A BOINC based System for Global Topographic Structure Extraction Using SRTM Digital Elevation Models

Li-li Zhang Jiang Zhu, Zhiming Zeng

Institute of Geographical Sciences and Natural Resource Research of CAS,
Beijing 100101,
Tel: +86-10 82736655-8605
FAX: +86-10-82734630
Email: zhangli@supermap.com

Abstract

A public computing project named Terrain@home, is built on BOINC for extracting terrain features from global scope DEM data. Shuttle Radar Topography Mission (SRTM), which is a near global scope database of high accuracy elevation data, is selected as test data for Terrain@home. Conventional algorithms for terrain features auto extracting for example slope gradient, delineating drainage networks and etc, have to be reorganized to adapt the public and distributed computing environment. The unique problem that presented in SRTM data is introduced in this paper. The testbed for Terrain@home is described and computing results from DEM blocks in Tibet plateau is established. Performance evaluation is the focus of the last section and it is confirmed that public computing is an effective method to handle massive geographic data.

1. Introduction

The rapidly increasing availability of large-scale/high-resolution DEM data with the dramatic development of earth observation technique stimulates the emerging of complex, data intensive geospatial analysis model. Analysis of automatically extracting terrain features based on DEM such as slope gradient, slope aspect, curvature, flow paths and delineate drainage networks, is always a kind of fundamental work for many geoscientists. Shuttle Radar Topography Mission (SRTM), which produced a near global scope database of high accuracy elevation data, offers an opportunity to extend terrain analysis from local to global. However, it is a major disincentive for geoscientist to deal with massive DEM data even in a limited interesting research area. For example, SRTM 3'' derivative contains 15 billion measurements. Even in a limited area such as Tibet Plateau, which is roughly 2,500,000 km^2, there are still more than 17000*36000 cells.

In fact, the marriage of problem size and computational complexity has yielded GIS applications with enormous appetites for computing power (Armstrong, 2000). Many efforts have been done for strategies of decomposing geographic data, and parallel techniques suited to most of the common geographical data models have been prototyped (Mineter, 2003) based on various HP platforms from supercomputers to clusters. However, an effective data sharing and HP computation of large scale geographical data with network of workstations or PCs is still a challenge of increasing importance for the whole geo-community when specialized proprietary parallel supercomputers are extraordinarily expensive.
BOINC (Berkeley Open Infrastructure for Network Computing) is a software system that makes it easy for scientist to create and operate public-resource computing projects(Anderson, 2004). Some geocomputation applications such as CLIMATE@home,

Cliematepredicton.net,
have benefited a lot from BOINC. But a task must be divisible into independent pieces whose ratio of computation to data is high (Anderson, 2004) before it can benefit from public computing. Fortunately, most computations of terrain feature extraction from regular grid data are embarrassingly parallel computation. So, BOINC is selected as a development platform for effective data sharing and HP computation on SRTM3 DEM data.

A public computing project named Terrain@home is developed based on BOINC to solve the near global scope terrain feature extraction computation on SRTM3 data, and Tibet plateau is the focus area to test the project.

2. Experimental Datasets

DEM data used in the paper is SRTM3 and 1:250,000-scale Topographic Database of the National Fundamental Geographic Information System of China (NFGTD-1:250k). In 2000 Jet Propulsion Laboratory's (JPL) SRTM flew for ten days, mapping the world for topographic elevation underneath the entire footprint of the Shuttle's path(-54 to +60 in latitude). This new database gives a grid posting every 1 arc second (~30m) with absolute elevation accuracy of ~5 meters. This data is far more accurate than large portions of the world have previously been known. Although the SRTM database combines global scope and high accuracy, there are a few challenges linked to the size and properties of the SRTM database. Firstly, the SRTM database is too large for traditional GIS tools to extract information. Secondly, there are many holes or voids in the data arising from areas of low radar backscatter and even topographic shadowing. So, DEM of NFGTD-1:250k, which has 3-arc-second resolution, is used to fill the voids of SRTM. The details of SRTM data filling will be introduced in the followed section. Finally, while the data is of high absolute accuracy and containing almost no systematic errors, the data noise is on the order of 3-5m – in areas of low slope there are many apparent sinks or local minima in the data. This will need extra computing power for conventional algorithm to solve the problem.

Tibet plateau is the focus research area. Shape of Tibet plateau is shown as figure1.

