IOP Conference Series: Materials Science and Engineering

Detecting Similarity in Flood Seasonality of Slovak and Austrian Catchments

To cite this article: Romana Marková *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **471** 022027

Detecting Similarity in Flood Seasonality of Slovak and Austrian Catchments

Romana Marková [1], Silvia Kohnová [1], Juraj Parajka [2]

[1] Slovak University of Technology, Department of Land and Water Resources Management, Radlinského 11, 810 05, Bratislava, Slovakia
[2] Vienna University of Technology, Institute of Hydraulic Engineering and Water Resources Management, Karlsplatz 13/222, A-1040 Vienna, Austria

romana.markova@stuba.sk

Abstract. The main aim of the study is the flood seasonality of Slovak and Austrian catchments in the 50-year period of 1961-2010. Flood seasonality analyses have become interesting topics for many authors worldwide. In the context of flood seasonality, the dates of annual maximum floods at 556 gauging stations across Austria (475) and Slovakia (81) were analyzed. The length of the time series of the selected gauging stations is variable, ranging from 30-50 years. We have focused on an investigation of the relationship between flood seasonality characteristic -r (the concentration of a flood around the mean date of a flood occurrence), the hydrological characteristics (long-term average air temperature (°C) for the period 1961-2010, the long-term average daily precipitation (mm) for the period 1961-2010), physiographic characteristics of catchments (catchment area (km^2), and the outlet elevations (m a.s.l.)). The pooling of the catchments was processed by a nonhierarchical k-means cluster analysis and boxplot analysis. The input for the k-means cluster analysis was r (the concentration of a flood around the mean date of a flood occurrence), the long-term average daily precipitation (mm) for the period 1961-2010 for the individual catchments, the catchment area (km^2), and the outlet elevations (m a.s.l.). The results show that the optimal number of clusters for the selected catchments is three and that these clusters show some similar features. Cluster one is located in the High Tatras, the Alps and the High Tauern (the long-term average air temperature is between -2 ÷ 8°C; the long-term average daily precipitation is between 3-6mm); cluster two is located in the Low Tatras and the Lower Tauern (the long-term average air temperature is between 1 ÷ 10°C, the long-term average daily precipitation between 2-4mm); and cluster three is located in the lowlands of Austria and Slovakia (the long-term average air temperature is between 5 ÷ 10°C, the long-term average daily precipitation is between 1-4mm).

1. Introduction

River floods are a common natural disaster in Europe and, along with storms, are the most significant natural hazards in Europe in terms of economic damage [1]. They are mainly caused by prolonged or heavy precipitation events or snowmelt [1]. River floods have become an interesting topic for many authors worldwide. More than 325 major river floods have been reported in Europe since 1980, of which more than 200 have been reported since 2000 [1]. Climate change and its impact on floods around the world has been a subject in many recent studies. In this study we have focused on statistical

IOP Conf. Series: Materials Science and Engineering 471 (2019) 022027 doi:10.1088/1757-899X/471/2/022027

analyses of floods and used two basic methods for our survey, i.e., circular statistics and a k-means clustering algorithm.

We used circular statistics for the flood seasonality analysis. Flood seasonality analysis is a method which is used for a better understanding of the causes of flooding. This method is based on calculating the average day within a year on which floods have occurred during the observation period for each station [2]. We also calculated the flood concentration index -r (a concentration of the flood timings/occurrences around the mean date of the flood occurrence), which was included into the cluster analysis. Multivariate statistical methods such as cluster analysis can assist in the identification of factors influencing flood behaviour [3].

