

PRECIPITATION USED AS KEY FACTOR IN TAPERED LINE BASED RIVER REPRESENTATION

Zsolt MAGYARI-SÁSKA¹ 

DOI : 10.21163/GT_2020.152.20

ABSTRACT:

GIS and cartography represent two different study areas but both of them share the map concept. Nowadays, most GIS software incorporates advanced representation and labeling possibilities, and thanks to the database, which is an essential component of any geoinformation system, there are multiple symbolization possibilities, using the stored data. Taking into account this symbiosis between the two components of modern GIS: analysis and advanced representation this study wants push further the cartographic possibilities of them taking advantage of their analysis and modelling part. Through this study we wanted to develop and test a method for river symbolization in which the river's line width to be proportional with the average precipitation amount at every part of the river. The idea of this study came by working with hand drawn historical maps in which this type of representation is common and very suggestive. Although there are some interesting approaches which want to symbolize river courses based on data, this type of approach presented in this study has not been found. The result obtained from a situation where inaccuracies are inevitable due to very different data sources: a recent DEM with 25m resolution and a map from the beginning of the 20th century with scale of 1: 300000, has shown that this method based on modeling, analysis and GIS operations can be used successfully to have a more accurate, more real and eye catching symbolization of watercourses.

Key-words: river symbolization; realistic representation; historical map; python scripting; QGIS

1. INTRODUCTION

Cartographic representations have several important roles. Maps as valuable information holding and presenting tools are used not just in orientation but also for illustrating content, phenomena, processes in an eye captivating manner (Cosgrove, 2005). This latter use of maps sometimes relies on exaggerating the reality, sometime simplifying it or in many cases just trying to reflect the reality as much as possible (Thorwer, 2007; Petersen, 2014). When it's about reality GIS can provide useful backend data analysis methods to precisely calculate data and based on these results to achieve a more suggestive cartographic representation (Hardy et al., 2005; Mei and Li, 2008; Brewer, 2015).

Our research intends to make a contribution in river mapping. The most common method for representing watercourses is the use of lines, which width suggests the importance and size of the river mainly based on its water flow quantity or in some cases based on river length. In such cases the width is maintained constant through the line length especially if it's created using GIS. In some vector illustration software may exists possibilities to gradually increase the line width but these approaches can use just the length of the line as base value.

If we look at historical maps in many cases, we can observe that the rivers width is gradually growing as it flows from their spring to inflow (**Fig. 1**). As in many other fields such in architecture or furniture design also in cartography meticulous work in the creation process was not intended just to reveal the beauty but to create a practical, a usable thing. A few years ago we started to analyze how we could digitally recreate in an automatized way the hachure lines used in relief representation (Magyari-Sáska, 2017). With this research we want to continue and to investigate the automatic recreation possibilities for historical and not just historical maps, presenting a GIS based method for river flows representation based on mean annual rainfall data.

¹ *Babes-Bolyai University Cluj, Faculty of Geography, Gheorgheni Extension, 535500 Gheorgheni, Romania, zsmagyari@gmail.com.*

