Sensitivity of Non-Parametric Trend Analyses to Multi-Year Extremes

Kenneth L. Wahl[1]

ABSTRACT

Non-parametric tests for trend are widely used in analyzing hydrologic data because they presume nothing about the underlying distribution and are relatively insensitive to the presence of individual outliers. However, hydrologic data often include multi-year sequences of wetter or drier than normal periods. Such sequences may be episodic in nature and may persist for several years, but may imply no systematic change. While trend results are relatively insensitive to individual outliers, multi-year sequences of extremes can have a significant effect on the test results if those sequences occur near the beginning or the end of the period of record used for the test. The sensitivity of trend results to a sequence of extremes depends on the placement of that sequence in the period of record being tested and on the length of the sequence of extremes relative to the length of the record.

INTRODUCTION

Non-parametric tests for trend are widely used in analyzing hydrologic data. These tests presume nothing about the underlying distribution and are relatively insensitive to the presence of individual outliers. Hydrologic data, however, often include sequences of wetter than normal or dryer than normal periods that may persist for several years. Such sequences are perturbations and are often viewed as representing outliers because they are episodic in nature and generally imply no systematic change.

The purpose of this paper is to examine the sensitivity of a commonly used non-parametric test in hydrology, Kendall's tau (Kendall, 1938, 1975; Hirsch and others, 1982), to an actual sequence of the extremes.

KENDALL'S TAU

Kendall's tau is designed to identify whether monotonic changes are occurring with time. The test is non-parametric; thus, the test variables need not have a known or assumed probability distribution. The test is insensitive to the presence of individual outliers and is applicable even when the data under test have values missing. Kendall's tau is determined in the following manner: Given a time series $x_1, x_2, \text{-------}, x_n$, of length n, the differences $d_{ij} = x_i - x_j$ are determined for $1 \le i < j \le n$. There are n(n-1)/2 differences. If P is the number of positive differences and N is the number of negative differences, then:

$$\text{tau} = (P - N) / [n(n-1)/2] \qquad (1)$$

If all differences are positive, tau = +1.00; if all the differences are negative, tau = -1.00. However, if the number of positive differences is equal to the number of negative differences (P = N), tau = 0. Tau is, therefore, a measure of the correlation between the series of x_i and time, and the sign of tau indicates whether x is increasing (+) or decreasing (-) with time. Probabilities of the tau distribution are given in statistics texts, including Gibbons (1976) and Helsel and Hirsch (1992).

The Kendall tau test works only on the signs of differences between pairs of data. Because Kendall's tau works only with relative rankings of the data, the test is non-parametric, and non-parametric methods are known to be relatively insensitive to the presence of individual outliers. However, differences are calculated for each value with n-1 other values. If the initial value in the series is the smallest (largest) data value in the set, it will generate n-1 number of positive (negative) signs when compared to other values. The inverse is true if the extreme is the final value in the series. The same data value in the middle of the data set would result in an equal number of negative and positive differences and tau would be zero. Thus, the leverage of extremes near either end of the period of record must be considered in evaluating results. The amount of leverage is, of course, a function of the sample size.

---

[1] Hydrologist, U.S. Geological Survey, Lakewood, CO

Presented at the Western Snow Conference, Snowbird, UT, 1998

The leverage effect of drought data at the end of the period of analysis was noted by Wahl (1991, 1992). He examined trends for California streamflow data through 1991 while the 1987-93 drought was in progress. Wahl (1992) determined Kendall's tau for both the entire period of record through 1991 and by limiting the data to pre-1987. Although record lengths were generally greater than 50 years, the trend results were strongly influenced by the data for the drought years of 1987-91.

APPROACH

Annual streamflow data for the Yellowstone River at Billings, Montana (USGS station 06214500), demonstrate the leverage effect on actual data. Streamflow data have been collected at this site since 1929. During the first 12 years data were collected (1929-41), the area was affected by a drought that extended until about 1941 (Merritt, 1991). Also, during the period 1967-76 annual average discharges were generally greater than the long-term average. The histogram of annual average discharge is shown in Figure 1a.

Tests were done in three different ways. First, a step-forward method of selecting data was used in which all records tested began in 1929. The minimum length of record tested was 10 years, and data were added in 5-year increments. Next, a fixed length of record of 25 years was tested in which the period was advanced incrementally, 5 years at a time, from the beginning of the data to the end. Finally, a step-backward procedure was used. Again, the minimum length of record was 10 years, and data were added in 5-year increments. However, in the step-backward approach all data included the most recent period ending in 1997. If monotonic trends are present in a given dataset, those trends should be in evidence regardless of which data selection process is used.

