INFRASTRUKTURA I EKOLOGIA TERENÓW WIEJSKICH
INFRASTRUCTURE AND ECOLOGY OF RURAL AREAS

No II/1/2019, POLISH ACADEMY OF SCIENCES, Cracow Branch, pp. 25-37
Commission of Technical Rural Infrastructure

DOI: https://doi.org/10.14597/INFRAECO.2019.2.1.003

RECENT TRENDS IN ANNUAL MAXIMUM FLOWS WITHIN THE UPPER VISTULA RIVER CATCHMENT

Mateusz Całek, Stanisław Węglarczyk
Cracow University of Technology

Abstract

All uninterrupted time series of annual maximum flows of size at least 30 recorded in the period 1951-2016 in the Upper Vistula River catchment, were taken into trend analysis. Each of the 138 time series ended not earlier than in 2012. To estimate the trend, the nonparametric Theil-Kendall linear regression method was used. After removing the trend, lag-1 Kendall rank autocorrelation coefficient was calculated and, if the coefficient was significant at 5% level, was used to correct the variance of the Kendall *S* statistic which otherwise remained unchanged. Finally, the variance-corrected Mann-Kendall trend test was used, detecting 22 significant (at 5% level) linear trends of which only two were the effect of autocorrelation. All 138 significant and non-significant trends showed certain areal clustering clearly visible on the map of the catchment, which suggested dividing the area into three parts according the direction of trend and/or the number of statistically significant trends. Generally, the trends in the southern of the Upper Vistula River catchment are increasing, the opposite is true for the northern part. This finding does not concern the north-west part of the catchment where both kinds of trends are observed, which may be explained by strong anthropogenic influence.

Key words: annual maximum flow, Upper Vistula River, Kendall-Theil linear regression, autocorrelation, variance correction approach

INTRODUCTION

Recently, geophysical (especially hydrometeorological) time series analysis is to a great extent aimed at the question of detecting nonstationarity of the real-world underlying processes. The importance of the nonstationarity question is connected with the importance of changes caused by human activities such as land use, water resources management, activities influencing climate changes, as well the changes of natural (e.g. seismic) origin. Usually, conclusions on nonstationarity are based on trend analysis based on relatively short time series. The relevant hydrometeorological literature is very extensive. Precipitation trend analysis may be found e.g. in Brath *et al*. (1999), Zhang *et al*. (2000), Trenberth *et al*. (2003), Cebulska *et al.* (2007), Prosdocimi *et al*. (2014), Cebulska (2015), Razavi *et al*. (2016), Meshram *et al*. (2016), Svoboda *et al*. (2016). Streamflow trends are discussed e.g., in Douglas *et al*. (2000), Burn and Hag Elnur (2002), Khaliq *et al.* (2009), Wrzesiński (2009), Prosdocimi *et al*. (2014), Zhang *et al*. (2014), Burn and Whitfield (2018) and Su *et al*. (2018).

In many cases trend analysis is performed with the use of ordinary least squares linear regression method. However, full application of the method (i.e., with trend significance testing) requires that the distribution of the trend residuals be normal. Such strong assumption may lead to false conclusions when the sample data is highly skewed, which is often the case when outliers are present in the sample. Nonparametric methods for trend estimation are a very useful alternative as they do not require the assumption of normality of residuals and are robust to outliers. The Kendall-Theil nonparametric linear regression (Helsel and Hirsch 2002), called also the Theil-Sen approach (Yue *et al.* 2002), based on using median instead of mean, is one of such methods.

To assess the significance of trend the Mann-Kendall trend test is usually applied. However, if the autocorrelation in the data exists, it changes the probability of false trend detection (Hamed and Rao 1998). The existence of positive/negative autocorrelation increases/decreases the variance of Mann-Kendall trend test statistic which may finally effect in significance or non-significance of the trend. Hamed and Rao (1998) corrected the variance for autocorrelation by replacing the sample size in the Mann-Kendall test with the effective sample size. Other methods for dealing with autocorrelation exists, e.g. pre-whitening method (Douglas *et al*. 2018, Su *et al*. 2018).

