ELSEVIER

Journal of Hydrology 259 (2002) 254–271

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series

Sheng Yue[a,*], Paul Pilon[a], George Cavadias[b]

[a]Meteorological Service of Canada—Ontario Region, Environment Canada, 867 Lakeshore Road, P.O. Box 5050, Burlington, Ont., Canada L7R 4A6
[b]D61, 1321 Sherbrooke West, Montreal, Que., Canada H3G 1J4

Received 24 July 2001; revised 26 November 2001; accepted 28 November 2001

Abstract

In many hydrological studies, two non-parametric rank-based statistical tests, namely the Mann–Kendall test and Spearman's rho test are used for detecting monotonic trends in time series data. However, the power of these tests has not been well documented. This study investigates the power of the tests by Monte Carlo simulation. Simulation results indicate that their power depends on the pre-assigned significance level, magnitude of trend, sample size, and the amount of variation within a time series. That is, the bigger the absolute magnitude of trend, the more powerful are the tests; as the sample size increases, the tests become more powerful; and as the amount of variation increases within a time series, the power of the tests decrease. When a trend is present, the power is also dependent on the distribution type and skewness of the time series. The simulation results also demonstrate that these two tests have similar power in detecting a trend, to the point of being indistinguishable in practice.

The two tests are implemented to assess the significance of trends in annual maximum daily streamflow data of 20 pristine basins in Ontario, Canada. Results indicate that the *P*-values computed by these different tests are almost identical. By the binomial distribution, the field significant downward trend was assessed at the significance level of 0.05. Results indicate that a higher number of sites show evidence of decreasing trends than one might expect due to chance alone. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Mann–Kendall test; Spearman's rho test; Non-parametric test; Trend analysis; Power of a test; Statistical hydrology

1. Introduction

The rank-based non-parametric Mann–Kendall (MK) statistical test (Mann, 1945; Kendall, 1975) has been commonly used to assess the significance of trends in hydro-meteorological time series such as water quality, streamflow, temperature, and precipitation. The main reason for using non-parametric statistical tests is that compared with parametric statistical tests, the non-parametric tests are thought to be more suitable for non-normally distributed data and censored data, which are frequently encountered in hydro-meteorological time series. The serial independence of a time series is still required in non-parametric tests. Examples of use of the MK test for detecting trend in hydrological and hydro-meteorological time series include the works by Steele et al. (1974), Hirsch et al. (1982), Hirsch and Slack (1984), Crawford et al. (1983), van Belle and Hughes (1984), Cailas et al. (1986), Hipel et al. (1988), Taylor and Loftis (1989), Demaree and Nicolis (1990), Gan

* Corresponding author. Fax: +1-905-336-4989.
E-mail address: sheng.yue@ec.gc.ca (S. Yue).

PII: S0022-1694(01)00594-7

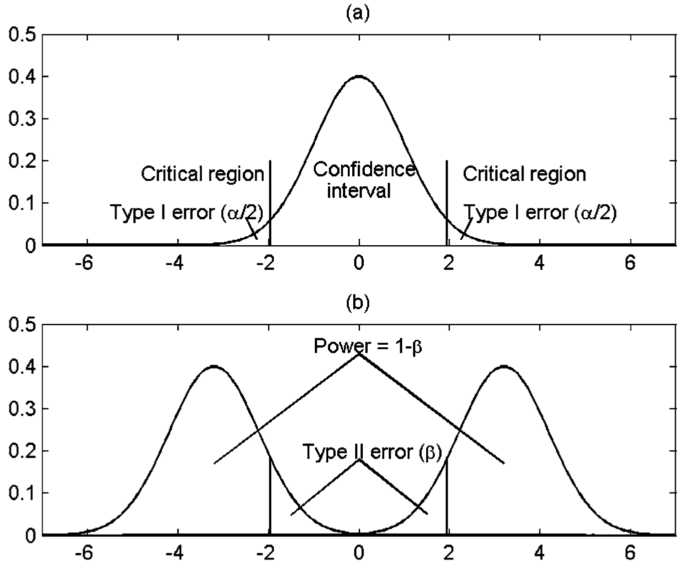

Fig. 1. (a) Schematic illustration of the confidence interval, critical regions, and Type I error for the two-tailed test; and (b) schematic illustration of the Type II error and the power of the test.

(1998), Zetterqvist (1991), Mcleod et al. (1991), Chiew and McMahon (1993), Yu et al. (1993), Lettenmaier et al. (1994), Burn (1994), Yulianti and Burn (1998), Lins and Slack (1999), Douglas et al. (2000), Zhang et al. (2000, 2001), Yue et al. (2002), and others. Although this test is widely used, a rather incomplete picture of the power of the MK test for the detection of trend under various circumstances is the current state of the art.

The Spearman's rho (SR) test is another rank-based non-parametric statistical test that can also be used to detect monotonic trend in a time series (Lehmann, 1975; Sneyers, 1990). However, since the appearance of the paper of Hirsch et al. (1982), the MK test has been popularly used to assess the significance of trends in hydro-meteorological time series. For whatever reason, the SR test is seldom used in hydro-meteorological trend analysis. Limited examples using the SR test include the works by Lettenmaier (1976), El-Shaarawi et al. (1983), Pilon et al. (1985), McLeod et al. (1991), and Hipel and McLeod (1994). This disproportionate number of applications between the two approaches may lead to an impression that the MK test is superior to the SR test on detection of trend in hydro-meteorological time series.

The objectives of this study are (i) to document the ability of both the MK test and the SR test to detect trend, as well as the influence of sample sizes and sample variations on the power of the tests; (ii) to explore the sensitivity of the power to the distribution type of sample data; (iii) to compare the power of the MK and SR tests; and (iv) to discuss the difference between statistical significance and practical significance. Both tests are also applied to test for trends in the serially independent annual maximum daily flow data of 20 pristine river basins in Ontario, Canada.

2. Power computation

The significance level, or a Type I error, α, is the probability of rejecting the null hypothesis, when it is true (see Fig. 1(a)). Significance levels are normally set quite low at values of 0.01, 0.05 or 0.10. The smaller the value of α, the more confidence there is that the null hypothesis is really false when it has been identified as such. A Type II error (β) is the probability of accepting a null hypothesis, when it is false (see Fig. 1(b)). The power of a test is the probability of correctly rejecting the null hypothesis, when it is false, which is equal to $1-\beta$ (see Fig. 1(b)).

When sampling from a population that represents the case where the null hypothesis is false, the power

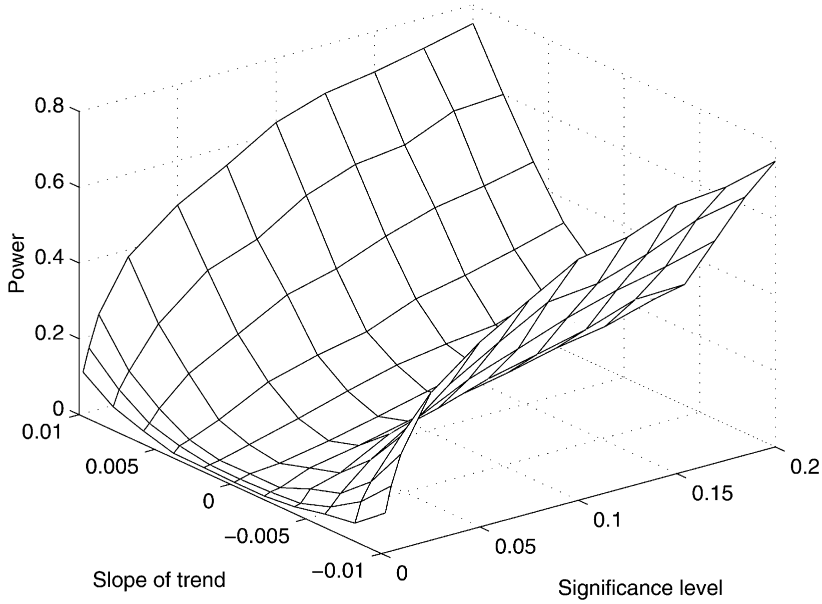

Fig. 2. Power–slope–significance level curve ($n = 50$ and $C_V = 0.5$).

can be estimated by

$$\text{Power} = \frac{N_{rej}}{N} \tag{1}$$

where N is the total number of simulation experiments and N_{rej} is the number of experiments that fall in the critical region (see Fig. 1(b)). The statistics of the MK and SR tests are presented in Appendix A. The critical regions of the MK statistic (S) can be approximately given by

$$S < z_{\alpha/2}\sqrt{V(S)} \text{ or } S > z_{1-\alpha/2}\sqrt{V(S)} \tag{2}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are, respectively, the $\alpha/2$ and $1-\alpha/2$ quantiles of the standard normal distribution; and $V(S)$ is the sample variance of the MK statistic S (see Appendix A). For the SR test, the critical region can be given by the similar inequalities as in Eq. (2) in which $V(S)$ is replaced by the variance, $V(D)$, of the SR statistic D (see Appendix A).