Floods and statistical analyses of flood characteristics are topics of interest to many authors, and recent studies have tried to shed light on floods and their connection with climate change; e.g., Hall and Blöschl [4] examined the spatial and temporal patterns of flood seasonality at a continental scale, using an extensive database that covers all climatic regions in Europe. They focused on an identification of regions with similar seasonal flood characteristics and on the description of the full temporal distribution of the flood events within the year. Jeneiová et al. [5] studied the spatial variability of seasonal flood occurrences in the Upper Danube region for the period 1961-2010. They focused on the understanding of the factors that control the spatial variability of winter and summer floods in 88 basins with different physiographic conditions. Parajka et al. [6] [7] used flood seasonality analysis for the identification of flood-producing processes. Parajka [7] derived eight clusters using flood characteristics (the mean date of the occurrence of annual floods, concentration of the flood timings/occurrences around the mean date of a flood occurrence), rainfall characteristics (the mean date of the occurrence of the annual precipitation maxima, concentration of the annual precipitation timings/occurrences around the mean date of the annual precipitation occurrence) and planar coordinates (X_{cor}, Y_{cor}) for the Alpine-Carpathian range. Gaál et al. [8] focused on a delineation of the homogeneous regions of rainfall gauging stations for a regional frequency analysis of k-day precipitation amounts (k=1 to 5 days) using objective and process-based logical pooling techniques. They derived three regions for the rainfall gauging stations in Slovakia. Gaál et al. [9] focused on reporting about a method for the estimation of intense rainfall events over a few stations in Switzerland by estimating the threshold intensity with the help of lightning data, which captured the convective character of these intense events. The aim of this study is to better understand flood processes in Alpine-Carpathian basins and detect which catchment characteristics influence these processes. This article is related to a previous article [3], where we mainly used physiographic catchment characteristics (catchment area, forest cover, mean catchment elevation) and r (concentration of the flood timings/occurrences around the mean date of a flood occurrence). This paper is organized as follows: introduction, data, methodology, results and discussion, and conclusions.

2. Data collections

For our analysis 556 gauging stations in Austria (475) and Slovakia (81) were selected. Over the analysed period of 1961-2010, the lengths of the hydrological time series vary from 30 to 50 years, depending on the availability of the data. The hydrological data (annual maximum floods, dates of the annual maximum floods) and physiographic catchment characteristics (catchment area (km^2), outlet elevation (m a.s.l.)) for Austria were obtained from the Austrian Hydrographic Service (HZB); the same hydrological data for Slovakia were obtained from the Slovak Hydrometeorological Institute (SHMI), Global Runoff Data Centre (GRDC) [10] and European Water Archive (EWA) [10]. All the selected gauging stations are displayed in figure 1.

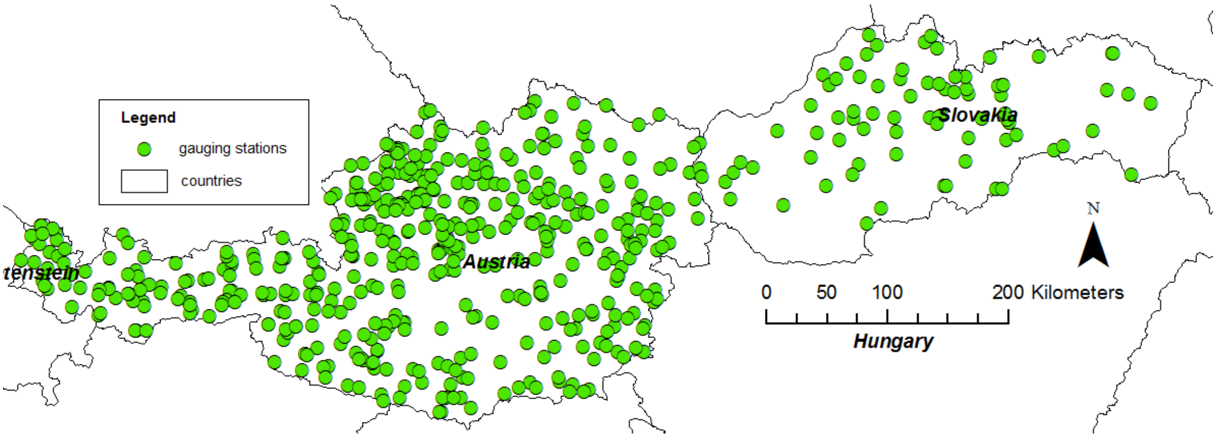

Figure 1. Location of selected gauging stations across Austria and Slovakia

A map of the long-term average daily precipitation totals (mm) for the period 1961-2010 was derived (figure 2) for the purpose of the k-means clustering analysis. This map was created using daily E-OBS station data [11].

