Automated Geoprocessing Workflow for Watershed Delineation and Classification for Flash Flood Assessment

Schröder, D.,[1] Omran, A. F.[2] and Bastidas, M. R. R.[1]
[1]Department of Geomatics, Computer Science and Mathematics, University of Applied Sciences Stuttgart, Schellingstraße 24, D-70174, Stuttgart, Germany
E-mail: dietrich.schroeder@hft-stuttgart.de, 42bara1mpg@hft-stuttgart.de
[2]Department of Science and Mathematical Engineering, Faculty of Petroleum and Mining Engineering, Suez University- Saleh Neseem Road, El Zaiteiat-Ataqa, Suez, Egypt, E-mail: adelfouad.omran@suezuniv.edu.eg

Abstract

The characteristics of a drainage basin have a major effect on the risk potential for flash floods. For a detailed study on the risk potential detailed topographic, geological, soil related, meteorological, and historical data is needed, which is not available in many regions of the world. On the other hand, high resolution Digital Elevation Models are ready to use, from which topographic parameters of a watershed related to the risk potential can be extracted. In this study a complete automated workflow based on ArcGIS ModelBuilder 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 model has been applied to the upper reach of the Akhangaran river south-east of Tashkent in Uzbekistan, which covers a total area of about 3600 km2. For the study area according to Strahler order 4 sub-basins of 6th order, 20 of 5th order, 71 of 4th order , 326 of 3rd order, and 1559 of 2nd order have been classified. The sub-basins have been evaluated according to the stream bifurcation ratio, the drainage density, the stream frequency, the circular ratio, the elongation ratio, and some other parameters.

1. Introduction

According to reinsurance companies like Münchener Rück (Munich, 2014), in the last decades losses due to damages on infrastructure as well as casualties caused by extreme weather events including flash floods have considerably increased. Flash floods are the most lethal form of natural hazard (based upon the ratio of fatalities to people affected), and cause millions of dollars in property damage every year. As reported by the Intergovernmental Panel on Climate Change (IPCC, 2012), it is expected that the number of extreme weather events leading to flash floods will increase even more due to climate change in the future. On the other hand, the vulnerability of many areas has increased due to population growth and rural depopulation leading to more intensive use of peri-urban areas including land with higher risk potential on flash flood, mudslides and landslides. Flash floods are rapid-onset hydrologic events that can be difficult to forecast. A combination of a high rainfall rate with rapid and often efficient runoff production processes is common to most flash flood events. Hydrologic influences of the ground surface can have a major impact on the timing, location, and severity of flash flooding.

Although rainfall is often considered the most important factor for the occurrence floods, other factors influencing the run-off characteristics are soil and basin properties. In this study, soil characteristics like moisture, permeability, or profile will be not considered, but the focus will be on the basin and the stream characteristics. The physical properties of a basin and its stream network have an effect on the amount and the timing of runoff and therefore the likelihood of flash flooding in the basin's outlet. Any factor that increases the speed and efficiency of runoff production processes can make a particular basin more prone to flash flooding. Thus fewer stream meanders, steep slopes, less surface roughness, and high stream density are some of the factors that may increase the susceptibility of a basin to flash floods. For a detailed study on the risk potential detailed topographic, geological, soil related, meteorological, and historical data is needed, which is not available in many regions of the world. On the other hand, high resolution Digital Elevation Models like GDEM2 data from ASTER are ready to use, from which topographic parameters of a watershed related to the risk potential can be extracted.

DEM based classification of basins may not show as many details as a field study would give. But it can be very well used for a statistical comparison of watershed characteristics and a relative risk potential for flash floods can be determined. There are many studies where for watersheds in different areas of the world an assessment based on the basin characteristics related to flash floods has been done, e.g. El-Moustafa and Mohamed (2013) for the Sinai Peninsula, Azab (2009) for three basins in Egypt draining to the Red Sea, Omran (2013) for Wadi Dahab on Sinai Penisula, Esper Angillieri for the Carrrizal basin in Argentina, Pareta and Pareta (2011) for the Yamuna Basin in India, just to mention some of the recent studies. Most of the studies are using a Digital Elevation Model (DEM) as input data for a GIS based evaluation of the basin and stream characteristics. To process regions of larger size, where maybe hundreds of watersheds have to be delineated and classified, the overall process has to be automated. GIS like ESRI's ArcGIS offer a number of sophisticated tools which help in this analysis. But there are still some gaps and pitfalls to be considered, if the tools are combined to a geoprocessing model to automate the complete assessment workflow. Thus usual in the study mentioned above some cumbersome manual work is involved and the studies are focusing on a restricted number of watersheds involved. This paper will focus on the automated geoprocessing workflow to extract basin and stream network parameters to produce a flash flood hazard risk map based on the results of the morphometric analysis.

