WATER RESOURCES RESEARCH, VOL. 27, NO. 5, PAGES 803–813, MAY 1991

Selection of Methods for the Detection and Estimation of Trends in Water Quality

ROBERT M. HIRSCH, RICHARD B. ALEXANDER, AND RICHARD A. SMITH

U.S. Geological Survey, Reston, Virginia

One result of increased scientific and public interest in water quality over the past few decades has been the gradual accumulation of reliable long-term water quality data records and an interest in examining these data for long-term trends. This paper summarizes and examines some of the major issues and choices involved in detecting and estimating the magnitude of temporal trends in measures of stream water quality. The first issue is the type of trend hypothesis to examine: step trends versus monotonic trend. The second relates to the general category of statistical methods to employ: parametric versus nonparametric. The third issue relates to the kind of data to analyze: concentration data versus flux data. The fourth relates to issues of data manipulation to achieve the best results from the trend analysis. These issues include the use of mathematical transformations of the data and the removal of natural sources of variability in water quality due to seasonal and stream discharge variations. The final issue relates to the choice of a trend technique for the analysis of data records with censored or "less than" values. The authors' experiences during the past decade with the development of several trend detection techniques and application of these techniques to a large number of water quality records provide insight into the issues related to a choice of a statistical test for trend in water quality.

Introduction

During the past decade, various nonparametric and seminonparametric techniques for the detection of trends in water quality data were developed and applied by many U.S. Geological Survey investigators. Theoretical investigations that compared the performance of these techniques with their parametric counterparts were conducted [*Hirsch et al.*, 1982; *Hirsch and Slack*, 1984; *Hirsch*, 1988]. The trend methods were also applied to an extremely large number of stream water quality and atmospheric deposition records as part of several investigations [*Smith et al.*, 1982, 1987; *Alexander and Smith*, 1988; *Schertz and Hirsch*, 1985; *Hirsch and Gilroy*, 1985]. The experiences gained during the development and application of these methods provide valuable insight into the various decisions related to the choice of a statistical test for trend in water quality data.

The purpose of this paper is to examine some of the major issues and choices involved in selecting a method for evaluating changes in stream water quality over time. The discussion draws heavily on the authors' experiences in conducting theoretical investigations of particular methodological choices, applying trend detection techniques to numerous water quality records, and advising others on applications of trend techniques. Specific statistical or data manipulation techniques for dealing with issues related to trend detection are either described or references on the methods are provided in the paper. This discussion is not intended to provide a comprehensive review of trend detection methodologies. Instead, the objective is to provide guidance in the selection and use of available statistical techniques for trend detection based on our experiences with a wide variety of trend detection methods.

The choices involved in the selection of a trend detection method discussed here include: (1) the type of trend hypothesis to examine (step trend versus monotonic trend), (2) the general category of statistical methods to employ (parametric versus nonparametric), (3) the kind of water quality data to analyze (concentration versus flux), (4) various data manipulation choices related to the use of mathematical transformations and the removal of natural sources of variability (discharge, seasonality) in water quality, and (5) the choice of a trend detection technique for water quality records with censored data.

Paper number 91WR00259.

Sample Collection and Analytical Methods

It is assumed for purposes of this examination that one or more sets of data, which were collected over a period of years in a consistent and reliable manner, are available to the investigator. This means that the rules for the timing of sample collection must be known (convenience sampling is not acceptable), the methods of sample collection, handling, shipment, preservation, laboratory measurement, and data reporting conventions (rounding and reporting limits) must be constant over the period of record. There can be exceptions to this requirement of constancy. Specifically, if changes have been documented to have no effect on the resulting data, or if changes result in known biases and these biases are subsequently corrected in the data undergoing analysis, then the procedures described here may legitimately be used to examine the data for trend.

Step Trend Versus Monotonic Trend

Two primary types of trends can be considered in hypothesis testing and in trend estimation. One is the step trend hypothesis. This hypothesis assumes that the data collected before a specific time are from a distinctly different population than the data collected after that time. The difference between the populations is assumed to be one of location (e.g., mean or median) but not necessarily of scale (e.g., variance or interquartile range). The other trend hypothesis is that the population shifts monotonically (i.e., no reversals

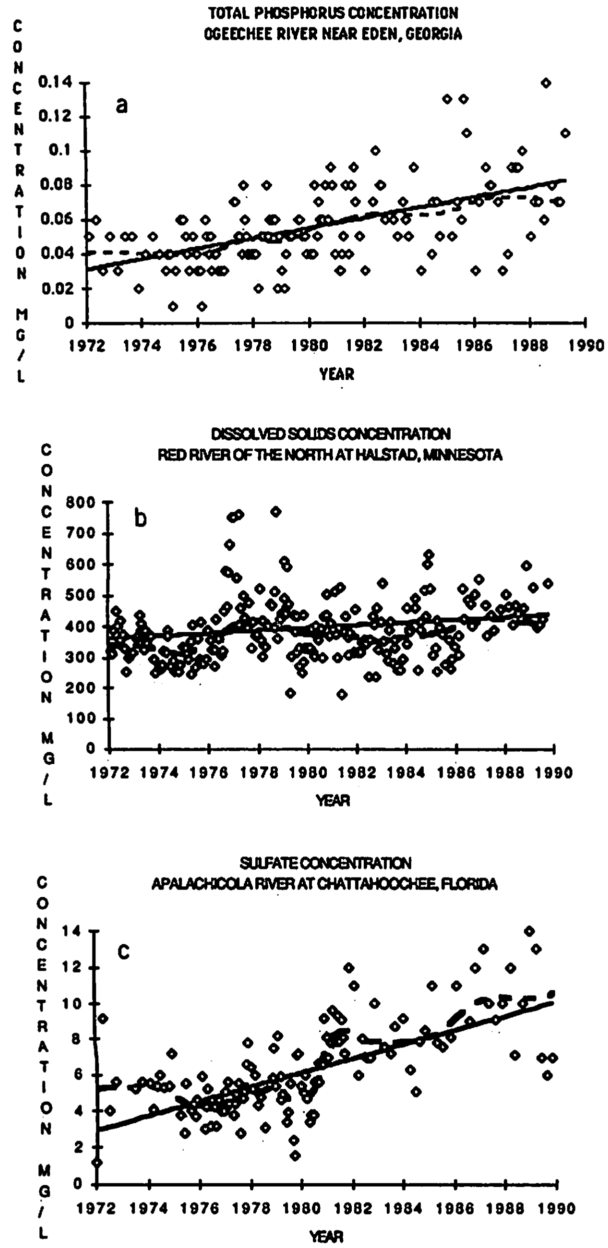

Fig. 1. (*a*) A relatively continuous, monotonic increase ($p < 0.001$) is detected in total phosphorus concentrations for the time period 1972–1989. Trend test is Seasonal Kendall. Solid line is regression estimate. Dashed line is LOWESS. (*b*) An increase ($p = 0.003$) is detected in dissolved solids concentrations for the time period 1972–1989. Within the overall increase there are notable decreases around 1974 and during the period 1978–1983. Trend test is Seasonal Kendall. (*c*) An increase ($p < 0.001$) is detected in sulfate concentrations for the time period 1972–1989. Much of this increase in sulfate concentration occurred as an abrupt rise in 1981. Trend test is Seasonal Kendall.

of direction) over time, but does not specify if this occurs continuously, linearly, in one or more discrete steps, or in any other specific pattern (see examples in Figure 1). The step trend hypothesis is much more specific than the monotonic trend hypothesis. It requires that a particular fact, the time of the change, is known prior to any examination of the data.

