COMPARISON OF TESTS FOR TREND IN LOCATION AND SCALE PARAMETERS IN HYDROLOGICAL AND PRECIPITATION TIME SERIES

Agnieszka Rutkowska[1✉], Silvia Kohnová[2]

[1] Department of Applied Mathematics, Faculty of Environmental Engineering and Land Surveying, University of Agriculture in Krakow, Balicka 253C, Krakow

[2] Slovak University of Technology, Faculty of Civil Engineering, Department of Land and Water Resources Management, Radlinskeho 11, 810 05 Bratislava, Slovak Republic.

ABSTRACT

Aim of the study

The objective was the comparison of properties of tests for trend in location and scale parameters for hydrological and precipitation time series, in particular (i) to review the non-parametric tests known from literature for various trends, whereby the tests were studied from the point of view of their ability of detecting the existing trend (their power); (ii) to study the non-parametric tests for change in scale the power of which has not been estimated yet in application to hydrological and precipitation series (the Ansari-Bradley (AB), Siegel-Tukey (ST), Mood (M) tests); and (iii) to assess the differences between the tests.

Material and methods

The study was based on the series of pseudo-numbers and on realizations of historical hydrological and precipitation time series. The Monte Carlo simulations and comparison of properties of the tests (AB, ST, M) that have not been studied yet in application to hydrological and precipitation series were conducted. The tests known in literature from applications in hydrological and precipitation series methods were also compared.

Results and conclusions

Results show an increase of the power of the AB, ST and M tests with sample length and with strength of the step trend in scale, insensitivity to changes in the coefficient of variation, a low sensitivity to changes in type of the distribution, and the highest power of the M test. The step trend in scale in two exemplary series was identified. The AB, ST, and M tests can be applied to verify the hypothesis about step trend in scale in hydrological or precipitation time series.

Keywords: hydrological and precipitation time series, monotonic or step trend, the trend in location or scale, statistical test, power of the test

INTRODUCTION

Environmental processes exhibit alterations, the sources of which are natural or human-induced. Recognition of the changes is crucial, for example, to correctly understand climate change implications or anthropogenic activity and flood frequency analysis.

Suppose that a process is represented by a time series of random variables and that the type of their distributions remains unchanged. In that case, the changes can reflect alterations in characteristics such as (a) mean, median, quantiles, and (b) variance, standard deviation, and range. These changes are equivalent to changes in parameters of (a) location, and (b) scale.

✉e-mail: rmrutkow@cyf-kr.edu.pl

Changes in other characteristics are also possible but rarely considered because they are usually reflected in the changes mentioned above. Usually, two types of changes are considered: monotonic and step. Monotonic trends are usually observed in natural processes where a gradual evolution of environmental conditions implies incremental changes in characteristics, for example, the increase of the mean temperature within the last several dozen years due to climate change, low-frequency climate fluctuations, changes in channel morphology, etc. Step (abrupt) trends in river flows can result from a sudden change caused by anthropogenic activity such as dam construction, water abstraction, or a natural rapid episode, e.g., by a catastrophic hurricane. Another example is water chemistry. A step trend can be induced by the release of dioxin into the water reservoir.

Statistical non-parametric methods for identifying long-term trends in characteristics of the location and/or dispersion of the series of independent random variables from the field of hydrology are addressed in the paper. Monotonic and abrupt deterministic changes are considered. The paper's objective is to review and discuss existing methods for trend identification in the time series from the point of view of hydrology. The primary attention is paid to the ability of the tests to detect existing trends in such time series, which includes (i) discussion about this ability based on results from literature, (ii) assessment of this ability using Monte Carlo simulations for the tests for trend in scale that have not been studied yet in hydrology, and (iii) assessing differences between the tests. We assume that there are no seasonal fluctuations in the series. The distribution of each random variable is positively skewed, which is common in environmental time series. Thus, we do not deal with symmetric distributions, such as normal distribution, where the well-known parametric statistical methods can be applied.

Two types of errors can be made in statistical inference, the type I error, and the type II error. Robust tests are tests that maintain the type I and type II errors. The type I error is the probability of rejecting the null hypothesis if it is true. In contrast, the type II error is the probability of not rejecting the null hypothesis if the alternative is true. The maximum of the type I error is the size of the test, which is assumed equal to a fixed value – the significance level. In most statistical tests the assumption is fulfilled. However, a more serious concern is the high value of the type II error, leading to an erroneous conclusion from hypothesis testing. Usually, the test's power is considered instead, which equals 1-type II error and belongs to the interval [0, 1]. The power is the measure of the test's ability to detect a trend if the trend exists; the higher the power, the greater the chance of trend detection. This is evidence that the test's power should be known before its application. The powers of many tests for the trend are well known from the subject literature, and the results are shown in the next sections. However, for several tests for a sudden change in scale the power has not been studied yet, when applied to positively-skewed distributions from the field of hydrology. The power of these tests was estimated using the Monte Carlo simulations.

DATA

Calculations are made based on artificial time series of random numbers drawn from populations with various distribution functions. Details are given in section on Methods. Additionally, the daily discharges in the Nida River at Pińczów and the series of daily precipitation totals in Skoczów were obtained from the Institute of Meteorology and Water Management – National Research Institute, Poland (IMGW-PIB). Examples of the step trends in scale were shown in the series.