3. Power of the MK test for detecting trend

Monte Carlo simulation is conducted to observe the power of the MK test. The experiment generates 2000 independent normally distributed time series (R_t) for each sample size $n = 10$ (10) 100 with the mean $E(R_t) = 1.0$ and different variance $V(R_t) = (0.1 \times i)^2$ where $i = 1$ (1) 10. The corresponding standard deviations ($\mathrm{SD}(R_t)$) and coefficients of variation ($C_V = \mathrm{SD}(R_t)/E(R_t)$) are 0.1 (0.1) 1.0. Some selected particular linear trend scenarios ($T_t = bt$, $b = -0.01$ (0.002) 0.01, $t = 0, 1, 2, \ldots, n$) are superimposed onto each of the generated series.

The relationship among power, significance level, and the magnitude or the slope of trend for the sample size $n = 50$ and coefficient of variation $C_V = 0.5$ is summarized in Fig. 2. Similar patterns emerge for other sample sizes. The sample size of 50 is chosen simply for illustrative purposes. For a fixed significance level $\alpha = 0.002$, 0.005, 0.01, 0.025 (0.025) 0.20, the power of a test is an increasing function of the absolute slope of trend. For a fixed slope of the trend, increasing the significance level also increases the power.

Fig. 3 depicts the relationship among power, slope of trend, and sample size for the given significance level of 0.05 and coefficient of variation $C_V = 0.5$. The power of the test is an increasing function of both the absolute slope and the sample size. In other words, as the sample size increases, the power of the test increases leading to an increased ability to discern the existence of trend.

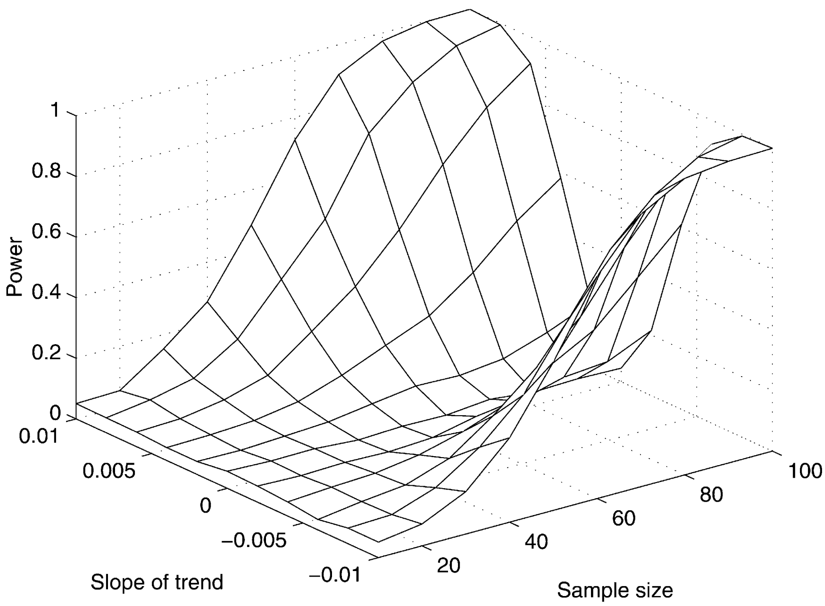

Fig. 3. Power–slope–sample size curve ($\alpha = 0.05$ and $C_V = 0.5$).

The relationship between the power, slope, and coefficient of variation with sample size of 50 and significance level of 0.05 are illustrated in Fig. 4. It is evident that for a fixed slope, the power of a test is a decreasing function of the coefficient of variation of a time series. That is, as the amount of variation within a time series increases, the power of the test decreases, implying it is more

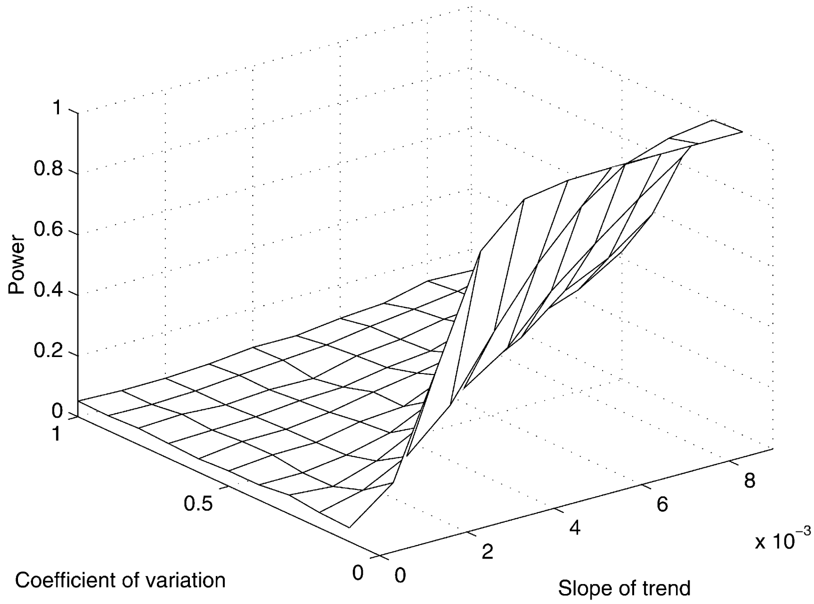

Fig. 4. Power–coefficient of variation–slope curve ($\alpha = 0.05$ and $n = 50$).

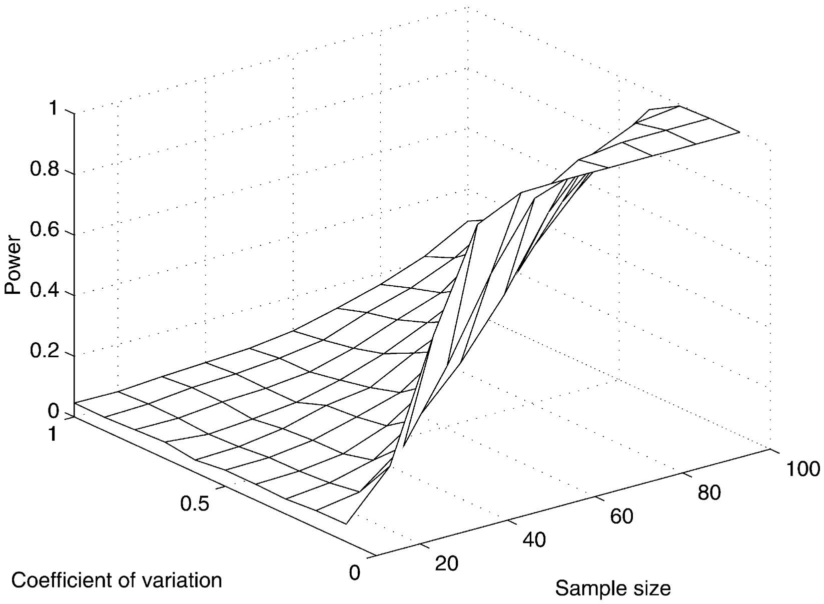

Fig. 5. Power–coefficient of variation–sample size curve ($\alpha = 0.05$ and $b = 0.005$).

difficult to detect the existence of trend. In essence, variation within a series masks the existence of trend. Similarly, Fig. 5 displays the relationship among the power, sample size, and coefficient of variation for the trend $b = 0.005$ and a significance level of 0.05.