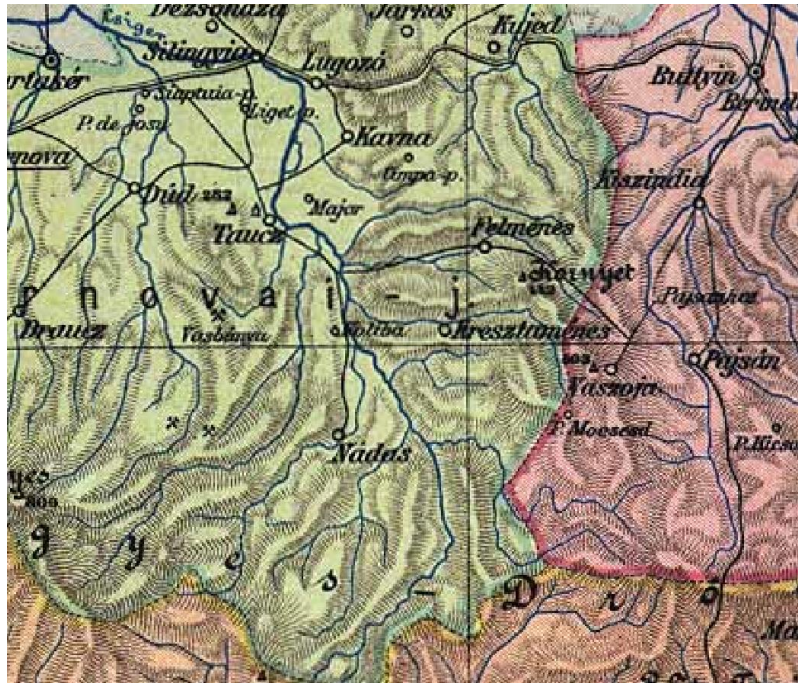


Fig. 1. Gently thickening river representation in a historical map

Although there are studies related about the processing (Piovan, 2019; Timár and Biszak, 2010; Rumsey and Punt, 2004) and automatic vectorization of old maps (Iosifescu et al., 2016), there is no such research or method regarding symbolization, although the importance of suggestive representation is shown up in several studies (Punt et al., 2006; Yamg and Lin, 2008).

2. METHODS AND RESULTS

The objective of the research was to develop a method for creating vector layers representing rivers, which are suitable for the desired symbolization. These layers should have an attribute value based on which the gradually growing line representation can be made. The symbolization should respect the reality considering that the line width should grow in correspondence with that rainfall which contributes to the river water quantity at every part of the river.

The existing vector data-model doesn't provide many possibilities to achieve the desired representation as a line feature could have a single (even if it's calculated from several fields) value based on which its width is determined by scaling it to different intervals. The existing symbolization methods can be grouped as below:

- methods in which a single symbology (including width) is used for a line depending on line attributes or their combinations. This is the most widely used method, with drawback, that the width is constant through all course (**Fig. 2a**).
- methods in which the line forming segments are identified in the symbolization process and treated separately in generating the width (**Fig. 2b**). This method is presented by A. Graser in 2017 in an on-line article *Better river styles with tapered lines* and further developed by M.R. Lombardo in 2019 in another on-line article: *Improved tapered rivers in QGIS*.

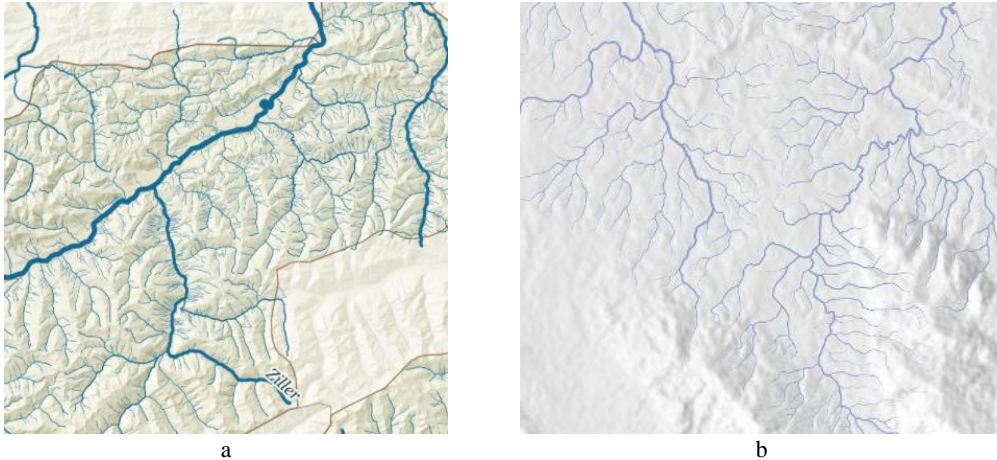


Fig. 2. Existing river symbolization ways: a) with constant width (source: Graser, A. – Tirolo riverset) b) with tapered lines (source: Lombardo, M.R. – Cuencas de los ríos Sansón y Sansoncito)

This second method category catches the main idea but the actually varying factors that controls the line width are: the current segment number divided to the total number of segments which forms the line, and values (hard coded or attribute based) which are constant for a line. This means that the middle line segment will have approximately the average width of the first and last line segments' width. Moreover, this approach being a software dependent possibility, doesn't represent a general solution in symbolization process. With all this, the method (if present in given GIS.) has the advantage of not exploding the line to segments, from which we can benefit at feature labeling.

