Turkish J. Eng. Env. Sci.
27 (2003) , 247 – 251.


# The Power of Statistical Tests for Trend Detection

**Bihrat ÖNÖZ, Mehmetçik BAYAZIT**
*İstanbul Technical University, Faculty of Civil Engineering,*
*80626 Maslak, İstanbul-TURKEY*
*e-mail: onoz@itu.edu.tr*

Received 29.06.2002

**Abstract**

The existence of a trend in a hydrological time series is detected by statistical tests. The power of a test is the probability that it will reject a null hypothesis when it is false. In this study, the power of the parametric t-test for trend detection is estimated by Monte Carlo simulation for various probability distributions and compared with the power of the non-parametric Mann-Kendall test. The t-test has less power than the non-parametric test when the probability distribution is skewed. However, for moderately skewed distributions the power ratio is close to one. Annual streamflow records in various regions of Turkey are analyzed by the two tests to compare their powers in detecting a trend.

**Key words:** Trend detection, Power of a test, t-test, Mann-Kendall test.

## Introduction

The time series of some random variables exhibit a trend such that there is a significant change over time. The trend analysis of hydrological series is of practical importance because of the effects of global climate change. Statistical procedures are used for the detection of the gradual trends over time.

The purpose of trend testing is to determine if the values of a random variable generally increase (or decrease) over some period of time in statistical terms (Helsel and Hirsch, 1992). Parametric or non-parametric statistical tests can be used to decide whether there is a statistically significant trend.

All statistical tests involve two kinds of errors. These are the so-called type I error (rejecting the null hypothesis when it is true), and the type II error (not rejecting the null hypothesis when it is false). It is important to consider the power of a test, where power is defined as one minus the probability of type II error. A powerful test will reject a false hypothesis with a high probability. In this paper, the power of parametric and non-parametric tests for trend detection are compared for various probability distributions.

## Tests for Trend Detection

The null hypothesis $H_o$ that there is no trend, is to be tested against the alternative hypothesis $H_1$, that there is a trend. Parametric or non-parametric tests can be used for this purpose.

The parametric test considers the linear regression of the random variable Y on time X. The regression coefficient $b_1$ (or the Pearson correlation coefficient r) is computed from the data. It is known (Haan, 1977) that the statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{b_1}{s\,/\,\sqrt{SS_x}} \quad (1)$$

follows Student's t distribution with degrees of freedom n-2, where n is the sample size, s is the standard deviation of residuals, and $SS_x$ is the sums of squares of the independent variable (time in trend analysis). The hypothesis $H_o : \rho = 0$ (or $\beta_1 = 0$) is tested against the hypothesis $H_1 : \rho \neq 0$ (or $\beta_1 \neq 0$) at a chosen level of significance $\alpha$, where $\rho$ and $\beta_1$ are the population values, respectively, of the correlation coefficient and regression coefficient. The hypothesis that there is no trend is rejected when the t value computed by Eq. (1) is greater in absolute value than the critical value $t_{\alpha/2}$.

There are two non-parametric tests for trend analysis. The Mann-Kendall test is based on the statistic S. Each pair of observed values $y_i$, $y_j (i > j)$ of the random variable is inspected to find out whether $y_i > y_j$ or $y_i < y_j$. Let the number of the former type of pairs be P, and the number of the latter type of pairs be M. Then S is defined as

$$S = P - M \tag{2}$$

For $n > 10$, the sampling distribution of S is as follows. Z follows the standard normal distribution where

$$Z = \begin{cases} (S-1)/\sigma_s & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ (S+1)/\sigma_s & \text{if } S < 0 \end{cases} \tag{3}$$

$$\sigma_s = \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

There is a correction for ties when $y_i = y_j$ (Salas, 1993). The null hypothesis that there is no trend is rejected when the computed Z value is greater than $z_{\alpha/2}$ in absolute value.

Another non-parametric test for trend detection is the Spearman's rho test. This test will not be discussed here because Yue *et al.* (2002) showed that it provides results almost identical to those obtained for the Mann-Kendall test.

### Comparison of the Tests

Parametric tests assume that the random variable is normally distributed and homosedastic (homogeneous variance). Non-parametric tests make no assumption for probability distribution.

The t-test for trend detection is based on linear regression, and therefore checks only for a linear trend. There is no such restriction for the Mann-Kendall test.