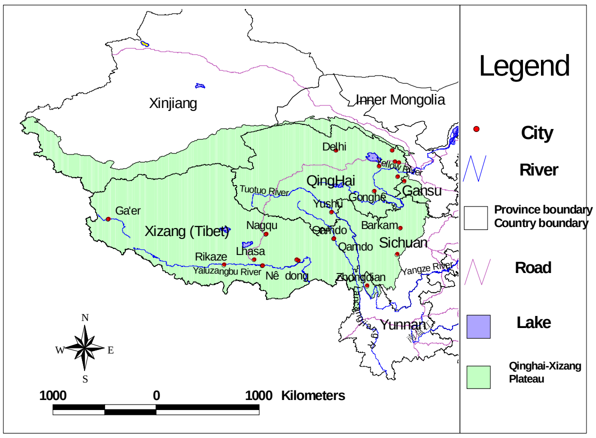

Figure 1. Shape of Tibet Plateau

3. Algorithms for Topographic Structure Extraction

Algorithms of Automatic topographic structure extraction from DEM have a long history and a well-established scientific utility. But there is still some work that has to be done to reorganize the algorithms to adapt the distributed computing environment. In the followed sections, some conventional DEM analysis algorithms will be reviewed firstly. Then, raster data decomposition which is important for raster data parallel processing is described. Thirdly, DEM void filling, the unique problem that presented by SRTM is illustrated. Finally, the paper focuses on the components of Terrain@home Project.

3.1 Conventional DEM Analysis Algorithms

Drainage network delineating which includes a set of algorithm, is the most deliberately researched procedures and has been implemented in some GIS software and described in working manuals. The most commonly used procedures which are described by Jenson and Domingue (1988) is illustrated in Figure 2. The Dinf method for specifying flow direction described by Tarboton(1997) significantly addressed the weakness of D8 method which assigns flow from each cell to one of its eight neighbours, either adjacent or diagonal, in the direction with steepest downward slope. But the assumption of all of these algorithms is that the memory of computer is not limited, so these methods only work on small DEM data with no more than 3000*3000 cells. Lars Arge(2001) described an I/O efficient method for hydrological structure extraction from DEM and the algorithm is implemented as a module of GRASS named terraflow. The efficiency of this method is described as Table 2 . Based on computing resources of NASA's Information Power Grid (IPG), Curkendall and Fielding described a computational-Grid Based System for Continental Drainage Network extraction using SRTM digital elevation models.

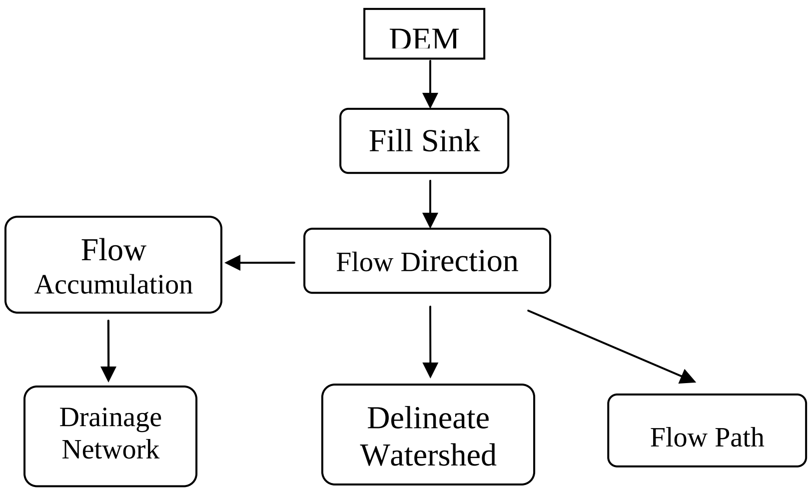

Figure2 Drainage Network Delineating Procedures described by Jenson and Domingue

Table2 Terraflow running times using 512MB of main memory (except for the Washington dataset where 1GB was used).