Examples of techniques tailored to the step trend alternative include parametric tests like the two sample *t* test [*Iman and Conover*, 1983] and estimates of change magnitude based on the difference in sample means. The nonparametric alternatives to these are the Mann-Whitney-Wilcoxon Rank Sum test [*Bradley*, 1968] and the associated Hodges-Lehmann estimator of trend magnitude [*Hodges and Lehmann*, 1963]. The parametric procedures for the monotonic trend alternative are regression analysis [*Montgomery and Peck*, 1982] of the water quality variable as a function of time. Regression provides a measure of significance based on a hypothesis test on the slope coefficient (or alternatively the correlation coefficient) and a measure of magnitude, the estimated slope. The nonparametric approach would be to use the Mann-Kendall test for trend [*Mann*, 1945; *Kendall*, 1975], which is functionally identical to Kendall's (tau) test for correlation [*Kendall*, 1975], and the associated slope estimate developed by *Sen* [1968]. Numerous variations are possible for each of the procedures mentioned. Several of these are discussed below in the sections on seasonal variation and flow variation.

The step trend procedures should only be used in two specific types of cases. The first is when the record (or records) being analyzed are naturally broken into two distinct periods with a relatively long time gap between them. There is no specific rule to determine how long the gap should be to make this the preferred procedure. If the length of the gap is more than about one-third the entire period of data collection, then the step trend procedure may be best (see Figure 2*a*) even if the actual trend was linear. In general, if the within-period trends are small in comparison to the between-period differences, then the step trend procedures should be used. The other situation is when there is a known event that occurred at a specific time during the record and is likely to have resulted in a change in water quality. The record should be divided into "before" and "after" periods at the time of this known event. The event could be the introduction of a new source of contaminants, reduction in some contaminant due to completion of treatment plant improvements, or the closing of some facility (see example in Figure 2*b*). It is imperative that the decision to use step trend procedures not be based on examination of the data (i.e., the analyst notices an apparent step but had no prior hypothesis that it should have occurred) or on a computation of the time which maximizes the difference between periods. To use such a two-step procedure would have the result of biasing the significance level of the test. Step trend procedures require a highly specific situation, and the decision to use them should be made prior to any examination of the data.

If there is no prior hypothesis of a time of change or if records from a variety of stations are being analyzed in a single study, the monotonic trend procedures are most appropriate. In multiple record studies, even when some of the records have extensive but not identical gaps, the monotonic trend procedures are generally best because comparable periods of record can be more easily examined among all the records. In fact, the frequent problem of multiple starting dates, ending dates, and gaps in a group of records presents a significant practical problem in trend analysis studies. In order to correctly interpret the data, records examined in a multiple station study must be concurrent. For example, it is pointless to compare a 1975–1985 trend at one station to a 1960–1980 trend at another. The difficulty arises in selecting a period which is long enough to

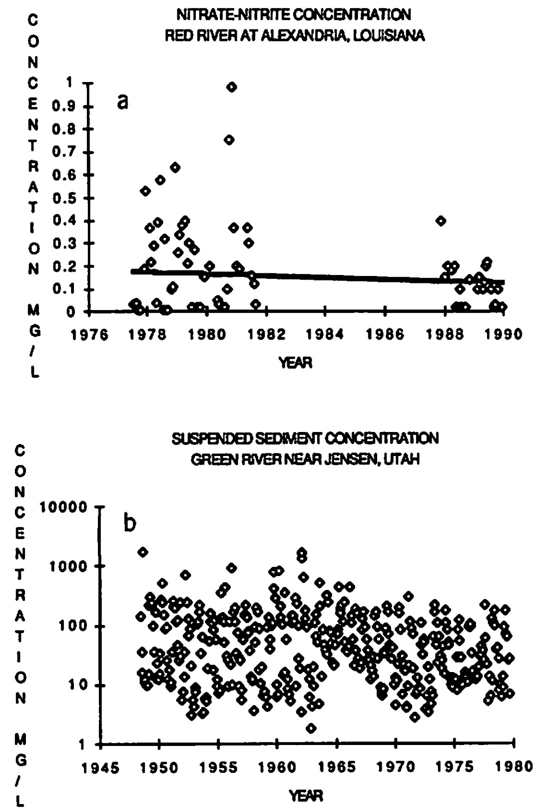

Fig. 2. (*a*) For the Red River at Alexandria, Louisiana, nitrate-nitrite concentrations measured during the 1988–1989 period are found to be significantly ($p = 0.085$) lower than those measured during 1977–1981 when tested with a step trend procedure, the Mann-Whitney Rank Sum test. A monotonic trend procedure, the Seasonal Kendall test, does not detect a significant trend ($p = 0.167$) in concentration for the period of record (1977–1989). The *Sen* [1968] estimate of linear trend associated with the Seasonal Kendall test is shown as a solid line. (*b*) A weakly significant ($p = 0.105$) reduction occurred in suspended sediment concentration in the Green River near Jensen, Utah, following the completion of the Flaming Gorge reservoir (located 93 miles (150 km) upstream of the station) in late 1962. In an application of the Mann-Whitney Rank Sum test, monthly flow-weighted concentrations of suspended sediment were separated into preconstruction (1948–1962) and postconstruction (1963–1979) time periods.

be meaningful but does not exclude too many shorter records.

A further difficulty involves deciding just how complete a record must be to be included in the analysis. For example, if the study is for 1970–1985 and there is a record that runs from 1972 through 1985 it is probably prudent to include it in the study. Furthermore, a 1- or 2-year gap in the middle of the record should not disqualify it from the analysis. More difficult are questions such as inclusion of a 1976–1984 record, or inclusion of a record that covers 1970–1975 and 1982–1985. One reasonable objective rule for deciding to include a record would be as follows: (1) divide the study period into thirds (three periods of equal length), (2) determine the coverage in each period (e.g., if the record is generally monthly, count the months for which there are data), (3) if any of the thirds has less than 20% of the total coverage then the record should not be included in the analysis.

Parametric Versus Nonparametric Methods

The parametric procedures for trend testing are regression in the case of monotonic trend and the two sample t test [*Iman and Conover*, 1983] for step trends. Associated estimators of trend magnitude are the regression slope and the difference in the means, respectively. Nonparametric alternatives to these procedures are the Mann-Kendall test [*Mann*, 1945; *Kendall*, 1975] and the Rank Sum test [*Bradley*, 1968], respectively, and their estimators of trend magnitude are the *Sen* [1968] slope estimator and the Hodges-Lehmann estimator [*Hodges and Lehmann*, 1963]. The Sen slope estimator is the median of all pairwise slopes in the data set. The Hodges-Lehmann estimator is the median of all differences between data in the first data set and data in the second data set. The parametric step trend procedures are special cases of the parametric monotonic trend procedures, and similarly, the nonparametric step trend procedures are special cases of the nonparametric monotonic trend procedures. To apply the monotonic trend procedures in the step trend case, the time variable is treated as a zero for the first data set, and one for the second data set.