From this analysis, it is possible to establish the minimum required sample size in order to detect a

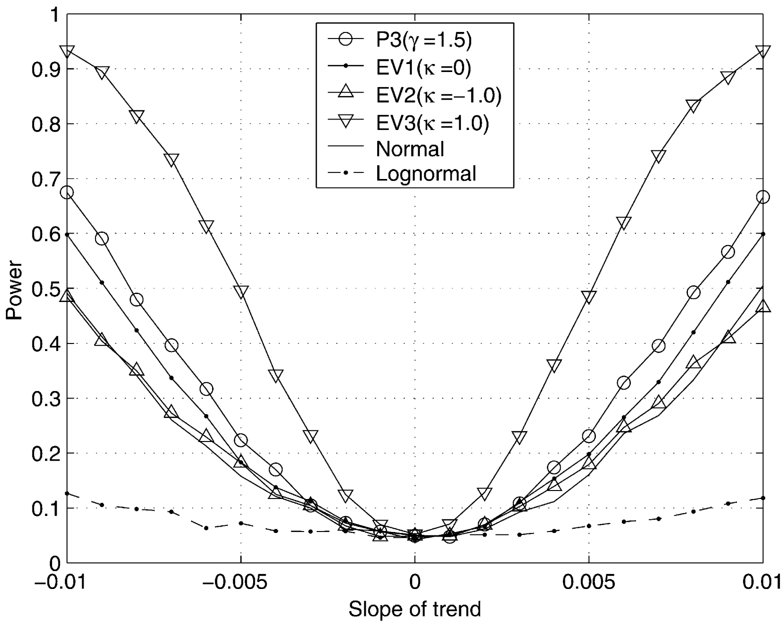

Fig. 6. Power of the test for different distribution types.

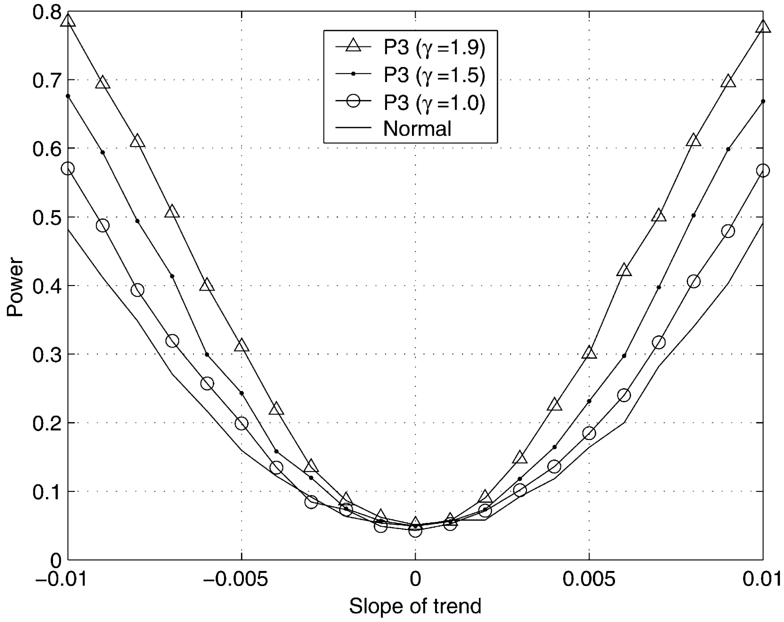

Fig. 7. Power–slope curve of P3-distributed series with different positive skewness (γ).

pre-specified population trend at a pre-assigned significance level. This analysis may also serve to provide information for hydro-meteorological observation network design, wherein one purpose of a network is to monitor potential gradual changes that may be due to growing concentration of greenhouse gases (Slack and Landwehr, 1992; Environment Canada, 1998).

4. Power of the MK test for detecting trend in non-normally distributed series

In Section 3, the power of the MK test was illustrated on the basis that the generated series are normally distributed. However, hydro-meteorological time series tend to be skewed and seldom follow the normal distribution. This section further examines the ability of the MK test to detect a trend in non-normally distributed time series. A few commonly used distribution types in hydrology, namely the extreme value distribution (EV1, EV2, and EV3), the Pearson type 3 distribution (P3), and the lognormal distribution types are employed to examine the power of the test. For the generation of series with these distribution types, the reader is referred to the work of Stedinger et al. (1993).

Similar to Section 3, 2000 time series with sample size of 50, mean of 1.0 and standard deviation of 0.5, and with each of the above selected distribution types were generated. The power-slope curves with different distribution types are displayed in Fig. 6. In the case of no trend, the power of the test remains the same for different distribution types and is equal to the pre-assigned significance level of 0.05. This indicates that the null distribution of the MK test statistic is not sensitive to the distribution type of time series. However, in the case when some trend exists, the power of the test is dramatically different for different distribution types. This is not due to smaller sample size and/or smaller number of simulated samples. The EV3 has the highest power while the lognormal distribution has the lowest power. This finding indicates that when a trend does exist, the power of the MK test is also dependent on the distribution type, which is in contrast to the common thought that the MK test is rank-based and would be distribution-free.

To illustrate that the power of the test is also dependent on the shape parameter of the probability distributions that sample data follow, the power of the test

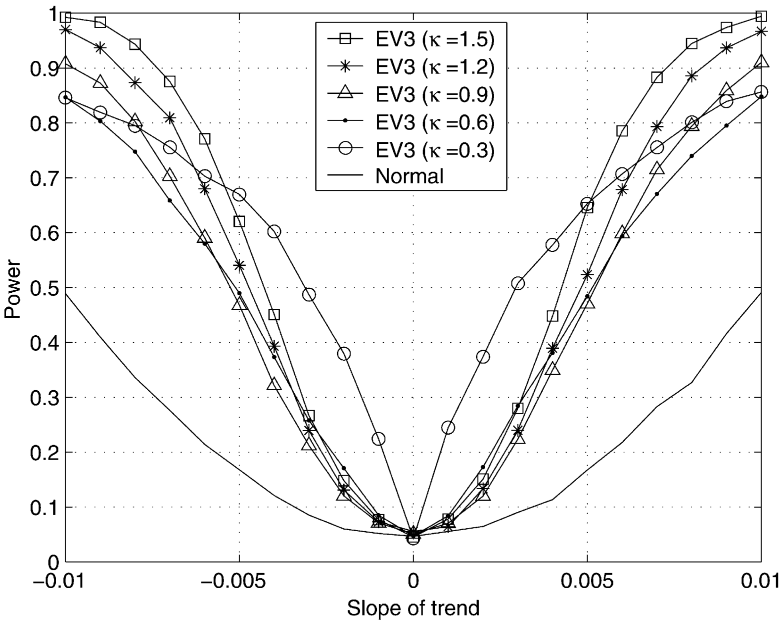

Fig. 8. Power–slope curve of EV3-distributed series with different shape parameter (κ).

for P3 distributed time series with positive skewness is shown in Fig. 7 and for EV3-distributed series in Fig. 8. Figs. 7 and 8 demonstrate that the power of the test is also influenced by the shape parameter of a distribution. This point has also been observed by Kingman and Gary (1994) and Levy and McCuen (2000).

The increase/decrease in power as described earlier, could impact on the overall interpretation of the significance of results of a number of sites, termed field significance (Livezey and Chen, 1983). This would particularly be the case when one is making inferences for various variables within the hydrological cycle in the evaluation of the significance of trend in, for example, trend-detection studies (Lettenmaier et al., 1994; Zhang et al., 2001). Even though two series such as precipitation and streamflow might possess similar trend components, the ability of the test statistic to discern the presence of trend will be impacted upon by the site's statistical properties such as skewness. These properties will vary depending on the variable being analyzed. This makes a comparison of the trend patterns for the various variables more complex than one might have expected. These properties would also have an impact on the overall assessment of field significance given that it reflects the individual site's significance. Let us look to a hypothetical example to illustrate these points. Two networks with 100 sites each have the same magnitude of trend ($b = 0.005$). If the sites of the first network were normally distributed, while the second were P3-distributed ($\gamma = 1.9$), then we would expect approximately a doubling in the detection rate based on the results shown in Fig. 7. Such an increase in the detection rate could impact upon the conclusion regarding evidence of trend. In essence, the power of the test is dramatically affected by the site's characteristics when the trend exists.

