THE COMPUTER PROGRAM ESTIMATE TREND (ESTREND), A SYSTEM FOR THE DETECTION OF TRENDS IN WATER-QUALITY DATA

By Terry L. Schertz, Richard B. Alexander, and Dane J. Ohe

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 91-4040

THE COMPUTER PROGRAM ESTIMATE TREND (ESTREND), A SYSTEM FOR THE DETECTION OF TRENDS IN WATER-QUALITY DATA

By Terry L. Schertz, Richard B. Alexander, and Dane J. Ohe

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 91-4040

Reston, Virginia

1991

U.S. DEPARTMENT OF THE INTERIOR

MANUEL LUJAN, Jr., Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information: write to:	Copies of this report can be purchased from:
Chief, Branch of System Analysis U.S. Geological Survey 410 National Center Reston, Virginia 22092	U.S. Geological Survey Books and Open-File Reports Section Box 25425, Mail Stop 517 Federal Center Denver, Colorado 80225-0425

CONTENTS

FIGURES

TABLES

THE COMPUTER PROGRAM ESTIMATE TREND (ESTREND), A SYSTEM FOR THE DETECTION OF TRENDS IN WATER-QUALITY DATA

By Terry L. Schertz, Richard B. Alexander
and Dane J. Ohe

ABSTRACT

Computerized statistical and graphical procedures were developed for use in U.S. Geological Survey (USGS) investigations of trend in stream water-quality data. These procedures, identified as EStimate TREND (ESTREND), are described in this paper to assist USGS investigators involved in multiple-station studies of water-quality trends. Additional discussion focuses on certain statistical and operational decisions required in multiple-station analysis of trends. The statistical methods used in ESTREND overcome common statistical problems encountered by conventional statistical trend techniques in the analysis of water-quality data. The problems include data that are non-normal and seasonally varying and water-quality records with missing values, "less-than" (censored) values, and outliers, all of which adversely affect the performance of conventional statistical techniques. Parametric and nonparametric statistical trend tests are used in ESTREND. A nonparametric method, the Seasonal Kendall test, is used for data that have few less-than values or data that have been censored at only one reporting limit. A parametric test for trend involving a maximum likelihood estimation method is used for data that have been censored at multiple reporting limits. The Seasonal Kendall test for uncensored data allows for the removal of flow variability in water-quality data which improves the performance of the statistical trend tests. Menu-driven procedures in ESTREND allow the user to easily retrieve water-quality data, analyze data for trend, and view tabular and graphical results of analyses.

INTRODUCTION

Techniques for trend estimation that perform well with the inherent characteristics of water-quality data have been developed by the U.S. Geological Survey (USGS). Procedures that use these techniques have been incorporated into a software system. The system, called <u>ES</u>timate <u>TREND</u> (ESTREND), and the methods used by ESTREND are described in this report.

<u>Background</u>

The U.S. Geological Survey (USGS) has been involved for the past 10 years in the analysis of trends in water-quality data collected from its stream water-quality networks. Initial work focused on the development of trend detection methods that would perform well for water-quality data. Because water-quality data are commonly non-normal and seasonally varying and records of water quality may have missing values, "less-than" values, and outliers, alternatives to conventional parametric trend techniques (which are adversely affected by these characteristics) were investigated by Hirsch and others (1982). They found that a nonparametric trend test, the Seasonal Kendall test, performed significantly better than its parametric counterparts for data sets that exhibited these statistical characteristics.

Less-than or censored values, which are frequently reported for trace metals, nutrients, and pesticides, commonly bias the results of conventional parametric tests for trend because their use is dependent upon the arbitrary selection of a value (for example, zero or the reporting limit) to represent less-than values. Censored values do not present any problem in applying the Seasonal Kendall test provided that a single reporting limit exists for the record (Hirsch and others, 1982). For water-quality records with multiple reporting limits, a parametric procedure involving a maximum likelihood estimation (MLE) method (Cohen, 1976) can be used to estimate unbiased parameters of a linear regression model relating water-quality to time. An adjusted maximum likelihood estimation method (called Tobit), which gives less biased estimates than conventional MLE methods, was recently developed (Cohn, 1988) and adapted for use in trend testing.

Most recently, discussions about the major issues of choosing an appropriate trend detection procedure such as (1) step trend versus monotonic trend, (2) parametric versus nonparametric tests, (3) concentration versus load data, (4) censored versus uncensored data, and (5) removal of sources of variability in the data have been consolidated in a paper by Hirsch and others (1991).

Several trend investigations of national and state-level stream water-quality records have been conducted using the Seasonal Kendall trend test and the Tobit procedure. A trend study by Smith and others (1987) and Alexander and Smith (1988) applied the Seasonal Kendall trend test to water-quality data collected from a USGS national network of surface water-quality stations. More recently, both the Seasonal Kendall trend test and the Tobit procedure were applied to USGS stream water-quality data in trend studies in Texas (Schertz, 1990) and New Jersey (Hay and Campbell, 1991). These latter two studies were initiated to augment and extend the results of the national study.

Purpose and Scope

The purpose of this report is to document a software system for the analysis of time trends in water-quality data. The system, EStimate TREND (ESTREND), was initially developed to assist in the national (Smith, and others, 1987) and state-level (Schertz, 1990; Hay and Campbell, 1991) USGS investigations of trends in stream water quality. Certain statistical and operational issues such as (1) the choice of a statistical test for trend, (2) methods of adjusting for flow and seasonal variations, (3) the choice of a minimum number of observations for trend testing, and (4) the selection of water-quality stations are discussed in this report. Decisions related to these issues are based on methods investigations cited in the background section and experience from national and state-level trend investigations.

The data manipulation tools and the statistical and graphical procedures used in ESTREND were adapted primarily to address circumstances encountered in the earlier USGS trend investigations. In many respects, however, the water-quality data examined in these studies were sufficiently complex (in terms of sample collection frequency, and censoring levels, for example) to provide a good basis for the design and testing of the software. ESTREND is best suited to multiple-station investigations where comparisons of water-quality trends and other summary statistics among stations are of interest. Statistical procedures other than those in ESTREND (e.g, regression) may be more appropriate for detailed statistical investigations of a small number of water-quality records.

For potential non-USGS users of ESTREND, it is important to note that the software is written to conform to the version of FORTRAN available on Prime mini-computer systems[1], which are currently used by the Water Resources Division of the USGS. Additionally, certain components of ESTREND rely on mathematical and graphical modules that are available on these computer systems. ESTREND will not currently operate on other computer systems, and its transport to other systems will require modifications to the existing computer code.