Dataset	Grid size	Fill		Flow		Total time	
		Terraflow	ArcInfo	Terraflow	ArcInfo	Terraflow	ArcInfo
Kaweah	$1.6\ 10^6$	2:19 min	1:17 min	1:02 min	0:31 min	3:21 min	1:48 min
Puerto Rico	$5.9\ 10^6$	7:04min	2:36min	1:23min	0:48min	8:27min	3:24min
Sierra Nevada	$9.5\ 10^6$	15:32mini	12:36min	11:13min	4:00min	26:44min	16:36min
Hawaii	$28.2\ 10^6$	33:23min	9:20min	3:19min	2:58min	36:42min	12:18min
Cumberlands	$66.7\ 10^6$	1:32hr	1:52hr	25:52min	1:06hr	1:58hr	2:58hr
Lower New England	$77.8\ 10^6$	1:55hr	1:47hr	38:43min	0:30hr	2:34hr	2:16hr
Central Appalachians	$122.0\ 10^6$	4:53hr	5:25hr	2:41hr	2:54hr	7:30hr	8:20hr
East-Coast USA	$245.7\ 10^6$	6:44hr	8:30hr	2:02hr	69:00hr	8:46hr	77:30hr
Midwest USA	$280.5\ 10^6$	10:21hr	13:30hr	6:05hr	18:55hr	16:26hr	32:25hr
Washington State	$1066.0\ 10^6$	31:30hr	N/A	31:23hr	N/A	62:53hr	N/A

Besides drainage network delineating, morphometric feature based on regular grid data that proposed by Wood, is also an important applied field of DEM. Figure3 illustrates six morphometric feature types based regular grid data, and the classification criteria are shown in table 3. Most of the conventional DEM analysis algorithms are simple, and always focus on the cell's eight neighbours.

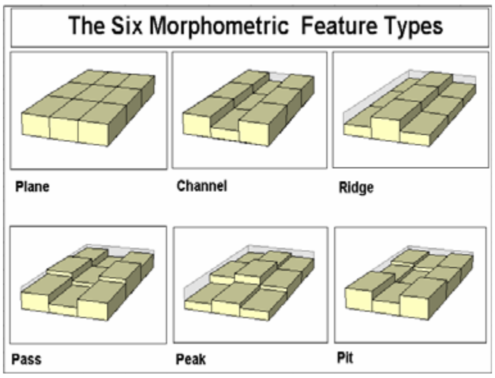

Figure 3 Six Morphometric Feature Types (Source: Wood, 1996)

Table 3 Simplified feature classification criteria (Source: Wood, 1996)

Feature	slope	cross-sectional curvature	maximum curvature	minimum curvature
Peak	0	#	+	+
Ridge	0	#	+	0
	+	+	#	#
Pass	0	#	+	-
Plane	0	#	0	0
	+	0	#	#
Channel	0	#	0	-
	+	-	#	#
Pit	0	#	-	-

3.2 Raster Data Decomposition

Parallelism requires that a problem is decomposed into concurrent subproblems and the method of decomposition used is a crucial determinant of processing performance. Three general methods, which are complete decomposition, domain decomposition and control decomposition, are elaborated by Ding(1996). Regular decomposition is that dividing a global scope DEM data into many blocks with same column and row count. The point is the border between neighbour blocks must be overlapped, illustrated as figure 4. Heuristic partitioning is a method of decomposition that investigated by Lee & Hamdi(1995), in which the goal is to minimise block border length. This decomposition is accomplished by recursively dividing the dataset into two sub-images, and in each division the direction of the new border is chosen to minimise the new border length.

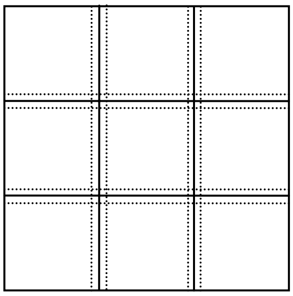

Figure4 Regular Raster Data Decomposition with 2 Rows/Columns Overlapper

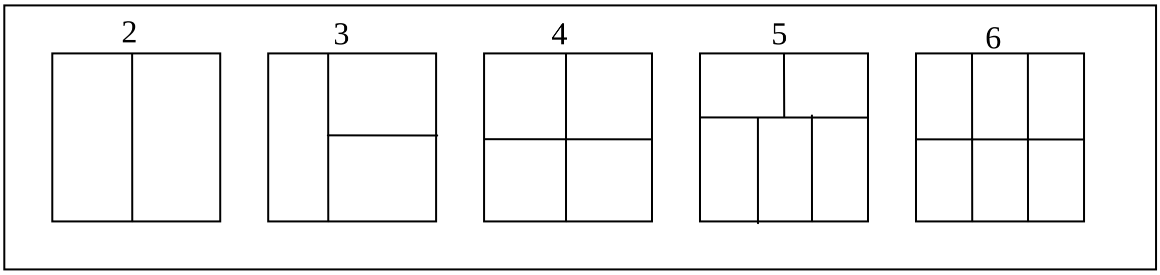

Figure5 Heuristic partitioning method for Raster Data Dividing

In fact, either SRTM or NFGTD-1:250k has been divided into many blocks with 1200*1200 cells per file. So, it is not necessary to implement the raster data decomposition algorithm.