5. Comparison of the power of the MK and SR tests

The MK and SR non-parametric tests have been compared by other authors (e.g. Daniel, 1978). These studies have found that very little basis usually exists for choosing one over the other. However, Daniel (1978) mentioned a few points of interest when comparing the MK and SR test statistics. He noted that the distribution of *S approaches normality* more rapidly than does *D*; and *S* provides an unbiased

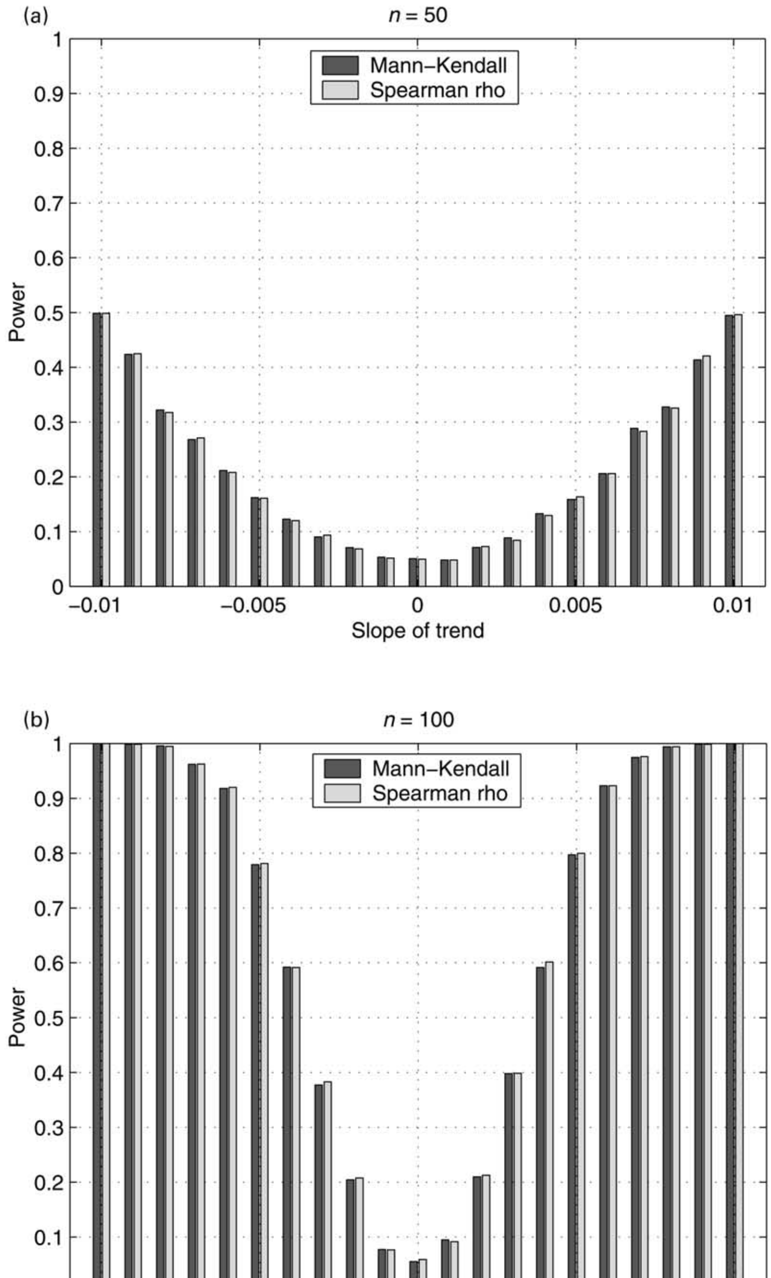

Fig. 9. Comparison of the power of the MK test and the Spearman's rho test for different slopes of series with $\alpha = 0.05$ and $C_V = 0.5$: (a) $n = 50$ and (b) $n = 100$.

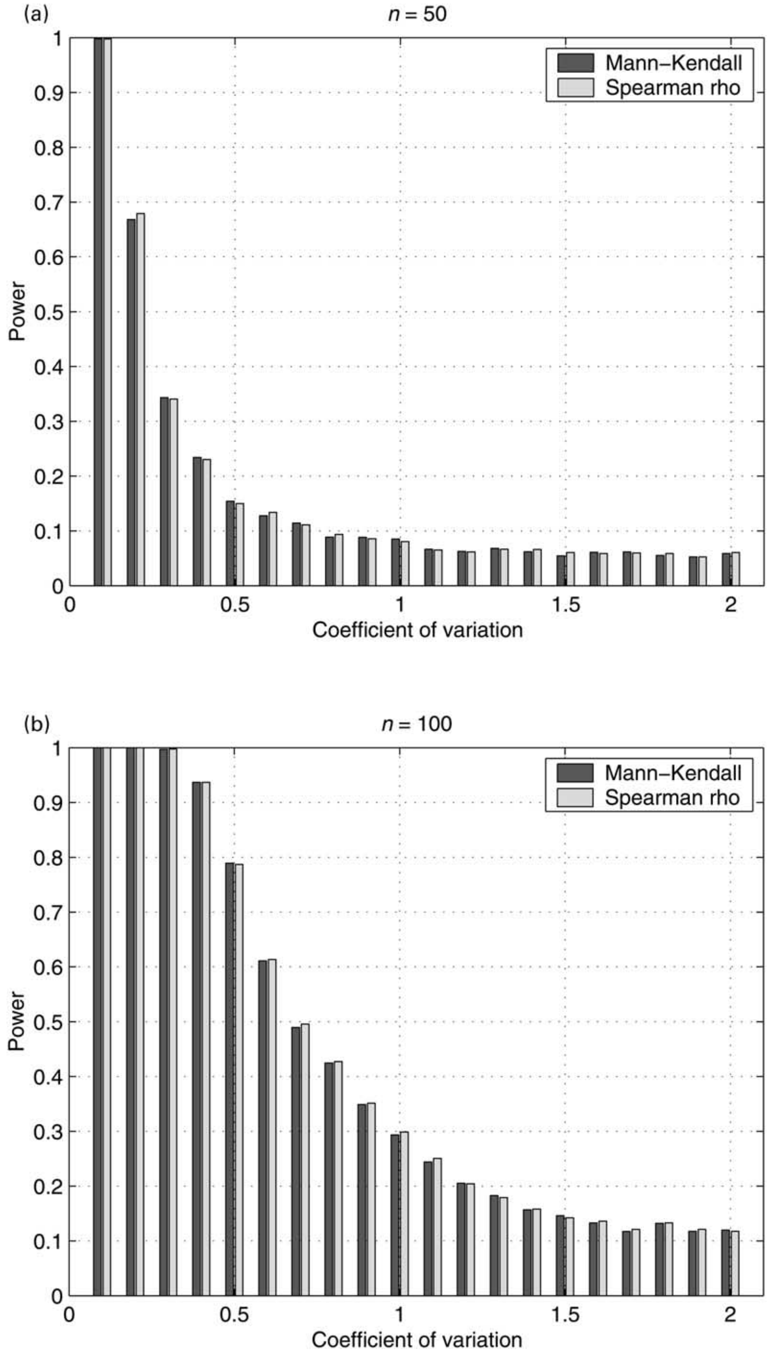

Fig. 10. Comparison of the power of the MK test and the Spearman's rho test for different coefficients of variation of series with $\alpha = 0.05$ and $b = 0.005$: (a) $n = 50$ and (b) $n = 100$.

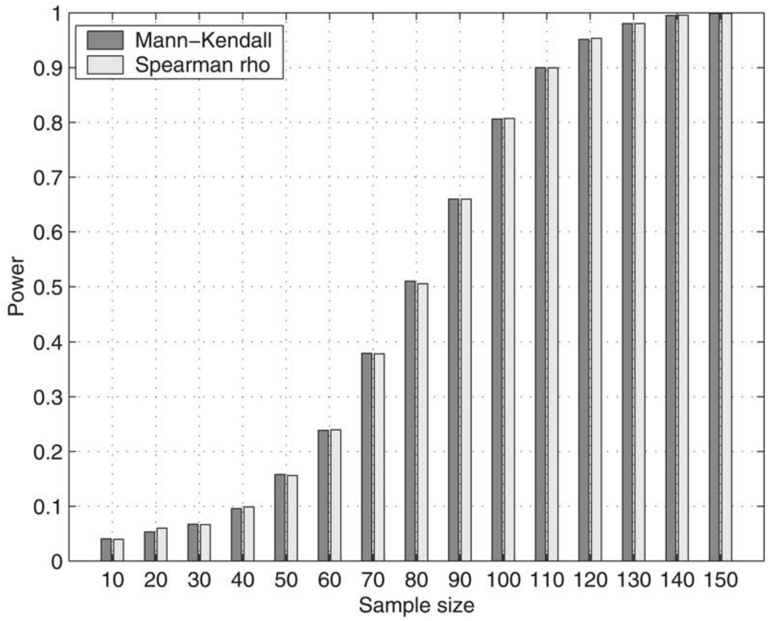

Fig. 11. Comparison of the power of the MK and SR tests for different sample sizes ($\alpha = 0.05$, $C_V = 0.5$ and $b = 0.005$).

estimate of the population parameter, while D does not, and therefore, S is more interpretable.

