A genetic algorithm for tributary selection with consideration of multiple factors

Ling Zhang[1, 2], Eric Guilbert[3]

[1] Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, P. R. China

[2] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, P. R. China

[3] Dept. of Geomatic Sciences, Laval University, Quebec G1V 0A6, Canada

The running head: A genetic algorithm for tributary selection

Keywords: Map Generalization; River network; Tributary selection; Drainage pattern; Genetic algorithm.

Corresponding author: Ling Zhang, lingzhang.sky@gmail.com, School of Geography Science, Nanjing Normal Univ., No. 1 Wenyuan Road, Qixia District, Nanjing 210023, P. R. China.

Abstract

Drainage systems are important components in cartography and Geographic Information System (GIS), and achieve different drainage patterns based on the form and texture of their network of stream channels and tributaries due to local topography and subsurface geology. The drainage pattern can reflect the geographical characteristics of a river network to a certain extent. In order to preserve drainage pattern during the generalization process, this paper proposes a solution to deal with multiple factors, such as the tributary length and the order altogether in river tributary selection. This leads to a multi-objective optimization problem solved with a Genetic Algorithm. In the multi-objective model, different weights are used to aggregate all objective functions into a fitness function. The method is applied on a case study to evaluate the importance of each factor for different types of drainage and results are compared with a manually generalized network. The result can be controlled by assigning different weights to the factors. From this work, different weight settings according to drainage patterns are proposed for the river network generalization.

1 Introduction

In both Geographic Information System (GIS) and terrain analysis, drainage systems are important components. Due to the 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 is "*the arrangement in which a stream erodes the channels of its network of tributaries*" (Chernicoff & Whitney, 2006). It can reflect the geographical characteristics of a river network to a certain extent because it depends on the topography and geology of the land. Whether in cartography or GIS, hydrography is one of the most important feature classes to be generalized to produce representations at various levels of detail. In general, there are two typical operations in river network generalization: tributary selection and scale-driven generalization (Li, 2007). There are many methods for tributary selection, but few of them consider the drainage pattern in the first place. Tributaries are selected based on the stream order and on local parameters such as their length or catchment area. Drainage pattern and other global factors measured on the network have not been considered before in the process.

Recently, many researchers have paid more attention on geospatial patterns in cartographic generalization (Heinzle et al., 2006; Mackaness & Edwards, 2002; Zhang, 2012). Drainage classification based on their patterns was introduced by Zhang and Guilbert (2013). This paper proposes a river network generalization method with consideration of different factors according to the drainage pattern. For that purpose, a Genetic Algorithm (GA) is designed and implemented for tributary selection. The method is applied to networks following different patterns and the importance of each factor is evaluated in each case so as to provide a proper weight setting for each drainage pattern in river tributary selection.

The remainder of the paper is organized as follows. *Section* 2 reviews related work about tributary selection. In *Section 3*, a tributary selection model is presented with consideration of different factors, and the objective function is provided. *Section 4* introduces basic concepts of GA and explains how they are applied to tributary selection. In *Section 5*, the selection method is applied for each type of pattern, and results showing the importance of different factors are analyzed. *Section 6* is the conclusion and the last section is limitations and future work.

2 Related work

Tributary selection consists, in river network generalization, in keeping or removing river segments according to their importance and the scale of the map. Rusak Mazur and Castner (1990) gave four possible options for the selective elimination of river tributaries based on the number of tributaries to be kept. Richardson (1993) presents a method to select rivers based on the Horton order (Horton, 1945) and the river length. Thomson and Brooks (2000) apply the *Gestalt* recognition principles to river network generalization to emphasize the main channels and omit less important channels. A mainstream is detected based on strokes using their Horton order and their length but determining the main stream using the longest path on clipped river network leads to errors. Touya (2007) presents a method that relies on the organization of river strokes in a hierarchy. His work allows the building of strokes on a clipped area where some sources are not natural, such as irrigation zones. However, it only focuses on the geometric factors of river networks, and it does not simplify the river network with consideration of geomorphologic structures.

As the structure and shape of a river is constrained by the underlying terrain, several authors developed generalization supported structures integrating terrain information to add knowledge in the selection process. Wolf (1988) builds a weighted network data structure to determine the significance of a river. Different from Horton's work, the weight of surface networks takes pits, passes and peaks together with the connecting ridges and courses into consideration. Wu (1997) investigates the characteristics of the river tree and develops a method based on spatial buffer analysis to establish the river tree structure. Ai et al. (2006) present a selection method where the importance of a channel is not defined by the geometric characteristics of the river stream but by the area of its watershed. Density and upstream drainage area are also used to prune the river network (Stanislawski, 2008, 2009). In the case of man-made ditches, Sandro et al. (2011) present a typification method for generalization of groups of ditches, which are represented as regular patterns of straight lines. In order to consider different geographical factors, such as river length, river tributaries spacing, catchment area, and river network density, there is a need for a multi-objective optimization (also known as multi-criteria or multi-attribute optimization) process in river tributary selection. Zhai et al. (2006) built a river data structure model representing the river system's spatial knowledge, and selected the river tributaries automatically based on a genetic multi-objective optimization algorithm. In their model, indicators such as the river length, the river importance and the distance between proximity rivers[1] were taken into account during the selection.

Although these works mentioned above take into account further parameters in river network generalization, they are only applied locally to express knowledge at river segment level. The structure of a river network is the result of complex geomorphologic processes that shaped the terrain and so the pattern exhibited by a network provides knowledge at a more global level that should be considered and preserved during the generalization process. However, little research has been done 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 consider it are given. In order to maintain the main hydrographical properties, Jiang et al. (2009) just present a simple result of river networks preserving the patterns after a selection operation but they did not go further to explain how these patterns were preserved. In different drainage patterns, different factors should be considered during the river network generalization. Sen et al. (2014) proposed a method using self-organizing maps (SOMs) for the selection of hydrographic model generalization. Geometric attributes, such as length and sinuosity, and topologic attributes, such as degree, betweenness and closeness, are used as input variables to the SOM. The method is tested on dendritic and modified basic pattern composed mostly of trellis (grid-like structures) and in part of rectangular networks, but the drainage pattern is not taken into account.

