Dave

Admin #243

Studies in Environmental Science 49

# STATISTICAL METHODS IN WATER RESOURCES

**D.R. Helsel and R.M. Hirsch**

*U.S. Geological Survey, Water Resources Division,*
*Reston, Virginia 22092, USA*

ELSEVIER
Amsterdam • Lausanne • New York • Oxford • Shannon • Tokyo

ELSEVIER SCIENCE B.V.
Sarah Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

First edition 1992
Second printing 1993
Third printing 1995

ISBN 0-444-81463-9



Printed in The Netherlands on acid-free paper.

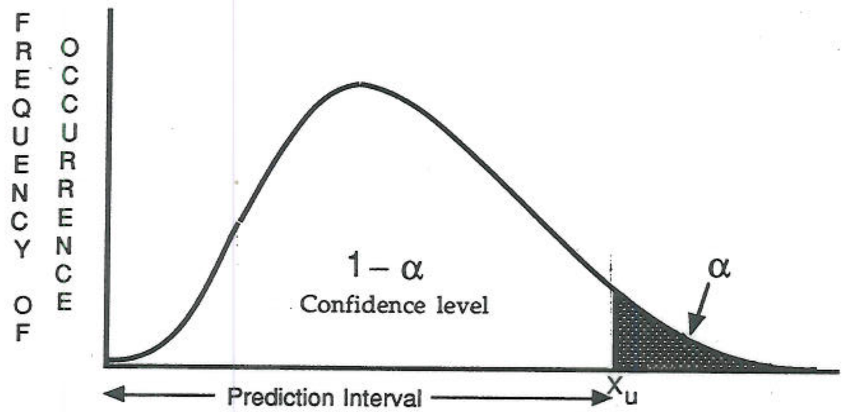



Figure 3.8 Confidence level and alpha level for a 1-sided prediction interval Probability of obtaining a new observation greater than $X_u$ when the distribution is unchanged is $\alpha$.

<u>Example 2, cont.</u>
An arsenic concentration of 350 ppb is found in a New Hampshire well. Does this indicate a change to larger values as compared to the distribution of concentrations for the example 2 data? Use $\alpha = 0.10$.

As only large concentrations are of interest, the new data point will be considered larger if it exceeds the $\alpha = 0.10$ one-sided prediction interval, or upper 90th percentile of the existing data. $X_{0.90 \bullet 26} = X_{23.4}$. By linear interpolation this corresponds to a concentration of

$$X_{23} + 0.4 \bullet (X_{24} - X_{23}) \;=\; 300 + 0.4 \bullet (40) \;=\; 316.$$

In other words, a concentration of 316 or greater will occur approximately 10 percent of the time if the distribution of data has <u>not</u> increased. Therefore a concentration of 350 ppb is considered larger than the existing data at an $\alpha$ level of 0.10.

### 3.6 Parametric Prediction Intervals

Parametric prediction intervals are also used to determine whether a new observation is likely to come from a different distribution than previously-collected data. However, an assumption is now made about the shape of that distribution. This assumption provides more information with which to construct the interval, as long as the assumption is valid. If the data do not approximately follow the assumed distribution, the prediction interval may be quite inaccurate.

$$\text{Prob}(W \le 3) + \text{Prob}(W \ge 13) = 0.048 + 0.048 = 0.095 \le 0.1.$$

Note that for a two-sided test, the critical values are farther from the expected value than in a one-sided test at the same $\alpha$ level.

It should be recognized that p-values are also influenced by sample size. For a given magnitude of difference between the x and y data, and a given amount of variability in the data, p values will tend to be smaller when the sample size is large. In the extreme case where vast amounts of data are available, it is a virtual certainty that p values will be small even if the differences between x and y are what might be called "of no practical significance."

Most statistical tables are set up for one-sided tests. That is, the rejection region $\alpha$ or the p-value is given in only one direction. **When a two-sided test at significance level $\alpha$ is performed, the tables must be entered using $\alpha/2$.** In this way rejection can occur with a probability of $\alpha/2$ on either side, and an overall probability of $\alpha$. **Similarly, tabled p-values must be doubled to get p-values for a two-sided test.** Modern statistical software often reports p-values with its output, eliminating the need for tables. Be sure to know whether it is one-sided or two-sided p-values being reported.