The aim of this paper is to estimate linear trends in annual maximum flows Q_{max} recorded in the Upper Vistula River catchment, south-eastern Poland, in the period 1951-2016, with the Kendall-Theil nonparametric linear regression and test their significance, taking into consideration the possible existence of autocorrelation in the data. The effect of the existence of significant autocorrelation on final result of the Mann-Kendall trend test is also examined.

MATERIALS AND METHODS

Area under study and data

The Upper Vistula River is defined as the part of the Vistula River down to the Zawichost gauging station. Its length is 394 km, its catchment covers the area of 50732 km^2 of which about 91% lies within Poland territory (Chełmicki 1991).

According to information of the Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB) – the provider of the data – out of all river gauging stations within the catchment, 190 stations have at least 30-year Q_{max} time series records from the period 1951-2016. However, as is shown in Fig. 1a, some series are broken, some come from the very early years of the period. To trend analysis only recent series were selected, where ‚recent' series means a series ending not earlier than in 2013. Additionally, only uninterrupted series were chosen (Fig. 1b).

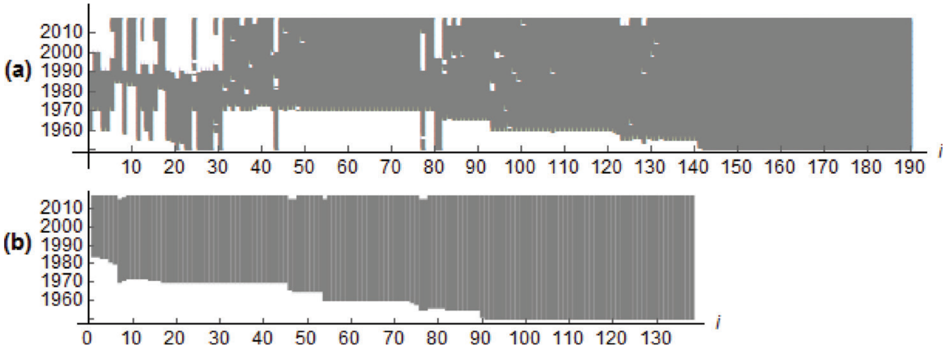

Figure 1. Temporal coincidence of at least 30-year Q_{max} time series in the area under study, ranked according to their size in non-descending order. (a) 190 series, some of them are broken, some are ending in the 1990s; (b) all those 138 series out of the 190 which are uninterrupted and end in the years 2013-2016.

Finally, 138 Q_{max} time series were used to trend analysis. Locations of gauging stations with times series sizes is shown in Figure 2.

Catchment areas down to gauging station cross-section ranged from 24 to almost 51,000 km^2. (Fig. 3). Almost 70% of the areas (96 out of 138) has the size not exceeding 1000 km^2, the areas of 9 catchments exceed 10,000 km^2. Elevation of gauging stations vary from 134 m a.s.l. to 966 m a.s.l. (Fig. 3). About 67% (92) of elevations of do not exceed 300 m a.s.l.

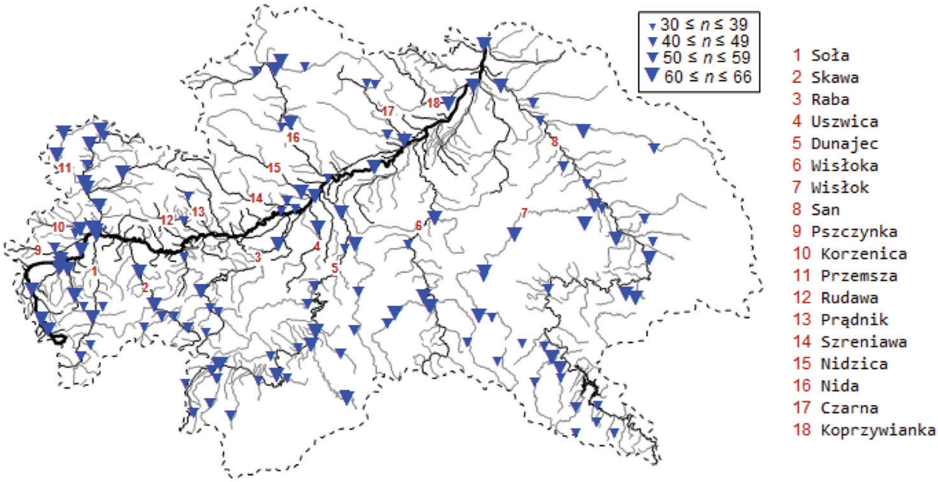

Figure 2. Location of flow gauging stations within the Upper Vistula River catchment. Numbers 1 through 18 indicate some larger tributaries of the Upper Vistula River; n denotes time series size.

