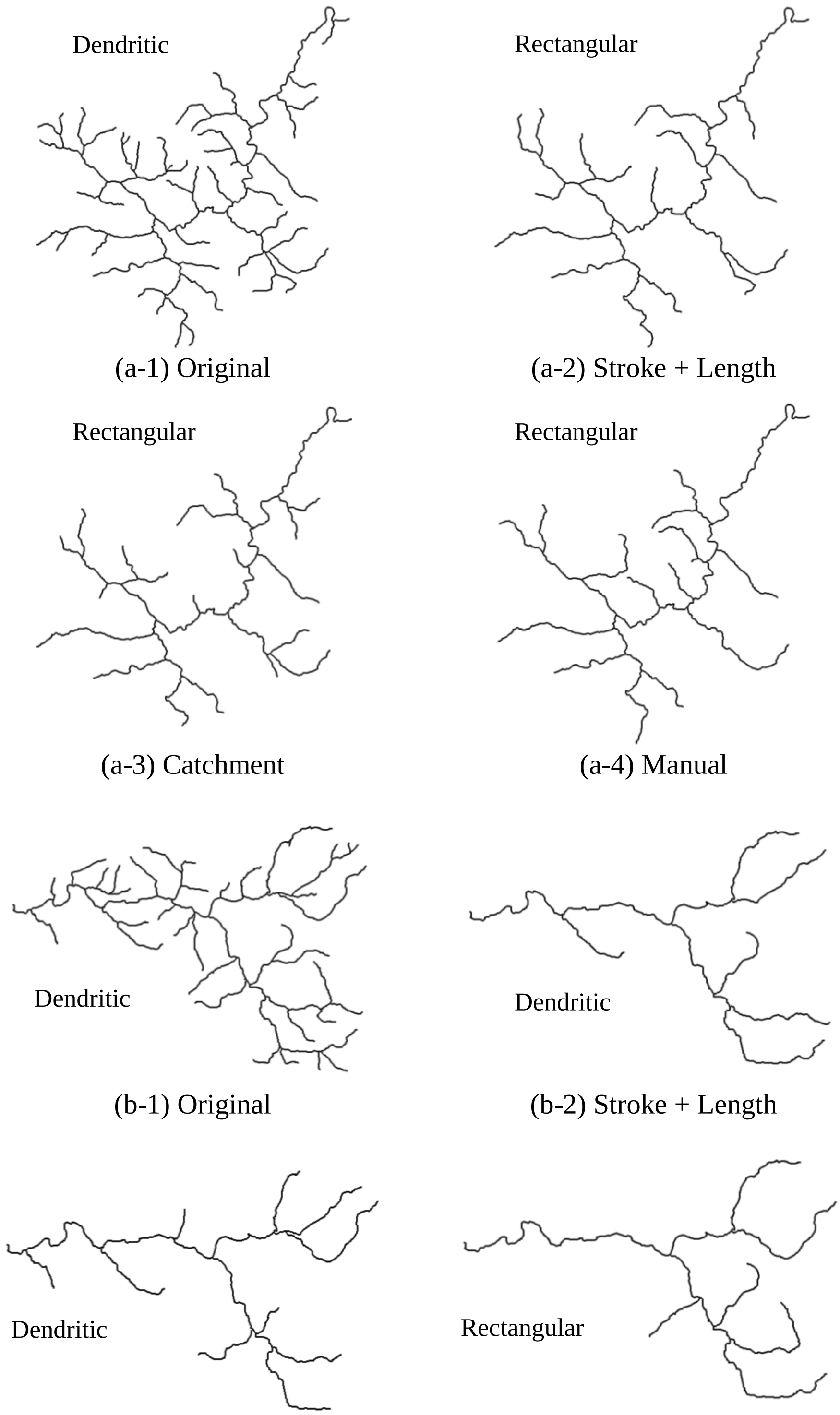


Figure 4.12 Some generalized river networks with changed patterns

Some examples of river networks that change their patterns after generalization are illustrated in Figure 4.12. In the figure, (a-1) and (b-1) are original networks; (a/b-2), (a/b-3) and (a/b-4) are generalized river networks by stroke and length, catchment and manual work. Table 4.17 shows the assessment results of the generalized river networks. From the table, network (a-1) is dendritic, but generalized networks (a-2), (a-3) and (a-4) are changed to rectangular. For another example, network (b-4) generalized by manual work alters the pattern from dendritic to rectangular, and networks (b-2) and (b-3) still maintain the pattern.

Table 4.17 Assessment for river networks in Figure 4.12

Network	Indicator				Membership Value			
	α	β	γ	δ	D	P	T	R
(a-1)	83.52 °	10%	1.43	1.42	**0.041**	0	0	0.020
(a-2)	108.65 °	29%	0.84	1.42	0	0	0.053	**0.169**
(a-3)	100.80 °	22%	0.95	1.55	0	0	0.004	**0.093**
(a-4)	104.98 °	23%	0.97	1.46	0	0	0.001	**0.102**
(b-1)	73.15 °	10%	1.38	1.45	**0.281**	0	0	0.018
(b-2)	72.56 °	27%	0.89	1.58	**0.300**	0	0.025	0.149
(b-3)	75.21 °	21%	0.60	1.61	**0.216**	0	0.188	0.089
(b-4)	76.40 °	45%	0.97	1.59	0.183	0	0.001	**0.396**

From Table 4.17, the membership value of each river network is not so large. Therefore, we check the membership values of river networks that preserved or changed their patterns. For river networks that preserved their patterns after generalization, the average membership value of original river networks is **0.57**; the average value is **0.30** only for river networks that changed patterns. It indicates that, in general, if the source river network has high membership value of a pattern, which has a significant characteristic of a pattern, it would be easier to preserve its pattern after generalization than a river network with low membership value.

4.7 Conclusion

From the experiment results, several conclusions can be given as follows.

(1) In general, the evaluation method based on the membership degree of a fuzzy rule for a drainage pattern is useful. From a large scale to a small scale, to a generalized river network, the drainage pattern preserves better

if the membership value is high. However, sometimes, the membership value will be not so robust at small scales especially when there are not enough river segments left because proposed indicators, such as average junction angle (α), bended tributaries percentage (β), average length ratio (γ), are statistical features.

(2) By evaluating generalized river networks from the point of drainage patterns, the method based on stroke and length is better than based on watershed partitioning. In addition, networks generalized manually are always with high membership values and preserve a good drainage pattern. A good generalized result does not only depend on one or two factors; many factors such as tributary spacing and balance are involved in manual generalization process.

(3) One limitation of the proposed evaluation method is focused on the drainage pattern only. Some other aspects simply cannot be assessed by the membership value. For example, for network (f) in Table 4.7 at 1:500K scale, although the membership value is 0.896 that is much greater than (g), it is not an ideal result as the tributaries in the dashed circle are crowded together.

(4) Another limitation is that the evaluation method is more available and accurate in source river networks with order 3 or higher, but not the higher the better because sub-networks can be classified as different patterns inside a large river network. A small river network with order 2 does not have enough river segments to provide robust indicators.

4.8 Summary

In this chapter, a quality assessment method based on fuzzy logic is provided to evaluate a generalized river network by the tributary selection operation. The quality evaluation is by checking a membership value from a fuzzy rule for a drainage pattern. Four drainage patterns are evaluated in this study: dendritic, trellis, parallel and rectangular. The method was applied to evaluate different tributary selection methods, such as by stroke and length, by watershed portioning and by manual work. The experimental data is the Russian river from the RRIIS at 1:24K scale, and the NHD at 1:100K scale.

From the experimental results, when the membership value is higher, the generalized river network is better. During the generalization, the membership of generalized river networks can be higher than the original. That is because a generalized river network can have more characteristics of the pattern than the original network after generalization. This method is appropriate for evaluating a generalized river network from the aspect of drainage patterns. The advantage of this research is that evaluating a generalized river network based on fuzzy logic is easy to understand and implement. The limitations of the research are: (1) evaluation is focused on the drainage pattern only according to the membership value, other criteria may also be proposed; (2) the method is more suitable for a river network with order 3 and 4 as a small network does not have enough river segments and a large network can have many sub-networks with different patterns inside.

The existing methods of tributary selection do not consider the pattern in the first place, although they can preserve the pattern of a generalized river network sometimes. Considering the pattern is an important factor in river networks, it should be taken into account in river network generalization. In order to provide a better generalized river network, the next chapter proposes a tributary selection method with consideration of drainage patterns. From the experiment of this work, only focusing on drainage patterns cannot generalize a river network as good as manual work. The indicators influencing the drainage pattern can be considered in the generalization, however, other factors are also needed, such as tributaries balance and spacing. The next chapter proposes a solution to deal with multiple factors at the same time during the river network generalization.

Chapter 5 A genetic algorithm for tributary selection with consideration of different factors

5.1 Introduction

The drainage pattern can reflect the characteristics of a river network to a certain extent. Therefore, drainage pattern preservation should be integrated in the river network generalization process. Each drainage pattern has its own characteristics, so different factors may be considered according to the drainage pattern. Existing methods such as those presented in the previous chapter are based on the consideration of different factors, among which the tributary length and the order are the most important.

Optimizing river selection according to different factors at the same time is a multi-objective optimization problem. In recent years, genetic algorithms (GAs) have been applied in multi-objective optimization problems (e.g. Coello, Lamont, & Van Veldhuizen, 2007; Hajela & Lin, 1992; Konak, Coit, & Smith, 2006). A genetic algorithm is a class of adaptive stochastic optimization algorithms that simulates the process of natural evolution, and usually it is used to create available solutions to optimization and search problems (Mitchell, 1996). GA is a sub-class of evolutionary algorithms (EAs), which are inspired by Darwin's theory about evolution and widely noticed since 1960s (Rechenberg, 1973). GAs were proposed by John Holland in the early 1970s (Holland, 1975), and developed by him and his students (Holland, 1992). Research in GAs stayed in the theoretical realm until the mid-1980s, then GAs started to be applied in many fields such as bioinformatics (e.g. Kikuchi et al., 2003; Kosakovsky Pond et al., 2006; Notredame, 1996), computational science (e.g. Deb et al., 2002; Della Croce, Tadei, & Volta, 1995; Goldberg, 1989), and engineering (e.g. Gen & Cheng, 1999; Johnson & Rahmat-Samii, 1997; Shopova & Vaklieva-Bancheva, 2006). Van Dijk, Thierens, & De Berg (2002) use GAs to resolve GIS problems, such as map labeling and generalization while preserving the data structure, and line simplification; Ware, Wilson, & Ware (2003) focus on spatial conflict between objects after scaling achieving near optimal solutions within practical time constraints.

River network generalization is usually conducted in two steps: first is the selective omission of river segments; second, line simplification and typification is performed. In this chapter, only selective omission is discussed. In order to get a better generalized river network, several factors are considered and introduced. For different drainage patterns, considered factors are different. A genetic algorithm is designed and implemented for tributary selection, and the method is implemented and tested in the Russian river. The chapter is organized as follows: Section 5.2 introduces basic concepts of GA and explains how they are applied to omissive selection of tributaries. Section 5.3 presents the different factors assessed in our experiments and the objective function evaluating the generalization. In Section 5.4, the selection method is applied for each type of patterns and results showing the importance of different factors are analyzed. The last section provides recommendations for each pattern and discusses the performances of the method.

5.2 Tributary selection using a genetic algorithm

5.2.1 Fundaments of genetic algorithm

To apply a GA, the solution (called individual) to the problem should be represented by a chromosome (or genome). Usually, a solution is represented by series of ones and zeros, but there are also other possible encodings (Whitley, 1994). Then, a set of solutions called population is generated, and genetic operators such as selection, crossover and mutation are applied to evolve the solutions in order to find the best one(s) by evaluating the fitness of every individual in the population. In general, a GA has 5 elemental components as follows: encoding, population initialization, fitness evaluation, genetic operators, and parameters setting (population size, probabilities of applying genetic operators, etc.).

(1) *Encoding.* The encoding of chromosomes depends on the problem. There are some common encoding types: binary, ordered, and valued encodings. The most common one is the binary encoding represented as an array of bits, 0 or 1. The ordered encoding is a string of numbers which is useful for ordering problems. In the valued encoding, every chromosome is permutation of real values, which can be related to special problems represented as real numbers, characters, or some objects. Examples of chromosomal encoding with applications are illustrated in Table 5.1.

Table 5.1 Examples of chromosomal encodings

Type	Encoding example	Application
Binary	A: 100101010101010 B: 011101011110101	Knapsack problem
Ordered	A: 132547689 B: 145236798	Travelling salesman problem (TSP)
Valued	A: 1.23 4.35 6.89 4.56 B: ABDEJEIFJLSIEJD C: (red), (yellow), (green)	Finding weights for neural networks

(2) *Population initialization*. A population is a group of candidate solutions, and the initial population is the first generation. Traditionally, the initial population to a problem is generated randomly from the entire search space containing all possible solutions.

(3) *Fitness evaluation*. It is a process evaluating all potential solutions by a fitness function that returns a "*fitness*" value which can reflect how optimal a solution is. Usually, the higher the "*fitness*", the better the solution. "*Fitness*" plays an important role in the generation of new populations.

(4) *Genetic operators*. Selection, crossover, and mutation are three basic genetic operators of GAs, which are used to create new offspring forming the next generation. The selection operation selects good chromosomes to be parents

to multiply offspring according to their fitness. There are several methods to select chromosomes, such as roulette wheel selection, rank selection and elitism. These methods are following the rule of "*survival of the fittest*". In crossover operation, a new offspring is created by two partial genes from parents. Mutation randomly changes a gene in the new offspring, which can prevent a new generation falling into a local optimum. Figure 5.1 shows the simplest way of crossover and mutation to binary chromosomes, the symbol "|" represents a crossover point.

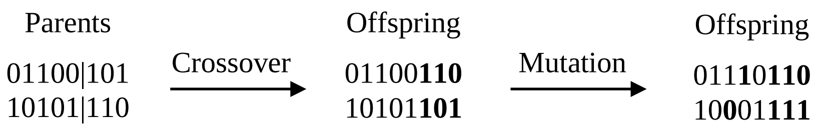


Figure 5.1 Crossover and mutation operation

(5) *Parameters setting*. ① Crossover probability (p_c) is the chance of crossover happening. If p_c is 0%, the new generation is an exact copy of the last one. If it is 100%, all offspring are created by crossover. ② Mutation probability (p_m) is the probability of a gene in a chromosome is mutated. If p_m is 100%, all genes in a chromosome are changed, if it is 0%, no gene is changed. ③ Population size indicates the number of chromosomes in a population.

The process of a basic GA is shown in the figure as follows.

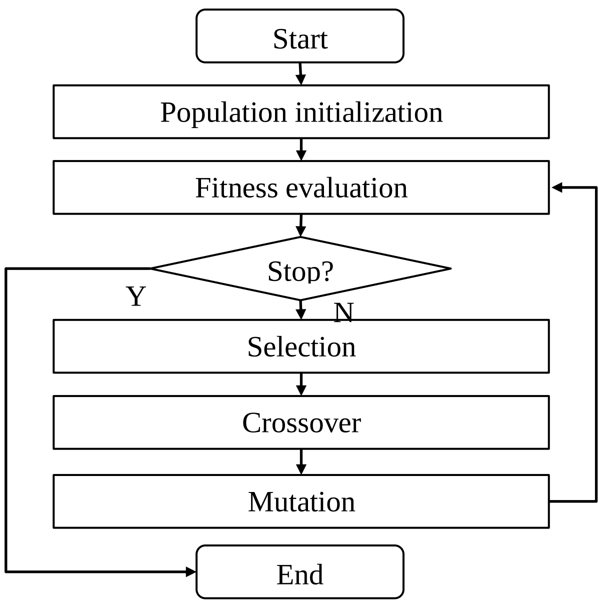
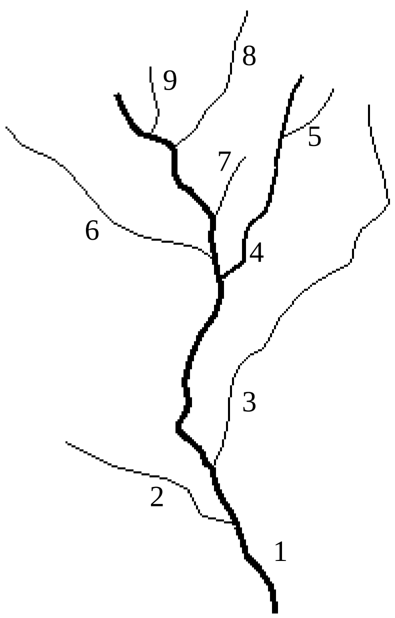


Figure 5.2 A basic GA process

5.2.2 Encoding of a river network

The binary encoding method is adopted as it is easy to indicate that 0 is omission and 1 is selection for a tributary. The Horton-Strahler order scheme after upstream routine, which is used to establish strokes in a river network by the method in Section 4.2.2.1 (p. 69), is regarded as the closest generalization decisions made by a human cartographer. Therefore, strokes are applied to build a chromosome. The length of a chromosome is the number of all strokes in a river network. In order to preserve the topology of a river network, a stroke cannot be omitted if its upper rivers are selected.

Figure 5.3 shows a river network and some examples of chromosomes with binary encoding of a river network.



(a) A river network

Stroke ID	1	2	3	4	5	6	7	8	9
Chromosome 1	1	1	1	1	1	1	1	1	1
Chromosome 2	0	0	0	0	0	0	0	0	0
Chromosome 3	1	0	1	1	0	1	0	0	0

(b) Binary encoding examples

Figure 5.3 Examples of chromosomes with binary encoding of a river network

In Figure 5.3(a), a simulated river network is illustrated, where the number is the ID of a stroke. In Figure 5.3(b), in chromosome 1, all strokes are selected, while all strokes are omitted in chromosome 2. In chromosome 3, only strokes with IDs 1, 3, 4 and 6 are selected, and others are omitted.

5.2.3 Initialization

In the initialization process, the number of selected strokes (marked as *Ns*) can be calculated by the "*Radical Law*" or other methods. Some rules should be followed:

(1) The number of genes set to 1 in a chromosome is equal to *Ns*.