RESULTS

The step-forward analysis yields a statistically significant trend of increasing flows for all record lengths of 20 or more years (fig. 1b). However, when tau is computed using 25 years of record at a time, only the periods ending before 1960 and that ending in 1975 show significant trends of increasing flows. The 25-year periods ending in 1990 and 1995 actually show significant trends of decreasing flows (fig. 1c). The cyclical pattern of tau in figure 1c is in contrast to the relatively stable 25-year moving average (fig. 1a). The cyclical pattern (fig. 1c) results in part from the relative positioning of the 1929-41 period of less than normal flows and the 1967-76 period when flows were greater than normal. The large positive tau begins to decline as the period of analysis advances and some of the lower flows of the 1929-41 period are dropped. By 1965, most of the early drought years have been dropped, and tau shows there to be no trend. The 25-year analyses ending in 1970 and 1975 again show increasing tau, reflecting the effects of the greater-than-normal flows of 1967-76 that are now on the forward end of the period being analyzed. As the computation progresses forward 5 years at a time, these same flows (1967-76) gravitate toward the back end of the period being analyzed, and tau declines to show a significant trend of decreasing flows in 1990 and 1995.

The comparison of the step-forward computations and the step-backward computations (fig. 2) show the influence of the 1929-41 period on trend analyses that use the entire period of record. In the step-forward procedure, the initial years of record are present in all analyses, and a statistically significant trend of increasing flows is present for all analyses of 20 or more years. By contrast, the step-backward analyses include the bulk of that initial period only when the record length is greater than about 60 years and shows no significant trends for record lengths of 15 to 60 years.

These tests show that trend results are highly dependent on both the length and relative positioning of multi-year sequences of extremes. The Kendall tau test is relatively insensitive to individual outliers. However, multi-year sequences that are wetter or dryer than normal have a significant effect on the results of trend tests if those sequences occur near the beginning or the end of the period of record used for the test. The effects of multi-year sequences of extremes need to be recognized when applying trend tests. One mechanism for recognizing these effects is to compare the results of step-forward and step-backward analyses.