METHODS

We assume that a time series consists of random variables X_t, $t = 1, 2, ..., n$ where t is time. The distributions of X_t are of the same type, but their parameters can differ. Two parameters are assumed to have changed, namely the location parameter and/or the scale parameter. In most tests, the additional assumption about the lack of autocorrelation must hold; however, some tests also have their versions for autocorrelated variables. The tests are used to verify the following hypotheses:

H_0: No trend exists in the series $\{X_t, t = 1, 2, ..., n\}$,
H_1: There is a trend in the series $\{X_t, t = 1, 2, ..., n\}$.

The trend can be monotonic or step. This is the two-sided version of H_1. The one-sided versions of

H_1 are: there is an increasing/decreasing trend in the series $\{X_t, t = 1, 2, ..., n\}$. If H_1 is true, then the characteristics and the parameters of X_t must differ. The significance level was fixed at $\alpha = 0.05$.

The following groups of non-parametric tests were considered:

Group 1 Tests for monotonic trends in the parameter of location: the Mann-Kendall (MK) test, the test based on the Spearman's rho (S), and the Cox-Stuart (CS) test (Mann 1945, Kendall 1938, Cox and Stuart 1955). The MK test was used in a huge number of studies, examples of which are Hirsch et al. (1982), Młyński et al. (2019), Nowak et al. (2019), Jeneiova et al. (2014), Kohnová et al. (2018), Meresa et al. (2017). The examples for the CS test in hydrology are Fatichi et al. (2009), Sen and Niedzielski (2010), Niedzielski and Kosek (2011), Rutkowska (2015).

Group 2 Tests for step trend in the parameter of location: the Pettitt (P) test, the CUSUM test, the test based on the Cramer von Mises method (CM) (Pettitt, 1979, Page, 1954, Woodyer, 1974, Anderson, 1962, Holmes et al. 2013). The P test was applied to streamflow and sediment discharge in the Yellow River (Gao et al. 2010), to precipitation totals in Western Hungary (Kocsis et al., 2020) and in Taiwan (Yu et al., 2006), while the CUSUM test was applied to rainfall in the middle region of the Yellow River (Baddoo et al., 2015). The CM test was recommended to detect an abrupt change in annual maximum discharges (Guidance, 2017). The time point of change should be known in advance in each test. If not, the maximum value of the test statistic for various divisions of the series is computed.

Group 3 Tests for step trend in the parameter of scale (heteroscedasticity between two groups): the Siegel-Tukey (ST) test (Siegel and Tukey 1960), the Ansari-Bradley (AB) test (Ansari and Bradley, 1960), and the Mood (M) test (Mood, 1954). The AB test was used by Fleming and Weber (2012) in studies on inflow volume to reservoirs in Canada. The ST test was considered by McCuen (2003). It was used to assess step-change in river flows after dam construction and after catchment's urbanization. The M test was recommended to detect an abrupt change in variance in maximum annual flows (Guidance, 2017).

Group 4 Tests for step trends both in location and scale parameters: the Cucconi (C) test (Cucconi, 1968; Marozzi, 2012), and the Lepage (L) test. The L test was used by Zhang et al. (2009) in studies on the variability of water resources in the Yellow River, while the C test was applied to river flows by Rutkowska and Banasik (2016) to several series of annual river discharges from Poland, USA, and Canada.

It should be emphasized that the tests in Groups 1–4 do not cover the whole range of various tests for trend detection (Kundzewicz and Robson, 2000, 2004; Kundzewicz and Radziejewski, 2006). However, the tests most frequently used in hydrology were selected in 1–4.

The power of most tests from groups 1, 2, and 4 is known from the literature. Results are presented in the next section. However, to the authors' knowledge, the power of the tests from Group 3 (step trend in scale) has not yet been studied when applied to time series from the field of hydrology and meteorology. The formulas for Group 3 are shown in Appendix A. These tests assume that the medians in the two groups are similar. If the assumption is not fulfilled, the median adjustment should be carried out before testing.

To estimate the power of the tests, the typical positively-skewed distribution functions of river discharges and precipitation totals were selected first, namely the Pearson III (PE3), the two-parameter lognormal (LOGN), and the Generalized Extreme Value type I (Gumbel, GUM) and type II (GEV). The power was estimated using the Monte Carlo simulation method. Without loss of generality, each distribution in the simulation was assumed to have the expected value equal to 1 because the tests are invariant to the transformation $\frac{X_t}{EX_t}$ where EX_t is the expected value of X_t. The assumption was made to reduce the number of cases in simulations and consider only various dispersions of the PE3, LOGN, and GUM distributions and various dispersions and shape parameters of the GEV distribution. A range of dispersions and shape parameters that most frequently occur in such series were considered. The number of simulations was $N = 5 \cdot 10^3$. The simulations were conducted in the following steps: (i) generation of random numbers of length $n = 20$ (20) 100 representing samples from various distribution functions with the coefficient of variation CV varying from 0.1 to 1.2, (ii) trend superimposition on the second part of the series, i.e., from $k = \frac{n}{2}$ to n or from

$k = \dfrac{n}{4}$ to n where k is the time point of change, (iii) testing for a trend for two-sided H_1. Additionally, the shape parameter of the GEV distribution equal to 0.2 (GEV02) and 0.4 (GEV04) was assumed in (i). The CV values from 0.1 to 1.2 in (i) were selected because such values are usually observed in river discharges over the world as the median of the CV is around 0.3 in North America, and about 0.2 in Europe and Asia (McMahon 1982, Yevjevich 1963), and in precipitation totals over central, northern and eastern Europe (Dankers and Hiederer 2008), e.g., in Poland (Ziernicka-Wojtaszek and Kopcińska, 2020), and, additionally, the CV can be increased or decreased by a trend. Trends with various strengths were superimposed on the series in (ii). The strength is reflected in the λ coefficient in the model (1) in Appendix B, where $\lambda \in (0, +\infty)$. If λ is very low or very high, then the strength of the step trend is high, and if λ is not far from 1, then the strength is low. Finally, the power of the test was estimated using the formula $\dfrac{N_{rej}}{N}$ where N_{rej} was the number of the series for which H_0 was rejected. It should be mentioned that if $\lambda = 1$ (no trend), then $\dfrac{N_{rej}}{N}$ is just the size estimate. More details and formulas used in the simulation are given in Appendix B. All calculations were carried out in R (R Core Team 2020, Stephenson 2002).