Deciding to use one procedure in preference to another should be based on considerations of power and efficiency in the kinds of cases one expects to encounter with actual data. Power is the probability of rejecting the null hypothesis (of no trend) given a particular type and magnitude of actual trend. Efficiency is a measure of estimation error. In particular, a procedure's relative efficiency can be measured by the ratio of the mean square error of a competing procedure to the mean square error of the particular procedure under consideration. For any given significance level, the most powerful test is the parametric procedure if residuals are normally distributed. Similarly, the relative efficiency of these procedures is higher when the residuals are normally distributed. However, what should be at issue in selecting a procedure is not performance under some ideal set of conditions (i.e., normality) but the range of performance abilities that occur for the types of distributions likely to exist in the data to be analyzed.

Hirsch et al. [1982] demonstrated that water quality data are commonly skewed. It is widely recognized that nonparametric procedures can have significantly higher power (or efficiency) than parametric procedures in cases where there is a substantial departure from normality and the sample size is large (see, for example, *Helsel and Hirsch* [1988]). However, there is less confidence among water quality statistics practitioners regarding the effectiveness of nonparametric procedures in cases of minor departures from normality and/or small sample sizes. Many of these practitioners are inclined to consider the parametric procedure as the standard method and only use nonparametric procedures when the data clearly demonstrate that the normal distribution assumption is invalid. Thus is it particularly important to consider cases where the departure from normality is sufficiently small such that visual inspection of the data distribution or formal tests of normality are unlikely to provide evidence for the lack of normality.

The following Monte Carlo analysis compares the performance of parametric and nonparametric methods in cases of small departures from normality and/or small sample sizes. The results of the analysis illustrate that nonparametric methods show modest advantages in terms of efficiency and

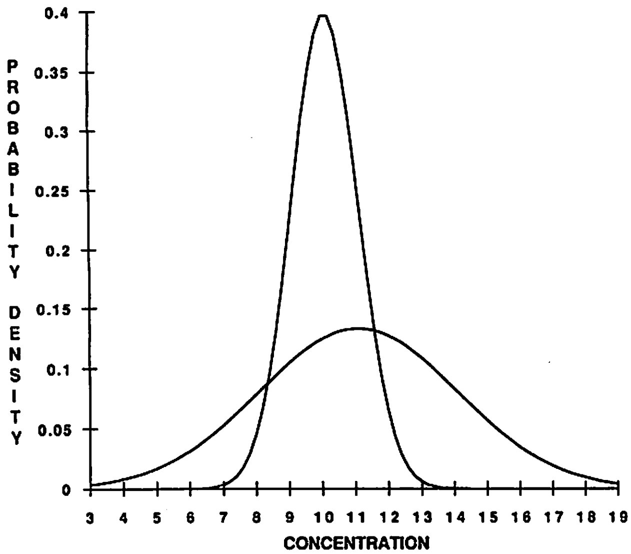

Fig. 3. Normal distributions used in a Monte Carlo analysis of parametric and nonparametric monotonic trend procedures. The first distribution has a mean of 10 and a standard deviation of 1; the second distribution has a mean of 11 and a standard deviation of 3.

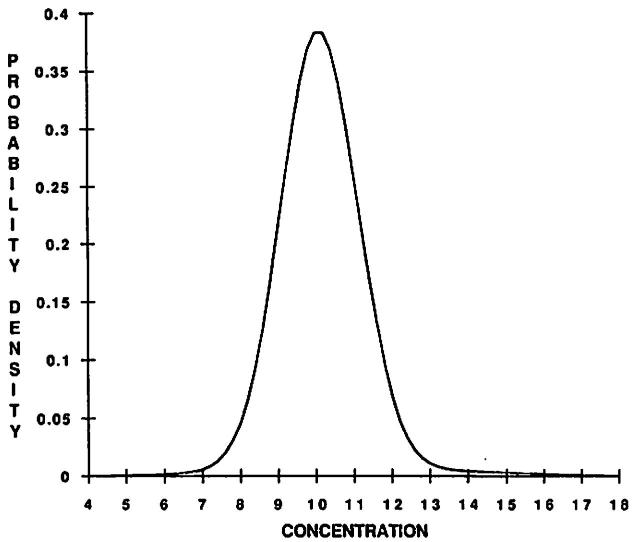

Fig. 5. A normal distribution used in a Monte Carlo analysis of monotonic trend procedures and consisting of a mixture of data from distribution 1 (80%) and from distribution 2 (20%) shown in Figure 3.

power over parametric methods for data sets that depart only slightly from normality. For the experiment, the data are assumed to be distributed as a mixture of two normal distributions. The predominant distribution has a mean of 10 and a standard deviation of 1, the second distribution has a mean of 11 and a standard deviation of 3. Figure 3 displays the two individual distributions and Figure 4 displays a mixture consisting of 95% from the first distribution and 5% from the second. Visual examination reveals only the slightest departure from symmetry. Given the sampling variability that exists in an actual data set, it would be unlikely that samples from this distribution would be identified as nonnormal. Figure 5 displays a more substantial departure from normality; it is a mixture of 80% of the first distribution and 20% of the second. There is a notable difference in the shape of the two tails of the distribution, but, again, the nonnormality is not highly noticeable.

Random samples were generated from each of several different mixture distributions denoted by the percentage of the second distribution in the mixture. The mixtures considered were 0, 1, 2, 3, 4, 5, 7, 10, and 20%. Sample sizes generated were either (N=) 6 or 36. Each sample was treated as a time series, and for each series a slope of the data (versus a time index) was computed by each of two methods: regression and the Sen slope estimator. The population value of the slope of each series was zero so the root mean square error (RMSE) for each estimator is simply the square root of the sums of squares of the estimates over the 1000 Monte Carlo trials considered. The results, expressed as the ratio of RMSE for the Sen estimator to the RMSE of the regression estimator, are shown in Figure 6. These results show that for the larger sample size (N = 36) the regression estimator is more efficient (by less than 10%) when the data are normal,

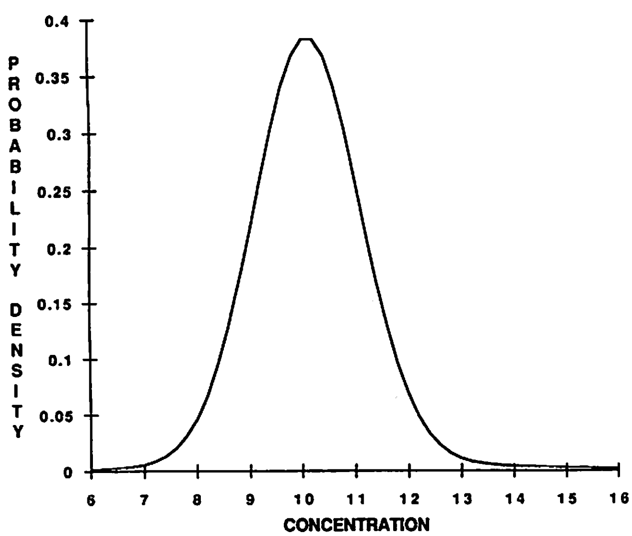

Fig. 4. A normal distribution used in a Monte Carlo analysis of monotonic trend procedures and consisting of a mixture of data from distribution 1 (95%) and distribution 2 (5%) shown in Figure 3.