(2) A gene cannot be assigned 0 if it breaks the topology of a river network, that is to say, a stroke cannot be omitted if its upper strokes are selected.

(3) Strokes with the higher order have priority to be selected; otherwise, strokes with lower order will be omitted first.

The steps of initialization are given as follows.

Step 1: Set a list *Ls* storing all strokes,
Na is the number of all strokes,
No is the omitted number given by *Na* – *Ns*,
Nc is current omitted number initialized as 0,
and a candidate list *Lc* storing omitted strokes.

Step 2: Randomly choose a stroke *Rs* from *Ls* (strokes with lower order have high priority to be chosen), the number of its upper strokes is *Ns* (including *Rs*).

Step 3: If *Nc* + *Ns* > *No*, then remove *Rs* from *Ls*, and go back to step 2; otherwise,
Nc += *Ns*,
add *Rs* and its upper strokes into *Lc*,
and remove *Rs* and its upper strokes from *Ls*.

Step 4: Repeat from step 2 until *Nc* = *No*.

Step 5: All strokes in *Lc* are omitted ones, assign their corresponding genes in the chromosome as 0, and others are 1.

5.2.4 Selection

In this work, the elitist model is used for the selection operation in the GA. Elitism directly copies the best chromosome to a new population without any other reproduction operations. This method can rapidly increase the performance of the GA, and it preserves the best solution all the time.

5.2.5 Reproduction

The reproduction for tributary selection using GA should be customized following similar rules to initialization. They are:

(1) After reproduction, the number of selected genes must be equal to *Ns*.
(2) The reproduction cannot break the topology of a river network, which cannot omit a stroke if it has upper strokes.

5.2.5.1 Crossover

In order to obey the rules of the reproduction for the tributary selection using GA, the crossover operation cannot be applied normally as one-point-crossover or two-point-crossover. The information exchange between the two parent chromosomes should be controlled to follow the rules. Here, a mask, which is represented as a chromosome with same length, is used to determine which genes are inherited from which parents. An offspring is generated as indicated in the mask: a gene is from the first parent chromosome if the mask gene is 1 and from the second parent if it is 2.

The steps of crossover are described as follows.

Step 1: Determine the parent chromosomes *C1* and *C2*;
Set a mask *M* with all genes with 1.

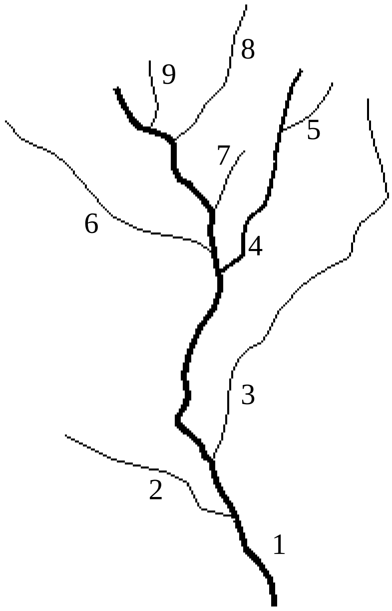
Step 2: Search *C1* and *C2*, and find positions that the allelic genes are different and store them in a list *Lp*.

Step 3: Do list traversal in *Lp*, for each position *P*, if the genes change in this position, the network topology would be broken, then remove *P* from *Lp*.

Step 4: The *M* is set as paired genes are exchanged between *C1* and *C2*: one is from 0 to 1, and another should be from 1 to 0.

Step 5: According to *M*, generate a new offspring *O*.

Step 6: Loop to step 2 until two different offspring are generated.



(a) a river network

	1	2	3	4	5	6	7	8	9
C1	1	1	1	0	0	1	0	0	1
C2	1	0	1	1	1	0	1	0	0
M	1	2	1	1	1	1	2	1	1
O	1	0	1	0	0	1	1	0	1

(b) crossover process

Figure 5.4 An example of crossover

Figure 5.4 shows an example of crossover process with two chromosomes. The solution is to select six strokes from a river network in Figure 5.4(a). In Figure 5.4(b), *C1* and *C2* are parents, and six alleles are different. However, the allelic genes in positions 4 and 5 are invalid. If genes exchange in position 4 or 5, stroke 5 would be separated from the river network in *C2* or *C1* respectively. In order to keep the number of genes with 1 value after the crossover, only positions 2 and 7 in the chromosome are marked to exchange information between the parents. So, a new offspring is generated by taking genes in positions 2 and 7 from *C2*. The rest is from *C1*.

5.2.5.2 Mutation

The mutation operation is to change a gene in a chromosome in order to make a solution jump out of a local optimum. Here, only a gene to be changed cannot satisfy the rule 1 of the reproduction. If a gene is changed from 0 to 1, another gene needs to be changed from 0 to 1. In the process, only one paired genes are supposed to be changed.

The modified mutation for selecting tributaries is given as follows.

Step 1: For a chromosome C,

set a list *L1* that stores all gene positions of 1,

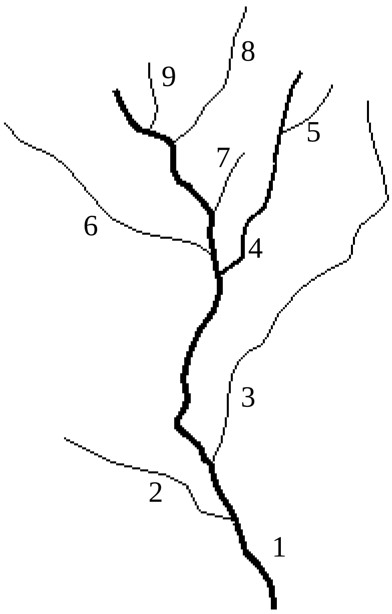
set a list *L0* that stores all gene positions of 0.

Step 2: Do list traversal in *L1*, some positions cannot be changed to 0 because their upper stokes are selected, remove these position from *L1*.

Step 3: Randomly choose a position *P1* from *L1*, change the gene to 0.

Step 4: Do list traversal in *L0,* some positions cannot be changed to 1 because their lower strokes are omitted, remove these positions from *L0*.

Step 5: Randomly choose a position *P0* from *L0*, change the gene to 1.



(a) a river network

	1	2	3	4	5	6	7	8	9
C	1	0	1	0	0	1	1	1	1
Step 1	L1: 1 3 6 7 8 9 L0: 2 4 5								
Step 2	L1: 3 6 7 8 9 L0: 2 4 5								
Step 3	1	0	0	0	0	1	1	1	1
Step 4	L1: 3 6 7 8 9 L0: 2 4								
C'	1	0	0	1	0	1	1	1	1

(b) mutation process

Figure 5.5 An example of mutation

In Figure 5.5, an example of mutation is illustrated. *C* is a chromosome that represents a solution of selecting six strokes for the river network in Figure 5.5(a). After the step 1, *L1* and *L0* are established. In the step 2, stroke 1 is removed from *L1*. Step 3 chooses stroke 3 to be changed from 1 to 0. Then, in the step 4, stroke 5, which cannot be omitted, is removed from *L0*. Finally, stroke 4 is chosen to be changed from 0 to 1. The *C'* is the chromosome after the mutation.

5.2.6 Termination

The GA process does not stop until a termination condition has been satisfied. For the termination of this problem, there are two methods used. The first one is to set the number of generations. Another one is number of generations that the best chromosome is not changing. In the experiments, both termination conditions are set for stopping the GA process.

5.3 Tributary selection modeling

5.3.1 Geometric factors and objective functions

(1) Drainage pattern membership

In drainage pattern recognition, the membership degree is applied to classify the pattern of a river network. The higher it is, the more characteristic the pattern is. In order to consider the drainage pattern in tributary selection in the first place, the pattern membership can be regarded as an important factor. Before generalization, the pattern of a river network or a sub-network can be identified first. Then, as an objective function, the membership degree can be applied to the generalization according to its pattern.

The membership values of dendritic, parallel, trellis and rectangular patterns were introduced in the previous chapter. Equations (4.5), (4.6), (4.7) and (4.8) (pp. 72-73) are used to calculate the values for dendritic, parallel, trellis and rectangular patterns respectively. The pattern membership is defined to get a degree from 0 to 1 to describe how much a river network belongs to a pattern. Therefore, for different drainage patterns, the objective functions are different. Here, the used MFs in each pattern are going to apply the settings in Table 4.2 (p. 75). The objective function of the drainage pattern membership can be given as follows:

$$\text{maximize} \quad F_M = \begin{cases} \min(z(\alpha;45°,90°), z(\delta;1,3)), & \text{dendritic} \\ \min\begin{pmatrix} z(\alpha;30°,60°), 1-s(\beta;0,1), \\ s(\gamma;0,1), s(\delta;1,3) \end{pmatrix}, & \text{parallel} \\ \min\begin{pmatrix} g(\alpha;10°,90°), 1-s(\beta;0,1), \\ z(\gamma;0,1), s(\delta;1,3) \end{pmatrix}, & \text{trellis} \\ \min(g(\alpha;10°,90°), s(\beta;0,1)), & \text{rectangular} \end{cases} \tag{5.1}$$

where α, β, γ, and δ are the average junction angle, the bended tributaries percentage, the average length ratio and the catchment elongation respectively, and $F_M \in [0,1]$.

(2) Stream order

The stream order is a way to define the size of perennial and recurring streams based on a hierarchy of tributaries. As reviewed in Chapter 2, there are several ordering schemes, and the Horton-Strahler scheme (Strahler, 1957) and the Shreve scheme (Shreve, 1966) are the famous ones. In this chapter, the Horton-Strahler order after upstream routine will be used for tributary selective omission as it can provide a generalized river network close to a decision made by human (Rusak Mazur & Castner, 1990).

In river network generalization, the selective omission operation, in general, starts from the tributaries with small order. The tributaries with large order have more opportunity to be shown in a map after selection omission. So, the objective function of the stream order is designed as

$$\begin{gathered} \text{maximize} \quad F_o = \sum_{i=1}^{n} O_i, \\ 0 < n \le N \end{gathered} \tag{5.2}$$

where F_o is the total order of selected tributaries; O_i is the order of the selected tributary *i*.

(3) Stream length

In digital map, a stream is stored as a set of points, and the length can be calculated approximately by the additive value of all distance of these points.

$$L = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} , \tag{5.3}$$

where L is the length of a stream composed of n points (x_i, y_i) ($1 \le i \le n$).

For the stream length factor, in a certain extent a tributary with longer length implies that this one is more important. In order to select longer tributaries preferentially, the follow objective function goes after the purpose for maximizing the length value of all selected rivers. This objective function F_L is as follows,

$$\begin{aligned} \text{maximize} \quad & F_L = \sum_{i=1}^{n} L_i , \\ & 0 < n \le N \end{aligned} , \tag{5.4}$$

where F_L is the total length of selected tributaries, n is the selected number of tributaries, which should not be bigger than the original number of tributaries N and L_i is the length of selected tributary i.

(4) Balance coefficient

Balance coefficient is the difference between the total length of streams on the left side of the mainstream and the total length on the right side. It shows the uneven degree of a drainage system. The larger the value, the more balanced the water quantities flowing from two sides of the mainstream. The balance coefficient B is calculated as:

$$B = \begin{cases} 1, & \sum_{i=1}^{m} L_i = 0 \text{ and } \sum_{j=1}^{n} L_j = 0 \\ \sum_{i=1}^{m} L_i \Big/ \sum_{j=1}^{n} L_j , & \sum_{i=1}^{m} L_i \le \sum_{j=1}^{n} L_j \\ \sum_{j=1}^{n} L_j \Big/ \sum_{i=1}^{m} L_i , & \sum_{i=1}^{m} L_i > \sum_{j=1}^{n} L_j \end{cases} , \tag{5.5}$$

where m, n are the numbers of the left and right side of the mainstream respectively; L_i is the length of stream i on the left side ($1 \le i \le m$), and L_j is the length of stream j on the right side ($1 \le j \le n$).

From the calculation of the balance coefficient, we can know that $B \in [0,1]$. $B = 1$ is an ideal status that the river is in balance of receiving the water from both side. The objective of balance coefficient is to maintain the balance after generalization. Therefore, for the objective function of balance coefficient, it is defined by the Gaussian function, equation (3.3) (p. 33), as follows,

$$\text{maximize} \quad F_B = \sum_{i=1}^{m} g(B_i'; 0.1, B_i) \Big/ m \,, \tag{5.6}$$

where m is the number of streams with the order > 1 (a stream should have upper streams); B_i is the balance coefficient of stream i before generalization, B'_i is the balance coefficient of stream i after generalization; and $F_B \in [0,1]$. In the Gaussian function, the center is B_i, and the standard deviation is set to 0.1. So, the closer B'_i to the center, the greater the value to 1.

(5) Tributary spacing

Tributary spacing is the distance between two adjacent tributaries which are on the same side of a main stream. It can reflect the local distribution of a river network, and it is an important factor for river networks generalization. As the adjacent tributaries are not parallel in general, the calculation of the distance is complicated. For two polygonal curves, the distance can be given by the Frechet distance (Alt & Godau, 1995). Ai et al. (2006) proposed a weighted distance computation method. Here, the application of the tributary spacing is more relevant to the trellis and parallel pattern, where the tributaries are more or less parallel. The shortest distance between two tributaries is used for tributary spacing. The advantage of using the shortest distance is that it prevents tributaries from being too close when the scale becomes smaller and so is preferred to other distances.

If two polygonal curves A and B are at some distance from each other, for any point a of A and any point b of B, the distance D, which is similarly regarded as the spacing S, between A and B is defined by:

$$S \approx D(A,B) = \min_{a \in A} \left\{ \min_{b \in B} \{ d(a,b) \} \right\} , \tag{5.7}$$

where $d(a, b)$ is the distance between a and b.

As to the objective function of the tributary spacing, it is given as

$$\text{maximize} \quad F_s = \min(S_i), \quad i = 1, 2, \ldots k \quad , \tag{5.8}$$

where k is the number of spacing of tributaries after selection, and S_i is the tributary spacing i. This function is to maximize the smallest spacing between tributaries.

5.3.2 Multi-objective modeling with consideration of drainage pattern

For multi-objective problems, the weighted sum method is the most convenient and simplest approach, which aggregates a number of objective functions into a single one by multiplying each function with a weight value (Deb, 2001). It can be written as (Hajela & Lin, 1992),

$$F(X) = \sum_{i=1}^{k} W_i F_i(X) \quad , \tag{5.9}$$

where k is number of objective functions; W_i is the weight of each objective function F_i, and the weights should satisfy the requirement of $\sum_{i=1}^{k} W_i = 1$. As the magnitude of each objective function might be different, the fitness function scaling should be applied, and the final formula is as follows:

$$\bar{F}(X) = \sum_{i=1}^{k} W_i F_i^*(X) \quad , \tag{5.10}$$

where F_i^* are the scaled objective functions. Usually, the normalization method is used for function scaling, and F_i^* is given by

$$F_i^*(X) = \frac{F_i(X) - F_i^{\min}}{F_i^{\max} - F_i^{\min}} \quad . \tag{5.11}$$

For all objective functions, the multi-objective functions are aggregated for the fitness in the GA process. It is given as follows.

$$F(X) = w_M F_M(X) + w_O F_O^*(X) + w_L F_L^*(X) + w_B F_B(X) + w_S F_S^*(X), \qquad (5.12)$$

where $w_M + w_O + w_L + w_B + w_S = 1$.

In the fitness function, X is a solution for the tributary selection represented as a chromosome. w_M, w_O, w_L, w_B, and w_S are the weight values for the objectives of pattern membership, stream order, stream length, balance coefficient, and tributary spacing respectively. Magnitudes of objective functions are different, so the objective functions are scaled to [0, 1].

5.4 Experiments and results

In the experiment, the approach was implemented in C# language. AForge.NET[8], which is an open source C# framework designed for developers and researchers in the field of computer vision and artificial intelligence, was used to establish the experiment platform. The evolution programming library - "AForge.Genetic" was used to implement the generalization process of the tributary selection. There are several datasets tested in this experiment, such as the Russian river used in Chapter 3, and the NHD of USA in Chapter 4. The Russian river was tested to show the generalized results of river networks after the selection omission by the GA, and the NHD was used to assess the results.

The river flow dataset from the NHD at 1:100K scale is not generalized automatically from a small scale data. It is regarded as an ideal generalized result by manual work. Therefore, the NHD data at 1:100K scale is used as a standard to check generalized river networks from Russian river by comparing the similarity, which is calculated by an overlap ratio. Supposing a river network from the NHD is composed of N river segments and an automatic generalized river network has M river segments overlapped with the NHD data, the overlap ratio is calculated as

[8] http://www.aforgenet.com/

$$Similarity = \left(\sum_{i=1}^{M} Len_i \Big/ \sum_{j=1}^{N} Len_j \right) \times 100\% \quad (5.13)$$

,

where Len_x is the length of a river segment.

In Figure 5.6, supposing network (a) is a generalized network and network (b) is a network from the NHD, the overlapped river segments are shown in bold gray shadow in (c). So, the similarity is the length of segments in shadow divided by the total length of network (b).

(a) (b) (c)

Figure 5.6 An example of overlap

The experiment process is as follows:

(1) Get a sub-network from Russian river, and identify its drainage pattern.
(2) Get the same river network from the NHD, and build strokes to obtain the number of selected strokes.
(3) According to the pattern, set weights for the fitness function.
(4) Get a generalized river network by applying GA.
(5) Calculate the similarity with the river network from the NHD.
(6) Repeat above steps for all sub-networks.

