**Doctoral Dissertation Research: Automating Flow Accumulation Thresholds for Extracting Stream Networks from Digital Elevation Models**

**Project Summary:** DEMs are widely used for stream network extraction. Assuming that water flows along the path of steepest descent, DEM-based methods track the simulated flow of water to calculate flow direction and flow accumulation. For stream initiation a threshold for flow accumulation is defined. Choosing an appropriate threshold is pretty significant because different thresholds result in considerably different stream networks in terms of total stream channel length, stream order, and drainage density. To date, the proper method to choose the appropriate threshold is still undefined. This proposal aims to automate the process of stream networks extraction from DEMs by incorporating terrain derivatives and landscape characteristics into flow accumulation threshold selection procedure. Previous work has examined the effects of terrain derivatives and landscape characteristics on stream formation and flow accumulation threshold, but none have completely been effective in taking the spatial variability of terrain into account to choose an appropriate threshold. Three key questions are central to the proposed research:

* How can the choice of a flow accumulation threshold be automated?
* How can terrain derivatives and landscape characteristics inform automatic extraction of stream networks from DEMs?
* How can spatial variations in the accuracy of DEM-derived stream networks be quantified?

**Intellectual Merit:** The proposed research demonstrates how flow accumulation threshold can be optimally selected. This technique provides a new method for stream extraction form DEMs. The results of this research will fill a critical gap in the literature regarding the impacts of threshold on headwaters (i.e., first order tributaries). Decision makers commonly use USGS topographic maps or National Hydrography Dataset (NHD) in their decisions. However, these datasets underestimate and do not reflect a substantial portion of headwaters. An understanding of uncertainty in the threshold is especially necessary for hydrologist to determine how this inaccuracy leads to uncertainty in their models. Results of this research are needed in order to improve the scientific value of these datasets that are widely used in most of the hydrologic applications. Furthermore, this research is highly interdisciplinary, bridging research in Geography, Civil Engineering, Geomorphology, Water Resources, and Environmental Studies. So, this research will result in major contributions to a variety of fields, including hydrological modeling, terrain modeling, hillslope processing, topographic mapping, etc. Finally, it should be mentioned that the principal investigator of this project is highly qualified to perform the proposed project, given her more than eight years of research on generalization of NHD.

This proposal seeks to automate the choice of flow accumulation area threshold through the following:

* Incorporating terrain parameters and environmental characteristics into flow accumulation area threshold selection procedure.
* Accurate identification of headwaters in order to find the best way of finding flow accumulation area threshold
* Efficient and accurate evaluation of DEM-derived stream networks in order to find the optimum threshold.

**Broader Impacts:** Among the broader impacts associated with this project, four stand out as potentially transformative: (1) Currently USGS is quite interested to extract streams from DEMs and improve the NHD. Close collaboration with employees of the USGS will guarantee the distribution of the research results to the relevant agencies. (2) This research has a real-world need. Although this research is developed based on subbasins in the United States, it will provide a model workflow for other parts of the world. In fact, the research goals of this proposal have the potential to result in a more efficient and cost-effective solution to a world's hydrology problem. (3) Research findings will be integrated into core Geography courses at the University of Colorado including fluvial geomorphology, hydrology, and GIS modeling courses, allowing students to broaden their understanding of the application of GIS in hydrology and geomorphology. In order to develop a stronger curriculum, also, the request for proposing a new course titled “GIS in water resources” will be proposed. (4) This project started as a class project collaborating with three undergraduate students that was a rewarding experience. They raised interesting questions in the project and made significant contributions to the research. This proposal also seeks summer funding for two undergraduate students to participate in research for preparing the data and validating the results.

**PROJECT DESCRIPTION: Automating Flow Accumulation Thresholds for Extracting Stream Networks from Digital Elevation Models**

**1. Introduction**

Water is one of the earth’s most valuable natural resources. Due to the cultural, economic, and political significance of water in the world today, the study of water throughout the earth is of great importance (Koundouri & Papandreou 2013). Therefore, it is mandatory that the surface water datasets used for topographic mapping, hydrological modeling, and water management contain a complete and accurate representation of stream channels. The focus of this research is on stream networks which provide a geographic and functional context for many water-related purposes. The accuracy of stream networks can affect many environmental applications, such as ecological water distribution analysis, source pollution management plans, and wildfire remediation. Field surveying is the most accurate way for mapping stream networks, but it is extremely time-consuming and costly. The path of streams through the land is formed based on topography, and by increasing the quality of Digital Elevation Models (DEMs), automated extraction of stream networks from DEMs is considered as a great alternative. The proposed research aims to automatically extract streams from DEMs in order to develop an accurate and comprehensive database of stream networks.

DEMs are widely used for stream network extraction (Montgomery & Foufoula‐Georgiou 1993, Pelletier 2013, Bai et al. 2015). Assuming that water flows along the path of steepest descent, DEM-based methods track the simulated flow of water to calculate flow direction and flow accumulation (Tarboton et al. 1991). For stream definition a threshold for flow accumulation is defined; and all cells in the flow accumulation matrix with a value greater than the threshold are defined as a stream cell. The problems of stream network extraction from DEMs (e.g., pits, flat terrain, grid bias, threshold selection, etc.) has been studied for several decades (Peucker & Douglas 1975, O'Callaghan & Mark 1984, Fairfield & Leymarie 1991, Luo et al. 2014, Mao et al. 2014). In general, accuracy of DEM-derived streams are a function of the DEM’s spatial resolution (McMaster 2002), the algorithm used to extract streams (Heine et al. 2004), and landscape characteristics (Data et al. 2001, Coblentz et al. 2014, Hastings & Kampf 2014).

The quality of the extracted stream networks is greatly influenced by the spatial resolution of the DEM (McMaster 2002, Chaubey et al. 2005). Currently, various levels of spatial resolution are offered by different datasets, such as Light Detection and Ranging (LiDAR) data (1 m resolution), Shuttle Radar Topography Mission (SRTM) data (30 and 90 m resolution), and the National Elevation Dataset (NED) (100, 30, 10, and 3 m resolution). Various studies have been done on the influence of DEM resolution on the accuracy of extracted stream networks (Wang & Wang 2014, Yang et al. 2014, Rosim et al. 2015). In this research the extraction of stream networks from fine resolution elevation data and its effects on flow accumulation threshold are investigated. We cannot assume that finer resolution DEMs lead to better results; some models have been designed to work with coarse resolution data (Data, Characteristics et al. 2001), and the proposed work expects to provide a detailed analysis of the stream properties extracted form DEMs with different resolutions and take into account the scale of the processes that are modeled.