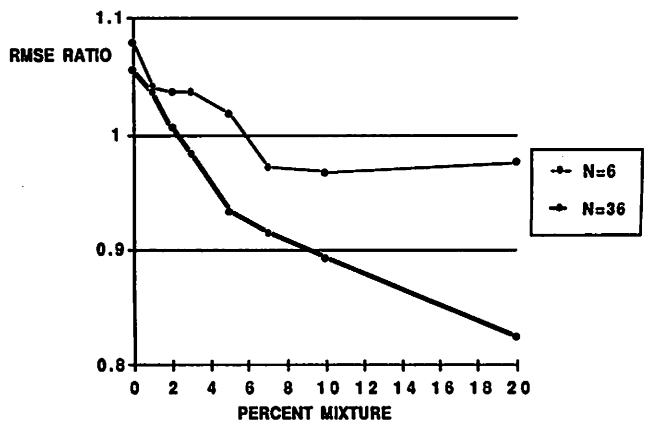

Fig. 6. The relative efficiency of the Sen slope estimator as compared with a regression slope estimator. The efficiency is expressed as the ratio of the RMSE of the Sen slope estimator to the RMSE of the regression estimator and expressed as a function of population mixture and record length.

but with even modest amounts of mixtures the Sen estimator becomes more efficient. In fact, at a 20% mixture the Sen estimator is almost 20% more efficient. Interestingly, when the sample size is very small ($N = 6$, smaller than one would typically have in a trend study), the efficiency remains virtually identical for the two estimators.

In light of these kinds of results, which show that the nonparametric procedures suffer only small disadvantages (in terms of efficiency or power) in the normal case, potentially modest advantages when the data depart slightly (perhaps imperceptibly) from normality, and large advantages when they depart a great deal from normality [see *Hirsch et al.*, 1982]: we have chosen to apply nonparametric procedures routinely in studies involving multiple data sets. It is often argued that one should attempt to transform the data to normality and then carry out the procedure on the transformed data. Such transformations are not always possible (at least with the common, simple transformations) due to heavy tails on the distribution. When such transformations are possible, it may be desirable to do so because the parametric approaches do allow one to consider simultaneously (through multiple regression or analysis of covariance) the effects of multiple exogenous effects such as flow variation or temperature, along with temporal trend. Such simultaneous considerations of effects are difficult with nonparametric techniques.

The use of parametric techniques on transformed data is not well suited to analyses of multiple data sets. The transformation appropriate to one data set may not be appropriate to another. If different transformations are used on different data sets then comparisons among results are difficult, if not impossible. Another reason to avoid the transformation to normality approach is that it contains an element of subjectivity (in the choice of transformation). The argument of the skeptic that "You can always reach the conclusion you want if you manipulate the data enough" is not without merit. The credibility of results is enhanced if a single statistical method is used for all data sets in a study. The parametric methods, to be properly applied, require that judgments be made about model fit, undue influence of outliers, and distribution of residuals. Use of nonparametric methods avoids both the effort and the potential for real or perceived biases being imparted by the data analyst. Consequently, we have used nonparametric procedures in virtually all of the multirecord trend analysis studies we have conducted. However, in an analysis of an individual record, parametric methods, including use of transformations, can be very suitable. Their use requires careful checking of model fit and residuals. They are often more informative than the nonparametric procedures in more complex applications.

CONCENTRATION VERSUS FLUX

Many time series of water quality data consist of a sequence of instantaneous concentration measurements (generally of a large number of chemical species) and concurrent measurements of river discharge. In fact, the existence of concurrent discharge data can be of great value in the interpretation of concentration data as will be discussed below under "removal of variance due to discharge." If discharge and concentration are both available, then the choice can be made between examining trends in concentrations or trends in flux, the product of discharge and concentration. Determination of which variable should be analyzed for trend depends on the question to be answered. For example, if the question is one of ambient quality in the stream, then the concentration would certainly be the appropriate variable to evaluate for trends. The exposure of organisms that reside in the stream to potentially harmful (or beneficial) chemicals is determined by concentrations and the time over which they persist. Flux is of no concern in this example. Similarly, if the question is one of exposure of some facility or population that withdraws water from the stream, then concentration is again the variable of interest.

However, if storage of the water and its constituents is an important factor, then flux may well be the appropriate variable to analyze. For example, the flux of relatively conservative constituents may be of interest in situations where the sampling site is upstream of a reservoir, lake, or estuary where the water has a long residence time (months to years) and the exposure to chemicals by aquatic organisms or populations that ingest the water is of concern. Studies focused on mass balances (changes in the sources and sinks of chemical species in watershed) should also lead to analyses of flux. In addition, if rates of denudation of the landscape or rates of deposition in a large downstream water body are of interest then analyses of flux would be appropriate.

It may be appropriate to evaluate trends in both concentration and flux if there are multiple objectives for the study. Knowing the trends in one of these measures will not necessarily provide a clear indication of the trends one can expect in the other measure. For example, one may find a general upward trend in concentration in a case where there are large increases in concentration occurring at low discharges (associated with increased point source contributions of a contaminant), but at high flows the trends are either nonexistent or so small that they are obscured by the high variability of concentrations typical of high-flow conditions. A trend analysis of flux would be dominated by the ambiguous high-flow information and the large changes in concentration at low flow would be viewed as inconsequentially small.

TRANSFORMATION OF VARIABLES

One feature that is common to a great deal of water quality data is that they depart substantially from a normal distribution. In many cases the concentration or flux data are positively skewed with many of the observations lying close to the lower bound of zero and a few observations lying one or more orders of magnitude above the lower values. If the extent of the analysis to be undertaken is simply a test for trend over time, then the decision to make some monotonic transformation of the data (to render them more nearly normal) is of no consequence provided that a nonparametric test is used. The nonparametric trend tests are invariant to monotonic transformation (such as the logarithm or square root). That means that in terms of significance levels the test results will be identical whether the test was applied to the raw data or the transformed data. The decision to transform data is, however, highly important in terms of fitting various models that are useful in trend analysis such as flow adjustment (discussed in a later section), for computing signifi-

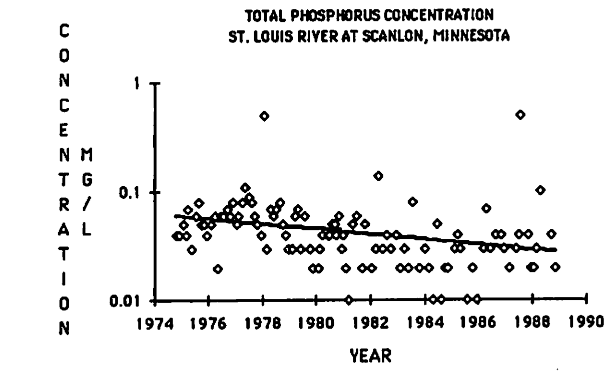

Fig. 7. For the highly skewed concentrations of total phosphorus in this example, a trend is not evident in the raw concentrations ($p = 0.432$ for trend test based on regression). A statistically significant decline ($p = 0.001$) is detected using regression on log-transformed concentrations (as shown by the solid trend line).

cance levels of a parametric test (see Figure 7), and for computing and expressing slope or step size estimates.

Although a monotonic trend is unlikely to approximate a linear pattern over time, one may still want to express as a single linear equation, the history of the trend. This is true particularly in the context of a multiple station trend analysis where the comparison of trend slopes may be of interest. If the actual trend is nonlinear (say, exponential or quadratic) it is quite possible that a linear trend line fitted to the data would predict negative values during some part of the period of record. A fit of this type is certainly not a reasonable approximation of the long-term trend.