Following the same procedure as in the preceding sections, the relationships among the power, slope, sample size, and coefficient of variation for the SR test were also examined. Very similar results for the two approaches were obtained and for the sake of brevity, the results are not reported in the paper.

A comparison of the power of the two tests was performed and is presented herein. For normally distributed random variables, the power-slope relationships with $C_V = 0.5$ and $\alpha = 0.05$ for sample size $n = 50$ and 100 are displayed in Fig. 9(a) and (b), respectively. The power–C_V relationship with $b = 0.005$ and $\alpha = 0.05$ for $n = 50$ and 100 are depicted in Fig. 10(a) and (b), respectively. Fig. 11 shows the power of the two tests for various sample sizes for $C_V = 0.5$, $b = 0.005$, $\alpha = 0.05$. It can be seen that both tests have almost the same power for detecting the designed linear trend.

For non-normally distributed time series, the power of the two tests for the P3 distribution with a skew coefficient of 1.5 for sample sizes of 50 and 100 is illustrated in Fig. 12(a) and (b). This was also performed for a negative skewness of 1.5 with results being very similar to that shown in Fig. 12. The comparison of the power of the two tests for the other distribution types was also made. Results indicate that these two tests also have similar power for detecting trend in highly skewed time series. The power of the two tests with other sample sizes was also observed and is similar to that shown in Fig. 11.

6. Case study

This section applies both tests to assess the significance of trend in annual maximum daily streamflow data of 20 pristine river basins, located in Ontario, Canada. Daily average flow data of these basins are recorded in HYDAT CD-ROM by Environment Canada (1999). The drainage area, record length, mean, coefficient of variation, coefficient of skewness, and coefficient of kurtosis of annual maximum streamflows for these basins are presented in columns (3)–(8) of Table 1, respectively. The spatial distribution of the stations is illustrated in Fig. 13, in which the stations are identified by their numbers. For a normally distributed random series, its skewness and kurtosis should be equal to 0 and 3, respectively. It is evident from Table 1 that the data are positively skewed and are in

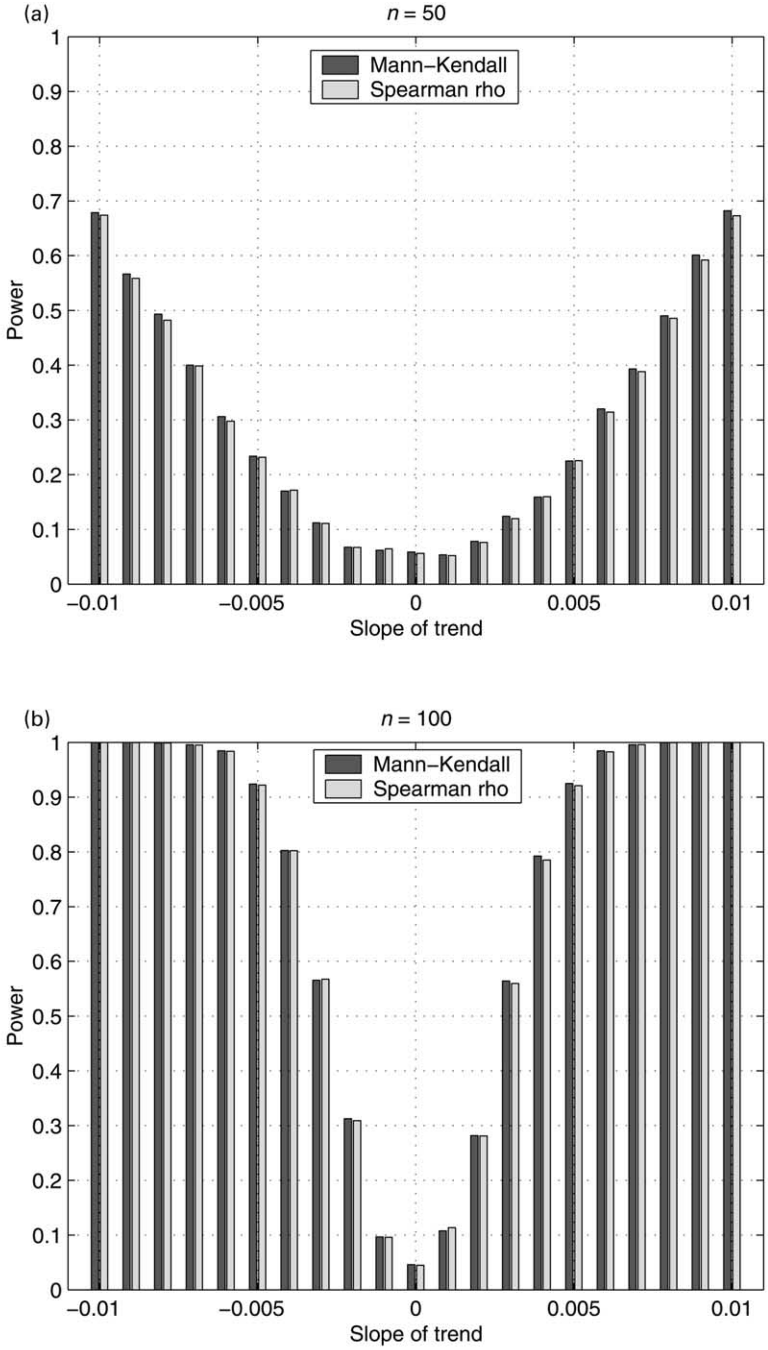

Fig. 12. Comparison of the power of the MK test and the Spearman's rho test for P3-distributed series with $\alpha = 0.05$, $C_V = 0.5$, and $\gamma = 1.5$: (a) $n = 50$ and (b) $n = 100$.

Table 1
Basic properties and *P*-values of annual maximum daily streamflow in 20 unregulated river basins in Ontario, Canada

Station		Area (km^2)	Record length (years)	Mean (m^3/s)	C_V	C_S	C_K	Correlation			Slope		*P*-value	
No.	Identifier							r_1	Limits		(m^3/s/year)	(12)/(5) (per year)	MK	SR
									Upper	Lower				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	02AA001	1550	75	125.2	0.417	1.075	4.219	0.006	0.176	−0.203	−0.41475	−0.00331	0.039	0.040
2	02AB008	187	45	23.9	0.579	1.048	3.859	0.158	0.222	−0.268	−0.12889	−0.00540	0.181	0.200
3	02BF002	1160	31	171.0	0.364	0.958	4.526	−0.277	0.262	−0.329	−1.73913	−0.01017	0.051	0.038
4	02CF008	179	22	23.8	0.410	0.482	2.862	−0.179	0.303	−0.398	−0.10000	−0.00419	0.400	0.417
5	02EA005	321	83	45.2	0.371	1.755	8.561	−0.033	0.168	−0.193	0.05714	0.00126	0.181	0.178
6	02EC002	1520	83	130.0	0.248	0.221	3.548	0.129	0.168	−0.193	0.06250	0.00048	0.316	0.298
7	02FB007	181	53	29.6	0.416	0.781	3.644	0.036	0.207	−0.245	−0.09626	−0.00326	0.208	0.210
8	02FC001	3960	84	498.6	0.348	0.431	3.212	−0.007	0.167	−0.192	0.47723	0.00096	0.262	0.214
9	02GA010	1030	50	185.7	0.413	0.335	2.088	0.021	0.212	−0.253	−0.42857	−0.00231	0.241	0.237
10	02HL004	712	41	65.7	0.304	0.769	4.379	0.145	0.232	−0.282	−0.05395	−0.00082	0.416	0.401
11	02JC008	1780	26	164.6	0.256	0.408	2.531	−0.131	0.282	−0.362	−2.07692	−0.01262	0.018	0.024
12	02KB001	4120	83	219.5	0.371	0.520	3.041	0.167	0.168	−0.193	1.15094	0.00524	0.001	0.002
13	02LB007	246	49	45.8	0.410	0.936	4.116	0.097	0.214	−0.256	−0.28794	−0.00629	0.062	0.037
14	04DA001	5960	32	241.4	0.497	0.978	3.393	−0.118	0.258	−0.323	−4.05991	−0.01682	0.023	0.043
15	04GA002	5390	27	87.0	0.449	1.686	6.755	−0.074	0.278	−0.355	1.05417	0.01211	0.091	0.093
16	04GB004	11,200	26	274.5	0.459	1.045	3.84	0.018	0.282	−0.362	4.17647	0.01522	0.059	0.039
17	04JC002	2410	48	124.2	0.311	0.917	3.804	−0.053	0.216	−0.259	−0.09722	−0.00078	0.371	0.344
18	04KA001	4250	25	423.0	0.539	0.418	2.045	−0.213	0.287	−0.370	−2.86154	−0.00676	0.304	0.299
19	04LJ001	8940	78	880.0	0.314	0.493	3.204	0.050	0.173	−0.199	−0.45455	−0.00052	0.413	0.483
20	04MF001	6680	32	692.7	0.364	0.370	3.187	−0.021	0.258	−0.323	0.95833	0.00138	0.461	0.459