The algorithm used for stream extraction from DEM has a key role in the accuracy of the final product. Main challenges of DEM-based stream networks extraction algorithms are: (1) Flow direction calculation in flat terrain; this issue has been practically resolved by considering the fact that water flow is towards lower terrain and away from higher terrain (Garbrecht & Martz 1997, Barnes et al. 2014). (2) Inconsistencies between the extracted streams and the available mapped streams due to differences in data sources, equipment and human processing; a problem which can be addressed by artificially decreasing the elevations of stream pixels using another data source (this process is called stream burning) (Tarboton 1997). (3) ‘‘Grid bias’’ in a single flow direction algorithm (assigning flow from each pixel to one of its eight neighbors, in the direction of steepest downward); Multiple flow direction algorithms have been proposed to resolve the issue (Freeman 1991, Tarboton 1997). (4) Difficulties in choosing the appropriate flow accumulation threshold; finding the optimum threshold and accurately locating headwaters (i.e., “the farthest upslope location of a channel with well-defined banks” (Montgomery & Dietrich 1988)) remains unsolved. The conventional practice for delineating a stream network from a DEM is to define a constant threshold for flow accumulation. The most common way to determine a threshold is by trial and error, which can be time consuming, inconsistent from one location to another, and possibly erroneous (if an inappropriate threshold value is selected) (Bhowmik et al. 2015). Choosing an appropriate threshold is difficult, because stream initiation depends on different processes such as groundwater in a low‐slope area or a combination of slope and soil erosion (Fig. 1). The proposed research will address this issue by automating the process of stream networks extraction from DEMs and incorporating terrain derivatives and landscape characteristics into flow accumulation threshold selection procedure.

Three key questions are central to the proposed research:

* How can the choice of a flow accumulation threshold be automated?
* How can terrain derivatives and landscape characteristics inform automatic extraction of stream networks from DEMs?
* How can spatial variations in the accuracy of DEM-derived stream networks be quantified?

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
| **Fig 1**: Examples of headwater formation based on different processes (Passalacqua et al. 2010, Clubb et al. 2014) | | |

**2. Background**

Numerous techniques have been proposed to extract the stream network from DEM represented either in the form of a Triangular Irregular Network (TIN) (Jones et al. 1990, Nelson & others 1994), or as a regular raster surface (Mower 1994). In this research, streams are extracted from DEM, and TINs may be used for validating the results. In this section, flow direction, flow accumulation and stream initiation is discussed, and then the factors (terrain derivatives, landscape characteristics, and stream geometry) that can enhance this process are explained.

**2.1. Flow Direction, Flow Accumulation and Stream Initiation**

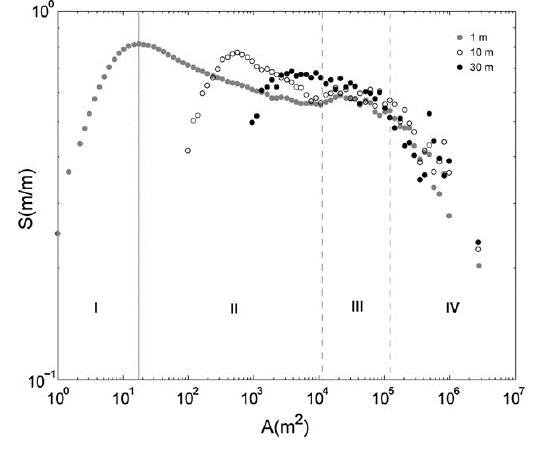
**Flow Direction:** Flow direction matrix whose cells indicate the direction of flow from each grid cell to its downhill neighbor. The simplest and most widely used flow direction method called D8 which assigns flow from each pixel to one of its eight neighbors, in the direction of steepest descent (O'Callaghan & Mark 1984). Although this method is simple and generates connected networks, choosing only one of the eight possible directions, is one of the disadvantages of D8 method. It leads to a bias in flow path (Fairfield & Leymarie 1991), because it cannot model divergent flow over convex slopes (Freeman 1991). To solve the limitations of D8, some other methods such as multiple flow direction method (assigning flow to each lower neighbor in proportion to the slope) (Freeman 1991), random direction method (using stochastic rules) (Fairfield & Leymarie 1991) and plane flow methods (fitting a plane to the elevations of pixel corners and assigning flow in the direction of maximal plane slope) (Lea 1992, Costa‐Cabral & Burges 1994) have been proposed. Another very famous method that assigns flow between one or two downslope pixels is D-infinity (Tarboton 1997). In this method water flows in the direction of steepest decent on the eight triangular facets; If steepest decent direction be between the two neighbor cells, flow will be assigned to both of them based upon the angles between the steepest decent direction and the direction to each of the two cells. In this research, D8 and D-infinity will be employed and their results will be compared and discussed.

**Flow Accumulation:** Flow accumulation is calculated based on the flow direction matrix. Each pixel in the flow accumulation is assigned a value equal to the number of upstream cells have water flowing to it. If the number of each cell is multiplied by its area the result would be upstream drainage area.

**Stream Initiation (Flow Accumulation Threshold Selection):** For stream definition a threshold for flow accumulation is defined; all cells in the flow accumulation matrix with a value greater than the threshold are defined as a stream cell and assigned the value “1” in the stream matrix. All other cells are assigned the value of "0" (O'Callaghan & Mark 1984). The delineation of stream networks is usually based on a constant threshold. Such an approach assumes a homogeneous terrain and stream density for the entire study area. For large areas and/or heterogeneous terrain, however, this assumption is an unrealistic hypothesis, and ignores the spatial variability of terrain. Optimum flow accumulation threshold may change significantly in different basins and even within a single subbasin. A central goal of this project is to propose a method which accounts for terrain heterogeneity. Therefore, hydrologically relevant terrain derivatives (e.g., slope, curvature, roughness, and wetness index) and landscape characteristics (e.g., climate, soil, geology, and vegetation) should be taken into account as ancillary data in order to finding the optimum threshold and preserve differences in stream density.