One way to ensure that this will not occur is to take a log transformation of the data prior to analysis of the trend. The trend slope will then be expressed in log units. A linear trend in the log units translates to an exponential trend in the original units. To use the log transformation is not equivalent to asserting that the trend is exponential, rather it provides an exponential trend approximation to the actual trend in the data (see example in Figure 8 and Table 1). To make these trend slopes more interpretable, these log concentration slopes can be expressed in percent per year. If B is the estimated slope of a linear trend in natural log units then the

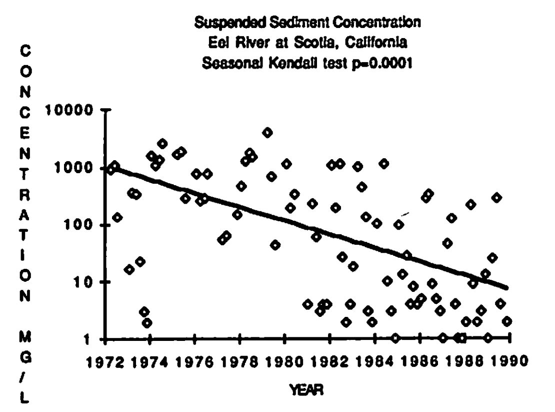

Fig. 8. A linear trend fitted to log-transformed suspended sediment concentrations. In real concentration units this results in an exponential trend. A linear trend fit to the actual concentration data would result in negative fitted concentration values in the period 1986–1990. If the Seasonal Kendall test is used the significance of the trend is identical for tests of concentration or log concentration.

TABLE 1. Predicted Concentrations of Suspended Sediment for the Eel River at Scotia, California, for Selected Years

Equation	Predicted Concentration, mg/L 1972	1980	1990
(1) $C = 249.8 - 18.7T^*$	249	100	−87
(2) $\ln C = 6.924 - 0.279T^*$	1016	109	7

Predictive equations (computed according to Sen) are based upon actual and log-transformed concentrations for the 1972–1990 time period. Equation (1) provides a linear estimate of trend in the actual concentrations. Equation (2) is linear in the logarithms of concentration. Note that equation (1) estimates a negative concentration in 1990.

$^*T$ is the difference in decimal years between the time of interest and the base year of 1972.

percentage change from the beginning of any year to the end of that year will be $(e^B - 1) \times 100$. If the trend is a step trend rather than a monotonic trend and the data were transformed prior to estimating the step size B, then the step size in percentage terms will be $(e^B - 1) \times 100$. If slopes or step sizes in original concentration units are preferred, then rather than multiply by 100 in these expressions one can multiply by some measure of central tendency in the data (a mean or median) to express the slope or step in the original units.

Our experience has been that more resistant and robust results can be obtained if log transformations are used for data that typically have ranges of more than an order of magnitude at a given station (see Figures 7 and 8). We have used transformations in conjunction with parametric tests and with nonparametric tests when the range of variation is quite large. However, in multiple record analyses, the decision to transform was made on the basis of the characteristics of the class of variables being studied and not on a case-by-case analysis. Variables on which log transforms should typically be taken include: concentrations of sediment; total concentration (suspended plus dissolved) for a constituent when the suspended fraction is substantial (for example, phosphorus and some metals); concentrations or counts of organisms; concentrations of substances that arise from biological processes (such as chlorophyll); and flux for virtually any constituent.

REMOVAL OF VARIANCE DUE TO DISCHARGE

In many cases a great deal of the variance in a water quality variable (concentration or flux) is a function of river discharge. This comes about as a result of two different kinds of physical phenomena. One is dilution: a solute may be delivered to the stream at a reasonably constant rate (due to a point source or groundwater discharge to the stream) as discharge changes over time. The result of this situation is a decrease in concentration with increasing flow (see Figure 9*a*). This is typically seen in most of the major dissolved constituents (the major ions). The other process is wash-off: a solute, sediment, or a constituent attached to sediment can be delivered to the stream primarily from overland flow from paved areas or cultivated fields, or from streambank erosion. In these cases concentrations as well as fluxes tend to rise with increasing discharge (see Figure 9*b*). Some constituents can exhibit combinations of both of these kinds of behavior. One example is total phosphorus. A portion of the phospho-

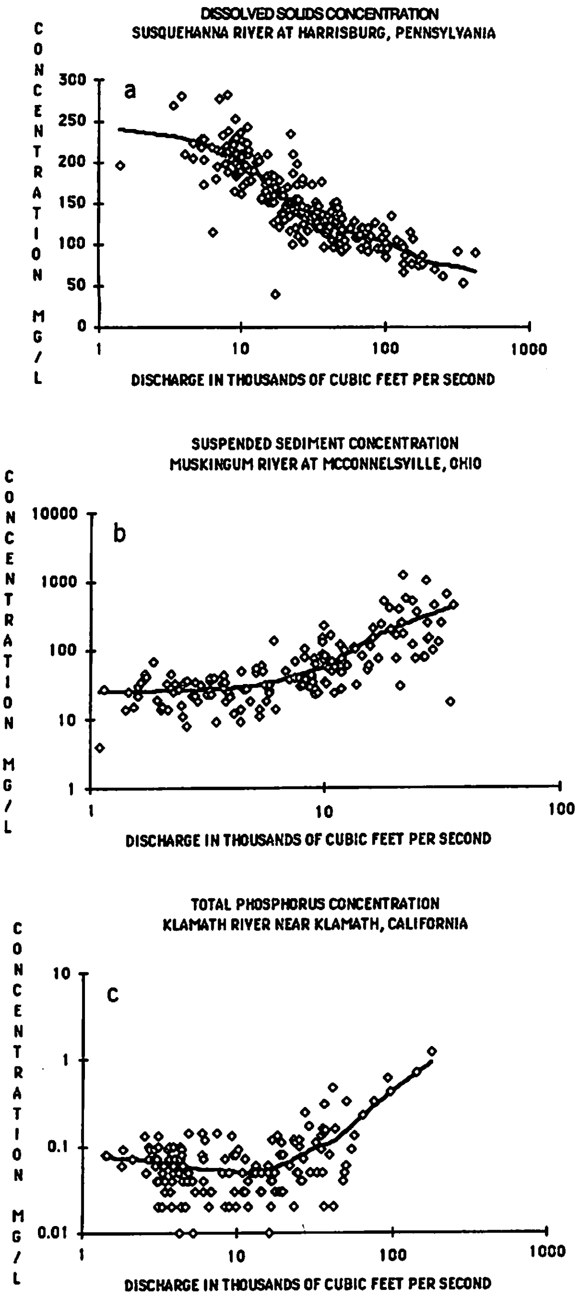

Fig. 9. (*a*) LOWESS curve showing concentrations of major ions released to the stream at a relatively constant rate, diluted by increases in stream discharge. (*b*) LOWESS curve showing concentrations of suspended sediment increasing stream discharge due to the wash-off and transport of larger quantities of suspended sediment with increasing flow. (*c*) LOWESS curve showing that at low to moderate levels of flow, phosphorus is released to the stream at a relatively constant rate, and concentrations of total phosphorus are diluted with increasing discharge. At higher levels of flow, the wash-off and transport of greater quantities of phosphorus lead to increases in concentrations of total phosphorus with increasing discharge (1000 cubic feet per second = 28.3 m^3/s).