2. Study Area and Input Data

As study area, the sub-basin of the upper reach of the Akhangaran river south-east of Tashkent in Uzbekistan has been used, which covers a total area of about 3500 km^2. The highest peak in the watershed is about 4050 m.a.s.l. whereas the outlet point used in the study has an elevation of about 500 m.a.s.l. The study area includes also the Angren reservoir with a designed storage capacity of 260Mm^3 and a full storage level at 1075 m.a.s.l. Geologically, the area is covered by upper Paleozoic volcanic rocks from porphyries and porphyrites rocks. Sands and sandstone of Suzak Formation and limestone of Alay Formation with layers of clay overlay the volcanic rocks. The upper most Quaternary deposits that cover the rock comprise prolluvial silts with a thickness from 3m – 5m to 10m - 45 m and boulder pebble deposits of Akhangaran River and silty gravel.

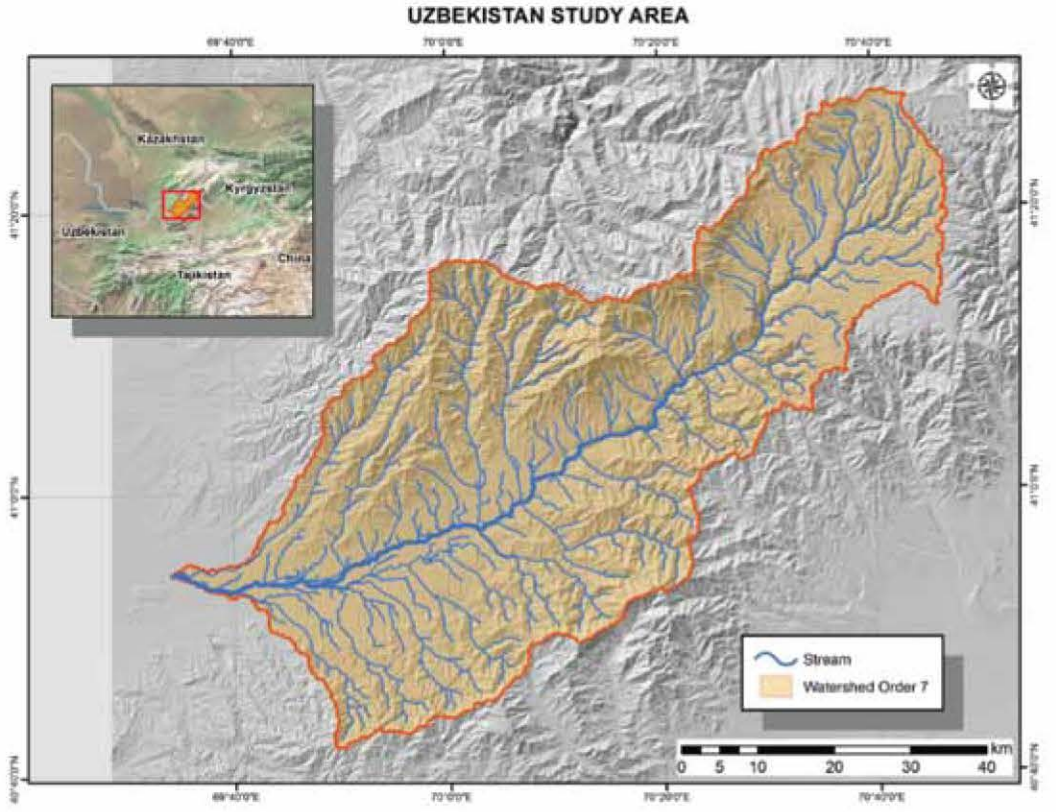

Figure 1: Location of study area

According to Djumaev (2010), most of the study area is classified as having mud and flash floods hazards frequently, i.e. one to three times a year. As input data for the processing ASTER GDEM 2 has been used, which is a product of METI and NASA (METI and NASA (2011)). GDEM 2 has the same gridding and tile structure as GDEM 1.0 (1 arc-second elevation grid divided and distributed as 1° x 1° tiles based on 1.2 million scene-based DEMs of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) on the National Aeronautic and Space Administration (NASA) spacecraft Terra), but benefits from the inclusion of additional scenes to reduce artifacts, higher horizontal resolution using a smaller correlation kernel (5 x 5 versus 9 x 9 used for GDEM 1.0), and an improved water mask. GDEM 2 has an overall accuracy of around 17 m at the 95% confidence level, and a horizontal resolution on the order of 75 m. GDEM 2 includes a lot of improvements for hydrological applications with respect to the predecessor GDEM 1, as in particular pits and spikes and other artifacts have been removed from the DEM. Thus from GDEM 2 – in particular in open and bare land cover areas - the overall drainage pattern can be extracted and based on that the watersheds characteristics can be calculated. The tiles for the study area have been mosaicked, clipped and re-projected to UTM24N WGS84 using ESRI's ArcGIS desktop before further geoprocessing (Figure 1).

3. Methodology