RESULTS OF THE POWER STUDIES FROM LITERATURE

Results of the power study of the MK test completed by Yue et al. (2002a, 2002b, 2003) showed that the power of the MK test is an increasing function of the absolute slope of the trend and the sample size n, and is a decreasing function of the coefficient of variation CV. If $n = 50$, then the power varies from 0.05 to 0.7 for the slope less than 0.01 (weak and moderate slope) for PE3 distribution with $CV = 0.5$ (moderate CV). If $n = 100$, then the power is much higher; it exceeds 0.8 for a weak slope. If $n = 20$, then the MK test's power is very low even for not very low slopes and moderate CVs. The S test's power is lower than that of the MK test for various slopes, sample sizes, and distribution types (Yue et al. 2002a). It is larger for the PE3 distribution than for the LOGN and GEV distri-

butions. Önöz and Bayazit (2003) stated that the MK test's power for the LOGN distribution is comparable to the other distributions' powers for various slopes and sample sizes.

The CS test study showed similar properties to the MK test, namely the increasing power with sample size and slope and the decreasing power with increasing CV (Rutkowska, 2015). Usually, the CS test has slightly lower power than the MK test for various distribution types. For the small sample size, $n = 20$, the CS test is somewhat more powerful than the MK test.

Studies on the P test's power for step trends in the location parameter were based on the PE3 distribution function in application to hydro-climatological series (Mallakpour and Villarini, 2016). They showed that the power increases with sample size and with the magnitude of change. It is higher when abrupt changes occur near the centers of the series rather than near the ends. Extensive comparative analysis of the CUSUM, P, and CM tests for a sudden change in hydrological time series was carried out by Zhou et al. (2019). It was found that the P and CM tests have better power properties than the CUSUM test. Both the P and CM tests have very similar power apart from heavy-tailed GEV distributions when the CM performs better.

Studies on the L and C tests for step trend both in location and scale showed that the tests are comparable in power with a slight predominance of the C test for short series with low variability.

RESULTS OF THE SIMULATION POWER STUDY FOR TESTS FOR THE TREND IN SCALE

For the need of presenting the results concisely, some rules were introduced, namely (i) the power p is low if $p < 0.3$, moderate if $0.3 \leq p < 0.7$ and high if $p \geq 0.7$; (ii) the variability of river flows is low if $CV < 0.1$, moderate if $0.1 \leq CV < 0.6$, and high if $CV \geq 0.6$; and (iii) the strength of the step trend is low if $\lambda = 0.7, 1.2, 1.4$, moderate if $\lambda = 0.5, 1.6, 1.8, 2$, and high if $\lambda = 2.2, 2.4, 2.6$.

The results of the ST, AB, and M tests' power study showed that the ability of the tests to detect a step trend in scale depends on sample size to the highest degree. This is visible in Fig. 1, where the AB test's power for sample size 20, 40, …, 100 was shown for var-

ious distribution functions with moderate variability ($CV = 0.3$), assuming that $\lambda = 2$. The power of the AB test is low if $n = 20$, moderate if $n = 40$ and $n = 60$, and high if $n = 80$ and $n = 100$. The difference in power between distributions is not high because the maximum difference is 0.18.
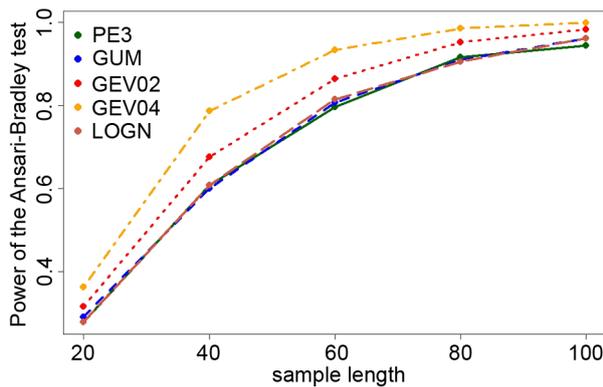

Fig. 1. The power of the Ansari-Bradley test as a function of sample length. The power is shown for various distribution functions of variables in the time series, assuming that $CV = 0.3$, $\lambda = 2$, and $k = n/2$

The power of each test also depends on the strength of the trend, reflected in the λ coefficient. Values of λ close to 1 make the tests nearly unable to detect an existing trend. On the other hand, all tests can detect a trend with ease, if the strength is moderate or high. The power of the Mood test for various λ's and sample length $n = 50$ is shown in Fig. 2. It can be observed that the power is high for moderate or high strength of the step trend and that the power quickly decreases if the strength is low. It can also be seen in Fig. 2 that the size of the M test ($\lambda = 1$) is always very close but not higher than 0.05, which confirms good size properties of the M test even for various distribution functions. Similarly, the AB and ST tests also have good size properties.