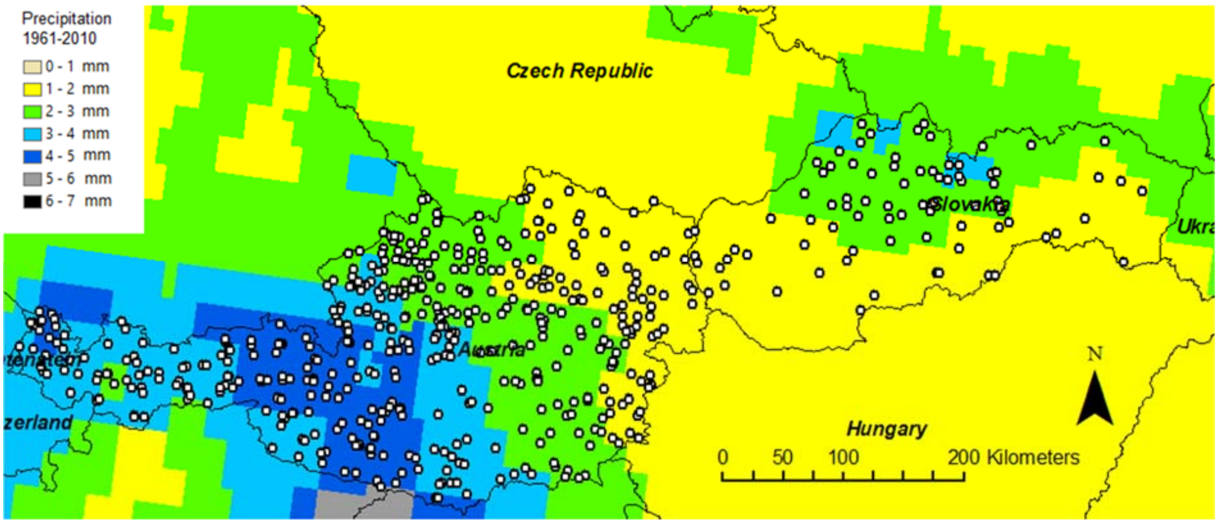

Figure 2. Map of the long-term average daily precipitation totals (mm) for the period 1961-2010

3. Methodology

The statistical analysis was processed using directional statistics [12] [13] [14] and k-means clustering [15].

3.1. Assessment of the flood seasonality

The assessment of the flood seasonality was computed using directional statistics.

The date of each flood event has been plotted on a unit circle, where the position of the event θ_i, is defined as:

$$\theta_i = D_i \frac{2\pi}{m_i}, \tag{1}$$

where $D_i = 1$ corresponds to January 1 and $D_i = m_i$ to December 31, and where m_i is the number of days in that year. The average date of the occurrence $\overline{D}$ of a flood at a station is calculated as follows [12] [13]:

$$\overline{D} = \begin{cases} tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right).\frac{\bar{m}}{2\pi}, & \bar{x} > 0, \bar{y} \geq 0 \\ tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right).\frac{\bar{m}}{2\pi} + \pi, & \bar{x} \leq 0 \\ tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right).\frac{\bar{m}}{2\pi} + 2\pi, & \bar{x} > 0, \bar{y} < 0 \end{cases} \tag{2}$$

with

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i), \tag{3}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\sin(\theta_i), \tag{4}$$

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i, \tag{5}$$

where $\bar{x}$ and $\bar{y}$ are the cosine and sine components of the average date; $\bar{m}$ is the average number of days per year (365.25); and n is the total number of flood peaks at that station [2].

The concentration of the flood timings/occurrences around the mean date of the flood occurrences is characterized by the length of the vector r. This parameter is calculated as follows [12]:

$$r = \sqrt{\bar{x}^2 + \bar{y}^2}. \tag{6}$$

Parameter r ranges from r=0 (high variability of the date of the occurrence, low seasonality) to r=1 (all flood events occur on the same day, low variability of the date of the occurrence, high seasonality) [7]

3.2. K-means clustering method

The purpose of a cluster analysis is to place objects into groups so that the objects in each group have the highest degree of similarity to each other, while the objects in different clusters have a maximum degree of dissimilarity [16]. The k-means clustering method belongs among the most often used pooling methods in the hydrological sciences.

The optimal number of clusters was computed in an R Studio software environment [17] with the function fviz_nbclust (package factoextra [18]). This function determines and visualizes an optimal number of clusters using different methods. We applied the average silhouette width [19] in our study:

$$s(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{7}$$

where a_i is the average dissimilarity between i and all the other points of the cluster to which i belongs, and b_i is the lowest average dissimilarity between i and any other cluster of which i is not a member.

The average silhouette width Sj over a cluster j is defined as follows:

$$Sj = \frac{1}{m}\sum_{i=1}^{m} s(i), \tag{8}$$

where $s(i)$ is the silhouette width of each datum, and m is the amount of data assigned to this cluster. It can be interpreted as follows [20]: 0.71-1.00 (very strong clustering structure), 0.51-0.70 (strong clustering structure), 0.26-0.5 (weak clustering structure), ≤0.25 (no clustering structure).

