Hydrol. Earth Syst. Sci. Discuss., 7, 441–459, 2010
www.hydrol-earth-syst-sci-discuss.net/7/441/2010/

Hydrology and
Earth System
Sciences
Discussions

A novel algorithm with heuristic information for extracting drainage networks from raster DEMs

W. Yang[1,2], K. Hou[1,3], F. Yu[2], Z. Liu[4], and T. Sun[1]

[1]College of Computer Science, Northeast Normal University, Changchun, 130017, China
[2]College of Computer Science and Technology, Changchun Normal University, Changchun, 130032, China
[3]College of Computer Science and Technology, Jilin University, Changchun, 130012, China
[4]College of Urban and Environmental Sciences, Northeast Normal University, Changchun, 130024, China

Received: 29 December 2009 – Accepted: 6 January 2010 – Published: 20 January 2010

Correspondence to: T. Sun (suntl@nenu.edu.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Extracting drainage networks from raster DEMs is a necessary requirement in almost all hydrological and environmental analyses and determining surface water flow direction is a fundamental problem. In a raster environment, surface water flow direction of each cell can be directed to the neighboring cell with the steepest downslope drop (The basic D8, deterministic eight-neighbour method), which is inadequate for routing flow over pits and flats. Several improved algorithms are proposed to find the outlet of pits and flats, which typically use entirely different procedures for processing pits and flats. Being different from others, this paper presents a new method to route flow through the pits and flats by searching for the outlet using the heuristic information to compensate inadequate searching information of other methods. Heuristic information can reveal the general trend of the DEM and help the proposed algorithm find the outlet of pits and flats accurately. Furthermore, the proposed algorithm can handle pits and flats effectively in one procedure. This new algorithm is implemented in Pascal and experiments are carried out on actual DEM data. It can be seen from the comparison of the drainage networks generated by the proposed algorithm and ArcGIS 9.2, the proposed algorithm with heuristic information can get a closer match result with existing river networks and avoid the generation of the unrealistic parallel drainage lines, unreal drainage lines and spurious terrain features.

1 Introduction

A digital elevation model (DEM) in raster format, carries elevation information for many applications in hydrologic studies (Freeman, 1991; Vieux, 1993; Arun et al., 2005). Cell-wise elevation information in a DEM can be used to find the drainage structure, which gives the basic information for runoff analysis and sediment transport studies (Jana et al., 2007; Mandlburger, 2009).

The basic and important problem of hydrological and environmental analyses is to determine flow direction of cells in a raster DEM. The most commonly used algorithm is known as D8 (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Martz and Garbrecht, 1998; Lin et al., 2008), but the continuity of drainages does not extend across pits and flats.

Over the past 20 years, there have been many improved methods for routing flow through pits and flats. But when searching for the outlet, most of these methods checked only the eight adjacent cells of pit or flat and did not consider the general trend of the DEM.

In this paper, we present a novel algorithm to handle pits and flats with heuristic information. Heuristic information is often applied to finding the shortest path (Huang et al., 2007; Zeng and Church, 2009). While searching for the outlet, the proposed algorithm not only checks the eight adjacent cells of pit or flat, but also takes the general trend of the DEM into considerations.

2 Overview of the methods used for drainage delineation

Numerous methods have been proposed for automated drainage recognition in raster DEMs. The basic D8 (deterministic eight-neighbor method) algorithm is probably the most popular method (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Martz and Garbrecht, 1998). However, it has a number of deficiencies. In this method, the flow direction of each cell in a raster DEM is determined by the comparison between the cell's elevation and its eight adjacent neighbors, the cell with the maximum drop is identified as the flow direction. The generation of parallel flow lines in flat areas that restricts the formation of concentrated channel flow has been identified as a limitation of this method. In particular, pits and flats, both real and spurious present challenges for the automated derivation of fully connected drainage patterns (Kenny et al., 2008).

Pits and flats are areas where the cell's flow direction cannot be determined with reference to its eight neighbouring cells. A pit is defined as a point none of whose eight neighbors have lower elevations, flats are areas of level terrain (slope=0). Pits and flats can be naturally occurring but are generally viewed as spurious features that result from elevation data capture errors (Oimoen, 2000; Lindsay and Creed, 2005) or errors introduced during DEM generation (O'Callaghan and Mark, 1984; Band, 1986). Pits and flats can be found in all DEMs (Kenny et al., 2008).