We can see that much work has been done on river networks generalization relating to tributary selection. However, most of them focus on geometric properties only, and do not consider the pattern in the first place. Considering that the drainage pattern is an important piece of information to preserve in river network generalization, a specific method adapted to the drainage may be designed. As the pattern can be characterized by factors such as the tributary balance and spacing (details are in *Section 3.1*), this

[1] Proximity rivers are adjacent rivers on the same side of a main stream (Ai et al., 2006).

paper presents a multi-objective tributary selection method where the importance of each factor depends on the drainage pattern. The method is based on a genetic algorithm in order to optimize the selection.

3 Tributary selection modeling

This section presents a list of factors considered in the description of each drainage. Each factor is characterized by a value assigned by an objective function. These functions are later combined into a multi-objective function.

3.1 Geometric factors and objective functions

3.1.1 Drainage pattern membership

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 several drainage patterns, such as dendritic, parallel, and trellis. 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 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). 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).

In this research, these three drainage patterns are considered and tested. A list of characteristics for each of them is proposed and shown in *Table 1*.

Attributing a pattern to a network is a subjective operation as it is based on a combination of qualitative characteristics. Zhang and Guilbert (2013) proposed a fuzzy logic approach in which a membership degree for each pattern is assigned to a 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 objective function of the drainage pattern membership can be given as follows:

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

where $\alpha, \beta, \gamma,$ and δ are the average junction angle, the bended tributaries percentage, the average length ratio and the catchment elongation respectively, and $F_M \in [0,1]$. The details of drainage pattern membership value are shown in our previous work (Zhang & Guilbert, 2013).

3.1.2 Stream order

The stream order is a way to define the size of perennial and recurring streams based on a hierarchy of tributaries. There are several ordering schemes. The Horton-Strahler scheme (Strahler, 1957) and the Shreve scheme (Shreve, 1966) are the most famous ones. In this paper, the Horton-Strahler order after upstream routine, which is the process to determine the main stream (Li, 2007), will be used for tributary selection as it can provide a generalized river network close to human-made (Rusak Mazur & Castner, 1990).

In river network generalization, the selection operation, in general, starts from tributaries at lower order. Tributaries at larger orders have higher opportunity to be shown on the map after selective omission. So, the objective function of the stream order is designed to evaluate the stream order of all network elements:

$$F_o = \sum_{i=1}^{n} O_i, \quad 0 < n \le N, \tag{2}$$

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

3.1.3 Stream length

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

$$L = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}, \tag{3}$$

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

The stream length factor implies in a certain extent that a longer tributary is more important. In order to select longer tributaries preferentially, the following objective function F_L aims at maximizing the length value of all selected rivers.

$$F_L = \sum_{i=1}^{n} L_i, \quad 0 < n \le N, \tag{4}$$

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*.

3.1.4 Balance coefficient

In order to avoid that only tributaries on one side of a river are eliminated, the tributary balance between two sides of a river should be maintained. 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:

$$B = \begin{cases} 1, & \sum_{i=1}^{m} L_i = 0 \text{ and } \sum_{j=1}^{n} L_j = 0 \\ \sum_{i=1}^{m} L_i \Big/ \sum_{j=1}^{n} L_j, & \sum_{i=1}^{m} L_i \le \sum_{j=1}^{n} L_j \\ \sum_{j=1}^{n} L_j \Big/ \sum_{i=1}^{m} L_i, & \sum_{i=1}^{m} L_i > \sum_{j=1}^{n} L_j \end{cases}, \tag{5}$$

where m and n are the numbers of tributaries on 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 $(1 \le j \le n)$.

The calculation of the balance coefficient shows that $B \in [0,1]$. $B = 1$ corresponds to a river that receives as much water from both side. The objective of the balance coefficient is to maintain the balance after generalization. Therefore, the objective function of the balance coefficient is defined by the Gaussian function as follows,

$$F_B = \sum_{i=1}^{m} g(B_i'; 0.1, B_i) \Big/ m \,, \tag{6}$$

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.

3.1.5 Tributary spacing

Tributary spacing is the distance between two adjacent tributaries which are on the same side of a main stream. As 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:

$$S \approx D(A,B) = \min_{a \in A} \left\{ \min_{b \in B} \left\{ d(a,b) \right\} \right\}, \tag{7}$$

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

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

$$F_s = \min(S_i), \quad i = 1, 2, \ldots k \,, \tag{8}$$

where k is the number of spacing of tributaries after selection, and S_i is the tributary spacing of tributary i. This function should be maximized to avoid tributaries clustering together.

3.2 Multi-objective modeling with consideration of the drainage pattern

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

$$F(X) = \sum_{i=1}^{k} w_i F_i(X) \,, \tag{9}$$

where k is the number of objective functions; w_i is the weight of each objective function F_i, and the weights satisfy the requirement of $\sum_{i=1}^{k} w_i = 1$. As the magnitude of each objective function may be different, they shall be rescaled, and the final formula is as follows:

$$\bar{F}(X) = \sum_{i=1}^{k} w_i F_i^*(X), \qquad (10)$$

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

$$F_i^*(X) = (F_i(X) - F_i^{\min}) / (F_i^{\max} - F_i^{\min}). \qquad (11)$$

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

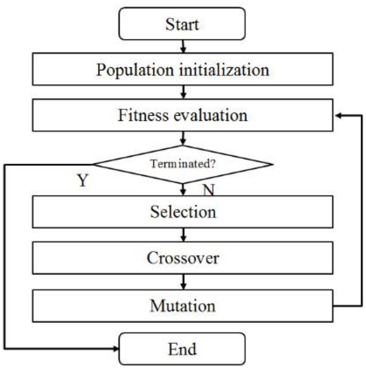
$$F(X) = w_M F_M(X) + w_O F_O^*(X) + w_L F_L^*(X) + w_B F_B(X) + w_S F_S^*(X), \qquad (12)$$

where $w_M + w_O + w_L + w_B + w_S = 1$, w_M, w_O, w_L, w_B and w_S are weights for drainage pattern membership, stream order, stream length, balance coefficient and tributary spacing respectively.