The workflow implemented is based on tools available in the ArcGIS 10.2 toolbox; in particular tools from the Spatial Analyst extension are used. The individual tools have been combined using the ArcGIS ModelBuilder for batch processing. Missing tools related to control the geoprocessing order and to manipulate vector data have been implemented in Python, whereas tools for raster manipulation have been implemented in Java based on ArcObjects. The major steps are as follows:

(i) preparing the DEM by filling the pits and depressions using the *Fill* tool
(ii) calculating the local drainage direction raster using the *Flow Direction* Tool
(iii) calculating the upstream drainage area for each raster cell using the *Flow Accumulation* tool
(iv) extracting a binary stream network by setting a threshold value for the catchment area defining the initial source of a stream
(v) segmenting and assigning Strahler order numbers for each link using the *Strahler Order* tool
(vi) re-ordering of the stream links according to their Strahler order
(vii) calculating pour points based on the confluence points of the tributaries according to their Strahler Order number
(viii) delineating the watersheds based on the pour points using the *Watershed* tool
(ix) extracting for each watershed the maximum flow length, the maximum and minimum elevation etc.
(x) calculating the risk parameters based on the extracted streams and delineated watersheds, elevation and slope raster
(xi) normalizing the parameter values p_i: $(p_{max} - p_i)/(p_{max} - p_{min})$, where p_{max} and p_{min} are the maximum respectively the minimum values for this parameter
(xii) summarizing the risk parameters and classify them in low, moderate and high risk

3.1 Stream Network Extraction and Strahler Ordering

In hydrology, different stream order and related watershed order systems are in use. One of the systems most used is the ordering system according to Strahler. The Strahler ordering system is based on the hierarchy of tributaries. Initial links will be of order 1. Following downstream, the order number will be increased at the confluence of two or more tributaries of the same order. Considering the topological structure of a stream network, the order numbers are defined by the following rules:

- if the node has no children, the Strahler order of its downstream link is 1, e.g. all streams having a singular source are of order 1
- if a node has one and only one tributary of order i while all the other tributaries have order less than 1, the order of its downstream link will not change
- if a node has two or more tributaries with highest order i, then the Strahler order of the downstream link will be i+1.