In order to find the outlet, many methods are proposed to assign flow direction of either pit or flat. One of most common methods is proposed by Jenson and Domingue (1988). The method works in two stages. In the first stage, pits are "filled" by increasing the elevations of the nodes within each depression to the level of the lowest node on the depression boundary. The procedure essentially converts every depression into a flat area that has at least one external node on its border which has a lower elevation than the flat. The second stage assigns flow directions to the nodes in flat areas by an iterative procedure. In each iteration, nodes are assigned flow directions that point to a neighboring node if the neighboring node has a flow direction. Flow direction assignments iteratively grow into the interior of the flat from the outlet node(s).

DEM filling assumes that pits are created by an underestimation of elevation values within the pits. However, Outlet breaching method modifies DEM filling by assuming that the pits are created by an elevation overestimation at the obstruction (Martz and Garbrecht, 1999). Outlet breaching involves identifying the breach points for each pit and lowering the elevation values in these areas, subject to user-specified tolerances. An advantage of outlet breaching is that it typically involves making fewer alterations to a DEM than filling and much of the original interpolated elevation values, and therefore flow directions within a pit, are preserved (Kenny et al., 2008).

Imposing relief across flats is a different method for flat areas. Flats in a raster DEM are not exactly level in nature because DEM can not represent elevation difference less than vertical resolution (Martz and Garbrecht, 1993). The method introduced a relief algorithm to impose a finite gradient over the flats (Martz and Garbrecht, 1992). This method required less computer memory compared to the algorithm proposed by Jenson and Domingue (1988). First, this algorithm searched the perimeter of a flat for valid downslope cells (exit points). These exit points were used as a source and neighboring cells within the flat were assigned a small fixed elevation value slightly higher than the outlet cell(s). The process is repeated iteratively until the entire flat had been consumed, thereby creating a gradient across flats based on Euclidian distance from the outlets. In more advanced forms of relief imposition, a secondary shallow gradient was imposed to force flow from the higher terrain surrounding a flat (Garbrecht and Martz, 2000). This second gradient ensured that flow within the flat was consistent with the topography surrounding the flat surface and avoided parallel flow paths across flats that could be witnessed in the methods proposed by Jenson and Domingue (1988) and Martz and Garbrecht (1993) (Kenny et al., 2008).

A unique partial solution to the problem of pits in a raster DEM is to utilize the ANU-DEM (Australian National University DEM) thin-plate spline interpolation method (called TOPOGRID in ArcInfo) (Hutchinson, 1989, 1996, 1997). The method uses a flow-directed, single-line hydrology network as a boundary condition during interpolation, ensuring a downslope gradient for all cells along the hydrology network (Kenny et al., 2008).

A new method was developed for flow tracing, in which the pits were handled in a more realistic way (Tribe, 1992). Spurious pits and real pits were identified and treated separately. The method involved defining a main flow path through the flat to an outlet and directing the flow paths for nearby points in the flat towards the main path. However, because the main flow path was defined as a straight line, it could cut through an area of higher elevation (Jones, 2002).

A stochastic approach was proposed in order to solve the directional problem of the D8 algorithm (Fairfield and Leymarie, 1991). The method solved the directional limitation of D8, but it brought new problem, it was less efficient and could introduce undesirable feature in other areas (Jones, 2002). Multiple flow directions were also a method (Tarboton, 1997). Drawbacks of the method were the elimination of the unimodal link between the flow directions and the complications in defining the catchment boundary due to the multiple flow direction from a cell (Jana et al., 2007).

The algorithm proposed by Jones (2002) could effectively process both pits and flats. When processing a flat or pit cell the algorithm searched for a nearest cell with lowest elevation (exit point) and found a shortest drainage path between the two cells. After finding the exit point and shortest drainage path, the algorithm would lower the elevation of all points along the shortest drainage path to create a consistent gradient downslope drainage path between the original flat or pit point and the outlet point (CatchmentSIM, 2007). But in case of very large flat areas, the algorithm did not incorporate the aspect into consideration and flow was randomly directed to the neighbouring cell satisfying the priority criteria locally (Jana et al., 2007).