Another characteristic that might have an impact on power is the coefficient of variation CV. Results of the simulation study showed, however, that the power is not sensitive to changes of the CV. The power of the ST test for various CVs, assuming that the variables follow the GUM distribution function, were depicted in Fig. 3. It can be observed that the power can remain moderate or high even for high CV only if the strength

of the trend is sufficiently high ($\lambda \in \{0.5, 2, 2.5\}$). However, if the strength is low ($\lambda \in \{0.8, 1.2\}$), the tests are nearly unable to detect an existing trend.

The power depends on the time point k of a jump because all tests can detect trend more clearly if $k = \frac{n}{2}$ than if $k = \frac{n}{4}$, e.g., the power of the AB, M, and ST test is equal to 0.89/0.66 (i.e., 0.89 for $k = \frac{n}{2}$ and 0.66 for $k = \frac{n}{4}$), 0.91/0.63, and 0.88/0.66 for the GEV04 and 0.73/0.56, 0.81/0.61, and 0.73/0.57 for the GUM, respectively, assuming that $\lambda = 2$, $CV = 0.3$ and $n = 50$.
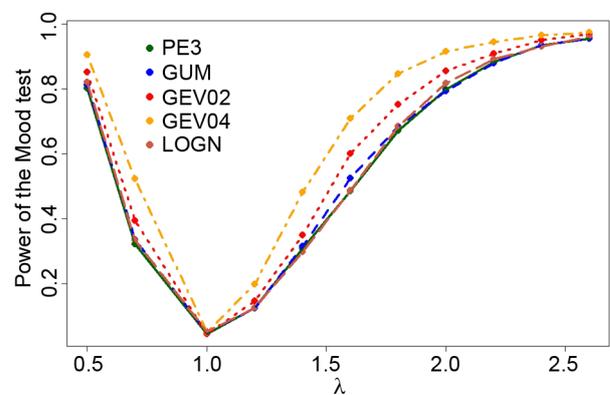
Fig. 2. The power of the Mood test as a function of the λ parameter for various distribution functions assuming that $CV = 0.3$, $n = 50$, and $k = n/2$
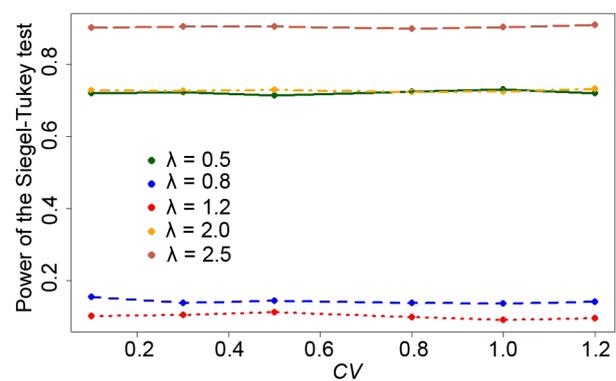
Fig. 3. The power of the ST test as a function of the coefficient of variation CV of the series before change for various λ values assuming that the distribution follows the GUM and that $k = n/2$. The power is low if the strength of the trend is low ($\lambda = 0.8$ or $\lambda = 1.2$) but is high otherwise ($\lambda = 0.5$ or $\lambda = 2$ or $\lambda = 2.5$)

Results showed that the power slightly differs for various distribution functions (PE3, LOGN, GUM, and GEV). Usually, the highest power was obtained for the GUM and GEV distributions and the lowest for the LOGN distribution. The higher power for the GEV is visible, for example, in Fig. 1 and Fig. 2. The absolute differences between the ST, AB, and M powers for various distributions vary from 0 to 0.2.

While remembering that the two-sided hypothesis H_1 was considered in this study, it should be noted that the powers of the tests can be higher if H_1 is one-sided.

The comparison of the ST, AB, and M tests showed that, although the three tests' powers are comparable, the best power properties were observed for the M test for all n, CV, and λ values. Therefore, the use of the M test is recommended in testing the hypothesis of a step trend.

EXAMPLES OF HYDROLOGICAL SERIES WITH STEP TREND IN SCALE

The series of maximum daily precipitation totals in February in Skoczów, and the series of mean winter daily discharges in the Nida River at Pińczów were tested for the step trend in scale for various time points of change. All points of change for which the AB test accepted the alternative hypothesis were selected. The lowest p-value indicated the final point of change. Additionally, the ST and M tests were performed for the same point of change. All subseries (after and before the change) had been median-adjusted before the tests were performed. The estimate of the λ parameter was the ratio of the standard deviation in the second and first subsample. The three tests detected the decreasing step trends in the series of maximum daily precipitation totals in February and in the series of mean winter river discharges. The approximate year of change, the p-values of the tests at the year of change, the CV before change, the estimate of the λ parameter, and the type of the trend were shown in Table 1. It can be observed that the CV values fall within the range from 0.1 to 1.2 that had been considered in the Monte Carlo simulation.

In Fig. 4 the step trend in scale was depicted in time series plots of the maximum daily precipitation totals in February in Skoczów (see: Fig. 4a), and for the mean winter discharges in the Nida River at Pińczów (see: Fig. 4b). The trend in scale is reflected in the decrease of standard deviation (or in variance) in (a) and (b). The ratio of the standard deviation (after and before the change) is the estimate of the λ parameter. Both series were also tested for monotonic and step trend in the parameter of location using the Mann-Kendall test and the Pettit test, respectively. However, no trend in location parameter was found.

