Anomalies and remedies in non-parametric seasonal trend tests and estimates

Graham McBride, NIWA (Hamilton, NZ), March 2000

The anomalies discussed here can arise when using non-parametric trend tests, as implemented for example in WQStat Plus (IDT Ltd. 1998) and in its predecessors. These tests are now being used on New Zealand environmental data at national scale (Smith *et al*. 1996) and at regional scale (e.g., Stansfield 1997; Vant & Wilson 1998), so an awareness of these issues and the appropriate remedies should prove useful.

The following is adapted from material presented in McBride & Loftis (1994). The main points are given in the following two sections (TWO ISSUES and SUMMARY). Technical details then follow.

TWO ISSUES

Results of the two-sided seasonal Kendall trend test and its associated seasonal Kendall slope estimator (as described by Gilbert 1987:225–228 and implemented in the WQStat II and WQStat Plus software, IDT 1998) can occasionally give one of two unexpected results:

(a) a non-zero slope estimate that *cannot* be significant (its apparent p-value is 100%, indicating certainty that there is no trend);

(b) a zero slope estimate that *is* significant (i.e., its apparent p-value is very low).

SUMMARY

Case (a) arises whenever the test statistic is computed to be zero and the large sample approximation is used to obtain its critical value (i.e., it is assumed that the test statistic follows the unit normal distribution). This approximation is inappropriate when that statistic is zero. A simple way to resolve the problem is to report the test's p-value as CEILING(p^*), where p^* is the largest p-value that can be attained when the test statistic is non-zero and CEILING raises p^* to the next (higher) integer. Because p^* is always less than 100% this procedure works well. For example, for trend analysis of monthly sampling for 5 years we have $p^* = 94.4\%$ and so the result should be reported as $p < 95\%$. For ten years of monthly sampling we have $p^* = 97.94\%$ and so the result should be reported as $p < 98\%$.

Case (b) arises when there is disproportionately large number of ties (i.e., equal values). In such a case the slope estimator may not be appropriate; it may be better to take the median of the remaining non-zero slopes as the estimate (the estimate does not form part of the test; it is an adjacent calculation).

TECHNICAL DETAILS

Three questions arise when seeking reasons for results (a) and (b), and when proposing remedies. Without loss of generality, their answers are well illustrated by taking the example of trend analysis for $n = 5$ years of monthly data.

<u>How can p = 100% for a two-sided seasonal Kendall test?</u>

To test the null hypothesis that there was no trend in concentration of some variable (denoted by c) over the five years the test procedure is:

1. Select a particular month (call it month i).
2. Compute the signs of all the possible differences in concentrations between years for that month (cross-month differences are not considered). The resulting 10 differences are: $\text{sgn}(c_{i2}–c_{i1})$, $\text{sgn}(c_{i3}–c_{i1})$, $\text{sgn}(c_{i4}–c_{i1})$, $\text{sgn}(c_{i5}–c_{i1})$, $\text{sgn}(c_{i3}–c_{i2})$, $\text{sgn}(c_{i4}–c_{i2})$, $\text{sgn}(c_{i5}–c_{i2})$, $\text{sgn}(c_{i4}–c_{i3})$, $\text{sgn}(c_{i5}–c_{i3})$, $\text{sgn}(c_{i5}–c_{i4})$.
3. Replace sgn(x) by +l if the sign is positive or by –1 if it is negative, and by 0 if there is no difference (i.e., there is a tie, so that $x = 0$). Then sum the 10 values and call that sum S_i.

4. Compute the variance of S_i from $[n(n-1)(2n+5) - \sum t_{ip}(t_{ip}-1)(2t_{ip}+5)]/18$, where t_{ip} is the number of tied data in the pth tied group for the ith month, and the summation is over the number of tied groups for that month.[1]
5. Sum the S_i series over all months; i.e., $S = \Sigma S_i$, for $i = 1, 2,\ldots, 12$. (So S is the sum of 120 terms.)
6. Compute the variances of S from $\mathrm{Var}(S) = \Sigma[\mathrm{Var}(S_i)]$, for $i = 1, 2,\ldots, 12$.
7. Compute the test static (Z_{sK}) from the large sample normal approximation (with continuity correction of one unit), i.e.,
   $Z_{sK} = (S-1)/\sqrt{[\mathrm{Var}(S)]}$ if $S > 0$ (1)
   $Z_{sK} = 0$ if $S = 0$ (2)
   $Z_{sK} = (S+1)/\sqrt{[\mathrm{Var}(S)]}$ if $S < 0$ (3)
8. Reject the null hypotheses if $|Z_{sK}| > Z_{\alpha/2}$, where α is the chosen significance level and $Z_{\alpha/2}$ is the value of the abscissa that cuts off an area = $\alpha/2$ in the right tail of the unit normal distribution.