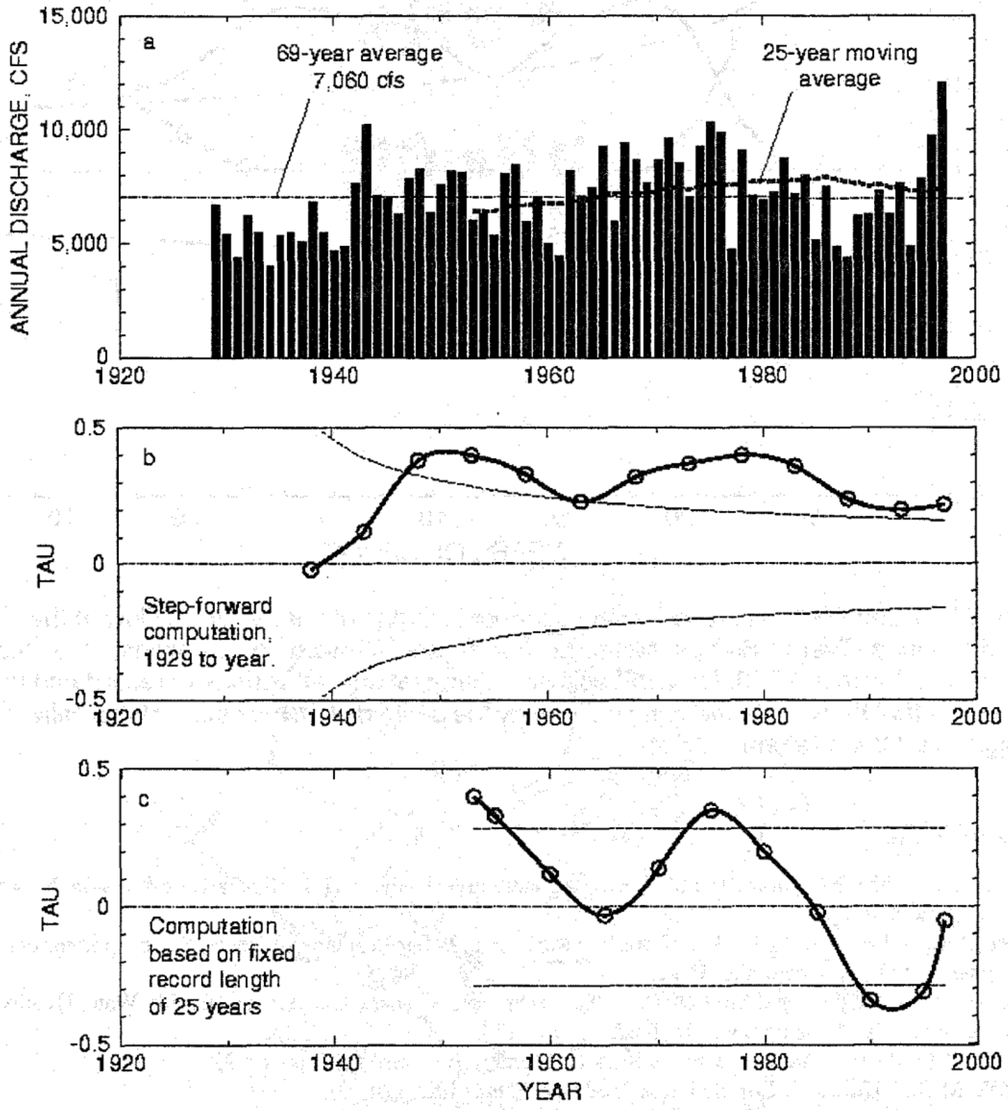

Figure 1. a) Annual average discharge, cubic feet per second, for the Yellowstone River at Billings, Montana; b) Kendall's tau results in a step-forward fashion, using all the data available as of the year shown; c) Kendall's tau results using a fixed record length of 25 years, ending at the year shown. [Values that lie outside the dashed lines on figure 2b and 2c indicate a statistically significant trend (alpha = 0.05)].

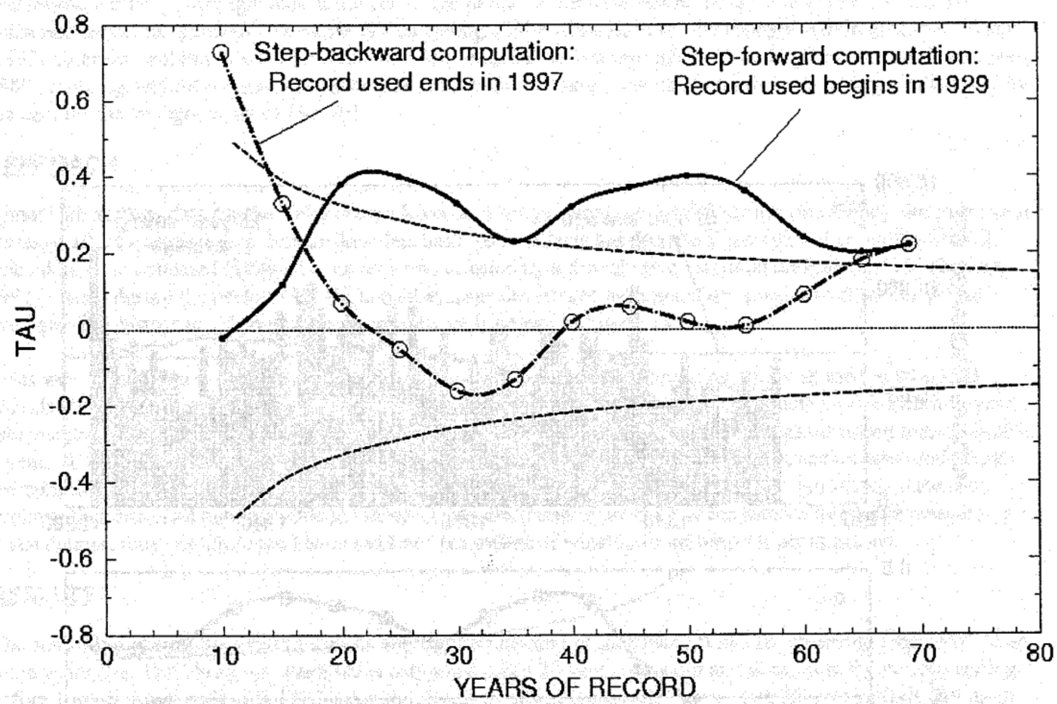

Figure 2. Kendall's tau results for various record lengths of annual streamflow at the Yellowstone River at Billings, Montana. For the step-forward computations, all periods of record begin in 1929; for step-backward computations, all periods of record end in 1997. [Values that lie outside the band defined by the dashed lines indicate a statistically significant trend (alpha = 0.05)].

REFERENCES

Gibbons, J.D., 1976, Nonparametric Methods for Quantitative Analysis: Holt, Rinehart and Winston, New York, New York, 463 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical Methods in Water Resources: Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 522 p.

Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of Trend Analysis for Monthly Water Quality Data: Water Resources Research, vol. 18, no. 1, pp. 107-121.

Kendall, M.G., 1938, A New Measure of Rank Correlation: Biometrika, 30, pp. 81-93.

Kendall, M.G., 1975, Rank Correlation Methods: Charles Griffin, London.

Wahl, K.L., 1991, Is April to July Runoff Really Decreasing in the Western United States?: Proceedings of the Western Snow Conference, April, 1991, Juneau, Alaska, pp. 67-78.

Wahl, K.L., 1992, Evaluation of Trends in Runoff in the Western United States: American Water Resources Association Proceedings of Symposium on Managing Water Resources during Global Change, Reno, Nevada, November 1992, pp. 701-710.

Merritt, L.A., 1991, Montana Floods and Droughts in National Water Summary 1988-89 – Hydrologic Events and Floods and Droughts: U.S. Geological Survey Water-Supply Paper 2375, pp. 369-376.