U.S. DEPARTMENT OF THE INTERIOR
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

September 2002



---

Publication available at:
http://water.usgs.gov/pubs/twri/twri4a3/

sources of water dominant at different times of the year, such as snow melt versus intense rainfall. Seasonal rise and fall of ground water can also influence water quality. A given discharge in one season may derive mostly from ground water while the same discharge during the another season may result from surface runoff or quick flow through shallow soil horizons. The chemistry and sediment content of these sources may be quite different.

Techniques for dealing with seasonality fall into three major categories (table 12.3). One is fully nonparametric, one is a mixed procedure, and the last is fully parametric. In the first two procedures it is necessary to define a "season". In general, seasons should be just long enough so that there is some data available for most of the seasons in most of the years of record. For example, if the data are primarily collected at a monthly frequency, the seasons should be defined to be the 12 months. If the data are collected quarterly then there should be 4 seasons, etc. Tests for trend listed in table 12.2 have analogs which deal with seasonality. These are presented in table 12.3.

| | Not Adjusted for X | Adjusted for X |
|---|---|---|
| **Nonparametric** | Seasonal Kendall test for trend on Y (Method I) | Seasonal Kendall trend test on residuals from LOWESS of Y on X (Method I) |
| **Mixed** | Regression of deseasonalized Y on T (Method II) | Seasonal Kendall trend test on residuals from regression of Y on X (Method I) |
| **Parametric** | Regression of Y on T and seasonal terms (Method III) | Regression of Y on X, T, and seasonal terms (Method III) |

Table 12.3 Methods for dealing with seasonal patterns in trend testing

12.4.1 The Seasonal Kendall Test

The seasonal Kendall test (Hirsch et al., 1982) accounts for seasonality by computing the Mann-Kendall test on each of m seasons separately, and then combining the results. So for monthly "seasons", January data are compared only with January, February only with February, etc. No comparisons are made across season boundaries. Kendall's S statistic $S_i$ for each season are summed to form the overall statistic $S_k$.

$$S_k = \sum_{i=1}^{m} S_i \qquad [12.1]$$



When the product of number of seasons and number of years is more than about 25, the distribution of $S_k$ can be approximated quite well by a normal distribution with expectation equal to the sum of the expectations (zero) of the individual $S_i$ under the null hypothesis, and variance equal to the sum of their variances. $S_k$ is standardized (eq. 12.2) by subtracting its expectation $\mu_k = 0$ and dividing by its standard deviation $\sigma_{Sk}$. The result is evaluated against a table of the standard normal distribution.

$$Z_{Sk} = \begin{cases} \dfrac{S_k - 1}{\sigma_{Sk}} & \text{if } S_k > 0 \\ 0 & \text{if } S_k = 0 \\ \dfrac{S_k + 1}{\sigma_{Sk}} & \text{if } S_k < 0 \end{cases} \qquad [12.2]$$

where $\mu_{Sk} = 0$,

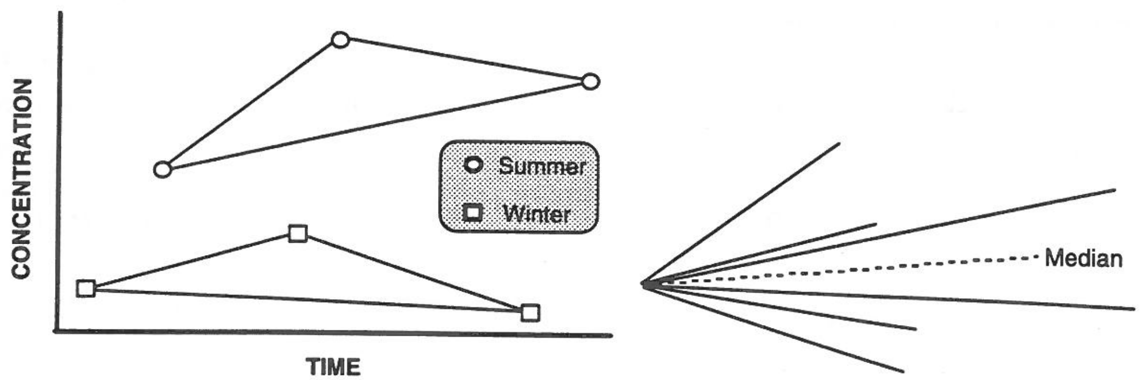
$$\sigma_{Sk} = \sqrt{\sum_{i=1}^{m} (n_i/18)\cdot(n_i-1)\cdot(2n_i+5)}\ , \quad \text{and}$$

$n_i$ = number of data in the ith season.