**2.2. Terrain Derivatives**

Here the terrain derivatives that can be used to find the optimum threshold are listed: (1) **Slope**: The relationship between local slope and flow accumulation area may be used to find a proper threshold for different landscape types. That is, the transition from hillslope to stream can be identified by investigating the ‘local slope versus flow accumulation area’ plot (Tarboton et al. 1992). Although the slope-dependent threshold method leads to higher drainage density in steeper terrain, as is found in natural terrain, there is not a clear relationship between the results of this method and field-mapped headwaters in some locations and some other factors such as bedrock characteristics have a greater influence on stream location (Jaeger et al. 2007, Clubb, Mudd et al. 2014). Moreover, this is worth-mentioning that DEM resolution tremendously affects the results of this method, and accurate estimation of local slope necessitates a fine resolution DEM (Fig. 2). (2) **Roughness**: Terrain roughness is a measure of the texture of a surface defined as the variability of elevation values (i.e., absolute standard deviation of all values within a window) of a topographic surface (Frankel & Dolan 2007). Terrain roughness demonstrates the geologic nature of rocks (Hobson 1972), and has a direct relationship with stream sinuosity and in inverse relationship with stream density (Coblentz, Pabian et al. 2014). (3) **Curvature (Plan & Profile):** “Curvature is the second derivative of a surface, or the slope of the slope.” (Kimerling et al. 2009). The curvature can be positive (peaks or ridges), negative (valleys or channels) or zero (flat surface or saddle). Terrain curvature provides useful information about the initiate location of streams.



**Fig 2**: DEM resolution affects the results of area-slope method (Tarolli et al. 2012)

**2.3. Landscape Characteristics**

Here the landscape characteristics that can be used to find the optimum threshold are listed: (1) **Climate**: Climate plays a significant role in the spatial patterns of density and stream networks. Generally there is a positive correlation between stream density and rainfall (Tucker & Bras 1998). Precipitation effects on density are also a function of vegetation and temperature. There is a positive relationship between Precipitation and density in semiarid climates and a negative relationship in humid climates (Kirkby 1987). Also, there is a negative correlation between stream density and rainfall in densely vegetated terrain (Morisawa & Clayton 1985). This study aims to identify the effect of dry, humid, and transitional climates on flow accumulation threshold. (2) **Soil**: Hydraulic conductivity (a measure describes the ease with which water can pass through soil or rock), soil transmissivity (a measure describing the amount of water that can be transmitted horizontally), and soil erodibility (a measure describing the susceptibility of soil to erosional processes) influence the density of streams. These factors are inversely related to the stream density (Data, Characteristics et al. 2001). (3) **Geology**: Bedrock geology influences the density of streams; density is lower in resistant rocks (Strahler 1964). (4) **Vegetation**: vegetation can have various effects on the stream density based on the climate and soil characteristics. Although humid climates result in an increase in stream density, an increase in vegetation leads to an increase in surface resistance and consequently a decrease in stream density (Istanbulluoglu & Bras 2005).

**2.4. Stream Geometry:**

The 2-dimensional geometry of stream networks contains implicit information about the topographic characteristics of a landscape. A stream network involves many tributaries, which are merged together and form a hierarchical structure. Assuming that water always flows along the path of steepest descent, the primary branches of the hierarchy are located at the utmost elevation and they make second-class branches at a lower altitude. So, stream orders can identify upstream and downstream locations and reflect the cumulative decrease of altitude in stream branches. Identifying headwaters (i.e., first order streams) has been recognized as the main problem of stream extraction from DEM (Mark 1983). The total stream network length, stream order, and drainage density can significantly change because of small errors in identifying headwaters (Garbrecht & Martz 2000). Heine et al. (2004) contend that “meeting the challenge of locating headwaters is the key to accurate mapping of stream networks.” Several authors have attempted locate headwaters using DEM, because most of the available stream data underestimate the lower order streams (Clubb, Mudd et al. 2014, Bai, Li et al. 2015, Płaczkowska et al. 2015). To compensate for the missing headwaters, the proposed work expects to propose a method to extract headwater locations and test against a robust data set across various landscapes.

Stream density (also called drainage density) also can be used to find an optimum threshold. Stream density is the sum of the length of all the streams in a drainage basin divided by the drainage basin area (Horton 1945). Stream density is a function of both climate and physical characteristics of the terrain. In fact, wet weather, impermeable bedrock, rugged terrain, etc. decrease surface water runoff and consequently stream density (Tucker & Bras 1998). A central goal of this project is to define homogeneous drainage density regions using environmental factors in order to enhance the extracted streams.

**3. Research Design**

**3.1. Data Set and Study Area**

Methods will be tested on about 30 subbasins distributed across the United States. The reasons for working on this problem within the United States are three-fold. First, water is a very important resource in the United States. Based on the USGS website: “Water is one of six science mission areas of the U.S. Geological Survey (USGS). Water's mission is to collect and disseminate reliable, impartial, and timely information that is needed to understand the Nation's water resources”. Second, the United States is characterized by varying terrain and climate conditions; so the density of streams and also the type of errors happened in the DEM-derived streams may differ dramatically, providing a varied set of subbasins with which to work, test and validate the methods described here. Third, the existence of high quality data is quite important in this research, and accessing to this data in the United States is feasible.

The methodology of this research is based on the combination of DEM and landscape parameters in a GIS environment. DEM is used to calculate flow direction, flow accumulation, slope, curvature, and roughness. Then in a raster-based process the DEM-derivatives and other ancillary data are combined in a way to find the best threshold. To this end, the following data will be used:

1) DEM, source: USGS (National Elevation Dataset (NED)), resolution: 30 m

2) Stream networks, source: USGS (National Hydrography Dataset (NHD)), scale: 1:24K

3) Hydrological units, source: USGS (Watershed Boundary Dataset (WBD)), scale: 1:250K

4) Soil data, source: Soil Survey Geographic database (SSURGO), Scale: 1:24K

5) Runoff, source: source: USGS, resolution: 5 km

7) Land cover, source: USGS (National Land Cover Database 2011), resolution: 30 m

**3.2. Research Objectives**

The research objective of this proposal is to accurately extract streams from DEM. To accomplish this, this study plans to (1) automate the choice of a flow accumulation threshold for extracting stream networks from DEMs, (2) compare the effects of different flow direction methods on flow accumulation threshold, (3) investigate the influence of DEM-derived terrain properties (e.g., slope, curvature, roughness), landscape characteristics (e.g., soil, vegetation, climate), and stream network geometry (e.g., density, orders) on flow accumulation threshold, (4) investigate methods for large-scale mapping of stream networks and test the practicality of and the possibilities for the implementation of an automated USA-wide method, (5) propose a method for identifying headwaters (the location of headwater illustrates a threshold point where enough water has accumulated to form stream channels) as the main problem of stream extraction from DEM in response to the need for accurate stream maps, and finally (6) propose a comprehensive method for evaluating the accuracy of DEM-derived stream networks.