4 Tributary selection using a genetic algorithm

Optimizing river selection according to different factors at the same time is a multi-objective optimization problem. A genetic algorithm (GA) is a class of adaptive stochastic optimization algorithms that simulates the process of natural evolution, and is used to find available solutions to optimization and search problems (Mitchell, 1996). Van Dijk, Thierens, and De Berg (2002) showed that GAs can solve GIS problems, such as map labeling, generalization and line simplification. Ware, Wilson, and Ware (2003) applied GA to solving spatial conflict between map objects after scaling.

In GA, the solution (called individual) to the problem is 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). An initial set of solutions called population is first generated, and genetic operators such as selection, crossover and mutation are applied to generate new solutions in order to find the best one(s) by evaluating the fitness of every individual in the population. The process of a basic GA



is shown in the *Figure* 1.

4.1 Encoding of a river network

In the proposed method, a chromosome corresponds to a generalized river network. The chromosome is composed by genes and each gene is associated to a tributary in the network. A gene is set to 1 if the tributary is selected in this network and to 0 if it is not. Following Thomson and Brooks (2000), tributaries are defined by strokes in the network and ordered according to the Horton-Strahler order scheme after upstream routine.

In *Figure 2(a)*, a simulated river network is illustrated, where the number is the ID of a stroke. In *Figure 2(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. As strokes must remain connected to the network, a stroke cannot be omitted if strokes at higher order are selected.

The process requires first to fix the number of strokes to be selected and is initialized by generating a population as possible solutions. At each step, a new population is generated through reproduction and the best chromosomes are selected so as to get a population containing better solutions.

4.2 Initialization

In the initialization process, the number of selected strokes (noted *Ns*) can be calculated by the "*Radical Law*" (Topfer & Pillewizer, 1966) or other methods. As the number of strokes to select is fixed beforehand and the consistency of the network must be preserved, we define the following rules:

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, i.e., 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.

These rules translate into the following algorithm 1 which is applied to generate each chromosome of the initial population.

Algorithm 1- Initialization of a chromosome

```
Input: La the list of all strokes; Ns the number of strokes to be selected
Output: Ls the list of selected strokes
Set Lo empty list of strokes
Set Na = card(La) // total number of strokes
Set No = Na – Ns // number of strokes to be omitted
Set Nc = 0 // number of strokes currently omitted
Set Lt = La
While Nc<No
   Randomly choose a stroke Rs from Lt
   Set Ns' the number of upper strokes of Rs (including Rs)
   If Nc + Ns' > No then // Rs cannot be omitted
      Remove Rs from Lt
   Else
      Nc = Nc + Ns'
      Add Rs and its upper strokes to Lc
      Remove Rs and its upper strokes from Lt
   End if
End while
Ls = La – Lc
```

4.3 Selection

In this paper, the elitist model (Mitchell, 1996) is used for the selection operation in the GA. Elitism directly copies the best chromosomes 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.

4.4 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, and cannot omit a stroke if it has upper strokes.

4.4.1 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 of the same length, is used to determine which genes are inherited from which parents. An offspring is generated as indicated in the mask (see also Figure 2): a gene is from the first parent chromosome if the mask gene is 1 and from the second parent if it is 2. The crossover algorithm is given by algorithm 2.

Algorithm 2 – Chromosome crossover

```
Input: Two chromosomes C1 and C2
Output: the offspring chromosome O
Set the mask M as a list of Na values set to 1
Set Lp as an empty list
For i = 1 to Na
   If C1(i) != C2(i) // genes have different values in C1 and C2
      Add i to Lp
For each i of Lp
   If changing the value of C1(i) breaks the network topology
      Remove i from Lp
While Lp is not empty
   Randomly select i1 and i2 in Lp such that C2(i1) != C2(i2)
   If no i1 and i2 can be selected Then
      Exit While
   Set M(i1) = 2
   Set M(i2) = 2
   Remove i1 and i2 from Lp
End While
For i = 1 to Na
   If M(i) = 1
      Set O(i) = C1(i)
   Else
      Set O(i) = C2(i)
```

Figure 2(c) shows an example of crossover process with two chromosomes. *C1* and *C2* are parents and O is the chromosome obtained by crossover. Genes 1, 3 and 8 of C1 and C2 are identical and so are copied directly to O. Other six allelic genes 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. *M* is one of the workable masks determined by random selection. In order to follow rule 1, positions 2 and 7 in the chromosome are marked to exchange information between the parents. A new offspring is generated by taking genes in positions 2 and 7 from *C2* and other genes from *C1*.

4.4.2 Mutation

The mutation operation changes a gene in a chromosome in order to avoid a too early convergence towards a local instead of a global optimum. Here, if only one gene is changed, the rule 1 of the reproduction cannot be satisfied. If a gene is changed from 0 to 1, another gene needs to be changed from 1 to 0. In the process, only one pair of genes is supposed to be changed.

The modified mutation for selecting tributaries is given in the following algorithm.

Algorithm 3 – Chromosome mutation

Input: a chromosome *C*
Output: the mutated chromosome *C*
Set *L1* the list of genes valued to 1 // list of selected tributaries
Set *L0* the list of genes valued to 0 // list of omitted tributaries
For each *g* in *L1*
    If setting *g* to 0 breaks the topology
        Remove *g* from *L1*
Randomly choose a gene in *L1* and set it to 0
For each *g* in *L0*
    If setting *g* to 1 breaks the topology
        Remove *g* from *L0*.
Randomly choose a gene in *L0* and set it to 1

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

4.5 Termination

The GA process does not stop until a termination condition is satisfied. For the termination of this problem, two methods are used. The first one sets the number of iterations. The second sets a number of iterations where the best chromosome does not change. In the experiments, both termination conditions are set to stop the GA process.

