

COMPARISON OF TWO APPROACHES FOR AN ESTIMATION OF THE MEAN ANNUAL FLOOD AT UNGAUGED SITES IN SLOVAKIA

¹ Zuzana NÉMETOVÁ*, ² Silvia KOHNOVÁ, ³ Romana MARKOVÁ

^{1,2,3} Department of Land and Water Resources Management, Faculty of Civil Engineering
Slovak University of Technology, Radlinského 11, 810 05 Bratislava, Slovak Republic
e-mail: ¹zuzana.nemetova@stuba.sk, ²silvia.kohnova@stuba.sk, ³romana.markova@stuba.sk

Received 25 April 2019; accepted 13 September 2019

Abstract: Regional flood frequency analysis is considered to be an important and popular method for estimating different hydrological variables at ungauged sites. The estimation of the index flood is the essential problem when this method is applied. The objective of the study is a comparison of the estimation of the mean annual flood (or index flood) by using two approaches based on the ‘so-called’ index flood method and top-kriging. The concept behind these methods permits estimating the mean annual flood at ungauged locations using information taken from gauged sites located within the same homogeneous pooling groups. The study area comprises 104 gauging stations on the whole territory of Slovakia. The observation period of the annual maximum discharges of the selected stations was from 1961-2010. The identification of the homogeneous pooling group was performed using a non-hierarchical k-means clustering algorithm. The optimal number of clusters is determined by the Silhouette method. As a result, eight homogeneous pooling group clusters were identified. Finally, the results of the estimated mean annual floods using the index flood method and top-kriging were compared with the observed data. Top-kriging provided better results than the classical index flood method for estimating the mean annual flood at ungauged sites.

Keywords: Mean annual flood, Cluster analysis, Regional flood frequency analysis, Top-kriging

* Corresponding Author

1. Introduction

Due to the substantial economic and environmental impacts of recent floods, flood frequency analyses and estimations of design discharges continue to attract a great deal of interest in contemporary engineering hydrology. The recent extreme floods in Slovakia and Central Europe have raised concerns about the reliability of flood frequency estimates, especially in ungauged basins, and also increased interest in regional flood frequency analysis [1].

The estimation of hydrological variables for ungauged basins is considered an even more important factor than the characterization of the stream-flow regime for gauged basins [2]. The prediction of hydrological characteristics for ungauged basins represents one of the most significant problems in hydrology and related sciences [3], [4], and is one of the fundamental tasks of engineering hydrology. The growing number of stations with longer records makes it possible to test how some of the new concepts of regional homogeneity and regional flood frequency analysis reported in recent literature, e.g. [5], [6], [7], and [4], perform in the estimation of design discharges in small and medium-sized catchments. Some of these methods were developed under specific conditions, and modifying them may be necessary under the rather heterogeneous geological and geomorphological conditions of Slovakia. One of the most important and popular methods is represented by index flood methods [8], [9]. According to [10], the index flood method is considered to be an efficient tool for pooling summary statistics from a data sample to evaluate the parameters of frequency distributions. A significant element of the index flood method depends on its application to a specific site condition and the information available [11]. The assumption of the index flood method is that sites are grouped into homogenous regions [12], [13].

There have been many methodologies developed for the identification of homogeneous regions [14], [15], [16]. The methods have been consecutively tested for their practical applicability and compared with the traditional approaches [17], [18], [19]. A number of new methods for the regional estimation of design discharges have been incorporated into the guidelines for engineering practice in many countries, i.e. Flood Estimation Handbook [20] in Great Britain, German Association for Water Management [21] in Germany, or the Australian Rainfall Runoff [22] in Australia.

The study deals with an estimation of the mean annual flood using two approaches, i.e. the index-flood method and top-kriging. The main concept behind these methods is an estimation of the mean annual flood at ungauged locations using information taken from gauged sites located within the same homogeneous pooling groups. The first method applied represents the index flood method, which was developed for regional flood frequency analysis by Dalrymple [23] and later improved by Hosking and Wallis [4].

The second approach applied is the top-kriging method for the prediction of hydrological characteristics. The method is predominantly used in geo-statistical analyses and has been developed by [24].

Finally, the results provided will help evaluate, which of these methods would provide better results in comparison to the observed values and could be recommended for the engineering practice for estimating design floods.