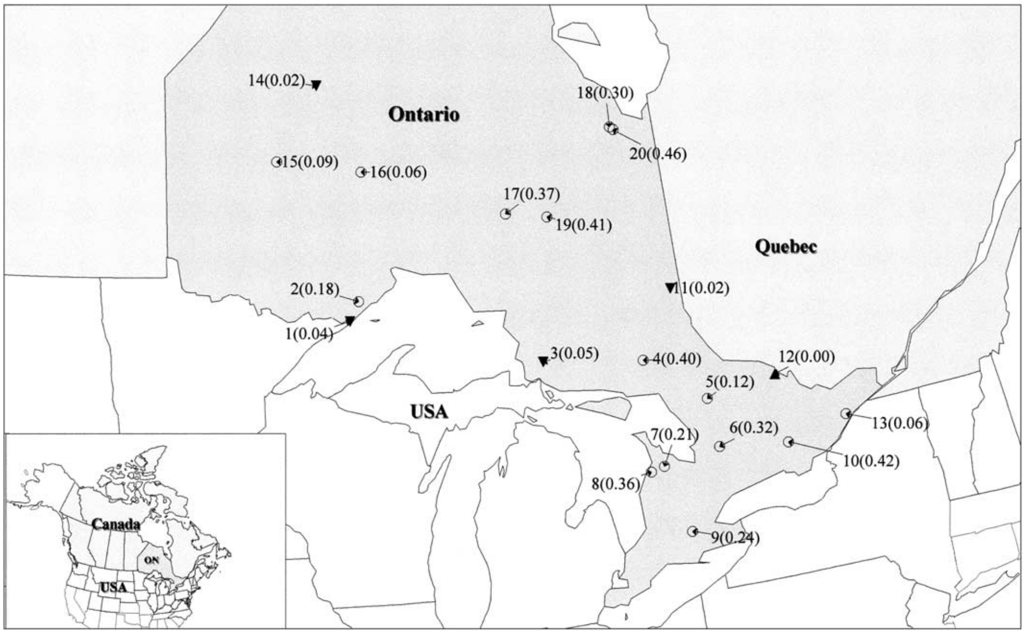

Fig. 13. Spatial distributions of 20 streamflow stations of pristine river basins of Ontario, Canada.

all likelihood poorly described by a normal distribution.

The lag-1 serial correlation coefficient (r_1) and its upper and lower limits of the confidence interval at the significance level of 0.10 of the two-tailed test for the annual maximum daily streamflows are presented in columns (9)–(11) of Table 1, respectively (see Anderson, 1942; Yevjevich, 1972; Salas et al., 1980). It can be seen that all the annual maximum flow series are not significantly serially correlated at the significance level of 0.10. Hence the MK and SR tests can be executed without need of correction for the effect of serial dependence on the tests (von Storch, 1995).

The magnitude of the slope was estimated using the approach by Theil (1950) and Sen (1968) (Appendix B), which is provided in column (12) of Table 1. This slope divided by the mean of the annual maximum daily flows is presented in column (13), which is termed a unit slope. The *P*-values by Eq. (A10) for the MK test are presented in columns (14), which are also presented in parentheses in Fig. 13. Among these 20 stations, four demonstrated a significant downward trend while only one station (station 12) shows a positive trend at the significance level of 0.05. The percentage of sites displaying evidence of a significant downward trend is 20%, which is much higher than one might expect at the pre-selected level of significance (5%).

The mean cross-correlation of the cross-correlation coefficients among the 20 streamflow observation sites is 0.14, which implies that the cross-correlation among the sites could be ignored. Thus, the application of the binomial probability distribution to assess the field significance of trend would be appropriate (Livezey and Chen, 1983) as the binomial distribution for assessing the field significance of trend requires that there is no cross-correlation among the sites in a region. The probability being or exceeding four sites showing downward trend by chance at a significance level of 0.05 is 1.6% by the binomial distribution (see Appendix C). Thus, there is some evidence that the downward trends may not be due to chance alone. In other words, annual maximum daily flow in Ontario region might be experiencing a downward trend, without considering the influence of cross-correlation on the test results.

The SR test was also executed to detect trends in the annual maximum daily streamflows. The *P*-values by the SR test are presented in column (15) of Table 1. The

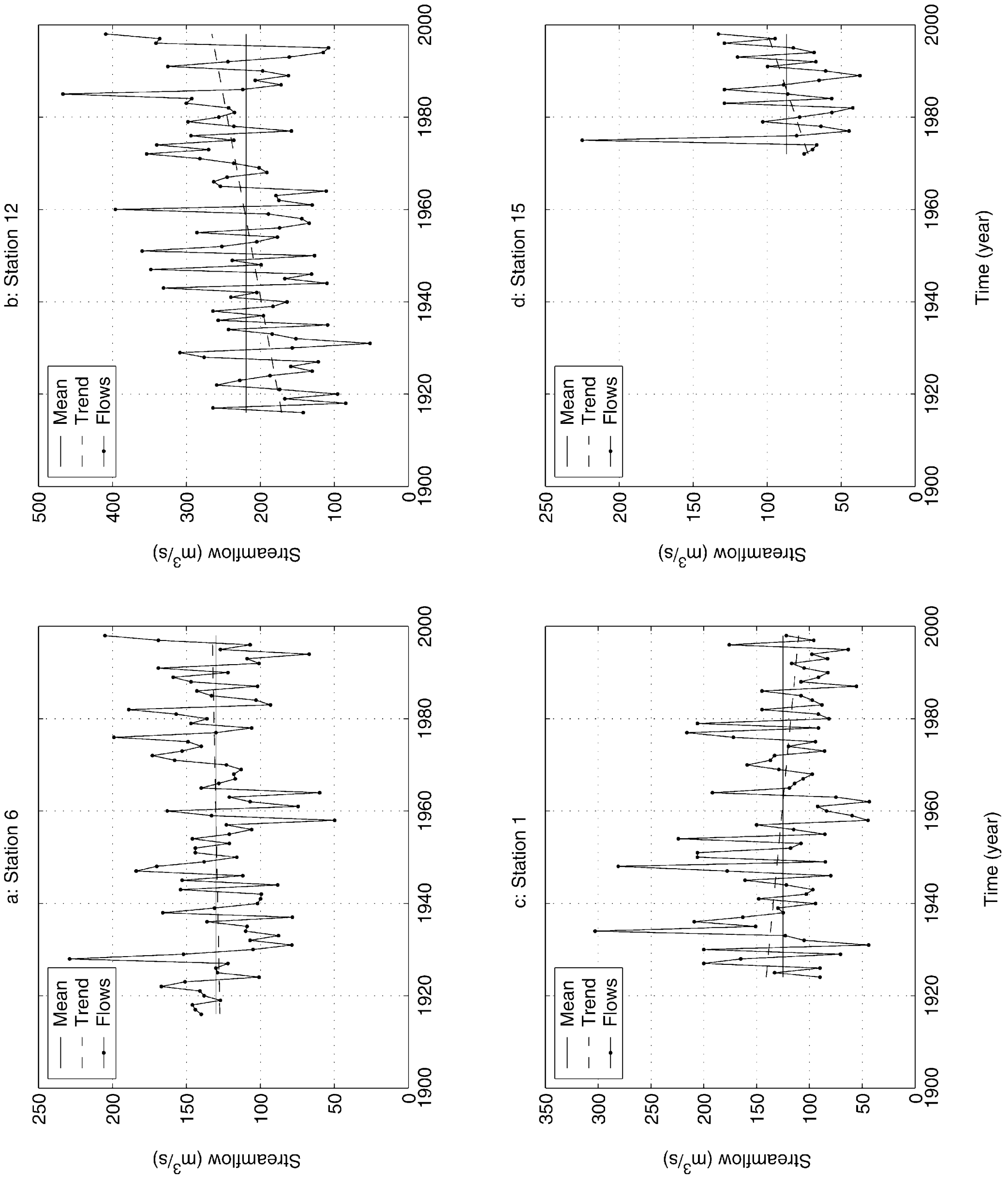

Fig. 14. Graphical illustration of annual maximum streamflow, its long-term mean, and the linear trends for stations 6, 12, 1, and 15.

