ELSEVIER

Journal of Hydrology 204 (1998) 182–196

Journal of Hydrology

A modified Mann–Kendall trend test for autocorrelated data

Khaled H. Hamed, A. Ramachandra Rao*

Purdue University, West Lafayette, IN 47907, USA

Received 12 March 1997; accepted 29 September 1997

Abstract

One of the commonly used tools for detecting changes in climatic and hydrologic time series is trend analysis. A number of statistical tests exist to assess the significance of trends in time series. One of the commonly used non-parametric trend tests is the Mann–Kendall trend test. The null hypothesis in the Mann–Kendall test is that the data are independent and randomly ordered. However, the existence of positive autocorrelation in the data increases the probability of detecting trends when actually none exist, and vice versa. Although this is a well-known fact, few studies have addressed this issue, and autocorrelation in the data is often ignored. In this study, the effect of autocorrelation on the variance of the Mann–Kendall trend test statistic is discussed. A theoretical relationship is derived to calculate the variance of the Mann–Kendall test statistic for autocorrelated data. The special cases of AR(1) and MA(1) dependence are discussed as examples. An approximation to the theoretical relationship is also presented in order to reduce computation time for long time series. Based on the modified value of the variance of the Mann–Kendall trend test statistic, a modified non-parametric trend test which is suitable for autocorrelated data is proposed. The accuracy of the modified test in terms of its empirical significance level was found to be superior to that of the original Mann–Kendall trend test without any loss of power. The modified test is applied to rainfall as well as streamflow data to demonstrate its performance as compared to the original Mann–Kendall trend test. © 1998 Elsevier Science B.V.

Keywords: Time series analysis; Trend analysis; Autocorrelation; Rainfall; Streamflow

1. Introduction

Trend detection in hydrologic and water quality time series has received considerable attention in the recent past. In a number of studies on water quality data in lakes and streams (Lettenmaier, 1976, 1988; Hirsch et al., 1982; Van Belle and Hughes, 1984; Hirsch and Slack, 1984; Hipel et al., 1988; Taylor and Loftis, 1989; Zetterqvist, 1991; Bouchard and Haemmerli, 1992; Yu et al., 1993) and streamflow data (World Meteorological Organization, 1988; Mitosek, 1992; Chiew and McMahon, 1993; Burn, 1994) a number of parametric and non-parametric tests have been applied for trend detection. Both parametric and non-parametric tests are commonly used. Parametric trend tests are more powerful than non-parametric ones, but they require data to be independent and normally distributed. On the other hand, non-parametric trend tests require only that the data be independent and can tolerate outliers in the data.

One of the widely used non-parametric tests for detecting trends in the time series is the Mann–Kendall test (Mann, 1945; Kendall, 1955). The Mann–Kendall trend test is derived from a rank correlation test for two groups of observations

* Corresponding author.

PII S0022-1694(97)00125-X

proposed by Kendall (1955). In the Mann–Kendall trend test, the correlation between the rank order of the observed values and their order in time is considered. The null hypothesis for the Mann–Kendall test is that the data are independent and randomly ordered, i.e. there is no trend or serial correlation structure among the observations. However, in many real situations the observed data are autocorrelated. The autocorrelation in observed data will result in misinterpretation of trend tests results. Cox and Stuart (1955) state that: 'Positive serial correlation among the observations would increase the chance of significant answer, even in the absence of a trend.' A closely related problem that has been studied is the case where seasonality exists in the data (Hirsch et al., 1982). By dividing the observations into separate classes according to seasons and then performing the Mann–Kendall trend test on the sum of the statistics from each season, the effect of seasonality can be eliminated. This modification is called the seasonal Kendall test (Hirsch et al., 1982; Hirsch and Slack, 1984; Zetterqvist, 1991). Although the seasonal test eliminates the effect of dependence between seasons, it does not account for the correlation in the series within seasons (Hirsch and Slack, 1984). The same problem exists when yearly data are analyzed, since they are often significantly autocorrelated.

In this paper, theoretical results are presented about the evaluation of the mean and variance of the Mann–Kendall trend test statistic in the presence of autocorrelation. Based on these theoretical results, as well as on an empirical approximation, a modified Mann–Kendall trend test which is robust in the presence of autocorrelation is suggested and its empirical significance level and power are investigated. The test is applied to rainfall and streamflow data to demonstrate its performance.

This paper is organized as follows. First, the original Mann–Kendall trend test is outlined in Section 2, and an alternative procedure is developed for evaluating the variance of the test statistic. The effect of autocorrelation on the Mann–Kendall trend test is then investigated in Section 3. The special cases of MA(1) and AR(1) autocorrelation are discussed in Sections 4 and 5. In Sections 6 and 7 computational considerations are discussed and an approximation is given for evaluating the variance of the Mann–Kendall trend test statistic for autocorrelated data. The asymptotic normality of the test statistic for the case of autocorrelated data is discussed in Section 8. In Sections 9 and 10, a modified Mann–Kendall test is proposed and its accuracy and power are investigated. An application to rainfall and streamflow data is given in Section 11 to demonstrate the performance of the test.

2. The original Mann–Kendall trend test

The rank correlation test (Kendall, 1955) for two sets of observations $X = x_1, x_2, \cdots, x_n$ and $Y = y_1, y_2, \cdots, y_n$ is formulated as follows. The statistic S is calculated as in Eq. (1):

$$S = \sum_{i<j} a_{ij} b_{ij} \tag{1}$$

where

$$a_{ij} = \operatorname{sgn}(x_j - x_i) = \begin{cases} 1 & x_i < x_j \\ 0 & x_i = x_j \\ -1 & x_i > x_j \end{cases} \tag{2}$$

and b_{ij} is similarly defined for the observations in Y. Under the null hypothesis that X and Y are independent and randomly ordered, the statistic S tends to normality for large n, with mean and variance given by:

$$E(S) = 0 \tag{3}$$

$$\operatorname{var}(S) = n(n-1)(2n+5)/18 \tag{4}$$

If the values in Y are replaced with the time order of the time series X, i.e. $1, 2, \cdots, n$, the test can be used as a trend test (Mann, 1945). In this case, the statistic S reduces to that given in Eq. (5):

$$S = \sum_{i<j} a_{ij} = \sum_{i<j} \operatorname{sgn}(x_j - x_i) \tag{5}$$

with the same mean and variance as in eqns (3) and (4). Kendall (1955) gives a proof of the asymptotic normality of the statistic S. The significance of trends is tested by comparing the standardized test statistic $Z = S/[\operatorname{var}(S)]^{0.5}$ with the standard normal variate at the desired significance level.