2. Materials and methods

2.1. Regional flood frequency analysis

There are a number of approaches used to regionalize information about floods. One of the techniques often applied in hydrology is represented by the index flood method, which was originally developed for flood frequency analysis [23], [4]. For the proper use of the index flood method, several crucial steps must be taken. First, homogeneous pooling groups must be identified; then their internal homogeneity must be tested. One of the most commonly used methods applied for the identification of pooling groups, i.e. the cluster analysis, is a numerical method for grouping a set of data objects into groups (clusters) by taking into account the physical and geographical characteristics of river basins that have a significant impact on the spatial variability of hydrological properties.

The clustering method should produce high quality clusters with a high intra-class similarity and a low inter-class similarity. The clustering is a widely used tool and is based on grouping objects where the objects in one group (cluster) are similar, and the objects from different groups are dissimilar [25]. The quality of the clustering results depends on both the similarity measured by the method and its implementation.

The cluster analysis is divided into several types; in this paper one of the non-hierarchical clustering methods, i.e. the k-means method, is used.

The k-means method is one of the simplest and most used clustering algorithms. The k-means clustering algorithm uses iterative refinement to determine the final results; the algorithm inputs are the number of clusters k and the data set. The method was developed by [26], and its main concept is to define k centroids, i.e. one for each cluster. Then the new centroid can be calculated according to the following equation:

$$\frac{1}{|S_i|} \sum_{x_k \in S_i} x_k = C_i, \quad (1)$$

where C_i is a centroid of the i -th cluster; S_i is an i -th cluster; x_k is the k -point belonging to the class of points, which are centroids within S_i .

The procedure is repeated until the centroids no longer change. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The optimal number of clusters was determined using the Silhouette method. The values of the Silhouette were determined using the following equation:

$$\frac{(b_i - a_i)}{\max\{a_i, b_i\}} = S_i, \quad (2)$$

where S_i is the value of the Silhouette; a_i is the mean distance of the i -th point from the other points in the cluster; b_i is the average distance from the i -th point to a point in another cluster.

2.2. Estimation of the index flood

A stepwise multiple regressions were used to determine the relationship between the climatic and physiographic basin characteristics and the index flood values in the pooling group. In order to minimize the effect of multicollinearity, attention was paid to the choice of predictors with a low mutual dependence. Several subjectively chosen and hydrologically reasonable starting combinations of the independent variables were used as seeds in the discrimination process. Because of large number of catchments (i.e. 104 catchments) and computational reasons related the number of predictors was restricted to between 2 and 4. This restriction seemed to be the best possible alternative for precise data processing and providing the most satisfactory results.

2.3. Top-kriging method

The top-kriging method was developed on the basis of the classical kriging method. The top-kriging method takes into account the topology of the stream networks and the nested catchments, which is contrary to the classical kriging method.

The kriging method is one of the interpolation methods, where an unknown value z^* of the variable at position x_0 can be calculated as the weighted average of the variable measured in the neighborhood:

$$\sum_{i=1}^n w_i z(x_i) = z^*(x_0), \quad (3)$$

where n is the number of the neighboring measurements used for the interpolation; $z(x_i)$ is the observation selected from the surrounding area x_0 ; $z^*(x_0)$ is the estimation of the variable at point x_0 ; w_i is the weights determined using a selected semivariogram model.

The measurements are not point values, and they are defined over a non-zero catchment area. In geostatistical terminology, A means only the support, and the point variable $z(x)$ represents the average over an area A according to the equation [24]:

$$\frac{1}{A} \int_A w(x) z(x) dx = \bar{z}(A), \quad (4)$$

where $\bar{z}(A)$ is the spatially averaged variable; $w(x)$ is the weighting function; A is the area of the catchment.

3. Study site and input data

The input data consists of 104 (in abstract 104, please, correct them) selected catchments located in the Slovak Republic (*Fig. 1*). The maximum annual peak flows were available for each water gauge station from 1961-2010. An estimation of the mean annual flood was conducted for the two periods selected, i.e. 1961-2010 and 1985-2010. The catchment areas range from 8.4 km² (Vydrná hydrological station, Petrinovec River) to 3821 km² (Banská Bystrica hydrological station, Hron River). The minimum elevation (106.01 m a.s.l.) was at the

Horovce hydrological station, Ondava River, and the maximum elevation (2606.4 m a.s.l.) was determined for the Poprad-Matejovce basin station together with the Poprad River. The maximum and minimum elevations were calculated using ArcGIS and based on a digital elevation model with a spatial resolution of 25 x 25 meters.