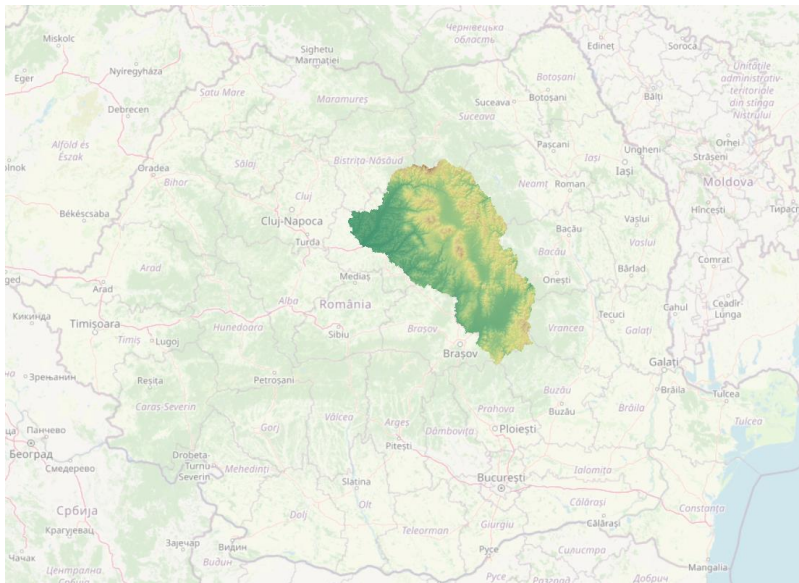


Fig. 3. Test area in Romania's central region

Our developed method wants to assign different values for different line segments proportional to the water quantity that could be present at a given line segment. As this is practically impossible we had to split every line feature representing a river, obtaining in fact as much features as it's the number of segments that forms the original line feature. The attached values indicating the water quantity can be modelled at various complexity levels. As our research is more exploratory then applicative for a given region, we don't want to focus on achieving the best possible modelling of rivers runoff values we did not take into account the pedagogical characteristics of soil nor the

vegetation's impact or evapotranspiration which may contribute to the runoff value (Crăciun et al., 2009).

As we will present below the study has the following main parts: achieving a fair spatial interpolation of average annual rainfall, calculating and attaching a rainfall based attribute to each exploded line segment to use it in symbolization. Even if the current study is methodological, we selected a test area, which was the central region of Romania covering part of Harghita and Covasna and Mureş counties as presented in **figure 3**.

2.1. Rainfall interpolation

Starting from the simplest IDW to complex geostatistical methods such as Kriging, there are various techniques for rainfall interpolation which are in use and compared in many studies (Hartkamp et al., 1999; Naoum and Tsanis, 2004; Taesombat and Sriwongsitanon, 2009; Mair and Fares, 2011; Noori et al., 2014). The efficiency of these methods depends on many factors which has the origin mainly in the number and spatial distribution of measurement sites (Di Piazza et al., 2011; Chen et al., 2017;). Because in our case there are just a few meteorological stations in the study region, we considered more appropriate to take all measurement data existing at different meteorological stations in Transylvania region and apply a less common but suitable method based on multiple regression (Efroymson, 1960; Hastie and Pregibon, 1992). Due to the fact that we don't want to interpolate the rainfall value obtained in a short time period (hours or days) to see the spatial distribution of values, but to characterize the annual average precipitation distribution over a multi-decade period, we have considered that identifying factors that influence the precipitation could be a fair try.

Accordingly, we used the backward stepwise multiple linear regression in R statistical software package (Venables and Ripley, 2002). The method takes all linear combination of the specified factors, named predictors and based on various statistical criteria (in case of R the Akaike Information Criterion is used) identifies that combination which has the highest correlation value to predict the given result dataset (Chatterjee and Hadi, 2012).

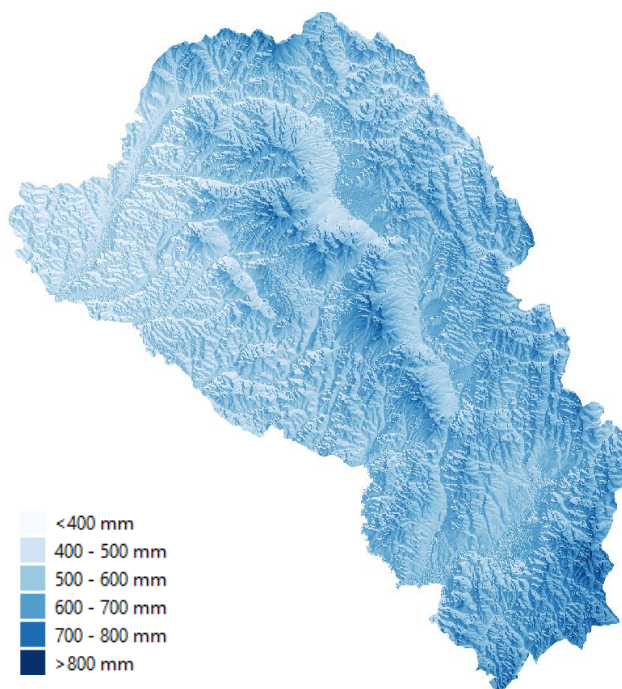


Fig. 4. Annual average rainfall map interpolated using backward stepwise multiple regression