3.3 Algorithm for Linking Results in Separated Blocks

It is clear that the process of extracting terrain features, in some ways, almost a perfectly data-parallel process. But there undeniably are very sequential aspects of this process as well in linking accumulation and stream order results in separated blocks. Curkendall and Fielding (2003) described an algorithm which could link flow accumulation, watershed and stream order in separated blocks as one consistent result. This algorithm is implemented in Terrain@home project to assimilate computing result.

3.4 Algorithm for SRTM Fill

It has been presented that there are many holes or voids in SRTM data arising from areas of low radar backscatter and even topographic shadowing. A comprehensive solution to the problem is something complex. In Terrain@home project, NFGTD-1:250k is used to fill voids in SRTM3 data and a simple method of filter is used to smooth the data. Computing result is illustrated by Figure 6.

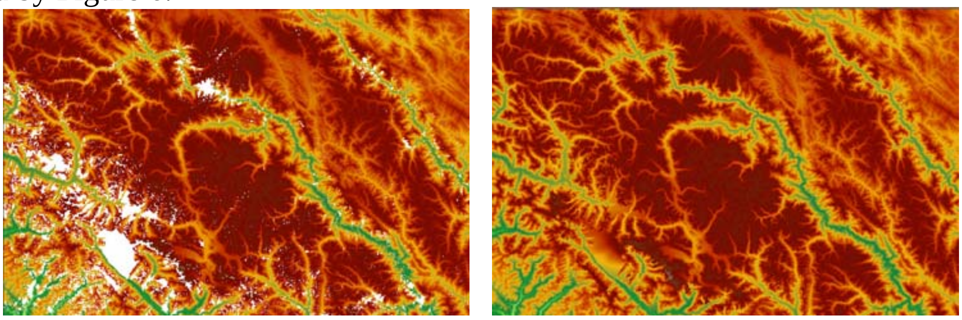

Figure 6 Original SRTM Data vs SRTM Data void filled

3.5 The components of Terrain@home Project

Although many efforts have been put into development of methods for extracting information from DEM data, little attempts have been done to construct a desktop GRID or use public computing resource to solve the massive DEM data analyst problem. And it is obviously that network of workstations or PCs is a more economical strategy than specified supercomputers.

BOINC is a software platform for distributed computing using volunteered computer resources. BOINC provides features that simplify the creation and operation of distributed computing projects. Based on the general BOINC components, the components of Terrain@home project are shown as Figure 7.