The derivation of the mean and variance of S is discussed in detail by Kendall (1955), Chapter 5). A slightly different approach, applicable to the

autocorrelated case, is developed in this paper. If X is normally distributed with mean μ and variance σ^2, then $(x_j - x_i)$ will also be normally distributed with mean zero and variance $2\sigma^2$. We thus have

$$E(S)=E\left[\sum_{i<j} a_{ij}\right]=E\left[\sum_{i<j}\mathrm{sgn}(x_j-x_i)\right]$$
$$=\sum_{i<j}E[\mathrm{sgn}(x_j-x_i)]=P(x_j-x_i>0)$$
$$-P(x_j-x_i<0)=0 \quad (6)$$

The result in Eq. (6) follows from the fact that $(x_j - x_i)$ is normally distributed with mean zero and variance $2\sigma^2$, which is symmetric around the origin. The variance of S is given by:

$$\mathrm{var}(S)=E(S^2)=E\left[\sum_{i<j} a_{ij}\right]^2=E\left[\sum_{i<j_{k<l}} a_{ij}a_{kl}\right] \quad (7)$$

For a time series with n observations, the sum in Eq. (7) involves $n^2(n-1)^2/4$ terms (Kendall, 1955). It can be shown that if the elements in X are independent and randomly ordered, the sum in Eq. (7) reduces to the three terms in Eq. (8) (Kendall, 1955):

$$\mathrm{var}(S)=\binom{n}{2}E(a_{ij}^2)+6\binom{n}{3}E(a_{ij}a_{ik})$$
$$+6\binom{n}{4}E(a_{ij}a_{kl}) \quad (8)$$

The first expectation in Eq. (8) accounts for terms with identical suffixes i and j. The second expectation in Eq. (8) accounts for terms with one common suffix. The third expectation in Eq. (8) accounts for terms with distinct suffixes. For $i \neq j$, $a_{ij}^2 = 1$ and thus $E(a_{ij}^2)$ is always equal to unity. Also, under the null hypothesis, x_i, x_j, x_k, x_l are independent for different suffixes i, j, k, l, so that $Y = (x_j - x_i)$ and $Z = (x_l - x_k)$ are also independent and $E(a_{ij}a_{kl}) = E(a_{ij})E(a_{kl}) = 0$, similar to Eq. (6). Now for the middle term in the RHS of Eq. (8), consider $Y = (x_j - x_i)$ and $Z = (x_k - x_i)$, each of which is $N(0,2\sigma^2)$. Together, Y and Z are jointly normal with correlation $\rho = \pm 0.5$ depending on the order of the common suffix in Y and Z. It can be shown that for three given distinct indices i, j and k, the reverse order of the common suffix ($a_{ij}a_{ki}$ or $a_{ji}a_{ik}$) occurs one-third of the time, giving rise to negative correlation, while the correct order ($a_{ij}a_{ik}$ or $a_{ji}a_{ki}$) occurs two-thirds of the time, giving rise to positive correlation. For example if $n = 3$, the sum involves the terms $a_{21}a_{31}$, $a_{21}a_{32}$ and $a_{31}a_{32}$, from which the first and third terms are in the correct order while the second term is in the reverse order. Since a_{ij} for $i \neq j$ takes only the values 1 or −1, the expected value of $a_{ij}a_{ik}$ is given by:

$$E(a_{ij}a_{ik})=P(a_{ij}a_{ik}=1)-P(a_{ij}a_{ik}=-1) \quad (9)$$

The difference in probabilities in Eq. (9) is equal to the difference between the probabilities of positive and negative quadrants, which is given by Kendall and Stuart (1976) for the bivariate normal distribution as in Eq. (10):

$$E(a_{ij}a_{ik})=\frac{2}{\pi}\sin^{-1}(\rho) \quad (10)$$

In Eq. (10), ρ is the correlation between the two binormal variates Y and Z. For $\rho = \pm 0.5$, Eq. (10) reduces to $E(a_{ij}a_{ik}) = \pm 1/3$. Since the positive value of ρ occurs two-thirds of the time while the negative value of ρ occurs one-third of the time, we will have $E(a_{ij}a_{ik}) = (1/3 \cdot 2/3 - 1/3 \cdot 1/3) = 1/9$. The sum in Eq. (7) can now be calculated using only three terms in the RHS of Eq. (8), where $E(a_{ij}^2) = 1$, $E(a_{ij}a_{ik}) = 1/9$, $E(a_{ij}a_{kl}) = 0$:

$$\mathrm{var}(S)=\binom{n}{2}\cdot 1+6\binom{n}{3}\cdot\frac{1}{9}+6\binom{n}{4}\cdot 0$$
$$=\frac{n(n-1)(2n+5)}{18} \quad (11)$$

which is the same expression given by Kendall (1955).

The assumption of normally distributed X was used to derive Eq. (11), but in fact the test is non-parametric and does not depend on the distribution of X. In fact, Kendall's proof does not take the distribution of X into account, but rather assumes that all a_{ij} are equally probable. In order to see that the variance of S does not depend on the distribution of X, consider the identity in Eq. (12),

$$a_{ij}=\mathrm{sgn}(x_j-x_i)=\mathrm{sgn}[R(x_j)-R(x_i)] \quad (12)$$

where $R(x_i)$ is the rank of x_i. According to Eq. (12) the value of a_{ij} does not depend on the actual distribution of X, because if the data were to be transformed to normality using a suitable technique, the

ranks of X values would not change. Indeed, all three expectations used in Eq. (11) to evaluate the variance of S do not depend on the distribution of X. Therefore, transformation of the data to normal, although possible (cf. Grigoriu, 1995), would not change the expected value of the statistic S or its variance. It is understood that the above comments apply only to monotonic transformations, since non-monotonic transformations would destroy trend information.

3. The effect of autocorrelation

A numerical example is given to illustrate the effect of serial correlation on the Mann–Kendall trend test. Fig. 1 shows two time series X and Y each of length $n = 100$ observations. Visual inspection of the two time series would not indicate a large difference in the apparent trends for the two series. In fact, series X is stationary white noise, while series Y is generated as an AR(1) series with $\phi = 0.4$ using series X as the input noise. Thus both series are stationary without trend. The test statistics for the two series from Eq. (5) are $S(X) = 557$ and $S(Y) = 793$. With $n = 100$, the variance from Eq. (4) is given by var(S) = 112 750 in both cases. The standardized test statistics are $Z(X) = S(X)/[\text{var}(S)]^{0.5} = 1.66$ and $Z(Y) = S(Y)/[\text{var}(S)]^{0.5} = 2.36$. This result would indicate that at 5% significance level ($Z_{cr} = 1.96$), series Y has a significant positive trend. In fact this result is only due to the effect of autocorrelation between Y values, since $Z(X)$ for the uncorrelated series X is not significant. The reason for $Z(Y)$ being larger is that the variance of S is underestimated when the data are positively autocorrelated. This aspect is discussed later.