**3.3. Research Questions**

**Question 1**. How can the choice of a flow accumulation threshold be automated?

The following predictions will be tested by using terrain derivatives and landscape characteristics as ancillary data.

Prediction 1-1: There is a strong relationship terrain derivatives and landscape characteristics with the place of headwaters, and this relationship is used to choose the optimum threshold.

Prediction 1-2: DEM-derived parameters (slope, curvature, roughness, stream density, stream order, etc.) are sufficient to establish a rule for a fair threshold approximation in different terrains.

Prediction 1-3: The choice of a flow accumulation threshold cannot be completely automated due to the difficulties of mapping headwaters, and terrain heterogeneity.

**Question 2.** How can terrain derivatives and landscape characteristics inform automatic extraction of stream networks from DEMs?

To address the predictions below, terrain derivatives and landscape characteristics will be used in different ways to inform the stream extraction process; Terrain derivatives and landscape characteristics will be used to create homogeneous drainage density regions, and find a rule of assigning an optimum threshold to each region. Also, flow accumulation threshold will be weighted based on terrain derivatives and landscape characteristics. Furthermore, these characteristics will be utilized in a regression analysis to find the optimum threshold.

Prediction 2-1: Terrain derivatives are quiet useful to extract streams in some places, but cannot be employed for all of the cases without using landscape characteristics.

Prediction 2-2: Data integration would be a challenge due to the differences in scale and resolution of DEM and ancillary data.

**Question 3**. How can spatial variations in the accuracy of DEM-derived stream networks be quantified?

It is clear that the accuracy of DEM-derived stream networks varies through spatially. I this research, the accuracy of extracted streams are assessed using different accuracy measures such as stream length, stream order, stream frequency, stream density, etc., and the spatial variations in these measures are investigated. The following predictions will be tested:

Prediction 3-1: The accuracy of DEM-derived stream networks can be evaluated based on DEM information such as the stream cross sections from DEM, or structure lines form DEM.

Prediction 3-2: The accuracy of extracted streams vary based on the landscape type. Extracted streams in dry and low relief terrains have the lowest accuracy.

Prediction 3-3: Headwaters are underestimated in the available datasets such as NHD, and overestimated in the DEM-derived stream networks.

**4. Methods**

**4.1. DEM Analysis:**

Prior to investigating the effects of threshold on extracted streams, DEM is processed in the following steps. First, DEM is preprocessed in order to remove sinks (pit or depression — is a cell surrounded by higher elevation cells, in which the water is trapped and cannot flow). Sinks are often created because of data noise, interpolation errors and systematic production errors in DEM elevation values (Garbrecht & Martz 2000). Sinks cause difficulties for flow direction algorithms and should be corrected in a preprocessing step. Removing sinks called sink filling can be done by smoothing DEM (O'Callaghan & Mark 1984) or increasing the elevation of sink cell to the lowest boundary cell to be consistent with their neighboring cells (Jenson & Domingue 1988). However, it is difficult to distinguish real sinks form spurious ones, and because of that the sink filling process is often done by some level of approximation. Second, a burning-in procedure is done; DEM errors could results in some discrepancies between the extracted streams and the available mapped streams, particularly in low-relief landscapes. In order to avoid these inconsistencies, the benchmark streams can be forced into the DEM (this process is called burning-in). This way, the extracted streams exactly match a benchmark stream, and this method also would be especially useful in flat terrain where the available algorithms do not work properly. Burning-in procedure is usually done by raising non-stream pixels with an arbitrary elevation (Maidment 1996) or artificially decreasing the elevations of stream pixels using another data source (Tarboton 1997). This procedure make a hydrologically corrected DEM and results in a more accurate stream network. However, the quality of extracted stream networks is dependent on the quality of the stream map, and using a stream map compiled at a different scale than the DEM results in misplacement of the extracted streams (Hastings 2012). In a third step, flow directions are calculated. In this research, D8 and D-infinity algorithms are employed as the most commonly used methods to investigate their effects on flow accumulation matrix and consequently on threshold.

**4.2. Threshold Analysis:**

The focus of this work is on the choice of an optimum flow accumulation threshold. Different methods have been proposed to select the threshold. The performance of common methods will be evaluated in this research. In fact, the aim of this research is it to develop a flow accumulation threshold model based on the existing methods, and to test model performance in extracting stream networks.

**Constant threshold method:** In this method, which is the most common method, the stream networks are extracted when the values of flow accumulation matrix exceed a threshold value. In fact, this threshold is the minimum drainage area required to initiate a stream. This threshold is often chosen based on visual comparison between the extracted streams and a benchmark.

**Area-slope method:** In this method, for a range of thresholds, streams are extracted and the slope of headwater streams against their flow accumulation areas are plotted. Then the best threshold is chosen based on the change in the slope of regression line. Montgomery and Foufoula-Georgiou (1993) propose an area–slope relationship in real landscapes to extract headwaters. This relationship is:

Acr = 1790 (tan ϴ)-1.84

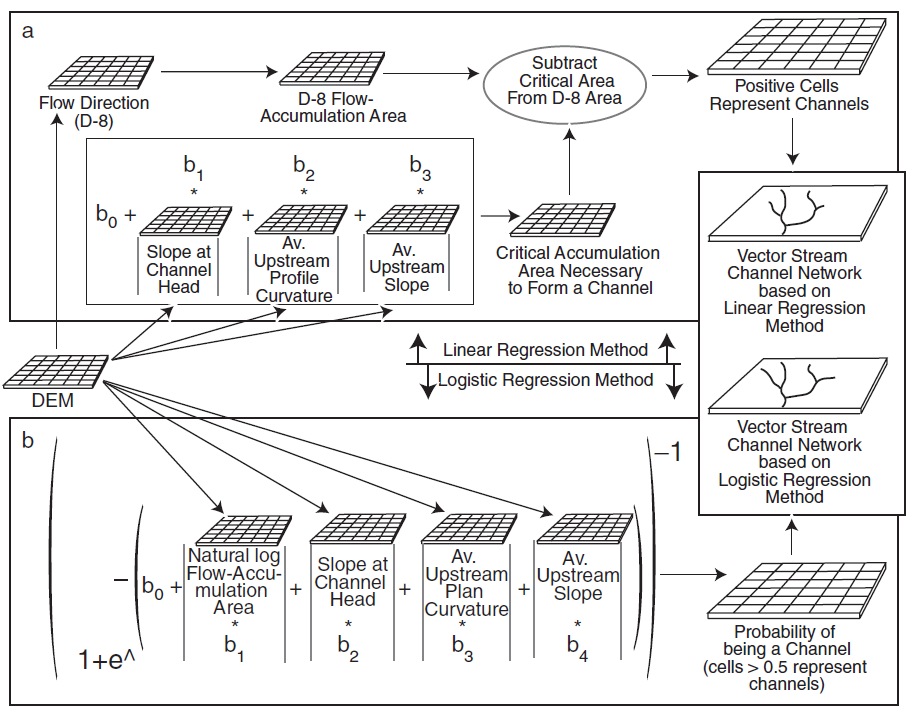
Acr is the ‘‘critical area’’ (i.e., flow accumulation threshold), and tanϴ is the slope at the headwater. Based on this technique, the greater the slope at the headwater, the smaller the flow-accumulation threshold required to cause a stream.