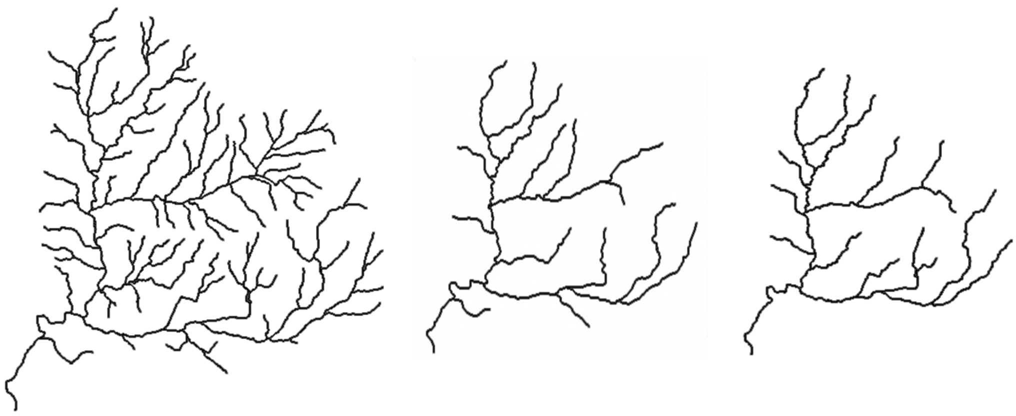
In the experiment, the used parameters are set as follows: the population size is set to 100; for termination, the total number of generation is 500, and the iteration would stop if the best solution does not change during 20 generations. These settings are empirical values. A small population size will easily lead to a local optimal solution, but the algorithm runs faster; otherwise, chromosomes are various due to a large population size, but it needs more time to execute the algorithm. However, an overpopulated setting cannot get a better solution from the experiment process.

There are several objectives to achieve in the experiment: one is to test the availability of each factor, and another is to rank the factors in the order of importance for the multi-objective function according to different drainage patterns. So, different weights are set up to test in the GA process. Two group tests are needed to be done for each drainage pattern. The first one is to set a weight of a factor to 0.6 and others to 0.1, and the second one is to set a weight to 0 and others are 0.25. The importance of a factor can be validated through these tests. Then, according to tested results, other schemes of weights setting can be examined to obtain a feasible setting for a drainage pattern. For each pattern, a case study will be used to show the results of different weight settings, and then these settings are tested in all sub-networks in the Russian river.

5.4.1 Dendritic networks in Russian river

(1) Dendritic case study

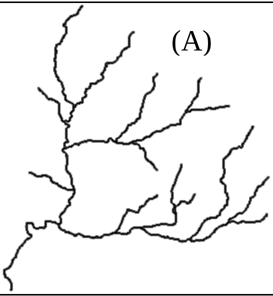
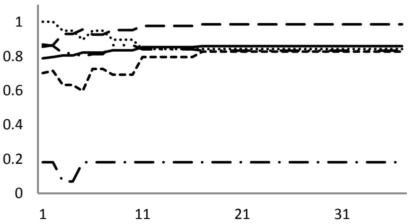
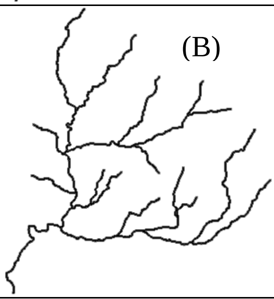
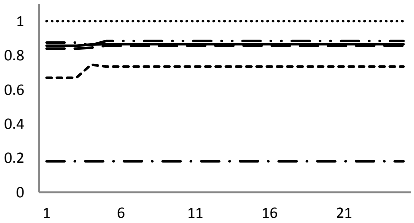
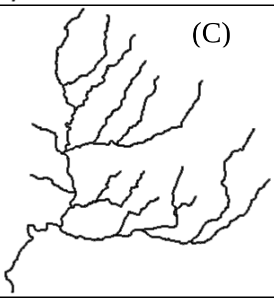
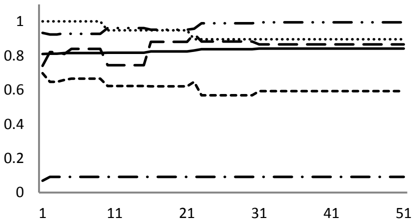
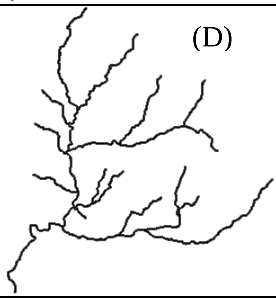
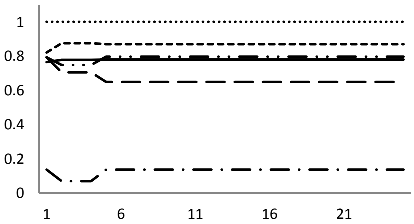
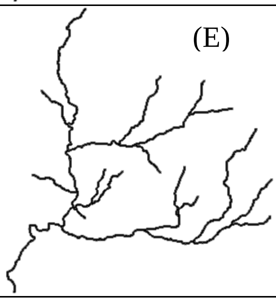
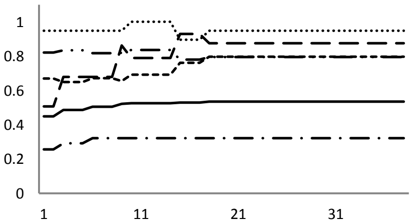
The tested dendritic river network from Russian river at 1:24K scale is shown in Figure 5.7(a), and the network from the NHD at 1:100K scale is illustrated in Figure 5.7(b). The network in Figure 5.7(c) is the generalized result by the method of stroke and length. It has the same number of strokes with the network from the NHD, and the similarity between them is 73.2%. The result of this case study is given in Table 5.2.



(a) 1:24K from Russian river (b) 1:100K from NHD (c) By stroke and Length

Figure 5.7 Tested network for dendritic case

Table 5.2 Generalized results for dendritic case

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0.6/0.1/ 0.1/0.1/0.1)	(A) 		66.2%
(0.1/0.6/ 0.1/0.1/0.1)	(B) 		65.6%
(0.1/0.1/ 0.6/0.1/0.1)	(C) 		81.0%
(0.1/0.1/ 0.1/0.6/0.1)	(D) 		60.7%
(0.1/0.1/ 0.1/0.1/0.6)	(E) 		61.7%

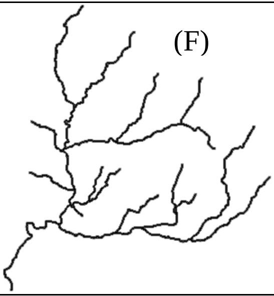
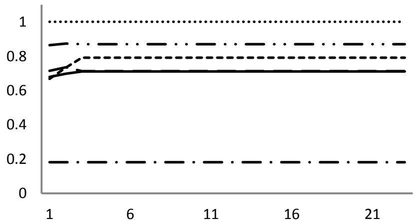
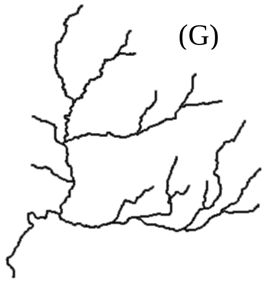
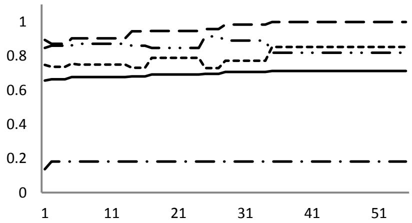
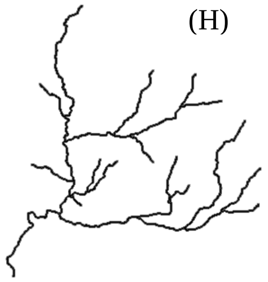
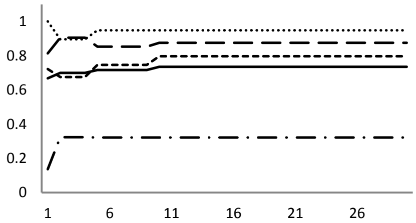
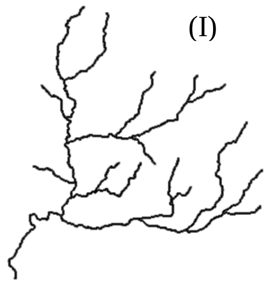
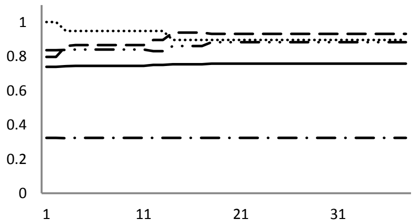
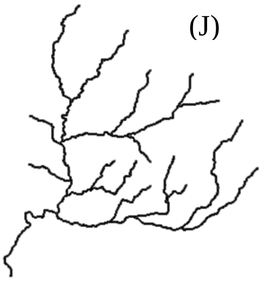
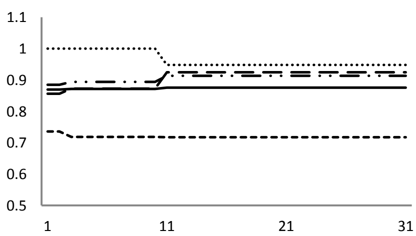
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·—·· Length --------- Balance —·—·— Spacing ··········· Order — — — — Membership

——— Fitness

(*Continued*)

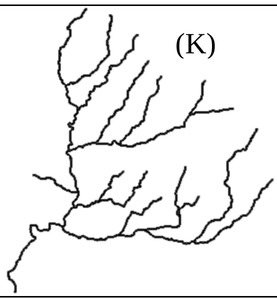
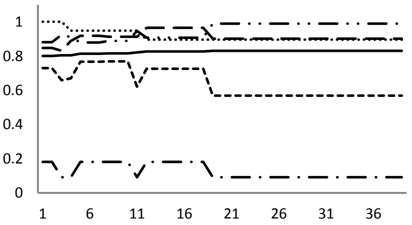
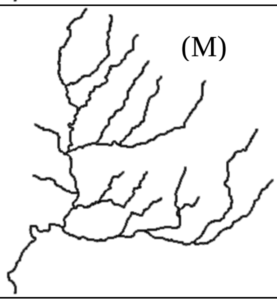
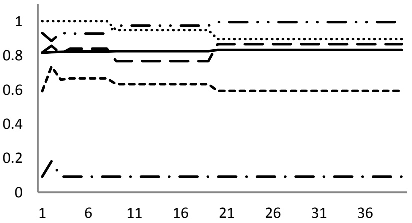
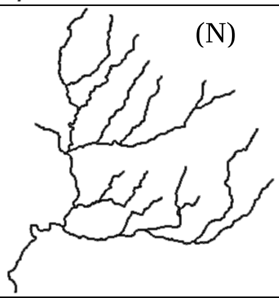
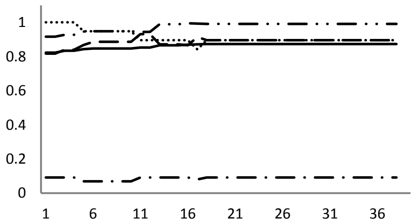
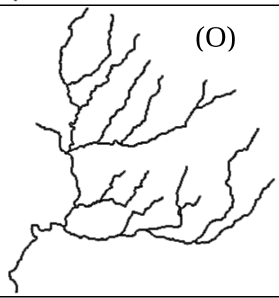
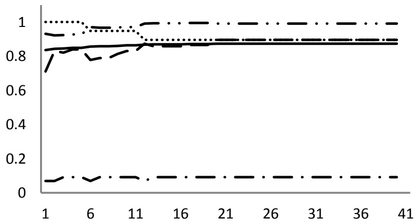
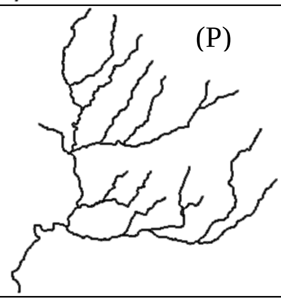
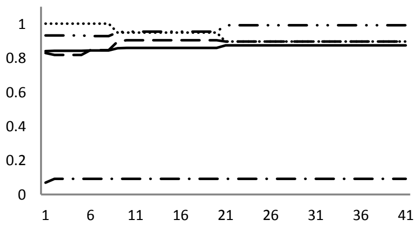
Table 5.2 (*Continued*)

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0/0.25/0.25/ 0.25/0.25)	(F) 		65.6%
(0.25/0/0.25/ 0.25/0.25)	(G) 		60.6%
(0.25/0.25/0/ 0.25/0.25)	(H) 		61.7%
(0.25/0.25/ 0.25/0/0.25)	(I) 		73.9%
(0.25/0.25/ 0.25/0.25/0)	(J) 		69.9%

*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is
—·—·· Length --------- Balance —·—·— Spacing ··········· Order — — — — Membership
——— Fitness

(*Continued*)

Table 5.2 (*Continued*)

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0.2/0.1/ 0.5/0.1/0.1)	 (K)		81.0%
(0.1/0.2/ 0.5/0.1/0.1)	 (M)		81.0%
(0.2/0.1/ 0.6/0/0.1)	 (N)		83.5%
(0.1/0.2/ 0.6/0/0.1)	 (O)		83.5%
(0.15/0.15/ 0.6/0/0.1)	 (P)		83.5%

*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—··—·· Length --------- Balance —·—·— Spacing ·············· Order — — — — Membership

——— Fitness

In Table 5.2, the weights settings are shown in the first column. In the table, the column of the GA process records the value of fitness function and all participated objective functions at each generation during the process. The value of stream order function is decreasing, and others are increasing. This is because a selection solution is initialized based on stream order. The lower order streams are chosen to be eliminated first, and then, after the GA operations, some streams with the lower order would be resumed to be selected due to the influence of other factors. The pattern membership value is increasing which can guarantee the pattern is preserved during the process.

From Table 5.2, the similarities of generalized networks (A) to (E) are 66.2%, 65.6%, 81.0%, 60.7% and 61.7% respectively. Network (C) is the result of setting the weight of the length factor with 0.6, and it has the greatest similarity among the group tests, which is also bigger than the similarity of the generalized network by the method of the stroke and length (81.0% > 73.2%). The similarities with the manual generalized network at 1:100K scale of other networks are fairly low. Then, we can see, the length is an important factor to a dendritic pattern. After the second group test, without considering the length, network (H) is generalized by setting $w_L = 0$, and the similarity decreases to 61.7% from 81.0%. For other factors, without the membership or the order, similarities of networks (F) and (G) also decrease from 66.2% to 65.6% and from 65.6% to 60.6% respectively. However, without factors of the balance coefficient or the tributary spacing, the increased similarities indicate that these two factors are not so important to the dendritic pattern. From the result, the preliminary rank of the factors is $w_L > w_M$, $w_O > w_S$, w_B, the length is definitely the most important one.

In the following tests, w_L is set as a high value to 0.5 or 0.6, and other factors are given by different values to fix the importance between w_M and w_O, and between w_S and w_B. From networks (K) and (M) in Table 5.2, although w_M and w_O are different similarities are the same. It is hard to say which factor has priority between w_M and w_O. However, with these two weight settings, similarities are still not greater than network (C). For the reason, considering w_B has influenced the similarity a lot, it is set as 0 to get networks (N), (O) and (P). The results show that $w_S > w_B$ because all similarities of (N), (O) and (P) have been improved. From this case study, the importance between w_M and w_O cannot be told. It should be fixed by testing in all dendritic river networks in the Russian river.

(2) Generalized dendritic networks

The setting of weights for the fitness function used in case study is also tested on all dendritic river networks in the Russian river. The statistic result is listed in Table 5.3.

Table 5.3 Generalized dendritic networks results

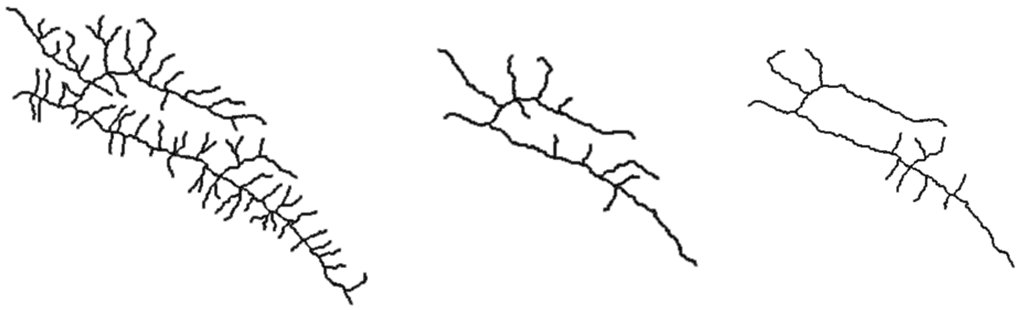
Weights Setting (w_M/w_O/w_L/w_B/w_S)	Average Similarity				Average Membership			
	Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
(0.6/0.1/0.1/0.1/0.1)	71.8%	76.2%	70.4%	72.6%	0.50	0.71	0.59	0.55
(0.1/0.6/0.1/0.1/0.1)	74.7%	82.5%	74.0%	76.1%	0.33	0.56	0.42	0.38
(0.1/0.1/0.6/0.1/0.1)	77.4%	87.3%	81.1%	79.4%	0.32	0.53	0.48	0.37
(0.1/0.1/0.1/0.6/0.1)	70.9%	70.1%	63.3%	70.4%	0.28	0.47	0.44	0.33
(0.1/0.1/0.1/0.1/0.6)	73.2%	77.5%	70.3%	73.9%	0.34	0.64	0.53	0.41
(0/0.25/0.25/0.25/0.25)	73.9%	82.5%	72.8%	75.4%	0.27	0.49	0.19	0.31
(0.25/0/0.25/0.25/0.25)	73.4%	78.9%	69.6%	74.3%	0.40	0.68	0.63	0.46
(0.25/0.25/0/0.25/0.25)	69.9%	72.0%	64.5%	70.0%	0.37	0.66	0.50	0.43
(0.25/0.25/0.25/0/0.25)	74.4%	81.6%	77.3%	75.9%	0.39	0.67	0.52	0.44
(0.25/0.25/0.25/0.25/0)	75.4%	80.8%	73.2%	76.3%	0.36	0.64	0.55	0.42
(0.2/0.1/0.5/0.1/0.1)	78.1%	86.2%	80.0%	79.7%	0.35	0.61	0.53	0.40
(0.1/0.2/0.5/0.1/0.1)	78.0%	87.5%	79.6%	79.8%	0.32	0.54	0.41	0.36
(0.2/0.1/0.6/0/0.1)	78.2%	86.6%	83.1%	79.9%	0.35	0.60	0.56	0.41
(0.1/0.2/0.6/0/0.1)	78.3%	87.8%	82.7%	80.3%	0.32	0.52	0.47	0.36
(0.15/0.15/0.6/0/0.1)	78.4%	88.4%	83.0%	80.4%	0.33	0.53	0.48	0.37
Stroke + Length	77.5%	86.4%	78.4%	79.2%	0.25	0.48	0.28	0.29
River networks at 1:24K scale from the Russian river					0.46	0.53	0.38	0.47
River networks at 1:100K scale from the NHD					0.28	0.43	0.27	0.31

From Table 5.3, comparing the first five and the second five tests, it shows the same conclusions with the dendritic case study except for the factor of drainage pattern membership. Without considering the drainage pattern, the average similarity increases from 72.6% to 75.4%. The similarity is computed by comparing with the manual work, so it can illustrate that the manual river networks are not generalized with consideration of the dendritic pattern to some extent. From the results, the similarities can be improved by setting w_L with high values. It shows that the length is the most important factor among all five proposed factors. Weight settings for w_M, w_O, w_L, w_B and w_S with 0.15, 0.15, 0.6, 0 and 0.1 get the greatest average similarity (80.4%) among all settings. It confirms that the settings are more appropriate for dendritic river networks. For the pattern membership value, the average membership of all dendritic networks before generalization at 1:24K scale is 0.47. After generalization, although the average membership of networks generalized by the setting of (0.15/0.15/0.6/0/0.1) only is 0.37, which is smaller than 0.47, it is greater than the value of manual generalized river networks (0.31). Only the average membership of generalized networks in order 2 is smaller than the original one (0.33 < 0.46) because many generalized networks in order 2 do not have enough tributaries for the computation of the drainage pattern membership. So, with consideration of the drainage pattern, the drainage pattern membership value is even increased after generalization. In addition, sub-networks in order 3 and 4 have higher average similarities than in order 2, so do average memberships. It illustrates that, to some extent, networks in lower order do not have enough tributaries to calculate each factor value in the fitness function, and there is no need to apply a complicated generalization method to a network with few tributaries.