3 Proposed algorithm

3.1 Heuristic information

While searching for outlet of pit or flat, the methods described above only check the eight adjacent cells of pit or flat and do not consider the general trend of the DEM. In other words, they have no additional information about states beyond that provides in the problem definition. All they can do is generate successors and distinguish a goal state from a nongoal state. These methods can find the outlet, but they are incredible and inefficient in most cases. Unrealistic parallel drainage lines, unreal drainage lines and spurious terrain features are most likely to be generated.

DEM digitizes the continuous and complex three-dimensional surfaces at discrete points on a regular grid. In the digitizing processing, a certain amount of information is lost. The pits and flats could be created by overestimation and underestimation of elevation values. Checking only the eight adjacent neighbors of pit and flat can not find the optimal outlet.

Heuristic information about the state space can prevent algorithms from blundering about in the dark (Stuart and Peter, 2003). The proposed algorithm uses the heuristic information to find the optimal outlet.

3.2 Proposed method

3.2.1 Calculating the incipient flow direction using the basic algorithm

One of the keys to deriving hydrologic characteristics about a surface is the ability to determine the direction of flow from every cell in the raster DEMs. A 3-by-3 moving window is used over a DEM to locate the flow direction for each cell.

The direction of flow is determined by finding the direction of maximum drop from each cell to its eight adjacent neighbors. When a direction of maximum drop is found, the output cell is coded with the value representing that direction. If all eight adjacent neighbors are no lower than the central cell, the central cell is a pit or flat and has an undefined flow direction. The cell with an undefined flow direction is then processed by the algorithm proposed by this article.

3.2.2 Finding the optimal outlet of pit or flat using the proposed algorithm

After calculating the incipient flow direction, the cell with an undefined flow direction (pit and flat cells) need advanced processing with the proposed algorithm. The proposed method is based on heuristic information. The main purpose of the algorithm is to handle pits and flats with heuristic information in one procedure.

The method would require two data structure, closed list and open list. The closed list stores every checked node and the open list keeps the unchecked nodes.

The proposed algorithm uses the heuristic information to evaluate the importance of the node n. The heuristic information consists of two elements, actual cost and estimated cost, according to:

$$f(n) = g(n) + h(n) \tag{1}$$

Where, $f(n)$ is heuristic information of node n, $g(n)$ is actual cost, the difference of elevation between the starting node and the node n, $h(n)$ is estimated cost, the arithmetic average elevation of the cells in 5-by-5 window centred the node n.

The proposed algorithm is initialized by putting the cell with an undefined flow direction in the closed list. The first node added to the closed list is the starting node and all eight adjacent nodes are added to the open list. The nodes in the open list are evaluated by their heuristic information.

In turned, the node with the maximum heuristic information is added to the closed list and removed from the open list and its adjacent nodes are added to the open list. The iteration continues until a node in the open list satisfies the terminating criteria, which is an "outlet" for the drainage. The terminating criteria require only one node with the maximum heuristic information and it has a lower elevation than the starting node. Finally, the nodes in the closed list are traced from the outlet back to the starting node in the closed list and assign the flow direction to the node on the path from the starting node to the outlet.

While searching for the outlet of the cell with an undefined flow direction, checking only the nearest eight neighbors is not enough. Since a very small area cannot reveal the general trend of the DEM. The proposed algorithm always selects the node in the open list with the maximum heuristic information as the next node to be checked. This heuristic information ensures the proposed algorithm only tries to select nodes that most likely to lead to the direction towards the outlet. The heuristic information predicates the general trend of the slope, which help find the best outlet with least time. Schematic diagram of the proposed algorithm is shown in Fig. 1.

When all cells with undefined flow direction are processed by the proposed algorithm, we can get the flow direction matrix. Using the flow direction matrix as input, the flow accumulation matrix is determined. The drainage networks can be derived from the flow accumulation matrix by specifying a constant threshold.

4 Experimental results

To examine the suitability and performance of the proposed algorithm, we implement it in Pascal and make experiments on actual DEM data. For presentation purpose all raster maps are converted into vector format.