5 Experiments and results

In the experiment, datasets of the Russian river, California are tested (*Figure 3*). Data at 1:24K are provided by the Russian River Interactive Information System (RRIIS[1]) while data at 1:100K were obtained from the National Hydrography Dataset (NHD[2]). The river flow dataset at 1:100K scale was not generalized automatically from detailed data. It can be regarded as a manually generalized dataset. Therefore, the NHD data at 1:100K scale is used as a standard to check generalized river networks by comparing the similarity, which is calculated by an overlap ratio.

Similarity is a complicated problem (Lv et al., 2012). In our experiment, as the number of selected segments is fixed, we look only at segments common to both

[1] http://www.rrwatershed.org

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

datasets. In addition, the segment length participated was considered in the overlapping ratio to give more importance to long segments. 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

$$Similarity = \left(\sum_{i=1}^{M} Len_i \Big/ \sum_{j=1}^{N} Len_j \right) \times 100\% , \tag{13}$$

where Len_x is the length of a river segment. An overlap example is shown in *Figure 4*. The experiment is conducted as follows:

1) Get a sub-network from Russian river at 1:24k, 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, 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 for 20 generations. These settings are empirical values. The algorithm runs faster if the population size is small, but this more often leads to a local optimal solution. However, it is not good to set the population too large because it will slow the algorithm and the result would not be better than a suitable population size.

There are several objectives to achieve in the experiment: one is to test the importance of each factor by setting a weight to 0 and others are 0.25, and another is to rank the factors in the order of importance for the multi-objective function according to different drainage patterns by setting a weight of a factor to 0.6 and others to 0.1. So, different weights are set in the GA. 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.

As the stroke and length method was regarded as the method that "*most closely approximates the generalisation decisions made by a human cartographer*" (Thomson & Brooks, 2000), it is applied in the experiment for comparison. Taking a stroke as an entity, there are two steps in the generalization process: ①remove the lower order stroke first; ②remove the shorter strokes if they are in the same order.

Different sets of weights have been set for each drainage pattern. These weights were tested by performing a selection on each sub-network of the Russian river. Appendix A presents the detailed results for dendritic networks. Appendix B presents commented results obtained for trellis and parallel networks. A summary of best results obtained for each pattern is presented in *Table 2*.

In *Table 2*, for all patterns, the length is the most important factor, because the length weights (w_L) are 0.6, 0.6 and 0.5 for the dendritic, trellis and parallels patterns respectively. They are much bigger than other weights. In the dendritic pattern, the pattern membership and the order are the second important factors (w_M = 0.15, w_O = 0.15); 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 (w_S = 0.2), 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. Hence, the order should not be considered as an important factor for the preservation of trellis. In the parallel pattern, the membership factor (w_M = 0.2) is more important than others except the length. Compared to the stroke and length method, the GA method provides a better network both in similarity and drainage pattern preservation.

6 Conclusion

In this paper, 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 geometric 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 pattern membership (w_M). If the weight of river order (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 avoid tributaries to cluster together.

The proposed GA 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 and test the importance of the drainage pattern. The most important factor is the length. In general, during manual generalization, the length is indeed the most considerable factor. Although the drainage pattern does not change much in the manual work, it can be preserved better if the pattern membership participates in the GA process.

For each pattern, a proper set of weights is given to achieve a greater similarity with the manual generalized river networks. *Table 3* illustrates the approximate settings for weights.

7 Limitations and future work

One limitation of the research is that the similarity is not improved obviously comparing to the stroke and length method. There are some reasons. ①The 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. ②Some 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 in the geography so it should be preserved whether it is short or long. Some examples are shown in *Figure 5*. Another limitation is that the only involved map scales in this paper are 1:24K and 1:100K scales. Other datasets and other scales should be tested in future work.

Further from this, some work still can be done to improve this study. The first one is that other factors can be considered. 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 in the selection process according to their geographical meanings. The second one is that rectangular drainages (*Figure 6*) should also be considered. The third one is to provide a method to calculate the weights of objective functions. This work requires more datasets at different scales. Apart from cartography, drainages are also important in studying the geomorphology of an area. Characterizing and preserving patterns can be useful in applications where terrain characteristics are studied at different scales.

8 Acknowledgements

This research was supported by the National Natural Science Foundation of China (41171350, 41501496), and Jiangsu Planned Projects for Postdoctoral Research Funds (1402061B).

9 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.

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

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

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.

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.

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.

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

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

Lv, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics Reports*, *519*(1), 1–49. doi:10.1016/j.physrep.2012.02.006

Mackaness, W., & Edwards, G. (2002). The importance of modelling pattern and structure in automated map generalisation. In *Joint ISPRS/ICA Workshop on Multi-Scale Representations of Spatial Data* (pp. 7–8). Ottawa.

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

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. Retrieved from http://www.earthonlinemedia.com/ebooks/tpe_3e/title_page.html

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. In A. Ruas (Ed.), *Advances in Cartography and GIScience* (Vol. 1, pp. 425–437). Paris: Springer Berlin Heidelberg.

Sen, A., Gokgoz, T., & Sester, M. (2014). Model generalization of two different drainage patterns by self-organizing maps. *Cartography and Geographic Information Science*, *41*(2), 151–165. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/15230406.2013.877231

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

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.

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

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.

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

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.

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

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. In *18th ICC*. Stockholm, Sweden.

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.

Zhang, L., & Guilbert, E. (2013). Automatic drainage pattern recognition in river networks. *International Journal of Geographical Information Science*, *27*(12), 2319–2342.

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

Appendix

A. Dendritic example

A tested dendritic river network is shown in *Figure 7*. The result of this example is given in *Table 4*. 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 participating objective functions at each generation during the process. The value of the stream order function is decreasing, and others are increasing. This is because a selection solution is initialized based on stream order. Lower order streams are eliminated first, and then, after the GA operations, some streams with lower order would be selected back 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 4*, 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 length (w_L) > drainage pattern membership (w_M) $\approx$ order (w_O) > tributary spacing (w_S) $\approx$ balance (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 different values to fix the importance between w_M and w_O, and between w_S and w_B. For networks (K) and (M) in *Table 4*, although w_M and w_O are different, the 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 for network (C). For this reason, considering w_B has influenced the similarity a lot, it is set to 0 for networks (N), (O) and (P). The results show that $w_S > w_B$ because all similarities of (N), (O) and (P) have improved.