In the case of autocorrelation between the values of X, the second and third terms in Eq. (8) will be different than they are for the uncorrelated case. In particular, the third expectation is no longer equal to zero. Starting with normally distributed X, consider again $Y = x_j - xi$ and $Z = x_l - x_k$. For convenience, assume that $\text{var}(x_i) = 1$ and $\text{cov}(x_i, x_j) = \rho(j - i)$. Note that ρ is a symmetric function, $\rho(i - j) = \rho(j - i)$ and $\rho(0) = 1$. We thus have

$$\text{var}(Y) = 2 - 2\rho(j - i) \tag{13}$$

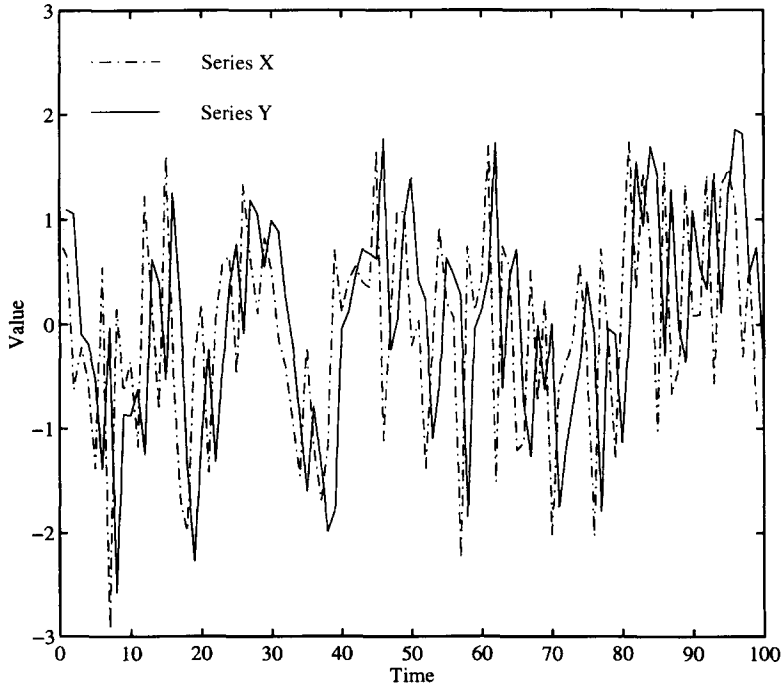

Fig. 1. The effect of autocorrelation on the Mann-Kendall test.

$$\text{var}(Z) = 2 - 2\rho(l-k) \tag{14}$$

$$\text{cov}(Y,Z) = \rho(j-l) - \rho(i-l) - \rho(j-k) + \rho(i-k) \tag{15}$$

The correlation between Y and Z is thus given by:

$$\text{corr}(Y,Z) = r_{ijkl} = \frac{\rho(j-l) - \rho(i-l) - \rho(j-k) + \rho(i-k)}{2\sqrt{[1-\rho(j-i)][1-\rho(l-k)]}} \tag{16}$$

Since X is normally distributed, Y and Z will be distributed as bivariate normal with correlation given by Eq. (16). The result in Eq. (10) can therefore be used to get the expression for $E(a_{ij}a_{kl})$ in Eq. (17):

$$E(a_{ij}a_{kl}) = \frac{2}{\pi}\sin^{-1}(r_{ijkl}) \tag{17}$$

The expected value of S is given by Eq. (18), since $x_i - x_j$ is $N(0,2 - 2\rho(j - i))$ which is still symmetric around zero:

$$E(S) = E\left(\sum_{i<j} a_{ij}\right) = \sum_{i<j} E(a_{ij}) = 0 \tag{18}$$

The variance of S is given by:

$$\text{var}(S) = E(S^2) = E\left[\sum_{i<j_{k<l}} S_{ij}a_{kl}\right] = \sum_{i<j_{k<l}} E(a_{ij}a_{kl}) \tag{19}$$

where $E(a_{ij}a_{kl})$ is given by Eq. (17).

In general, the distribution of X may not necessarily be normal in real data. The expected value of S will not change for different distributions. This is due to the fact that $x_i - x_j$ is always symmetric around zero, regardless of the distribution of X. In fact, the variance of S will not depend on the distribution of X either, for the same reasons mentioned in Section 2. However, in order to evaluate the variance of S for X with an arbitrary distribution, two pieces of information are needed. The first is the autocorrelation between the data, and the second is the equivalent of the relationship in Eq. (10) for that particular distribution. Such relationships are not available for distributions other than the bivariate normal distribution. Fortunately, data need not be transformed to normal, for all that is sought is the autocorrelation structure of the normalized data, not the transformed data themselves. An estimate of the autocorrelation of the normalized data can be obtained by first calculating the autocorrelation between the ranks of the observations, denoted $\rho_S(i)$. Since the ranks of the observations are used, $\rho_S(i)$ is independent of the distribution of X. The expression in Eq. (20), given by Kendall (1955), which relates the parent correlation ρ for two normal samples to their rank correlation ρ_S, is then used to transform rank autocorrelation to normalized data autocorrelation:

$$\rho(i) = 2\sin\left(\frac{\pi}{6}\rho_S(i)\right) \tag{20}$$

Using the values of $\rho(i)$ in Eq. (20) with eqns (16)–(19), the correct variance of S can be calculated independent of the distribution of X, by using the autocorrelation of the ranks. This result is applicable to any general autocorrelation structure in the data. Next we give some particular results for specific forms of the autocorrelation function $\rho(i)$ as examples.

4. The MA(1) model

The MA(1) model is in the form of:

$$x_t = e_t + \theta e_{t-1} \tag{21}$$

where e_t is $N(0,\sigma_e^2)$. In this case the autocorrelation function $\rho(i)$ is given by:

$$\rho(i) = \begin{cases} 1 & i = 0 \\ \dfrac{\theta}{(1+\theta^2)} & i = 1 \\ 0 & i > 1 \end{cases} \tag{22}$$

Due to the fact that $\rho(i)$ vanishes for $i > 1$, many of the terms in the sum in the RHS of Eq. (19) will vanish, leaving only 16 terms to be evaluated. These 16 terms are the result of different combinations of the terms $E(a_{ij}a_{kl}) = E[\text{sgn}(x_j - x_i)\cdot\text{sgn}(x_l - x_k)]$. For example, if $j = l$ and $i = k$, then from Eq. (16) $r_{ijkl} = 1$; if $j = l$, $|i - k| > 1$, $|i - j| > 1$ and $|k - j| > 1$, then $r_{ijkl} = 0.50$; and if $j = l$, $|i - k| = 1$, $|i - j| > 1$ and $|k - j| > 1$, then $r_{ijkl} = (1 + \rho)/2$. Other terms can be identified in a similar manner by considering the relationship between the different suffixes. By using eqns (16)–(20) the variance of S in the MA(1) case is given by:

$$\text{V}(S) = \sum_{j=1}^{16} n_j \cdot \frac{2}{\pi}\sin^{-1} r_j \tag{23}$$

In Eq. (23) n_j and their corresponding r_j are functions

Table 1
Theoretical values of V(*S*)/var(*S*) for the MA(1) model for different values of θ and n

n	θ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
10	1.13	1.25	1.37	1.47	1.55	1.62	1.66	1.69	1.71	1.71
20	1.16	1.30	1.44	1.55	1.65	1.72	1.77	1.80	1.82	1.82
30	1.17	1.32	1.47	1.59	1.68	1.76	1.81	1.84	1.86	1.87
50	1.17	1.34	1.49	1.62	1.72	1.79	1.85	1.88	1.90	1.90
100	1.18	1.35	1.51	1.64	1.74	1.82	1.88	1.91	1.93	1.93
120	1.18	1.36	1.51	1.64	1.75	1.83	1.88	1.91	1.93	1.94
n	θ									
	−0.1	−0.2	−0.3	−0.4	−0.5	−0.6	−0.7	−0.8	−0.9	−1.0
10	0.87	0.76	0.65	0.57	0.50	0.45	0.41	0.39	0.38	0.37
20	0.85	0.70	0.57	0.46	0.38	0.31	0.26	0.24	0.22	0.22
30	0.83	0.68	0.54	0.42	0.33	0.26	0.21	0.18	0.16	0.16
50	0.83	0.66	0.51	0.39	0.29	0.22	0.17	0.13	0.12	0.11
100	0.82	0.65	0.49	0.36	0.26	0.18	0.13	0.10	0.08	0.07
200	0.82	0.64	0.49	0.36	0.26	0.18	0.12	0.09	0.07	0.07

of the time series length n and $\rho(i)$ in Eq. (22) (Hamed and Rao, 1997).

Table 1 gives the theoretical values of V(*S*)/var(*S*) for different values of θ and for n = 10, 20, 30, 50, 100 and 120, based on V(*S*) given by Eq. (23) and var(*S*) given by Eq. (4) for the uncorrelated case. To verify the validity of the theoretical values in Table 1, numerical simulation was used. One hundred sets, each with 200 realizations of the time series in Eq. (21) were obtained and the variance of S was calculated by using the simulated time series. The results from these simulations were found to confirm the validity of the theoretical approach (Hamed and Rao, 1997).

As can be seen from Table 1, positive autocorrelation will result in the increase of V(*S*) and thus the variance of S will be underestimated by the original var(*S*). Thus, if the original Mann–Kendall test is used for testing trends in positively autocorrelated data, it will indicate significant trends, when actually no trends exist. Negative autocorrelation in the data, on the other hand, produces an opposite effect by reducing the variance of S.

5. The AR(1) model

The AR(1) model is of the form of:

$$x_t = \phi x_{t-1} + e_t \tag{24}$$

In this case, the autocorrelation function is given by:

$$\rho(i) = \phi^{|i|} \tag{25}$$

From the results presented in the previous section, it is seen that the number of terms to be evaluated in the sum in the RHS of Eq. (19) reduces to only 16 terms for the MA(1) model, because $\rho(i)$ vanishes for $i > 1$. This is not, however, the cae for an AR(1) model or a model with a general autocorrelation structure, and all $n^2(n-1)^2/4$ terms in Eq. (19) will be non-zero. Expectations needed for the sum in the RHS of Eq. (19) must be calculated for each term using eqns (16) and (17). Each set of subscripts i, j, k and l will have a different value of r_{ijkl} in Eq. (16) depending on the values of the indices. Table 2 gives the theoretical values of V(*S*)/var(*S*) in the AR(1) case, for different values of ϕ and for n = 10, 20, 30, 50, 100 and 120, where V(*S*) is calculated by using eqns (16), (17) and (19), and $\rho(i)$ in Eq. (16) is given by Eq. (25). The results from numerical simulation again confirm the validity of the theoretical approach (Hamed and Rao, 1997).

The same results obtained in the MA(1) case concerning the effect of autocorrelation are observed for the AR(1) model. Positive autocorrelation increases the variance of S and vice versa. However, the effect in the case of the AR(1) model is much larger than in the case of the MA(1) model, because the autocorrelation extends beyond the first lag for the AR(1)

Table 2
Theoretical values of V(*S*)/var(*S*) for the AR(1) model for different values of θ and n

n	ϕ								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	1.14	1.30	1.49	1.72	1.99	2.33	2.74	3.25	3.89
20	1.17	1.38	1.64	1.96	2.38	2.96	3.76	4.94	6.78
30	1.18	1.41	1.69	2.06	2.55	3.23	4.24	5.88	8.85
50	1.19	1.44	1.74	2.14	2.69	3.47	4.70	6.85	11.48
100	1.20	1.46	1.78	2.21	2.80	3.67	5.08	7.72	14.42
120	1.20	1.46	1.79	2.22	2.82	3.71	5.15	7.88	15.01
n	ϕ								
	−0.1	−0.2	−0.3	−0.4	−0.5	−0.6	−0.7	−0.8	−0.9
10	0.88	0.78	0.69	0.60	0.53	0.46	0.40	0.35	0.31
20	0.86	0.73	0.63	0.53	0.45	0.37	0.31	0.25	0.19
30	0.85	0.72	0.61	0.51	0.42	0.34	0.27	0.21	0.15
50	0.84	0.70	0.59	0.49	0.39	0.31	0.24	0.17	0.11
100	0.83	0.69	0.57	0.47	0.37	0.29	0.22	0.15	0.09
200	0.83	0.69	0.57	0.46	0.37	0.29	0.21	0.14	0.08

model. For example, the variance of S for AR(1) with $n = 120$ and $\phi = 0.9$ is 15 times as large as that for uncorrelated data.

6. Computational considerations

For an arbitrary autocorrelation function, similar to the case of the AR(1) model, the number of terms required to evaluate the variance of S by using Eq. (19) will be of order $n^4/4$. Therefore, time series with a large number of observations n will require a large number of calculations to evaluate the variance of S. This will not be suitable for use as a routine calculation, especially when a large number of time series are to be analyzed. Nevertheless, the values obtained by using the theoretical approach are valuable as a guide for selecting an approximation to the theoretical values which requires less computational effort. In the next section an empirical approximate formula is developed for calculating the variance of S for the case where autocorrelation exists in the data. The approximate formula requires a few calculations of order n at most, while providing a satisfactory approximation to the computationally demanding theoretical formula in Eq. (19).