5.4.2 Trellis networks in Russian river

(1) Trellis case study

Figure 5.8(a) shows the tested trellis river network from the Russian river at 1:24K scale. Figure 5.8(b) is the network from the NHD at 1:100K scale, and Figure 5.8(c) illustrates the generalized network by stroke and length, which has a similarity of 74.3% with the network in (b).



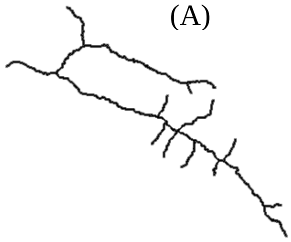
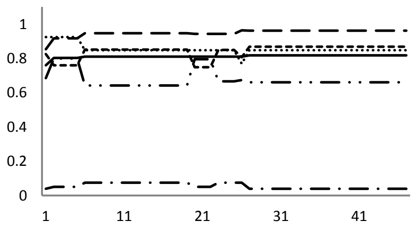
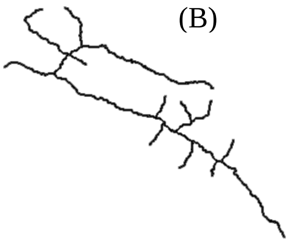
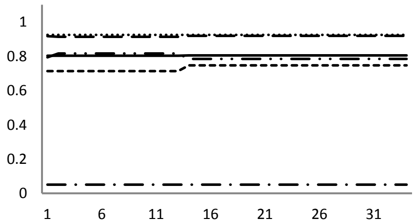
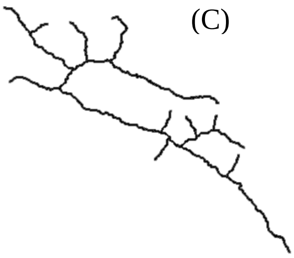
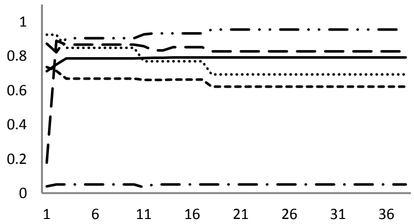
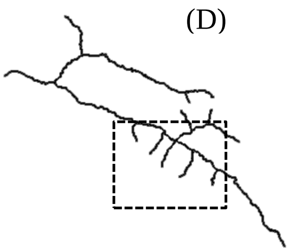
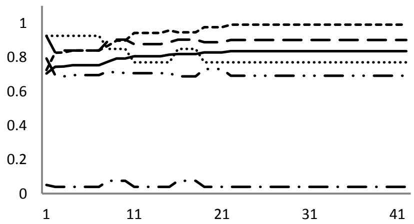
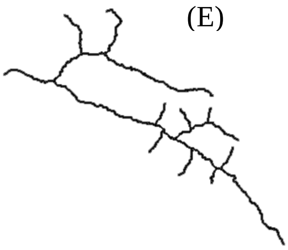
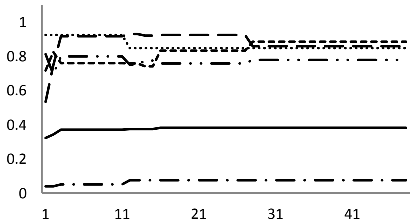
(a) 1:24K from Russian river (b) 1:100K from NHD (c) By stroke and Length

Figure 5.8 Tested network for trellis case

In Table 5.4, from the first ten tests, the similarity varies between 84.5% (network C) to 68.3% (network H). Without considering the factor of length, the generalized network is far from the manual one. Length is again the most important factor in the generalization. The same situation happens on factors of the order and the tributary spacing, so these two factors are also considerable in the process. Drainage pattern and balance coefficient have limited consideration from the case of the trellis network. However, they do influence the generalization process when they are highly participated. The membership trend line is almost close to 1 after almost 25 generations for network (A). In network (D), the tributaries of the right side of the main stream (in the dashed box) are preserved to balance the network, and in the GA process, the trend line of the balance coefficient is almost close to 1 after 20 generations. From these two group tests, the ranking of the five factors by importance is $w_L > w_O$, $w_S > w_M$, w_B according to changes in similarities.

In order to consider the drainage pattern, w_M is assigned with a low value (0.1 or 0.2) to take part in the GA process in the following tests. In Table 5.4, among generalized networks (K), (M) and (N), network (N) has the greatest similarity (83.8%) which is generalized by assigning w_S bigger than other weights except for w_M. So, tributary spacing is better considered in the trellis pattern than other factors. The similarity of network (K) is larger than (M) (81.3% > 80.1%). It indicates that the pattern membership may be more considerable than the factor of order. In addition, the factor balance possibly has no contribution to improve the similarity. Therefore, w_M, w_O, w_L, w_B and w_S are assigned with 0.2, 0.1, 0.5, 0 and 0.2 respectively to try to prove the hypothesis. The similarity of network (O) is 81.6% which is smaller than 83.8% of network (N). This setting is not a suitable one to get a higher similarity network with the manual one. Considering that the similarity of network (C) is 84.5%, the length factor should be assigned with a higher value ($w_L > 0.5$). Network (P) confirms that assigning 0.6 to w_L is preferable. Weight settings of 0.1, 0.1, 0.6, 0 and 0.2 to w_M, w_O, w_L, w_B and w_S respectively are suitable.

Table 5.4 Generalized results for trellis case

Weights (w_M/w_O/ w_L/w_B/w_S)	Generalized network	GA process*	Similarity
(0.6/0.1/ 0.1/0.1/0.1)			68.3%
(0.1/0.6/ 0.1/0.1/0.1)			74.3%
(0.1/0.1/ 0.6/0.1/0.1)			84.5%
(0.1/0.1/ 0.1/0.6/0.1)			69.6%
(0.1/0.1/ 0.1/0.1/0.6)			77.9%

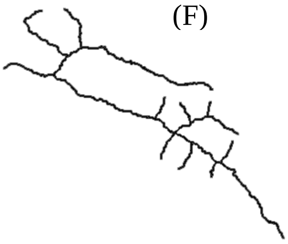
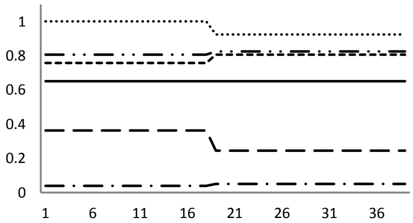
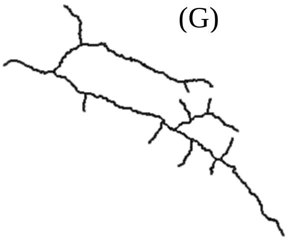
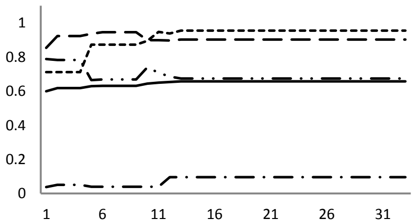
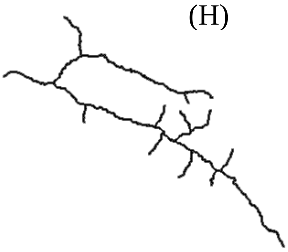
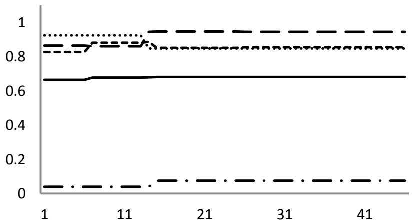
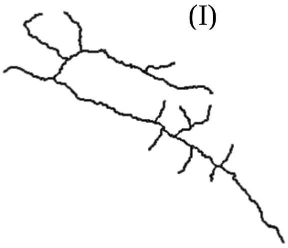
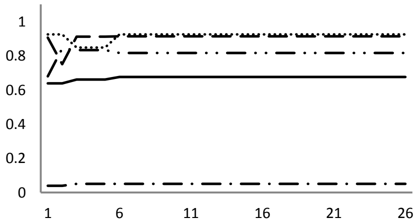
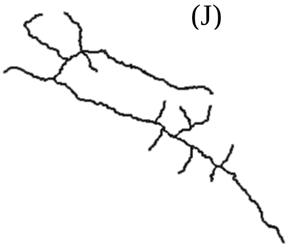
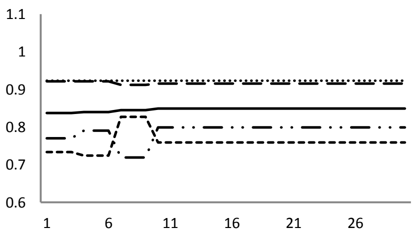
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·—·· Length --------- Balance —·—·— Spacing ········ Order — — — — Membership

——— Fitness

(*Continued*)

Table 5.4 (*Continued*)

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0/0.25/0.25/ 0.25/0.25)	(F) 		78.0%
(0.25/0/0.25/ 0.25/0.25)	(G) 		69.6%
(0.25/0.25/0/ 0.25/0.25)	(H) 		68.3%
(0.25/0.25/ 0.25/0/0.25)	(I) 		74.3%
(0.25/0.25/ 0.25/0.25/0)	(J) 		76.9%

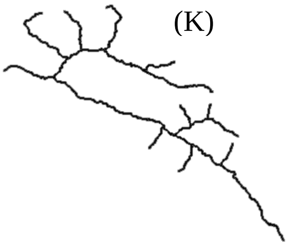
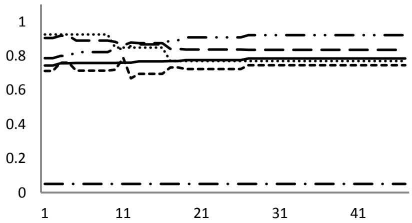
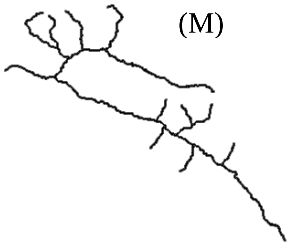
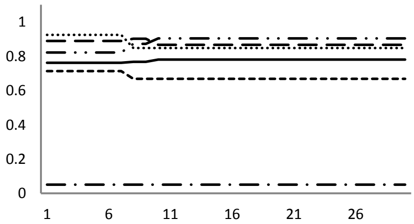
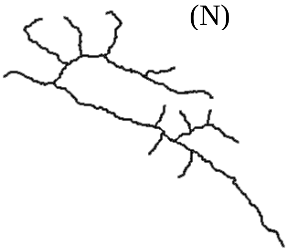
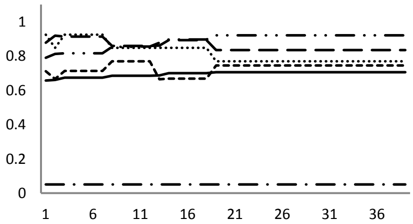
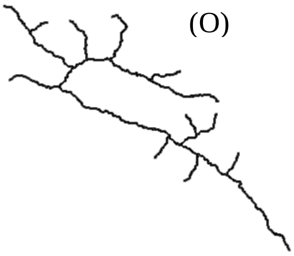
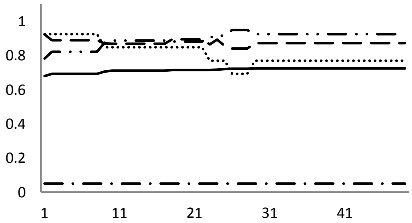
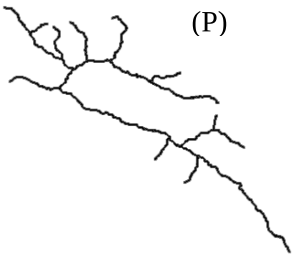
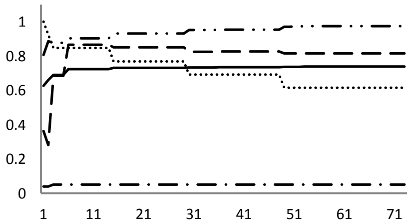
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—··— Length --------- Balance —·—·— Spacing ·········· Order — — — — Membership

——— Fitness

(*Continued*)

Table 5.4 (*Continued*)

Weights (w_M/w_O/ w_L/w_B/w_S)	Generalized network	GA process*	Similarity
(0.2/0.1/ 0.5/0.1/0.1)	(K) 		81.3%
(0.1/0.2/ 0.5/0.1/0.1)	(M) 		80.1%
(0.1/0.1/ 0.5/0.1/0.2)	(N) 		83.8%
(0.2/0.1/ 0.5/0/0.2)	(O) 		81.6%
(0.1/0.1/ 0.6/0.0/0.2)	(P) 		85.3%

*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·—·· Length --------- Balance —·—·— Spacing ·········· Order — — — — Membership

——— Fitness

(2) Generalized trellis networks

For all trellis river networks in Russian river, the statistic results which are tested by the weight settings used in case study are listed in Table 5.5.

Table 5.5 Generalized trellis networks results

Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity				Average Membership			
	Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
(0.6/0.1/0.1/0.1/0.1)	62.9%	73.0%	67.7%	65.7%	0.74	0.76	0.96	0.76
(0.1/0.6/0.1/0.1/0.1)	74.3%	82.2%	76.5%	76.4%	0.32	0.42	0.93	0.37
(0.1/0.1/0.6/0.1/0.1)	79.7%	84.8%	81.2%	81.0%	0.28	0.49	0.89	0.36
(0.1/0.1/0.1/0.6/0.1)	64.3%	64.6%	67.9%	64.5%	0.35	0.65	0.90	0.45
(0.1/0.1/0.1/0.1/0.6)	73.7%	79.7%	73.2%	75.1%	0.29	0.66	0.91	0.41
(0/0.25/0.25/0.25/0.25)	74.9%	80.3%	76.7%	76.3%	0.24	0.33	0.86	0.29
(0.25/0/0.25/0.25/0.25)	73.3%	79.4%	70.7%	74.7%	0.38	0.68	0.95	0.49
(0.25/0.25/0/0.25/0.25)	66.6%	76.2%	66.9%	69.0%	0.45	0.72	0.92	0.54
(0.25/0.25/0.25/0/0.25)	75.9%	80.7%	77.6%	77.2%	0.33	0.66	0.93	0.44
(0.25/0.25/0.25/0.25/0)	75.0%	78.6%	75.0%	75.9%	0.35	0.67	0.93	0.46
(0.2/0.1/0.5/0.1/0.1)	79.1%	84.3%	80.6%	80.5%	0.29	0.66	0.92	0.41
(0.1/0.2/0.5/0.1/0.1)	78.5%	83.1%	80.9%	79.8%	0.27	0.66	0.90	0.40
(0.1/0.1/0.5/0.1/0.2)	80.3%	83.1%	82.8%	81.1%	0.28	0.58	0.89	0.39
(0.2/0.1/0.5/0/0.2)	77.3%	85.9%	79.0%	79.5%	0.32	0.65	0.92	0.43
(0.1/0.1/0.6/0.0/0.2)	79.7%	89.2%	82.2%	82.2%	0.26	0.52	0.87	0.35
Stroke + Length	78.0%	82.2%	84.0%	79.4%	0.20	0.40	0.86	0.28
River networks at 1:24K scale from the Russian river					0.21	0.18	0.03	0.19
River networks at 1:100K scale from the NHD					0.12	0.15	0.002	0.12

From Table 5.5, statistical results support the conclusions in the case study. Without considering the length, the average similarity reduces from 81.0% to 69.0% (the decreasing amplitude is 14.8 points). The length factor influences the similarity a lot. The first test shows that pattern membership has been raised after generalization by assigning w_M with 0.6 and others with 0.1. Although the average similarity is low, the pattern has been preserved a lot. So, it is useful to take the factor of membership into consideration to preserve or even improve the drainage pattern during the generalization process. From the last five tests, the average similarity is bigger than the method of the stroke and length. The multi-objective method is better than the stroke and length method especially in the preservation of the drainage pattern. The weight settings for w_M, w_O, w_L, w_B and w_S with 0.1, 0.1, 0.6, 0 and 0.2 respectively are the most proper settings from the statistic results. In addition, generalized networks in order 3 have better similarity and membership than others. The order 3 can be recommended for working.