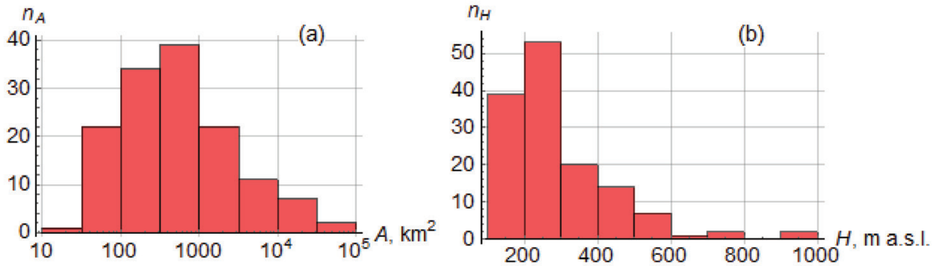

Figure 3. Distribution of (a) the number n_A of catchment areas A down to gauging station, and (b) the number n_H of gauging station elevations H.

In Fig. 4 three basic statistics: mean Q_{max} flow $\bar{Q}_{max}$, coefficient of variation C_V, and skewness coefficients C_S, of 138 time series of annual maximum flows versus catchment area are presented. As expected, mean flow exhibits visible influence of catchment area magnitude. Most of coefficients of variations do not exceed 100% and are higher than 50%. Their values tends to decrease with catchment area up to ca. 5000 km^2 and stabilize afterwards.

Skewness coefficients C_S seem to be least affected by the catchment area; a considerable number of C_S exhibit large values which suggests non-normal distribution of Q_{max} and suggests therefore using a nonparametric method for trend estimation.

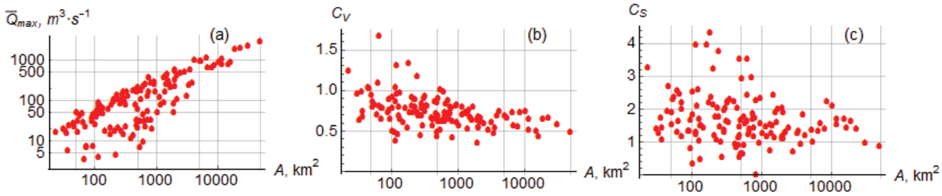

Figure 4. Basic characteristics of 138 time series of Q_{max} versus catchment area A: (a) mean, $\overline{Q}_{max}$, (b) coefficient of variation, C_V, and (c) skewness coefficient, C_S.

Methods: trend estimation and testing

For each time series a linear model of time trend of annual maximum flow, Q_{max}, was assumed in the following form:

$$E[Q_{max}(t)] = \alpha_0 + \alpha_1 t \tag{1}$$

where $E[Q_{max}(t)]$ denotes the expected value of an annual maximum flow in year t. The regression coefficients were estimated with the Kendall-Theil method (Theil 1950, Sen 1968, Helsel and Hirsch 2002) which does not require the normality assumption on the distribution of residuals.

The Kendall-Theil nonparametric linear regression method gives the following formulas for regression coefficients (Helsel and Hirsch 2002):

$$\hat{\alpha}_1 = \underset{\substack{i=1,\ldots,n-1 \\ j=i+1,\ldots,n}}{\text{median}} \left(\frac{x_j - x_i}{t_j - t_i} \right) \tag{2}$$

$$\hat{\alpha}_0 = \underset{i=1,\ldots,n}{\text{median}}\, x_i - \hat{\alpha}_1 \underset{i=1,\ldots,n}{\text{median}}\, t_i \tag{3}$$

where x_k is the recorded value of an annual maximum flow in year t_k, $Q_{max}(t_k)$, and n is the sample size. The $\hat{\alpha}_1$ coefficient (2), called the Theil slope (Helsel & Hirsch 2002) or Theil-Sen slope (Su *et al.* 2018) is a robust estimate of the magnitude of a trend.