The possible cause of the decreasing trend in the Nida river was the river training works undertaken in the 1980s that changed hydrological regimes in the catchment (Strużyński 2011). In the 2000s, several actions were undertaken to restore the Middle delta Nida river's unique ecosystems. An example is the project Life4Delta_PL (2019–2024). The cause of the decrease in the maximum daily precipitation total in the winter month of February might be ascribed to the impact of climate change, but this hypothesis needs further studies.

Table 1. Results of the AB, ST, and M tests for the two time series

Data	Years	p-values of the two-sided AB, ST, and M tests	Year of the change	Coefficient of variation CV, before the change	Estimate of the λ parameter[1]	Trend
Maximum daily rainfall total in February at the Skoczów station	1951–2019 (calendar years)	pAB = 0.016 pST = 0.015 pM = 0.015	1989	0.50	0.71	decreasing
Mean winter flow in the Nida River at Pińczów	1951–2019 (hydrological years)	pAB = 0.003 pST = 0.003 pM = 0.004	1986	0.37	0.59	decreasing

[1] The estimate of the parameter is the ratio of the standard deviation of the second and first subsample (after and before the change).
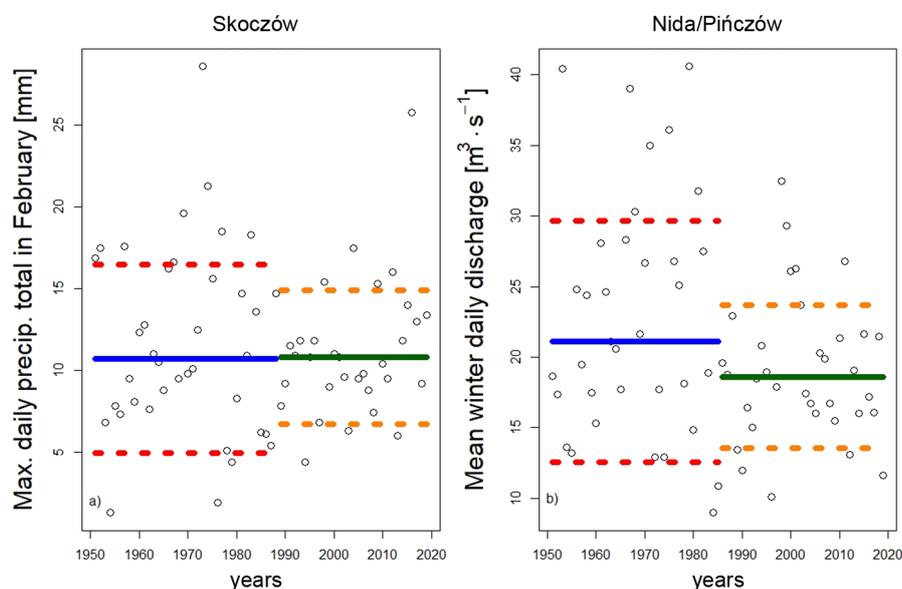
Fig. 4. The time series plots of (a) the maximum daily precipitation total (February) in Skoczów, and (b) the winter river flows in the Nida River at Pińczów. The continuous blue and the dotted red lines and the continuous green and the dotted orange lines reflect the median value and median+/-standard deviation values before and after the change, respectively. The decrease in the standard deviation is apparent in (a) and (b)

DISCUSSION

All groups of tests 1–4 can be complemented by other methods. However, a careful consideration of test assumptions should be undertaken before applying the given method. For example, the classical t-test and the F-test can also be included in the Group 2 and Group 3, respectively, but the tests require normality that is only seldom valid for hydrological variables. Another issue is that some methods are more general in the sense that they can detect a change in the distribution, like for example the sign test and Kruskall-Wallis test. These two tests can be included in all four groups because change in the distribution can be reflected in changes in mean, quantile, variance, etc. It is worth noting that the tests for a change in distribution often have weaker power properties than the tests from the Groups 1–4. Therefore, only the tests for changes in a specific parameter were considered in this paper. Sometimes, because various statistical methods complement each other, they should be simultaneously applied.

The ability to detect trend in time series of asymmetric variables, a weak sensitivity to the type of the distribution, and the good size properties are the main advantages of the tests from the Groups 1–4.

One has to pay attention to assumption of independence of random variables in the time series for the tests from Groups 1–4. If the variables were serially dependent, the tests might not be able to detect trend even if the trend really existed. Moreover, the tests from the Group 1 are inefficient in detecting a non-monotonic trend, for example an oscillating behavior. Regarding Groups 2–4, the tests are unable to detect several abrupt changes in the location or scale parameter. These are limitations of the tests.

Lowenstein (2015) studied the robustness of the M and ST tests. Various distributions from the field of social and behavioral science with artificial data from psychology and education were considered there, with only a few asymmetric data sets (8.6%). Good power properties of the tests for large sample sizes were confirmed for most cases, with some exceptions for asymmetric variables.

In this study, the M test was proved to outperform the AB and ST tests as regards power. The property was valid for various asymmetric distributions. This result

coincides with results of Lemeshko et al. (2010) where a family of symmetric distributions was considered.

The test statistics of the tests from the Group 2 and Group 3 can be also used to construct confidence intervals for the parameter they concern. One example is found in the paper by Abd-Elfattah (2013) who obtained confidence intervals for the dispersion parameter through the inversion of the M and ST tests, assuming that the variable of interest follows a symmetric (normal, Laplace, uniform) or an asymmetric Extreme Value distribution.