B. Generalized networks in Russian River

B.1. Generalized dendritic networks

The setting of weights for the fitness function used in *Appendix A* is also tested on all dendritic river networks in the Russian river. The whole river network can be decomposed in different sub-networks at different orders. According to the Horton-Strahler order of its main stream, a sub-network can belong to different orders from 2 to 4. The statistic result is listed in *Table 5*.

From *Table 5*, comparing the first five and the second five tests, it shows the same conclusions with the dendritic case study except for setting the weight of drainage pattern membership to 0 (Row 6). 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 in the manually generalized river networks, the type of pattern is preserved to some extent but is not emphasised. On the opposite, putting a heavy weight on the pattern membership function tends to increase this membership and to caricature the pattern. From the results, the similarities can be improved by setting length factor (w_L) with high values. It shows that the length is the most important factor among all five proposed factors. Weight settings for the pattern membership (w_M), order (w_O), length (w_L), balance (w_B) and tributary spacing (w_S) with 0.15, 0.15, 0.6, 0 and 0.1 get the greatest average similarity (80.4%) among all settings. The result is same with the case study, so it can confirm that the settings are more appropriate for dendritic river networks. The average pattern 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 (0.15/0.15/0.6/0/0.1) is only 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.

B.2. Generalized trellis networks

For all trellis river networks in Russian river, the statistic results which are tested by different weight settings are listed in *Table 6*.

From *Table 6*, 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 increased after generalization by assigning drainage pattern membership(w_M) with 0.6 and others with 0.1. Although the average similarity is low, the pattern has been preserved a lot from the average drainage pattern membership values in row 1. So, it is useful to take the factor of membership into consideration to preserve or even improve the drainage pattern representation during the generalization process. From the last five tests (rows 11-15), the average similarity is bigger than the method of the stroke and length. The multi-objective method does 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.

B.3. Generalized parallel networks

In the dataset, there are no parallel networks in order 4, and most networks are in order 2. The statistic result is listed in *Table 7*.

In *Table 7*, comparing with the result by assigning the weight of balance (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 contributes to the similarity to some extent. 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 the weight of length (w_L) with 0.5 and adjusting weights of other factors, the pattern membership should be set up as the second important factor. 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. 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.

Table 1. Drainage pattern characteristic. There are three patterns addressed in this research. The geometric and topologic characteristics are summarized from their definitions in Ritter (2006)'s book.

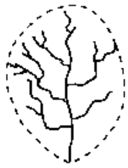
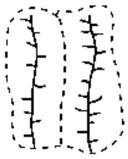
Drainage pattern	Schematic Diagram	Geometric and Topologic Characteristic
Dendritic		-Tributaries joining at acute angle
Parallel		- Parallel-like - Elongated catchment - Long straight tributaries - Tributaries joining at small acute angle
Trellis		- Short straight tributaries - Tributaries joining at almost right angle

Table 2. The best weight settings from the experiment results. The experiment details are show in the appendix. w_M, w_O, w_L, w_B and w_S are the weights of drainage pattern membership, order, length, balance and tributary spacing respectively.

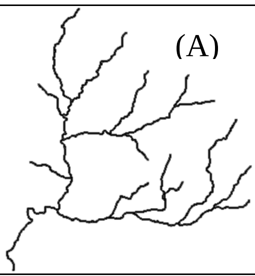
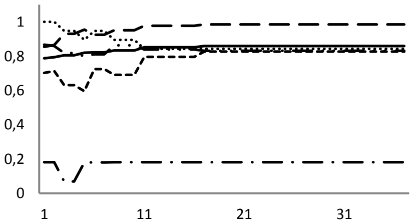
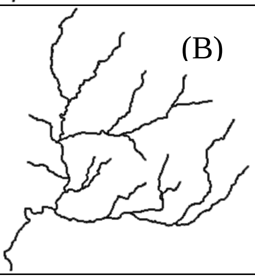
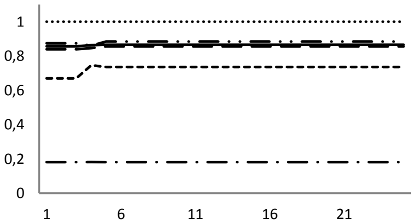
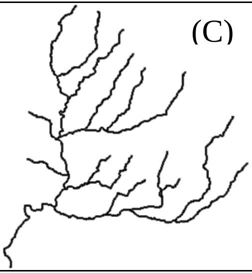
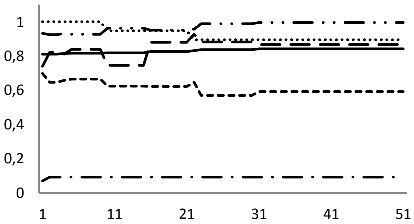
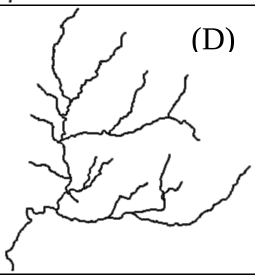
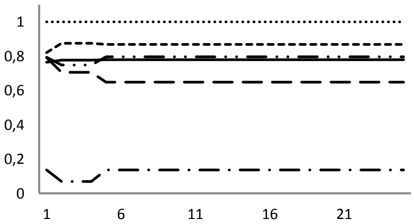
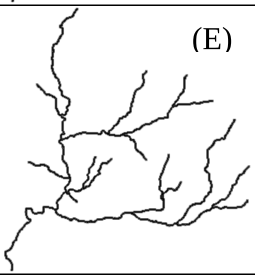
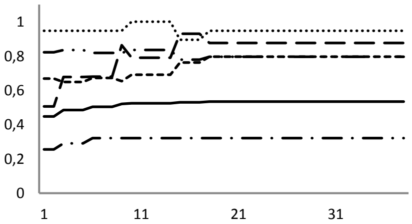
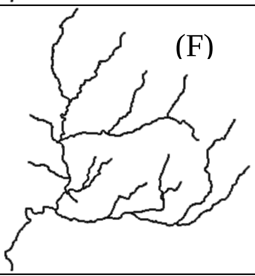
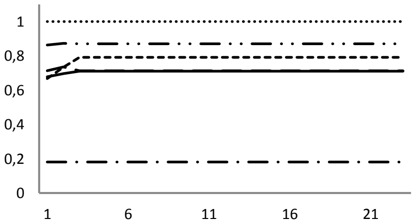
Drainage Pattern	Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity				Average Dranage Pattern Membership			
		Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
Dendritic	(0.15/0.15/0.6/0/0.1)	78.4%	88.4%	83.0%	80.4%	0.33	0.53	0.48	0.37
	Stroke + Length	77.5%	86.4%	78.4%	79.2%	0.25	0.48	0.28	0.29
Trellis	(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
Parallels	(0.2/0.1/0.5/0.1/0.1)	64.7%	91.6%	-	66.9%	0.47	0.73	-	0.50
	Stroke + Length	62.9%	91.6%	-	65.3%	0.44	0.73	-	0.47