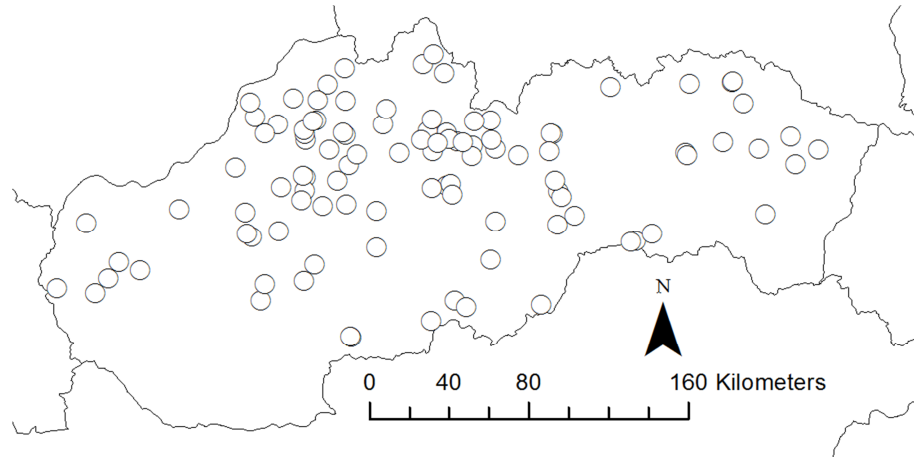


Fig. 1. Location of the selected hydrometeorological stations in Slovakia

4. Results

The estimation of the mean annual flood was performed using two pooling scheme approaches, i.e. the index-flood and the top-kriging method. In the first step of the index-flood method, it was necessary to pool the catchments into homogenous pooling groups. The optimal number of homogeneous pooling groups was determined using the k-means clustering method and the following input characteristics: average slope of the catchment ($^{\circ}$), the long-term daily rainfall (mm/day), the seasonal concentration index (representing the flood variability index), and the forested area (representing the percentage of the total catchment area). The optimal number of pooling groups was estimated using the Silhouette method. Eight pooling groups were finally selected as the best result. *Fig. 2* shows the location of the catchments in the regions identified by the k-means clustering method.

In the following section, the results for pooling group number 5 are presented (*Fig. 3*). For an indirect estimation of the index flood using multiplicative formulae, the correlation matrix for the periods selected (1961-2010; 1985-2010) was derived by taking into account the physical, geographic and climatic characteristics available (*Fig. 4*). Correlation matrix provides a basis for comparison of the regions and visualizes different regions in the catchment.

The final formulas derived for the index flood estimation for the two observation periods are:

$$Q_{1985-2010} = e^{14.9873} \cdot AREA^{0.7094} \cdot TEMPERATURE^{-7.5429}, \quad (5)$$

$$Q_{1961-2010} = e^{14.9873} \cdot AREA^{0.7094} \cdot TEMPERATURE^{-7.5429}. \quad (6)$$