It follows from the literature overview that the tests studied in this paper can be used for various distributions and that they can have also other applications, such as in constructing confidence intervals.

CONCLUSIONS

The conclusion can be drawn from the literature review that the MK test has the best power properties among Group 1 tests (i.e. tests for a monotonic trend in the location parameter). However, for a small sample size, the CS test should also be considered. Among the tests from Group 2 (i.e. tests for a step trend in the parameter of location), the CM and P tests can be recommended. In contrast, among the tests from Group 4 (i.e. tests for a step trend both in location and scale), both L and C tests have comparable power properties.

The conclusion can be drawn from the results of the Monte Carlo simulation study of tests from Group 3 (i.e. tests for a step trend in the parameter of scale) that the AB, ST, and M tests have high power only if $n \geq 60$, and when the strength λ is moderate or high. The type of distribution and the CV does not influence the power of the tests. The step trend is more easily detectable if the point of change is in the middle rather than near the end of the series. Among the three tests: the AB, ST, and M test, the M test is the best choice because of its best power properties and because the test has a simple formula for the test statistic.

APPENDIX A

We assume that $X = (x_1, x_2, ..., x_m)$ and $Y = (y_1, y_2, ..., y_n)$ are samples of sizes m and n of independent observations from populations with continuous cumulative distribution functions F and G, respectively. The medians of F and G are identical (if the assumption is not valid, then the medians should be subtracted from the samples). The F and G are of the same type but differ at most by the scale parameter θ, i.e., $G(u) = F(\theta u)$ where $\theta > 0$. The hypotheses for testing the difference in scale are:

H_0: $\theta = 1$, H_1: $\theta \neq 1$ (or $\theta > 1$ or $\theta < 1$).

To perform the Ansari-Bradley (AB) test, the new sample $Z = X \cup Y$, $Z = (z_1, z_2, ..., z_{m+n})$, should be ranked in increasing or decreasing order. The AB statistic equals $WA = \sum_{i=1}^{p} i\delta_i + \sum_{i=p+1}^{m+n} (m+n+1-i)\delta_i$ where $\delta_i = 1$ if $z_i \in X$ and $\delta_i = 0$ otherwise, and where p is the largest integer not more extensive than $\frac{m+n+1}{2}$, i.e., $p = \left[\frac{m+n+1}{2}\right]$ (Ansari and Bradley, 1960; Bauer, 1972).

The test statistic of the Mood square rank test equals $WM = \sum_{i=1}^{n} \left(r_i - \frac{m+n+1}{2}\right)^2$ where r_i is the rank of the i th observation from the Y sample in the ordered sample $Z = X \cup Y$ (Mood 1954).

In the Siegel-Tukey test, the center point is selected in the ordered sample $Z = X \cup Y$. The center point is the median if $m + n$ is odd (case (i)) and $\frac{m+n}{2}$ or $\frac{m+n}{2} + 1$ if $m + n$ is even (case (ii)). Then the center point gets a score of 1. Next, one of the adjacent elements gets a score of 2, two of the next elements from the other side get scores 3 and 4, and the next two elements from the previous side get 5 and 6, etc. (case (i)) or scores 2 and 3 are given to two elements nearest to the center from one side, then 4 and 5 to the elements nearest to the center from the other side, etc. (case (ii)).

Critical values of the three tests are given in Bauer (1972), Mood (1954), McCuen (2003).

APPENDIX B

We assume that a time series consist of n independent random variables under consideration, e.g., river discharges, precipitation totals, etc, We also assume that the mean value of all such variables is a and that the standard deviation changed at the time point k but the type of the distribution remained unchanged. Thus

each variable can be decomposed into the sum $Z_t + a$ where $EZ_t = 0$ for $t = 1, 2, ..., n$, and the standard deviation of Z_t also changed at the time point k. Additionally, Z_t can be written as

$$Z_t = \begin{cases} X_t - a & \text{for } t = 1, ..., k-1 \\ \lambda(X_t - a) & \text{for } t = k, ..., n \end{cases} \qquad (1)$$

where λ is a fixed, positive value while X_t are independent and identically distributed random variables such that $EX_t = a$ for $t = 1, 2, ..., n$. If $\lambda > 1$, then the standard deviation of Z_t increases, and if $\lambda < 1$, then it decreases at time point k. It should also be noted that the relative change of the variance of Z_t at time point k is $\dfrac{D^2 Z_k - D^2 Z_{k-1}}{D^2 Z_{k-1}} = \lambda^2 - 1$.

For the need of the power analysis, the model $Y_t = 1 + \dfrac{1}{a} Z_t$ was considered in the Monte Carlo simulation. It should be pointed out that the relative change of the variance of Y_t at time point k also equals $\lambda^2 - 1$. Moreover, the type of distribution of Y_t is the same as that of Z_t. In this sense, the model for Y_t is equivalent to the model (1). Additionally, because $EY_t = 1$ for $t = 1, 2, ..., n$, we can generate random numbers from a distribution with an expected value equal to 1. In this way, the model for Y_t is independent of the mean value of the variable under consideration.

ACKNOWLEDGMENTS

REFERENCES

Abd-Elfattah, E.F. Saddlepoint approximations for tests of dispersion. Comput Stat 28, 289–306 (2013). DOI: https://doi.org/10.1007/s00180-011-0299-4.

Ansari, A. R., Bradley, R. A. (1960). Rank-sum tests for dispersions. Ann. Math. Stat. 31, 1174–1189.