As candidate factors, we used the height (source: EU-DEM v.1.1, resolution: 25m) which is considered an important factor (Modallaldoust et al., 2008; Huang and Hu, 2009), the coordinates, and the aspect. From the mentioned 4 predictors only the x coordinate was considered by the model not influential for the given region and for the annual average precipitation value. Based on the obtained predictors we made the interpolation and got the rainfall distribution map (Fig. 4).

2.2. Calculating accumulated precipitation values

In the beginning we performed the necessary hydrological operations such as fill sinks to can create a correct flow accumulation image. In creating the flow accumulation, the interpolated precipitation values were considered as weighting factor. Analyzing the accumulation values the channel network can be identified, which has an important role in watershed basins generation. The presented process was realized in SAGA (Fig. 5 – blue part).

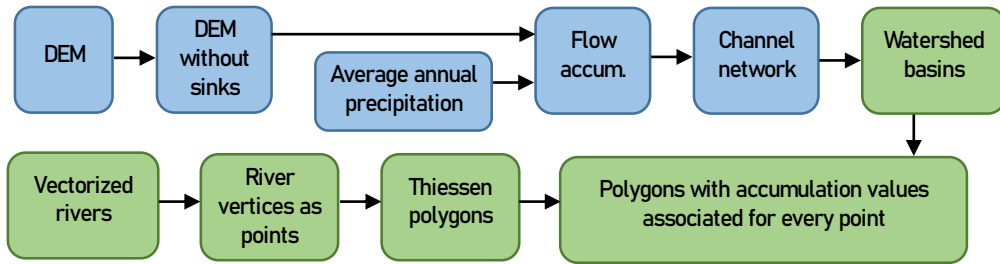


Fig. 5. Determining flow accumulation for every vertices representing river line segments endpoints (raster layers with blue, vector layers with green background)

Having the accumulated precipitation values it seems to be very easy to attach these values to the vertices defining the rivers. Due to the fact that the DEM on which the hydrological analysis was done and the historical map from which the rivers were vectorized could not have proper overlay the vertices in many cases were positioned on such cells which had very improper accumulated precipitation value. We had to find a way to attach an appropriate cell value even if this are relatively far from the vertex but its value surely is most correct than those below the vertex.

We needed a spatial splitting method. For this we created Thiessen polygons based on river vectorization vertices. In such way we've got an own region for every river vertex which of course could pass over several catchments. To overcome this unwanted situation, we made an intersection between the created Thiessen polygons and the watershed basins layer. (Fig. 5 – green part) From the obtained polygons we always took the highest accumulated precipitation value, considering that in fact this value should have been under the vertices in case of spatially precisely corresponding data layers. Even if for our test case the dislocation of map elements maybe it's more pronounced than in other situations, in most cases this kind of inaccuracies appears and the presented method could represent a solution for it.

2.3. Attaching accumulated precipitation values to line segments

After we got the accumulated precipitation values in polygons representing intersection between catchment areas and the Thiessen polygons defined by the endpoints of river segments, we had added their values to the proper line segments. For this we the mean precipitation values present at the end points of the line segments were used. Because this process could not be performed with existing QGIS operations we had to develop a Processing script in Python which has two input: a line layer with vectorized rivers and a polygon layer with the accumulated precipitation values for the catchment area slices.