The following discussion covers appropriate station and data characteristics for trend testing and a description of the statistical trend methods and data management software.

STATION AND DATA CHARACTERISTICS

Water-quality records should be selected for a trend study to meet the criteria defined in the study objectives and to satisfy the data-requirements of the trend test. The geographic location, sample collection schedule, sample collection methods, analytical methods and period of record are elements that influence the data collected at each station and also determine the suitability of the data for trend testing. These factors should be evaluated in the station selection process.

Geographic Locations of Stations

The geographic locations of stations selected for a trend study should cover the area specified by the objectives of the study. If specific water-quality issues are identified in the study objectives, stations with basin characteristics that are relevant to these issues should be selected. For example, an investigation of the effect of changes in agricultural activity on trends in surface water-quality should primarily examine water-quality records for stations from predominantly rural, agricultural stream basins. If, on the other hand, water-quality trends in the major rivers of a region or state are of interest, stations should be selected without consideration of specific problem areas, and should give a balanced geographic coverage of the region's or state's waters. The stations used in a recent study of water-quality trends in New Jersey (Hay and Campbell, 1991) were part of a sampling network that was established to cover the major tributaries and segments of the rivers in New Jersey (fig. 1). Specific water-quality problems were not considered in the site selection of the stations included in the network.

---

[1]The use of trade names is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

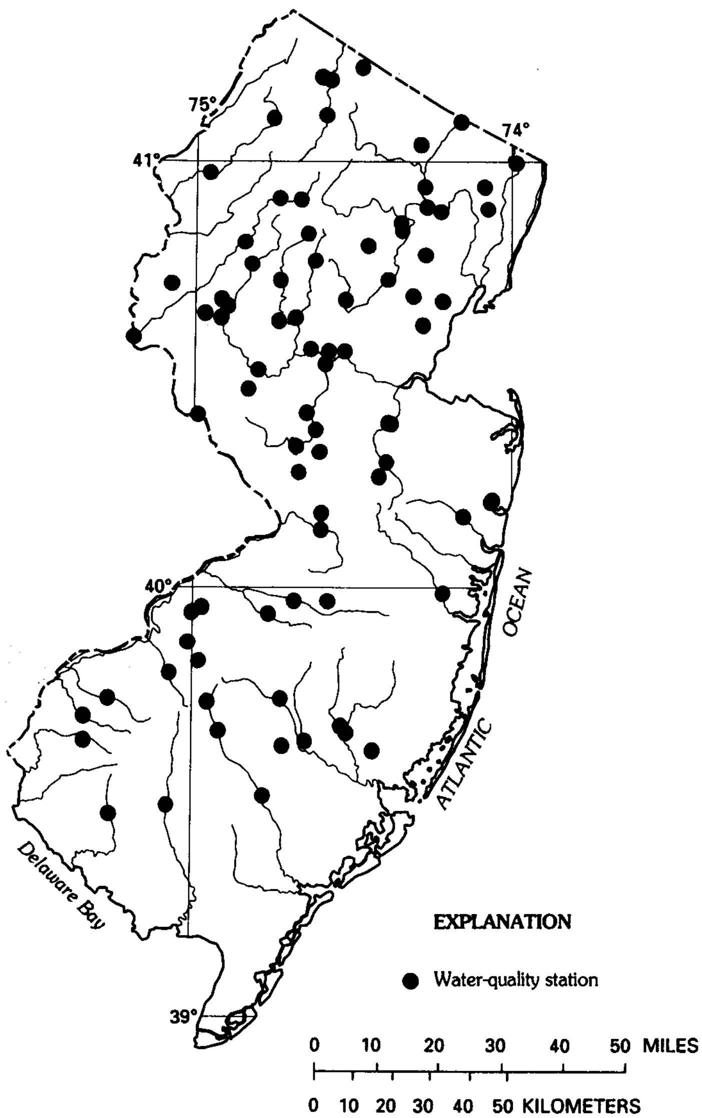

Figure l.—Location of New Jersey water-quality stations used in trend study.

Similarly, the stations used in a recent investigation of water-quality trends in Texas (Schertz, 1990) were selected from more than 400 possible stations to obtain a balanced coverage of the state (fig. 2). Stations were generally selected without knowledge of specific pollution sources. A few stations that had been moved upstream or downstream during the study period were eliminated to maintain consistency in the data collection process for trend detection purposes. Relocation of a station, even a minor relocation, could result in a trend that simply reflects a change due to differences in water quality at the two station locations rather than an actual change in the ambient water quality at a single river location. However, overlapping record that demonstrates similarity between the data from the two locations may be used to qualify a relocated station that is vital to a study.

Sample Collection

The collection of samples at a station can be scheduled to satisfy many objectives. For example, if the objective of sample collection is to study the effects of storm-water runoff at the selected site, then the collection of samples would be intermittent based on the occurrence of rainfall. The majority of the samples from this approach would be collected at streamflow levels greater than normal. If the objective of the sample collection is to study the daily fluctuations of dissolved oxygen at the selected site, then samples would be collected at regularly spaced intervals during a 24-hour period. Only the streamflow levels that occur within the selected 24-hour period would be represented in the sample data.

An investigation of water-quality trends should use station water-quality records with samples collected at regularly spaced intervals within a year to preclude any temporal bias in the sample data. The sample collection schedule should not include any planned bias toward any particular level of streamflow. The number of samples collected within each year may vary (although consistency in sampling frequency is desirable) as long as water-quality samples are collected according to a regular, fixed schedule. The methods used to collect, preserve, handle, ship, and analyze samples and report data values should remain constant during the period of study for a trend analysis study as well. Changes in these methods may confound the detection of trends in ambient water quality or may even directly cause trends to appear in water-quality records.

Period of Record

The period of record for stations included in a trend study should be at least five years for monthly data or longer for less dense data. Decision rules related to record length and the minimum number of samples required for monotonic trend testing are given in sections 4 and 5. If trend results are to be compared between stations and constituents, water-quality records selected for trend analysis should be from approximately concurrent periods.