Baddoo, T., Guan, Y., Zhang, D., Andam-Akorful, S.A. (2015). Rainfall Variability in the Huangfuchuang Watershed and Its Relationship with ENSO. Water 7 (7) 3243–3262, DOI: https://doi.org/10.3390/w7073243.

Bauer, D. F., 1972. Confidence Sets Using Rank Statistics. Journal of the American Statistical Association, 67(339), 687–690.

Cox, D. R., Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. Biometrica 42, 80–95.

Cucconi, O. (1968). Un nuovo test non parametrico per il confronto tra due gruppi campionari. Giornale degli Economisti. XXVII: 225–248.

Dankers, R., Hiederer, R. (2008). Extreme Temperatures and Precipitation in Europe: Analysis of High-Resolution Climate Change Scenario, Report of European Commission Joint Research Centre, Institute for Environment and Sustainability, http://publications.jrc.ec.europa.eu/repository/bitstream/JRC44124/eur_23291_en.pdf

Fatichi, S., Barbosa, S. M., Caporali, E., Silva, M. E. (2009). Deterministic versus stochastic trends: Detection and challenges. J. Geophys. Res., 114 (1–2), 80–95.

Fleming, S.W., Weber, F.A. (2012). Detection of long-term change in hydroelectric reservoir inflows: Bridging theory and practise, J. Hydrol., 470–471, 36-54, DOI: https://doi.org/10.1016/j.jhydrol.2012.08.008.

Gao, P., Zhang, X.-C., Mu, X.-M., Wang, F., Li, R., Zhang, X. (2010). Trend and change-point analyses of streamflow and sediment discharge in the Yellow River during 1950–2005. Hydrol. Sci. J. 55(2), 275–285.

Guidance for detection of nonstationarities in annual maximum discharges, 2017. Department of the Army, U.S. Army Corps of Engineers, Washington, DC 20314-1000, 2017.

Hirsch, R. M., Slack, J. R., Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18, 107– 121.

Jeneiová, K., Sabo, M., Kohnová, S. (2014). Trend detection in long term maximum annual discharge series in catchments of Slovakia. Acta Hydrol. Slov. 15, 161–170.

Kendall M.G. (1938). A new measure of rank correlation. Biometrika 30, 81–93.

Kocsis, T., Kovács-Székely, I., Anda, A. (2020). Homogeneity tests and non-parametric analyses of tendencies in precipitation time series in Keszthely, Western Hungary. Theor. Appl. Climatol. 139 (3), 849–859. DOI: https://doi.org/10.1007/s00704-019-03014-4.

Kohnová, S., Vasilaki, M., Hanel, M., Szolgay, J., Hlavčová, Loukas, A., Földes, G. (2018). Detection of future changes in trends and scaling exponents in extreme short-term rainfall at selected stations in Slovakia. Contrib. to Geoph. Geodesy, 48/3, 207–230.

Kundzewicz, Z.W., Robson, A. (Editors) (2000). Detecting Trend and Other Changes in

Hydrological Data. World Climate Program – Water, WMO/ UNESCO, WCDMP-45, WMO/TD 1013, Geneva.

Kundzewicz, Z.W., Robson, A. (2004). Change detection in hydrological records – a review of the methodology, / Revue méthodologique de la détection de changements dans les chroniques hydrologiques. Hydrolog. Sci. J., 49(1), 7–19. DOI: 10.1623/hysj.49.1.7.53993.

Kundzewicz, Z.W, Radziejewski, M. (2006). Methodologies for trend detection. Climate Variability and Change – Hydrological Impacts. Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006, IAHS Publ. 308.

Lepage, Y. (1971). A combination of Wilcoxon's and Ansari–Bradley's statistics. Biometrika 58, 213–217.

Mann, H.B. (1945). Non-parametric test against trend. Econometrica 13, 245–259.

Lemeshko, B.Y., Lemeshko, S.B. & Gorbunova, A.A. Application and power of criteria for testing the homogeneity of variances. Part II. Nonparametric criteria. Meas Tech 53, 476–486 (2010). https://doi.org/10.1007/s11018-010-9530-x.

Life4Delta_PL, 2019-2024. LIFE17 NAT/PL/000018 – Re-naturalisation of inland delta of Nida River. https://life4delta.pl/ (accessed 08.10.2020).

Lowenstein, L. C., Robustness And Power Comparison Of The Mood-Westenberg And Siegel-Tukey Tests (2015). Wayne State University Dissertations. Paper 1151.

Marozzi, M. (2012). A modified Cucconi test for location and scale change alternatives. Colombian J. Stat. 35, 369–382.

McCuen, R. H. (2003). Modeling Hydrologic Change. Boca Raton. FL: Lewis Publishers.

McMahon, T. A. (1982). Hydrological Characteristics of Selected Rivers of the World. Paris: Unesco.

Meresa, H.K., Romanowicz, R.J. Napiórkowski, J.J. (2017). Understanding changes and trends in projected hydroclimatic indices in selected Norwegian and Polish catchments, Acta Geophys., 65, 829–848. DOI:10.1007/s11600-017-0062-5.

Młyński, D.P., Wałęga, A., Bugajski, P., Operacz, A., Kurek, K. (2019). Verification of empirical formulas for calculating mean low flow in reflect to affecting on disposable water resources. Acta Sci. Pol. Formatio Circumiectus, 18(2), 83–92, DOI: 10.15576/ASP.FC/2019.18.2.83.

Mood, A.M. (1954). On the asymptotic efficiency of certain non-parametric two-sample tests. Ann. Math. Stat. 25, 514–522.