The accumulated precipitation values along every vectorized river, this should form a monotonically increasing data series. Unfortunately, this isn't true due to the following situations:

- the flow may have incorrect vectorized direction from inflow to spring resulting a theoretical monotonically decreasing series. Although this may be considered a vectorization error it could frequently appear.
- the vectorized river path may cross through different catchments (**Fig. 6**) for different reasons (scale, resolution or precision difference between the digital elevation model and original map sheet)

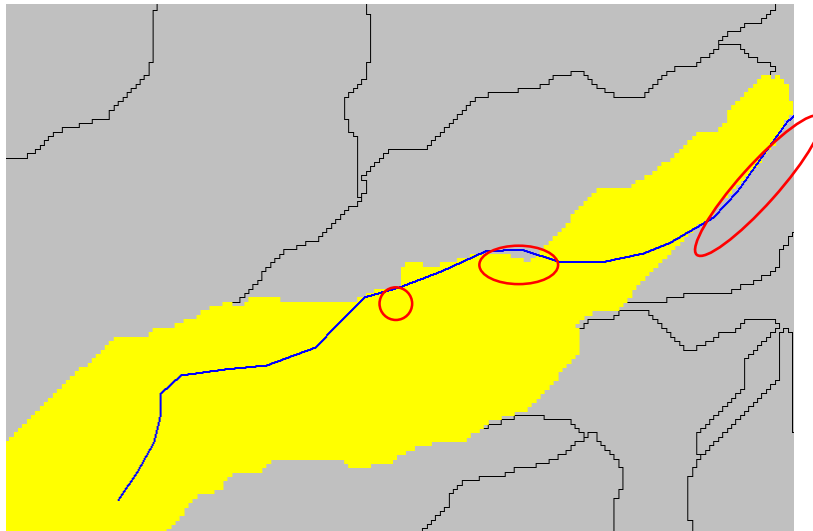


Fig. 6. Vectorized rivers touches multiple catchments resulting improper accumulation data

While the first situation obviously should be corrected by detecting and inverting the direction the second situation raises the question of should the streamflow path corrected or not? The answer of course depends on the effective situation. In our case the concept was to create a digital version of a historical map with the highest representation fidelity. In this case even if the original map had a weak precision, we wanted to keep its look, maintaining the internal positional proportions between the cartographic elements. Considering this, for both problematic situations we developed one method at each to overcome them as follows.

Incorrect streamline direction

The direction of a vectorized line is stored inside the associated data structure and can visualized by using arrow symbol in visual representation. Even if most of the time the line direction is not important in several cases it could be crucial. Such applications are network routing with one-way connections or river flow modelling. In our case we should have a method to detect whether the vectorization direction was correct or not. This could be done using gradient analysis along the line but as we already had the accumulated precipitation values, we decided to make use of it.

The detection method is simple based on signum function with the assumption that the number of vertices positioned in proper catchments are higher than for those in incorrect catchments. In this case by counting and comparing the number of cases at which the accumulated precipitation values doesn't follow an increasing trend to total number of vertices we succeeded to identify with high probability the lines vectorized on incorrect direction. The ratio was calculated for every river using equation 1

$$\frac{\sum_{i=1}^{n-1} \text{sgn}(v_{i+1} - v_i)}{n-1} \quad (1)$$

where n indicates the number of vertices that forms the line and v_i is the accumulated precipitation value at vertex i . A line was considered having an incorrect direction is the value of the ratio in equation 1 was higher than 0.6.

Anomalies in precipitation accumulation data series

As mentioned before due to spatial dislocation between DEM and original map sheets, which are caused especially by scale/resolution differences, in many cases a streamflow touches or crosses other catchment areas than its own. In these cases, the accumulate precipitation values are inappropriate for the current flow and can significantly differ from the real values. **Figure 7** shows an example for such cases. There are observable that many vertices which are positioned in other catchments. For this example, in the majority of cases the precipitation values of wrong catchments are much lower, undershooting the values for the normal course of the river and in a few cases the vertices are positioned in catchment with higher precipitation value, overshooting the values. As these misplaced values has negative influence on the desired symbolization we should identify and treat them properly.

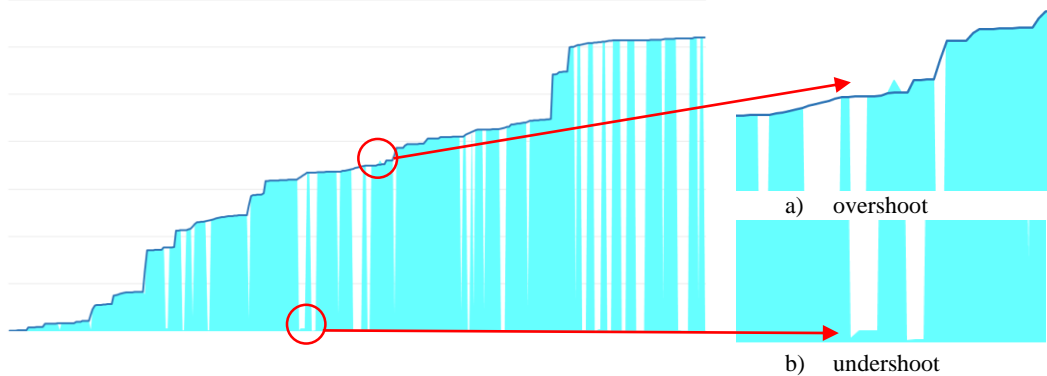


Fig. 7. Incorrect data series (with cyan) resulted from accumulated precipitation value due to vertices in incorrect catchment areas. Correct flow values are presented with the top blue line.