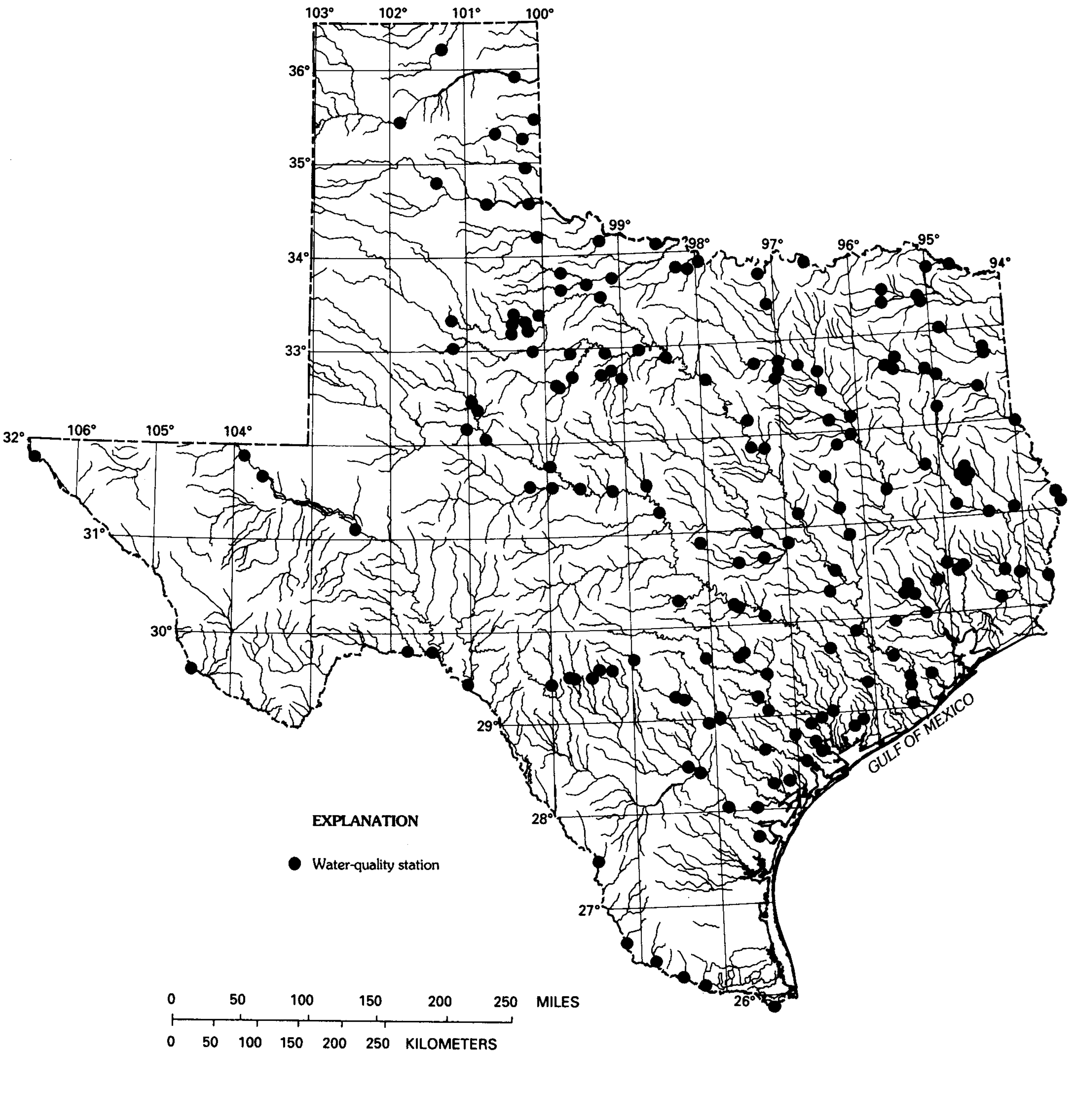

Figure 2.—Location of Texas water-quality stations used in trend study.

OVERVIEW OF STATISTICAL PROCEDURES FOR DETECTION OF TRENDS IN WATER-QUALITY DATA

Scope of the Methods

The statistical trend procedures in ESTREND are for the detection of monotonic changes in water quality with time, which may occur gradually over time or as an abrupt change. While the ESTREND procedures provide important summary statistics on the degree of monotonic trend in water-quality for a selected time period, these trend statistics may not adequately describe short-term changes that may be present in a water-quality record. A more complete picture of water-quality trends may be obtained by testing different portions or time intervals of a water-quality record for trend through multiple runs of the ESTREND procedures. In addition, visual examinations of water-quality records supplemented with smoothing techniques such as LOWESS (Locally Weighted Scatterplot Smoothing; Cleveland, 1979), which is included as part of the graphical features of ESTREND, provide the most direct information about short-term changes in water quality. The example time series plot shown in figure 3 underscores the importance of examining different portions of a water-quality record for trend. A linear model fit to total phosphorus concentrations displayed in figure 3 shows that no trend occurred during the period 1975 to 1989. However, closer examination of the data and nonlinear LOWESS smooth suggest that a slight increase occurred in concentration from 1975 to 1981 followed by a decrease and leveling off of concentration after 1981.