**Local curvature method:** Local curvature can be used as a weight grid in a flow accumulation function in order to preserve spatial variability in stream density, and differentiate between hillslopes and streams (Tarboton & Ames 2001). Plan and profile curvature are mainly used to investigate the effect of topography on streamflow. Profile curvature is the curvature of the surface in the direction of maximum slope. It influences erosion and deposition by affecting the acceleration or deceleration of flow. Plan curvature, on the other hand, is perpendicular to the direction of the maximum slope. It influences convergence and divergence of flow (Kimerling, Buckley et al. 2009). Heine et al. (2004) investigated the effects of profile and plan curvature over the entire upslope flow-accumulation area. They found that there is positive relationship between both profile and plan curvature and upslope flow accumulation area at the headwaters.

**Stream ordering method:** There are several methods for assigning orders to stream networks (e.g., Strahler and Shreve). In the Shreve method, the order of each stream is equal to the number of up streams in the network (Shreve 1966). All streams with no tributaries are assigned an order of one. At the point where two streams intersect, their orders are added and assigned to the downslope stream. For example, the intersection of orders 2 and 4 form an order 6 stream. In the Strahler method, headwater streams are assigned an order of one, and order increases at the point where two streams of equal order join (Strahler 1957). But the order does not increase when two streams with non-equal order join. A threshold can be defined in order to extract streams with an order greater than the threshold (Peckham 1995).

**Stream density method:** Calculating stream density can be done by first approximating catchment area sizes by Thiessen polygons (Stanislawski et al. 2007), next delineating density partitions by a rasterization, buffering and re-vectorization sequence, and then dividing the channel length (km) by the estimated catchment size (km2) within each density partition (Stanislawski & Buttenfield 2011). A power law relationship proposed by Moglen et al. (1998) illustrates the relationship between stream density (Sd) and flow accumulation threshold (As):

**Variable threshold method using regression analysis**: Some studies have tried to incorporate terrain derivatives and landscape characteristics into the threshold selection process. A variable flow accumulation threshold method using topographic characteristics as independent variables can improve stream network delineation results. Regression analysis is usually used to establish a relationship between flow accumulation area and related factors and find the initial points of streams. Fig. 3 illustrates two variable flow accumulation threshold methods using multiple liner regression and logistic regression. In this study, a relationship has been established between flow accumulation area at stream heads and four key independent variables (slope at headwater, average upstream slope, average upstream plan curvature, and average upstream profile curvature), which can be extracted from DEMs (Heine, Lant et al. 2004). Previous literature shows that the logistic regression model leads to the most accurate results for extracting stream networks (Heine et al., 2004; Jaeger et al., 2007). Other variables, such as climate, soil, vegetation, and geology, also influence the existence of streams, but were not incorporated in their research, which will be investigated in this study.

****

**Fig 3**: Variable flow accumulation estimation using: (a) linear regression method, and (b) logistic regression method (Heine, Lant et al. 2004)

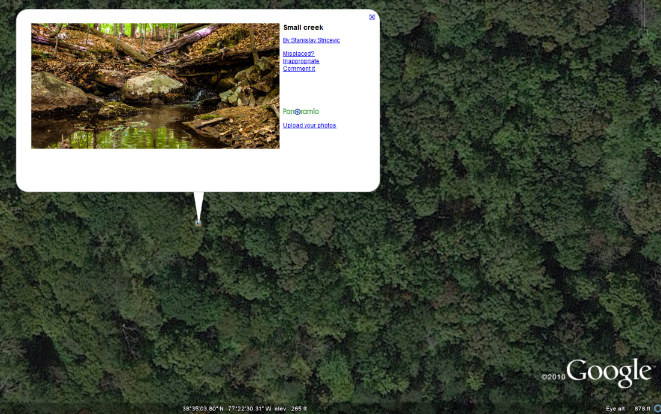
**Weighted flow accumulation matrix:** Flow accumulation matrix can be weighted based on the terrain derivatives and landscape characteristics. A spatially variable raster weight grid is computed based on surface runoff, soil permeability, soil depth, groundwater contributions to streams, and terrain slope to make weighted flow accumulation matrix (Stanislawski et al. 2012, Stanislawski et al. 2014). This method reflect natural drainage density, and can substantially improve the results of stream extraction.

However, previous studies investigating the effects of a subset of terrain factors in stream network extraction for a few study area have not examined all the corresponding factor together for many different landscapes at the same time, and such research is currently needed. Thus, the proposed study will tailor stream network extraction methods to the topographic and environmental characteristics, as well as stream geometry.

**4.3. Verification of Extracted Stream Networks:**

As discussed earlier, several methods are employed for choosing the appropriate flow accumulation threshold. The most widely used are the constant threshold method, and area-slope method, which lead to a high amount of inaccuracy. The number, length and density of extracted streams are highly dependent on the selected threshold. Finding a method to evaluate the accuracy of DEM-derived streams is another challenge that this study is going to address.

The accuracy of extracted streams can be evaluated by performing a considerable amount of field surveys that are time-consuming and costly. Mapping streams from aerial photographs or satellite images is another option (Yang et al. 2014). Although this method is considered an accurate way of mapping streams, there are some limiting factors such as river size (in comparison to the spatial resolution of the sensor), water quality, stream depth, and vegetation obstructions and shadows (Fig. 4). Existing stream network datasets also can be used. The National Hydrography Dataset (NHD) is maintained by the U. S. Geological Survey (USGS) to represent surface water. It is a digital spatial dataset of surface water features (e.g., rivers, streams, canals, lakes, ponds, coastlines, dams, and stream gages) for the entire United States. The NHD has a key role in topographic mapping, hydrological modeling, and water management. However, this dataset contains incomplete or missing stream channels and discontinuities in some subbasins, especially in headwater areas. Heine et al. (2004) demonstrate that NHD under-represented the streams by 64%. Therefore, accurate mapping and evaluation of headwater streams is problematic.