5.4.3 Parallel networks in Russian river

(1) Parallel case study

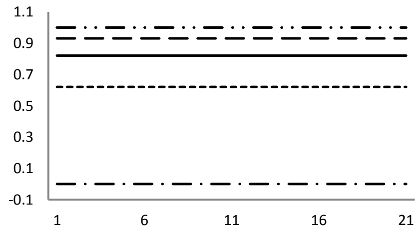
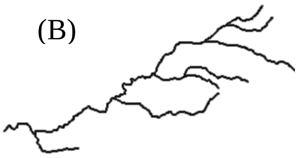
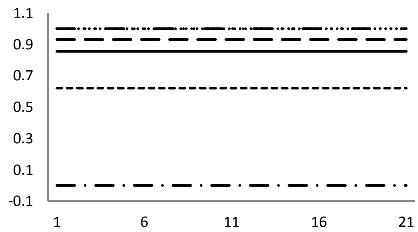
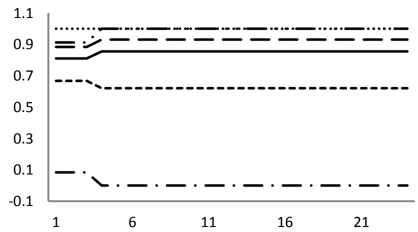
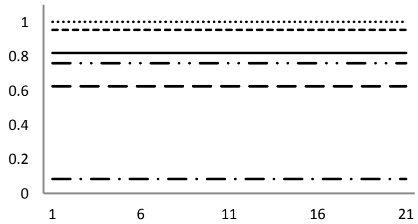
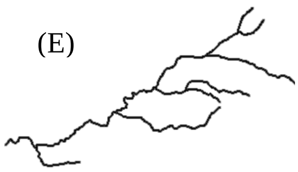
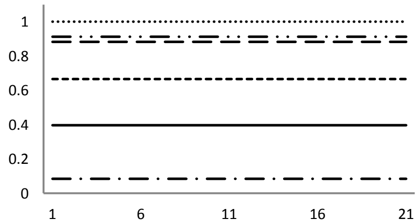
From the experiment results in Chapter 3, most of the parallel networks are in order 2 and few in order 3. The number of tributaries in the parallel networks is small, so the population size is changed to 50 to accelerate the algorithm. The tested parallel river network from Russian river at 1:24K scale is shown in Figure 5.9(a), and the network from the NHD at 1:100K scale is illustrated in Figure 5.9(b). Figure 5.9(c) shows the generalized network by stroke and length, and its similarity with (b) is 100%.

(a) 1:24K from Russian river (b) 1:100K from NHD (c) By stroke and Length

Figure 5.9 Tested network for parallel case

Table 5.6 shows the results for the case study of the parallel pattern.

Table 5.6 Generalized results for parallel case

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0.6/0.1/ 0.1/0.1/0.1)	(A) 		100.0%
(0.1/0.6/ 0.1/0.1/0.1)	(B) 		100.0%
(0.1/0.1/ 0.6/0.1/0.1)	(C) 		100.0%
(0.1/0.1/ 0.1/0.6/0.1)	(D) 		87.0%
(0.1/0.1/ 0.1/0.1/0.6)	(E) 		93.7%

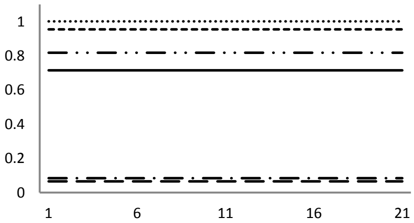
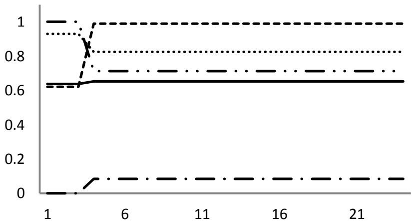
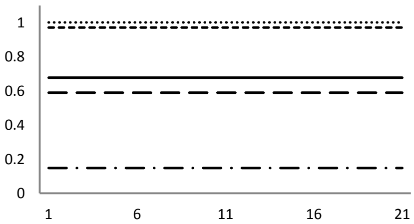
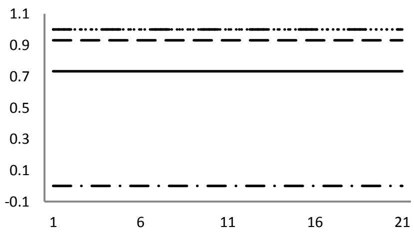
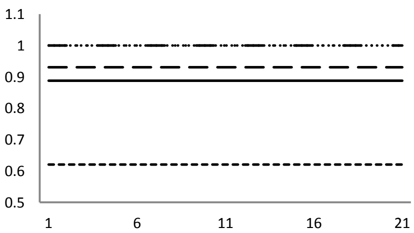
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—··—·· Length --------- Balance —·—·— Spacing ·········· Order — — — — Membership

——— Fitness

(*Continued*)

Table 5.6 (*Continued*)

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0/0.25/0.25/ 0.25/0.25)	(F) 		88.8%
(0.25/0/0.25/ 0.25/0.25)	(G) 		87.5%
(0.25/0.25/0/ 0.25/0.25)	(H) 		72.2%
(0.25/0.25/ 0.25/0/0.25)	(I) 		100.0%
(0.25/0.25/ 0.25/0.25/0)	(J) 		100.0%

*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·–·· Length --------- Balance —·—·— Spacing ·············· Order — — — — Membership

——— Fitness

In Table 5.6, similarities of many generalized network are 100.0% due to the limited number of stokes in the network at 1:24K scale. If there are not enough tributaries in a river network, the result would be not so robust to draw a conclusion. Even so, from the tests, without considering the factor of length, network (H) has the lowest similarity of 72.2% among all tests. The similarity of network (C) is 100%, which assigns the length (w_L) with a high weight (0.6). Then, the factor of length is important during the generalization. Comparing networks (A) and (F), similarities reduce from 100% to 88.8%. From the tests, the order factor also should be considered. From the GA processes, we can see, a generalized network can reach the similarity of 100% if value of objective functions of the length and order were 1. The pattern membership is also important because it can preserve the network with the characteristics of the parallel pattern such as very acute junction angles and a long elongation. For factors of the balance coefficient and the tributary spacing, they are not as important as other factors due to the increased similarities without the consideration of them. So, from the tests in the table, the preliminary rank of the factor importance is: $w_L > w_M$, $w_O > w_B$, w_S.

The number of strokes in the network is only 10. It is easily generalized as the same as the manual network at 1:100K scale so that it cannot tell the different importance of all factors. Therefore, the following tests are run on all parallel networks in the Russian river.

(2) Generalized parallel networks

In the experiment, there are no parallel networks in order 4, and most networks are in order 2. The statistic result is listed inTable 5.7.

Table 5.7 Generalized parallel networks results

Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity			Average Membership		
	Order 2	Order 3	Total	Order 2	Order 3	Total
(0.6/0.1/0.1/0.1/0.1)	61.5%	82.3%	63.2%	0.54	0.90	0.57
(0.1/0.6/0.1/0.1/0.1)	62.9%	79.4%	64.3%	0.49	0.69	0.51
(0.1/0.1/0.6/0.1/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
(0.1/0.1/0.1/0.6/0.1)	63.6%	85.1%	65.4%	0.48	0.58	0.49
(0.1/0.1/0.1/0.1/0.6)	61.8%	76.3%	63.0%	0.48	0.67	0.49
(0/0.25/0.25/0.25/0.25)	60.0%	73.8%	61.2%	0.45	0.26	0.43
(0.25/0/0.25/0.25/0.25)	62.9%	73.2%	63.8%	0.49	0.64	0.50
(0.25/0.25/0/0.25/0.25)	60.2%	60.5%	60.2%	0.51	0.52	0.51
(0.25/0.25/0.25/0/0.25)	62.9%	79.4%	64.3%	0.49	0.69	0.51
(0.25/0.25/0.25/0.25/0)	63.8%	91.6%	66.1%	0.49	0.73	0.51
(0.2/0.1/0.5/0.1/0.1)	64.7%	91.6%	66.9%	0.47	0.73	0.50
(0.1/0.2/0.5/0.1/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
(0.1/0.1/0.5/0.2/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
(0.1/0.1/0.5/0.1/0.2)	62.2%	91.6%	64.6%	0.45	0.73	0.47
(0.3/0.1/0.5/0.1/0)	64.0%	91.6%	66.3%	0.48	0.73	0.50
Stroke + Length	62.9%	91.6%	65.3%	0.44	0.73	0.47
River networks at 1:24K scale from the Russian river				0.56	0.66	0.56
River networks at 1:100K scale from the NHD				0.64	0.46	0.63

In Table 5.7, from the first ten tests, it corresponds to the case study except for the balance coefficient. Comparing with the result by assigning w_B with 0.6, the average similarity reduces to 64.3% from 65.4% by assigning w_B with 0. It shows that the factor of balance has contribution to the similarity. The length is still the most important factor during the generalization as the average similarities would decrease (65.3% to 60.2%) if the length is not considered. By assigning w_L with 0.5 and adjusting weights of other factors, the pattern membership should be set up as the second important factor. Because the average similarity is the greatest by assigning w_M = 0.2, w_O = 0.1, w_L = 0.5, w_B = 0.1 and w_S = 0.1. It is a suitable setting for weights to get a higher similarity with manual generalized networks. The average similarity is not larger than 70% due to the influence of low similarities of networks in order 2. The average similarity of generalized networks in order 3 is 91.6%. The same situation happens in the average membership in order 3.

5.4.4 Rectangular networks in Russian river

(1) Rectangular case study

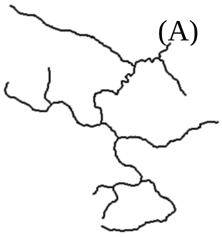
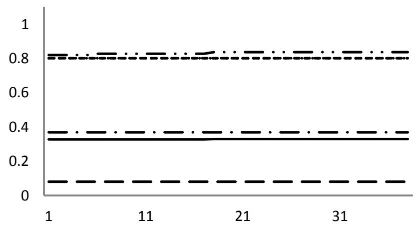
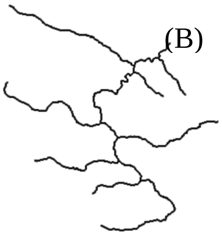
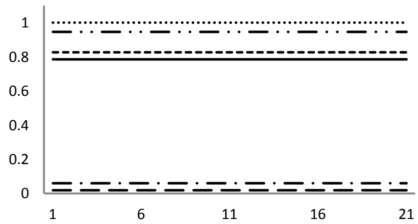
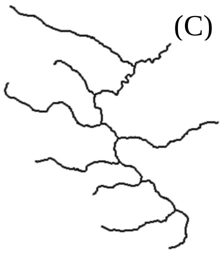
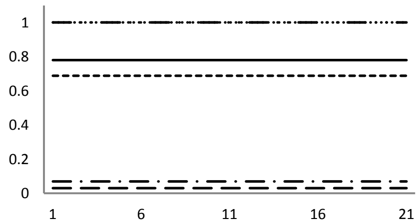
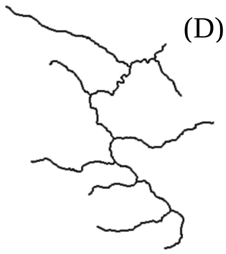
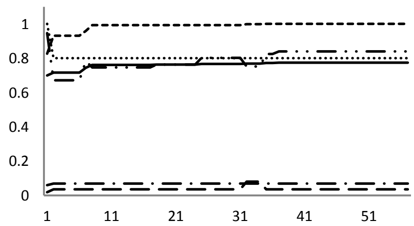
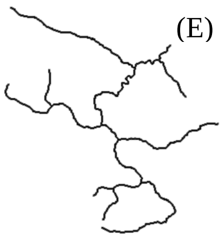
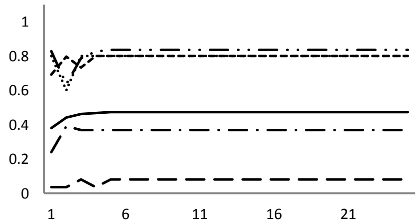
There are few rectangular river networks in the Russian river. The tested rectangular river network from Russian river at 1:24K scale is shown in Figure 5.10(a), and the network from the NHD at 1:100K scale is illustrated in Figure 5.10 (b). Figure 5.10(c) illustrates the network generalized by stroke and length, and its similarity is 93.6% compared with the network from the NHD.

(a) 1:24K from Russian river (b) 1:100K from NHD (c) By stroke and Length

Figure 5.10 Tested network for rectangular case

Table 5.8 shows the generalized results for the rectangular case. The membership of the network in Figure 5.10(a) is only 0.003. The characteristic of the rectangular pattern is not so obvious. From the result, the membership value increases by assigning w_M with values. Network (A) in the table looks more like a rectangular river network. However, the similarity is only 75.0% which is smaller than the network generalized by stroke and length (93.6%). So, during the generalization by human work, the order and length are more considered.

Table 5.8 Generalized results for rectangular case

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0.6/0.1/ 0.1/0.1/0.1)	(A) 		75.0%
(0.1/0.6/ 0.1/0.1/0.1)	(B) 		82.6%
(0.1/0.1/ 0.6/0.1/0.1)	(C) 		93.6%
(0.1/0.1/ 0.1/0.6/0.1)	(D) 		81.2%
(0.1/0.1/ 0.1/0.1/0.6)	(E) 		75.0%

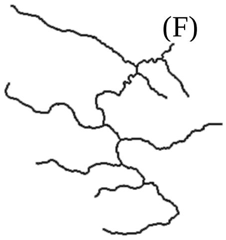
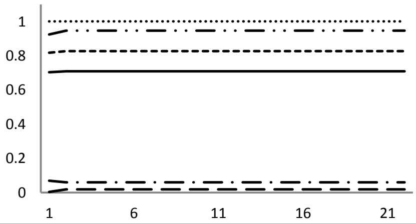
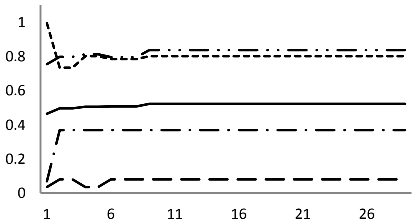
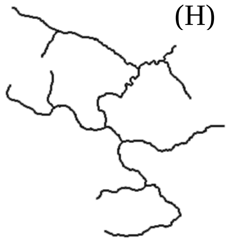
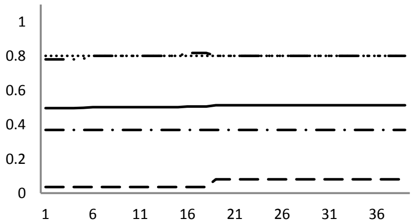
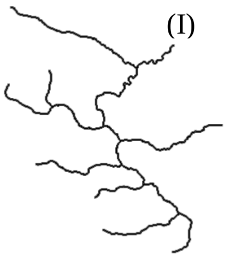
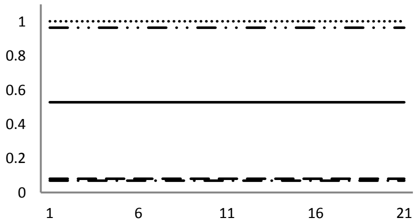
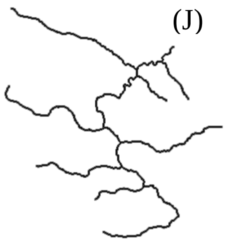
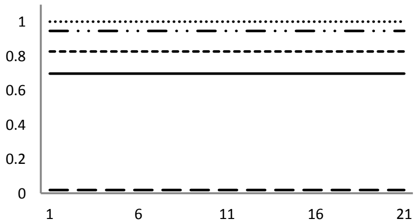
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·—·· Length --------- Balance —·—·— Spacing ··········· Order — — — — Membership

——— Fitness

(*Continued*)

Table 5.8 (*Continued*)

Weights (w_M/w_O/ w_L/w_B/w_S)	Generalized network	GA process*	Similarity
(0/0.25/0.25/ 0.25/0.25)	(F) 		82.6%
(0.25/0/0.25/ 0.25/0.25)	(G) 		75.0%
(0.25/0.25/0/ 0.25/0.25)	(H) 		75.0%
(0.25/0.25/ 0.25/0/0.25)	(I) 		88.0%
(0.25/0.25/ 0.25/0.25/0)	(J) 		82.6%

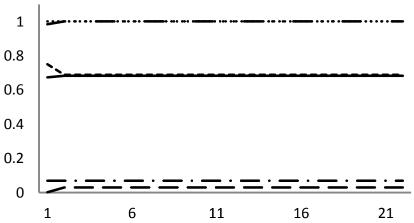
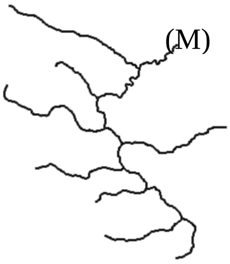
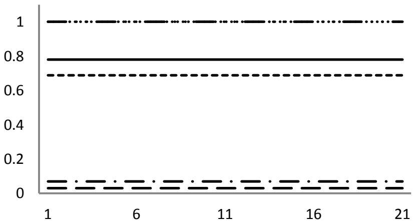
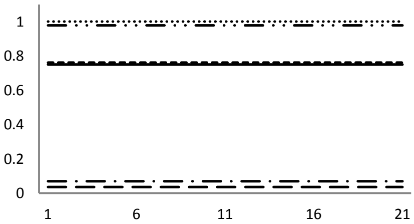
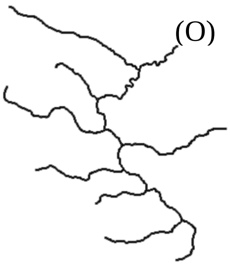
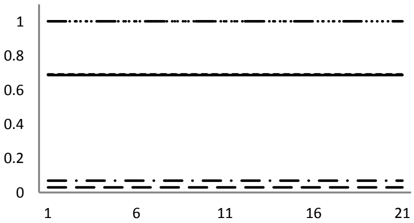
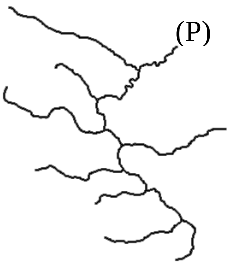
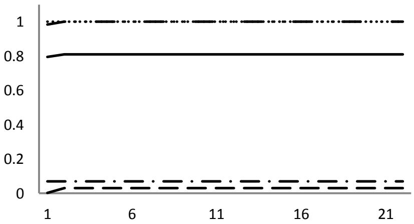
*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—··— Length --------- Balance —·—·— Spacing ··········· Order — — — — Membership

——— Fitness

(*Continued*)

Table 5.8 (*Continued*)

Weights ($w_M/w_O/$ $w_L/w_B/w_S$)	Generalized network	GA process*	Similarity
(0.2/0.1/ 0.5/0.1/0.1)			93.6%
(0.1/0.2/ 0.5/0.1/0.1)			93.6%
(0.1/0.1/ 0.5/0.2/0.1)			88.0%
(0.1/0.1/ 0.5/0.1/0.2)			93.6%
(0.1/0.1/ 0.7/0/0.1)			93.6%

*The vertical axis is the value of theobjective function; the horizontal axis is the number of iterations. The legend is

—·· — ·· Length --------- Balance —·—·— Spacing ·········· Order — — — — Membership

——— Fitness

From Table 5.8, length and order factors contribute to the similarity a lot because similarities of network (G) and (H) decreased (82.6% to 75.0% and 93.6% to 75.0%) by setting w_O = 0 and w_L = 0 respectively. Moreover, similarities increased without considering other factors. So length and order are important to generalize a rectangular network as well as manually. However, the drainage pattern is not considered during the generalization. Assigning w_M = 0.2, network (K) still gets a similarity of 93.6%. So, the factor of membership can help to preserve the drainage pattern and do not reduce the value of similarity. From generalized network (K) to (P), similarities are the same and so we cannot rank all factors. From the case study, the preliminary importance rank of factors is: w_O, w_L > w_M, w_S > w_B.