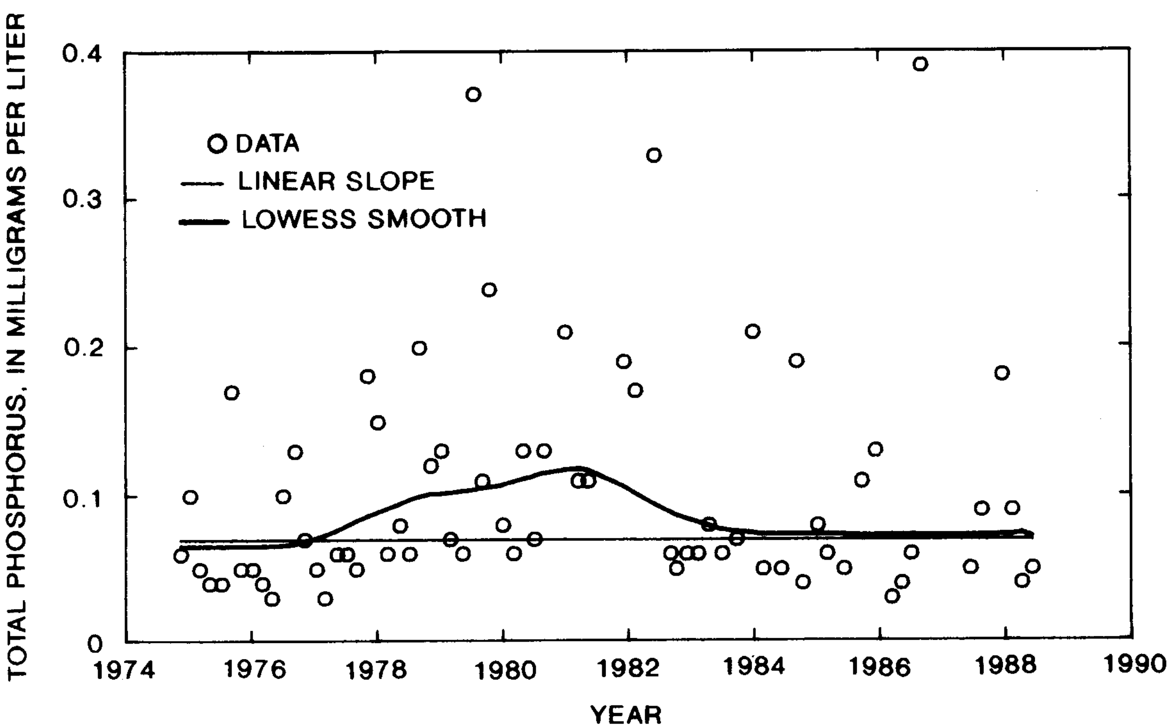

Figure 3.—Non-monotonic variations in total phosphorus for the Kissimmee River at S-65E near Okeechobee, Florida.

Water-quality records tested by ESTREND procedures may have been collected continuously with time according to a relatively fixed schedule or may have a moderately large gap in the middle of the record that separates data collection periods. The statistical tests described here perform adequately in both situations, although other statistical procedures may be more appropriate (i.e., more powerful[2]) for the detection of Water-quality records tested by ESTREND procedures may have been collected continuously with time according to a relatively fixed schedule or may have a moderately large gap in the middle of the record that separates data collection periods. The statistical tests described here perform adequately in both situations, although other statistical procedures may be more appropriate (i.e., more powerful[2]) for the detection of abrupt shifts or step changes in water quality with time for which there is a known reason for the change (see Hirsch and others, 1991). If the trend results are to be compared among many stations where the causes of trend and the timing of these causes are quite variable and possibly not well understood, then the results of monotonic trend procedures are more easily compared and summarized from one site to another than similar results of a step trend procedure.

Summary of the Methods

For trend investigations involving the analysis of many water-quality constituents collected at numerous stations, nonparametric tests for trend have distinct advantages over their parametric counterparts. Conventional parametric statistical procedures assume that the sample data are from a symmetric, bell-shaped Gaussian (normal) distribution. Nonparametric statistical procedures, which involve comparisons of ranks rather than actual data values, are not subject to distributional assumptions, and are typically more powerful than their parametric counterparts for data that violate normality assumptions (Helsel, 1987). Water-quality data for constituents such as suspended sediment, nutrients, bacteria, and some common dissolved ions are frequently asymmetrically distributed (fig. 4). Thus, nonparametric tests for trend are likely to be more powerful than conventional parametric techniques in the analysis of data for many of these water-quality constituents.

In the case of multiple-station trend studies, the large number of data records prohibit case-by-case checking to verify that the assumptions of parametric tests are satisfied because this process is too time consuming. Moreover, tests for normality of sample data are often unsatisfactory as they are unlikely to detect any but the most extreme violations if the sample size is small (Hirsch and others, 1982). Nonparametric tests have nearly as much power (approximately 95 percent as efficient) as parametric tests when distributional assumptions are satisfied (Hirsch and others, 1982). In addition, nonparametric tests are more easily applied to the large numbers of data records examined in many trend investigations because extensive verification of assumptions is not required.

Conventional parametric tests for trend are difficult to apply to water-quality records with large numbers of "less-than" values. The data for certain water-quality constituents such as trace metals, pesticides, and in some cases, nutrients frequently occur in very low concentrations in natural waters such that concentrations are often measured below the analytical reporting limit. When applied to records with censored values, conventional parametric trend methods require the arbitrary selection of a value (e.g., the reporting limit or zero) to represent less-than values. This substitution is likely to produce biased estimates of trend slopes, and give unreliable estimates of the Type I error of hypothesis tests (Helsel and Cohn, 1988).

[2]Power is defined as the probability that the test indicates trend (reject the null hypothesis of no trend) when the generating process that is sampled actually has trend.

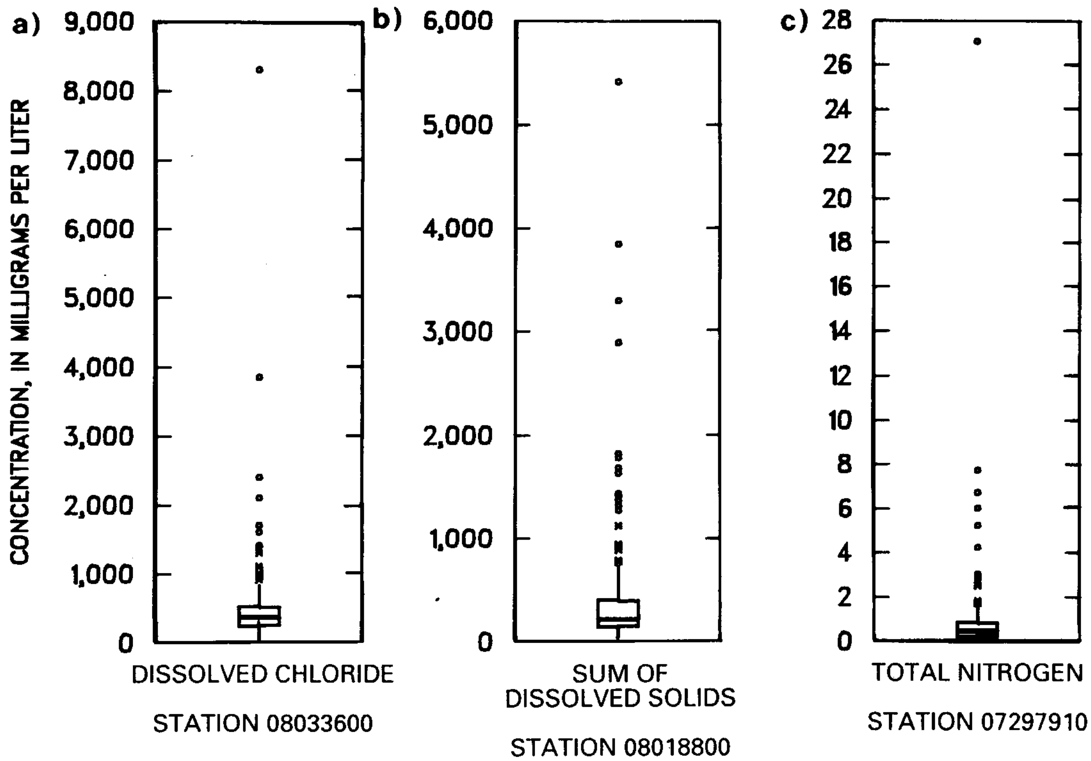

Figure 4.—Examples of non-normal water-quality data; (a) dissolved chloride, (b) dissolved solids, and (c) total nitrogen.