In our example, the lowest non-zero value of $|Z_{sK}|$ is 0.0707. This value occurs when there are no ties and the number of terms of opposite sign differs by 2, so that $S = \pm 2$ and Var $(S) = \Sigma(5\text{x}4\text{x}15/18) = 12\text{x}5\text{x}4\text{x}15/18 = 200$. Equations 1 or 3 (whichever is appropriate) then give $|Z_{sK}| = 1/\sqrt{200} = 0.0707$. The associated p-value is 94.36% (an algorithm for computing normal p-values is given in the last section). Slightly larger values of the test statistic, and hence smaller p-values, are obtained if $S = \pm 2$ in the presence of ties [because then $\mathrm{Var}(S) < 200$].

However there are a number of situations for which it is possible to obtain $Z_{sK} = 0$ and hence, for a two-sided test, $p = 100\%$:

i. there are no ties and an equal number (60) of terms in the S series have opposite signs (so that $S = 0$ and equation 2 applies);
ii. there is an even number of ties and an equal number (<60) of terms in the S series have opposite signs (so that S = 0 and equation 2 applies);
iii. there is an odd number of ties and the remaining terms in the S series differ by 1 in their numbers of opposite signs (so that $S = 1$ in equation 1 or $S = -1$ in equation 3, and the numerator of the right-hand of the equations is zero).

Reporting $p = 100\%$ is untenable—in statistics nothing is certain. This wrong result arises because the large sample approximation is not appropriate when $Z_{sK} = 0$; there is a non-zero probability of getting $S = 0$ or $+1$ from a finite sample. When the null is true, that probability can be computed, albeit laboriously (see Hollander & Wolfe 1973, p. 190). A simple remedy in our example is to report $p >$ 95%, because 95% is just greater than the highest p-value attainable for a non-zero test statistic (94.36%), but is less than 100%. Note however that if we were analysing 10 years of monthly data then the lowest non-zero value of the test statistic would be $|Z_{sK}| = 1/\sqrt{1500} = 0.0258$, with associated p-value = 97.94%. The p-value corresponding to $Z_{sK} = 0$ could then be reported as $p > 98\%$.

<u>How can the estimated trend slope be non-zero when the test statistic is zero?</u>

Consider the seasonal Kendall slope estimate. It is the median of all possible slopes within each month (i.e., no cross-month slopes contribute to the estimate). The procedure is:

1. Select a particular month (call it month i).
2. Compute all possible annual slopes between years for that month. There are 10 of these, i.e.,
   $$\frac{c_{i2}-c_{i1}}{1}, \frac{c_{i3}-c_{i1}}{2}, \frac{c_{i4}-c_{i1}}{3}, \frac{c_{i5}-c_{i1}}{4}, \frac{c_{i3}-c_{i2}}{1}, \frac{c_{i4}-c_{i2}}{2}, \frac{c_{i5}-c_{i2}}{3}, \frac{c_{i4}-c_{i3}}{1}, \frac{c_{i5}-c_{i3}}{2}, \frac{c_{i5}-c_{i4}}{1}$$
3. Repeat this procedure for all months (i.e., for $i = 1, 2,\ldots, 12$). This gives 120 slopes.
4. Rank these slopes and take the median as the estimate. This is the average value of the 60th and 61st slopes.

[1] The formula is more complicated if there are multiple data per season (Gilbert 1987:226; IDT Ltd. 1998:89), but we need not consider that here (the same conclusions emerge if one uses that formula).

Consider the situations (i)-(iii) above (covering all possibilities of obtaining $Z_{sK} = 0$). If (i) holds, the estimate is the average of the least negative and the least positive slopes. If these two slopes are equal in absolute value they will cancel and the estimate will be zero; otherwise it will be non-zero. If (ii) holds, the estimate must be zero (because the 60th and 61st medians must be zero). If (iii) holds and there are 3 or more ties, the estimate will be zero: if there is only one tie it will be non-zero.

So for $Z_{sK} = 0$, a non-zero slope estimate can occur in two cases. These occur when only one of the two values straddling the median is zero, or when the two values straddling the median are non-zero and their signs and absolute values are different. Consequently, any such non-zero slope would be the smallest possible non-zero value, as expected. This result is consistent with the test attaining the highest possible *p*-value, but less than 100% (as above).

