Appendix: Mann-Kendall Trend Tests

Introduction
Hipel & McLeod (1994) and McLeod et al. (1990) have used the Mann-Kendall trend test in the analysis of various types of environmental data.

Kendall Rank Correlation
Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a bivariate random sample of size n. The Pearson correlation coefficient provides an optimal measure of the degree of association between the X's and the Y's when the sample is drawn from a bivariate normal distribution. The Pearson correlation coefficient is reasonably robust for many other distributions as well.
The Kendall correlation coefficient, denoted by τ, provides a more general non-parametric measure of monotonic association. It is said to be monotonic since making a monotonic transformation on either the X's or the Y's does not change the numerical value of τ.
Kendall's rank correlation coefficient (Kendall, 1970, equation 1.5) may be written,

$$\tau = \frac{S}{D}. \tag{2}$$

where S, is the Kendall score given by

$$S = \sum_{i>j} \mathrm{sign}(X_j - X_i)\mathrm{sign}(Y_j - Y_i), \tag{1}$$

where sign($\bullet$) denotes the sign function and D is the maximum possible value of S. In the case where there are no ties among either the X's or the Y's,

$$D = \binom{n}{2}.$$

More generally, if there are n_x distinct ties of extent $t_i, i = 1, ..., n_x$ among the X's and n_y distinct ties of extent $u_i, i = 1, ..., n_y$ among the Y's then

$$D = \sqrt{\left(\binom{n}{2} - T\right)\left(\binom{n}{2} - U\right)},$$

where

$$T = \frac{1}{2}\sum_{i=1}^{n_x} t_i(t_i - 1),$$

and

$$U = \frac{1}{2} \sum_{i=1}^{n_y} u_i(u_i - 1).$$

In the case where there are no ties in either ranking, it is known (Kendall, 1975, p.51) that under the null hypothesis, the distribution of S may be well approximated by a normal distribution with mean zero and variance,

$$\mathrm{Var}\,(S) = \frac{1}{18} n(n-1)(2n+5),$$

provided that $n \geq 10$. Valz and McLeod (1990) have given a simplified derivation of this formula for $\mathrm{Var}\,(S)$.
In the case of ties, the variance of S is more complicated,

$$\begin{aligned}\mathrm{Var}\,(S) = \{\frac{1}{18} n(n-1)(2n+5) - \sum t_i(t_i-1)(2t_i+5) - \sum u_i(u_i-1)(2u_i+5)\} \\ + \frac{1}{9n(n-1)(n-2)} \{\sum t_i(t_i-1)(t_i-2)\}\{\sum u_i(u_i-1)(u_i-2)\} \\ + \frac{1}{2n(n-1)} \{\sum t_i(t_i-1)\}\{\sum u_i(u_i-1)\}.\end{aligned}$$

Valz, McLeod and Thompson (1994) have examined the adequacy of the normal approximation in this general case.
The test of the null hypothesis $\mathcal{H}_0 : \tau = 0$ is equivalent to testing $\mathcal{H}_0 : S = 0$. If there are no ties and if $n \geq 10$ the normal approximation based on $\mathrm{Var}\,(S)$ is adequate. When $n \leq 10$ and there are ties present in only of the variables then the efficient exact algorithm of Panneton & Robillard (1972a, 1972b) may be used. Otherwise if ties are present in both variables then the exact enumeration algorithm given by Valz (1990) may be used or alternatively bootstrapping (Efron and Tibshirani, 1993). Our S-Plus function `Kendall` implements these algorithms for computing τ and its significance level under a two-sided test.

Mann-Kendall Trend Test

Given n consecutive observations of a time series $z_t, t = 1, \cdots, n$, Mann (1945) suggested using the Kendall rank correlation of z_t with $t, t = 1, \cdots, n$ to test for monotonic trend. The null hypothesis of no trend assumes that the $z_t, t = 1, \cdots, n$ are independently distributed. Our S-Plus function, `MannKendall(z)` implements the Mann-Kendall test using `Kendall(x, y)` to compute τ and its signficance level under the null hypothesis.
The Mann-Kendall trend test has some desirable features. In the simple linear trend model with independent Gaussian errors, $z_t = \alpha + \beta t + e_t$, where e_t is Gaussian white noise, it is known that the Mann-Kendall trend test has 98% efficiency relative to the usual least squares method of testing $\beta = 0$. Also, an empirical simulation study of Hipel, McLeod and Fosu (1986) showed that the Mann-Kendall test outperformed the lag one autocorrelation test for detecting a variety of deterministic trends such as a step-intervention or a linear trend.