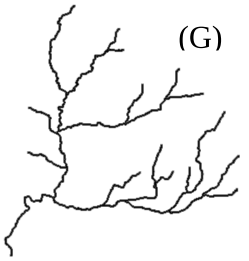
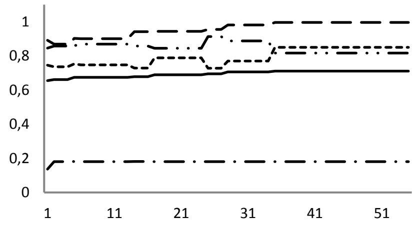
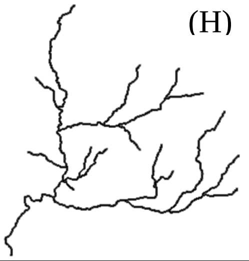
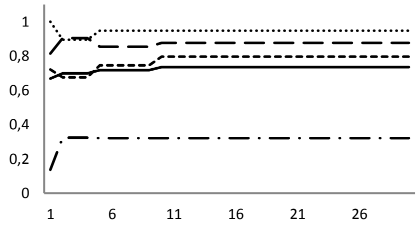
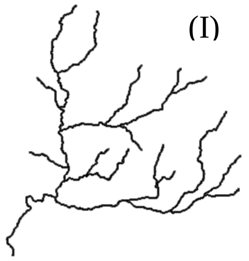
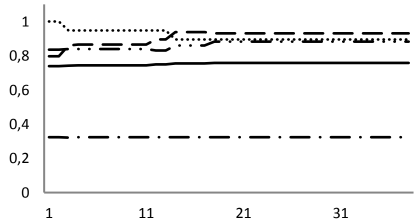
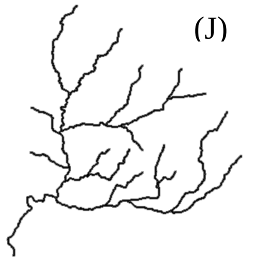
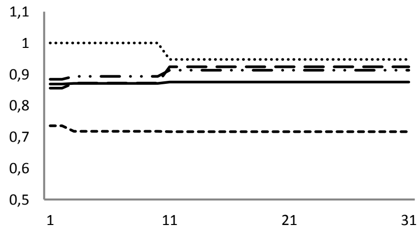
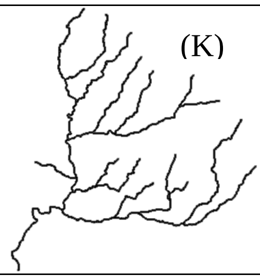
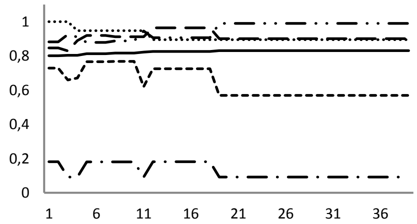
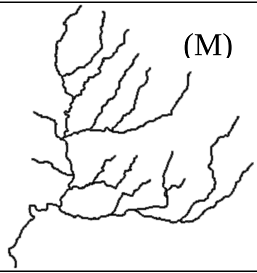
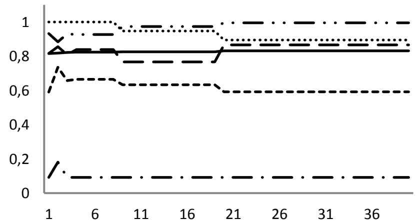
Table 3. The approximate weight settings for each drainage pattern. Balance coefficient cannot be considered in the dendritic and trellis patterns. The length is the most important factor for all patterns.

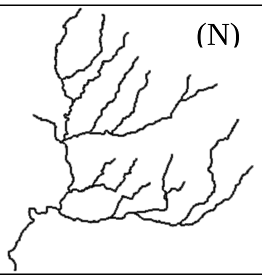
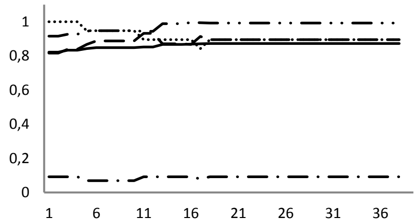
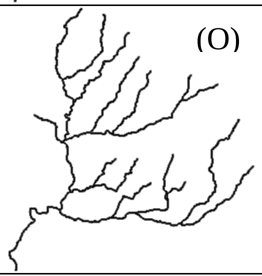
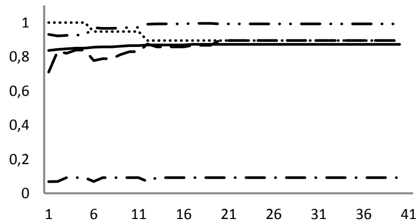
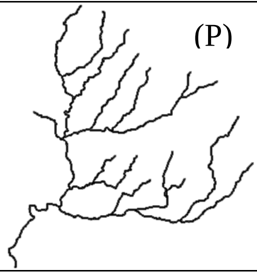
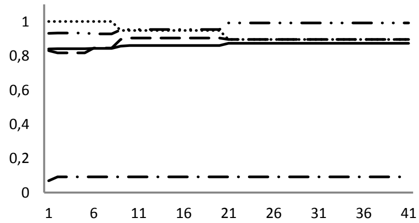
	Dendritic	Parallel	Trellis
Pattern membership (w_M)	●	●	○
Stream order (w_O)	●	●	○
Stream length (w_L)	●●	●●	●●
Balance coefficient (w_B)	×	○	×
Tributary spacing (w_S)	○	○	●

●● - more important ● - important ○ - not important × - not considered

Table 4. Generalized results for dendritic case. w_M, w_O, w_L, w_B and w_S are the weights of drainage pattern membership, order, length, balance and tributary spacing respectively.