### 4.4 Tests for Normality

The primary reason to test whether data follow a normal distribution is to determine if parametric test procedures may be employed. The null hypothesis for all tests of normality is that the data are normally distributed. Rejection of $H_0$ says that this is doubtful. Failure to reject $H_0$, however, does not prove that the data do follow a normal distribution, especially for small sample sizes. It simply says normality cannot be rejected with the evidence at hand. Use of a larger $\alpha$-level (say 0.1) will increase the power to detect non-normality, especially for small sample sizes, and is recommended when testing for normality.

The test for normality used in this book is the probability plot correlation coefficient (PPCC) test discussed by Looney and Gulledge (1985a). Remember from Chapter 2 that the more normal a data set is, the closer it plots to a straight line on a normal probability plot. To test for normality, this linearity is tested by computing the linear correlation coefficient between data and their normal quantiles (or "normal scores", the linear scale on a probability plot). Samples from a normal distribution will have a correlation coefficient very close to 1.0. As data depart from normality, their correlation coefficient will decrease below 1. To perform a test of $H_0$: the data are normal versus $H_1$: they are not, the correlation coefficient (r) between the data and their normal quantiles is tested to see if it is significantly less than 1. For a sample size of n, if r is smaller than the critical value r* of table B3 for the desired $\alpha$-level, reject $H_0$.

The test statistic S measures the monotonic dependence of y on x. Kendall's S is calculated by subtracting the number of "discordant pairs" M, the number of (x,y) pairs where y decreases as x increases, from the number of "concordant pairs" P, the number of (x,y) pairs where y increases with increasing x:

$$\boxed{S = P - M} \qquad [8.1]$$

where P = "number of pluses", the number of times the y's increase as the x's increase, or the number of $y_i < y_j$ for all $i < j$,

M = "number of minuses," the number of times the y's decrease as the x's increase, or the number of $y_i > y_j$ for $i < j$.

for all $i = 1,....(n - 1)$ and $j = (i+1),.....n$.

Note that there are $n(n-1)/2$ possible comparisons to be made among the n data pairs. If all y values increased along with the x values, $S = n(n-1)/2$. In this situation, the correlation coefficient $\tau$ should equal +1. When all y values decrease with increasing x, $S = -n(n-1)/2$ and $\tau$ should equal $-1$. Therefore dividing S by $n(n-1)/2$ will give a value always falling between $-1$ and +1. This then is the definition of $\tau$, measuring the strength of the monotonic association between two variables:

Kendall's tau correlation coefficient

$$\tau = \frac{S}{n(n-1)/2} \qquad [8.2]$$

To test for significance of $\tau$, S is compared to what would be expected when the null hypothesis is true. If it is further from 0 than expected, $H_0$ is rejected. For $n \leq 10$ an exact test should be computed. The table of exact critical values is found in table B8 of the Appendix.

### 8.2.2 Large Sample Approximation

For $n > 10$ the test statistic can be modified to be closely approximated by a normal distribution. This large sample approximation $Z_S$ is the same form of approximation as used in Chapter 5 for the rank-sum test, where now

$d = 2$ (S can vary only in jumps of 2),

$\mu_S = 0$, and

$\sigma_S = \sqrt{(n/18) \bullet (n-1) \bullet (2n+5)}$.

# TABLE B8 -- Quantiles (p-values) for Kendall's tau correlation coefficient

(from Kendall, 1975).