(2) Generalized rectangular networks

Setting of weights is also tested on all rectangular river networks in Russian river, although only seven rectangular networks were characterized which may be too few to draw a robust conclusion. Results are listed in Table 5.9. For the average membership, all weight settings are acceptable because they are all bigger than the original average membership. The problem here is that the original membership value is too low to classify a network as the rectangular pattern. Whatever, from the results, the highest average similarity is 88.8% by setting w_M = 0.1, w_O = 0.1, w_L = 0.6, w_B = 0.1 and w_S = 0.1. It is larger than the average similarity (86.7%) of network generalized by stroke and length. Although the similarity is not improved much, it also indicates that the generalized river networks are more like manually generalized networks by considering factors of the membership, the length, the order, the balance and the tributary spacing.

Table 5.9 Generalized rectangular networks results

Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity				Average Membership			
	Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
(0.6/0.1/0.1/0.1/0.1)	87.9%	83.9%	64.9%	83.5%	0.38	0.11	0.27	0.29
(0.1/0.6/0.1/0.1/0.1)	88.5%	78.7%	71.1%	83.2%	0.36	0.04	0.09	0.23
(0.1/0.1/0.6/0.1/0.1)	91.8%	88.3%	77.5%	88.8%	0.32	0.02	0.05	0.20
(0.1/0.1/0.1/0.6/0.1)	88.5%	78.0%	61.8%	81.7%	0.36	0.03	0.09	0.23
(0.1/0.1/0.1/0.1/0.6)	86.0%	71.9%	58.2%	78.0%	0.34	0.04	0.05	0.21
(0/0.25/0.25/0.25/0.25)	90.0%	78.7%	71.1%	84.0%	0.32	0.04	0.09	0.21
(0.25/0/0.25/0.25/0.25)	88.5%	71.9%	67.1%	80.7%	0.36	0.05	0.27	0.26
(0.25/0.25/0/0.25/0.25)	86.8%	74.9%	58.0%	79.3%	0.38	0.03	0.14	0.25
(0.25/0.25/0.25/0/0.25)	88.3%	78.4%	71.1%	83.0%	0.38	0.04	0.05	0.24
(0.25/0.25/0.25/0.25/0)	88.5%	78.7%	71.1%	83.2%	0.36	0.04	0.20	0.25
(0.2/0.1/0.5/0.1/0.1)	86.3%	91.2%	83.3%	87.2%	0.32	0.09	0.22	0.24
(0.1/0.2/0.5/0.1/0.1)	91.8%	88.3%	71.1%	87.9%	0.32	0.02	0.09	0.20
(0.1/0.1/0.5/0.2/0.1)	90.0%	85.5%	72.0%	86.1%	0.34	0.03	0.09	0.21
(0.1/0.1/0.5/0.1/0.2)	91.8%	84.2%	72.0%	86.8%	0.32	0.05	0.09	0.21
(0.1/0.1/0.7/0/0.1)	91.8%	81.2%	78.9%	87.0%	0.32	0.05	0.05	0.21
Stroke + Length	91.8%	84.2%	71.1%	86.7%	0.32	0.05	0.09	0.21
River networks at 1:24K scale from the Russian river					0.06	0.004	0.01	0.03
River networks at 1:100K scale from the NHD					0.01	0.15	0.14	0.07

5.5 Conclusion

In this chapter, we introduced a new genetic algorithm for river selection where the objective function includes different factors weighted according to their importance. Five factors corresponding to geographic characteristics of the networks were chosen (drainage pattern membership, order, tributary length, tributary balance and spacing between tributaries). Different results can be obtained by adjusting the weights of the multi-objective function. For example, the drainage pattern can be preserved by assigning more weight to w_M. If w_O was set to a higher value, the tributaries in lower order would be eliminated first. The length factor can preserve longer tributaries; the balance coefficient can keep the original balance of a tributary along a river; and the tributary spacing can let the tributaries not cluster together.

The method is used to assess the influence of different factors in the generalization process for each type of drainage. It was applied to the Russian river data and results were compared with manually generalized data with the goal to achieve similar results. The most important factor is the length. In general, during the manual generalization, the length is indeed the most considerable factor. For the manually generalized river networks, the drainage pattern is not considered well by comparing the pattern membership value before and after the generalization. The drainage pattern can be preserved better if the pattern membership participates in the GA process.

For each pattern, a proper setting for weights is given to achieve a greater similarity with the manual generalized river networks. One limitation is that not all settings of weights are tested. The obtained setting is an experimental approximation for a drainage pattern. First of all, for all patterns, the length is the most important factor. In the dendritic pattern, the pattern membership and the order are the second important factors; the factor of balance is not so important and even cannot be considered. For the trellis pattern, the tributary spacing is the second important factor, and the balance is also not important. Trellis tributaries are usually short streams of order 1. As a consequence, giving too much importance to the order tends to eliminate these tributaries first and loose the character of the network. As a consequence, the order should not be considered as an important factor for the preservation of trellis. In the parallel pattern, the membership factor is more important than others except the length. Table 5.10 illustrates the approximate settings for weights.

A river network in order 3 is more recommended for the practical application. No matter in the average similarity or the average pattern membership, river networks in order 3 perform better than in other orders.

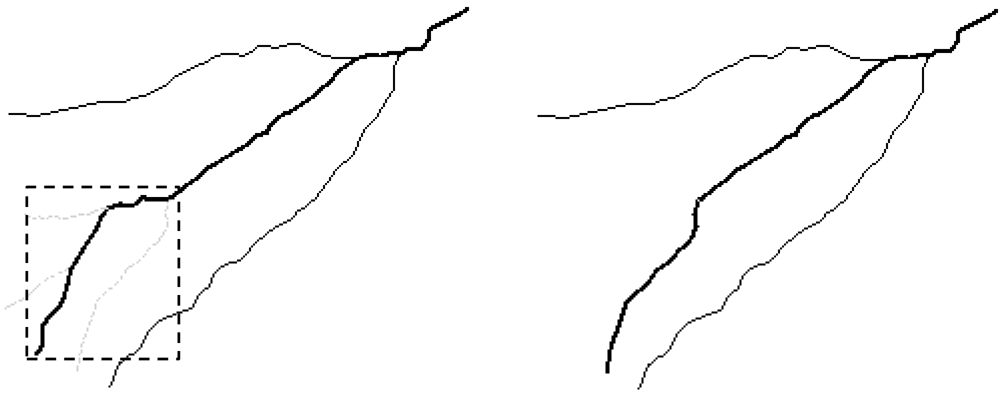
Table 5.10 The approximate weight settings for each drainage pattern

	Dendritic	Parallel	Trellis	Rectangular
Pattern membership (w_M)	●	●	□	□
Stream order (w_O)	●	●	□	□
Stream length (w_L)	★	★	★	★
Balance coefficient (w_B)	×	□	×	□
Tributary spacing (w_S)	□	□	●	□

★ - more important ● - important
□ - not important × - not considered

From the experiment, the order factor seems not so important because of the initialization of a chromosome. A new chromosome is created based on the order. The lower order strokes are removed first. So, in the GA process, the value of the order objective function always starts from 1. The initialization provides a good start for the GA. The advantage is that it can save the time for the GA to some extent. However, the order still needs to add to the objective function. If the order is not considered, some higher order strokes expected to be preserved would be removed in the GA process.

One limitation of the research is that the similarity is not improved obviously. There are some reasons. ① The GA is implemented by encoding the network with strokes. Correct strokes will help to increase the similarity, but sometimes strokes are not built as expected. As an example, in Figure 5.11, network (a) is from the Russian river at 1:24K scale and (b) is from the NHD at 1:100K scale. The bold line is the main stream obtained by the stroke. In the dashed box, the stroke is not the same as the stroke in network (b). So, no matter how to adjust the weights, network (a) cannot be generalized as (b). ② Some manual networks are not generalized as expected: some tributaries are short and in lower order, but they are still selected after generalization. It may happen that the tributary has some significant meanings on geography and should be preserved whatever it is short or long. In Figure 5.11, network (c) is a generalized network, where dashed tributaries are eliminated by considering the length. However, in the dashed circle, network (d), which is from the NHD, selects a shortest tributary.



(a) 1:24K scale from the Russian river (b) 1:100K scale from the NHD

(c) 1:24K scale from the Russian river (d) 1:100K scale from the NHD

Figure 5.11 Unexpected situations, dashed lines are eliminated tributaries.

Another limitation is that the involved map scales are 1:24K scale and 1:100K scale. The 1:100K scale dataset is used for the evaluation. In the Russian river, the parallel and rectangular networks are not so many. So, other datasets and other scales data should be used to do the test in the feature work.

In addition, the GA is useful in the multi-objective problem of the global optimization for the river network generalization. It can provide a fast method to get a solution. From the experiment, a solution is gained not more than 100 generations and sometimes less than 10 generations in the GA process. However, some faults of the GA itself bring some problems in the research. First, parameters should be set properly such as the population size and the terminal condition. In the research, these settings are empirical values. Second, solution is not unique sometimes. A group of solutions can be obtained to satisfy the requirements. Third, a local optimal solution might be received in the GA process.

Chapter 6 Summary and perspective

6.1 Summary

Drainage pattern is an important factor to describe the morphological structure of a drainage system and reflects the geographic and topological characteristics of a river network extracted from the drainage system. As a kind of semantic information about the river network at basin scale, it can be considered in river network generalization, terrain analysis or other aspects. However, automatic classification of drainage patterns is not well studied and so received limited interest in GIS applications. Therefore, the objective of this thesis is to develop an original drainage classification method for river networks. Knowledge about drainage patterns is used to enrich the generalization process and to evaluate the quality of generalized river networks, following the idea that map generalization is "*not a mere reduction of information*" but one of the challenges to preserve "*the geographic meaning*" (Bard & Ruas, 2005). This research focuses on river network generalization where the network is considered as a geographical entity whose meaning should be preserved. It is therefore a novel work in explicitly considering the geographical and hydrographical aspects of the cartographical data. It looks at the quality of river network generalization in geo-spatial analysis with the consideration of drainage pattern. It also participates to a research trend in map generalization (and more globally in GIS) to include on the top of the geometric aspect, geographical one as map generalization is the process of "information abstraction" rather than "data compression"(Ai et al., 2006). In this thesis, three original contributions are introduced.

The first and most important contribution is the development of a method for drainage pattern recognition. Chapter 3 proposed a new method based on geometric indicators, such as the junction angle, sinuosity and catchment elongation, to recognize the pattern of a river network automatically. In the method, fuzzy logic was applied to improve the robustness of the classification. Five patterns are classified in the research: dendritic, parallel, trellis, rectangular and reticulate pattern. From the result, the drainage pattern of a river network can be classified successfully, and the junction angle and the elongation are most significant. Some unclassified networks happened either because they could belong to several different patterns, or to none. In addition, a hierarchical structure is implemented to store a river network by the drainage patterns of its sub-networks at different levels. Such classification and organization can be used for generalization to select river tributaries according to the pattern of network. There are still some limitations in this research. ① Validation on the classification of drainage pattern is done by visual assessment of case studies, because there was no such information as a drainage classification in existing datasets. ② Currently, only five drainage patterns are addressed in the research. More patterns such as radial and centripetal patterns could be considered in further work. ③ The influence of quality and resolution of source data is not studied, but they may affect the number of tributaries represented in the network and the computation of indicators.

The second contribution concerns the evaluation of river selection methods in generalization. Chapter 4 provided a quality evaluation method to check whether a generalized river network preserves the drainage pattern or not. In the method, the membership value from a fuzzy rule for a drainage pattern is obtained and evaluated. Four drainage patterns are addressed: dendritic, trellis, parallel and rectangular as these four patterns are classified by fuzzy logic. The evaluation method quantifies pattern preservation. The generalized river network better preserves the pattern when the membership value is higher. Manually generalized networks are always better than other generalized networks in terms of pattern preservation. However, if there are not enough tributaries in the network, the membership value could be misleading because average junction angle, bended tributaries percentage and average length ratio are statistical parameters. Therefore, this method is more adapted for river networks at order 3 and 4. However, the membership value only can be used for the evaluation of the drainage pattern. It can be a supplement for the evaluation of river network generalization which can be used with other measures specific to river segments.

The third contribution is the development of a generalization method with consideration of different geographic and geometric factors. Chapter 5 introduced a multi-objective optimization method made use of genetic algorithm. Only selective omission is considered. The GA process is designed appropriately for tributary selection including the encoding, initialization, crossover and mutation stages. For the encoding, chromosomes are defined by strokes of the river network. Fitness of an individual is obtained by a multi-objective model where five factors are proposed: drainage pattern membership, stream order, length, balance coefficient and tributary spacing. Each factor is built as an objective function, and the multi-objective function aggregates all objective functions to compute the fitness of a chromosome. In the model, different weights are assigned and tested and the importance of each factor is evaluated. From the result, the most important factor yielding similar results to manually generalized network for all patterns is the length. With consideration of pattern membership, the drainage pattern can be preserved well in the GA process. In the dendritic pattern, the pattern membership and the order are the second most important factors. Tributary spacing is the second most important factor in trellis. For parallel drainages, the membership factor is more important than all others except length. In addition, river networks in order 3 are more recommended for practical application. However, considering several factors in the GA process at same time, the similarity between the generalized network and the manual network is not improved obviously. In this study, only two scales 1:24K and 1:100K datasets of the Russian river are tested, other datasets at different scales should be tested to confirm conclusions. The weights of different factors in the multi-objective function are not fixed, and only approximate weight settings are provided through the tests in the experiment.

6.2 Perspective

Although drainage basin and river are both elements which attracted a lot of attention in GIS, consideration for river networks as objects of their own and their organization in different patterns did not receive as much consideration. The drainage pattern recognition method and application to generalization presented in this thesis provided new concepts and original knowledge mainly applied to cartographic processes. As limited analyses have been conducted so far in this direction, we believe that further research can be done at both conceptual and application level.

In short term perspectives, some work can be done to improve the study in this thesis:

- *Automatic drainage pattern recognition*. A first direction is to develop additional indicators for the recognition of other drainage patterns such as radial and centripetal patterns. Such patterns depend on the spatial organization of a group of networks and topologic indicators expressing relationships between networks should be considered. The second one is to test the method on more dataset at different scales from different data sources (e.g. DTM or DEM). The computation of indicators, such as average junction angle, bended tributary percentage and average length ratio, relies on the number of tributaries represented in a network, which can be influenced by the map scale. More river networks from a micro scale to a macro scale should be tested. It can find out the influence of quality and resolution of source data on the drainage pattern classification. The fuzzy logic approach presented in our work was appropriate for dealing with geometric indicators. If other indicators relating with topology or semantic are to be introduced, other approaches such as neural networks or machine learning may be considered. These approaches need more river data, in which the pattern information is already known, to do the parameter fitting or the training. Validation is done visually by assessing classified river networks. A further validation can be conducted by experts such as cartographers and geographers through questionnaires or other survey methods. Indeed, knowledge of experts could also contribute in the fine-tuning of membership functions, improving classification results and helping distinguishing misclassified from unclassifiable networks.

- *Evaluation of pattern preservation*. The main further work is to develop the method to cover more drainage patterns, such as reticulate, radial, and centripetal patterns, in the quality assessment.
- *Multi-objective optimization in river network generalization*. The first one is that other factors can be considered. The five factors proposed in the thesis are logical to some extent but that was an empirical choice. Some semantic factors such as the name and geographical meanings of tributaries are not considered. The name of tributaries can help to establish correct strokes. Some tributaries have priorities to be preserved in the process of the selection omission according to their geographical meanings. Another one is to provide a method to calculate the weights of objective functions. This work needs more data of river networks at different scales to achieve.