**Fig 4**: A headwaters located in a forested watershed in the mid-Atlantic United States

In order to find the best threshold selection method, extracted streams are assessed using the following accuracy measures that address different components of the stream networks: a) stream length measured for the entire drainage basin, (b) stream order based on the Strahler method to classify streams based on the number of upstream segments, (c) stream density calculated by dividing the sum of the length of all the streams in a drainage basin divided by the drainage basin area, (d) stream frequency calculated by dividing the number of all the streams in a drainage basin divided by the drainage basin area, (e) the number of headwaters (first order streams), and (f) the Coefficient of Line Correspondence (CLC) described below.

The CLC measure proposed by Stanislawski et al. (2010) is used to estimate how well two sets of line features match. In fact, The CLC computes the conflation between the stream network generated from the DEM and a stream data used as a benchmark. Conflation occurs when the dissimilarity between two different objects appears to be lost, such as when a DEM-derived stream network and the benchmark stream network appear as one visually on a map. When the two datasets do not match up, this is the result of omissions or commissions. Omissions occur when DEM-derived stream length is less than the benchmark stream network; commissions occur when DEM-derived stream length is greater than the benchmark stream network. The equation used is:

CLC =

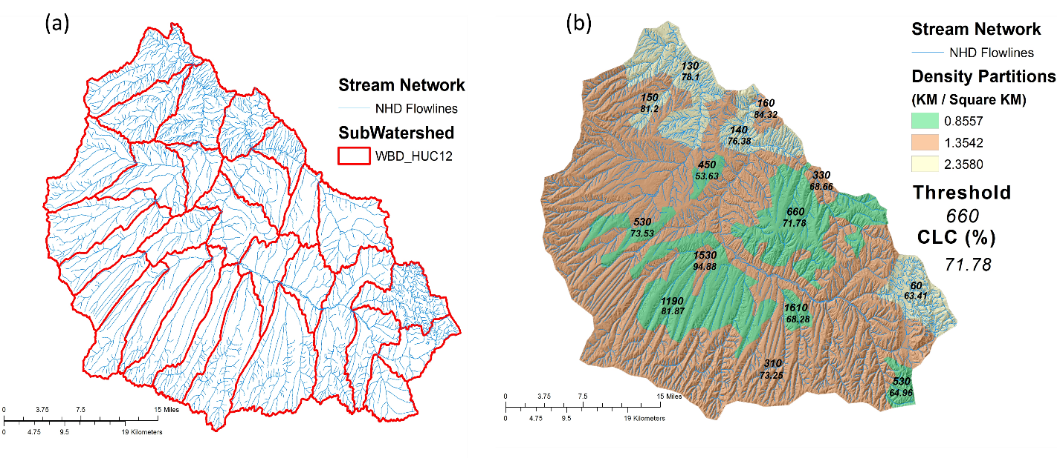
To find out which streams were being omitted from the modeled results, extracted streams are buffered based on the National Map Accuracy Standard (NMA) - 1/30th of an inch at scale (approximately 20 m for 1:24k). The buffer is twice this distance (40m) to capture stream divergence on both sides which remained within the standard. Lengths of either the modeled or benchmark streams which fall within 20 m of one another were assumed to be corresponding.

**5. Preliminary Results**

**5.1. Case Study: Density-Based Stream Network Extraction from Digital Elevation Models**

In this research, an automated workflow extracts a sequence of stream networks from a flow accumulation matrix using a progression of threshold values. Then, to find the optimum threshold, the extracted stream networks are compared to a benchmark data source (here, the 1:24,000 High Resolution NHD). Comparison is computed using a conflation ratio between matching stream channel lengths to matches and mismatches (here, the CLC is used), as well as an analysis of the spatial distribution of mismatches in the stream network (e.g., first order tributaries, larger order channels, or throughout the network). Although this basic workflow runs successfully, the final extracted stream networks do not account for local differences in stream density (Fig. 5-a), which have been shown in previous research to be cartographically and hydrologically important (Buttenfield, Stanislawski, & Brewer, 2010). Local density variations are important to hydrologic analysis, and using one single threshold value for extracting stream networks for region characterized by varying stream channel densities can distort hydrological modeling results in subsequent flow or length computations.

This research proposes a new method to determine the best thresholds within previously established channel density partitions of the hydrological unit of interest. To obtain the best threshold, flow direction and flow accumulation are first calculated for the whole study area. Next, the study area is categorized on the basis of stream network density that is calculated from the vector NHD. The selection of an optimal flow accumulation threshold value iterates through a progression of values selected through a binary search. Results are compared to the benchmark data set, until a threshold value achieves the stream channel that is closest to the benchmark data set (Fig. 5-b). The results illustrate that the proposed method provides a more geographically and hydrologically valid database representation of elevation-derived stream networks. One area that requires further investigation in developing this solution is that the vector NHD tends to miss most channels in headwater regions. One possible solution is to run benchmark comparisons against only 2nd order and higher channels in the network, and this forms one area to investigate. Another important constraint on the solution is that channel connectivity must be maintained in the final set of selected streams as the network crosses density partitions, and this will be addressed by topological checking of the final set of derived streams. To justify the development of a sound model for selecting density partition thresholds, four subbasins were tested to generate results in heterogeneous subbasins with diverse landscape differences. The results illustrate that the proposed method provides a more geographically and hydrologically valid database representation of elevation-derived stream networks. The highlights of the proposed approach for stream network extraction are: (1) It can be used for large hydrological units, (2) The selection of an optimal flow accumulation threshold value is done automatically, (3) The streams are extracted with a higher accuracy in a density partition, and (4) the results illustrate that there is a correlation between threshold and density. In the future, we will extend the approach to automatically extract the density partitions from flow accumulation matrix instead of using density partitions as an ancillary data. Also, one area that requires further investigation in developing this solution is that the vector NHD tends to miss most channels in headwater regions. One possible solution is to run benchmark comparisons against only 2nd order and higher channels in the network, and this forms one area to investigate. Another important constraint on the solution is that channel connectivity must be maintained in the final set of selected streams as the network crosses density partitions, and this will be addressed by topological checking of the final set of derived streams.