7. Approximate formula for calculating V(*S*)

Bayley and Hammersley (1946) studied the variance of the mean of a sample of size n when the data are autocorrelated. In the case of uncorrelated data the variance of the mean is given by:

$$\mathrm{var}(\bar{x}) = \frac{\sigma^2}{n} \tag{26}$$

where σ^2 is the variance of X. However, when the data are autocorrelated the variance is underestimated by Eq. (26). The correct variance of the mean in the case of autocorrelated data is shown by Bayley and Hammersley (1946) to be given by:

$$\mathrm{V}(\bar{x}) = \frac{\sigma^2}{n_b^*} = \mathrm{var}(\bar{x}) \cdot \frac{n}{n_b^*} \tag{27}$$

where $\mathrm{var}(\bar{x})$ is given by Eq. (26) and n_b^* is given by the expression in Eq. (28), in which $\rho(j)$ is the autocorrelation function of the data and n is the actual number of observations.

$$\frac{1}{n_b^*} = \frac{1}{n} + \frac{2}{n^2} \sum_{j=1}^{n-1} (n-j)\rho(j) \tag{28}$$

The value of n_b^* is considered as an 'effective' number of observations, to account for the autocorrelation in the data.

By analogy to the case considered by Bayley and Hammersley, we suggest an empirical formula for calculating the variance of S in the case of

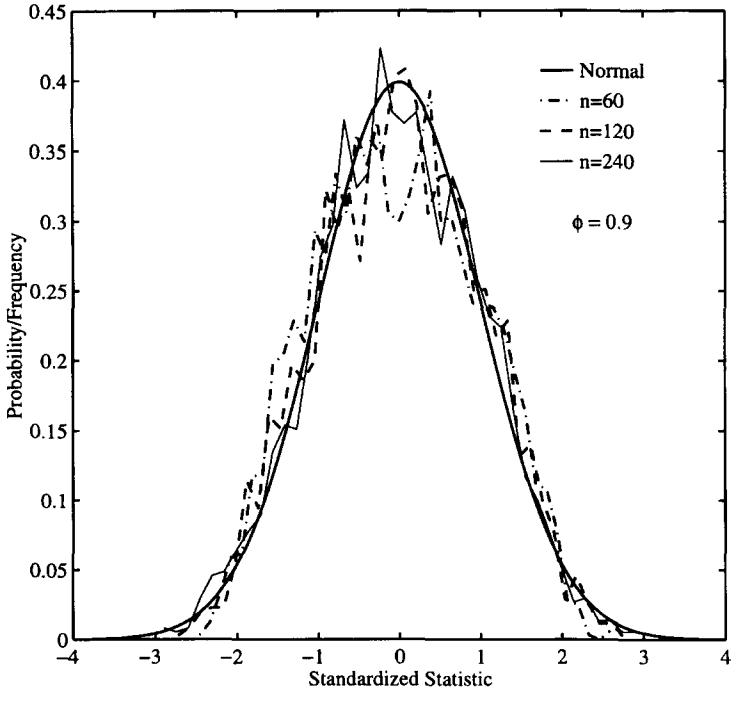

Fig. 2. The distribution of the statistic S as compared to the normal distribution for $\phi = 0.9$ and different sample sizes n.

autocorrelated data similar to that in Eq. (27):

$$V^*(S) = \text{var}(S) \cdot \frac{n}{n_S^*} = \frac{n(n-1)(2n+5)}{18} \cdot \frac{n}{n_S^*} \tag{29}$$

where n/n_S^* represents a correction due to the autocorrelation in the data. An expression is needed for evaluating n/n_S^*. Although suggested by Lettenmaier (1976), the use of the same expression in Eq. (28) for n_b^* to evaluate n_S^* does not produce values of $V^*(S)$ that are comparable to the theoretical values in the cases of the MA(1) and AR(1) models in Tables 1 and 2, respectively. A closer look at the structure of the statistic S indicates that the use of the autocorrelation between the ranks is more appropriate that using the autocorrelation from the data. Also, the number of terms involved in calculating S is not n, as in the case of Bayley and Hammersley, but rather $n(n-1)$. After considering a number of alternative forms to Eq. (28), the best approximation to the theoretical values was obtained by using n/n_S^* given by the empirical expression in Eq. (30):

$$\frac{n}{n_S^*} = 1 + \frac{2}{n(n-1)(n-2)} \times \sum_{i=1}^{n-1} (n-i)(n-i-1)(n-i-2)\rho_S(i) \tag{30}$$

where n is the actual number of observations and $\rho_S(i)$ is the autocorrelation function of the ranks of the observations. One advantage of using the approximation in Eq. (30) is that by using the ranks of the observations, the variance of S can be evaluated using eqns (29) and (30) without the need for either the normalized data or their autocorrelation function, as was the case with Eq. (19). Values of the variance calculated by using Eq. (29) were found to be in good agreement with the theoretical values in Tables 1 and 2 (Hamed and Rao, 1997). The accuracy of the approximation given by Eq. (29) was also found to improve as n increased. The approximation in Eq. (29) can therefore be used to obtain accurate estimates of the variance of S, based on the autocorrelation between the ranks of the observations, with much less computational effort.

8. Asymptotic normality

A mathematical proof of the asymptotic normality of S was not attempted in the case where the data are autocorrelated. However, the distribution of S was empirically investigated through numerical simulation. Fig. 2 shows the empirical distribution of the standardized normal test statistic $Z = S/[\text{var}(S)]^{0.5}$ for an AR(1) model with $\phi = 0.9$ and $n = 60$, 120 and 240 using 2500 data sets. Although $\phi = 0.9$ represents an extreme case, it is clear from Fig. 2 that as n increases the distribution of S tends to normality even for ϕ as large as 0.9. The tail probability values for various AR(1) and ARMA(1,1) simulated series were also found to approach those of the normal distribution as n increases (Hamed and Rao, 1997). From these results we conclude that the asymptotic normality assumption for S does not seem to be affected by the existence of autocorrelation in the data.

9. A modified Mann–Kendall test

A modified version of the Mann–Kendall test which is robust in the presence of autocorrelation is proposed, based on the modified variance of S given by Eq. (19) and its approximation in Eq. (29). The autocorrelation between the ranks of the observations, $\rho_S(i)$, is first evaluated. The values of $\rho_S(i)$, however, must be calculated after subtracting a suitable non-parametric trend estimator (Sen, 1968; Zetterqvist, 1991). Sample estimates of the autocorrelation have a variance of order $1/n$ (Kendall, 1955). Due to the nature of the calculations in eqns (19) and (29), which involve a large number of terms, it was found that insignificant values of $\rho_S(i)$ will have an adverse effect on the accuracy of the estimated variance of S. Therefore, only significant $\rho_S(i)$ values are used in Eq. (30). This is achieved by requiring a suitable pre-set significance level for the autocorrelations to be included in the calculations, which can be taken equal to that of the test.