Before the k-means clustering algorithm is applied, the variables are standardized to a mean of zero and a standard deviation of 1 (z-score standardisation) to reduce the effect of variables with small or large ranges [3].

4. Results and discussion

The main aim of the study is a better understanding of the flood processes in the Alpine-Carpathian basins, an identification of similar pooling groups in terms of flood seasonality, and their hydrological and physiographic catchment characteristics.

The pooling groups were identified using the k-means clustering method. The input variables for the k-means clustering were: r (the concentration of the flood timings/occurrences around the mean date of a flood occurrence), the long-term mean daily precipitation totals (mm) for the period 1961-2010, the catchment area (km^2), and the outlet elevations (m a.s.l.). The optimal number of clusters computed by the average silhouette width is three and is shown in figure 3, but according to figure 3, the optimal number of clusters could also be two. As stated in our previous article [3], we derived two clusters mainly using physiographic catchment characteristics (the catchment area, the forest cover, the mean catchment elevation) and r (the concentration of the flood timings/occurrences around the mean date of a flood occurrence). The main difference between the previous article [3] and this study is the use of not only different physiographic catchment characteristics, but also the hydrological characteristics (long-term mean daily precipitation totals (mm) for the period 1961-2010).

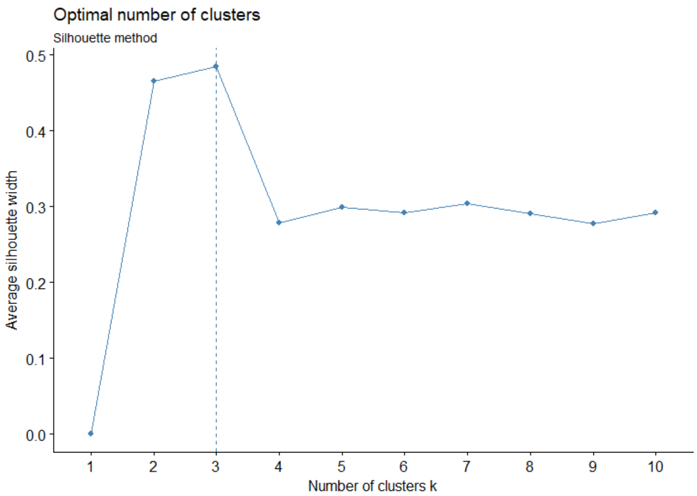

Figure 3. Optimal number of clusters computed by the average silhouette width

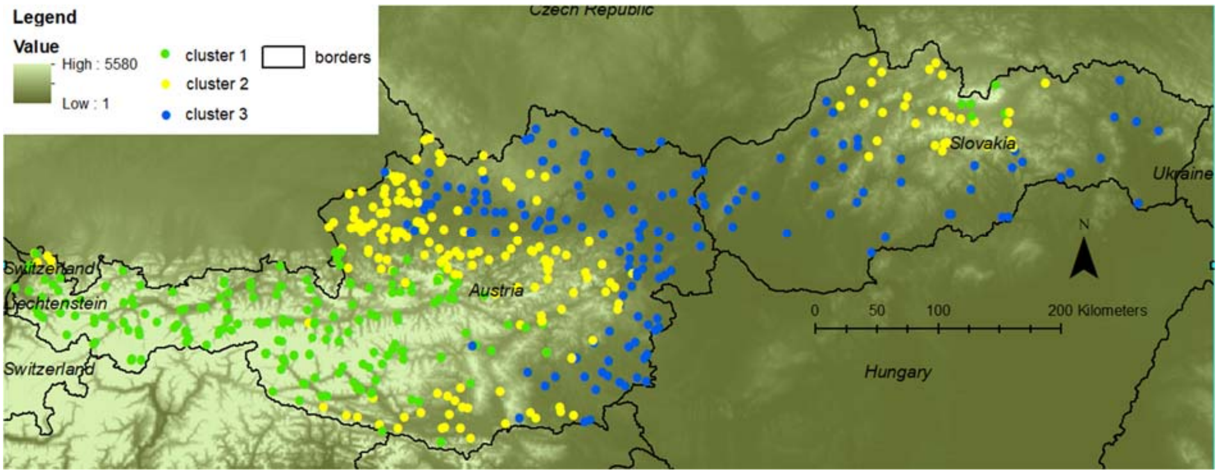

Figure 4. Spatial distribution of the catchments belonging to Pooling Groups 1, 2 and 3

Figure 4 visualises the location of catchments due to similarities in flood regimes in the derived pooling groups. The pooling groups differ from each other in climatological as well as morphological characteristics as presented in figure 6 and as well as in flood regimes as presented in figure 5.