**Fig 5.** The Piceance-Yellow River subbasin (HUC 14050006), located in northwestern Colorado. This region is defined by its dry climate and mountainous terrain, with a mean elevation of 2202 m and an average slope of 26.7%. (a) Subwatersheds (HUC 12s) in the sub-basin are identified. Choosing one threshold for each subwatershed leads to extracted stream networks that do not account for local density variations. (b) This subbasin has three density partitions that are applied to the fourteen separate polygons within the HUC 12 regions. The optimum threshold and its CLC is identified for each density polygon.

**6. Broader Impacts**

Among the broader impacts associated with this project, 4 stand out as potentially transformative: (1) Currently USGS is quite interested to extracts streams from DEMs and improve the NHD. Close collaboration with employees of the USGS will guarantee the distribution of the research results to the relevant agencies. (2) This research has a real-world need. Although this research is developed based on subbasins in the United States, it will provides a model workflow for other parts of the world. In fact, the research goals of this proposal have the potential to result in a more efficient and cost-effective solutions to a world's hydrology problem. (3) Research findings will be integrated into core Geography courses at the University of Colorado including fluvial geomorphology, hydrology, and GIS modeling courses, allowing students to broaden their understanding of the application of GIS in hydrology and geomorphology. In order to develop a stronger curriculum, also, the request for proposing a new course titled “GIS in water resources” will be proposed. (4) This project started as class project collaborating with three undergraduate students that was a rewarding experience. They raised interesting questions in project and made significant contributions to the research. This proposal also seeks summer funding for two students to participate in research. For validation, a method will be proposed based on structure lines in Triangular Irregular Network (TIN), which needs lots of work to do.

**7. Research Calendar**

*Summer 2015:* literature review

*Fall 2015:* literature review, and data preparation

*Spring 2016:* Compare the effects of different flow direction methods on flow accumulation threshold

*Summer 2016*: Investigate the influence of DEM-derived terrain properties, landscape characteristics, and stream network geometry on flow accumulation threshold

*Fall 2016:* Evaluating the accuracy of DEM-derived stream networks

*Spring 2017:* Proposing a method to automate the choice of a flow accumulation threshold for extracting stream networks from DEMs

*Summer 2017:* Present findings at the annual conference of the Association of American Geographers

*Fall 2017:* Present findings at the Water Resources Research journal

*Spring 2018:* Write doctoral dissertation; Prepare manuscripts for submittal to appropriate journals

**References**

Bai, R., T. Li, Y. Huang, J. Li and G. Wang (2015). "An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree." Geomorphology 238: 56-67.

Barnes, R., C. Lehman and D. Mulla (2014). "An efficient assignment of drainage direction over flat surfaces in raster digital elevation models." Computers & Geosciences 62: 128-135.

Bhowmik, A. K., M. Metz and R. B. Schäfer (2015). "An automated, objective and open source tool for stream threshold selection and upstream riparian corridor delineation." Environmental Modelling & Software 63: 240-250.

Chaubey, I., A. Cotter, T. Costello and T. Soerens (2005). "Effect of DEM data resolution on SWAT output uncertainty." Hydrological Processes 19(3): 621-628.

Clubb, F. J., S. M. Mudd, D. T. Milodowski, M. D. Hurst and L. J. Slater (2014). "Objective extraction of channel heads from high‐resolution topographic data." Water Resources Research 50(5): 4283-4304.

Coblentz, D., F. Pabian and L. Prasad (2014). "Quantitative Geomorphometrics for Terrain Characterization." International Journal of Geosciences 2014.

Costa‐Cabral, M. C. and S. J. Burges (1994). "Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas." Water resources research 30(6): 1681-1692.

Data, E., E. Characteristics, R. Colombo, J. Vogt, J. Vogt and F. Bertolo (2001). "Deriving drainage networks and catchment boundaries at the European scale." Joint Research Centre, European Commission, Ispra.

Fairfield, J. and P. Leymarie (1991). "Drainage networks from grid digital elevation models." Water Resources Research 27(5): 709-717.

Frankel, K. L. and J. F. Dolan (2007). "Characterizing arid region alluvial fan surface roughness with airborne laser swath mapping digital topographic data." Journal of Geophysical Research: Earth Surface (2003–2012) 112(F2).

Freeman, T. G. (1991). "Calculating catchment area with divergent flow based on a regular grid." Computers & Geosciences 17(3): 413-422.

Garbrecht, J. and L. W. Martz (1997). "The assignment of drainage direction over flat surfaces in raster digital elevation models." Journal of hydrology 193(1-4): 204-213.

Garbrecht, J. and L. W. Martz (2000). "Digital elevation model issues in water resources modeling." Hydrologic and hydraulic modeling support with geographic information systems: 1-28.

Hastings, B. (2012). Comparison of digital terrain and field-based channel derivation methods in a subalpine catchment, Front Range, Colorado. Master thesis, Colorado State University.

Hastings, B. E. and S. K. Kampf (2014). "Evaluation of digital channel network derivation methods in a glaciated subalpine catchment." Earth Surface Processes and Landforms 39(13): 1790-1802.

Heine, R. A., C. L. Lant and R. R. Sengupta (2004). "Development and comparison of approaches for automated mapping of stream channel networks." Annals of the Association of American Geographers 94(3): 477-490.

Hobson, R. (1972). "Surface roughness in topography: quantitative approach." Spatial analysis in geomorphology: 225-245.

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.

Istanbulluoglu, E. and R. L. Bras (2005). "Vegetation‐modulated landscape evolution: Effects of vegetation on landscape processes, drainage density, and topography." Journal of Geophysical Research: Earth Surface (2003–2012) 110(F2).

Jaeger, K. L., D. R. Montgomery and S. M. Bolton (2007). "Channel and perennial flow initiation in headwater streams: management implications of variability in source-area size." Environmental Management 40(5): 775-786.

Jenson, S. and J. Domingue (1988). "Extracting topographic structure from digital elevation data for geographic information system analysis." Photogrammetric engineering and remote sensing 54(11): 1593-1600.

Jones, N. L., S. G. Wright and others (1990). "Watershed Delineation with Triangle-Based Terrain Models." Journal of Hydraulic Engineering 116: 1232.

Kimerling, A. J., A. R. Buckley, P. C. Muehrcke and J. O. Muehrcke (2009). Map use: reading and analysis, Esri Press.

Kirkby, M. J. (1987). "Modelling some influences of soil erosion, landslides and valley gradient on drainage density and hollow development." Geomorphological models: theoretical and empirical aspects/Frank Ahnert (editor).