We select the SRTM for our study. The SRTM digital elevation data (Jarvis et al., 2008), produced by NASA originally, is a major breakthrough in digital mapping of the world, and provides a major advance in the accessibility of high quality elevation data for large portions of the tropics and other areas of the developing world. The SRTM 90 m DEMs have a resolution of 90 m at the equator. The SRTM digital elevation data has been processed to fill data voids and to facilitate its ease of use by a wide group of potential users.

The study area (Fig. 2), between longitude 67° to 68° E and latitude 34° to 35° N, locates in the middle of Afghanistan. The general elevation of the area varies from 2054 to 5048 m and 1200 rows by 1200 columns in the DEM. “Rivers Geospatial Data Presentation Form” (vector digital data) (USAID, 1990) are used to validate the results. Rivers depicts line river features from United States Defence Mapping Agency (USDMA; 1967–1988) 1:100 000 scale topographic maps.

Figure 3 shows the basic algorithm to extract drainage network from the study area which contains 1 440 000 nodes and 10 937 pits and flats with undefined flow direction. Due to the pits and flats, the basic algorithm fail to produce continuous flow lines and larger rivers are not delineated.

Figure 4 shows the drainage networks generated using ArcGIS 9.2 which is wide used by GIS practitioners. The hydrologic modelling functions in ArcGIS Spatial Analyst provide methods for extracting drainage networks form DEMs. Using the fill tool, flow direction tool and flow accumulation tool, we can creates a raster of accumulated flow to each cell by accumulating the weight for all cells that flow into each downslope cell. A constant threshold (500) is selected by trial and error. Flow accumulation data can be used to produce the final drainage networks when cells with values greater than the threshold with the con tool. It is obvious that major drainage could be delineated satisfactorily but the spurious parallel flow channels can be seen and the detailed information is lost. It is obvious in the red square area in Fig. 4. It is due to the error in selecting the flow direction of pits and flats.

Figure 5 shows the proposed algorithm for extracting drainage network from the study area. In order to compare to Fig. 4 the threshold is also 500. It can be seen from Fig. 5 that continuous drainage network and major drainage are delineated. The detailed information can be retrieved and the spurious parallel flow lines can not be found. In order to avoid these cases, the proposed algorithm takes heuristic information for hydrologic analysis and tries to capture the general trend of the study area. It is more obvious in Fig. 6.

The Fig. 6 shows the detail from the same square region of Figs. 4 and 5 along with the existing river networks (Rivers Geospatial Data Presentation Form) at the selected areas. It can be seen from the ten regions that the concentrated flow paths generated by the algorithm proposed by this article show a close match with the existing river networks. But the flow paths generated by the ArcGIS 9.2 show unrealistic parallel drainage lines, unreal drainage lines and spurious terrain features. From the comparison, it is be found that the proposed algorithm get better and closer match than the ArcGIS 9.2.

5 Conclusion and discussion

This paper presents a new method to flow through the pits and flats by searching for the outlet using the heuristic information to compensate inadequate searching information of other methods. Furthermore, the proposed algorithm can handle pits and flats effectively in one procedure.

The proposed algorithm has been tested with actual DEM data, and the experimental evidence suggests that the proposed algorithm can extract drainage networks better and fully than ArcGIS 9.2. From the comparison of the drainage network generated by the proposed algorithm and ArcGIS 9.2, it is be concluded that the extracted drainage network by the proposed algorithm cannot be entirely matched with the existing river networks but the propose algorithm shows a closer match and get better result than ArcGIS 9.2. We believe that if affecting factor of rainfall, soil, vegetation and others is added to the heuristic information, a closer match result may appear and this is our future work.

DEMs digitize continuous and complex three dimensional surfaces at discrete points on regular grid based on sampling data. However, sampling data are limited and DEMs are approximate simulation of real surfaces. Elevation sampling error, discrete sampling error, horizontal resolution, and vertical resolution may cause error of topography. Heuristic information can reveal the general trend of the DEM and help the proposed algorithm find the outlet of pits and flats accurately. The proposed algorithm is more appropriate for the processing of large-scale DEMs, which can be used for spatial decision-making with regard to large regional sustainable development projects, such as the planning of drainage areas, formulating responses to flood disasters, determining borderlines and other applications.

Acknowledgements. We would like to express my thanks and deepest appreciation to Xiangnan Liu. The work is supported by Science Foundation for Young Teachers of Northeast Normal University.

