Computers & Geosciences 94 (2016) 140–149

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

New ArcGIS tools developed for stream network extraction and basin delineations using Python and java script

Adel Omran [a,d,*], Schröder Dietrich [b], Abdou Abouelmagd [c], Märker Michael [d]

[a] Department of Science and Mathematical Engineering, Faculty of Petroleum and Mining Engineering, Suez University, Egypt
[b] Department of Geomatics, Computer Science and Mathematics, University of Applied Sciences Stuttgart, Germany
[c] Department of Geology, Suez Canal University, Ismailia 41522, Egypt
[d] Heidelberger Akademie der Wissenschaften; Geographisches Institut der Universität Tübingen, Germany

ARTICLE INFO

Article history:
Received 26 February 2016
Received in revised form
19 April 2016
Accepted 10 June 2016
Available online 13 June 2016

ABSTRACT

Damages caused by flash floods hazards are an increasing phenomenon, especially in arid and semi-arid areas. Thus, the need to evaluate these areas based on their flash flood risk using maps and hydrological models is also becoming more important. For ungauged watersheds a tentative analysis can be carried out based on the geomorphometric characteristics of the terrain. To process regions with larger watersheds, where perhaps hundreds of watersheds have to be delineated, processed and classified, the overall process need to be automated. GIS packages such as ESRI's ArcGIS offer a number of sophisticated tools that help regarding such analysis. Yet there are still gaps and pitfalls that need to be considered if the tools are combined into a geoprocessing model to automate the complete assessment workflow. These gaps include issues such as i) assigning stream order according to Strahler theory, ii) calculating the threshold value for the stream network extraction, and iii) determining the pour points for each of the nodes of the Strahler ordered stream network. In this study a complete automated workflow based on ArcGIS Model Builder using standard tools will be introduced and discussed. Some additional tools have been implemented to complete the overall workflow. These tools have been programmed using Python and Java in the context of ArcObjects. The workflow has been applied to digital data from the south-western Sinai Peninsula, Egypt. An optimum threshold value has been selected to optimize drainage configuration by statistically comparing all of the extracted stream configuration results from DEM with the available reference data from topographic maps. The code has succeeded in estimating the correct ranking of specific stream orders in an automatic manner without additional manual steps. As a result, the code has proven to save time and efforts; hence it's considered a very useful tool for processing large catchment basins.

1. Introduction

The evolution of a drainage system over space and time is affected by many variables such as lithology, tectonic lineaments, geomorphology, soil and the area's landcover. Many of these variables are mirrored in the landscape topography, which can be quantified and classified using concepts of geomorphometry. The measurement of shapes or geometries of any natural geomorphological features is termed as geomorphometry (Pike et al., 2008; Selvan et al., 2011). A detailed morphometric analysis of a basin greatly helps to characterize the impact of drainage morphometry on landforms and their features (Chandrashekar et al., 2015). Morphometric analyses are important in the context of the estimation of flash flood risk levels of watersheds. They can be used as an attempt to elucidate the surface water potentialities of basins in order to describe the basin's hydrological behavior (Angillieri, 2012; Omran, 2013) and to quantify the hydrological characteristics. Thus, the results of morphometric analysis will be a useful input for a comprehensive water resource management plan (Jawaharraj et al., 1998; Kumaraswami et al., 1998; Sreedevi et al., 2001). Hydrological models that are mainly based on morphometrical analytical results include the Instantaneous Unit Hydrograph (Nash, 2009), the Geomorphoclimatic Unit Hydrographs (Gupta et al., 1980) and the Geomorphic Unit Hydrograph (Rodriguez-Iturbe and Valdes, 1979). These models have been applied over ungauged basins in arid and semi-arid regions. The study of morphometric parameters mainly requires the delineation of both the drainage networks and the watershed line.

Therefore, today's state-of-the-art techniques should be applied. These include Remote Sensing (RS) data and processing

* Corresponding author at: Department of Science and Mathematical Engineering, Faculty of Petroleum and Mining Engineering, Suez University, Egypt.

http://dx.doi.org/10.1016/j.cageo.2016.06.012
0098-3004/© 2016 Elsevier Ltd. All rights reserved.

using Geographic Information Systems (GIS), with the added availability of high resolution digital elevation models (DEM) from earth observation satellites and the progress made in computer sciences. Due to the increased power to process these large amounts of data even on PCs or laptops and the progress statistical and mathematical methods, many difficulties have been resolved and new problems can be tackled (Evans et al., 2003). This leads to a renaissance of concepts that were already introduced 70 years ago. At that time several studies focused on mapping of drainage networks and their watersheds for hydrological studies (Horton, 1945; Strahler, 1964; Shreve, 1974). Tracing techniques were used in these studies to extract drainage network and delineate boundaries of basins for studying the characteristics of basins and their relationship to the geometriesof those basins. Using classical approaches for delineation the drainage network, these studies need to measure linear features directly in the field or retrieve from secondary sources, (e.g., digitized from topographic maps, aerial photographs and stereo images). In many areas of the world, topographic maps are still the basic traditional reference for drainage network analysis because of their availability, simplicity and affordability. However, the extraction of information, such as delineation of drainage and watershed from topographic maps, requires much time and expertize in cartography, resulting in subjective decisions. Moreover, the results of manual procedures such as tracing methods still have to be transferred to digital data for further processing. Limitations and subjectivity of manual procedures in defining stream networks highlight the need for a more precise and efficient approach in depicting landscape dissection. The widespread availability of digital data including DEMs, radar images, stereo photogrammetry, and Light Detection and Ranging (LiDAR) point clouds has opened new gates for more objective approaches to the delineation of channel networks (Sekulin et al., 1992; Bertolo, 2000; Lin et al., 2005; Afana, 2011). In the 1980s computational technologies were developed to use DEMs for the extraction and numerical analysis of drainage networks (O'Callaghan and Mark, 1984; Jensen, 1985). Nowadays, many GIS like ESRI's ArcGIS, QGIS or Saga include in their toolboxes standard tools to extract the stream segments and basin watersheds from DEMs. In the recent years, with the availability of data and processing power, it is possible to process bigger datasets for larger catchment areas in a reasonable time and the extracted results were used to study the morphometrical parameters for mapping flooding risk areas (Rudriaih et al., 2008; Al Saud, 2009; Nageswararao et al., 2010).