In long term perspectives, more work can be done to develop the research to other applications:

- First, the concept of fuzzy logic can be applied to the classification of other patterns, such as road patterns or building patterns where a classification can also be established (Heinzle, Anders, & Sester, 2006).
- Then, the drainage pattern information can be used for terrain analysis or classification, either as a direct application or by integrating this with other DTM/DEM information (Nardi et al., 2008). Different types of terrain information about e.g. the physiology and geology of the terrain can be added in a GIS and, overlaying them with the river data, assist in providing more knowledge about the river system and non-geometric indicators for a more accurate classification.
- For the quality of the generalization, only considering the drainage pattern in river network generalization is obviously not enough and more aspects should be taken into account in the evaluation system for generalization methods, such as the tributary balance and the spacing between two paralleled tributaries. In addition, spatial relationships with

other objects in cartography, e.g. roads or buildings, can also be considered into the evaluation system.

- Furthermore, the drainage pattern can be used in the simplification of river segments to add constrains for cartographic generalization at different scales (Gutman & Weaver, 2012) or to integrate rivers with other field or object data such as terrain or buildings (Gaffuri, 2007). Geometric indicators can also be used to caricature or exaggerate characteristics of a network to provide a schematic representation of the drainage (Nijssen, Lettenmaier, Liang, Wetzel, & Wood, 1997).

APPENDIXES

A. Data structures

Junction
Number id # JunctionType type # GeoPoint geometry # List<RiverSegment> connected_river_segment_list
+ AddRiverSegment(RiverSegment) + Boolean Equals(Junction)

RiverSegment
Number id # Number order # Junction from # Junction to # GeoLineString geometry
+ Boolean Equals(RiverSegment)

Stream
Number id # Number order # DrainagePattern pattern # GeoLinearRing catchment # Stream lower_stream # List<RiverSegment> inside_river_segment_list # List<Stream> upper_stream_list
+ Number GetLength() + Boolean Equals(Stream)

B. Algorithms

1. Algorithm for building the hierarchy of sub-networks.

Input : A stream and a handling river segment
Output : The root of the drainage tree

```
void Build(Stream st, RiverSegment rs)
   for each connected river segment rs' to rs
      if rs' in a reticulate part then
         set this reticulate part as a new reticulate network rn
         locate rn under root
         for each river segment rs'' connecting to this reticulate
            self-call Build(st, rs'')
         end for
      else if rs'.order = rs.order then
            add rs' to st
            self-call Build(st, rs')
      else if rs'.order ≠ 1 then
            new a stream st'
            add rs' to st'
            locate st' under st
            self-call Build(st',rs')
      end if
   end for
end
```

2. Algorithm for merging sub-networks according to their drainage pattern along a river stream.

Input : The root of the drainage tree before merging
Output : The root of the new drainage tree

```
void Merge(Stream st)
   if st is a leaf or st.pattern = reticulate then return
   if  st has same pattern with its sub-networks' then return
   for each sub-network st' of st
      self-call Merge(st')
   end for
   set N as the count of river segments {rs1, rs2...rsN} in st
   set i as 1
   while i ≤ N
      if all connected sub-networks have same pattern p then
         new a stream st''
         locate st'' under st
         set st'' pattern as p
         add river segment rsi to st''
         for j = i+1 to N
            for each sub-network st* connecting to rsj
```

```
                if st* is on the same side of st'' and st*.pattern = p then
                    add rsj to st''
                    locate st* under st''
                    set break = false
                else
                    set break = true
                end if
            end for
            if break = true then exit for
        end for
    else
            i++
    end if
  end while
end
```

3. Algorithm for removing redundant information in the drainage tree.

Input : The root of the drainage tree before removing
Output : The root of the new drainage tree

```
void Remove(Stream st)
  if st is a leaf or st.pattern = reticulate then  return
  if all sub-networks have same drainage pattern with st then
      remove sub-networks information
  else
      for each connected stream st' of rs
          self-call Remove(st')
      end for
  end if
end
```

C. DOT scripts for the drainage tree

1. The DOT script for Figure 3.20:

```
graph {
graph[nodesep=".05"];
node[margin=".01" shape=box color=black width=.1 height=.1 fontsize=10];
28[style=dashed,color=red]; 30[style=dashed,color=red];
1--2; 1--5; 1--6; 1--15; 1--20; 1--21; 1--25; 1--26; 1--27; 1--28;
1--30; 1--33; 1--37; 1--40; 1--43; 1--44; 1--45; 1--46; 1--47; 2--3;
2--4; 6--7; 6--8; 6--9; 6--10; 6--11; 6--12; 6--13; 6--14; 15--16;
15--17; 15--18; 15--19; 21--22; 21--23; 21--24; 27--29; 27--31; 27--32; 33--34;
33--35; 33--36; 37--38; 37--39; 40--41; 40--42;
}
```

2. The DOT script for Figure 3.21:

```
graph {
graph[nodesep=".05"];
node[margin=".01" shape=box color=black width=.1 height=.1 fontsize=10];
28[style=dashed,color=red]; 30[style=dashed,color=red];
1[style=filled,color=skyblue]; 2[style=filled,color=orange];
5[style=filled,color=orange]; 6[style=filled,color=yellow];
15[style=filled,color=skyblue]; 20[style=filled,color=orange];
21[style=filled,color=skyblue]; 25[style=filled,color=skyblue];
26[style=filled,color=orange]; 27[style=filled,color=skyblue];
33[style=filled,color=yellow]; 37[style=filled,color=skyblue];
40[style=filled,color=tomato]; 43[style=filled,color=skyblue];
44[style=filled,color=skyblue]; 45[style=filled,color=orange];
46[style=filled,color=orange]; 47[style=filled,color=orange];
3[style=filled,color=orange]; 4[style=filled,color=orange];
7[style=filled,color=skyblue]; 8[style=filled,color=orange];
9[style=filled,color=skyblue]; 10[style=filled,color=orange];
11[style=filled,color=orange]; 12[style=filled,color=orange];
13[style=filled,color=orange]; 14[style=filled,color=orange];
16[style=filled,color=orange]; 17[style=filled,color=skyblue];
18[style=filled,color=orange]; 19[style=filled,color=skyblue];
22[style=filled,color=orange]; 23[style=filled,color=skyblue];
24[style=filled,color=skyblue]; 29[style=filled,color=yellow];
31[style=filled,color=skyblue]; 32[style=filled,color=orange];
34[style=filled,color=skyblue]; 35[style=filled,color=orange];
36[style=filled,color=yellow]; 38[style=filled,color=skyblue];
39[style=filled,color=orange]; 41[style=filled,color=skyblue];
42[style=filled,color=skyblue];
1--2; 1--5; 1--6; 1--15; 1--20; 1--21; 1--25; 1--26; 1--27; 1--28;
1--30; 1--33; 1--37; 1--40; 1--43; 1--44; 1--45; 1--46; 1--47; 2--3;
2--4; 6--7; 6--8; 6--9; 6--10; 6--11; 6--12; 6--13; 6--14; 15--16;
15--17; 15--18; 15--19; 21--22; 21--23; 21--24; 27--29; 27--31; 27--32; 33--34;
33--35; 33--36; 37--38; 37--39; 40--41; 40--42;
}
```

3. The DOT script for Figure 3.23:

```
graph {
graph[ranksep="0.25" nodesep=".05"];
node[margin=".01" shape=box color=black width=.1 height=.1 fontsize=10];
```

```
28[style=dashed,color=red]; 30[style=dashed,color=red];
1[style=filled,color=skyblue]; 6[style=filled,color=yellow];
15[style=filled,color=skyblue]; 26[style=filled,color=orange];
27[style=filled,color=skyblue];33[style=filled,color=yellow];
37[style=filled,color=skyblue]; 40[style=filled,color=tomato];
46[style=filled,color=orange]; 50[label="50*",style=filled,color=orange];
51[label="51*",style=filled,color=orange];
52[label="52*",style=filled,color=skyblue];
7[style=filled,color=skyblue]; 8[style=filled,color=orange];
9[style=filled,color=skyblue]; 48[label="48*",style=filled,color=orange];
14[style=filled,color=orange]; 13[style=filled,color=orange];
12[style=filled,color=orange]; 11[style=filled,color=orange];
10[style=filled,color=orange]; 16[style=filled,color=orange];
17[style=filled,color=skyblue]; 18[style=filled,color=orange];
19[style=filled,color=skyblue]; 29[style=filled,color=yellow];
31[style=filled,color=skyblue]; 32[style=filled,color=orange];
34[style=filled,color=skyblue]; 35[style=filled,color=orange];
36[style=filled,color=yellow]; 38[style=filled,color=skyblue];
39[style=filled,color=orange]; 41[style=filled,color=skyblue];
42[style=filled,color=skyblue]; 2[style=filled,color=orange];
5[style=filled,color=orange]; 47[style=filled,color=orange];
3[style=filled,color=orange]; 4[style=filled,color=orange];
20[style=filled,color=orange]; 45[style=filled,color=orange];
21[style=filled,color=skyblue]; 44[style=filled,color=skyblue];
43[style=filled,color=skyblue]; 25[style=filled,color=skyblue];
22[style=filled,color=orange]; 49[label="49*",style=filled,color=skyblue];
24[style=filled,color=skyblue]; 23[style=filled,color=skyblue];
{rank=same;1;}
{rank=same;50;52;}
{rank=same;6;15;27;28;30;33;37;40;2;21;}
{rank=same;48;51;49;}
{rank=same;7;8;9;14;13;12;11;10;16;17;18;19;26;29;31;32;34;35;36;38;39;41;42;46
;3;4;5;47;20;45;22;24;23;44;43;25;}
1--6; 1--15; 1--26; 1--27; 1--28; 1--30; 1--33; 1--37; 1--40; 1--46;
1--50; 1--51; 1--52; 6--7; 6--8; 6--9; 6--48; 48--14; 48--13; 48--12;
48--11; 48--10; 15--16; 15--17; 15--18; 15--19; 27--29; 27--31; 27--32; 33--34;
33--35; 33--36; 37--38; 37--39; 40--41; 40--42; 50--2; 50--5; 50--47; 2--3;
2--4; 51--20; 51--45; 52--21; 52--44; 52--43; 52--25; 21--22; 21--49; 49--24;
49--23;
}
```

4. The DOT script for Figure 3.24:

```
graph {
graph[ranksep=".05"  nodesep=".05"];
node[margin=".01" shape=box color=black width=.1 height=.1 fontsize=10];
28[style=dashed,color=red]; 30[style=dashed,color=red];
1[style=filled,color=skyblue]; 6[style=filled,color=yellow];
15[style=filled,color=skyblue]; 26[style=filled,color=orange];
27[style=filled,color=skyblue];33[style=filled,color=yellow];
37[style=filled,color=skyblue]; 40[style=filled,color=tomato];
46[style=filled,color=orange]; 50[label="50*",style=filled,color=orange];
51[label="51*",style=filled,color=orange];
52[label="52*",style=filled,color=skyblue];
7[style=filled,color=skyblue]; 8[style=filled,color=orange];
9[style=filled,color=skyblue]; 48[label="*",style=filled,color=orange];
16[style=filled,color=orange]; 17[style=filled,color=skyblue];
18[style=filled,color=orange]; 19[style=filled,color=skyblue];
29[style=filled,color=yellow]; 31[style=filled,color=skyblue];
32[style=filled,color=orange]; 34[style=filled,color=skyblue];
35[style=filled,color=orange]; 36[style=filled,color=yellow];
38[style=filled,color=skyblue]; 39[style=filled,color=orange];
41[style=filled,color=skyblue]; 42[style=filled,color=skyblue];
21[style=filled,color=skyblue]; 44[style=filled,color=skyblue];
43[style=filled,color=skyblue]; 25[style=filled,color=skyblue];
22[style=filled,color=orange]; 49[label="49*",style=filled,color=skyblue];
{rank=same;1;}
{rank=same;50;52;}
{rank=same;6;15;27;28;30;33;37;40;21;}
{rank=same;48;51;49;}
{rank=same;7;8;9;16;17;18;19;26;29;31;32;34;35;36;38;39;41;42;46;22;44;43;25;}
1--6; 1--15; 1--26; 1--27; 1--28; 1--30; 1--33; 1--37; 1--40; 1--46;
1--50; 1--51; 1--52; 6--7; 6--8; 6--9; 6--48; 15--16; 15--17; 15--18;
15--19; 27--29; 27--31; 27--32; 33--34; 33--35; 33--36; 37--38; 37--39; 40--41;
40--42; 52--21; 52--44; 52--43; 52--25; 21--22; 21--49;
}
```

5. The DOT script for Figure 3.27:

```
graph {
{
node[margin=.01, width=.1, height=.1, shape=plaintext, fontsize=6];
a--b--c--d--e--f--g;
}
graph[ranksep=".25" nodesep=".03"];
node[margin=".01" shape=box color=black width=.1 height=.1 fontsize=6, label=""];
1[style=filled,color=skyblue]; 2[style=filled,color=skyblue];
3[style=filled,color=skyblue]; 86[style=filled,color=skyblue];
7[style=filled,color=skyblue]; 43[style=filled,color=yellow];
44[style=filled,color=skyblue]; 47[style=filled,color=tomato];
57[style=filled,color=orange]; 58[style=filled,color=skyblue];
67[style=filled,color=skyblue]; 79[style=filled,color=orange];
80[style=filled,color=skyblue]; 84[style=filled,color=yellow];
85[style=filled,color=orange]; 130[label="*",style=filled,color=orange];
8[style=filled,color=orange]; 17[style=filled,color=skyblue];
22[style=filled,color=orange]; 23[style=filled,color=skyblue];
121[label="*",style=filled,color=skyblue]; 122[label="*",style=filled,color=orange];
```

```
123[label="*",style=filled,color=skyblue]; 124[label="*",style=filled,color=orange];
125[label="*",style=filled,color=skyblue]; 21[style=filled,color=skyblue];
117[label="*",style=filled,color=orange]; 25[style=filled,color=skyblue];
118[label="*",style=filled,color=orange]; 41[style=filled,color=skyblue];
38[style=filled,color=skyblue]; 37[style=filled,color=skyblue];
15[style=filled,color=skyblue]; 36[style=filled,color=skyblue];
39[style=filled,color=skyblue]; 40[style=filled,color=orange];
32[style=filled,color=skyblue]; 27[style=filled,color=skyblue];
33[style=filled,color=skyblue]; 34[style=filled,color=yellow];
119[label="*",style=filled,color=orange]; 120[label="*",style=filled,color=skyblue];
45[style=filled,color=orange]; 46[style=filled,color=skyblue];
48[style=filled,color=yellow]; 49[style=filled,color=orange];
52[style=filled,color=yellow]; 126[label="*",style=filled,color=skyblue];
59[style=filled,color=skyblue]; 63[style=filled,color=skyblue];
66[style=filled,color=yellow]; 60[style=filled,color=orange];
61[style=filled,color=orange]; 62[style=filled,color=skyblue];
64[style=filled,color=orange]; 65[style=filled,color=skyblue];
68[style=filled,color=skyblue]; 76[style=filled,color=skyblue];
71[style=filled,color=orange]; 72[style=filled,color=skyblue];
127[label="*",style=filled,color=orange]; 128[label="*",style=filled,color=skyblue];
77[style=filled,color=orange]; 78[style=filled,color=skyblue];
82[style=filled,color=orange]; 129[label="*",style=filled,color=skyblue];
87[style=filled,color=skyblue]; 97[style=filled,color=skyblue];
116[style=filled,color=skyblue]; 88[style=filled,color=orange];
89[style=filled,color=skyblue]; 93[style=filled,color=skyblue];
96[style=filled,color=skyblue]; 92[style=filled,color=yellow];
131[label="*",style=filled,color=orange]; 94[style=filled,color=skyblue];
95[style=filled,color=orange]; 107[style=filled,color=skyblue];
133[label="*",style=filled,color=orange]; 134[label="*",style=filled,color=skyblue];
108[style=filled,color=skyblue]; 109[style=filled,color=skyblue];
110[style=filled,color=orange]; 104[style=filled,color=skyblue];
111[style=filled,color=skyblue]; 105[style=filled,color=orange];
106[style=filled,color=skyblue]; 113[style=filled,color=orange];
114[style=filled,color=orange]; 132[label="*",style=filled,color=orange];
{rank=same;a;1;}
{rank=same;b;3;86;}
{rank=same;c;7;58;67;87;97;}
{rank=same;d;123;125;130;134;}
{rank=same;e;17;23;38;32;27;44;47;59;63;68;76;80;89;93;107;104;111;}
{rank=same;f;117;118;121;122;124;119;120;126;127;128;129;131;133;132;}
{rank=same;g;2;8;21;22;25;41;39;40;37;15;36;33;34;43;45;46;48;49;52;57;60;61;62
;64;65;66;71;72;77;78;79;82;84;85;88;92;94;95;96;108;109;110;105;106;113;114;11
6;}
1--2; 1--3; 1--86; 3--7; 3--43; 3--44; 3--47; 3--57; 3--58; 3--67;
3--79; 3--80; 3--84; 3--85; 3--130; 7--8; 7--17; 7--22; 7--23; 7--121;
7--122; 7--123; 7--124; 7--125; 17--21; 17--117; 23--25; 23--118; 123--41; 123--38;
123--37; 123--15; 123--36; 38--39; 38--40; 125--32; 125--27; 32--33; 32--34; 27--119;
27--120; 44--45; 44--46; 47--48; 47--49; 47--52; 47--126; 58--59; 58--63; 58--66;
59--60; 59--61; 59--62; 63--64; 63--65; 67--68; 67--76; 68--71; 68--72; 68--127;
68--128; 76--77; 76--78; 80--82; 80--129; 86--87; 86--97; 86--116; 87--88; 87--89;
87--93; 87--96; 89--92; 89--131; 93--94; 93--95; 97--107; 97--133; 97--134; 107--108;
107--109; 107--110; 134--104; 134--111; 104--105; 104--106; 111--113; 111--114;
111--132;
}
```

REFERENCES

Ai, T., Liu, Y., & Chen, J. (2006). The Hierarchical Watershed Partitioning and Data Simplification of River Network. In A. Riedl, W. Kainz, & G. A. Elmes (Eds.), *Progress in Spatial Data Handling* (pp. 617–632). Springer Berlin Heidelberg.

Alt, H., & Godau, M. (1995). Computing the Frechet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, *5*(01n02), 75–91.