There are two important properties of a statistical test that are investigated to evaluate its performance. The first property is the empirical significance level of the test, defined as the probability of rejecting the null hypothesis H_0 of no trend when in fact H_0 is true. This probability is the percentage of samples rejected by the test under H_0 of no trend when in fact H_0 is true. This probability is the percentage of samples rejected by the test under H_0, and should be equal to the nominal significance level α of the test, provided that the normal distribution with mean zero and variance equal to var(S) is the correct distribution of S. The second property is the power of the test, defined as the probability of rejecting H_0 when the alternative hypothesis is true. This probability is the percentage of samples rejected by the test when a trend of a certain slope exists in the data. The power of a good test should increase rapidly as the slope of the trend departs from zero. A number of numerical simulations were used to evaluate the performance of the modified test compared to the original Mann–Kendall trend test. Samples of size 2000 from uncorrelated data, AR(1) and ARMA(1,1) models were generated. The AR(1) model is characterized by ϕ as in Eq. (24). The ARMA(1,1) model is of the form of:

$$x_t = \phi x_{t-1} + e_t + \theta e_{t-1} \tag{33}$$

The ARMA(1,1) model is characterized by ϕ and $\rho(1)$, where $\rho(1)$ is given by:

$$\rho(1) = \frac{(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2} \tag{34}$$

Table 3 gives the empirical significance level of the original Mann–Kendall trend test. The empirical significance levels in Table 3 are in large error compared to the nominal values of the significance level α. This is due to the underestimation of the variance of S when the autocorrelation in the data is ignored. For example, for $n = 120$ and AR(1) with $\phi = 0.6$ for a nominal significance level of $\alpha = 0.05$, one would expect to reject 50 samples out of 1000, while the original test rejects 186 samples. As expected, the error in the significance level increases as the autocorrelation increases, as seen from the results in Table 3.

Table 4 gives the empirical significance level of the modified test. The variance of S was calculated using the approximation in Eq. (29), where $\rho_S(i)$ is calculated from $\rho(i)$ using the inverse of Eq. (20). A significance level of $\alpha = 0.1$ for the autocorrelation of the ranks $\rho_S(i)$ was used, which produced the best overall empirical significance levels. Comparing the

Table 3
The original Mann-Kendall test empirical significance level (2000 samples)

n	ϕ	$\rho(1)$	Nominal significance level α				
			0.010	0.020	0.050	0.100	0.200
60	0.0	0.0	0.011	0.020	0.052	0.105	0.199
	0.2	0.2	0.028	0.045	0.087	0.144	0.233
	0.4	0.2	0.065	0.091	0.141	0.199	0.283
		0.4	0.057	0.083	0.142	0.205	0.282
	0.6	0.2	0.134	0.163	0.227	0.276	0.346
		0.4	0.133	0.166	0.221	0.269	0.348
		0.6	0.110	0.139	0.201	0.259	0.327
	0.9	0.2	0.299	0.328	0.363	0.386	0.423
		0.4	0.279	0.300	0.335	0.369	0.405
		0.6	0.301	0.323	0.362	0.387	0.423
120	0.0	0.0	0.006	0.021	0.055	0.104	0.207
	0.2	0.2	0.023	0.039	0.079	0.131	0.234
	0.4	0.2	0.076	0.107	0.169	0.229	0.322
		0.4	0.060	0.086	0.143	0.202	0.290
	0.6	0.2	0.139	0.173	0.233	0.288	0.358
		0.4	0.131	0.163	0.209	0.260	0.341
		0.6	0.110	0.134	0.186	0.244	0.326
	0.9	0.2	0.299	0.324	0.359	0.385	0.427
		0.4	0.292	0.311	0.349	0.386	0.428
		0.6	0.301	0.326	0.356	0.387	0.426
240	0.0	0.0	0.011	0.022	0.052	0.104	0.209
	0.2	0.2	0.026	0.046	0.148	0.206	0.298
	0.4	0.2	0.075	0.098	0.148	0.206	0.298
		0.4	0.060	0.082	0.142	0.207	0.302
	0.6	0.2	0.148	0.185	0.239	0.293	0.368
		0.4	0.141	0.171	0.227	0.275	0.346
		0.6	0.122	0.160	0.212	0.272	0.364
	0.9	0.2	0.290	0.315	0.351	0.382	0.414
		0.4	0.296	0.321	0.353	0.386	0.426
		0.6	0.301	0.326	0.356	0.387	0.426

Control limits $\alpha \pm 2[\alpha(1-x)/2000]^{1/2}$ for the empirical level; for a nominal level α are: $\alpha = 0.01$, 0.006-0.014; $\alpha = 0.02$, 0.014-0.026; $\alpha = 0.05$, 0.04-0.06; $\alpha = 0.10$, 0.087-0.113; $\alpha = 0.20$, 0.182-0.218.

results in Tables 3 and 4, it is seen that the modified test gives more accurate empirical levels, compared to the nominal significance level α, than does the original Mann–Kendall test. For example, when testing at a nominal significance level $\alpha = 0.02$ one would expect to reject 20 samples out of 1000. For $n = 120$, $\phi = 0.6$ and $\rho(1) = 0.4$ the number of samples rejected by the original Mann–Kendall test in Table 3 is 163, which is much larger than 20, while the number of samples rejected by the modified test in Table 4 is 31, which is closer to 20. The empirical significance levels in Table 4, however, become less accurate compared to the nominal significance levels when large autocorrelations are present and the number of observations n is small. This is not suprising because by almost any measure or technique it is very hard to distinguish strong persistence from trend (Hirsch and Slack, 1984).

10. Power of the modified test

The proposed modification of the Mann–Kendall trend test is based on the assumption that data are autocorrelated, and therefore the autocorrelation is estimated from the data. When the data are actually independent, the assumption of autocorrelation may lead, in some cases, to failure to identify true trends, thus reducing the power of the test. This is due to the uncertainties in evaluating the autocorrelation in

Table 4
Modified test empirical significance level (2000 samples)

n	ϕ	$\rho(1)$	Nominal significance level α				
			0.010	0.020	0.050	0.100	0.200
60	0.0	0.0	0.020	0.027	0.057	0.109	0.204
	0.2	0.2	0.022	0.038	0.069	0.121	0.221
	0.4	0.2	0.025	0.036	0.065	0.118	0.210
		0.4	0.024	0.041	0.077	0.136	0.234
	0.6	0.2	0.033	0.047	0.095	0.158	0.254
		0.4	0.029	0.046	0.095	0.158	0.254
		0.6	0.034	0.050	0.096	0.147	0.251
	0.9	0.2	0.081	0.118	0.180	0.247	0.325
		0.4	0.288	0.110	0.164	0.223	0.305
		0.6	0.089	0.125	0.188	0.246	0.330
120	0.0	0.0	0.009	0.024	0.059	0.108	0.209
	0.2	0.2	0.015	0.026	0.059	0.103	0.204
	0.4	0.2	0.017	0.029	0.165	0.128	0.233
		0.4	0.021	0.034	0.064	0.128	0.227
	0.6	0.2	0.026	0.038	0.077	0.126	0.244
		0.4	0.021	0.031	0.072	0.129	0.223
		0.6	0.026	0.042	0.071	0.124	0.216
	0.9	0.2	0.061	0.084	0.135	0.196	0.286
		0.4	0.054	0.077	0.126	0.193	0.275
		0.6	0.061	0.088	0.142	0.198	0.290
240	0.0	0.0	0.012	0.022	0.054	0.105	0.211
	0.2	0.2	0.013	0.022	0.054	0.103	0.211
	0.4	0.2	0.010	0.019	0.052	0.104	0.197
		0.4	0.018	0.031	0.061	0.113	0.222
	0.6	0.2	0.022	0.034	0.066	0.121	0.230
		0.4	0.019	0.034	0.063	0.124	0.224
		0.6	0.013	0.028	0.064	0.122	0.230
	0.9	0.2	0.035	0.047	0.093	0.153	0.249
		0.4	0.031	0.048	0.089	0.146	0.239
		0.6	0.034	0.053	0.110	0.157	0.245