The Mann-Kendall test is expected to be less affected by the outliers because its statistic is based on the sign of differences, not directly on the values of the random variable.

In general, parametric tests are more powerful for given n when the variable is normally distributed, but much less powerful when it is not, compared with the non-parametric tests (Hirsch *et al.*, 1991).

The power of a test can be determined only when the true situation is known. In a trend test, this requires knowledge of the trend. The probability of the rejection of a given trend (probability of type II error, $\beta$) can be computed by Monte Carlo simulation for a chosen level of significance $\alpha$. In fact $\alpha$ is the probability of type I error, or the probability of detecting a trend when no trend exists. The power equals $1 - \beta$.

Yue *et al.* (2002) computed the power of the Mann-Kendall test for detecting a trend by Monte Carlo simulation. They generated a large number of independent time series following a certain distribution (normal, generalized extreme value, Pearson type 3 and lognormal) for various sample sizes from 10 to 100 and of different variance. Then some particular linear trend scenarios were superimposed onto each of the generated series. The power of the test was computed for various significance levels.

The power of the test is an increasing function of the absolute slope of the trend, and of the significance level $\alpha$ for given n. The power approaches $\alpha$ as the slope of the trend goes to zero. As the sample size increases, the power of the test also increases such that the existence of a trend is more easily discerned. The power is a decreasing function of the coefficient of variation of a time series because the variation within a series masks the existence of a trend.

Yue *et al.* (2002) found that the power of the Mann-Kendall test is quite different for other distribution types when a trend exists. The EV3 distribution (type 3 generalized extreme value) has the highest power while the lognormal distribution has the lowest power. This is an interesting result showing that the power of the Mann-Kendall trend test is dependent on the distribution type, in contrast to common thinking that states that this test is rank based and would be distribution free. The power of the test is also dependent on the shape parameter of the probability distribution, such that it increases with the coefficient of skewness for the generalized extreme value and Pearson type 3 distributions.

The results of Yue *et al.* (2002) for the lognormal distribution are not in agreement with those of the other distributions that have roughly the same value of the coefficient of skewness. For this reason, their experiments are repeated. The results obtained differed from those of the authors for lognormal distribution. These results show that the power of the test for the lognormal distribution is comparable to the power for the other distributions with the same value of the coefficient of skewness. The modified results for this case will be used in the comparison with the power of the t-test.

Values of the shape factor k of the generalized extreme value distribution used in the simulation by Yue *et al.* (2002) are not realistic. Yue (personal communication) provided the results for some other values of k, which are used in this study for comparing the power of the two tests.

**Power of the Parametric t-Test**

In this study, the power of the t-test for trend analysis is investigated and compared with the power of the Mann-Kendall test. For this purpose, a Monte Carlo simulation is performed similar to that of Yue *et al.* (2002) described above.

Two thousand time series each of size n = 50 are generated with the mean $\mu = 1.0$ and coefficient of variation $C_v = 0.5$. Generated samples follow various probability distributions: normal (N), lognormal (LN), Gumbel, generalized extreme value (GEV) and Pearson type 3 (P3). The coefficient of skewness is $C_s = 0$ for the normal, $C_s = 1.625$ for the lognormal, $C_s = 1.14$ for the Gumbel, $C_s = 1.5$ (k = -0.05) or $C_s = 13.5$ (k = -0.3) for the extreme value and $C_s = 1.5$ for the Pearson type 3 distributions. Linear trends are superimposed onto each of the generated series such that the slope of the trend line is ± 0.0025, ± 0.005, ± 0.0075 or ± 0.01.

The t-test is applied to each series at the level of significance $\alpha = 0.05$. The power of the test is determined in each case as the percentage of the series for which the trend is detected (the null hypothesis of no trend is rejected).

Figure 1 shows the power of the t-test for various distributions as a function of the slope of the trend. The power increases with the absolute value of the slope, and is nearly the same for the equal positive and negative slopes. It is seen that the power increases with the coefficient of skewness of the distribution. The normal distribution ($C_s = 0$) has the lowest power. The lognormal, Gumbel, generalized extreme value and Pearson type 3 distributions that have a skew coefficient of about $C_s \cong 1.5$ lead to values of the power that are close to each other. The power of the test is highest for the generalized extreme value distribution with k = -0.3 ($C_s = 13.5$).