Alves, D. S. (1993). The Amazonia information system. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING*, *29*, 259.

Ansari, N., & Delp, E. J. (1991). On detecting dominant points. *Pattern Recognition*, *24*(5), 441–451.

Balboa, J. L. G., & López, F. J. A. (2000). Frequency filtering of linear features by means of wavelets. A method and an example. *The Cartographic Journal*, *37*(1), 39–49.

Bard, S. (2004). Quality Assessment of Cartographic Generalisation. *Transactions in GIS*, *8*(1), 63–81.

Bard, S., & Ruas, A. (2005). Why and how evaluating generalised data? In *Developments in Spatial Data Handling*. Berlin Heidelberg: Springer.

Biggs, N., Lloyd, E. K., & Wilson, R. J. (1986). *Graph Theory, 1736-1936*. Clarendon Press New York, NY, USA.

Binks, P. J., & Hooper, G. J. (1984). Uranium in Tertiary palaeochannels, “West Coast Area”, South Australia. In *Australasian Institute of Mining and Metallurgy* (Vol. 289, pp. 271–275). Darwin, N. T.

Bondy, J. A., & Murty, U. S. R. (2008). *Graph theory*. London: Springer.

Braden, B. (1986). The Surveyor’s Area Formula. *The College Mathematics Journal*, *17*(4), 326–337.

Brenner, N. (1969). Fast Fourier transform of externally stored data. *Audio and Electroacoustics, IEEE Transactions on*, *17*(2), 128–132.

Burghardt, D., & Meier, S. (1997). Cartographic displacement using the snakes concept. *Semantic Modeling for the Acquisition of Topografic Information from Images and Maps, Birkhaeuser Verlag*, 59–71.

Buttenfield, B. P., Stanislawski, L. V, & Brewer, C. A. (2010). Multiscale representations of water: Tailoring generalization sequences to specific physiographic regimes. In *Proceedings of GIScience 2010* (pp. 14–17).

Buyya, R. (2005). Cost-Based Scheduling of Scientific Workflow Application on Utility Grids. In *First International Conference on e-Science and Grid Computing (e-Science'05)* (pp. 140–147). IEEE.

Charlton, R. (2008). *Fundamentals of fluvial geomorphology*. Psychology Press.

Chen, J., Liu, W., Li, Z., Zhao, R., & Cheng, T. (2007). Detection of spatial conflicts between rivers and contours in digital map updating. *International Journal of Geographical Information Science*, *21*(10), 1093–1114.

Chernicoff, S., & Whitney, D. (2006). *Geology (4th Edition)* (p. 744). Prentice Hall.

Coello, C. A. C., Lamont, G. B., & Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer.

Coffman, D. M., & Turner, A. K. (1971). Computer Determination of the geometry and topology of stream networks. *Water Resources Research*, *7*(2), 419–423.

Colombo, R., Vogt, J. V., Soille, P., Paracchini, M. L., & de Jager, A. (2007). Deriving river networks and catchments at the European scale from medium resolution digital elevation data. *CATENA*, *70*(3), 296–305.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Comput*, *19*(90), 297–301.

De Wit, M. C. J. (1999). Post-Gondwana drainage and the development of diamond placers in western South Africa. *Economic Geology*, *94*(5), 721–740.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Wiley.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

Della Croce, F., Tadei, R., & Volta, G. (1995). A genetic algorithm for the job shop problem. *Computers & Operations Research*, *22*(1), 15–24.

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *10*(2), 112–122.

Fagan, S. D., & Nanson, G. C. (2004). The morphology and formation of floodplain-surface channels, Cooper Creek, Australia. *Geomorphology*, *60*(1), 107–126.

Fairfield, J., & Leymarie, P. (1991). Drainage networks from grid digital elevation models. *Water Resources Research*, *27*(5), 709–717.

Florinsky, I. V. (2009). Computation of the third - order partial derivatives from a digital elevation model. *International Journal of Geographical Information Science*, *23*(2), 213–231.

Fukuda, T., & Shibata, T. (1992). Hierarchical intelligent control for robotic motion by using fuzzy, artificial intelligence, and neural network. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks* (Vol. 1, pp. 269–274). IEEE.

Furnans, J., & Olivera, F. (2001). Watershed Topology - The Pfafstetter System. In *ESRI USER CONFERENCE* (p. Vol. 21).

Fürst, J., & Hörhan, T. (2009). Coding of watershed and river hierarchy to support GIS-based hydrological analyses at different scales. *Computers & Geosciences*, *35*(3), 688–696.

Gaffuri, J. (2007). Outflow preservation of the hydrographic network on the relief in map generalisation. In *International Cartographic Conference*,. Moscow.

Gen, M., & Cheng, R. (1999). *Genetic algorithms and engineering optimization*. Wiley-interscience.

Gleyzer, A., Denisyuk, M., Rimmer, A., & Salingar, Y. (2004). A fast recursive GIS algorithem for computing StarHler stream order in braided and nonbraided networks. *JAWRA Journal of the American Water Resources Association*, *40*(4), 937–946.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning* (1st ed.). Addison-Wesley Professional.

Grimaldi, S., Nardi, F., Benedetto, F. Di, Istanbulluoglu, E., & Bras, R. L. (2007). A physically-based method for removing pits in digital elevation models. *Advances in Water Resources*, *30*(10), 2151–2158.

Guilbert, E., & Lin, H. (2006). B-Spline curve smoothing under position constraints for line generalisation. In *In Proceedings of the ACM GIS'06* (pp. 3–10). ACM.

Gutman, M., & Weaver, C. (2012). Wavelet-based Automated River Network Generalization Categories and Subject Descriptors. In *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications* (p. 12). ACM.

Hajela, P., & Lin, C. Y. (1992). Genetic search strategies in multicriterion optimal design. *Structural and Multidisciplinary Optimization*, *4*(2), 99–107.

Heinzle, F., Anders, K., & Sester, M. (2006). Pattern recognition in road networks on the example of circular road detection. *Geographic Information Science*, *4197*, 153–167.

Hills, E. S. (1972). *Elements of structural geology*. Chapman and Hall.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Cambridge, MA: MIT Press.

Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, *56*(3), 275–370.

Howard, A. D. (1967). Drainage analysis in geologic interpretation: a summation. *American Association of Petroleum Geologists Bulletin*, *51*(11), 2246–2259.

Jia, Y., Wang, H., Zhou, Z., Qiu, Y., Luo, X., Wang, J., … Qin, D. (2006). Development of the WEP-L distributed hydrological model and dynamic assessment of water resources in the Yellow River basin. *Journal of Hydrology*, *331*(3-4), 606–629.

Jiang, L., Qi, Q., & Zhang, A. (2009). How to decide the units of drainage pattern of generalization (Vol. 2, pp. II–658–II–661). IEEE.

Joao, E. (1998). *Causes and consequences of map generalisation*. London: Taylor & Francis.

Johnson, J. M., & Rahmat-Samii, V. (1997). Genetic algorithms in engineering electromagnetics. *IEEE Antennas and Propagation Magazine*, *39*(4), 7–21.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, *1*(4), 321–331.

Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K., & Tomita, M. (2003). Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, *19*(5), 643–650.

Knighton, D. (1998). *Fluvial forms and processes: a new perspective*. Arnold, Hodder Headline, PLC.

Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, *91*(9), 992–1007.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics*, *22*(24), 3096–8.

Lambert, D. (2007). *The field guide to geology*. Checkmark Books.

Lang, T. (1969). Rules for robot draughtsmen. *Geographical Magazine*, *42*(1), 50–51.

Leekwijck, W. Van, & Kerre, E. E. (1999). Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, *108*(2), 159–178.

Leopold, L. B., & Wolman, M. G. (1957). *River Channel Patterns: Braided, Meandering and Straight*. Washington (DC): US Government Printing Office.

Lewin, J., & Brewer, P. A. (2001). Predicting channel patterns. *Geomorphology*, *40*(3), 329–339.

Li, T., Wang, G., & Chen, J. (2010). A modified binary tree codification of drainage networks to support complex hydrological models. *Computers & Geosciences*, *36*(11), 1427–1435.

Li, Z. (1988). An algorithm for compressing digital contour data. *Cartographic Journal, The*, *25*(2), 143–146.

Li, Z. (2007). *Algorithmic foundation of multi-scale spatial representation*. CRC.

Li, Z., Yan, H., Ai, T., & Chen, J. (2004). Automated building generalization based on urban morphology and Gestalt theory. *International Journal of Geographical Information Science*, *18*(5), 513–534.

Liao, G.-C., & Tsao, T.-P. (2004). Application of fuzzy neural networks and artificial intelligence for load forecasting. *Electric Power Systems Research*, *70*(3), 237–244.

Mackaness, W., & Edwards, G. (2002). The importance of modelling pattern and structure in automated map generalisation. In *Joint Workshop on Multi-Scale Representations of Spatial Data*. Ottawa.

Mackaness, W., & Ruas, A. (2007). Evaluation in the map generalisation process. In *Generalisation of Geographic Information : Cartographic Modelling and Applications* (pp. 89–111). Amsterdam: Elsevier.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Muller, J., Weibel, R., Lagrange, J., & Salge, F. (1995). Generalization: State of the art and issues. In *GIS and Generalization: Methodology and Practice.* (pp. 3–17). London: Taylor & Francis.

Nakos, B., & Miropoulos, V. (2003). Local length ratio as a measure of critical points detection for line simplification. In *5th ICA Workshop on Pregress in Automated Map Generalisation* (pp. 28–30). Paris: Citeseer.

Nardi, F., Grimaldi, S., Santini, M., Petroselli, A., & Ubertini, L. (2008). Hydrogeomorphic properties of simulated drainage patterns using digital elevation models: the flat area issue. *Hydrological Sciences Journal*, *53*(6), 1176–1193.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., & Wood, E. F. (1997). Streamflow simulation for continental-scale river basins. *Water Resources Research*, *33*(4), 711–724.

Notredame, C. (1996). SAGA: sequence alignment by genetic algorithm. *Nucleic Acids Research, 24*(8), 1515–1524.

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, *28*(3), 323–344.

Ortega, L., & Rueda, A. (2010). Parallel drainage network computation on CUDA. *Computers & Geosciences*, *36*(2), 171–178.

Paiva, J., & Egenhofer, M. J. (2000). Robust inference of the flow direction in river networks. *Algorithmica*, *26*(2), 315–333.

Pidwirny, M. (2006). The drainage basin concept. *Fundamentals of Physical Geography, 2nd Edition*.

Plazanet, C., Affholder, J. G., & Fritsch, E. (1995). The importance of geometric modeling in linear feature generalization. *Cartography and Geographic Information Science*, *22*(4), 291–305.

Poorten, P. M. Van Der, & Jones, C. B. (2002). Characterisation and generalisation of cartographic lines using Delaunay triangulation. *International Journal of Geographical Information Science*, *16*(8), 773–794.

Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, *369*(2), 853–866.

Rechenberg, I. (1973). *Evolutionsstrategie*. Stuttgart: Holzmann-Froboog.

Richardson, D. E. (1993). *Automated Spatial and Thematic Generalization Using a Context Transformation Model: Integrating Steering Parameters, Classification and Aggregation Hierarchies, Reduction Factors, and Topological Structures for Multiple Abstractions*. R&B Publications.

Ritter, M. E. (2006). The physical environment: An introduction to physical geography.

Rosenfeld, A., & Johnston, E. (1973). Angle detection on digital curves. *Computers, IEEE Transactions on*, *100*(9), 875–878.

Rusak Mazur, E., & Castner, H. W. (1990). Horton's ordering scheme and the generalisation of river networks. *Cartographic Journal*, *27*(2), 104–112.

Sandro, S., Massimo, R., & Matteo, Z. (2011). Pattern Recognition and Typification of Ditches. (A. Ruas, Ed.)*Advances in Cartography and GIScience*, *1*, 425–437.

Saux, E. (1998). B-spline curve fitting: Application to cartographic generalization of maritime lines. In *In Proceedings of the 8th International Conference on Computer Graphics and Visualization* (pp. 196–203). Keldysh Institute of Applied Mathematics.

Schumm, S. A. (1977). The fluvial system. *Publ. by: Wiley-Interscience*.

Schumm, S. A., Dumont, J. F., & Holbrook, J. M. (2002). *Active tectonics and alluvial rivers*. Cambridge Univ Pr.

Schumm, S. A., & Khan, H. R. (1972). Experimental study of channel patterns. *Geological Society of America Bulletin*, *83*(6), 1755–1770.

Serres, B., & Roy, A. G. (1990). FLOW DIRECTION AND BRANCHING GEOMETRY AT JUNCTIONS IN DENDRITIC RIVER NETWORKS*. *The Professional Geographer*, *42*(2), 194–201.

Shopova, E. G., & Vaklieva-Bancheva, N. G. (2006). BASIC—A genetic algorithm for engineering problems solution. *Computers & Chemical Engineering*, *30*(8), 1293–1309.

Shrestha, S., Kazama, F., & Newham, L. T. H. (2008). A framework for estimating pollutant export coefficients from long-term in-stream water quality monitoring data. *Environmental Modelling & Software*, *23*(2), 182–194.

Shreve, R. L. (1966). Statistical law of stream numbers. *The Journal of Geology*, 17–37.

Skopeliti, A., & Tsoulos, L. (2001). A methodology for the assessment of generalization quality. In *The Fourth ACI Workshop on Progress in Automated Map Generalization*. Beijing, China.

Smart, J. S. (1970). Use of topologic information in processing data for channel networks. *Water Resources Research*, *6*(3), 932–936.

Stanislawski, L. V. (2008). Development of a knowledge-based network pruning strategy for automated generalisation of the United States National Hydrography Dataset. In *The 11th ICA Workshop on Generalization and Multiple Representation*. Montpellier, France.

Stanislawski, L. V. (2009). Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. *Computers, Environment and Urban Systems*, *33*(5), 325–333.

Stanislawski, L. V, & Buttenfield, B. P. (2011). Hydrographic Generalization Tailored to Dry Mountainous Regions. *Cartography and Geographic Information Science*, *38*(2), 117–125.

Steiniger, S., & Meier, S. (2004). Snakes: a technique for line smoothing and displacement in map generalisation.

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, *38*(6), 913–920.

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, *33*(2), 309.

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological Processes*, *5*(1), 81–100.

Tarjan, R. E. (1974). A note on finding the bridges of a graph. *Information Processing Letters*, *2*(6), 160–161.

Teh, C. H., & Chin, R. T. (1989). On the detection of dominant points on digital curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *11*(8), 859–872.

Thomson, R. C., & Brooks, R. (2000). Efficient generalization and abstraction of network data using perceptual grouping. In *5th Int Conf on Geo-Computation* (pp. 23–25). University of Greenwich, UK.

Topfer, F., & Pillewizer, W. (1966). The Principles of Selection. *Cartographic Journal*, *3*(1), 10–16.

Touya, G. (2007). River Network Selection based on Structure and Pattern Recognition. In *ICC2007* (pp. 4–9). Moscow.

Trudeau, R. J. (1994). *Introduction to graph theory*. Dover Publications.

Twidale, C. R. (2004). River patterns and their meaning. *Earth-Science Reviews*, *67*(3–4), 159–218.

Van Dijk, S., Thierens, D., & De Berg, M. (2002). Using genetic algorithms for solving hard problems in GIS. *GeoInformatica*, *6*(4), 381–413.

Verdin, K. ., & Verdin, J. . (1999). A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, *218*(1-2), 1–12.

Visvalingam, M., & Whyatt, J. D. (1993). Line generalisation by repeated elimination of points. *Cartographic Journal, The*, *30*(1), 46–51.

Vogt, J. V, Colombo, R., & Bertolo, F. (2003). Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics. *Geomorphology*, *53*(3-4), 281–298.

Wang, G., Wu, B., & Li, T. (2007). Digital Yellow River Model. *Journal of Hydro-environment Research*, *1*(1), 1–11.

Wang, H. O., Tanaka, K., & Griffin, M. F. (1996). An approach to fuzzy control of nonlinear systems: stability and design issues. *IEEE Transactions on Fuzzy Systems*, *4*(1), 14–23.

Wang, L.-X. (1993). Stable adaptive fuzzy control of nonlinear systems. *IEEE Transactions on Fuzzy Systems*, *1*(2), 146–155.

Ware, J. M., Wilson, I. D., & Ware, J. A. (2003). A knowledge based genetic algorithm approach to automating cartographic generalisation. *Knowledge-Based Systems*, *16*(5), 295–303.

Weibel, R. (1995). Three essential building blocks for automated generalization. In *GIS and Generalization: Methodology and Practice.* (pp. 56–69). London: Taylor & Francis.

Weibel, R., & Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical Information Systems: Principles, techniques, management and applications* (pp. 125–155).

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65–85.

Wilmer, J., & Brewer, C. (2010). Application of the Radical Law in Generalization of National Hydrography Data for Multiscale Mapping. In *A special joint symposium of ISPRS Technical Commission IV & AutoCarto*. Orlando, Florida.

Wolf, G. W. (1988). Weighted surface networks and their application to cartographic generalization. In *Visualisierungstechniken und Algorithmen* (Vol. 182, pp. 199–212). Springer-Verlag.

Wu, H. (1997). Structured Approach to Implementing Automatic Cartographic Generalization. *18th ICC*. Stockholm, Sweden.

Yen, J., & Langari, R. (1998). Fuzzy logic: intelligence, control, and information.

Ying, H., Siler, W., & Buckley, J. J. (1990). Fuzzy control theory: A nonlinear case. *Automatica*, *26*(3), 513–520.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353.

Zhai, R. J., Wu, F., Deng, H., & Tan, X. (2006). Automated Elimination of River Based on Multi-Objective Optimization Using Genetic Algorithm. *Journal of China University of Mining & Technology*, *35*(3), 403–408.

Zhan, F. B., & Noon, C. E. (1998). Shortest Path Algorithms: An Evaluation Using Real Road Networks. *Transportation Science*, *32*(1), 65–73.

Zhang, X. (2012). *Automated evaluation of generalized topographic maps*. The University of Twente.