Control limits $\alpha \pm 2[\alpha(1-x)/2000]^{1/2}$ for the empirical level; for a nominal level α are: $\alpha = 0.01$, 0.00-0.014; $\alpha = 0.02$, 0.014-0.026; $\alpha = 0.05$, 0.04-0.06; $\alpha = 0.10$, 0.087-0.113; $\alpha = 0.20$, 0.182-0.218.

the data, especially with small samples. For example, the seasonal Kendall test (Hirsch and Slack, 1984) is less powerful when the correlation between seasons is taken into account than when the correlation is neglected. Table 5 gives the power of the modified test, denoted by M, using 500 simulated samples of independent data, each of size $n = 60$, compared to the power of the original Mann–Kendall trend test, denoted by O. The trend in Table 5 is expressed as a percentage of the standard deviation of the process. It can be seen from Table 5 that the power of the proposed modified test is comparable to that of the original Mann–Kendall test. Therefore, there is no loss of power when the proposed modification is adopted to account for the autocorrelation in the data.

11. Application to hydrologic data

Two examples are given here for the application of the modified test to hydrologic data. The first example involves rainfall data, while the second example involves streamflow data. Fig. 3 shows two annual divisional average rainfall time series from the National Climatic Data Center database (National Climatic Data Center, 1991) for divisions 2 in Indiana (IN02) and 1 in Ohio (OH01). The time series extend from 1895 to 1992 with a length of 98 years. Both the original and modified tests were applied to the time series at the 5% significance level. To apply the modified test, a non-parametric trend estimate (Sen, 1968) is first subtracted from the time series

Table 5
Power of the modified test using 500 samples of sample size $n = 60$ of independent data compared to the power of the original Mann-Kendall test

Trend (%)	Test	Significance level				
		0.010	0.020	0.050	0.100	0.200
0.00	M	0.022	0.040	0.076	0.102	0.192
	O	0.008	0.026	0.062	0.102	0.192
0.50	M	0.060	0.102	0.184	0.292	0.446
	O	0.042	0.072	0.178	0.280	0.418
1.00	M	0.198	0.264	0.408	0.526	0.684
	O	0.162	0.232	0.378	0.506	0.662
1.50	M	0.378	0.498	0.634	0.760	0.862
	O	0.332	0.452	0.616	0.762	0.854
2.00	M	0.612	0.724	0.846	0.926	0.958
	O	0.560	0.684	0.824	0.914	0.956
2.50	M	0.820	0.900	0.958	0.984	0.994
	O	0.786	0.880	0.964	0.986	0.996
3.00	M	0.944	0.964	0.980	0.992	0.998
	O	0.940	0.966	0.984	0.990	0.998
3.50	M	0.984	0.994	0.996	0.998	1.000
	O	0.978	0.992	0.998	1.000	1.000
4.00	M	0.994	0.996	0.998	1.000	1.000
	O	0.996	0.996	0.996	1.000	1.000

and the autocorrelation between the ranks of the observations is calculated. Autocorrelations which are significant at the 5% level are then used for evaluating the modified variance of S using eqns (29) and (30). The significant autocorrelations are given in Table 6 for each of the time series, as well as the ratio between the original and modified variance and the values of the original and modified test statistics. It can be seen in Table 6 that the existence of negative autocorrelation in the two time series results in the reduction of the variance of S to 0.40 and 0.62 of their original values, respectively. This, in turn, results in the failure of the original test to identify significant trends in these two time series at the 5% significance level.

The second example involves an annual average riverflow time series from the Cedar River at Cedar Rapids, IA (USGS station No. 05464500) (United States Geological Survey, 1992). The time series extends from 1903 to 1992 with a length of 90 years

Table 6
Test results using the original modified tests for two NCDC divisional rainfall times series and a USGS riverflow time series

Site name	Significant autocorrelationat 5% level		Value of $V^*(S)/var(S)$	Original statistic Z	Modified statistic Z
	Lag i	$\rho(i)$			
IN02	1	−0.1997	0.40	1.71	2.69*
	19	−0.2020			
	1	−0.1888			
OH01	11	0.1797	0.62	1.69	2.14*
	12	−0.1942			
CEDAR	1	0.2474			
	9	0.2142	1.52	1.97*	1.60
	16	−0.2397			

*Value is significant at significance level $\alpha = 0.5$.

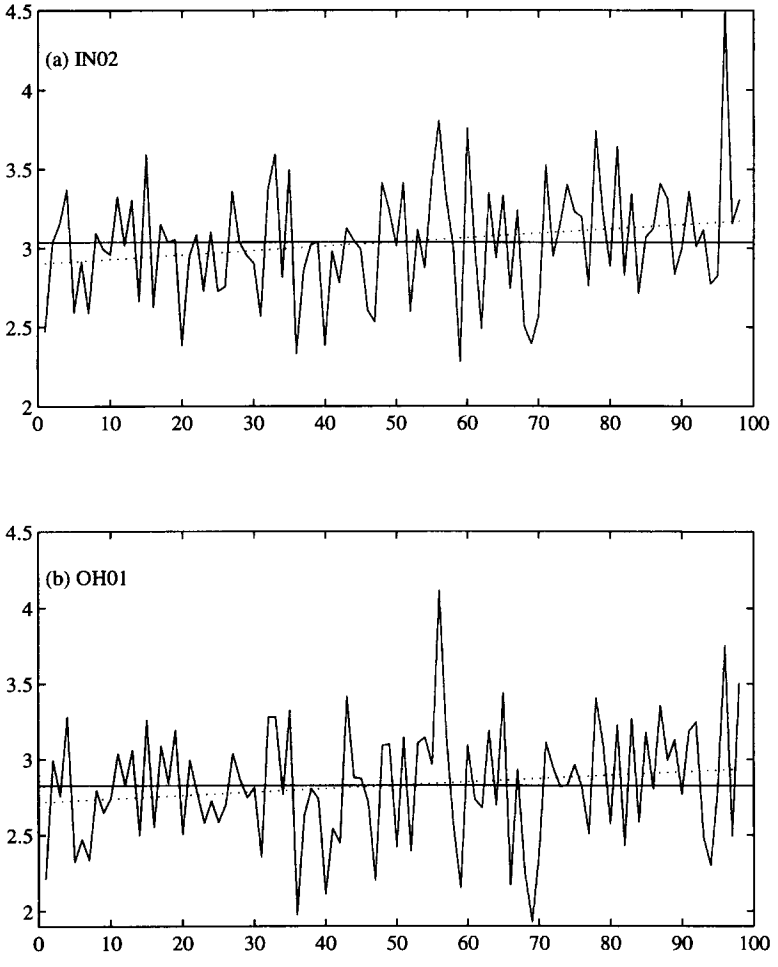

Fig. 3. NCDC divisional average time series (a) division 2 in Indiana, and (b) division 1 in Ohio.