p = Prob [S ≥ x] = Prob [S ≤ −x]

| | Number of data pairs = n | | | |
|---|---|---|---|---|
| x | 4 | 5 | 8 | 9 |
| 0 | 0.625 | 0.592 | 0.548 | 0.540 |
| 2 | 0.375 | 0.408 | 0.452 | 0.460 |
| 4 | 0.167 | 0.242 | 0.360 | 0.381 |
| 6 | 0.042 | 0.117 | 0.274 | 0.306 |
| 8 | | 0.042 | 0.199 | 0.238 |
| 10 | | 0.0083 | 0.138 | 0.179 |
| 12 | | | 0.089 | 0.130 |
| 14 | | | 0.054 | 0.090 |
| 16 | | | 0.031 | 0.060 |
| 18 | | | 0.016 | 0.038 |
| 20 | | | 0.0071 | 0.022 |
| 22 | | | 0.0028 | 0.012 |
| 24 | | | 0.0009 | 0.0063 |
| 26 | | | 0.0002 | 0.0029 |
| 28 | | | <0.0001 | 0.0012 |
| 30 | | | | 0.0004 |
| 32 | | | | 0.0001 |

| | Number of data pairs = n | | |
|---|---|---|---|
| x | 6 | 7 | 10 |
| 1 | 0.500 | 0.500 | 0.500 |
| 3 | 0.360 | 0.386 | 0.431 |
| 5 | 0.235 | 0.281 | 0.364 |
| 7 | 0.136 | 0.191 | 0.300 |
| 9 | 0.068 | 0.119 | 0.242 |
| 11 | 0.028 | 0.068 | 0.190 |
| 13 | 0.0083 | 0.035 | 0.146 |
| 15 | 0.0014 | 0.015 | 0.108 |
| 17 | | 0.0054 | 0.078 |
| 19 | | 0.0014 | 0.054 |
| 21 | | 0.0002 | 0.036 |
| 23 | | | 0.023 |
| 25 | | | 0.014 |
| 27 | | | 0.0083 |
| 29 | | | 0.0046 |
| 31 | | | 0.0023 |
| 33 | | | 0.0011 |
| 35 | | | 0.0005 |
| 37 | | | 0.0002 |

# 12

# Trend Analysis

Concentrations and loads of phosphorus have been observed at numerous tributaries to an important estuary over a 20-year period. Have concentrations and/or loads changed over time? Have concentrations changed when changing flow conditions are taken into account (the early years were during a very dry period), or are all changes simply due to more precipitation in the latter years? Is there an observable effect associated with a ban on phosphorus compounds in detergents which was implemented in the middle of the period of record?

Groundwater levels were recorded for many wells in a study area over 14 years. During the ninth year development of the area increased withdrawals dramatically. Is there evidence of decreasing water levels in the region's wells after versus before the increased pumpage?

Benthic invertebrate and fish population data were collected at twenty stations along one hundred miles of a major river. Do these data change in a consistent manner going downstream? What is the overall rate of change in population numbers over the one hundred miles?

Procedures for trend analysis build on those in previous chapters on regression and hypothesis testing. The explanatory variable of interest is usually time, though spatial or directional trends (such as downstream order or distance downdip) may also be investigated. Tests for trend have been of keen interest in environmental sciences over the last 10-15 years. Detection of both sudden and gradual trends over time with and without adjustment for the effects of confounding variables have been employed. In this chapter the various tests are classified, and their strengths and weaknesses compared.

## 12.1 General Structure of Trend Tests

### 12.1.1 Purpose of Trend Testing

A series of observations of a random variable (concentration, unit well yield, biologic diversity, etc.) have been collected over some period of time. We would like to determine if their values generally increase or decrease (getting "better" or "worse"). In statistical terms this is a determination of whether the probability distribution from which they arise has changed over time. We would also like to describe the amount or rate of that change, in terms of changes in some central value of the distribution such as a mean or median. Interest may be in data at one location, or all across the country. Figure 12.1 presents an example of the results of trend tests for bacteria at sites throughout the United States.