References

Arun, P. S., Jana, R., and Nathawat, M. S.: A rule base physiographic characterization of drought prone watershed applying remoter sensing and GIS, J. Indian Soc. Remote Sens., 32, 189–201, 2005.

Band, L. E.: Topographic partition of watersheds with digital elevation models, Water Resour. Res., 22(1), 15–24, 1986.

CatchmentSIM: http://www.toolkit.net.au/catchsim/Overview/PFSMethod/PFSMethod.htm, last access: 13 March 2007, 2007.

Fairfield, J. and Leymarie, P.: Drainage networks from grid digital elevation models, Water Resour. Res., 27, 709–717, 1991.

Freeman, T. G.: Calculating catchment area with divergent flow based on a regular grid, Comput. Geosci., 17, 413–422, 1991.

Huang, B., Wu, Q., and Zhan, F. B.: A shortest path algorithm with novel heuristic for dynamic transportation networks, Int. J. Geogr. Inf. Sci., 21, 625–644, 2007.

Hutchinson, M. F.: A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, J. Hydrol., 106, 211–232, 1989.

Hutchinson, M. F.: A locally adaptive approach to the interpolation of digital elevation models, in: Proceedings of the Third International Conference Workshop on Integrating GIS and Environmental Modelling, Santa Fe, NM, USA, National Center for Geographic Information and Analysis, Santa Barbara, CA, 1996.

Hutchinson, M. F.: ANUDEM Version 4.6, Users Guide, Canberra, Australia, p. 18, 1997.

Jana, R., Reshmidevi, T. V., Arun, P. S., and Eldho, T. I.: An enhanced technique in construction of the discrete drainage network from low-resolution spatial database, Comput. Geosci., 33, 717–727, 2007.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org/,last access: 13 October 2009, 2008.

Jenson, S. K. and Domingue, J. O.: Extracting topographic structures from digital elevation data from geographic information system analysis, Photogramm. Eng. Rem. S., 54, 1593–1600, 1988.

Kenny, F., Matthews, B., and Todd, K.: Routing overland flow through sinks and flats in interpolated raster terrain surfaces, Comput. Geosci., 34, 1417–1430, 2008.

Lin, W. T., Chou, W. C., Lin, C. Y., Huang, P. H., and Tsai, J. S.: WinBasin: Using improved algorithms and the GIS technique for automated watershed modelling analysis from digital elevation models, Int. J. Geogr. Inf. Sci., 22, 47–69, 2008.

Lindsay, J. and Creed, I. F.: Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEMs, Photogramm. Eng. Rem. S., 71, 1029–1036, 2005.

Mandlburger, G., Hauer, C., Höfle, B., Habersack, H., and Pfeifer, N.: Optimisation of LiDAR derived terrain models for river flow modelling, Hydrol. Earth Syst. Sci., 13, 1453–1466, 2009, http://www.hydrol-earth-syst-sci.net/13/1453/2009/.

Martz, L. W. and Garbrecht, J.: Numerical definition of drainage network and the sub-catchment areas from digital elevation models, Comput. Geosci., 18, 747–761, 1992.

Martz, L. W. and Garbrecht, J.: Automated extraction of drainage network and watershed data from digital elevation models, Water Resour. Bull., 29, 901–908, 1993.

Martz, L. W. and Garbrecht, J.: The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models, Hydrol. Process., 12, 843–855, 1998.

Martz, L. W. and Garbrecht J.: An outlet breaching algorithm for the treatment of closed depressions in a raster DEM, Comput. Geosci., 25, 835–844, 1999.

O'Callaghan J. F. and Mark, D. M.: The extraction of drainage networks from digital elevation data, Comput. Vision Graph., 28, 323–344, 1984.

Oimoen, M. J.: An effective filter for removal of production artifacts in US Geological Survey 7.5-minute digital elevation models, in: Proceedings of the 14th International Conference on Applied Geologic Remote Sensing, Las Vegas, NV, USA, Veridian ERIM International, Ann Arbor, MI, 311–319, 2000.

Richard, J.: Algorithms for using a DEM for mapping catchment areas of stream sediment samples, Comput. Geosci., 28, 1051–1060, 2002.