and is shown in Fig. 4. Similar to the previous example, the modified variance of S is calculated using eqns (29) and (30) and autocorrelations which are significant at the 5% level. The significant autocorrelations are given in Table 6 for the time series, as well as the ratio between the original and modified variance and the values of the original and modified test statistics. The modified variance in this case was found to be 1.52 times the variance of the original test as given by Eq. (4). Hence, while the original test statistic of 1.97 indicates a significant trend at the 5% level, the modified test statistic of 1.6 is not significant at the same significance level when autocorrelation is taken into consideration. In this case, the existence of positive autocorrelation in the data resulted in the trend in this time series being falsely identified as significant at the 5% level, while in fact the trend is insignificant at that level.

It is clear from the above two examples that the existence of either positive or negative autocorrelation in time series will interfere with proper identification of significant trends. The proposed trend test offers a simple, easy to calculate modification to account for autocorrelation in the data. As shown in the previous sections, the modification does not affect the power of the test while offering more accurate significance levels.

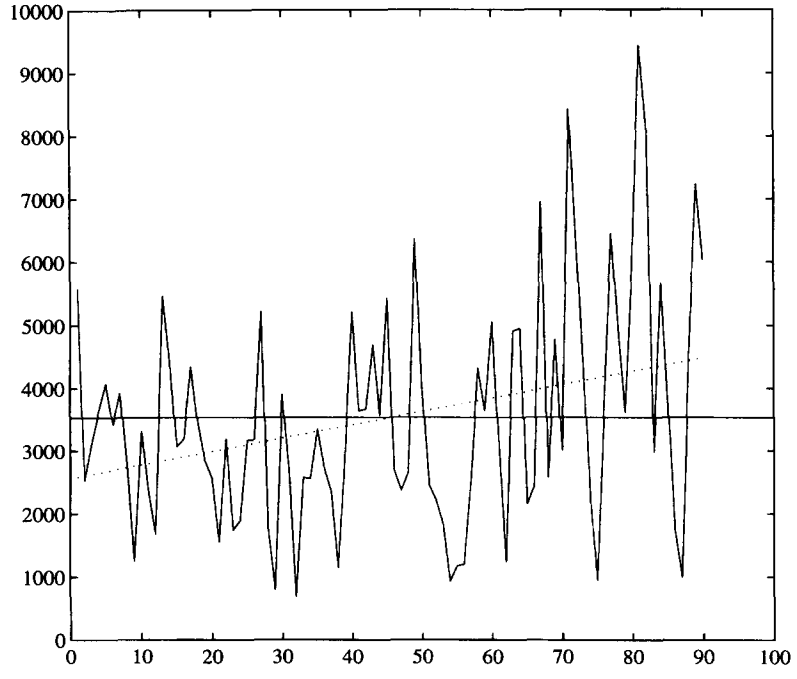

Fig. 4. Flow time series from USGS station No. 05464500, Cedar River at Cedar Rapids, IA.

12. Conclusions

A theoretical relationship is derived for evaluating the variance of the Mann–Kendall statistic S for data with autocorrelation. An approximation to the theoretical value of the invariance of S, which is useful for large samples, is also given. A modified trend test, which is robust in the presence of autocorrelation in the data is proposed based on the modified variance of S. The significance level and power of both the original and modified tests were compared through simulation. When autocorrelation exists in the data, the empirical significance levels of the original Mann–Kendall trend test are in large error compared to the nominal significance levels. On the other hand, the empirical significance levels of the proposed modified test are much closer to the correct nominal significance levels. The power of the proposed modified test is comparable to that of the original Mann–Kendall test when the data are actually independent. Examples for the application of the modified test to rainfall as well as streamflow time series were also presented, and the performance of the test was discussed.

References

Bayley, G.V., Hammersley, J.M., 1946. The 'effective' number of independent observations in an autocorrelated time series, J. Royal Stat. Soc., 8 (2), 184–197.

Bouchard, A., Haemmerli, J., 1992. Trend detection in water quality time series of LRTAP-Quebec network lakes, Water, Air, and Soil Pollution, 62, 89–110.

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

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

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

Grigoriu, M., 1995. "Applied Non-Gaussian Processes", Prentice-Hall, Englewood Cliffs, N.J.

Hamed, K.H., Rao, A.R., 1997. A trend test for autocorrelated data. Environmental and Hydrologic Engineering Report CE-EHE-97-7, School of Civil Engineering, Purdue University, West Lafayette, IN. 47906, USA.

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

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

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend

analysis for monthly water quality data, Water Resources Research, 18 (1), 107–121.

Kendall, M.G., 1955. Rank Correlation Methods. Griffin, London.

Kendall, M.G., Stuart, A., 1976. The Advanced Theory of Statistics, Vol. I: Distribution Theory. Griffin, London.

Lettenmaier, D.P., 1976. Detecting trends in water quality data from records with dependent observations, Water Resources Research, 12 (5), 1037–1046.

Lettenmaier, D.P., 1988. Multivariate nonparametric tests for trend in water quality, Water Resources Research, 24 (3), 505–512.

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

Mitosek, H.T., 1992. Occurrence of climate variability and change within the hydrologic time series: A statistical approach. CP-92-05, International Institute for Applied Systems Analysis, Laxenburg, Austria.

National Climatic Data Center (NCDC), 1991. Data selections from the disk resident database (DRD964x). National Climatic Data Center, Ashville, NC.

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

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

United States Geological Survey (USGS), 1992. USGS Daily Values. Computer Laser Optical Disks, EarthInfo, Boulder, CO.

Van Belle, G., Hughes, J.P., 1984. Nonparametric tests for trend in water quality, Water Resources Research, 20 (1), 127–136.

World Meteorological Organization (WMO), 1988. Analyzing long time series of hydrological data with respect to climate variability. WCAP-3, WMO/TD- No. 224, 12 pp.

Yu, Y., Zou, S., Whittemore, D., 1993. Nonparametric trend analysis of water quality data of rivers in Kansas, J. Hydrology, 150, 61–80.

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