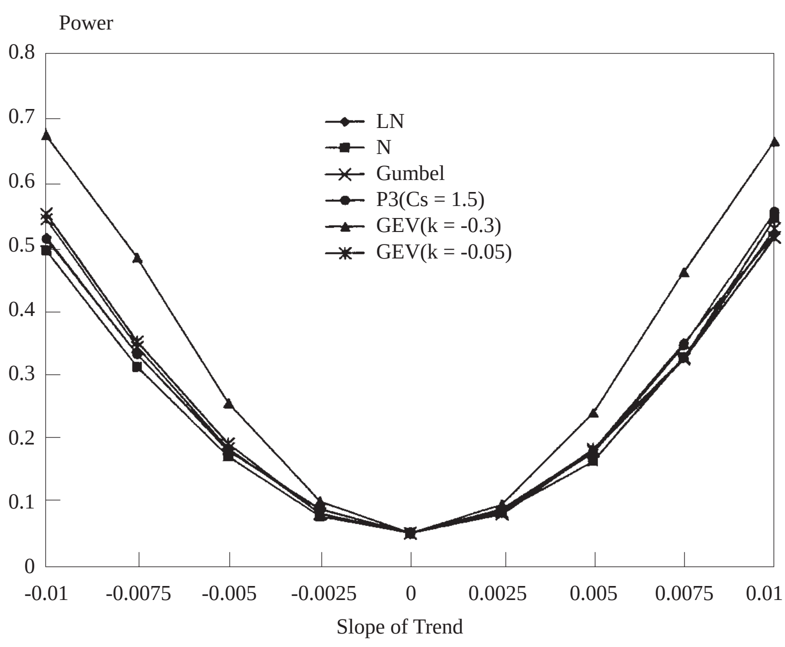


**Figure 1.** Power of the t-test for trend detection as a function of the slope of the trend for various probability distributions.

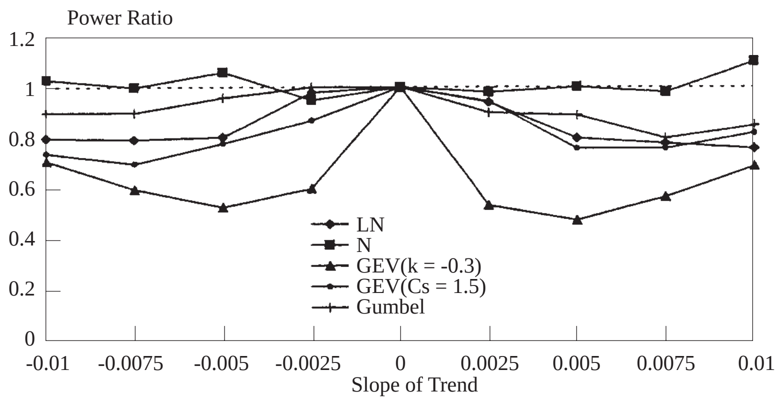


**Figure 2.** Ratio of the power of the t-test to the power of the Mann-Kendall test as a function of the slope of the trend for various probability distributions.

When there is no trend, a type II error is not relevant because the null hypothesis is true. In this case, the power of the test is undefined and the probability of rejection equals $\alpha$. For this reason, the curves in the figures are plotted to pass through the point $\alpha = 0.05$ for slope of trend $= 0$. This is not to be interpreted as the power of the test for no trend being equal to $\alpha$, but as the probability of rejection equaling $\alpha$ in this case.

The ratio of the power of the t-test to that of the Mann-Kendall test is plotted in Figure 2 for various probability distributions as function of the slope of the trend. As expected, this ratio is slightly above one for the normal distribution, implying that the t-test is more powerful than the Mann-Kendall test in this case. For all other (non-normal) distributions, the ratio is significantly less than one. For skewed distributions, the Mann-Kendall test is more powerful, especially when the coefficient of skewness is high. The power ratio is as low as 0.5 when $C_s$ is extremely high (GEV with $C_s = 13.5$).

## Application

Trend analysis of the annual streamflow series at 107 sites in various river basins of Turkey is performed by the t-test and Mann-Kendall test. All the streamflow records in Turkish streams that are free of human influence and are at least 25 years long are included in the study. The records have lengths in the range of 25-65 years. The coefficient of skewness of the data varies between $-2.0$ and 3.3. Tests are applied at the significance level $\alpha = 0.05$.

Both tests detected a trend in 29 series. At two sites the trend was detected only by the t-test, and at two other sites only by the Mann-Kendall test. All the detected trends are of a decreasing kind. Table 1 shows the sites where a trend is detected by the tests.