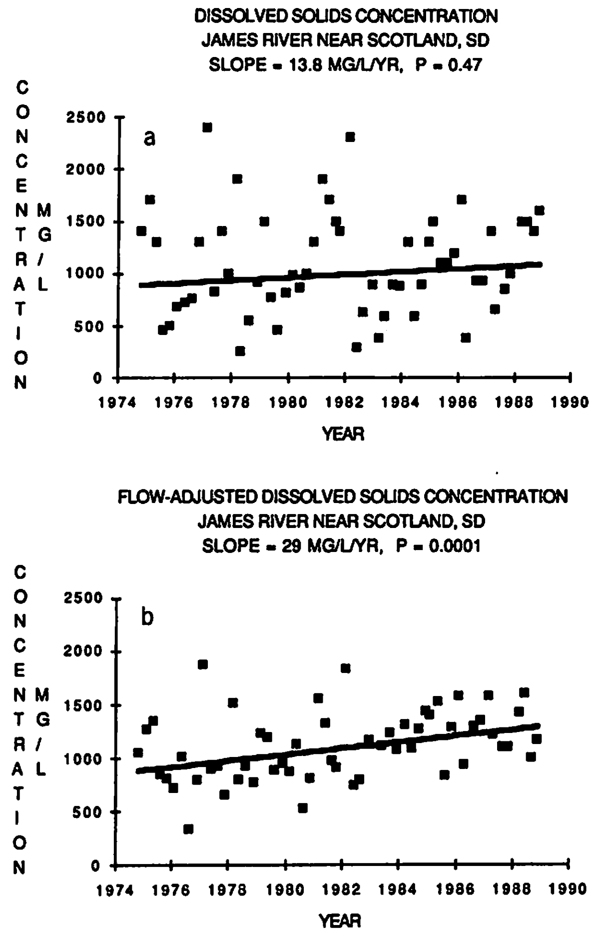

Fig. 10. (*a*) A test for trend in dissolved solids concentrations with the Seasonal Kendall test is not statistically significant (p = 0.47). The Sen estimate of linear trend associated with the Seasonal Kendall test is shown. (*b*) For the same data shown in Figure 10(*a*), following the removal of flow-related variability in dissolved solids concentrations, a test for trend with the Seasonal Kendall test is highly significant (p = 0.0001). The estimate of the magnitude of trend in flow-adjusted concentrations is about twice the estimate of the trend magnitude in raw concentrations. For purposes of constructing the plot, the residuals obtained from the concentration versus flow regression were added to the mean dissolved solids concentration for the period of record.

rus may come from point sources such as sewage treatment plants (dilution effect), but another portion may be derived from surface wash-off and be attached to sediment particles (see Figure 9*c*).

The power and efficiency of any procedure for detecting and estimating the magnitude of trends will be aided if the variance of the data can be decreased (see example in Figure 10). This can be done by removing discharge effects either stagewise or simultaneously [*Alley*, 1988]. In the case of the parametric procedure it is clearly preferable to simultaneously model the flow effect and the trend effect by using multiple regression (for monotonic trend) or analysis of covariance (for step trend). In either case, one uses discharge (or some suitable transformation of discharge) as a covariate. In the nonparametric case, the process must be conducted in stages. The variation due to discharge is modeled by a regression against discharge (or some transformation of discharge) or by some robust curve fitting procedure such as LOWESS (locally weighted scatterplot smoothing [*Cleveland*, 1979]. Then the trend analysis is conducted on the residuals from this relationship [see *Hirsch et al.*, 1982; *Alley*, 1988; *Smith et al.*, 1982]. LOWESS and linear regression fits of concentration and stream discharge are compared for an example data set in Figure 11.

The results of such a trend analysis become, in effect, an analysis of trends in the discharge–water quality relation-

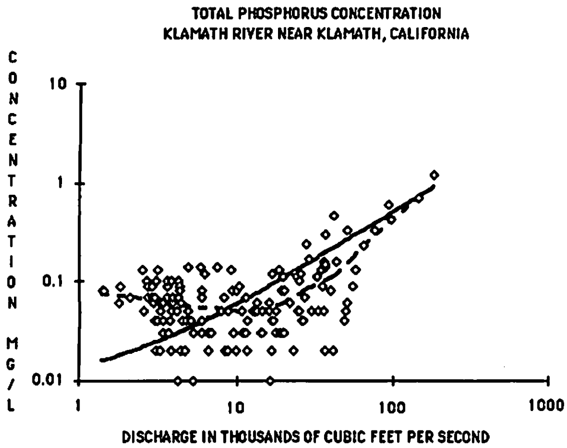

Fig. 11. A comparison of linear regression (solid curve) and locally weighted scatterplot smoothing (LOWESS) (dashed curve) fits of total phosphorus concentrations and stream discharge from the Klamath River at Klamath, California (1000 cubic feet per second = 28.3 m^3/s).

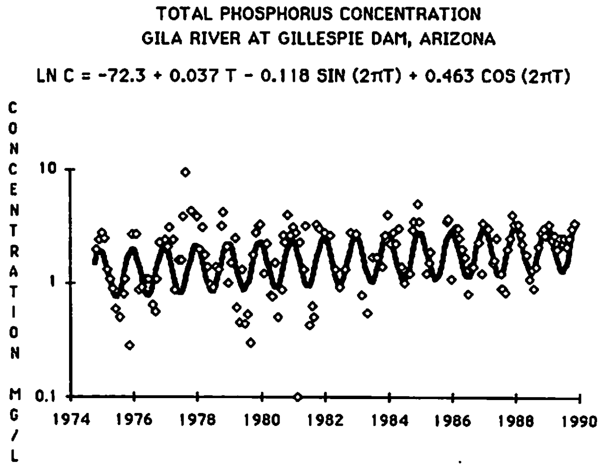

Fig. 12. Multiple linear regression with trigonometric functions of time of year as a test for trend in concentration. Total phosphorus concentrations, C, predicted from a multiple linear regression model involving time, T, and season are shown with a solid curve. A statistically significant ($p < 0.001$) increase in concentration is detected by the model. A linear regression model that does not account for seasonal variability in concentration also detects a significant increase in concentration, but with less statistical confidence ($p = 0.003$).

ship. If the discharge record is stationary (trend free) then the results of such an analysis of residuals becomes an efficient means of detecting and estimating the magnitude of trends in the water quality variable of interest. If the distribution of discharge has changed over the period of analysis, then trends in these residuals does not necessarily translate to a trend in the distribution of the water quality variable. Thus flow adjustment should not be used where human activity has altered the probability distribution of discharge, through changes in regulation, diversion, or consumption during the period of the trend analysis.

Removal of Seasonal Variability

An additional source of variation in water quality data may be described as seasonal variation. Some constituents are influenced by the changes in biological activity (both natural activity and managed activity such as agriculture) in the watershed and in the stream itself. This is certainly true of nutrients due to the seasonal application of fertilizers and the natural pattern of uptake and release by plants. Sediment is also seasonally variable, due to different sources of water dominant at different times of the year. For example, at a given discharge in the spring the source of water may be snowmelt, but in the summer it may be intense rainfall. The seasonal rise and fall of groundwater can also be influential. A given discharge in one season may derive mostly from groundwater while the same discharge during the season of low groundwater levels may derive from surface runoff or quick flow through shallow soil horizons. The chemistry and sediment content of these two sources may be quite different.