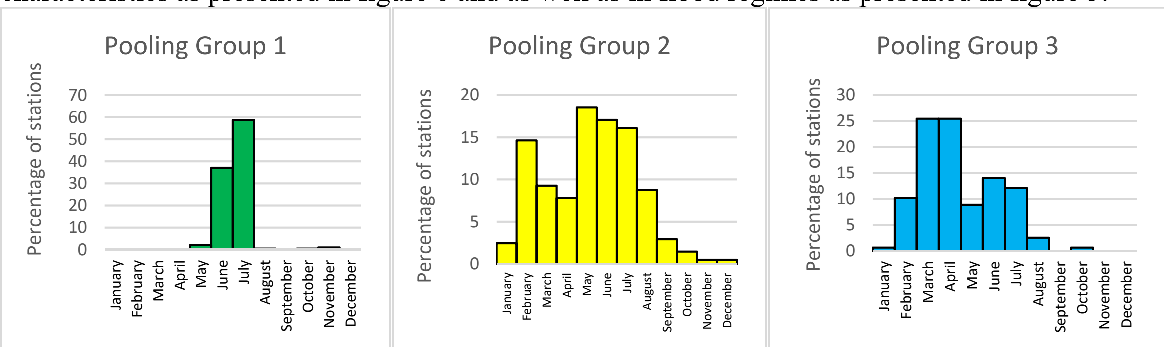

Figure 5. Histogram of stations per mean date of the occurrence of annual floods for the selected pooling groups

The main characteristics of the individual pooling groups can be summarized as follows:

Pooling group one is located in the High Tatras, the Alps and the High Tauern (figure 4). Floods occur mostly in June and July (figure 5). This pooling group is characterized by the highest mean catchment elevations (from 579 to 2887 m a.s.l.) and the highest outlet elevations (from 362 to 1995 m a.s.l.) (figure 6 A), B)). The long-term mean air temperature (°C) for the period 1961-2010 varies from -2 °C to 8 °C (figure 6 D))

Pooling group two is located in the Low Tatras and the Lower Tauern (figure 4). This pooling group is characterized by wide temporal spread flood timing due to a mixed regime of snowmelt and summer floods (figure 5). The mean catchment elevation ranges from 350 to 2046 m a.s.l., and the outlet elevations range from 256 m a.s.l. to 809 m a.s.l. (figure 6 A),B)). The long-term mean air temperature (°C) for the period 1961-2010 varies from 4 °C to 10 °C (figure 6 D)).

Pooling group three is located in the lowlands of Austria and Slovakia (figure 4). Floods mostly occur in March and April (figure 5).

Mean catchment elevations range from 154 m a.s.l. to 1316 m a.s.l., and the outlet elevations range from 91 m a.s.l. to 528 m a.s.l. (figure 6 A),B)). The long-term mean air temperature (°C) for the period 1961-2010 varies from 5 °C to 10 °C (figure 6 D)).