As these values can't be considered time series or series with a specific statistical distribution the well-known anomaly detection or outlier detection algorithms could not be applied successfully.

Taking in consideration the real-life imposed monotonicity that should be present for every stream we developed a rank-based algorithm where two type of ranks are associated to every vertex one based on their appearance order and the other based on accumulated precipitation value order. When the values of two rank type are identical, we got a monotonic increasing series and the goal was reached. The main steps of the algorithm are presented in **figure 8**, which was implemented in PyQGIS, which is a valuable scripting language to extend the processing capabilities of QGIS (Sherman, 2018).

```

set initial position-based ranks
set initial value-based ranks
while pairwise ranks are different
    identify and exclude from dataset the rightmost data with the highest rank difference
    set new position-based ranks
    set new value-based ranks
for all excluded data
    set their new values using linear trend based on not excluded head and end data

```

Fig. 8. Processing algorithm for detecting and correcting anomalous data in accumulation value series

2.4. Final result

Applying the PyQGIS processing script we got a vector layer in which every line feature corresponds to an original line segment of the vectorized rivers. As an attribute of the layer, for every feature the corrected accumulated precipitation value is present.

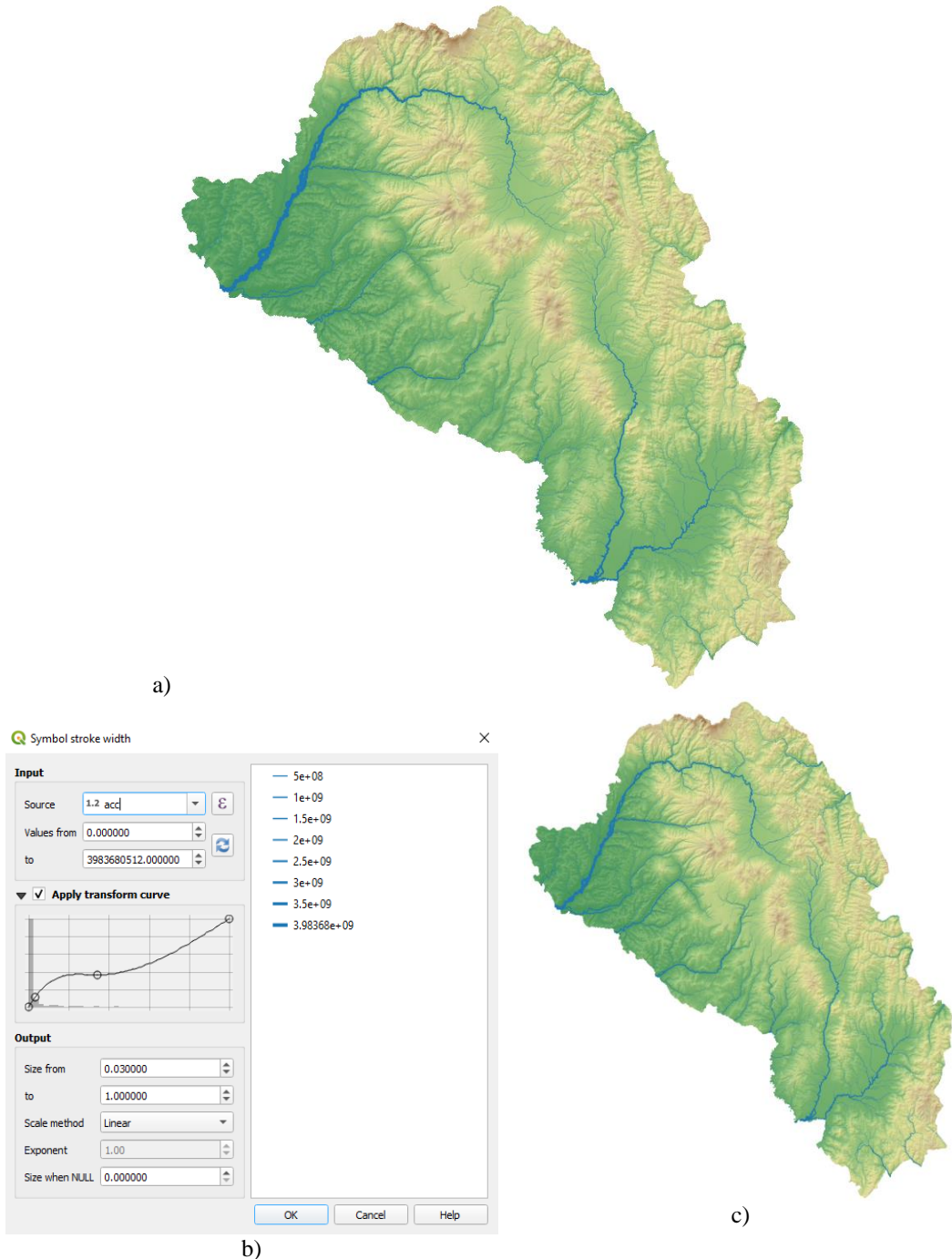


Fig. 9. Resulted river representation. Growing width is modelled based on precipitation in catchment areas.
 a) line width after linear scaling b) manually definable scaling curve in QGIS c) line width using manually adjusted scaling curve emphasizing smaller creeks and brooks

These values can be used in line symbolization. From now on the visual effect depends on the symbolization possibilities of the used GIS. The values can be scaled in a linear way, determining the representation width, or other functions (logarithmic, exponential) can be applied on data to emphasize higher or lower values before rendering them (**fig. 9**). QGIS also has an advanced method, called Assistant, in which the transformation curve between the existing data and the selected visualization width range can be manually adjusted.

Whatever the possibilities are in a given GIS the segments width now is proportional with the presumed water quantity that can be assumed for that river portion.

3. CONCLUSIONS