With respect to other ordering systems, Strahler order has the advantage that all catchments with streams are directed graphs, oriented from the root towards the leaves. On the other hand, the order number of the main stream is not clear defined, i.e. the highest order number for a study area depends on how the source nodes are defined. The Strahler order tool of ArcGIS assigns the order number to the stream segments according to the topological structure of the network, i.e. at each node a new link will start regardless whether the stream order number will change or not. Thus based on the topology and the Strahler order numbers a merging of the stream links is necessary. The algorithm to achieve this has been implemented as a Python script so it can be used as a tool in the Model Builder. The algorithm itself is explained in Omran et al., (2011). Another problem is how to fix the threshold value for the stream network extraction. Some of the parameters for the risk analysis like stream frequency or bifurcation ratio are highly depending on this value. Unfortunately, there is no easy way to specify a correct threshold value. Different approaches have been discussed in literature, ranging from “best practice” to using statistical approaches up to using expert knowledge. For this study, the latter has been used. By expert opinion, the most probable highest order of the total watershed has been fixed. This number is used as input for a model, see figure 2, which iterates starting from an initial threshold value until the algorithm results in streams according to the specified order number. This threshold value is used in the further analysis. The next issue to be discussed in the workflow is the determination of the pour points for each of the node of the Strahler ordered stream network. As the pour points have to be aligned with the flow accumulation layer for the automatic extraction of the watersheds, the manipulation of raster cells are necessary.

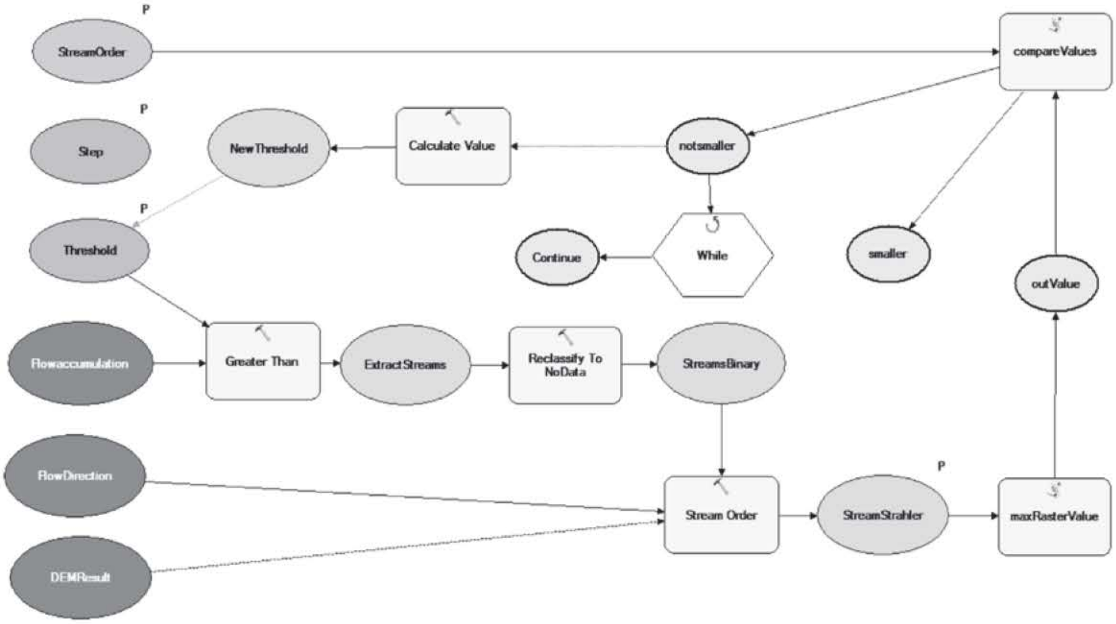

Figure 2: ArcGIS Model Builder showing the sub model to determine the threshold value

```
Input:
    Strahler order streams (vector)
    Strahler order streams (raster)

Algorithm PourPoint:
    listIntersectionPoints = list[ ]
    listPourPoints = list [ ]
    separate stream segments according to their Strahler order
    for loop iOrder = 1 to maxOrder:
        intersect stream segments of iOrder with the stream segments of iOrder to maxOrder
        convert the intersection points to raster
        for each pixel
            if pixel is intersection point: add the pixel location to listIntersectionPoints
        for each item in listIntersectionPoints
            for each of the 8 neighbors of the item's location in the Strahler stream raster
            if the pixel's value is of iOrder add the pixels to listPourPoints
        create a new raster for the items in listPourPoints
```