P-values for the SR test are almost identical to those of the MK test. This case study shows the high similarity of results when both tests are applied, which was confirmed earlier through the simulation experiments.

7. Statistical significance versus practical significance

Daniel (1978) has discussed the difference between statistical significance and practical significance. He noted that a statistically significant trend may not be practically significant and vice versa. Sufficiently large samples will reveal any change, no matter how small, through the use of a statistical test, but this may not be of any practical help. Likewise, small samples may fail to detect a change statistically, but the degree of change might be of practical significance.

In our study case, although 13 of the 20 stations have negative slopes, only stations 1, 3, 11, and 14 have 'statistically' significant downward trends (*P*-values ≤ 0.05) by the MK test. The slope of station 2 is not statistically significant yet its slope is greater than that of station 1. This difference in *P*-value is attributable to the longer record length for station 1. Similarly, of the seven stations having positive slopes, only station 12 was assessed to be statistically significant at the significance level of 0.05. Station 15 has a slope much larger than that of station 12, but the trend of station 15 is not statistically significant.

Fig. 14 shows four plots containing annual maximum daily streamflow versus year of occurrence for some selected sites listed in Table 1, which are used to assist in exploring the importance of statistical versus practical significance. Fig. 14(a)–(d) represents the streamflows at site 6 (02EC002—Black River near Washago), site 12 (02KB001—Petawawa River near Petawawa), site 1 (02AA001—Pigeon River at Middle Falls), and site 15 (04GA002—Cat River below Wesleyan Lake), respectively. Site 6 has a *P*-value of 0.32 and a unit slope of 0.0048. A visual inspection of Fig. 14(a) shows little evidence of trend in the data. Over a 50-year period, the slope yields an increase in the mean flow of 130 m^3/s by approximately 3.1 m^3/s or 2.4%. From a practical perspective, the estimated change is within the sampling error for the streamflow measurements and does not appear, given the available data, as being of practical importance.

As a further example, Fig. 14(b) and (c) show evidence of an increasing trend at sites 12 and a decreasing trend at site 1, respectively. The linear trend also appears to be a reasonable estimate of the gradual tendency. Both tendencies have been assessed to be statistically significant (Table 1). Over a 50-year period, the estimated slope for site 12 results in an increase of 57.6 m^3/s or 26% in the mean of 220 m^3/s. In comparison, site 1 shows a decrease of 20.7 m^3/s or 17% of the mean of 125 m^3/s. The tendency estimated for both these basins is quite dramatic and are of practical significance.

The trend in the data of site 15 is not statistically significant (Table 1). Should the slope depicted in Fig. 14(d) be accurate, it could be inferred that the data are displaying a slope of practical importance. Over a 50-year period, the estimated slope results in an increase of 52.7 m^3/s or 61% in the mean of 87.0 m^3/s. The increasing tendency for this basin is quite dramatic and is of practical significance. However, caution should be advocated in this case particularly due to the lack of statistical evidence to support the alternative hypothesis of trend.

It must be noted that the ultimate purpose of a test for trend is to provide information to the engineer in charge of the water resources project on the current attributes of the streamflow (e.g. annual minimum and maximum, mean). However, even if the test indicates a significant trend with a slope of practical importance, the engineer is justified in changing his assumption concerning the magnitude of the flow characteristic only if this change is permanent for the life of the project. To ascertain this, it is necessary to complement the detection test with an attribution study, i.e. to determine the nature of the changes in the input variables of the basin system (temperature, precipitation, basin characteristics, etc.) that result in streamflow trends. It is also necessary to establish if the nature of these changes is likely to persist over the design life of the project. Only after the attribution study can the engineer say whether the statistically significant test is also of practical significance.

8. Conclusions

This study investigated the power of the MK test and SR test for assessing the significance of trend in hydrological series. The simulation experiments have demonstrated that the power of these tests is an increasing function of the slope of trend, sample size, and pre-assigned significance level; while it is a decreasing function of the variation of the time series. The power of these tests is also dependent on the distribution type and its shape parameter that time series have, should a trend exist. Such increases/decreases in power may inadvertently affect the interpretation of field significance when comparing variables that have different statistical properties. This information provides practitioners with a better picture of the power of these tests for detecting trends, given that their power varies with certain distribution type, sample size, sample variation, as well as sample skewness. SR test provided results almost identical to those obtained for the MK test.

Both tests were also applied to annual maximum flow data for 20 sites of pristine river basins in Ontario, Canada. Of the 20 sites, four demonstrated significant downward trend while one demonstrated significant upward trend. The binomial distribution was used to assess the field significance of downward trend. The results indicate that the evidence of a downward trend in the annual maximum daily streamflows for Ontario streams is stronger than what one would expect due to chance alone.

The statistical and practical significance of the results was also discussed and analyzed for the Ontario streamflow data, given that statistically significant trend does not necessarily imply trend of practical significance and vice versa. Overall, it was found that in this case study, practical and statistical significance was in agreement. Given the short record lengths and highly variable nature of the data, caution is advised to those who might consider use of practical aspects without statistical consideration. Physical attribution of the factors forcing the trend in streamflow must be ascertained as well as their potential future states prior to establishing whether statistically significant test results are of practical utility in engineering design. It is suggested that both statistical and practical significance of trend be considered in an overall analysis of trend.

Acknowledgements

This study is part of the Canadian Climate Change Action Fund Project: Use of Statistical Techniques for Assessing the Impact of Climate Change on Canadian Streamflows. The authors thank their colleague, Paul Ford for his kind assistance on drawing Fig. 13. The thoughtful comments and corrections of Harry F. Lins and the anonymous reviewers are gratefully acknowledged.

Appendix A

A.1. Mann–Kendall test

The MK test is based on the test statistic S defined as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(x_j - x_i) \tag{A1}$$

where the x_j are the sequential data values, n is the length of the data set, and

$$\mathrm{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases} \tag{A2}$$

Mann (1945) and Kendall (1975) have documented that when $n \geq 8$, the statistic S is approximately normally distributed with the mean and the variance as follows:

$$E(S) = 0 \tag{A3}$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i i(i-1)(2i+5)}{18} \tag{A4}$$

where t_i is the number of ties of extent i. The standardized test statistic Z is computed by

$$Z_{MK} = \begin{cases} \dfrac{S-1}{\sqrt{\mathrm{Var}(S)}} & S > 0 \\ 0 & S = 0 \\ \dfrac{S+1}{\sqrt{\mathrm{Var}(S)}} & S < 0 \end{cases} \quad \text{(A5)}$$

The standardized MK statistic Z follows the standard normal distribution with mean of zero and variance of one.

A.2. Spearman's rho test

SR test is another non-parametric rank-order test. Given a sample data set $\{X_i, i = 1, 2, \ldots, n\}$, the null hypothesis H_0 of the SR test against trend tests is that all the X_i are independent and identically distributed; the alternative hypothesis is that X_i increases or decreases with i, that is, trend exists. The test statistic is given by (Sneyers, 1990)

$$D = 1 - \frac{6\sum_{i=1}^{n}[R(X_i) - i]^2}{n(n^2-1)} \quad \text{(A6)}$$

where $R(X_i)$ is the rank of ith observation X_i in the sample of size n.