The null hypothesis in the Mann-Kendall test is that no monotonic trend exists (which means, in particular, that α_1=0) or, in other words, that the data are independent and randomly ordered. However, if positive autocorrelation in the data exists, it increases the probability of detecting nonexistent trends (Hamed and Rao 1998). A remedy proposed by Hamed and Rao (1998) is to use the effective sample size as a correction to the variance of the Mann-Kendall trend test statistic.

The Mann-Kendall trend test statistic S is usually defined as

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_j - x_i) \tag{4}$$

where

$$\operatorname{sgn}(x) = \begin{cases} +1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases} \tag{5}$$

The variance, var(S), of the statistic S is

$$\operatorname{var}(S) = n(n-1)(2n+5)/18 \tag{6}$$

Hamed and Rao (1998) corrected the variance var(S) for autocorrelation:

$$\operatorname{var}^*(S) = \operatorname{var}(S) \cdot (n/n^*) \tag{7}$$

where the correction factor n/n^* equals

$$\frac{n}{n^*} = 1 + \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n-1} (n-i)(n-i-1)(n-i-2)\rho_S(i) \tag{8}$$

where $\rho_S(i)$ is the autocorrelation function of the ranks of the data. If the lag-i autocorrelations are positive/negative, then it follows from (8) that the corrected variance var*(S) of statistic S is larger/smaller than var(S), i.e., when no autocorrelation exists. Hamed and Rao (1998) suggested that only significant $\rho_S(i)$ values be taken into account. In the paper, significance level of 5% was adopted. Also, the lag-1 autoregression model was assumed. In this case $\rho_S(i) = \rho_S(1)^i$, i =1,2,..., and the correction factor (8) is

$$\frac{n}{n^*} = 1 + 2\frac{r_1^{n+1} - nr_1^2 + (n-1)r_1}{n(r_1-1)^2} \tag{9}$$

(Matalas and Langbein 1962), where r_1 is the lag-1 rank autocorrelation coefficient estimate of $\rho_S(1)$. Zetterqvist (1991) and Hamed and Rao (1998) suggest that the values of autocorrelation coefficients must be calculated after subtracting a suitable nonparametric trend estimator from the sample. This suggestion was adopted in the paper.

The whole procedure of trend estimation and testing was performed in the following steps. For each time series nonparametric linear regression coefficient (2) and (3) were calculated. Next, a each time series was detrended:

$$x'_t = x_t - \hat{\alpha}_1 t \tag{10}$$

Then, lag-1 rank autocorrelation coefficient, $\hat{\rho}_S(1) = r_1$, for the detrended series was calculated as the lag-1 Kendall rank correlation coefficient τ:

$$r_1 = \tau = \frac{2}{n'(n'-1)} \sum_{i=1}^{n'-2} \sum_{j=i+1}^{n'-1} \operatorname{sgn}(x'_j - x'_i) \operatorname{sgn}(x'_{j+1} - x'_{i+1}), \quad n' = n - 1 \tag{11}$$

The variance of the (11) may be estimated as

$$\operatorname{var} \tau = \frac{2(2n'+5)}{9n'(n'-1)} \tag{12}$$

(Abdi 2007). If *n'* exceeds 10, variable Z_τ

$$Z_\tau = \tau / \sqrt{\operatorname{var} \tau} \tag{13}$$

is approximately N(0,1) distributed (Abdi 2007), and the p-value for hypothesis H_0: $\rho_S(1) = 0$ may be calculated as for the two-sided test:

$$p_V = 2\left[1 - \Phi(z_\tau)\right] \tag{14}$$

where $\Phi(z)$ is the standard normal cumulative distribution function. If $p_V < 5\%$, r_1 is significant and the variance of the Mann-Kendall statistic *S* is computed according to equation (7) with correction factor (9). If $p_V > 5\%$, equation (6) is applied. In both cases, standardized statistic *S* is calculated:

$$Z_S = \left[S - \operatorname{sgn}(S)\right] / \sigma_S \tag{15}$$

where σ_S equals either $[\text{var*}(S)]^{1/2}$ or $[\text{var}(S)]^{1/2}$. For *n* exceeding 10, variable Z_S is approximately N(0,1) distributed, so Mann-Kendall test suggest trend existence when $p_V = 2[1-\Phi(z_S)]$ is less than 5%. In order to study the influence of autocorrelation on the final result of the Mann-Kendall test, also not variance-corrected version of the test was performed.