The nonparametric and parametric trend detection techniques available in ESTREND are well suited for the analysis of uncensored and censored water-quality records. Two trend detection techniques are provided in ESTREND; a nonparametric method (Seasonal Kendall) for use with uncensored data or data censored with only one reporting limit (Hirsch and others, 1982), and a parametric method (Tobit) for use with data censored with multiple reporting limits (Cohen, 1976; Cohn, 1988). For each water-quality constituent, one of these tests (as selected by the user) is applied to all station records so that the results of the trend procedure can be easily compared among stations.

The decisions involved in the choice of an ESTREND trend procedure are displayed in figure 5. The Seasonal Kendall procedure is assumed to be preferable in this decision tree because of the advantages that nonparametric methods typically offer in multiple-station water-quality investigations. For highly censored data records, the choice of the Tobit procedure may be preferable, although this choice will depend greatly upon the number of reporting limits

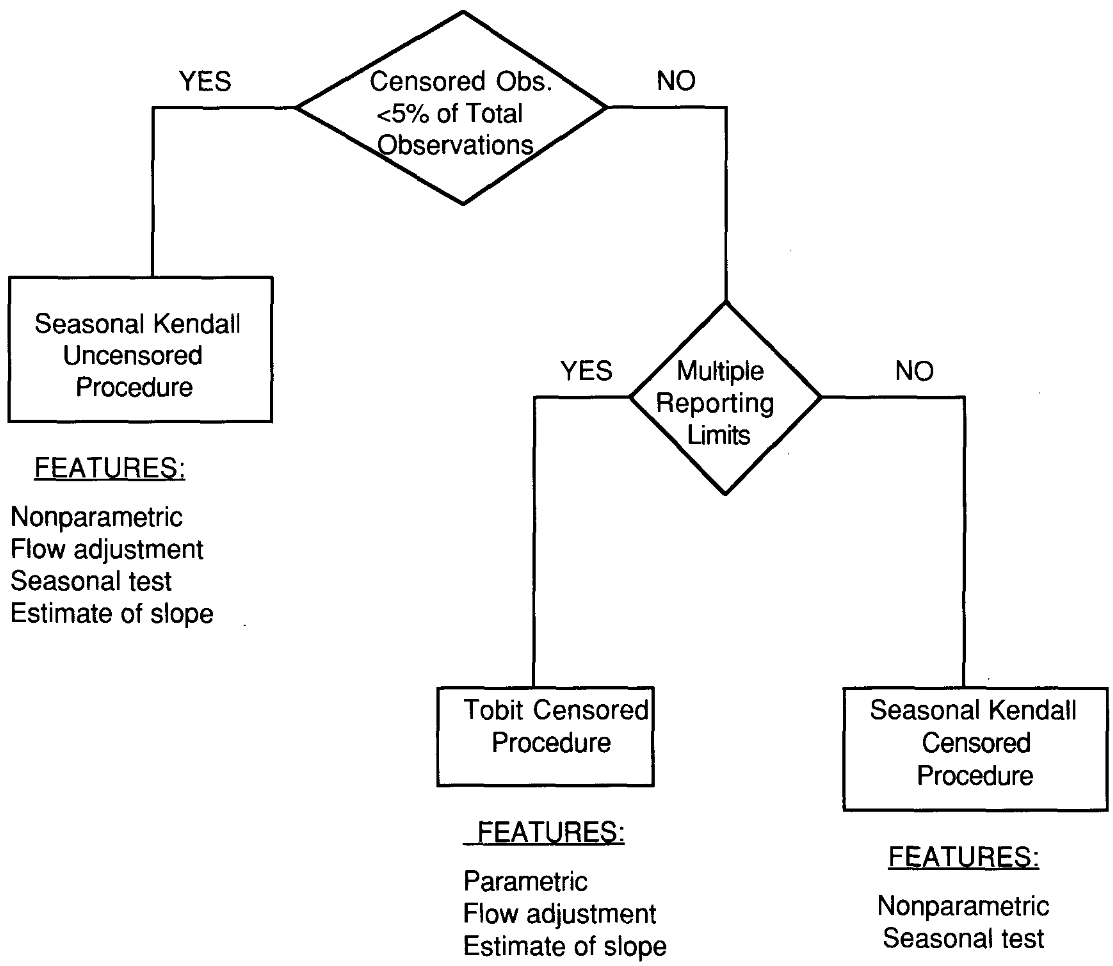

Figure 5.—Decision rules for the selection of a statistical trend test in ESTREND.

in the data record, the type of statistical trend information that the user requires (e.g., trend in flow-adjusted data or an estimate of the trend slope), and whether a nonparametric or parametric hypothesis test is desired. Three applications of the statistical trend tests may be selected by the user in ESTREND:

Seasonal Kendall Test for Uncensored Data

The Seasonal Kendall trend test is generally recommended for use on water-quality records with little or no censoring (less than about five percent of the data record censored). In this application of the test, censored values are assigned a value of one half of the reporting limit. If applied to records with relatively few less-than values, this recoding of censored values is not likely to significantly bias the results of the trend procedure. This application of the test accounts for seasonal-and flow-related variations in water-quality data which enhances the statistical power of the test. Trend results for non-flow adjusted data are also computed in this procedure.

Seasonal Kendall Test for Censored Data

For water-quality records with a large number of observations (more than about five percent of the record) censored at a single reporting limit, the Seasonal Kendall test may also be suitable for testing for trend. In this application of the test, all values (detected and nondetected) that are less than the specified reporting limit are considered tied (nondetected values are assigned a value of zero in the computer code to distinguish them from detected values that are rounded to the reporting limit). This application of the test legitimately evaluates the occurrence of trend in concentrations above the reporting limit. For multiply-censored records, the highest reporting limit in the record should be used in the application of the test. The required recoding of all detected and nondetected values below this reporting limit may, however, result in the loss of significant amounts of data. Variability caused by flow cannot be reliably removed from water-quality records with a large number of censored values, and removal is not attempted as part of this particular application of the trend procedure. The presence of a large number of censored values in the record also adversely affects the estimation of the trend slope, which may not be reliably reported with this application of the procedure.