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%
(0/0.25/0.25/ 0.25/0.25)	(F) 		65.6%

(0.25/0/0.25/ 0.25/0.25)			60.6%
(0.25/0.25/0/ 0.25/0.25)			61.7%
(0.25/0.25/ 0.25/0/0.25)			73.9%
(0.25/0.25/ 0.25/0.25/0)			69.9%
(0.2/0.1/ 0.5/0.1/0.1)			81.0%
(0.1/0.2/ 0.5/0.1/0.1)			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.

Legend: — · · — Length ------ Balance — · — Spacing ·········· Order — — Membership —— Fitness

Table 5. Generalized dendritic networks results. w_M, w_O, w_L, w_B and w_S are the weights of drainage pattern membership, order, length, balance and tributary spacing respectively.

	Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity				Average Drainage Pattern Membership			
		Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
1	(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
2	(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
3	(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
4	(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
5	(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
6	(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
7	(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
8	(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
9	(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
10	(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
11	(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
12	(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
13	(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
14	(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
15	(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
16	Stroke + Length	77.5%	86.4%	78.4%	79.2%	0.25	0.48	0.28	0.29
17	River networks at 1:24K scale from the Russian river					0.46	0.53	0.38	0.47
18	River networks at 1:100K scale from the NHD					0.28	0.43	0.27	0.31

Table 6. Generalized trellis networks results. w_M, w_O, w_L, w_B and w_S are the weights of drainage pattern membership, order, length, balance and tributary spacing respectively.

	Weights Setting (w_M/w_O/w_L/w_B/w_S)	Average Similarity				Average Drainage Pattern Membership			
		Order 2	Order 3	Order 4	Total	Order 2	Order 3	Order 4	Total
1	(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
2	(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
3	(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
4	(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
5	(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
6	(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
7	(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
8	(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
9	(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
10	(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
11	(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
12	(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
13	(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
14	(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
15	(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
16	Stroke + Length	78.0%	82.2%	84.0%	79.4%	0.20	0.40	0.86	0.28
17	River networks at 1:24K scale from the Russian river					0.21	0.18	0.03	0.19
18	River networks at 1:100K scale from the NHD					0.12	0.15	0.002	0.12

Table 7. Generalized parallel networks results. w_M, w_O, w_L, w_B and w_S are the weights of drainage pattern membership, order, length, balance and tributary spacing respectively.

	Weights Setting ($w_M/w_O/w_L/w_B/w_S$)	Average Similarity			Average Drainage Pattern Membership		
		Order 2	Order 3	Total	Order 2	Order 3	Total
1	(0.6/0.1/0.1/0.1/0.1)	61.5%	82.3%	63.2%	0.54	0.90	0.57
2	(0.1/0.6/0.1/0.1/0.1)	62.9%	79.4%	64.3%	0.49	0.69	0.51
3	(0.1/0.1/0.6/0.1/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
4	(0.1/0.1/0.1/0.6/0.1)	63.6%	85.1%	65.4%	0.48	0.58	0.49
5	(0.1/0.1/0.1/0.1/0.6)	61.8%	76.3%	63.0%	0.48	0.67	0.49
6	(0/0.25/0.25/0.25/0.25)	60.0%	73.8%	61.2%	0.45	0.26	0.43
7	(0.25/0/0.25/0.25/0.25)	62.9%	73.2%	63.8%	0.49	0.64	0.50
8	(0.25/0.25/0/0.25/0.25)	60.2%	60.5%	60.2%	0.51	0.52	0.51
9	(0.25/0.25/0.25/0/0.25)	62.9%	79.4%	64.3%	0.49	0.69	0.51
10	(0.25/0.25/0.25/0.25/0)	63.8%	91.6%	66.1%	0.49	0.73	0.51
11	(0.2/0.1/0.5/0.1/0.1)	64.7%	91.6%	66.9%	0.47	0.73	0.50
12	(0.1/0.2/0.5/0.1/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
13	(0.1/0.1/0.5/0.2/0.1)	62.9%	91.6%	65.3%	0.44	0.73	0.47
14	(0.1/0.1/0.5/0.1/0.2)	62.2%	91.6%	64.6%	0.45	0.73	0.47
15	(0.3/0.1/0.5/0.1/0)	64.0%	91.6%	66.3%	0.48	0.73	0.50
16	Stroke + Length	62.9%	91.6%	65.3%	0.44	0.73	0.47
17	River networks at 1:24K scale from the Russian river				0.56	0.66	0.56
18	River networks at 1:100K scale from the NHD				0.64	0.46	0.63

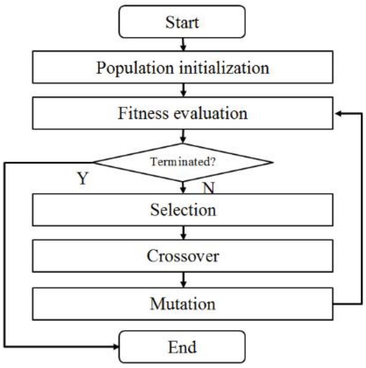


Figure 1. A Basic GA process.