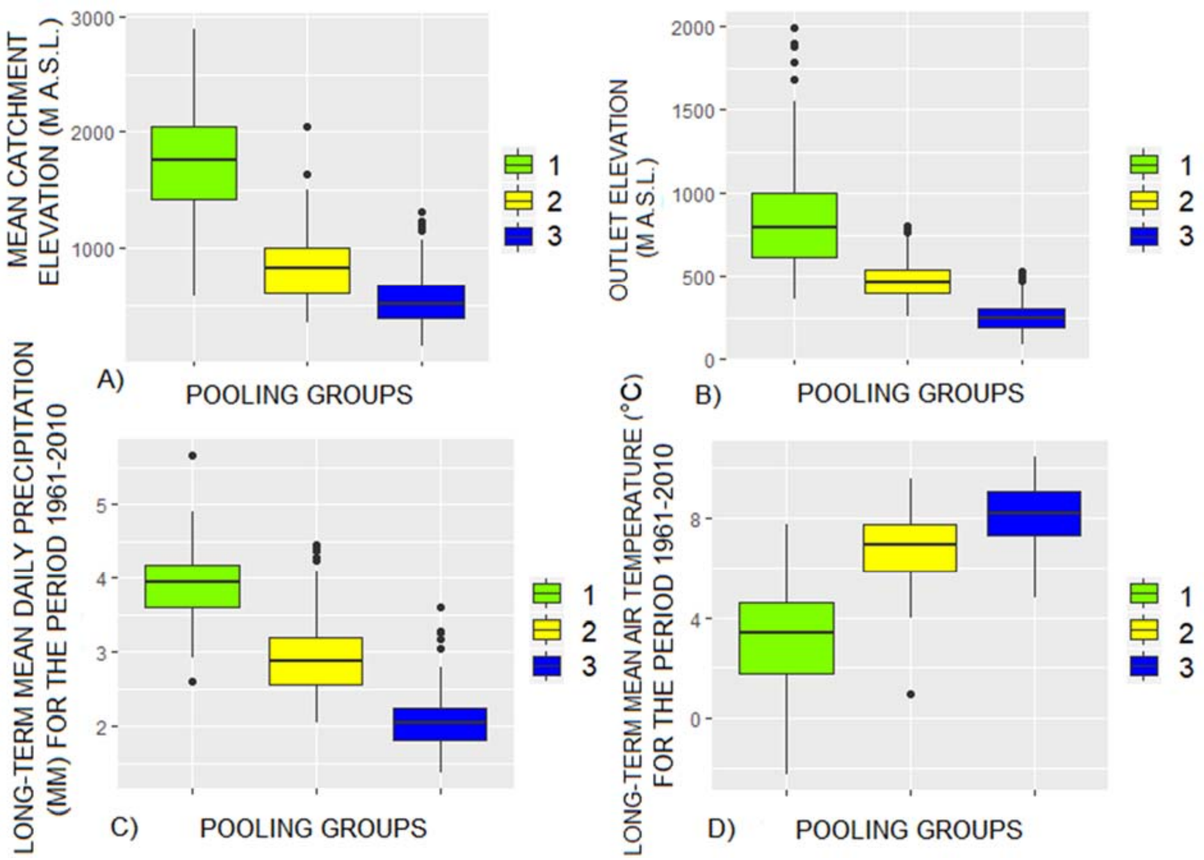

Figure 6. Box plots of the hydrological characteristics (the long-term mean precipitation totals (mm) for the period 1961-2010, the long-term mean air temperature (°C) for the period 1961-2010) and physiographic catchment characteristics (the mean catchment elevation (m a.s.l.) and the outlet elevation (m a.s.l.)) for the derived pooling groups

5. Conclusions

The aim of the study was to analyse the similarities in the flood seasonality of Slovak and Austrian catchments. The pooling of catchments in groups with similar flood seasonalities was processed using directional statistics and the k-means clustering method. Directional statistics was used for estimating the flood seasonality parameter r (the concentration of a flood timings/occurrences around the mean date of a flood occurrence), which was used together with other climatological and morphological characteristics such as the hydrological input for the k-means clustering analysis, which resulted in the formation of three pooling groups.

Pooling group one is characterized by the highest mean catchment and outlet elevation (m a.s.l.), the highest long-term average daily precipitation (mm) during the time period 1961-2010 and the lowest long-term average air temperature (°C) during the time period 1961-2010. This pooling group is located in the High Tatras, the central Alps, and the High Tauern and is characterized by extreme summer floods, which are mostly of a convective origin and which prevail in June and July.

Pooling group two represents catchments located in the Low Tatras and the Low Tauern. The values of the hydrological and physiographic catchment characteristics differ among the rest of the pooling groups. This pooling group is characterized by wide temporal spread flood timing.

Pooling group three is characterized by the lowest mean catchment and outlet elevations (m a.s.l.), the lowest long-term average daily precipitation (mm) during the time period 1961-2010, and the highest long-term average air temperature (°C) during the time period 1961-2010. This pooling group is located in the lowlands of Austria and Slovakia and is characterized by snowmelt floods, which occur mostly in March and April.

Acknowledgments
This work was supported by the Slovak Research and Development Agency under Contract No. APVV-15-0425 and VEGA 1/0891/17. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). We would like to thank the Slovak Hydrometeorological Institute in Bratislava, the Global Runoff Data Centre, 56002 Koblenz, Germany and the Austrian Hydrographic Service (HZB) for providing the data for this work.