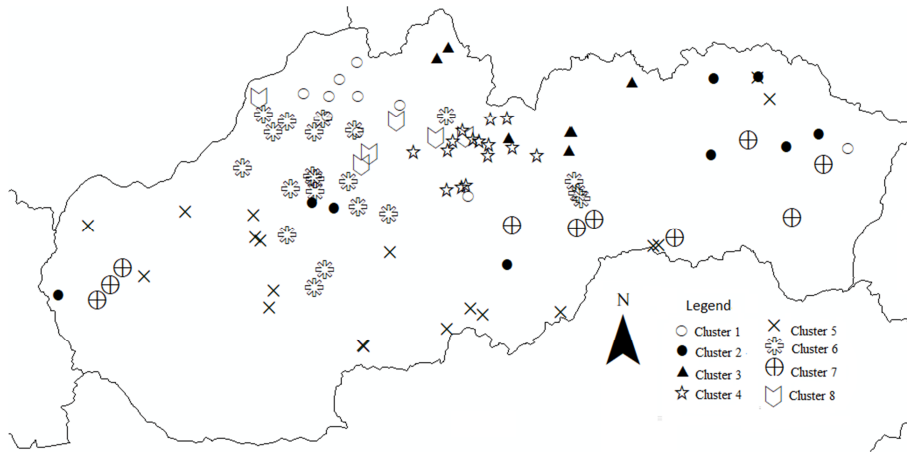


Fig. 2. Location of the catchments in the eight pooling groups

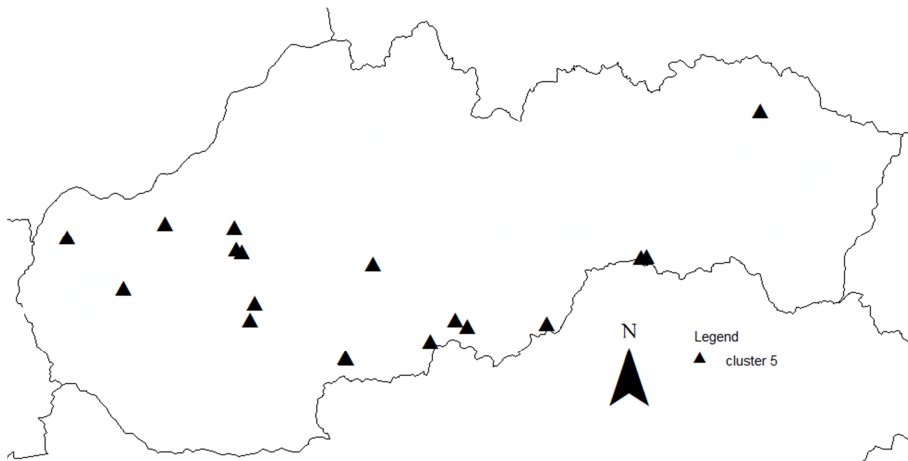


Fig. 3. Location of the catchment analysed in pooling group 5

Using the formulas (5) and (6), the index flood was estimated for all the catchments in pooling group 5 and for both observation periods. The comparison of the mean annual flood estimated with the top-kriging method and the index-flood method is shown in *Table I* and *Table II*. Based on the results, it can be concluded that the top-kriging method represents a more appropriate approach for the estimation of a mean annual flood as the values it estimated are closer to the data observed. This is true for both periods selected.

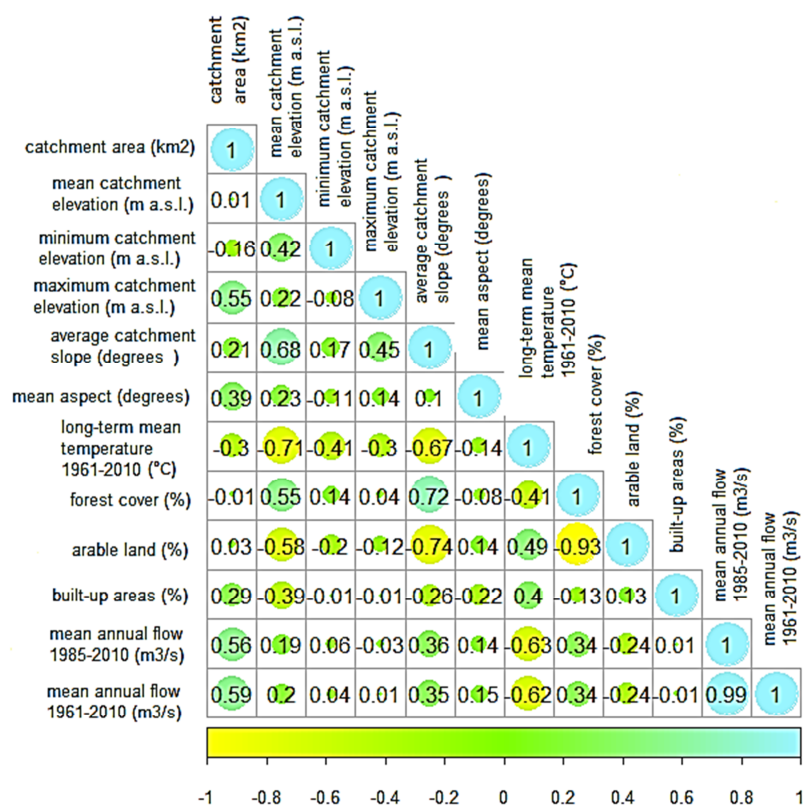


Fig. 4. The correlation matrix of the mean annual flood with the selected catchment and climatic characteristics for the two periods (1985-2010 and 1961-2010) evaluated