Some of the variation that may be initially viewed as seasonal variation can in fact be statistically explained in terms of variation in discharge. However, in many cases even after the discharge effects have been removed, seasonality remains in the data [see *Hirsch et al.*, 1982]. Consequently, whether or not flow effects have been removed, it is desirable to attempt to limit seasonal variations in the data. In parametric procedures this can be done by the use of trigonometric functions of time of year as explanatory variables to remove the effects of an annual cycle (see Figure 12). Another approach is the use of qualitative variables (0 or 1 values) to indicate if a particular data value is in a particular season. In the nonparametric procedures one can remove the effects of seasonality without attempting to explicitly model it as is done in the parametric case. This is accomplished by performing the test on each of the several individual seasons, summing the test statistics and summing their expectations and variances. The overall test for trend can then be carried out by using the summed test statistic and its expectation and variance. One application of this procedure is the Seasonal Kendall test [*Hirsch et al.*, 1982], and another is the Rank Sum test on grouped data [*Bradley*, 1968]. The estimators of trend magnitude are constructed by taking all slopes (in the case of the Sen estimator) or all differences (in the case of the Hodges-Lehmann estimator) within a given season, and finding the median of all of these values over all of the seasons (see *Hirsch* [1988] for a discussion of this seasonally based Hodges-Lehmann estimator).

Tests for Trend in Censored Water Quality Records

Water quality records of some metals and organic compounds (including pesticides) commonly have data values that are censored or reported as less than or equal to the reporting limit of a particular analytical method. This complicates the use of the previously discussed parametric procedures and the stagewise methods for trend detection described by *Alley* [1988] because the arbitrary choice of a value to represent censored values (e.g., zero or the reporting limit) can give inaccurate results for hypothesis tests and biased estimates of trend slopes or estimates of change magnitude.

A parametric approach to the detection of trends in censored water quality data is the estimation of the parameters of a linear regression model relating water quality to time and other explanatory variables through the method of maximum likelihood estimation (MLE), also referred to as Tobit estimation [*Hald*, 1949; *Cohen*, 1950]. The effect of time, discharge, season, and group (in the case of step trends) on water quality may be modeled simultaneously in this approach as can be done in a conventional multiple regression. Because the MLE method assumes a linear model with normally distributed errors, transformations (such as logarithms) of water quality variables and discharge are frequently useful to make the data more nearly normal and improve the fit of the MLE regression. Failure of the data to conform to these assumptions will tend to lower the statistical power of the test and give unreliable estimates of the model parameters. The type I error of the test is, however, relatively insensitive to violations of the normality assumption.

An extension of the MLE method was developed by *Cohen* [1976] to provide estimates of regression model parameters for data records with multiple censoring levels. An adjusted MLE method for multiply-censored data that is less biased in certain applications than the MLE method of *Cohen* [1976] was also recently developed by *Cohn* [1988]. The availability of multiply-censored MLE methods is noteworthy for the analysis of lengthy water quality records with censored values since these records frequently have multiple reporting limits that reflect improvements in the accuracy of analytical methods (and reductions in reporting limits) with time.

The nonparametric procedures (namely the Seasonal Kendall and the Rank Sum test) can be used for the detection of trend in censored water quality data, but their use is restricted to the analysis of non-flow adjusted or raw data values since residuals cannot be computed for censored values in either a regression or a LOWESS smooth. Because the Seasonal Kendall test and the Rank Sum test involve ranked comparisons of data values, only records with a single reporting limit may be tested for trend. However, these tests may be applied to water quality records with multiple reporting limits if all censored and uncensored values less than or equal to the highest reporting limit in the record are considered to be tied with one another. The application of the test under these circumstances may give unsatisfactory results for records where the maximum reporting limit exceeds many detected values in the record (as may occur if the reporting limit has changed significantly over time). This is because the required recoding of data may significantly increase the amount of censored data and possibly restrict the evaluation of trend to a range of concentrations that are rarely observed.

While the sign of the Sen estimate of trend magnitude associated with the Seasonal Kendall test (and Hodges-Lehmann estimator for step trends) is accurate for data records with a large number of censored values, the magnitude of the slope estimate is likely to be in error for highly censored records. The substitution of an arbitarily chosen value between zero and the reporting limit for censored values when applying one of these tests can give biased estimates of the trend slope. While the amount of bias cannot be stated precisely, the presence of only a few nondetected values in a record (less than about 5%) is not likely to affect the accuracy of the trend slope magnitude significantly.

TABLE 2. Options for Testing for Monotonic Trends in Uncensored Water Quality Data

	Not Flow Adjusted	Flow Adjusted
Fully parametric	Regression of C on time and season	Regression of C on time, season, and Q
Mixed	Regression of deseasonalized C on time	Seasonal Kendall on residuals from regression of C on Q
Nonparametric	Seasonal Kendall	Seasonal Kendall on residuals from LOWESS of C on Q

C is concentration, Q is streamflow (may use a transformation of flow), regression on season is using a periodic function of time of year, deseasonalizing can be done by subtracting seasonal medians, Seasonal Kendall test is Mann-Kendall test for trend done for each season (the Seasonal Kendall test statistic is the sum of the several test statistics), and LOWESS is locally weighted scatterplot smoothing.

SUMMARY

Statistical procedures for the detection of monotonic and step trends are summarized for uncensored data in Tables 2 and 3 and for censored data in Tables 4 and 5, respectively. These tables provide a convenient summarization of the various combinations of the techniques described in this paper (although a few specific methods in the tables are not explicitly described in the body of the paper).

The decision to examine water quality data for a step trend (Tables 3 and 5) should be made prior to examination of the data and should not be based on the observation of an abrupt change during the period of record. Analysis for step trends is most appropriate when a specific event occurred that is likely to have resulted in a change in water quality and the record may be clearly divided into a "pre" and "post" period. Testing for step trends may also be suitable in situations where two distinct data collection periods exist separated by years during which data collection was discontinued. In general, the monotonic trend procedures (Tables 2 and 4) are most appropriate for use if no prior hypothesis regarding the timing of a change is known or if multiple records that may be affected by different pollution sources are being analyzed.

Both the monotonic and the step trend procedures in

TABLE 3. Options for Testing for Step Trends in Uncensored Water Quality Data

	Not Flow Adjusted	Flow Adjusted
Fully parametric	Analysis of covariance C on season and group (before and after)	Analysis of covariance C on season, Q, and group
Mixed	Two-sample t test on deseasonalized C	Seasonal Rank Sum on residuals from regression of C on Q
Nonparametric	Seasonal Rank Sum	Seasonal Rank Sum on residuals from LOWESS of C on Q

C is concentration, Q is streamflow (may use a transformation of flow), regression on season is using a periodic function of time of year, deseasonalizing can be done by subtracting seasonal medians, Seasonal Rank Sum test is the Rank Sum test done for each season (the Seasonal Rank Sum test statistic is the sum of the several test statistics), and LOWESS is locally weighted scatterplot smoothing.

TABLE 4. Options for Testing for Monotonic Trends in Censored Water Quality Data

	Not Flow Adjusted	Flow Adjusted
Fully parametric	TOBIT regression of C on time and season	TOBIT regression of C on time, season, and Q
Nonparametric	Seasonal Kendall	No test available

C is concentration, Q is streamflow (may use a transformation of flow), TOBIT regression on season is using a periodic function of time of year, Seasonal Kendall test is Mann-Kendall test for trend done for each season (the Seasonal Kendall test statistic is the sum of the several test statistics).