RESULTS AND DISCUSSION

For all 138 time series of Q_{max}, linear trend slope $\hat{\alpha}_1$ (2) was calculated (Fig. 5a), then the series were detrended and the lag-1 Kendall rank autocorrelation coefficient was computed (Fig. 5b). Out of 128 coefficients, 22 were significant at 5% level and, therefore, were used for calculating the variance correction factor (9). Only two r_1-s increased var*S* so much that it made trends non-significant (cf. Fig. 5). In the remaining 106 time series, the variance was calculated by equation (6). Final results of the test comprise 22 significant (9 increasing and 13 decreasing) and 106 not significant trends.

Apart from absolute trend slopes $\hat{\alpha}_1$ also relative trend slopes $\hat{\alpha}'_1 = \hat{\alpha}_1 / \bar{Q}_{max}$ (Fig. 5c) were calculated because they allow to better compare the magnitude of slopes within the catchment, which may reveal some kind of spatial clustering. To this end, spatial distribution of slope magnitudes of divided into 8 classes is shown in Fig. 6, while Fig. 7 shows spatial distribution of significant and not significant linear trends.

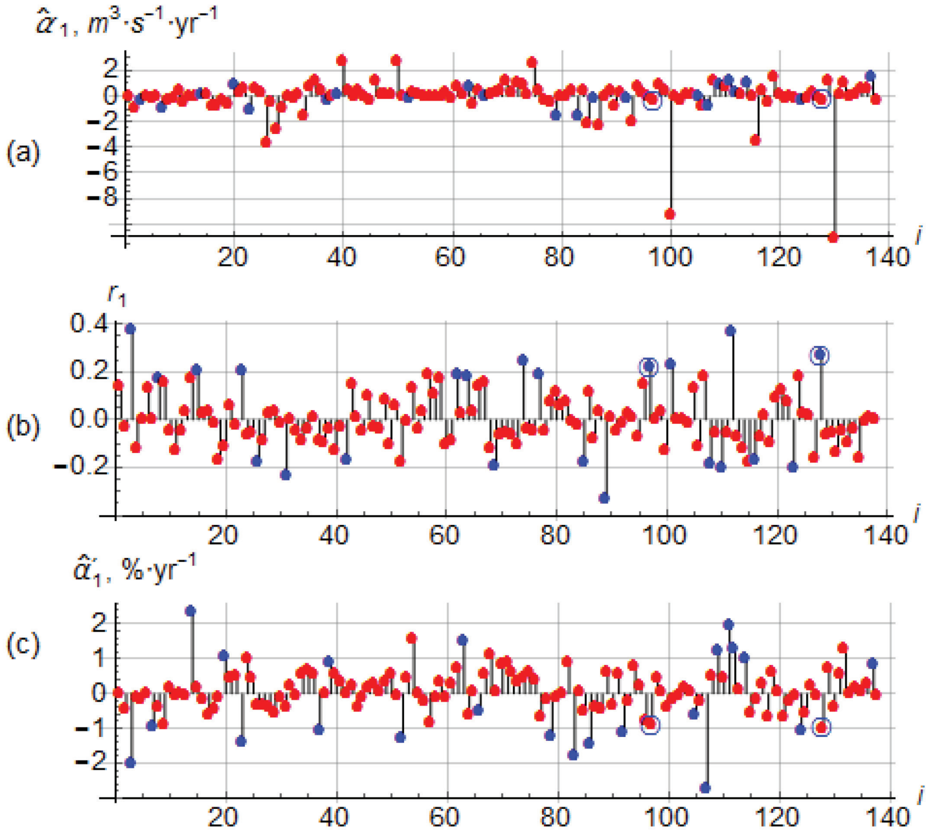

Figure 5. (a) absolute slopes α_1, (b) Kendall lag-1 rank autocorrelation coefficient, r_1, and (c) relative slopes $\hat{\alpha}'_1 = \hat{\alpha}_1 / \bar{Q}_{max}$ of 138 Q_{max} time series trends, ordered as in Fig. 1b. Blue/red dots denote significance/non-significance of α_1, r_1 and $\hat{\alpha}'_1$ at 5% level. Large circles show the cases when r_1 effectively increased var$S$ making trend non-significant.