Russel, S. J. and Norvig, P.: Artificial Intelligence – A Modern Approach, Pearson Education, Inc., New Jersey, USA, 2003.

Tribe, A.: Automated recognition of valley lines and drainage networks from grid digital elevation models: a review and a new method, J. Hydrol., 139, 263–293, 1992.

United States Agency for International Development (USAID): Afghanistan Information Management Sevice(AIMS), http://www.aims.org.af/, last access: 13 October 2009, 1990.

Vieux, B. E.: DEM aggregation and smoothing effects on surface runoff modelling, J. Comput. Civil. Eng., 7(3), 310–338, 1993.

Zeng, W. and Church, R. L.: Finding shortest paths on real networks: the case for A*, Int. J. Geogr. Inf. Sci., 23, 531–543, 2009.

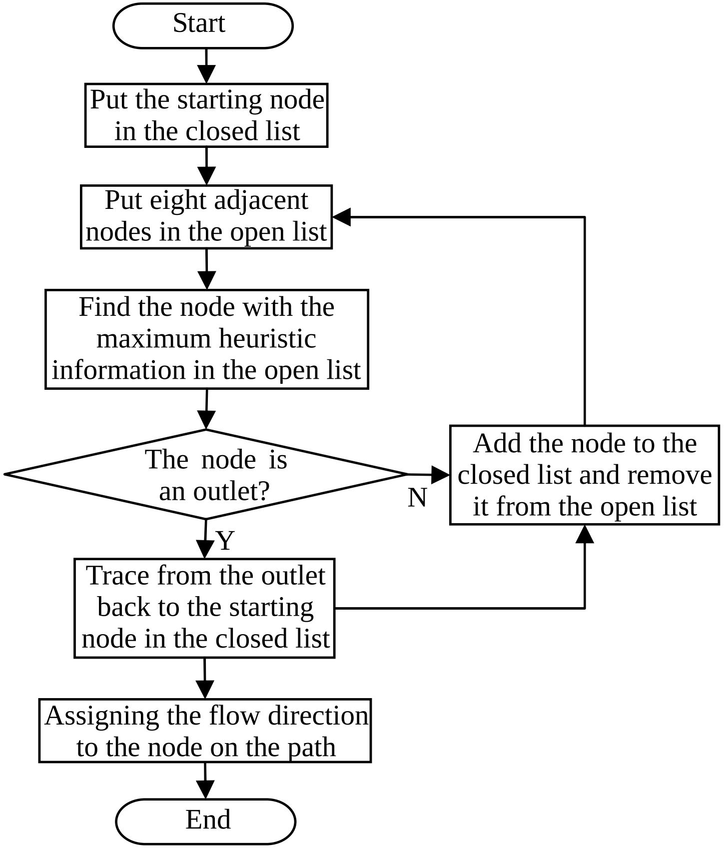

Fig. 1. Schematic diagram showing the proposed algorithm.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
|◀ ▶|
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

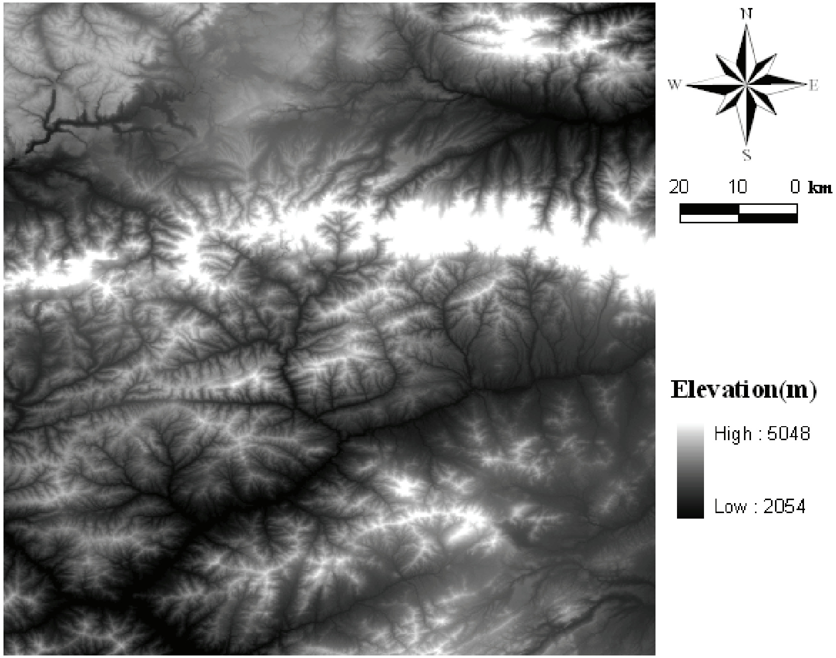

Fig. 2. Elevation distribution of the study area.