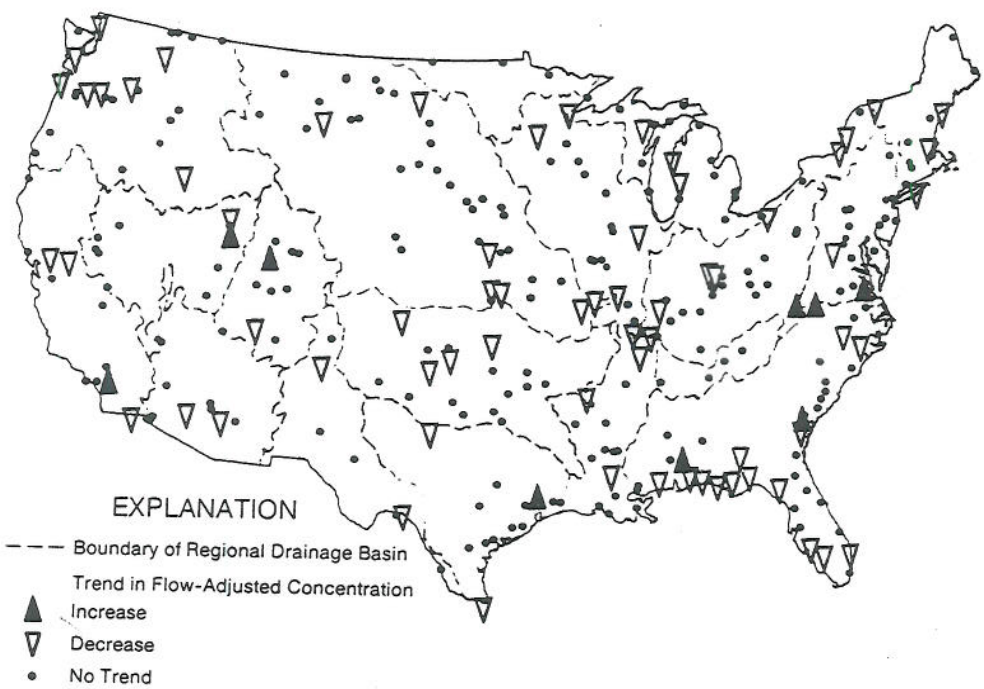



Figure 12.1 Trends in flow-adjusted concentrations of fecal streptococcus bacteria, 1974-1981 (from Smith et al., 1987).

The null hypothesis: $H_0$ is that there is no trend. However, any given test brings with it a precise mathematical definition of what is meant by "no trend", including a set of background assumptions usually related to type of distribution and serial correlation. The outcome of the test is a "decision" – either $H_0$ is rejected or not rejected. Failing



to reject $H_0$ does no
statement that the e
Table 12.1 summariz
analysis.

| Decision |
|---|
| Fail to reject $H_0$. "No trend" |
| Reject $H_0$. "Trend" |

Table 12.1 Pr
$\alpha$ = Prob (

The power $(1-\beta)$ for t
that actually exists is
need a test), so a test r
to be encountered. If
powerful than its alte
used. The test selecte
power over all situatio
Some of the characteri
this chapter, are:

Distribution (no
Outliers (wild v
Cycles (seasonal,
Missing values (
Censored data (l
Serial Correlatio

### 12.1.2 Approaches to T

Five types of trend tests
factors. The first, show
parametric, entirely non
(columns) is whether th
associated variables. Th

to reject $H_0$ does not mean that it was "proven" that there is no trend. Rather, it is a statement that the evidence available is not sufficient to conclude that there is a trend. Table 12.1 summarizes the possible outcomes of a statistical test in the context of trend analysis.

| | True Situation | |
|---|---|---|
| Decision | No trend. $H_0$ true. | Trend exists. $H_0$ false. |
| Fail to reject $H_0$. "No trend" | Probability = $1-\alpha$ | (Type II error) $\beta$ |
| Reject $H_0$. "Trend" | (Type I error) significance level $\alpha$ | (Power) $1-\beta$ |

Table 12.1 Probabilities associated with possible outcomes of a trend test.
$\alpha$ = Prob (reject $H_0$ | $H_0$ true) and $1-\beta$ = Prob (reject $H_0$ | $H_0$ false)

The power ($1-\beta$) for the test can only be evaluated if the nature of the violation of $H_0$ that actually exists is known. This is never known in reality (if it were we wouldn't need a test), so a test must be found which has high power for the kind of data expected to be encountered. If a test is slightly more powerful in one instance but much less powerful than its alternatives in some other reasonable cases then it should not be used. The test selected should therefore be robust – it should have relatively high power over all situations and types of data that might reasonably be expected to occur. Some of the characteristics commonly found in water resources data, and discussed in this chapter, are:

Distribution (normal, skewed, symmetric, heavy tailed)
Outliers (wild values that can't be shown to be measurement error)
Cycles (seasonal, weekly, tidal, diurnal)
Missing values (a few isolated values or large gaps)
Censored data (less-than values, historical floods)
Serial Correlation

### 12.1.2 Approaches to Trend Testing

Five types of trend tests are presented in table 12.2. They are classified based on two factors. The first, shown in the rows of the table, is whether the test is entirely parametric, entirely nonparametric, or a mixture of procedures. The second factor (columns) is whether there is some attempt to remove variation caused by other associated variables. The table uses the following notation:

Y = the random response variable of interest in the trend test,
X = an exogenous variable expected to affect the value of Y,
R = the residuals from a regression or LOWESS of Y versus X, and
T = time (often expressed in years).

Simple trend tests (not adjusted for X) are discussed in section 12.2. Tests adjusted for X are discussed in section 12.3.

| | Not Adjusted for X | Adjusted for X |
|---|---|---|
| **Nonparametric** | Mann-Kendall trend test on Y | Mann-Kendall trend test on residuals R from LOWESS of Y on X |
| **Mixed** | — | Mann-Kendall trend test on residuals R from regression of Y on X |
| **Parametric** | Regression of Y on T | Regression of Y on X and T |

Table 12.2 Classification of five types of tests for trend

If the trend is spatial rather than temporal, T will be downstream order, distance downdip, etc. Examples of X and Y include the following:

- For trends in surface water quality, Y would be concentration, X would be streamflow, and R would be called the flow-adjusted concentration;
- For trends in flood flows, Y would be streamflow, X would be the precipitation amount, and R would be called the precipitation-adjusted flow (the duration of precipitation used must be appropriate to the flow variable under consideration. For example, if Y is the annual flood peak from a 10 square mile basin then X might be the 1-hour maximum rainfall, whereas if Y is the annual flood peak for a 10,000 square mile basin then X might be the 24-hour maximum rainfall).
- For trends in groundwater levels, Y would be the change in monthly water level, X the monthly precipitation, and R would be called the precipitation-adjusted change in water level.

## 12.2 Trend Tests With No Exogenous Variable

### 12.2.1 Nonparametric Mann-Kendall Test

Mann (1945) first suggested using the test for significance of Kendall's tau where the X variable is time as a test for trend. This was directly analogous to regression, where the test for significance of the correlation coefficient r is also the significance test for a simple linear regression. The Mann-Kendall test can be stated most generally as a test for whether Y values tend to increase or decrease with T (monotonic change).

$H_0$: Prob $[Y_j > Y_i] = 0.5$, where time $T_j > T_i$.

$H_1$: Prob $[Y_j > Y_i] \neq 0.5$ (2-sided test).

No assumption of normality is required, but there must be no serial correlation for the resulting p-values to be correct. Typically the test is used for a more specific purpose – to determine whether the central value or median changes over time. The spread of the distribution must remain constant, **though not necessarily in the original units.** If a monotonic transformation such as the ladder of powers is all that is required to produce constant variance, the test statistic will be identical to that for the original units. For example, in figure 12.2 a lognormal Y variable is plotted versus time. The variance of data around the trend is increasing. A Mann-Kendall test on Y has a p-value **identical** to that for the data of figure 12.3 – the logarithms of the figure 12.2 data. The logs show an increasing median with constant variance. Only the central

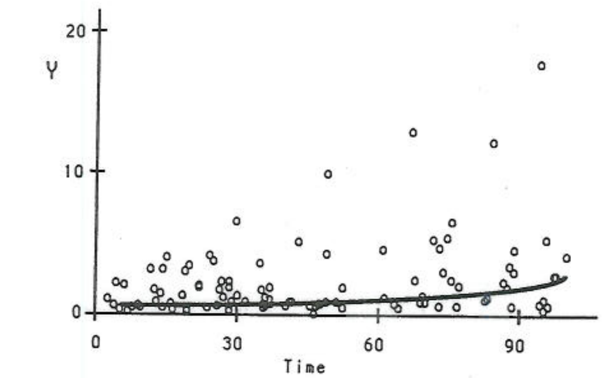



Figure 12.2 Y versus Time. Variance of Y increases over time.