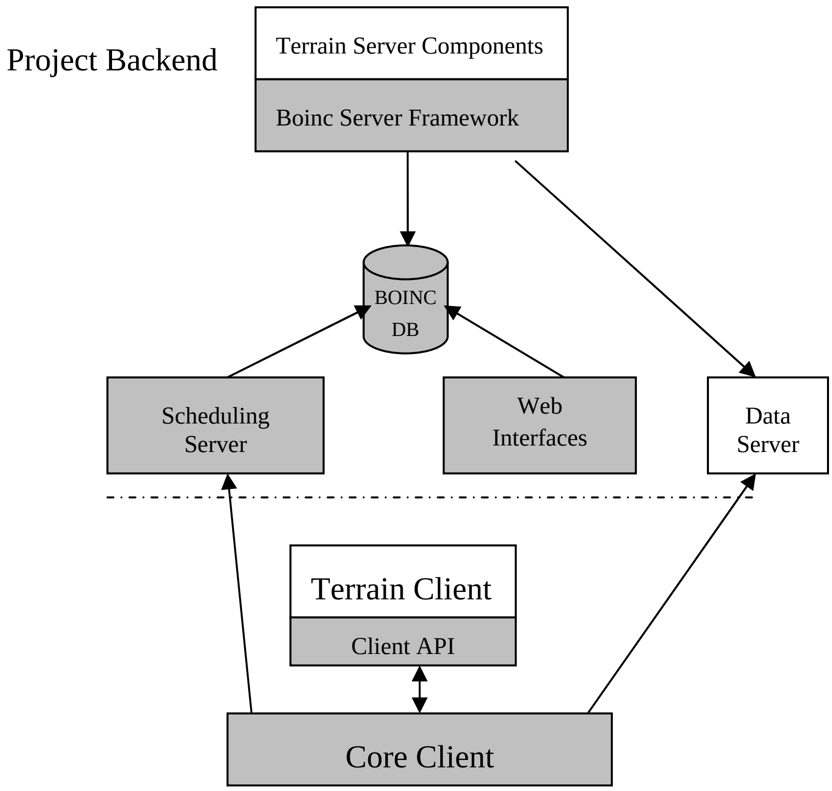

Figure 7 Components of Terrain@home project (Based on BOINC general project component)

Terrain server components are a set of Linux daemons include Work Generator, Validator, Assimilator, and File Deletion. Work generator is the program that sends the DEM block files to client, and Assimilator is in charge of storing the result to the Database. Terrain Client is the program that executes terrain analysis on computing node, and the algorithms for Topographic Structure Extraction are implemented here. When a volunteer downloads the core client and joins in the Terrain@home project through the Web Interfaces, the scheduling server will handle the request from the volunteer's computers, and a block of DEM will be send to the client. After the terrain analysis for example computing slope is finished, the result will be upload to the Data Sever. A geoscientist who is interested in the terrain analysis result can browse the map through WebGIS.

There are two scheduling strategies of BOINC project: scheduling policies in the BOINC Client and scheduling policies in the BOINC Server. In client side, there are also two interrelated scheduling policies: CPU scheduling policy and work fetch policy.

4. Computing Environment and Testbed

Figure 7 illustrates the architecture of Terrain@home. SRTM elevation data and the derived result are stored in data server. The scheduling server connects the data server through spatial database engine (SDE), and handles the request of computing node through the Web Interface.

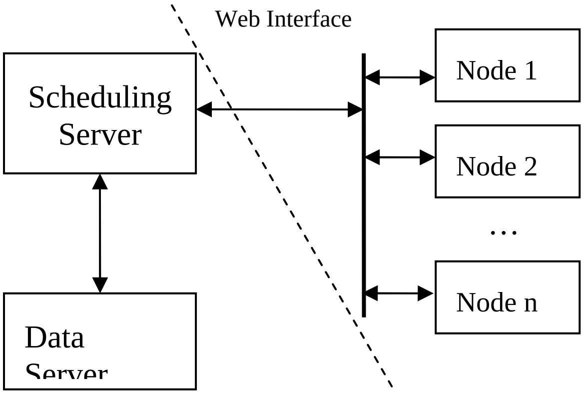

Figure 8 Terrain@home Architecture

5. Preliminary Results

The following derived data is being produced:

- Slope Map Grid

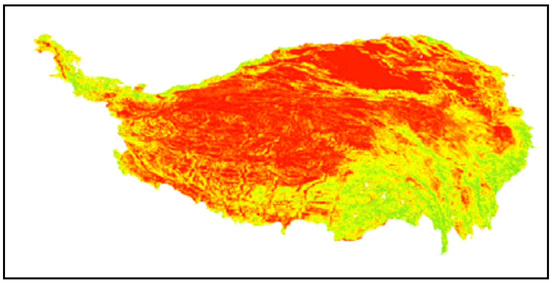

Figure9 SRTM3 Slope of Tibet

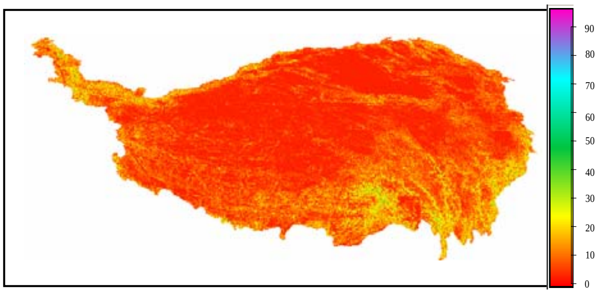

Figure10 GTOPO30 Slope of Tibet

Slope is the most important terrain feature, and it is also the basic for other computation. The SRTM3 Slope with degree unit of Tibet is show as Figure 8. GTOPO30 Slope of Tibet is also established for comparison. It is clear that SRTM3 slope is greatly different from GTOPO30 because of the more acute resolution. NFGTD-1:250k slope is almost same with SRTM3 in the full extent view, so it is not presented in the paper.