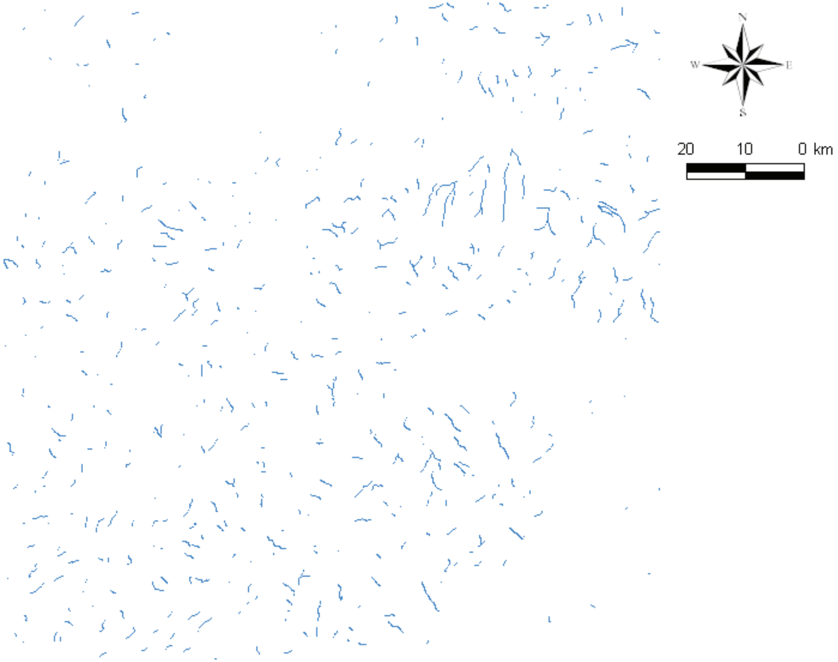

Fig. 3. Drainage networks generated using the basic algorithm.

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◄
►
Back
Close
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

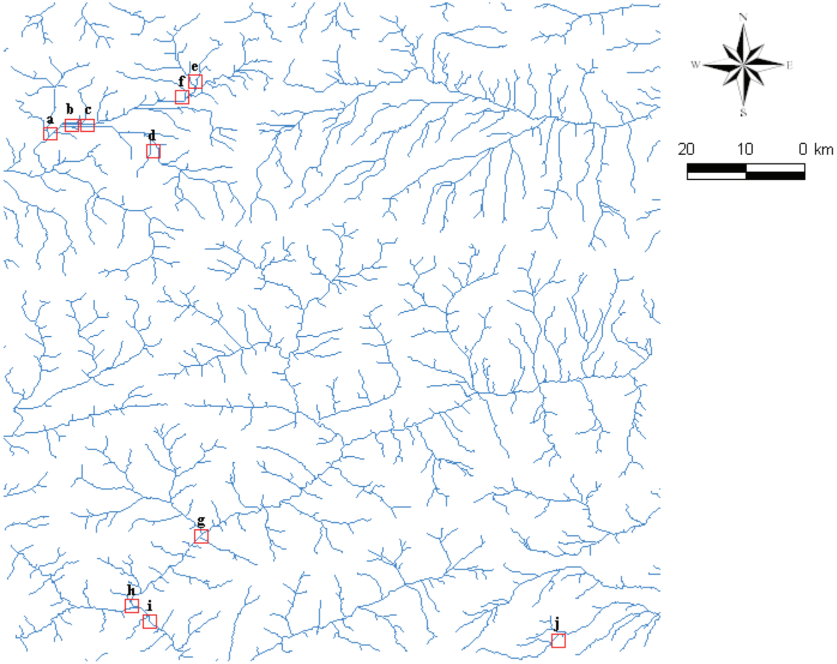

Fig. 4. Drainage networks generated using the ArcGIS 9.2.