Niedzielski, T., Kosek, W. (2011). Minimum time span of TOPEX/Poseidon, Jason-1 and Jason-2 global altimeter data to detect a significant trend and acceleration in sea level change. Adv. Space Res. 47, 1248–1255.

Nowak, B. M., Ptak, M., Sojka, M. (2019). Changes in ice regime of Jagodne Lake (North-Eastern Poland). Acta Scientiarum Polonorum Formatio Circumiectus, 18(1) 89–100, DOI:10.15576/ASP.FC/2019.18.1.89.

Önöz, B., Bayazit, M. (2003). The Power of Statistical Tests for Trend Detection, Turkish J. Eng. Env. Sci. 27, 247–251.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rutkowska, A. (2015). Properties of the Cox-Stuart test for trend in application to hydrological series: the simulation study. Commun. Stat – Simul C., 44(3), 565–579. DOI: 10.1080/03610918.2013.78498.

Rutkowska, A., Banasik, K. (2016). The Cucconi Test for Location-scale Alternatives in Application to Asymmetric Hydrological Variables. Commun. Stat – Simul C., 45(3), 986–1000. DOI: 10.1080/03610918.2014.911897.

Sen, A. K., Niedzielski, T. (2010). Statistical Characteristics of Riverflow Variability in the Odra River Basin, Southwestern Pol. J. Environ. Stud. 19(2), 387–397.

Siegel, S., Tukey, J. W. (1960). A non-parametric sum of rank procedure for relative spread in unpaired samples. J. Am. Stat. Assoc. 55, 429–444.

Stephenson, A. G. (2002). Evd: Extreme Value Distributions. R News, 2(2), 31–32, June 2002. URL: https://CRAN.R-project.org/doc/Rnews/.

Strużyński, A. 2011. Flood Protection in highly valuable river ecosystems – middle delta system of the Nida river, Electronic Journal of Polish Agricultural Universities, 14(2), http://www.ejpau.media.pl/volume14/issue2/art-06.html (accessed 08.10.2020)

Yevjevich, V. M. (1963). Fluctuations of wet and dry years. Part 1, research data assembly and mathematical models. Colorado State University Hydrology Paper 1.

Yu, P.S., Yang, T. C, Kuo, C. C. (2006). Evaluating long-term trends in annual and seasonal precipitation in Taiwan. Water Resour. Manag. 20, 1007–1023. DOI:10.1007/s11269-006-9020-8

Yue, S., Pilon, P., Cavadias, G. (2002a). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, J. Hydrol., 259, Issues 1–4, 254–271.

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

Yue, S., Pilon, P., Phinney, B. (2003). Canadian streamflow trend detection: impacts of serial and cross-correlation. Hydrol. Sci. J. 48(1), 51–64.

Zhang, Q., Xu, C., Yang, T. (2009). Variability of Water Resource in the Yellow River Basin of Past 50 Years, China. Water Resour. Manag. 23, 1157–1170 https://doi.org/10.1007/s11269-008-9320-2.

Zhou, C., van Nooijen, R., Kolechkina, A., Hrachowitz, M. (2019). Comparative analysis of non-parametric change-point detectors commonly used in hydrology Hydrological Sciences Journal, VOL. 64, NO. 14, 1690–1710. https://doi.org/10.1080/02626667.2019.1669792

Ziernicka-Wojtaszek, A., Kopcińska, J., 2020. Variation in Atmospheric Precipitation in Poland in the years 2001–2018. Atmosphere 11(794), DOI:10.3390/atmos11080794

PORÓWNANIE TESTÓW O ISTNIENIU TRENDU W PARAMETRACH POŁOŻENIA I SKALI W SZEREGACH HYDROLOGICZNYCH I OPADOWYCH

ABSTRAKT

Cel pracy
Celem pracy jest porównanie własności testów o istnieniu trendu w parametrach położenia i skali w szeregach hydrologicznych i opadowych, w szczególności (i) dokonanie przeglądu testów nieparametrycznych o istnieniu trendu różnego typu, znanych z literatury; testy były analizowane pod kątem ich zdolności do wykrycia trendu (tzn. mocy), (ii) zbadanie mocy tych testów nieparametrycznych, dla których moc jest nieznana w zastosowaniu do szeregów hydrologicznych i opadowych (testy Ansari-Bradleya (AB), Siegel-Tukeya (ST) i Mooda (M), (iii) ocena różnic między testami.

Materiał i metody
W badaniach wykorzystano ciągi liczb pseudolosowych oraz realizacje historycznych szeregów czasowych hydrologicznych i opadowych. Wykonano symulacje Monte Carlo i porównanie własności tych testów (AB, ST, M), które nie były dotychczas badane pod kątem zastosowania w szeregach hydrologicznych i opadowych.

Wyniki i wnioski
Wyniki wskazują na wzrost mocy testów AB, ST i M wraz ze wzrostem liczebności próby oraz wzrostem siły trendu skokowego w parametrze skali, niewrażliwość na zmiany współczynnika zmienności, małą wrażliwość na zmiany typu rozkładu zmiennej losowej oraz na największą moc testu M. Wykryto skokowy trend w dwóch przykładowych szeregach. Testy AB, ST i M mogą być stosowane do weryfikacji hipotezy o istnieniu skokowego trendu w parametrze skali w szeregach hydrologicznych i opadowych.

Słowa kluczowe: hydrologiczne i opadowe szeregi czasowe, trend monotoniczny lub skokowy w parametrze położenia lub skali, test statystyczny, moc testu