To distinguish how the estimations of the mean annual flood differ using these methods described above, the relative differences to the observed values were calculated. The deviations of the top-kriging method and the observed mean annual flood are fewer compared to the index flood method. The percentage deviations for the methods used and the periods selected are displayed in Fig. 5 and Fig. 6. A summary of the results calculated for both periods evaluated and both methods used are presented in Tables I. and Table II.

Tables I and Table II compare the values for the mean annual flood (m³/s) observed, the mean annual flood estimated by the index flood method, and the top-kriging method.

There are no significant differences in the mean annual flood observed between the two periods selected. When comparing the mean annual flood estimated by the methods used, it is clear that the top-kriging method provides values closer to the mean annual flood values observed. In this case, it can be concluded that the top-kriging method represents a more appropriate approach for the estimation of the mean annual flood.

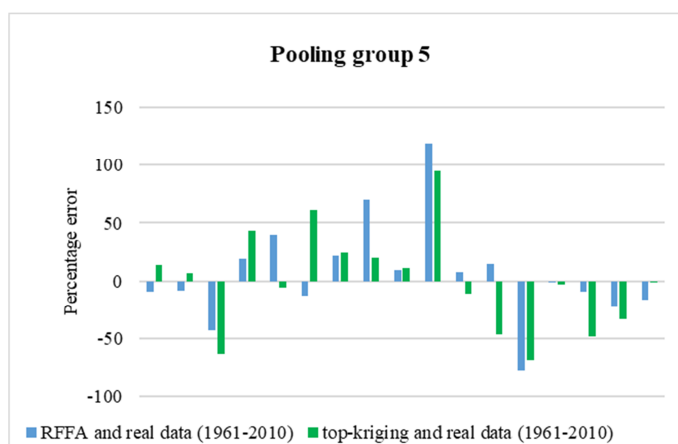


Fig. 5. The relative differences in the percentages of the mean annual floods (m^3/s) observed and those estimated (m^3/s) by the index flood and top-kriging methods for Pooling Group 5 and the period 1961-2010

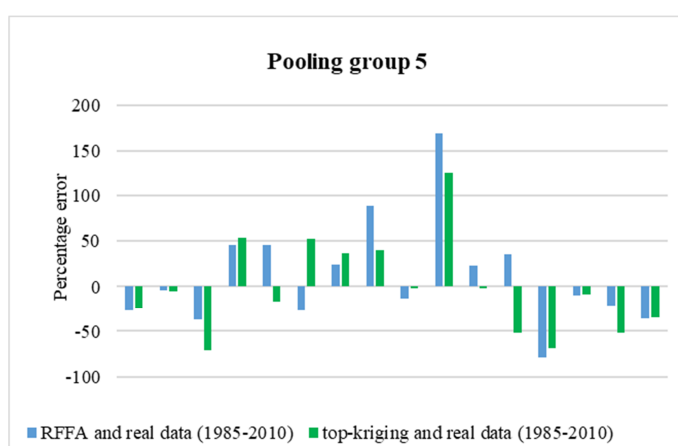


Fig. 6. The relative differences in the percentages of the mean annual floods (m^3/s) observed and those estimated (m^3/s) by the index flood and top-kriging methods for Pooling Group 5 and the period 1985-2010

5. Discussion and conclusion

The aim of the study was a comparison of two pooling scheme approaches for estimating the mean annual flood for the periods selected. The hydrometeorological stations are localized from the west to the east in Slovakia, and the majority of the stations are situated in Panonska panva. Based on the results calculated, it is obvious that the most significant percentage deviations were achieved for station 9580 (the

Ondava hydrological station). This is true for both methods used and both periods evaluated as well.