3. How can a zero estimated trend slope be statistically significant?

Consider a case where there are many ties. For example, analysis of 5 years monthly dissolved reactive phosphorus data for a site in the New Zealand River Water Quality network (site CH1 on the Hurunui River, Smith *et al.* 1996) gave 53 positive slopes, 27 negative slopes, and 40 ties (10 groups of 2, 8 groups of 3, and 1 group of 4). These ties cause the median slope to be zero (it is straddled by ties) and Var(S) to be reduced from 200 to 152. The software computed $S = 25$ and so the test statistic (equation 1) was $|Z_{sK}| = 25/\sqrt{152} = 2.028$. This is statistically significant at the 5% level (being greater than $Z_{\alpha=0.05/2} = 1.960$). Note that if the 40 ties were replaced by an equal number (20) of positive and negative differences, the test statistic would have been $25/\sqrt{200} = 1.768$ which would not be significant at the 5 % level.

The problem here is that the slope estimator is not very appropriate in the presence of many ties. It may be better to take the median of all non-tied values, or perform some other slope estimation procedure. If there are a large number of ties resulting from replacement of "less-than" data, it may not be meaningful to seek for trends at all, especially if the detection limit has changed with time.

CALCULATING NORMAL *P*-VALUES

The *p*-value for the seasonal Kendall trend test (and many other tests) is obtained from the unit normal distribution. It is obtained as $p = \text{Prob}[Z > |Z_{sK}|]$, where Z is the unit normal deviate with critical values $\pm|Z_{sK}|$. An approximate *p*-value can be found from standard tables. The accurate value (to at least four digits) can be obtained in a direct calculation by evaluating the complementary error function—called erfcc in Press *et al*. (1992). This function is related to the unit normal distribution function (using equations 7.1.2 & 26.2.29 of Abramowitz & Stegun (1965). The resulting equation is $P(x) = 1 - [\text{erfc}(x/\sqrt{2})]/2$, where $P(x)$ is the area under the unit normal distribution to the left of some value x. Two computer functions are required:

```
Function ProbN(x)
!Returns the area under the unit normal distribution to the left of any positive
!or negative value x.
y     = ABS (x)/SQRT (2.)
t     = 1./(1. + 0.5*y)
erfcc = t*exp(-y*y    -     1.26551223 + t*(1.00002368 + t*( 0.37409196
      + t*(0.09678418 + t*(-0.18628806 + t*(0.27886807 + t*(-1.13520398
      + t*(1.48851587 + t*(-0.82215223 + t* 0.17087277        )))))))))
If x < 0.0 erfcc = 2. - erfcc
ProbN = 1. – erfcc/2.
End
```

and

```
Function p(u)
!Returns p = Prob [Z>|ZsK|] where u = |ZsK| (u must be non-negative).
p     = 1. - (ProbN(u) – ProbN(-u))
end
```

References

Abramowitz, M.; Stegun, I.A. (1972). Handbook of Mathematical Functions. Dover, New York.

Gilbert, R.O. (1987). Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Hollander, M.; Wolfe, D.A. (1973). Nonparametric Statistical Methods. Wiley, New York.

IDT Ltd. (1998). WQStat Plus User's Guide. Intelligent Design Technologies Ltd., Longmont, Colorado (available from l.miller@niwa.cri.nz).

McBride, G.B.; Loftis, J.C. (1994). The most important statistical aspects. Proceedings of the international workshop *Monitoring Tailor-made*, RIZA (Netherlands Government Institute for Inland Water Management and Waste Water Treatment), M. Adriaanse *et al*. (eds), Beekbergen, The Netherlands: 153–161.

Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. (1992). Numerical Recipes in FORTRAN. 2nd ed. Cambridge University Press, Cambridge.

Smith, D.G.; McBride, G.B.; Bryers, G.G.; Wisse, J.; Mink, D.F. J. (1996). Trends in New Zealand's River Water Quality Network. *New Zealand Journal of Marine and Freshwater Research 30*: 485–500.

Stansfield, B. (1997). Wairarapa Rivers Water Quality Monitoring Programme 4 Year Summary November 1991–November 1995. Technical Report No. 97/1, Wellington Regional Council (Wairarapa Division), Masterton, New Zealand.

Vant, B.; Wilson, B. (1998). Trends in River Water Quality in the Waikato Region, 1980-97. Environment Waikato Technical Report 1998/13. Waikato Regional Council, Hamilton, New Zealand.