Koundouri, P. and N. A. Papandreou (2013). Water Resources Management Sustaining Socio-economic Welfare: The Implementation of the European Water Framework Directive in Asopos River Basin in Greece, Springer Science & Business Media.

Lea, N. (1992). "An aspect driven kinematic routing algorithm." Overland flow: hydraulics and erosion mechanics: 393-407.

Luo, W., X. Li, I. Molloy, L. Di and T. Stepinski (2014). "Web service for extracting stream networks from DEM data." GeoJournal 79(2): 183-193.

Maidment, D. R. (1996). GIS and hydrologic modeling-an assessment of progress. Third International Conference on GIS and Environmental Modeling.

Mao, Y., A. Ye, J. Xu, F. Ma, X. Deng, C. Miao, W. Gong and Z. Di (2014). "An advanced distributed automated extraction of drainage network model on high-resolution DEM." Hydrology and Earth System Sciences Discussions 11(7): 7441-7467.

Mark, D. M. (1983). "Relations Between Field‐surveyed Channel Networks and Map‐based Geomorphometric Measures, Inez, Kentucky." Annals of the Association of American Geographers 73(3): 358-372.

McMaster, K. J. (2002). "Effects of digital elevation model resolution on derived stream network positions." Water Resources Research 38(4): 13-11-13-18.

Moglen, G. E., E. A. Eltahir and R. L. Bras (1998). "On the sensitivity of drainage density to climate change." Water Resources Research 34(4): 855-862.

Montgomery, D. R. and W. E. Dietrich (1988). "Where do channels begin?" Nature 336(6196): 232-234.

Montgomery, D. R. and E. Foufoula‐Georgiou (1993). "Channel network source representation using digital elevation models." Water Resources Research 29(12): 3925-3934.

Morisawa, M. and K. M. Clayton (1985). Rivers: form and process, Longman London and New York.

Mower, J. E. (1994). "Data-parallel procedures for drainage basin analysis." Computers & Geosciences 20(9): 1365–1378.

Nelson, E. J. and others (1994). "Algorithm for Precise Drainage-Basin Delineation." Journal of Hydraulic Engineering 120: 298.

O'Callaghan, J. F. and D. M. Mark (1984). "The extraction of drainage networks from digital elevation data." Computer vision, graphics, and image ….

Passalacqua, P., P. Tarolli and E. Foufoula‐Georgiou (2010). "Testing space‐scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape." Water resources research 46(11).

Peckham, S. D. (1995). "New results for self‐similar trees with applications to river networks." Water Resources Research 31(4): 1023-1029.

Pelletier, J. D. (2013). "A robust, two‐parameter method for the extraction of drainage networks from high‐resolution digital elevation models (DEMs): Evaluation using synthetic and real‐world DEMs." Water Resources Research 49(1): 75-89.

Peucker, T. K. and D. H. Douglas (1975). "Detection of surface-specific points by local parallel processing of discrete terrain elevation data." Computer Graphics and image processing.

Płaczkowska, E., M. Górnik, E. Mocior, B. Peek, P. Potoniec, B. Rzonca and J. Siwek (2015). "Spatial distribution of channel heads in the Polish Flysch Carpathians." CATENA 127: 240-249.

Rosim, S., J. de Oliveira Ortiz, J. R. de Freitas Oliveira, A. C. Jardim and E. S. Abreu (2015). Drainage Network Definition for Low Resolution DEM Obtained from Drainage Network Extracted from High Resolution DEM Using Upscaling Processes. Engineering Geology for Society and Territory-Volume 3, Springer: 271-274.

Shreve, R. L. (1966). "Statistical law of stream numbers." The Journal of Geology: 17-37.

Stanislawski, L. and B. Buttenfield (2011). A raster alternative for partitioning line densities to support automated cartographic generalization. 25th International cartography conference, July.

Stanislawski, L., A. Doumbouya, C. Miller-Corbett, B. Buttenfield and S. Arundel (2012). Scaling stream densities for hydrologic generalization. 7th international conference on geographic information science, Columbus.

Stanislawski, L. V., B. P. Buttenfield and V. A. Samaranayake (2010). "Automated metric assessment of hydrographic feature generalization through bootstrapping." … of the 13th Workshop of ICA ….

Stanislawski, L. V., A. T. Doumbouya and B. P. Buttenfield (2014). Hydroloagically Consistent Pruning of the High-Resolution National Hydrography Dataset to 1:24,000-scale. American Water Resources Association Spring Specialty Conference. Salt Lake City UT.

Stanislawski, L. V., M. Finn, E. Usery and M. Barnes (2007). "Assessment of a rapid approach for estimating catchment areas for surface drainage lines." ACSM-IPLSAMSPS 2007: 9-12.

Strahler, A. N. (1957). "Quantitative analysis of watershed geomorphology." Eos, Transactions American Geophysical Union 38(6): 913-920.

Strahler, A. N. (1964). "Quantitative geomorphology of drainage basin and channel networks." Handbook of applied hydrology.

Tarboton, D. G. (1997). "A new method for the determination of flow directions and upslope areas in grid digital elevation models." Water resources research 33(2): 309-319.

Tarboton, D. G. and D. P. Ames (2001). Advances in the mapping of flow networks from digital elevation data. World water and environmental resources congress, Am. Soc Civil Engrs USA.

Tarboton, D. G., R. L. Bras and I. Rodriguez-Iturbe (1991). "On the extraction of channel networks from digital elevation data." Hydrological processes 5(1): 81-100.

Tarboton, D. G., R. L. Bras and I. Rodriguez-Iturbe (1992). "A physical basis for drainage density." Geomorphology 5(1): 59-76.

Tarolli, P., G. Sofia and G. Dalla Fontana (2012). "Geomorphic features extraction from high-resolution topography: landslide crowns and bank erosion." Natural Hazards 61(1): 65-83.

Tucker, G. E. and R. L. Bras (1998). "Hillslope processes, drainage density, and landscape morphology." Water Resources Research 34(10): 2751-2764.

Wang, L. and K. Wang (2014). "Impacts of DEM Uncertainty on Estimated Surface Solar Radiation and Extracted River Network." Bulletin of the American Meteorological Society.

Yang, K., M. Li, Y. Liu, L. Cheng, Y. Duan and M. Zhou (2014). "River delineation from remotely sensed imagery using a multi-scale classification approach."

Yang, P., D. P. Ames, A. Fonseca, D. Anderson, R. Shrestha, N. F. Glenn and Y. Cao (2014). "What is the effect of LiDAR-derived DEM resolution on large-scale watershed model results?" Environmental Modelling & Software 58: 48-57.