Tobit test for censored data

For records with large amounts of data censored at one or more reporting limits (more than about five percent of the record), the parametric trend procedure, Tobit, may be used to test for trend. This trend procedure gives a reliable estimate of the trend slope and significance levels when a reasonable fit to the data is obtained, and when the assumptions of the test are satisfied (i.e., regression residuals are approximately normally distributed). Because careful scrutiny of the Tobit model fit is recommended (primarily to check for the presence of outliers), scatterplots and the trend results should be examined for each station and constituent. The current version of this procedure in ESTREND does not allow for the reduction of flow- or seasonal related variability in water quality.

Testing criteria

For the three trend methods used in ESTREND, criteria are established to evaluate whether a water-quality record has a sufficient number of observations and sufficient density of data over the period of interest to test for trend and estimate other summary statistics such as moments. The selected data requirements for trend testing and estimation of moments are based on experiences in applying these tests as well as generally accepted requirements for the application of the statistical tests. These criteria typically represent minimum data requirements for trend testing. For certain trend testing criteria where the minimum data requirements are less clear, the user is given general guidance in the selection of appropriate minimum acceptable requirements for trend testing, but the choice of minimum requirements is the responsibility of the user.

STATISTICAL PROCEDURES FOR UNCENSORED WATER-QUALITY DATA

Seasonal Kendall Test

The Seasonal Kendall test (Hirsch and others, 1982) is a nonparametric test for monotonic trend in water quality. The test, which is a generalization of the Mann-Kendall test (Mann, 1945; Kendall, 1975), reduces the adverse effect that seasonal differences in concentration may have on trend detection by only making comparisons of data from similar seasons (see section 4-2). The null hypothesis for the Mann-Kendall test is that the probability distribution of the random variable has not changed over time. The test assumes only that the values of the random variable are independent and are from the same statistical distribution. The test makes all possible pairwise comparisons of a time ordered set of water-quality values. If a later value (in time) is larger, a plus is recorded; if the later value is smaller, a minus is recorded. The test statistic is computed as the difference between the total number of pluses (increases in time) and the total number of minuses (decreases in time) in the record. A test statistic of zero would indicate no change over time (the null hypothesis). As deviations of the test statistic from zero become larger, the likelihood of trend in the data is greater, and rejection of the null hypothesis is more likely. For each water-quality record, a measure of the likelihood of trend (the probability of incorrectly rejecting the null hypothesis of no trend), the p-value, is obtained from a standard normal distribution. All censored values are assigned one half of their reporting limit for this method.

Adjustment for Seasonal Effects in Water-Quality Data

A source of variability in water quality that may prevent the detection of trends is seasonality. A typical seasonal pattern of variations in total phosphorus concentrations is shown in figure 6. The variations in the concentrations of some water-quality constituents may reflect changes in biological activities, sources of nutrients, or sources of sediment that occur because of seasonal changes in land use. Activities associated with agriculture are an example of seasonal changes in land use. The predominant source of water in a stream may vary seasonally and influence the concentration of some constituents. For example, snow melt or rainfall may dominate the source of flow in a stream during some months and ground-water seeps may dominate the source of flow in the same stream in other months. The water quality of these two sources could be very different from each other. Methods for the reduction of flow-related variability in water quality (as described in section 4.5) partially compensate for the volume of streamflow, but may not entirely compensate for changes solely caused by the source of the streamflow.

A parametric approach to removal of the effects of seasonality in water-quality data is to include a cyclical function as an explanatory variable in a regression equation that relates water-quality data to time (Schertz and Hirsch, 1985). Use of the Seasonal Kendall test does not require that the seasonal variation be explicitly modeled. The Seasonal Kendall test (Hirsch and others, 1982) reduces the effect of seasonal variation by restricting the possible comparisons to those water-quality values that are collected during the same seasonal period of each year. The Seasonal Kendall test statistic is calculated as a summation of the Mann-Kendall test statistics from each seasonal period (Hirsch and others, 1982). The total variance of the Seasonal Kendall test statistic is estimated as the sum of the seasonal estimates of variance. Statis-

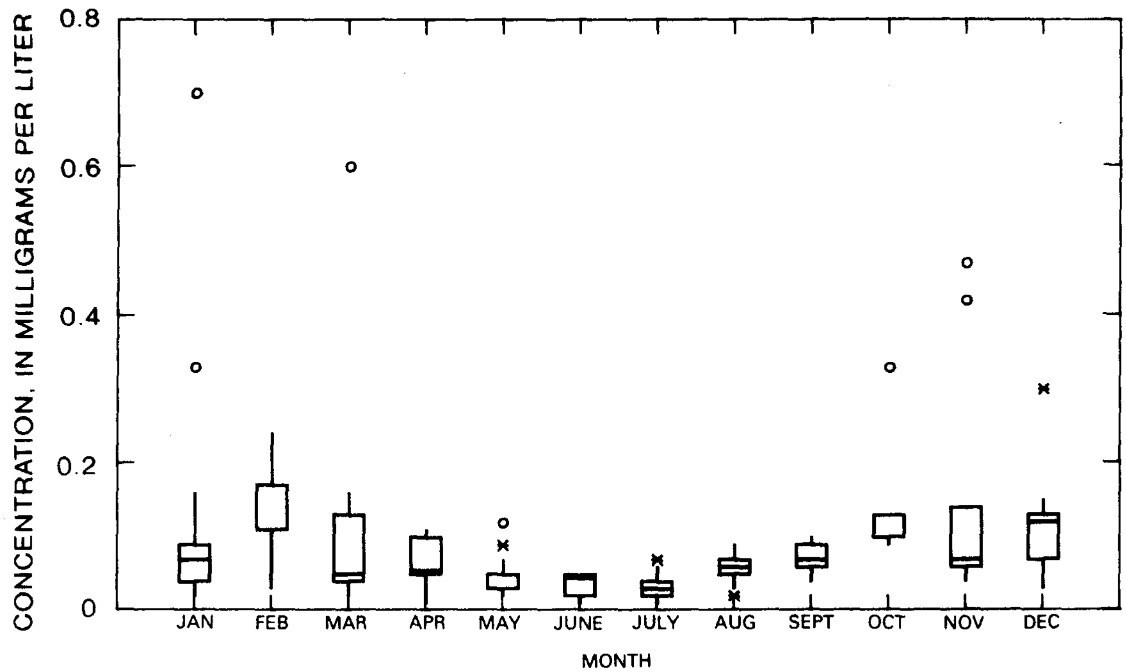

Figure 6.--Seasonal variations in total phosphorus for the KlamathRiver near Klamath, California.

tical significance, which is obtained from a standard normal distribution, is reported for the standardized Seasonal Kendall test statistic (Hirsch and others, 1982). These reported levels of statistical significance should closely approximate those from an exact distribution of the test statistic for sample sizes larger than 10.