Figure 3: Pour point algorithm in pseudo-code notation

Table 1: Morphometric parameters used for flash flood risk mapping

Morphometric Parameter	Formula
Basin Area Size (R_a)	$R_a = A$ A = Area of the basin inn km^2
Circularity ratio (R_c)	$R_c = 4 \pi A / P^2$ P = Perimeter of basin
Elongation ratio (R_e)	$R_e = 2/L_m \cdot \sqrt{A/\pi}$ L_m = Maximum basin length in km
Mean Slope (S_m)	$S_m = 1/n \sum_{all\ pixel\ in\ basin} s_i$ n = number of pixels in basin s_i = slope value of pixel in degree
Relief Ratio (R_r)	$R_r = (H_{max} - H_{min}) / L_m$ H_{max}, H_{min} the highest, respectively lowest elevation in km
Bifurcation ratio (R_b)	$R_b = N_u / N_{u+1}$ N_u = Total number of stream segments of order u. For sub-basin with several R_b the average will be used.
Drainage density (D)	D = L / A L = Total stream length of all orders in km
Stream frequency (F)	F = N / A N = Total number of streams of all orders
Ruggedness (R_{ru})	$R_{ru} = (H_{max} - H_{min}) \cdot D$ H_{max}, H_{min} = maximum, respectively minimum elevation value of basin
Slope variability (Vs)	$V_S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (s_i - s_m)^2}$

Accessing raster cells using Python is not possible in a straight forward manner, thus the algorithm necessary has been implemented in Java using the ArcObjects API. The algorithm uses the Strahler re-ordered streams as well as Strahler raster as input. As output of the algorithm a new raster with pour points is created. The resulting pour points are based on the confluence locations of the stream segments but are moved one pixel upstream for each segment of lower or same order than the order of the confluence point. The algorithm is given in pseudo code in figure 3.

3.2 Geomorphological Risk Parameter

After stream network extraction and watershed delineation, the risk parameters have to be calculated. The parameters used in this study are discussed in the following and are summarized in table 1.

Basin Size: The size of the contribut¬ing area of the rainfall in a basin has a direct effect on the total volume of runoff that drains from that basin. For a larger basin runoff starting from the most upstream area will take longer time to reach the outlet than for a smaller one with similar characteristics. In addition, a single heavy rainfall event will probably impact only a part of the large basin at any given time, but it could cover the entire small basin. According to UCAR (2010) most flash floods occur in small sub basins having an area size less than 80 km2.

Basin Shape: The runoff in a more circular basin will arrive more quickly at the basin outlet than in an elongated basin. In addition, runoff from different locations in this basin may arrive at the outlet at the same time, resulting in a greater peak flow. By contrast, in a longer, narrower basin, water from multiple locations is less likely to arrive at the same time. There are different parameters in use to describe the shape of a polygon. A common measure is the isoperimetric quotient, also called circularity parameter, the ratio of the area of the shape to the area of a circle (the most compact shape) having the same perimeter. Thus for circular watershed the value will be 1 as according to Miller (1953) circularity ratios in the range 0.4 to 0.5 indicates strongly elongated and highly permeable homogenous geologic materials. Another measurement used in many analyses is the elongation ratio.

According to Schumm (1956) is defined as the ratio of diameter of a circle of the same area as the basin to the maximum basin length. The maximum basin length can easily be calculated using the Flow Length tool. The varying shapes of watershed can be classified with the help of the index of elongation ratio, i.e. circular (0.9~1.0), oval (0.8~0.9), less elongated (0.7~0.8), elongated (0.5~0.7), and more elongated (less than 0.5).

Slope and Relief: Not only does slope affect the timing of runoff, but it also affects the amount of infiltration. The greater the slope, the lower the infiltration rate is, since gravity pulls less water into the land surface and more water across that surface. Both effects increase runoff. In general, the steeper the slope and the steeper the drainage channels, the quicker the flow response and the higher the peak discharges. Thus the relief of a basin promotes flow concentration along drainage lines, resulting in high discharges over a short period in time. Here in this study two parameters are calculated. By using the Zonal Statistics tool the average slope for each basin will be used as well as the relief ratio. For calculating the relief ratio for each basin the highest and lowest elevation are extracted and their difference is divided by the maximum flow length.