- Terrain Feature Map Grid

Full extent view Terrain Feature Map Grid of Tibet is show as Figure 10

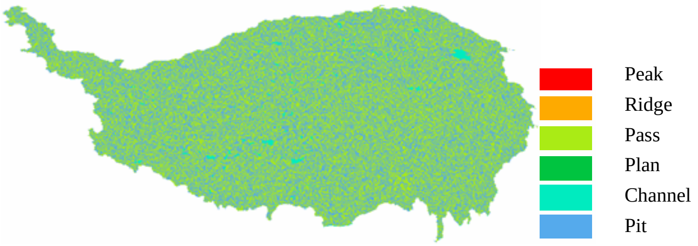

Figure11 Terrain Feature Map Grid

- Drainage Network Map Grid

Part region of drainage network extracted from SRTM3 is show as Figure 11. Cells in white colour is cells those with accumulation more than 200000.

Figure12 Drainage Network Map Grid

6. Concluding Discussion

A test project named Terrain@home is planned and implemented based on BOINC to extract topographic features from massively DEM data for example SRTM data. Conventional algorithms for terrain features auto extracting have been reorganized to adapt the public and distributed computing environment. The unique problem that presented in SRTM data is introduced in this paper, and a preliminary but high quality results is obtained through the project. But it should be noted that Terrain@home is not similar with other projects based BOINC for example SETI@home which are very large scale scientific computation using public computing resource through Internet. Terrain@home is a relatively small project compared with SETI@home or other BOINC base projects on Internet, and is only tested in LAN environment. Despite its preliminary character, this study can clearly indicate that network of workstations or PCs is a feasible and more economical strategy than specified supercomputers in high performance or high throughput geocomputation. And this work will be helpful and provide valuable experience to really large scale public computing project in geocomputation field for example spatial data mining.

7. References

Wood, Jo.， 1996,The Geomorphological Characterization of Digital Elevation Models,PhD thesis, Department of Geography, Leicester,UK: Univeristy of Leicester.

Jenson,S. K. and Domingue, J. O.,1988,Extracting topographic structure from digital elevation data for geographic information system analysis, Photogrammetric Engineering and Remote Sensing, 54:1593-1600.

Farr, T. G. and Kobrick,M.,2001,The shuttle radar topography mission,AGU Fall Meeting Abstracts.

Anderson,D.P.,2003, Public computing: reconnecting people to science, In Presented at the Conference on Shared Knowledge and the Web, Residencia de Estudiantes, Madrid, Spain. http://boinc.berkeley.edu/talks/madrid_03/madrid.html.

Anderson,D.P.,2004, Boinc: A system for public-resource computing and storage,In *5th IEEE/ACM International Workshop on Grid Computing*, Pittsburgh, USA.

Stainforth,D. et al,2002, Climateprediction.net:design principles for public-resource modeling research, In *Proceedings of the 14th IASTED International Conference PARALLEL and Distributed Computing and Systems*, Cambridge,USA.

Curkendall, D.W., Fielding E.J., Cheng T.-H., and Pohl J.M., 2003, A computational-grid based system for continental drainage network extraction using srtm digital elevation models, In *Proceedings of the HPSECA/IPCC conference*, Taiwan.

Armstrong, M.P.,2000, Geography and computational science, *Annals of the Association of American Geographers*, 90(1):146-156.

Mineter,M.J., 2003, A software framework to create vector-topology in parallel gis operations, *International Journal of Gegraphyical Infromation Science*, 17(3).

Lee,C.K. and Hamdi,M., 1995,Parallel image processing operations on a network of workstations,Parallel Computing,21,137-160

DING,Y. and DENSHAM ,P.J., Spatial strategies for parallel spatial modelling, *International Journal of Geographical Information Systems*, 10(6):669-689, 1996.

Tarboton, D. G., 1997, A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models, Water Resources Research, 33(2): 309-319.

Toma,L, Wickremesinghe, R., Arge, L., Chase,J. S., Vitter,J. S., Halpin,P. N. and Urban ,D., Flow computation on massive grid terrains. GeoInformatica, International Journal on Advances of Computer Science for Geographic Information Systems, 2001.