Four sites where the two tests gave different results are borderline cases. At sites 311 and 2505, the t-test would accept a trend at the levels $\alpha = 0.07$ and $\alpha = 0.15$, respectively. At sites 523 and 1314, a trend would be detected by the Mann-Kendall test at the levels $\alpha = 0.11$ and $\alpha = 0.07$, respectively.

## Conclusions

The power of the t-test for trend detection in stochastic time series is estimated by Monte-Carlo simulation and compared with that of the Mann-Kendall test. The t-test is slightly more powerful when the probability distribution is normal. The power ratio decreases with the increase of the coefficient of skewness. For moderately skewed distributions, the t-test is almost as powerful as the Mann-Kendall test. Therefore, these two tests can be used interchangeably in practical applications, with identical results in most cases, as was shown on an example concerning the trends of an annual streamflow series in Turkey. In borderline cases it is recommended to apply the tests at different levels of significance.

## Acknowledgment

The streamflow data used in this study is provided by the Elektrik İşleri Etüd İdaresi Genel Müdürlüğü under the İ.T.Ü. Araştırma Fonu Research Project “Türkiye Nehirleri Taşkın, Ortalama ve Düşük Akımlardaki Trendler”.

**Table 1.** Sites Tested and Sites with Trend.

| River Basin | Sites Tested | Set with Trend | |
|---|---|---|---|
| | | t-test | MK-test |
| Meriç | 105,106 | 105 | 105 |
| Susurluk | 302,311,314,328 | 302,328 | 302,311,328 |
| Aegean Sea | 407,408 | 407 | 407 |
| Gediz | 509,515,523 | 509,523 | 509 |
| K.Menderes | 601 | 601 | 601 |
| B.Menderes | 701,725 | 701 | 701 |
| W.Mediterranean | 809,812,815 | 809,812,815 | 809,812,815 |
| C.Mediterranean | 902,912,917 | - | - |
| Lake Burdur | 1003 | 1003 | 1003 |
| Afyon | 1102 | 1102 | 1102 |
| Sakarya | 1203,1222,1223,1224, 1233,1237,1239,1244 | 1203,1223,1224,1233, 1239,1244 | 1203,1223,1224,1233, 1239,1244 |
| W.Black Sea | 1314,1319,1327,1331, 1333,1334,1335 | 1314 | - |
| Yeşilırmak | 1401,1412,1418,1422, 1424 | - | - |
| Kızılırmak | 1501,1517,1524,1532, 1535,1536 | - | - |
| C.Anatolia | 1611,1612,1621,1622 | 1611,1622 | 1611,1622 |
| E.Mediterranean | 1714,1717,1719,1720, 1721 | 1714,1717,1719,1720, 1721 | 1714,1717,1719,1720, 1721 |
| Seyhan | 1801,1805,1820,1822 | 1801 | 1801 |
| Ceyhan | 2004,2006,2008,2015 | - | - |
| Fırat | 2102,2122,2123,2124, 2131,2133,2135,2145, 2147,2149,2154,2156, 2157,2158,2164,2166 | 2123,2131,2145 | 2123,2131,2145 |
| E.Black Sea | 2202,2213,2215,2218, 2228,2232,2233,2238, 2239,2245,2247 | - | - |
| Çoruh | 2304,2305,2315,2316, 2323 | - | - |
| Aras | 2402,2409,2415,2418 | - | - |
| Lake Van | 2505 | - | 2505 |
| Dicle | 2603,2610,2612,2618, 2620 | - | - |
| TOTAL | 107 | 31 | 31 |

## References


Haan, C.T., Statistical Methods in Hydrology, The Iowa State Univ. Press, Ames, 1977.

Helsel, D.R., and Hirsch, R.M., Statistical Methods in Water Resources, Elsevier, Amsterdam 1992.

Hirsch, R.M., Alexander, R.B., and Smith, R.A., Selection of methods for the detection and estimation of trends in water quality, Water Resources Res., 27, 803-814, 1991.

Salas, J.D., Analysis and modeling hydrologic time series, In Handbook of Hydrology, Maidment, D.R. (ed.), McGraw-Hill, New York, 19.1-19.72, 1993.

Yue, S., Pilon, P., and Caradias, G., Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, J. Hydrol., 259, 254-271, 2002.