Ruggedness: Ruggedness or roughness represents the heterogeneity of the surface and thus the spatial variability of the topography. There are different parameters in use to measure the topographic ruggedness. Ruggedness may be based on standard deviation of slope, standard deviation of elevation, slope convexity, variability of plan convexity (contour curvature), or some other measure of topographic texture. Widely used is the ruggedness number introduced by Strahler (1958) that relates the drainage density with the area or the relief as the product of both. Therefore areas of high drainage density and low relief are as rugged as areas of low drainage density and high relief. According to Patton and Baker (1976) areas of potential flash flooding might be expected to have the highest ruggedness numbers incorporating a fine drainage texture, with minimal length of overland flow across steep slopes, and high stream channel gradients. As ruggedness parameters in this study the standard deviation of the slope as well as Strahler's ruggedness number has been used.

Stream Density and Frequency: Stream density is one of the most important characteristics for evaluating potential runoff. Stream density is the length of all channels within the basin divided by the area of the basin. A drainage basin with a large number of tributaries has a higher stream density than a basin with very few tributary streams. Higher stream density allows the landscape to drain more efficiently following a storm event. More efficient drainage means that water moves into streams and creeks faster, causing peak storm flows to be larger and to occur sooner. A basin with a lower stream density usually indicates a deep, well-developed soil. In this case, water is more likely to infiltrate into the soil rather than become surface runoff and enter into the channel network. Another measure of topographic texture based on the ratio of the number of stream segments per unit area of the basin is the frequency segments. Similar to the stream density, a high value indicates that areas have higher potential risk of flash floods than low values.

4. Results and Discussion

The complete workflow has been applied to the study area. Using a threshold value of 150 pixels, i.e. about 0.1 km2, all over all 7487 watersheds of first order, 1534 of second order, 320 of third order, 70 of fourth order, 20 of fifth, and 4 of sixth order have been delineated and classified according to the risk parameters. The summed up results for the watersheds of order two to six are shown in table 2. Up to order four, the size of the basins is in the critical range of small watersheds less than 100 km^2. On the other hand, the mean values of the circularity ratio as well as of the elongation ratio indicate highly elongated watersheds. The slope values show with an average of 20 % rather high values corresponding to high values for the relief ratio. The mean value exceeds even the value stated by Marchi et al., (2010) for rugged catchments in the Alpine areas with 0.19. Slaymaker (2010) states that ruggedness numbers for mountainous areas having values more than 10 have a potential of hazard risk, whereas numbers >1 shows still an intermediate potential. For the study area the highest value for order 4 is 4.9 found for a watershed of order 4. The standard deviation of the slope shows rather high values as well, with a mean of 8.0°, having the highest value for watershed of order two with 16.8°. The mean bifurcation ratio is ranging from 2.0 to 14.0, where the mean for the different orders is decreasing with the increasing order numbers of the watersheds. The mean overall value of 3.3 is rather low compared to the results obtained of other mountainous regions; see e.g. Altaf et al., (2013). Drainage density and drainage frequency are following a similar pattern (Figure 4).

Table 2: Statistics of risk parameters for sub basins of order 2 to 6

WATERSHEDS 2 TO 6 ORDER - COUNT: 1948			
Parameter	**Minimum**	**Maximum**	**Mean**
Area (Km^2)	0.253	939.170	5.337
Circularity Ratio	0.0828	0.589	0.329
Slope (deg)	2.250	34.204	19.892
Elongation Ratio	0.217	0.818	0.508
Relief Ratio	0.00690	0.548	0.209
Ruggedness	0.0312	5.740	1.257
Stand. dev. Slope (deg)	1.290	16.729	7.978
Drainage Density	0.579	4.554	2.051
Mean Bifurcation Ratio	2.000	14.000	3.310
Drainage Frequency	1.003	11.868	3.785