References

[1] EEA, "Climate change, impacts and vulnerability in Europe 2012," *EEA Report No 12/2012, European Environment Agency, Copenhagen*. ISSN 1725-9177. 300 pp. doi:10.2800/66071, 2012
[2] Blöschl et al. "Changing climate shifts timing of European floods," *Science (New York, N.Y.)*, vol. 357, issue 6351, pp. 588-590. doi:10.1126/science.aan2506, 2017
[3] R. Marková, S. Kohnová, J. Parajka, J. Poórová, J. Hall, "Identification of catchments with similar flood characteristics in the Alpine-Carpathian basins," *SGEM 2017. 17th International Multidisciplinary Scientific GeoConference. Vol. 17. Water Resources. Forest, Marine and Ocean Ecosystems: conference processdings. Albena, Bulgaria, 29 June – 5 July 2017*. Sofia: STEF 92 Technology, pp. 293-300, ISSN 1314-2704, ISBN 978-619-7408-27-0. DOI: 10.5593/sgem2017H/33, 2017
[4] J. Hall, G. Blöschl, "Spatial patterns and characteristics of flood seasonality in Europe," *Hydrology and Earth System Sciences Discuss*., http://doi.org/10.5194/hess-2017-649,in review, 2017
[5] K. Jeneiová, S. Kohnová, J. Hall, J. Parajka, "Variability of seasonal floods in the Upper Danube River basin," *Journal of Hydrology and Hydromechanics*, vol. 64, issue 4, pp. 357-366, doi: 10.1515/johh-2016-0037, 2016
[6] J. Parajka, S. Kohnová, R. Merz, J. Szolgay, K. Hlavčová, G. Blöschl, "Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria," *Hydrological Sciences Journal*, vol. 54, issue 3, pp. 456-473, 2009.
[7] J. Parajka, S. Kohnová, G. Bálin, M. Barbuc, M. Borga, P. Claps, S. Cheval, A. Dumitrescu, E. Gaume, K. Hlavčová, R. Merz, M. Pfaundler, G. Stancalie, J. Szolgay, G. Blöschl, "Seasonal characteristics of flood regimes across the Alpine-Carpathian range," *Journal of Hydrology*, vol. 394, issue 1-2, pp. 78-89, doi:10.1016/j.hydrol.2010.05.05, 2010.
[8] L. Gaál, J. Szolgay, M. Lapin, P. Faško, "Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis," *Journal of Hydrology and Hydromechanics*, vol. 57, issue 4, pp. 226-249, doi: 10.2478/v10098-009-0021-1, 2010.
[9] L. Gaál, P. Molnár, J. Szolgay, "Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland," *Hydrol. Earth Syst. Sci*., vol. 18, issue 5, pp. 1561-1573, http://doi.org/10.5194/hess-18-1561-2014, 2014.
[10] GRDC, 2018. Global Runoff Database. Global Runoff Data Centre. Koblenz, Federal Institute of Hydrology, BgG, Germany. http://www.bafg.de/DE/02_Aufgaben/05_International/02_GRDC/GRDC.html?nn=180066
[11] M. R. Haylock, N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, M. New, "A European daily high-resolution gridded dataset of surface temperature and precipitation for 1950-2006," *J. Geophys. Res (Atmospheres)*, vol. 113, D20119, doi: 10.1029/2008JD10201
[12] A. C. Bayliss, R. C. Jones, "Peaks-over-threshold flood database: Summary statistics and seasonality. IH Report No. 121," *Institute of Hydrology*, Wallingford, UK, 1993
[13] K.V. Mardia, "Statistics of directional data," *Academic Press Inc. London*, 1972

[14] D.H. Burn, "Catchment similarity for regional flood frequency analysis using seasonality measures," *Journal of Hydrology*, vol. 202, issue 1, pp. 212-230, 1997
[15] J.B. Macqueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Pres. ,vol. 1, pp. 281-297, 1967
[16] S. Soltani, R. Modarres, "Classification of spatio-temporal patterns of rainfall in Iran using a hierarchical and divisive cluster analysis," *Journal of Spatial Hydrology,* vol. 6, No. 2, 2006
[17] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org
[18] A. Kassambara, F. Mundt (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. http://CRAN.R-project.org/package=factoextra
[19] L. Kaufman, P. J. Rousseeuw, "Findings Groups in Data: An Introduction to Cluster Analysis," Wiley, New York, 1990.
[20] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987