The aim of this study was to develop a methodology for representing rivers according to their water quantity. To test our methodology, we considered solely the interpolated precipitation values as contributing factor to line width, but the methodology permits to consider any other more sophisticated method to estimate the flow value. As only individual line features can have their width determined by attribute table values, we had to split every line to its line segments resulting multiple features for every river. This is the most disadvantageous element of our study because it affects the labelling possibilities of original line features. The most challenging part of the study was how to maintain data consistency of the accumulated values in case of rivers because the spatial position of vectorized river network in some cases doesn't correspond to catchment areas derived from DEM.

We think that the developed methodology presents a possibility not just to create more beautiful visual representations but also more realistic ones. Even if the real width of a river is not determined just by its water quantity in cartographic representations this is the key factor indicating the magnitude of the river.

The developed methodology, as shown in figure 5 not just gradually increases the river width according to the water amount, but inflows from other water sources are clearly observable as steps in the data series. Another benefic effect of this approach is that the line width is maintained even if the river splits around small islands which is usually vectorized as separate line features with different lengths.

Beside the fact that using our method the rivers width is not rendered based on their length or category as in most of the current representations, there are some other advantages. It does not matter the vectorization direction, because in case of wrong direction it is detected and changed accordingly. It does not matter if the lines are not merged together or if the river network topology is incorrect as these elements are not considered.

In the final the methodology can be used even if we want to illustrate just a smaller portion of the river network from a larger area, as the data used on which the representation is made relies on accumulated precipitation values and if we have this data, the accumulation values are propagated automatically to the interested locations where the vectorized rivers are present.

ACKNOWLEDGMENTS

The presented research was supported by the DOMUS scholarship program of the Hungarian Academy of Sciences.