Spatial distribution of trend magnitudes (Figs. 6 and 7) reveals a kind of clustering, suggesting dividing the catchment area into sub-areas with prevailing upward/downward trends.

In general, positive and negative slopes are grouped in the southern and northern halves of the catchment, respectively. Absolute values of most (ca. 80) relative slopes are smaller than 0.5%·yr[1]. As to larger positive slopes, the majority are situated in the mountain (i.e. the southernmost) parts of the catchment.

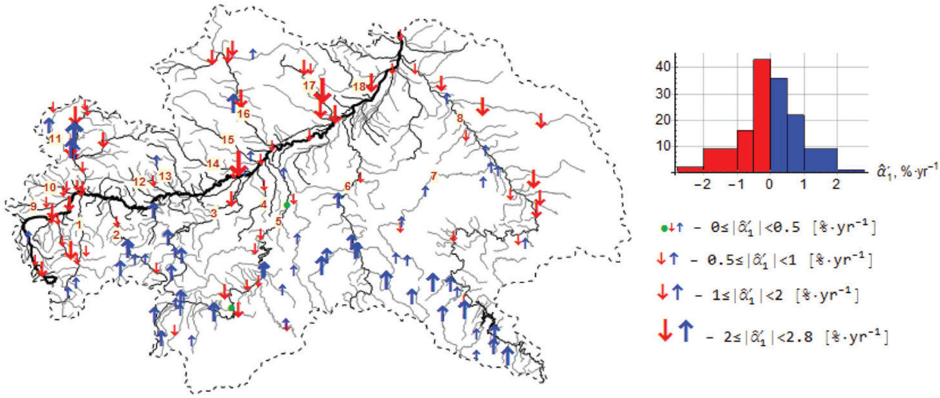

Figure 6. Spatial distribution of magnitude of 138 positive (↑) and negative (↓) relative slopes $\hat{\alpha}'_1$ divided into 8 classes (see histogram). Green dot (•) indicates two locations with estimated zero-slope.

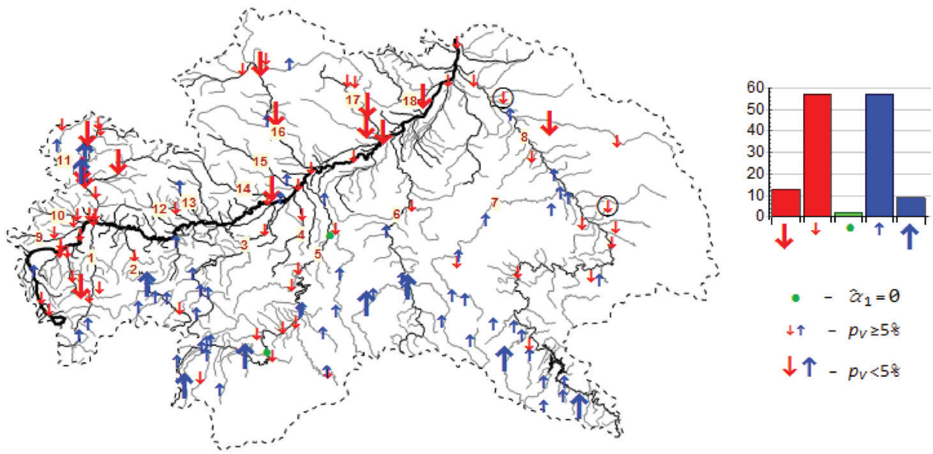

Figure 7. Increasing (↑) and decreasing (↓) significant and not significant trends in the Upper Vistula River catchment. Green dots (•) indicate two locations with estimated zero-slope. Histogram shows the size of each of 5 groups. Large circles are explained in Fig. 5.