In the case of no ties in the values of $z_t, t = 1, \cdots, n$ the Mann-Kendall rank correlation coefficient τ has an interesting interpretation. In this case, the Mann-Kendall rank correlation for a trend test can be written

$$\tau = \frac{S}{\binom{n}{2}},$$

where

$$S = 2P - \binom{n}{2},$$

where P is the number of times that $z_{t_2} > z_{t_1}$ for all $t_1, t_2 = 1, \ldots, n$ such that $t_2 > t_1$. Thus $\tau = 2\pi_c - 1$, where π_c is the relative frequency of positive concordance, i.e., the proportion of time for which $z_{t_2} > z_{t_1}$ when $t_2 > t_1$. Equivalently, the relative frequency of positive concordance is given by $\pi_c = 0.5(\tau + 1)$.

The Mann-Kendall test is essentially limited to testing the null hypothesis that the data are independent and identically distributed. Our time series data may diverge from this assumption in two ways. First there may be autocorrelation and second may be a seasonal component. To eliminate these factors we can use annual data but this has the effect of reducing the power. For strong positive autocorrelation in the series, the effect of using annual totals will reduce the effect of this autocorrelation substantially and the loss of power is, perhaps, not expected to be too much — this is something we will investigate further in a methodological study.

The method of Brillinger (1989) deals with both the problems of seasonality and autocorrelation but it also requires an estimate of the spectral density at zero. However the test of Brillinger (1989) is not suitable for testing for long-term trend with monthly data with a strong seasonal component since the running-average smoother used will not be useful in this case. Another model-building approach to trend analysis is intervention analysis (Box & Tiao, 1975; Hipel & McLeod, 1994) which can also handle both seasonality and autocorrelation. This assumes a known intervention time and the development of a suitable time series model.

Seasonal Mann-Kendall Trend Test

The Seasonal-Mann-Kendall trend test is a test for monotonic trend in a time series with seasonal variation. Hirsch et al. (1982) developed such a test by computing the Kendall score separately for each month. The separate monthly scores are then summed to obtain the test statistic. The variance of the test statistic is obtained by summing the variances of the Kendall score statistic for each month. The normal approximation may then be used to evaluate significance level. In this test, the null hypothesis is that the time series is of the form $z_t = \mu_m + e_t$ where e_t is white noise error and μ_m represents the mean for period m. The τ coefficient is defined by

$$\tau = \frac{\sum_{i=1}^{s} S_i}{\sum_{i=1}^{s} D_i},$$

where $S_i, D_i, i = 1, \cdots, s$ denote the Kendall scores and denominators for the i-th season and s is the seasonal period. We implemented this procedure in S-Plus in our function `SeasonalMannKendall(z)`

References

Box, G.E.P. & Tiao, G.C. (1975), Intervention Analysis, *Journal of the American Statistical Association.*

Brillinger, D.R. (1989), "Consistent detection of a monotonic trend superimposed on a stationary time series", *Biometrika*, Vol.76, pp.23–30.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Hipel, K.W. & McLeod, A.I. (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam: Elsevier. ISBN: 0–444–89270–2

Hipel, K.W., McLeod, A.I. & Fosu, P. (1983), Empirical power comparisons of some tests for trend, in *Statistical Aspects of Water Quality Monitoring, Developments in Water Science*, Vol.27, pp.347–362, Edited by A.H. El-Shaarawi and R.E. Kwiatkowski.

Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982), Techniques for trend assessment for monthly water quality data, *Water Resources Research*, 18, 107–121.

Kendall, M.G. (1970), *Rank Correlation Methods*, 2nd Ed., New York: Hafner.

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

McLeod, A.I., Hipel, K.W. & Bodo, B.A. (1990), Trend analysis methodology for water quality time series, *Environmetrics*, 2, 169–200.

Panneton, M. & Robillard, P. (1972a), The Exact Disribution of Kendall's S with Ties in One Ranking, *Applied Statistics*, 21, 321-323.

Panneton, M. & Robillard, P. (1972b), Algorithm AS 54: Kendall's S Frequency Distribution, *Applied Statistics*, 21, 345-348.

Valz, P.D. & McLeod, A.I. (1990), A simplified derivation of the variance of Kendall's tau, *The American Statistician,* 44, 39–40.

Valz, P., McLeod, A. I. and Thompson, M. E., (1995) Cumulant generating function and tail probability approximations for Kendall's score with tied rankings, *Annals of Statistics*, 23, 144–160.

Valz, P. (1990), Developments in Rank Correlation Procedures with Applications to Trend Assessment in Water Resources Research", Ph.D. Thesis, Department of Statistical and Actuarial Sciences, The University of Western Ontario.