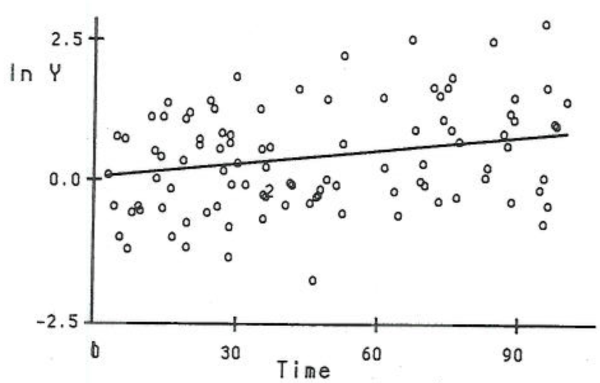



Figure 12.3 Logarithms of Y versus Time. The variance of Y is constant over time.

location changes. The Mann-Kendall test possesses the useful property of other nonparametric tests in that it is invariant to (monotonic) power transformations such as those of the ladder of powers. Since only the data **or any power transformation of the data** need be distributed similarly over T except for their central location in order to use the Mann-Kendall test, it is applicable in many situations.

To perform the test, Kendall's S statistic is computed from the Y,T data pairs (see Chapter 10). The null hypothesis of no change is rejected when S (and therefore Kendall's $\tau$ of Y versus T) is significantly different from zero. We then conclude that there is a monotonic trend in Y over time.

An estimate of the rate of change in Y is also usually desired. If Y or some transformation of Y has a linear pattern versus T, the null hypothesis can be stated as a test for the slope coefficient $\beta_1 = 0$. $\beta_1$ is the rate of change in Y, or transformation of Y, over time.

### 12.2.2 Parametric Regression of Y on T

Simple linear regression of Y on T is a test for trend

$$Y = \beta_0 + \beta_1 \bullet T + \varepsilon$$

The null hypothesis is that the slope coefficient $\beta_1 = 0$. Regression makes stronger assumptions about the distribution of Y over time than does Mann-Kendall. It must be checked for normality of residuals, constant variance and linearity of the relationship (best done with residuals plots -- see Chapter 9). If Y is not linear over time, a transformation will likely be necessary. If all is ok, the t-statistic on $b_1$ is tested to determine if it is significantly different from 0. If the slope is nonzero, the null hypothesis of zero slope over time is rejected, and we conclude that there is a linear trend in Y over time. Unlike Mann-Kendall, the test results for regression will not be the same before and after a transformation of Y.

### 12.2.3 Comparison of Simple Tests for Trend

If the model form specified in a regression equation were known to be correct (Y is linear with T) and the residuals were truly normal, then fully-parametric regression would be optimal (most powerful and lowest error variance for the slope). Of course we can never know this in any real world situation. If the actual situation departs, even to a small extent, from these assumptions then the Mann-Kendall procedures will perform either as well or better (see Chapter 10, and Hirsch et. al., 1991, p.805-806).

There are practical cases where the regression approach is preferable, particularly in the multiple regression context (see next section). A good deal of care needs to be taken to insure it is correctly applied and to demonstrate that to the audience. When one is forced, by the sheer number of analyses that must be performed (say a many-station, many-variable trend study) to work without detailed case-by-case checking of assumptions, then nonparametric procedures are ideal. They are always nearly as powerful as regression, and the failure to edit out or correctly transform a small percentage of outlying data will not have a substantial effect on the results.

Example 1

Appendix C10 lists phosphorus loads and streamflow during 1974-1985 on the Illinois River at Marseilles, IL. The Mann-Kendall and regression lines are plotted along with the data in figure 12.4. Both lines have slopes not significantly different from zero at $\alpha = 0.05$. The large load at the beginning of the record and non-normality of data around the regression line are the likely reasons the regression is considerably less significant. Improvements to the model are discussed in the next sections.