Trends situated on the southern part of the Upper Vistula River catchment area, beginning from the Skawa River (no. 2 in Figs. 6 and 7) eastwards, including the upper San River (no. 8) sub-catchment, exhibit mainly increasing behaviour; seven of them are statistically significant. Some exceptions (decreasing not significant trends) are visible in the sub-catchments of Dunajec (river no. 5) and San (no. 8).

Second area, even more homogeneous than the first one, covers the northern and south-western parts of the catchment with exception of the north-western part. Almost all trends within this area are decreasing. Seven statistically significant downward trends, out of all of nine within the area, are situated north of the Vistula River.

Third homogeneous area, the north-western and the smallest sub-part, shows mixed and strong trend behaviour: as many as 6 trends are significant, 4 of them are downward, 2 upward, all in the catchment of the Przemsza (no. 11 in Fig. 6) River. A very probable explanation of such intense grouping of significant trends lies in the fact that this part of the catchment is under strong anthropogenic (mostly mining) influence (cf. Dulias 2016).

FINAL REMARKS AND CONCLUSIONS

The problem of trend existence in hydrometeorological time series is still important and interesting both from theoretical and practical point of view. No doubt that trends in annual maximum flows Q_{max} belong to this category. The Upper Vistula river catchment is an area with the very high flooding potential, which additionally increases the interest in the Q_{max} trend analysis.

To the analysis, all 138 recent (i.e., ending not earlier than in 2012) uninterrupted time series of annual maximum flows of size at least 30, recorded in the period 1951-2016 in the Upper Vistula River catchment were taken.

Because the Q_{max} time series exhibit considerable asymmetry, nonparametric Mann-Kendall trend tests was applied. However, if the autocorrelation exists in the series, a modification to the test should be introduced as the positive/negative autocorrelation increases/decreases the variance of the Kendall S statistic. In the paper, the Hamed and Rao (1998) variance correction for autocorrelation was used. Out of all detrended series, 22 showed significant autocorrelation at 5% level, and were used to modify the Mann-Kendall trend test. However, only two autocorrelation coefficients turned out to be efficient making the trend not significant. Finally, the test detected 22 trends significant at 5% level.

Spatial distribution of 138 trend slope magnitudes shows a very suggestive clustering of downward/upward values which allowed to divide the area into three parts differing in direction of trend and/or the number of statistically significant trends. In general, trends in the southern half of the Upper Vistula River catchment (excluding the westernmost part) are mostly upward (although few of them are significant), those in the northern half and the westernmost of the southern half part are downward and a considerable number is significant.

The north-western, relatively small, part of the catchment is a certain exception, as there coexist significant both upward and downward trends which may be ascribed to the strong anthropogenic transformations in this area.

ACKNOWLEDGMENTS

The Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB) is kindly acknowledged for providing the data.

REFERENCES

Abdi, H. (2007). *Kendall rank correlation*. In: Salkind, N.J. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage.

Brath, A., Castellarin, A., Montanari, A. (1999). *Detecting non stationarity in extreme rainfall data observed in Northern Italy*, in: Proceedings of EGS – Plinius Conference on Mediterranean Storms, Maratea, 219–231.

Burn, D.H., Hag Elnur, M.A. (2002). *Detection of hydrologic trends and variability*. J. Hydrol. 255: 107–122.

Burn, D. H., Whitfield, P.H. (2018). *Changes in flood events inferred from centennial length streamflow data records*, Advances in Water Resources 121: 333–349.

Cebulska, M., Twardosz, R., Cichocki J. (2007). *Zmiany rocznych sum opadów atmosferycznych w dorzeczu górnej Wisły w latach 1881 – 2030*, [w:] K. Piotrowicz, R. Twardosz (red.), Wahania klimatu w różnych skalach przestrzennych i czasowych, Instytut Geografii i Gospodarki Przestrzennej UJ, Kraków, 383 – 390.

Cebulska, M. (2015). *Wieloletnia zmienność maksymalnych opadów dobowych w Kotlinie Orawsko–Nowotarskiej* (*1984-2013*), Czasopismo Inżynierii Lądowej, Środowiska i Architektury, 62 (3/I): 49-60.