To apply the complete workflow from DEM preprocessing to the extraction of morphometric parameters to larger regions, where perhaps hundreds of watersheds have to be delineated and classified, the process has to be automated. Although the complete process can be implemented straight forward, the extracted results still create different problems with regard to their credibility in using the data in geomorphological and hydrological models (Omran, 2013). Additionally, depending on the software which is used, there are still some gaps in the automated workflow which until now have to be filled by manual work. For example, using ArcGIS' Spatial Analyst Hydrology tool set, the extracted stream segments are attributed with the correct Strahler order, but they have to be merged according to their order number as the watershed characteristics like drainage density or frequency are based on them. Omran et al. (2012) describe an algorithm for the merging of stream segments based on Strahler's theory. Another problem to be addressed is the determination of the pour points of different sub basins according the merged segments.

Despite the validation of stream extraction from DEM has received considerable attention, the assessment of the achieved results still lags behind. The validation procedure for a drainage network should be carried out prior further processing, as in the case of hydrological models. Generally, there are two main approaches for drainage network validation: quantitative and qualitative methods (Chorley et al., 1984). Quantitative method includes geomorphometrical parameters that describe structural properties of a stream network. These properties are extracted from different sources (e.g., digital maps or extracted drainage networks) and then compared statistically. The qualitative method depends on expert knowledge based on field verification, visual interpretation of the resulting data, comparison with other data sources, such as orthophotos and 3D structures (Afana, 2011). Field work still forms one of the most precise approaches to validate channel network. The exact stream network can be examined in the field, but time and efforts make it impractical to check for stream validity, especially in large-scale catchments. As a surrogate for own field work, topographic maps can be considered. Based on geodetic field survey, a drainage network has been captured accurately and has been mapped following the cartographic rules of generalization, depending on the scale of the map. A good compromise for the level of details needed and cartographic generalization topographic maps of scale1: 25,000. On the other hand, the details of the drainage network extracted from a DEM are closely related to the DEM's resolution. For this study, related to the availability of maps and the DEM resolution, mid-scale maps (e.g., 1:50,000 and 1:100,000) were used for evaluation of the results. Stream network details are defined by a threshold value that determines where channels begin in the landscape, widely known as the "specific threshold area". This value represents the minimum drainage area required to drain to a point where a channel forms. It is the essence of stream extraction to select the appropriate threshold value. The choice of the appropriate value used to define the optimum channel network is highly related to the scale and resolution of the original data (Thompson et al., 2001). This value for channel initiation is usually specified arbitrarily although it is recognized that different threshold value will result in substantially different stream networks for the same basin (Helmlinger et al., 1993). The smaller the chosen threshold value, the more detailed the obtained channel network, and more initial sub-watersheds will be generated. On the other hand, depending on the algorithm used for determining the flow direction, artifacts will be introduced which will not reflect the real world situation. Different suggestions have been made to find an optimal threshold value. Generally, using a constant value for stream network delineation is an accepted means of determining where channels begin in the landscape (e.g., O'Callaghan and Mark, 1984; Band, 1986; Montgomery and Dietrich, 1992). However, drainage density has been shown to vary between regions due to different climatic regimes, natural landscape characteristics, and land-use impacts (Gandolfi and Bischetti, 1997; Tucker and Bars, 1998). Additionally, assigning a constant threshold value neglects the spatial variability of headwater source areas and may lead to significant differences between field observations and predicted conditions (Willgoose and Perera, 2001). One common approach to define the threshold value is calculating 1% of the maximum flow accumulation, which is considered a default method for displaying the stream network (Band, 1986; Tribe, 1992; Merwade and Ruddell, 2012). Deilami et al. (2013) statistically determined the threshold value to be the first break value from the standard deviation classification method for a flow accumulation raster layer. Another method to select the threshold value was developed and implemented in the TauDEM software (Tarboton, 2001; Shrestha and Miyazaki, 2006). Ariza-Villaverde et al. (2013) suggest a multifractal analysis for determining an optimal threshold value. In many other studies, the value has been determined based on trial and error, using the visual similarity between the extracted network and the lines depicted on topographic maps. This paper describes a new automated approach for selecting the optimum threshold value and comparing the results with these extracted from topographic maps.