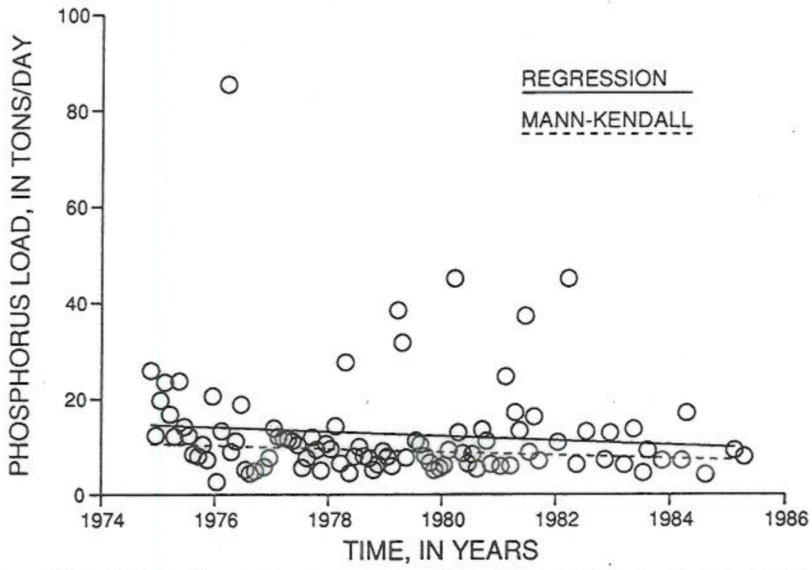



Figure 12.4 Mann-Kendall and regression trend lines (data in Appendix C10).

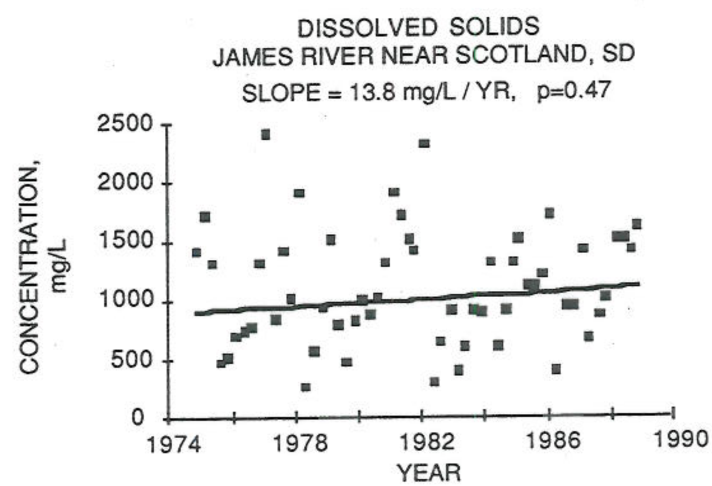
| | | | |
|---|---|---|---|
| Regression: | Load = 16.8 – 0.46•time | t = –1.09 | p = 0.28 |
| Mann-Kendall: | Load = 12.2 – 0.33•time | tau = –0.12 | p = 0.08. |

## 12.3 Accounting for Exogenous Variables

Variables other than time trend often have considerable influence on the response variable Y. These "exogenous" variables are usually natural, random phenomena such as rainfall, temperature or streamflow. By removing the variation in Y caused by these variables, the background variability or "noise" is reduced so that any trend "signal" present can be seen. The ability (power) of a trend test to discern changes in Y with T is then increased. The removal process involves modelling, and thus explaining, the effect of exogenous variables with regression or LOWESS (for computation of LOWESS, see Chapter 10). This is the rationale for using the methods in the right-hand column of table 12.2.