Chełmicki, W. (1991). *Położenie, podział i cechy dorzecza* in: Dynowska I., Maciejewski M. (red.) *Dorzecze górnej Wisły*. PWN, vol. I, Warszawa-Kraków.

Douglas, E.M., Vogel, R.M., Kroll, C.N. (2000). *Trends in floods and low flows in the United States: impact of spatial correlation*, Journal of Hydrology 240: 90–105.

Dulias R. (2016) *The Impact of Mining on the Landscape: A Study of the Upper Silesian Coal Basin in Poland*, Environmental Science and Engineering, Springer, doi: 10.1007/978-3-319-29541-1

Hamed, K.H., Rao, A.R. (1998). *A modified Mann-Kendall trend test for autocorrelated data*. J. Hydrol. 204 (1–4): 182–196.

Helsel D.R., Hirsch R.M (2002). *Statistical Methods in Water Resources*, U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, Chapter A3.

Kendon E.J., Roberts N. M., Fowler H. J., Roberts M. J., Chan S. C., Senior C.A. (2014). *Heavier summer downpours with climate change revealed by weather forecast resolution model*, Nature Climate Change 4: 570–576. doi:10.1038/nclimate2258.

Khaliq, M.N., Ouarda, T.B.M.J., Gachon, P. (2009). *Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers: The effect of short – and long-term persistence*, J. Hydrol. 369: 183–197.

Matalas, N. C., Langbein, W. B. (1962), *Information content of the mean*, J. Geophys. Res. 67(9): 3441–3448.

Meshram, S.G., Singh, V.P., Meshram, C. (2016). *Long-term trend and variability of precipitation in Chhattisgarh State, India*, Theor Appl Climatol. doi:10.1007/s00704-016-1804-z.

Prosdocimi, I., Kjeldsen, T. R., Svensson, C. (2014). *Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK*. Natural Hazards and Earth System Sciences 14: 1125-1144.

Razavi, T., Switzman, H., Arain, A., Coulibaly, P. (2016). *Regional climate change trends and uncertainty analysis using extreme indices: A case study of Hamilton*, Canada, Climate Risk Management 13: 43-63.

Sen, P.K. (1968). *Estimates of the Regression Coefficient Based on Kendall's Tau*, Journal of the American Statistical Association 63 (324): 1379-1389.

Svoboda, V., Hanel, M., Máca, P., Kyselý, J. (2016). Projected changes of rainfall event characteristics for the Czech Republic. Journal of Hydrology and Hydromechanics 64, doi: 10.1515/johh-2016-0036.

Theil, H. (1950). *A rank-invariant method of linear and polynomial regression analysis, Part I*, In the *Proceedings of the Royal Netherlands Academy of Sciences*, 53: 386-392.

Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. (2003). *The changing character of precipitation*, Bull. Am. Meteorol. Soc. 84: 1205–1217.

Wrzesiński, D. (2009). *Tendencje zmian przepływu rzek polski w drugiej połowie XX wieku*. Badania Fizjograficzne Seria A – Geografia Fizyczna 60: 147-162.

Yue, S., Pilon, P., Phinney, B., Cavadias, G. (2002). *The influence of autocorrelation on the ability to detect trend in hydrological series*. Hydrol Process. 16 (9): 1807-1829.

Zetterqvist, L. (1991). *Statistical estimation and interpretation of trends in water quality time series*, Water Resources Research 27 (7): 1637-1648.

Zhang, Q., Gu, X., Singh, V.P., Xiao, M., Xu, C-Y. (2014). *Stationarity of annual flood peaks during 1951-2010 in the Pearl River basin, China*. J. Hydrol. 519: 3263-3274.

Zhang, X., Vincent, L. A., Hogg, W. D., Niitsoo, A. (2000). *Temperature and precipitation trends in Canada during the 20th century*. Atmosphere–Ocean 38 (3): 395–429.

Corresponding author: Prof. Stanisław Węglarczyk, PhD. DSc
Mgr inż. Mateusz Całek, PhD student

Cracow University of Technology
Faculty of Environmental Engineering
ul. Warszawska 24
31-155 Kraków
e-mail: sweglar@pk.edu.pl

Received: 14 of January 2019
Accepted: 22 of February 2019