Tables 2–5 are differentiated on the basis of their parametric and nonparametric characteristics as well as by whether they remove variation due to discharge. Those procedures classified as "mixed" in Tables 2 and 3 have both parametric and nonparametric components that are typically executed in separate steps. In general, the nonparametric and mixed procedures perform appreciably better (greater power and efficiency) than the parametric procedures for the highly nonnormal distributions commonly encountered for many water quality constituents. Even for small departures from normality, the performance of the nonparametric procedures is similar to or better than that for the parametric procedures. The nonparametric and mixed techniques are particularly convenient to use in investigations of multiple data sets because exhaustive checking of distributional assumptions is not required. Moreover, they offer greater comparability of trend results among multiple records than may exist in the use of the parametric procedures possibly requiring different transformations. The parametric methods are frequently more suitable in detailed studies of an individual record where careful verification of the model fit and residuals can be made.

The choice of a procedure involving flow adjustment should be based primarily on the study objectives. If the purpose of the study is to assess the effect of trends in ambient concentrations on the suitability of water for use by humans or aquatic organisms rather than to investigate the cause of trend, then the removal of variability in concentration due to flow (or other natural causes) may not be desirable.

Adjustment for seasonal variability is made in one of three possible ways in the tests for use with uncensored data described in Tables 2 and 3: (1) the use of trigonometric functions of the time of year for the fully parametric procedures; (2) the deseasonalization of concentration prior to trend testing for the nonflow adjusted mixed procedures; and (3) the summation of seasonal test results for the flow-adjusted mixed procedures and for the nonparametric procedures.

TABLE 5. Options for Testing for Step Trends in Censored Water Quality Data

	Not Flow Adjusted	Flow Adjusted
Fully parametric	TOBIT analysis of covariance of C on season and group	TOBIT analysis of covariance of C on season, Q, and group
Nonparametric	Seasonal Rank Sum	No test available

C is concentration, Q is streamflow (may use a transformation of flow), TOBIT regression on season is using a periodic function of time of year, and the Seasonal Rank Sum test is the Rank Sum test done for each season (the Seasonal Rank Sum test statistic is the sum of the several test statistics).

The trend techniques for the analysis of censored data are classified as either fully parametric or nonparametric in Tables 4 and 5. Mixed procedures involving the stagewise methods as described in Tables 2 and 3 are not applicable to censored records because residual values cannot be computed for censored data values. For the same reason, censored data records cannot be flow adjusted with nonparametric procedures. Flow adjustment is only possible with the parametric procedure, Tobit, through the inclusion of discharge as a regression model term. Adjustment for seasonal variability is made either through the use of trigonometric functions of the time of year for the parametric procedure or through the summation of seasonal test results for the nonparametric procedures.

Given the widespread interest in environmental quality, water quality assessments will continue to be an important area of hydrologic investigation. A part of such assesment activity is the collection of water quality data using large networks of stations at which samples are collected and analyzed according to a standard protocol. The proper interpretation of these data for trends requires strict standardization of methods, adequate quality assurance, and the proper application of statistical techniques suited to the characteristics of the data and to the public policy questions of interest.

The statistical tests and estimators described here, along with the use of exploratory data analysis procedures (including some of the types of graphics shown in this paper), can be of great use in providing insights about water quality trends at a given site and about water quality trends over entire regions. The techniques presented here, and suggestions about their applicability, are based on the authors' collective experience with a wide variety of data sets over a period of about a decade. There will continue to be needs to develop and test new methods that improve on these. Two particular issues that need additional development are methods that make the best possible use of existing data (in light of potentially strong serial correlation in the data) in cases where sampling frequencies have changed substantially over time and robust approaches to analysis of data sets with multiple censoring thresholds.

REFERENCES

Alexander, R. B., and R. A. Smith, Trends in lead concentrations in major U.S. rivers and their relation to historical changes in gasoline-lead consumption, *Water Resour. Bull., 24,* 557–569, 1988.

Alley, W. M., Using exogenous variables in testing for monotonic trends in hydrologic time series, *Water Resour. Res., 24,* 1955–1961, 1988.

Bradley, J. V., *Distribution-Free Statistical Tests,* Prentice-Hall, Englewood Cliffs, N. J., 1968.

Cleveland, W. S., Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc., 74,* 829–836, 1979.

Cohen, A. C., Jr., Estimating the mean and variance of normal populations for singly truncated and doubly truncated samples, *Ann. Math. Stat., 21,* 557–569, 1950.

Cohen, A. C., Jr., Progressively censored sampling in the three parameter log-normal distribution, *Technometrics*, *18*, 99–103, 1976.

Cohn, T. A., Adjusted maximum likelihood estimation of the moments of lognormal populations from type I censored samples, *U.S. Geol. Surv. Open File Rep.*, *88–350*, 1988.

Hald, A., Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point, *Skand. Aktuarietidskrift*, *32*, 119–134, 1949.

Helsel, D. R., and R. M. Hirsch, Discussion of paper by R. H. Montgomery and J. C. Loftis, "Applicability of the t-test for detecting trends in water quality variables," *Water Resour. Bull.*, *24*, 201–204, 1988.

Hirsch, R. M., Statistical methods and sampling design for estimating step trends in surface-water quality, *Water Resour. Bull.*, *24*, 493–503, 1988.

Hirsch, R. M., and E. J. Gilroy, Detectability of step trends in the rate of atmospheric deposition of sulfate, *Water Resour. Bull.*, *21*, 773–784, 1985.

Hirsch, R. M., and J. R. Slack, A nonparametric trend test for seasonal data with serial dependence, *Water Resour. Res.*, *20*, 727–732, 1984.

Hirsch, R. M., and J. R. Slack, and R. A. Smith, Techniques of trend analysis for monthly water quality data, *Water Resour. Res.*, *18*, 107–121, 1982.

Hodges, J. L., Jr., and E. L. Lehmann, Estimates of location based on rank tests, *Ann. Math. Stat.*, *34*, 598–611, 1963.

Iman, R. L., and W. J. Conover, *A Modern Approach to Statistics*, John Wiley, New York, 1983.

Kendall, M. G., *Rank Correlation Methods*, 4th ed., Charles Griffin, London, 1975.

Mann, H. B., Non-parametric test against trend, *Econometrica*, *13*, 245–259, 1945.

Montgomery, D. C., and E. A. Peck, *Introduction to Linear Regression Analysis*, John Wiley, New York, 1982.

Schertz, T. L., and R. M. Hirsch, Trend analysis of weekly acid rain data, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *85–4211*, 1985.

Sen, P. K., Estimates of the regression coefficient based on Kendall's Tau, *J. Am. Stat. Assoc.*, *63*, 1379–1389, 1968.

Smith, R. A., R. M. Hirsch, and J. R. Slack, A study of trends in total phosphorus measurements at NASQAN stations, *U.S. Geol. Surv. Water Supply Pap.*, *2190*, 1982.

Smith, R. A., R. B. Alexander, and M. G. Wolman, Water quality trends in the nation's rivers, *Science*, *235*, 1607–1615, 1987.

R. B. Alexander, R. M. Hirsch, and R. A. Smith, U.S. Geological Survey, MS 436, Reston, VA 22092.

(Received October 10, 1990;
revised December 24, 1990;
accepted January 18, 1991.)