Selection of Seasonal Water-Quality Values

For each season, a single value is selected for use in the Seasonal Kendall test. For seasons with multiple values, the most central value with respect to time that is also paired with discharge is selected to represent the season (see example in table 1). In contrast to the use of a mean or median to represent seasons with multiple values, this selection rule maintains a more constant variance in seasonal values for data records where the sampling frequency has changed over time. The maintenance of relatively constant variance during the testing period is desirable because more accurate statistical tests are likely under these conditions.

Season	Sample collection date	Date of selected sample
Fall/Winter (October 1-March 31)	October 31	December 11
	November 17	
	December 11	
	February 3	
Spring/Summer (April 1-September 30)	May 22	June 19
	June 19	
	September 2	

Selection of Seasonal Periods

The definition of seasonal periods for use in the Seasonal Kendall test is dependent upon knowledge of the seasonal variability in water quality in the study region. At some monitoring stations, the seasonal fluctuations in water quality may be best reflected by the established data-collection schedule (or sampling frequency), which, in effect, defines both the number of seasons per year and the length of each season. At other stations, the historical sampling frequency may only represent the best available or most convenient estimate of the seasonal cycle. As part of the ESTREND applications of the Seasonal Kendall test (Methods I and 2), the user may define both the number of seasons per year and the length of each season.

Water-quality data that are collected at a constant sampling frequency during a given time period will consist of a uniform number of values that can be easily compared in a trend test. Under these sampling conditions, the number of seasons used in the Seasonal Kendall test is set to the number of annual samples (e.g., 12 or monthly) that were regularly collected during the period of record. The selection of seasons becomes more complicated, however when the sampling frequency has changed during the period of record. In Texas, for example, the sampling frequency at stations included weekly, monthly, bimonthly, and quarterly sampling frequencies over the past 20 years because of changes in the network design and funding constraints (see fig. 7) (Schertz, 1990).

To compensate for variations in data density resulting from changes in sampling frequency, the number of annual seasons used in the Seasonal Kendall test should be selected by the user to reflect the year(s) with the smallest sampling frequency. Because no more than one value may be selected for each season in applying the Seasonal Kendall test, the choice of the smallest sampling frequency as the number of seasons gives uniform coverage of the whole study period without allowing any bias toward years with denser data collection. The number of seasons assigned to stations in the Texas trend investigation is shown in figure 8.

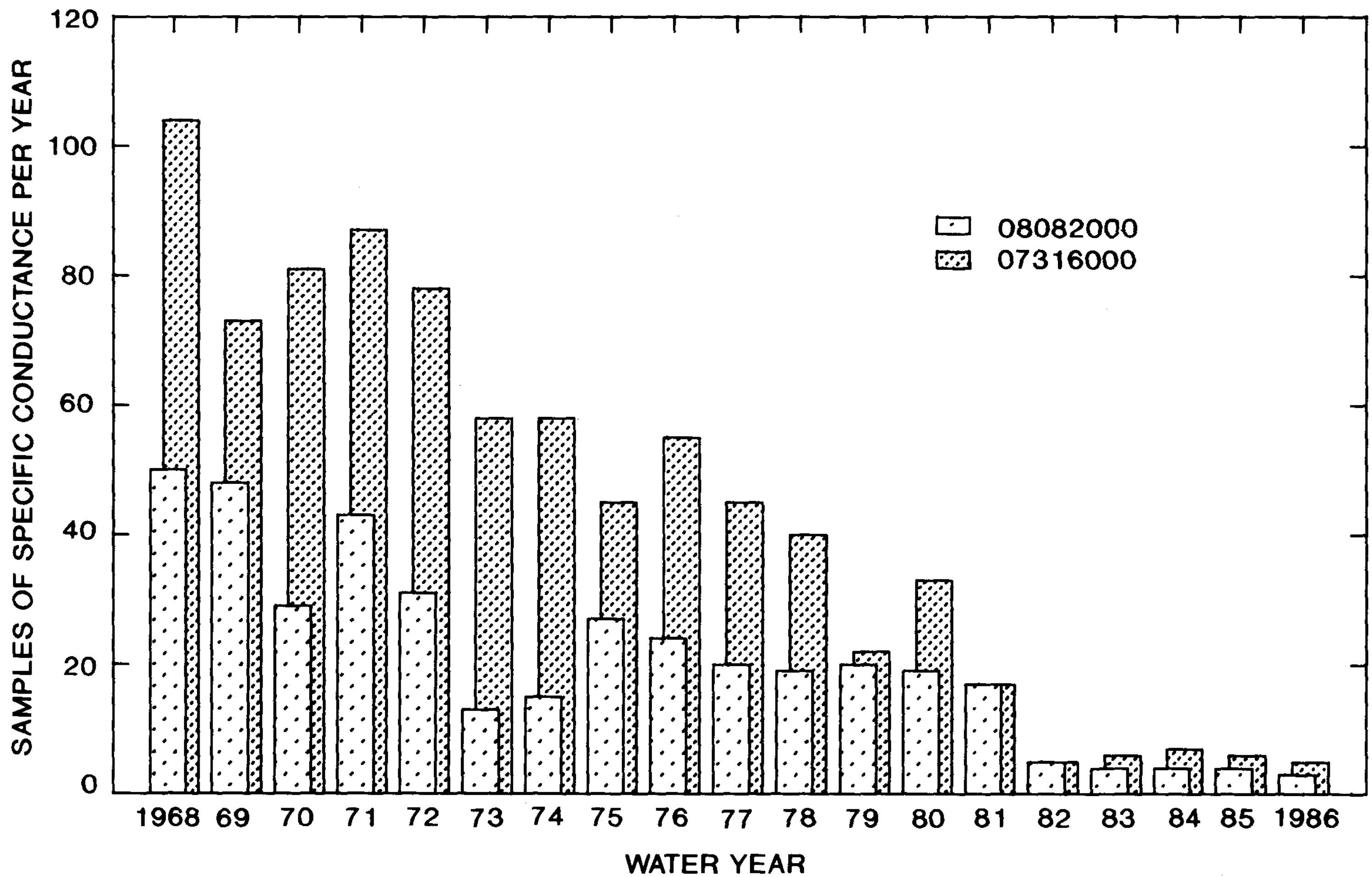

Figure 7.—Examples of sampling frequency for the 1968-86 water years for two Texas water-quality stations.