Table I

Comparison of the observed and estimated mean annual floods for the catchments in Pooling Group 5 and the 1985-2010 period

Station number	River	Observed mean annual flood (m ³ /s)	The estimation of the mean annual flood by the index flood method (m ³ /s)	The estimation of the mean annual flood by the top-kriging method (m ³ /s)
5030	Myjava	35.25	25.90	73.73
5230	Trnávka	2.45	3.90	8.57
6470	Jablonka	13.28	9.10	8.62
6640	Nitra	107.37	167.30	164.37
6690	Bebrava	30.06	31.50	31.79
6710	Bebrava	58.36	61.20	51.65
6800	Hostiansky potok	8.52	6.90	8.67
6820	Zitava	23.96	16.50	28.56
7228	Neresnica	21.75	29.60	14.28
7440	Ipeľ	33.54	42.60	69.53
7450	Tuharsky potok	7.55	6.10	5.54
7500	Tisovník	31.98	35.30	34.98
7600	Litava	26.85	14.20	19.13
7885	Blh	17.31	20.00	17.64
9020	Morava	23.46	32.10	31.25
9060	Turňa	5.15	24.70	16.24
9580	Ondava	86.32	32.00	38.25

On the other hand, lower values were calculated for station 9060 (the Turňa hydrological station) and 5030 (the Myjava hydrological station), when compared to the flow observed. This is also true for both periods evaluated and methods used.

Considering the mean annual flood observed, it can be stated that the differences between the mean annual floods observed do not represent significant values.

Because the top-kriging method provides better results when estimating the mean annual flood compared to the index flood method for both periods selected, it can be concluded that the top-kriging method represents a more appropriate approach for determining a mean annual flood. The methods used can be usefully applied to estimate mean annual flood in ungauged catchments within Slovakia. However, for further development, it is desirable to extend the physical-geographical characteristics and thus

innovate the results to a high level providing a tool useful not only for scientific purposes but also for design practice.

Table II

Comparison of the observed and estimated mean annual floods for the catchments in Pooling Group 5 and the 1961-2010 period

Station Number	River	Observed mean annual flood (m ³ /s)	The estimation of the mean annual flood flow by the index flood method (m ³ /s)	The estimation of the mean annual flood by the top-kriging method (m ³ /s)
5030	Myjava	35.46	30.98	66.15
5230	Trnávka	3.05	5.3	8.24
6470	Jablonka	12.73	10.7	8.89
6640	Nitra	105.82	136.09	157.52
6690	Bebrava	32.56	35.63	30.55
6710	Bebrava	57.01	68	57.22
6800	Hostiansky potok	8.64	8.02	9.73
6820	Zitava	26.35	18.9	27.78
7228	Neresnica	26.3	30.2	16.28
7440	Ipeľ	42.99	47.1	83.61
7450	Tuharsky potok	8.28	6.8	6.63
7500	Tisovník	37.8	37.9	38.75
7600	Litava	27.32	16.1	22.67
7885	Blh	24.22	22.1	21.73
9020	Morava	32.16	35.46	28.22
9060	Turňa	6.47	28.3	20.88
9580	Ondava	76.9	35.2	39.38

Acknowledgements

This study was supported by the Slovak Research and Development Agency under Contract No. APVV-15-0497 and VEGA Grant Agency No 1/0891/7. The authors thank the agencies for their research support.

References

- [1] Kelčík S., Pindjaková T., Šoltész A. Assessment and design of the flood protection measures in the district of Levice (Slovakia), *Pollack Periodica*, Vol. 11, No. 1, 2016, pp. 35–41.
- [2] Loucks, D. P., Van Beek E., Stedinger J. R., Dijkman J. P., Villars M. T. *Water resources systems planning and management: an introduction to methods, models and applications*, Paris, UNESCO Publishing, 2005.