Under the null hypothesis, the distribution of D is asymptotically normal with the mean and variance as follows (Lehmann, 1975; Sneyers, 1990)

$$E(D) = 0 \quad \text{(A7)}$$

$$V(D) = \frac{1}{n-1} \quad \text{(A8)}$$

The P-value of the SR statistic (d) of the observed sample data is estimated using the normal cumulative distribution function (CDF) as its statistics are approximately normally distributed with mean of zero and variance of $V(D)$ for the SR statistic. Using the following standardization,

$$Z_{SR} = \frac{D}{\sqrt{V(D)}} \quad \text{(A9)}$$

the standardized statistic Z follows the standard normal distribution $Z \sim N(0, 1)$.

The P-value (probability value, p) of both the MK statistic (S) and the SP statistic (D) of sample data can be estimated using the normal CDF,

$$p = 0.5 - \Phi(|Z|)\ (Z = Z_{MK}, Z_{SR}) \quad \text{(A10)}$$

$$\left(\Phi(|Z|) = \frac{1}{\sqrt{2\pi}}\int_0^{|Z|} e^{-t^2/2}\,dt\right)$$

If the P-value is small enough, the trend is quite unlikely to be caused by random sampling. At the significance level of 0.05, if $p \le 0.05$, then the existing trend is considered to be statistically significant.

Appendix B

The magnitude of the slope of trend is estimated using the approach by Theil (1950) and Sen (1968). The slope is estimated by

$$b = \mathrm{Median}\left(\frac{x_j - x_l}{j - l}\right) \qquad \forall\, l < j \quad \text{(B1)}$$

where b is the estimate of the slope of trend and x_l is the lth observation. The slope determined by Eq. (C1) is a robust estimate of the magnitude of monotonic trend.

Appendix C

The binomial distribution is

$$P(k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \quad \text{(C1)}$$

where $P(k)$ is the probability of k occurrences in n trials, and p is a certain probability associated with each occurrence. The probability of k or more occurrences being an anomaly or accidental result is

$$P'(x \ge k) = 1 - \sum_{i=0}^{k-1} P(k) \quad \text{(C2)}$$

References

Anderson, R.L., 1942. Distribution of the serial correlation coefficients. Ann. Meth. Stat. 13 (1), 1–13.

Burn, D.H., 1994. Hydrologic effects of climatic change in West Central Canada. J. Hydrol. 160, 53–70.

Cailas, M.D., Cavadias, G., Gehr, R., 1986. Application of a nonparametric approach for monitoring and detecting trends in water quality data of the St. Lawrence River. Can. Water Poll. Res. J. 21 (2), 153–167.

Chiew, F.H.S., McMahon, T.A., 1993. Detection of trend or change in annual flow of Australian rivers. Int. J. Clim. 13, 643–653.

Crawford, C.G., Slack, J.R., Hirsch, R.M., 1983. Nonparametric tests for trend in water quality data using the statistical analysis system. Open Report no. 83-550, US Geological Survey, USA.

Daniel, W.W., 1978. Applied Nonparametric Statistics, Boston.

Demaree, G.R., Nicolis, C., 1990. Onset of Sahelian drought viewed as a fluctuation-induced transition. Q. J. R. Meteorol. Soc. 116, 221–238.

Douglas, E.M., Vogel, R.M., Knoll, C.N., 2000. Trends in flood and low flows in the United States: impact of spatial correlation. J. Hydrol. 240, 90–105.

El-Shaarawi, A.H., Esterby, S.R., Kuntz, K.W., 1983. A statistical evaluation of trends in the water quality of the Niagara River. J. Great Lakes Res. 9, 234–240.

Environment Canada, 1999. HYDAT CD-ROM version 98-1.05.8: surface water and sediment data.

Gan, T.Y., 1998. Hydroclimatic trends and possible climatic warming in the Canadian Prairies. Water Resour. Res. 34 (11), 3009–3015.

Hipel, K.W., McLeod, A.I., 1994. Time Series Modelling of Water Resources and Environmental Systems. Developments in Water Science, vol. 45. Elsevier, Amsterdam.

Hipel, K.W., McLeod, A.I., Weiler, R.R., 1988. Data analysis of water quality time series in Lake Erie. Water Resour. Bull. 24 (3), 533–544.

Hirsch, R.M., Slack, J.R., 1984. Non-parametric trend test for seasonal data with serial dependence. Water Resour. Res. 20 (6), 727–732.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18 (1), 107–121.

Kendall, M.G., 1975. Rank Correlation Methods. Griffin, London.

Kingman, A., Gary, Z., 1994. Some power considerations when deciding to use transformations. Stat. Med. 13, 769–783.

Lehmann, E.L., 1975. Nonparametrics, Statistical Methods Based on Ranks. Holden-Day, Inc, California.

Lettenmaier, D.P., 1976. Detection of trends in water quality data from records with dependent observations. Water Resour. Res. 12 (5), 1037–1046.

Lettenmaier, D.P., Wood, E.F., Wallis, J.R., 1994. Hydro-climatological trends in the continental United States, 1948–88. J. Clim. 7, 586–607.

Levy, B.S., McCuen, R.H., 2000. Skew-dependent detection of change of central tendency in hydrologic data. ASCE, J. Hydrol. Engng. 5 (1), 43–51.

Lins, H.F., Slack, J.R., 1999. Streamflow trends in the United States. Geophys. Res. Lett. 26 (2), 227–230.

Livezey, R.E., Chen, W.Y., 1983. Statistical field significance and its determination by Monte Carlo techniques. Mon. Weath. Rev. 111, 46–59.

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245–259.

McLeod, A.I., Hipel, K.W., Bodo, B.A., 1991. Trend assessment of water quality time series. Water Resour. Bull. 19, 537–547.

Pilon, P.J., Condie, R., Harvey, K.D., 1985. Consolidated frequency analysis package (CFA), User Manual for Version 1—DEC PRO Series. Water Resources Branch, Inland Water Directorate, Environment Canada, Ottawa, Canada.

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied Modelling of Hydrologic Time Series. Water Resources Publications, Littleton, CO, USA.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Slack, J.R., Landwehr, J.M., 1992. Hydro-climatic data network (HCDN): a U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874–1988, U.S. Geological Survey. Open File Report 92-129, USGS Water Supply Paper No. 2406, Reston, Virginia, 171 p.

Sneyers, R., 1990. On the Statistical Analysis of Series of Observations. World Meteorological Organization, Technical Note no. 143, WMO no. 415.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, D.R. (Ed.). Handbook of Hydrology. McGraw-Hill, New York Chapter 18.

Steele, T.D., Gilroy, E.J., Hawkinson, R.O., 1974. Techniques for the assessment of areal and temporal variations in streamflow quality. Open File Report, US Geological Survey, Washington, DC.

Taylor, C.H., Loftis, J.C., 1989. Testing for trend in lake and groundwater quality time series. Water Resour. Bull. 25 (4), 715–726.

Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis, I, II, III. Nederl. Akad. Wetensch. Proc. 53, 386–392 see also pp. 512–525 and 1397–1412.

van Belle, G., Hughes, J.P., 1984. Nonparametric tests for trend in water quality. Water Resour. Res. 20 (1), 127–136.

von Storch, H., 1995. Misuses of statistical analysis in climate research. In: Storch, H.V., Navarra, A. (Eds.). Analysis of Climate Variability: Applications of Statistical Techniques. Springer, Berlin, pp. 11–26.

Yevjevich, V., 1972. Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, CO, USA.

Yu, Y.S., Zou, S., Whittemore, D., 1993. Non-parametric trend analysis of water quality data of river in Kansas. J. Hydrol. 150, 61–80.

Yue, S., Pilon, P.J., Phinney, B., Cavadias, G., 2002. Patterns of trend in Canadian streamflow. Proceedings of the 58th Annual Eastern Snow Conference, CGU, May 14–17, Ottawa, Ontario, Canada (in press).

Yulianti, J.S., Burn, D.H., 1998. Investigating links between climatic warming and low streamflow in the Prairies Region of Canada. Can. Water Resour. J. 23 (1), 45–60.

Zetterqvist, L., 1991. Statistical estimation and interpretation of trends in water quality time series. Water Resour. Res. 27 (7), 1637–1648.

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

Zhang, X., Harvey, K.D., Hogg, W.D., Yuzyk, T.R., 2001. Trends in Canadian streamflow. Water Resour. Res. 37 (4), 987–998.