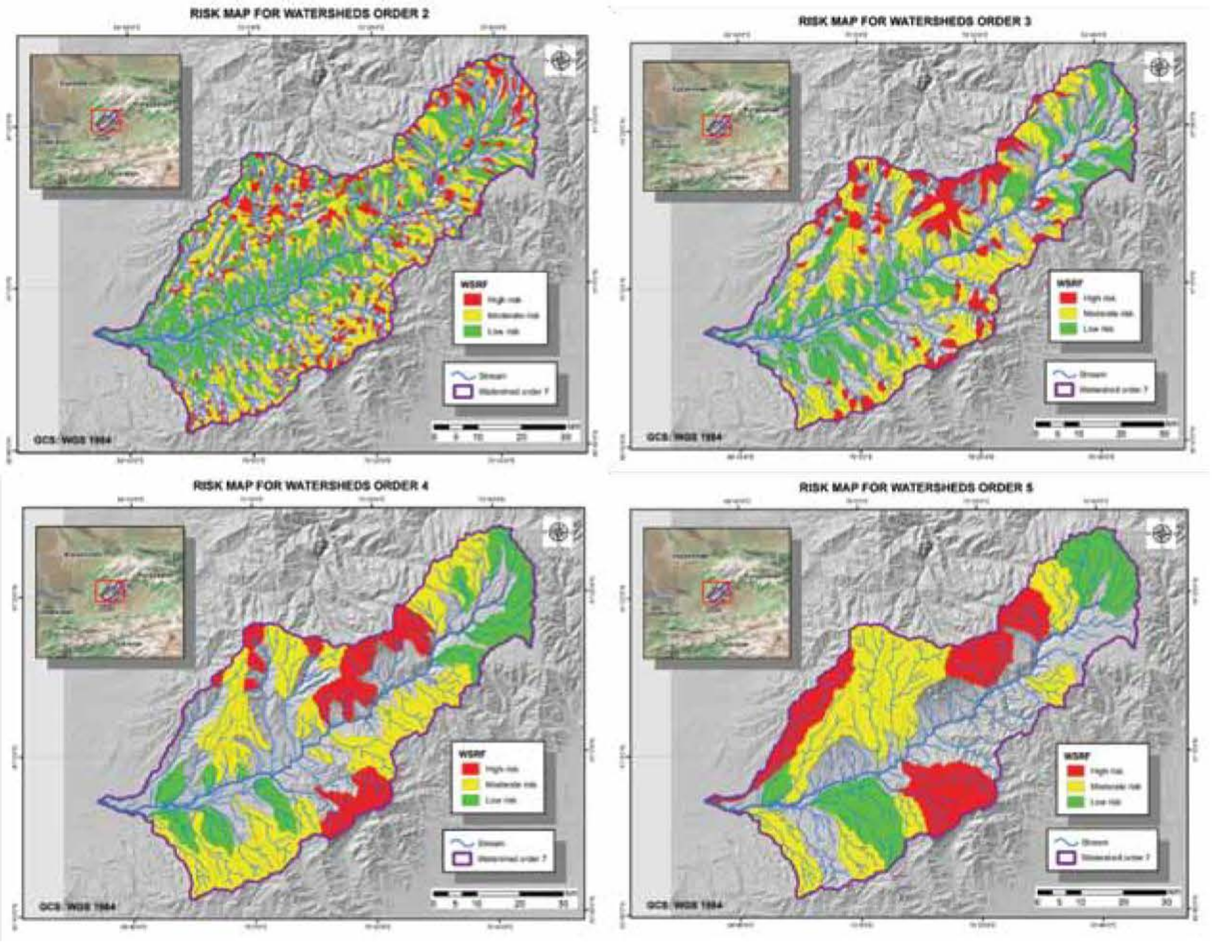

Figure 4: Watersheds of order 2 to 5 classified according to their risk potential

5. Conclusion

It has been shown, that with the automated process a larger region containing many sub basins can be classified in a rather short time according to morphometric parameters related to a flash flood risk potential. The overall processing time for the study area took about 3 hr on a state-of-the-art laptop, where most of the time was spend on the re-order of the Strahler streams and the pour point generation. Concerning the risk parameter, the model can be easily extended by additional ones, as the basic stream network and basin characteristics are available for each of the individual watersheds. It is planned for a more detailed study, to extend the model by other layers like land use to analyze additional potential non-geomorphic risk factors as well. The model will be extended also to analyze the vulnerability of the watersheds, i.e. considering settlement structure, land use, infrastructure etc.