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

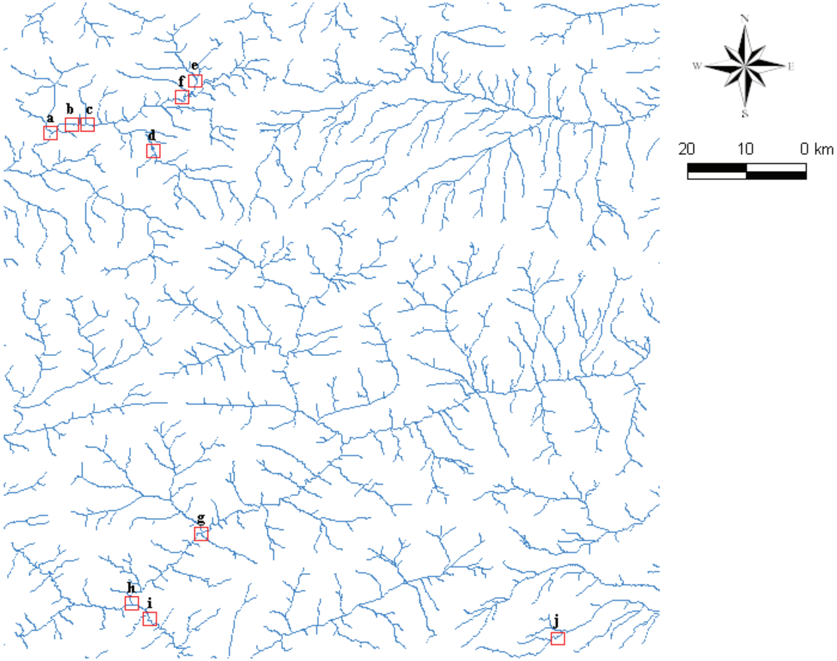

Fig. 5. Drainage networks generated using the proposed algorithm.

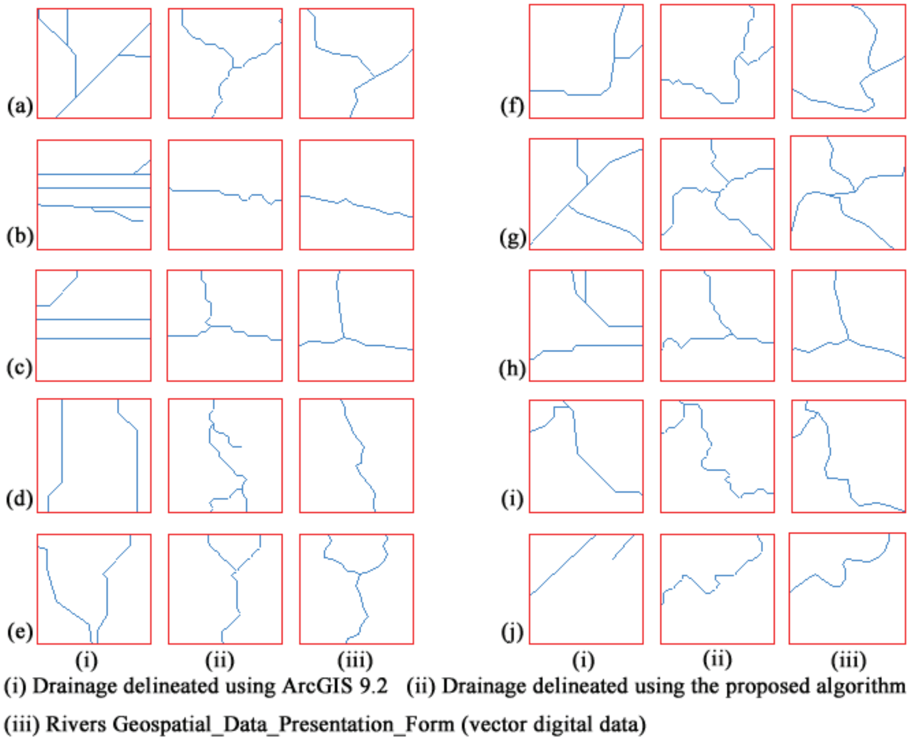

Fig. 6. Detail of square regions.