The null hypothesis is rejected at significance level $\alpha$ if $|Z_{Sk}| > Z_{crit}$ where $Z_{crit}$ is the value of the standard normal distribution with a probability of exceedance of $\alpha/2$. When some of the Y and/or T values are tied the formula for $\sigma_{Sk}$ must be modified, as discussed in Chapter 8. The significance test must also be modified for serial correlation between the seasonal test statistics (see Hirsch and Slack, 1984).

If there is variation in sampling frequency during the years of interest, the data set used in the trend test may need to be modified. If variations in sampling frequency are random (for example if there are a few instances where no value exists for some season of some year, and a few instances when two or three samples are available for some season of some year) then the data can be collapsed to a single value for each season of each year by taking the median of the available data in that season of that year. If, however, there is a systematic trend in sampling frequency (monthly for 7 years followed by quarterly for 5 years) then the following type of approach is necessary. Define the seasons on the basis of the lowest sampling frequency. For that part of the record with a higher frequency define the value for the season as the observation taken closest to the midpoint of the season. The reason for not using the median value in this case is that it will induce a trend in variance, which will invalidate the null distribution of the test statistic.

An estimate of the trend slope for Y over time T can be computed as the median of all slopes between data pairs within the same season (figure 12.11). Therefore no cross-season slopes contribute to the overall estimate of the Seasonal Kendall trend slope.




A. B.

Figure 12.11 A. All pairwise slopes used to estimate the Seasonal Kendall trend slope (two seasons -- compare with figure 10.1).
B. Slopes rearranged to meet at a common origin

To accommodate and model the effects of exogenous variables, directly follow the methods of section 12.3 until the final step. Then apply the Seasonal Kendall rather than Mann-Kendall test on residuals from a LOWESS of Y versus X and T versus X
(R versus T*).

12.4.2 Mixture Methods

The seasonal Kendall test can be applied to residuals from a regression of Y versus X, rather than LOWESS. Keep in mind the discussion in the previous section of using adjusted variables T* rather than T. Regression would be used only when the relationships exhibit adherence to the appropriate assumptions.

A second type of mixed procedure involves deseasonalizing the data by subtracting seasonal medians from all data within the season, and then regressing these deseasonalized data against time. One advantage of this procedure is that it produces a description of the pattern of the seasonality (in the form of the set of seasonal medians). However, this method has generally lower power to detect trend than other methods, and is not prefered over the other alternatives. Subtracting seasonal means would be equivalent to using dummy variables for m−1 seasons in a fully parametric regression. Either use up m−1 degrees of freedom in computing the seasonal statistics, a disadvantage which can be avoided by using the methods of the next section.



12.4.3 Multiple Regression With Periodic Functions

The third option is to use periodic functions to describe seasonal variation. The simplest case, one that is sufficient for most purposes, is:

$$Y = \beta_0 + \beta_1 \cdot \sin(2\pi T) + \beta_2 \cdot \cos(2\pi T) + \beta_3 \cdot T + \text{other terms} + \varepsilon \qquad [12.3]$$

where "other terms" are exogenous explanatory variables such as flow, rainfall, or level of some human activity (e.g. waste discharge, basin population, production). They may be continuous, or binary "dummy" variables as in analysis of covariance. The trend test is conducted by determining if the slope coefficient on T ($\beta_3$) is significantly different from zero. Other terms in the equation should be significant and appropriately modeled. The residuals $\varepsilon$ must be approximately normal.

Time is commonly but not always expressed in units of years. Table 12.4 lists values for $2\pi T$ for three common time units: years, months and day of the year.

| | | | |
|---|---|---|---|
| The expression | $2\pi T$ | $= 6.2832 \cdot t$ | when t is expressed in years. |
| | | $= 0.5236 \cdot m$ | when m is expressed in months. |
| | | $= 0.0172 \cdot d$ | when d is expressed in day of year. |

Table 12.4 Three values for $2\pi T$ useful in regression tests for trend

To more meaningfully interpret the sine and cosine terms, they can be re-expressed as the amplitude A of the cycle (half the distance from peak to trough) and the day of the year Dp at which the peak occurs:

$$\beta_1 \cdot \sin(2\pi t) + \beta_2 \cdot \cos(2\pi t) = A \sin[2\pi(t + t_0)] \qquad [12.4]$$

where $A = \sqrt{\beta_1^2 + \beta_2^2}$ [12.5]

The phase shift $t_0 = \tan^{-1}(\beta_2 / \beta_1)$ ,

$t_0' = t_0 \pm 2\pi$ if necessary to get $t_0$ within the interval $0 < t_0 < 2\pi = 6.2832$

and $Dp = 58.019 \cdot (1.5708 - t_0')$ [12.6]