References

Altaf, F., Meraj, G. and Ramshoo, S. A., 2013, Morphometric Analysis to Infer Hydro-logical Behaviour of Lidder Watershed, Western Himalaya, India. *Geography Journal*, Vol. 2013, Article ID 178021, 14, 2013. doi:10.1155/2013/178021.

Azab, M. A., 2009, Flood Hazard between Marsa Alam – Ras Banas, Red Sea, Egypt. *Proceedings Forth Environmental Conference, Faculty of Science, Zagazig University*, 17-35.

Djumaev, N. M., 2010, Ministry of Emergency Situations of Republic of Uzbekistan. *Asian Conference on Disaster Reduction 2010*. http://www.adrc.asia/cdr/2010kobe/documents/S2-1_04_Uzbekistan.pdf [accessed on March 18, 2015].

El-Moustafa, A. M. and Mohamed, M. M., 2013, Flash Flood Risk Assessment using Morphological Parameters in Sinai Peninsula. *Open Journal of Modern Hydrology*, Vol.3 No.3,http://dx.doi.org/10.4236/ojmh.2013.33016

Esper Angillieri, M. Y., 2012, Morphometric Characterization of the Carrizal Basin Applied to the Evaluation of Flash Floods Hazard, San Juan, Argentina. *Quaternary International,* 253, 74-79.

IPCC, 2012, Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption. *Special Report of the Intergovernmental Panel on Climate Change*. http://www.ipcc-wg2.gov/SREX/images/uploads/SREX-All_FINAL.pdf [accessed on March 18, 2015].

Marchi, L., Borga, M., Preciso, E. and Gaume, E., 2010, Characterisation of Selected Extreme Flash Floods in Europe and Implications for Flood Risk Management. *Journal of Hydrology,* 394, 118–133.

METI and NASA, 2011, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) Version 2. *Earth Remote Sensing Data Analysis Center ASTER GDS User Service* http://www.gds.aster.ersdac.or.jp/gds_www2002/service_e/inq.c_e/set_inq.c_e.html [accessed January 5, 2015].

Miller, V. C., 1953, A Quantitative Geomorphic Study of Drainage Basin Characteristics in the Clinch Mountain area Virginia and Tennessee. *Tech. Rept. 3*, Dept. of Geol., Columbia Univ., New York.

Munich RE, 2014, Annual Statistics. https://www.munichre.com/touch/naturalhazards/en/natcatservice/annual-statistics [accessed 17/02/2015].

Omran, A., Schröder, D., Rayes, A. and Geriesh, M, 2011, Flood Hazard Assessment in Wadi Dahab – South Eastern Sinai –Egypt Based on Basin Morphometry using GIS Technique. In *Geospatial Crossroads @ GI_Forum '11*, edited by Car, Griesebner, Strobl, Wichmann Verlag.

Omran, A., 2013, Application of GIS and Remote Sensing for Water Resource Management in Arid Area – Wadi Dahab Basin – South Eastern Sinai – Egypt (Case Study), Doktoral Dissertation University Tübingen, IDN.1044121289. http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-71102.

Patton, P. and Baker, V. R., 1976, Morphometry and Floods in Small Drainage Basins Subject to Diverse Hydrogeomorphic Controls. *Water Resources Research*, Vol. 12, No. 5.

Pareta, K. and Pareta, U., 2011, Quantitative Morphometric Analysis of a Watershed of Yamuna Basin, India using ASTER (DEM) Data and GIS. *International Journal of Geomatics and Geosciences*, Vol. 2, No 1.

Schumm, S. A., 1956, Evolution of Drainage Systems and Slopes in Badlands at Perth Amboy. N. J., Bull. *Geol. Soc. Am.*, 67, 597-646.

Slaymaker, O., 2010, Mountain Hazards. In *Geomorphological Hazards and Disaster Prevention*, edited by Alcántara-Ayala, I., Goudie, A.S., Cambridge University Press.

Strahler, A. N., 1958, Dimensional Analysis Applied to fluvially Eroded Landforms. *Geological Society of America Bulletin*, 69, 279-300.

UCAR, 2010, Flash Flood Early Warning System Reference Guide, *University Corporation for Atmospheric Research*, http://www.meted.ucar.edu/hazwarnsys/haz_fflood.php [accessed 15/2/2015]