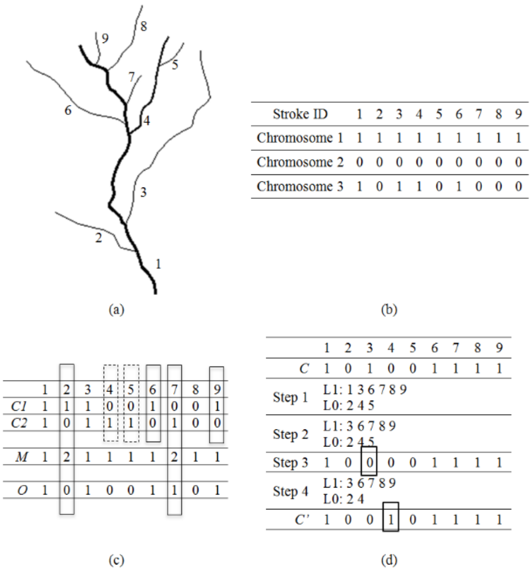


Figure 2. Tributary selection by GA. (a) An example of a river network, the numbers is the stroke IDs. (b) Some examples of chromosomes with binary encoding of a river network. (c) An example of crossover process. (d) An example of mutation process.

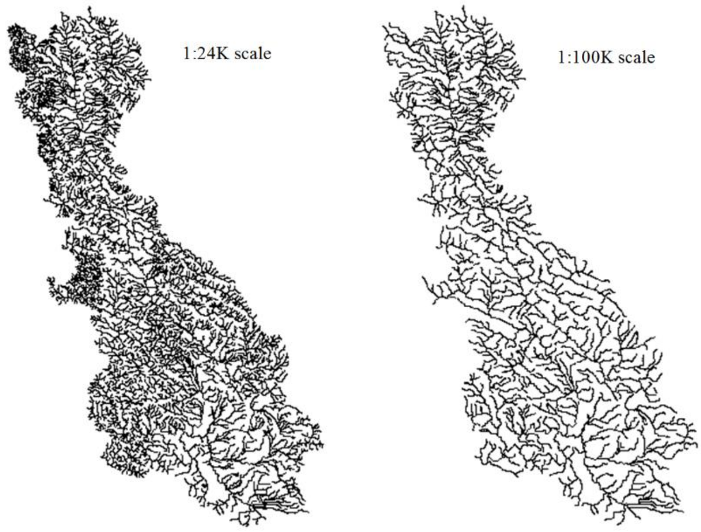


Figure 3. Russian river, California, USA. Left: 1:24,000-scale (1:24K) from RRIIS. Right: 1:100,000-scale (1:100K) from NHD.

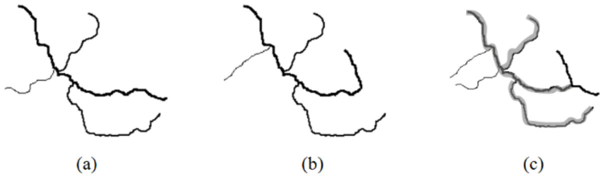


Figure 4. An example of overlap. Supposing network (a) is a generalized network and network (b) is a network from the NHD, the overlapped river segments are shown in bold gray shadow in (c). So, the similarity is the length of segments in shadow divided by the total length of network (b).

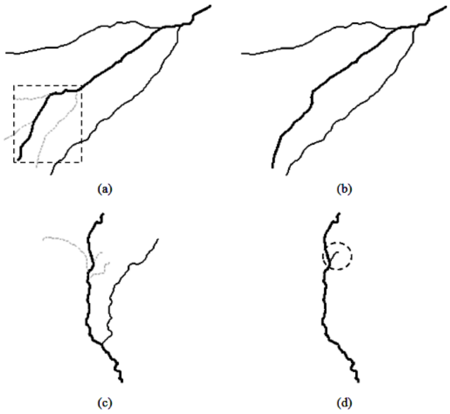


Figure 5. Unexpected situations, dashed lines are eliminated tributaries. Network (a) and (c) are from the Russian river at 1:24K scale. Network (b) and (d) are from the NHD at 1:100K scale. The bold line is the main stream obtained by the stroke. In the dashed box, the stroke is not the same as the stroke in network (b) because segments at 1:100k do not exist at 1:24k. No matter how the weights are adjusted, network (a) cannot be generalized as (b). Network (c) is a generalized network, where dashed tributaries are eliminated by considering the length. However, in the dashed circle, network (d) from the NHD set, a shortest tributary was selected.

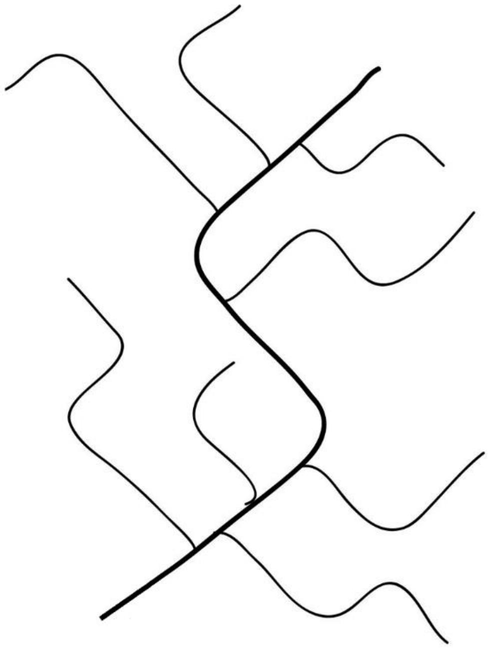


Figure 6. An example of rectangular drainage pattern.

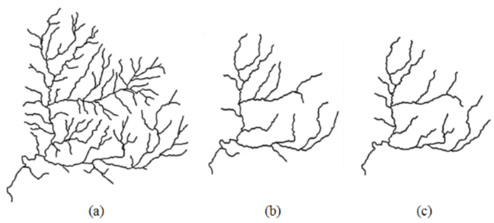


Figure 7. Tested network for dendritic case study. Network (a) is from Russian river at scale 1:24K. Network (b) is from the NHD at 1:100K scale. Network (c) is generalized by stroke and length method. It has the same number of strokes with the network from the NHD, and the similarity between them is 73.2%.