For example, figure 12.5a presents a test for trend in dissolved solids at the James River in South Dakota. No adjustment for discharge was employed. The p-value for the test equals 0.47, so no trend is able to be seen. The Theil estimate of slope is plotted,

a)




b)

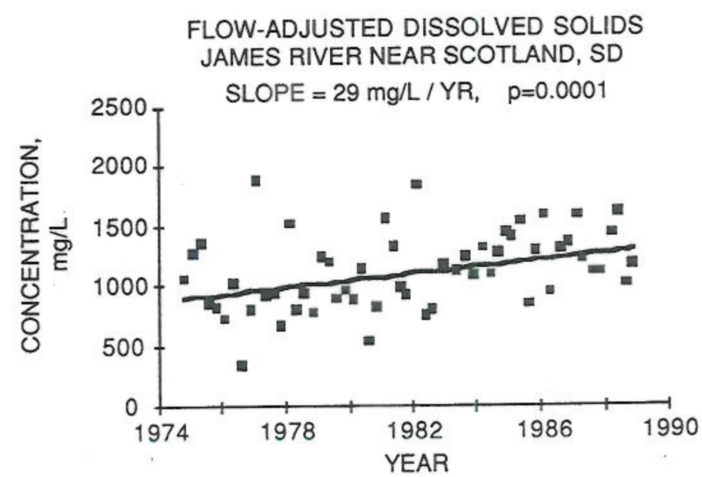
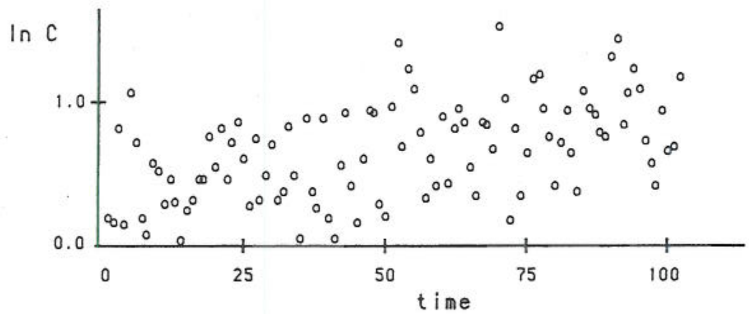


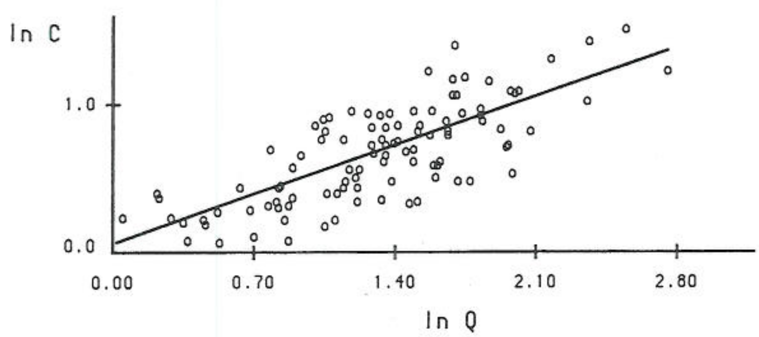

Figure 12.5 Trend tests a) before adjustment for flow. b) after adjustment for flow. (from Hirsch et al., 1991)

showing the line to be surrounded by a lot of data scatter. In figure 12.5b, the same data are plotted after using regression to remove the variation due to discharge. Note that the amount of scatter has decreased. The same test for trend now has a p-value of 0.0001; for a given magnitude of flow, dissolved solids are increasing over time.

When removing the effect of one or more exogenous variables X, the probability distribution of the Xs is assumed to be unchanged over the period of record. Consider a regression of Y versus X (figures 12.6a and 6b). The residuals R from the regression describe the values for the Y variable "adjusted for" exogenous variables (figure 12.6c). In other words, the effect of other variables is removed by using residuals – residuals express the variation in Y over and above that due to the variation caused by changes in the exogenous variables. A trend in R implies a trend in the relationship between X and Y (figure 12.6d). This in turn implies a trend in the distribution of Y over time while accounting for X. However, if the probability distribution of the Xs has changed over the period of record, a trend in the residuals may not necessarily be due to a trend in Y.




12.6a. Log of concentration vs. time. Trend is somewhat difficult to see.




12.6b. Ln of concentration vs. exogenous variable: ln of streamflow (Q).
Strong linear relation shown by regression line.
Expect higher concentrations at higher flows.