REFERENCES

- Brewer, C. A. (2015). *Designing Better Maps: A Guide for GIS Users*, 2nd ed. Esri Press, Redlands, CA.
- Chatterjee, S., Hadi, A. (2012), *Regression Analysis by Example*. 5th ed. N.p.: John Wiley & Sons
- Chen, T., Ren, L., Yuan, F., Yang, X., Jiang, S., Tang, T., Liu, Y., Zhao, C., Zhang, L. (2017) Comparison of Spatial Interpolation Schemes for Rainfall Data and Application in Hydrological Modeling, *Water*, 9(5), 342, DOI: <https://doi.org/10.3390/w9050342>
- Cosgrove, D. (2005) Maps, Mapping, Modernity: Art and Cartography in the Twentieth Century, *Imago Mundi*, 57(1), 35-54, DOI: 10.1080/0308569042000289824
- Crăciun, I., Haidu, I., Magyari-Sáska, Zs., Imbroane, A.I. (2009), Estimation of Runoff Coefficient According to Soil Moisture Using GIS Techniques, *Geographia Technica*, 4(2), 1-10
- Di Piazza, A., Lo Conti, F., Noto, L.V., Viola, F., La Loggia, G. (2011), Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy, *International Journal of Applied Earth Observation and Geoinformation*, 13(3), 396-408, <https://doi.org/10.1016/j.jag.2011.01.005>
- Efroymson, M.A. (1960) Multiple regression analysis. In Ralston, A. and Wilf, HS, editors, *Mathematical Methods for Digital Computers*. Wiley.
- Hardy, P., Eicher, C., Briat, M., Kressmann, T. (2005) Database-stored Representations and Overrides, Supporting Automated Cartography with Human Creativity, *Auto-Carto 2005 Proceedings*, Las Vegas
- Hartkamp, D., De Beurs, K., Stein, A., White, J.W. (1999), *Interpolation Techniques for Climate Variables*, International Maize and Wheat Improvement Center, Mexico, 26
- Hastie, T.J., Pregibon, D. (1992) *Generalized linear models*. Chapter 6 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.
- Huang B., Hu T. (2009) Spatial Interpolation of Rainfall Based on DEM. In: *Advances in Water Resources and Hydraulic Engineering*. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-540-89465-0_15
- Iosifescu, I., Tsorlini, A., Hurni, L. (2016), Towards a comprehensive methodology for automatic vectorization of raster historical maps, *e-Perimetron*, 11(2), 57-76
- Magyar-Sáska, Zs. (2017) Automatic generation of hachure lines, *Geographia Technica* 12(1), 78-81, DOI: 10.21163/GT_2017.121.08
- Mair A., Fares A. (2011), Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island, *Journal of Hydrologic Engineering* 16(4), 371-383, DOI: 10.1061/(ASCE)HE.1943-5584.0000330
- Mei Y., Li, L. (2008), A Cartographic data model for Better Geographical Visualization based on Knowledge, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B2. 611-616
- Modallaldoust, S., Bayat, F., Soltani, B., Soleimani, K., (2008), Applying Digital Elevation Model to Interpolate Precipitation. *Journal of Applied Sciences*, 8: 1471-1478, DOI: 10.3923/jas.2008.1471.1478
- Naoum, S., Tsanis, I.K. (2004), Ranking Spatial Interpolation Techniques Using a GIS-Based DSS, *Global Nest Journal*. 6(1), 1-20, DOI: <https://doi.org/10.30955/gnj.000224>
- Noori, M.J., Hassan, H.H., Mustafa, Y.T. (2014), Spatial Estimation of Rainfall Distribution and Its Classification in Duhok Governorate Using GIS, *Journal of Water Resource and Protection*, 6, 75-82, DOI: 10.4236/jwarp.2014.62012
- Petersen, G. (2014). *GIS Cartography: A Guide to Effective Map Design* (2nd edition) CRC Press, Boca Raton, FL.
- Piovan, S. (2019). Historical Maps in GIS. *The Geographic Information Science & Technology Body of Knowledge* (1st Quarter 2019 Edition), John P. Wilson (Ed.).
- Rumsey, D., Punt, E. M. (2004) *Cartographica Extraordinaire: The Historical Map Transformed*. Redlands, CA: ESRI Press.
- Sherman, G. (2018), *PyQGIS Programmer's Guide*, Locate Press, 252
- Taesombat, W., Sriwongsitanon, N. (2009), Areal rainfall estimation using spatial interpolation Techniques, *ScienceAsia*, 35, 268-275, DOI: 10.2306/scienceasia1513-1874.2009.35.268
- Thrower, N. J. W. (2007). *Maps and Civilization: Cartography in Culture and Society* (3rd edition), University of Chicago Press, Chicago.
- Tímár, G., Biszak, S. (2010), Digitizing and georeferencing of the historical cadastral maps (1856-60) of Hungary, *5th International Workshop on Digital Approaches in Cartographic Heritage*, Vienna, Austria, 22–24 February 2010
- Venables, W.N., Ripley, B.D. (2002) *Modern Applied Statistics with S*. New York: Springer (4th ed).