- [3] Sivapalan M. Prediction in ungauged basins: A grand challenge for theoretical hydrology, *Hydrological Processes*, Vol. 17, No. 15, 2003, pp. 3163–3170.
- [4] Hosking M. J. R., Wallis J. R. *Regional frequency analysis: an approach based on L-moments*, Cambridge University Press, 1997.
- [5] Acreman C. M., Sinclair D. C. Classification of drainage basins according to their physical characteristics - an application of flood frequency analysis in Scotland, *Journal of Hydrology*, Vol. 84, No. 3, 1986, pp. 365–380.
- [6] Zrinji Z., Burn H. D. Flood frequency analysis for ungauged sites using a region of influence approach, *Journal of Hydrology*, Vol. 153, No. 1-4, 1994, pp. 1–21.
- [7] Meigh J. R., Farquharson F. A. K., Sutcliffe J. V. A worldwide comparison of regional flood estimation methods and climate, *Hydrological Science Journal*, Vol. 42, No. 2, 1997, pp. 2225–2244.
- [8] Portela M. M., Dias A. T. Application of the index-flood method to the regionalization of flood peak discharges on the Portugal mainland, *WIT Transactions on Ecology and the Environment*, Vol. 83, 2005, pages 11.
- [9] Malekinezhad H., Zare-garizi A. Regional frequency analyses of daily rainfall extremes using L-moments approach, *Atmosfera*, Vol. 27, No. 4, 2014, pp. 411–427.
- [10] Basu B., Srinivas V. V. Evaluation of the index-flood approach related regional frequency analysis procedures, *Journal of Hydrologic Engineering*, Vol. 21, No. 1, 2016, pages 1–12.
- [11] Bocchiola D., De Michele C., Rosso R. Review of recent advances in index flood estimation, *Hydrology and Earth System Sciences*, Vol. 7, No. 3, 2003, pp. 283–296.
- [12] Forestieri A., Lo Conti F., Blenkisop S., Cannarozzo M., Fowler J. H., Noto V. L. Regional frequency analysis of extreme rainfall in Sicily (Italy), *International Journal of Climatology*, Vol. 38, No. 1, 2018, pp. 698–716.
- [13] Hosking J. R. M., Wallis J. R. Some statistics useful in regional frequency analysis, *Water Resources Research*, Vol. 29, No. 2, 1993, pp. 271–281.
- [14] Kheirfam H., Vafakhah M. Assessment of some homogeneous methods for the regional analysis of suspended sediment yield in the south and southeast of the Caspian Sea, *Journal of Earth System Science*, Vol. 124, No. 6, 2015, pp. 1247–1263.
- [15] Nathan R. J., McMahon T. A. Identification of homogeneous regions for the purposes of regionalisation, *Journal of Hydrology*, Vol. 121, No. 1-4, 1990, pp. 217–238.
- [16] Kyselý J., Píček J., Huth R. Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic, *Studia Geophysica et Geodaetica*, Vol. 51, No. 2, 2007, pp. 327–344.
- [17] Kohnová S., Szolgay J. Regional estimation of index flood and standard deviation of a summer maximum flows in the Tatras, *Journal of Hydrology and Hydromechanics*, Vol. 51, No. 4, 2003, pp. 241–255.
- [18] Kohnová S., Szolgay J., Hlavčová K. Regional flood frequency analysis of annual maximum floods from the winter season in Slovakia, *Meteorological Journal*, Vol. 11, No. 1-2, 2008, pp. 65–70.
- [19] Kohnová S., Szolgay J., Solin L., Hlavčová K. *Regional methods for prediction in ungauged basins*, Ostrava: KEY Publishing, 2006.
- [20] *Flood Estimation Handbook*, Part 3. Statistical procedures for flood frequency estimation, *IH Wallingford*, 1999.
- [21] DVWK Regeln 101/1999, Choice of design flood. Recommendation for calculating the flood probability, (in German) Verlag Paul Parey, Hamburg, 1999.
- [22] *Australian Rainfall and Runoff*, A guide to flood estimation, Institutions of Engineers Australia, Book VI, Vol. 1, 1998.
- [23] Dalrymple T. *Flood frequency methods*, US Geological Survey, 1960.

- [24] Skøien J. O., Merz R. Bloschl G. Top-kriging - geostatistics on stream networks, *Hydrology and Earth System Sciences Discussions*, European Geosciences Union, Vol. 10, No. 2, 2006, pp. 277–287.
- [25] Nagy D., Aszalós L., Mihálydeák T. Finding the representative in a cluster using correlation clustering, *Pollack Periodica*, Vol. 14, No. 1, 2019, pp. 15–24.
- [26] MacQueen J. Some methods for classification and analysis of multivariate observations, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, USA, 18-21 July 1965, University of California Press, Vol. 1, 1967, pp. 281–297.