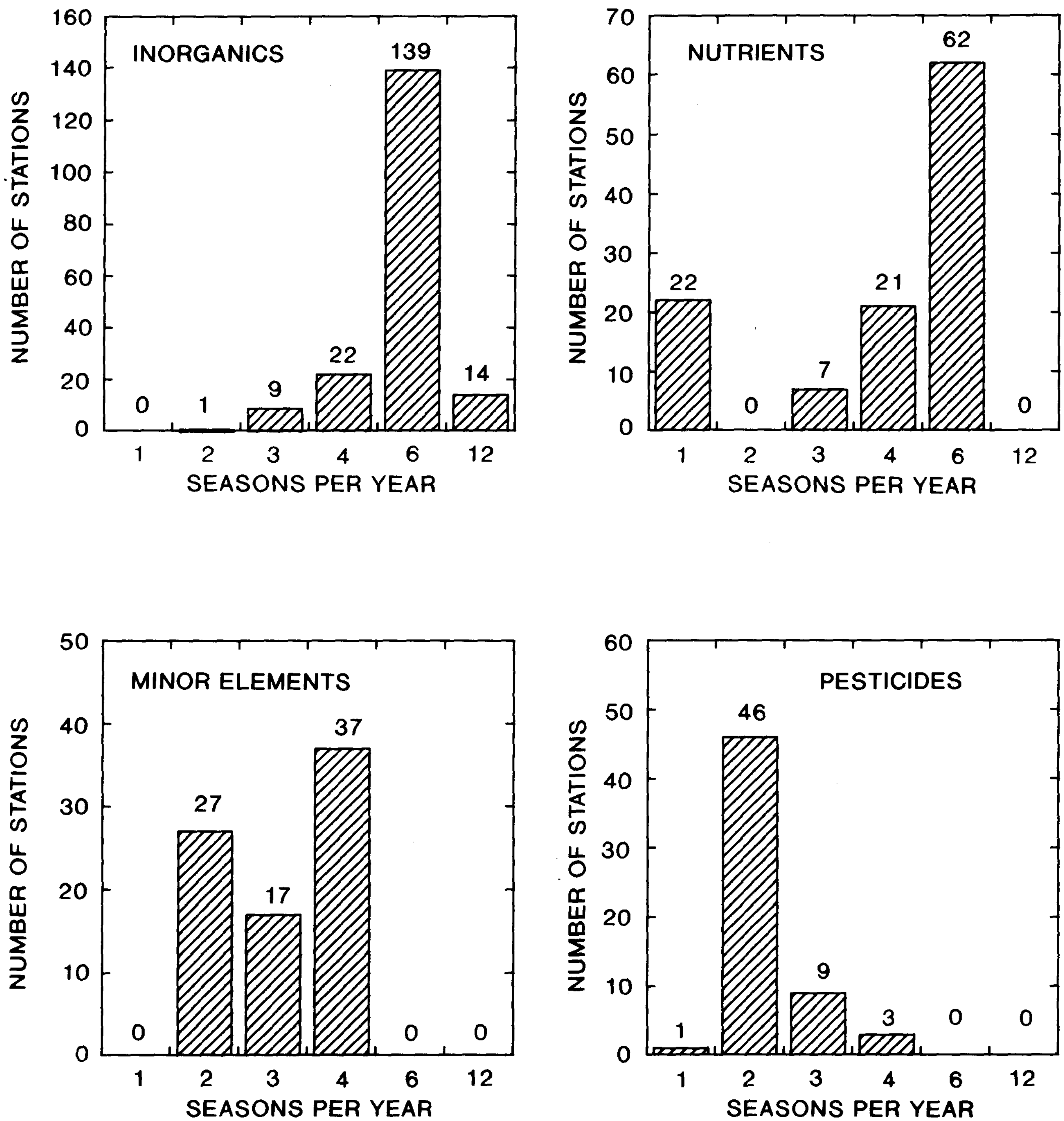

Figure 8.—Number of Texas water-quality stations in seasonal categories by constituent groups.

The number of seasons chosen for use in the Seasonal Kendall test is restricted to a maximum of 12 per year for two reasons. First, potential statistical problems exist when water-quality values used for trend testing are so close in time that the values are not independent of one another (i.e., the values are serially correlated). In such cases, the sample size is actually smaller than would appear because of redundancy in the data. This may result in significance levels, which are dependent upon sample size, that are inaccurate. Although the Seasonal Kendall test performs better than the parametric alternatives, it cannot be considered an exact test in the presence of serial correlation (Hirsch and Slack, 1984). No specific corrections were made for serial correlation in the trend tests used in the most recent studies (Schertz, 1990; Hay and Campbell, 1991; Smith and others, 1987), but the effects were minimized by limiting the data density to no more than 12 values per year. The second reason for restricting the maximum number of seasons to 12 is that differences in water-quality conditions rarely exist between seasons much shorter than a month. For example, the water-quality conditions for the first week in January probably would not be consistently different enough from the water-quality conditions for the second week in January to warrant separate comparisons for each year in the study.

An automated procedure is provided in ESTREND to assist the user in estimating how well the actual sampling frequency of a water-quality record supports user defined seasons. The procedure may be used to identify those seasons that are uniformly sampled throughout the period of record. For a given seasonal frequency, the procedure indicates whether the beginning and ending fifths of the record and, separately, the middle three-fifths, contain sufficient data such that a majority of the possible number of seasonal pairwise comparisons examined in the Seasonal Kendall test are available for most seasons. The beginning and ending portions of the record are given emphasis in this automated procedure to insure that the data adequately spans the period of interest. The procedure performs as follows:

1) The selected record is divided into beginning and ending portions with each consisting of about one-fifth of the entire record and the remaining middle portion consisting of about three-fifths of the record. The length in years of first and last fifths of the record is defined as the length in years of the selected period of record divided by five and rounded up to the nearest integer year. The difference between the total number of integer years in the record and the number of integer years in the beginning and ending fifths of the record is the number of integer years for the middle portion of the record.

2) Seasonal pairwise comparisons are made with user defined seasons between data in the beginning fifth of the record and the ending fifth of the record. For records with N years in the beginning and ending fifths, the maximum number of possible comparisons per season is N2.

3) Seasonal pairwise comparisons are also made with user defined seasons on data in the remaining middle portion of the record (integer years that constitute slightly less than three-fifths of the record). For records with N years in the middle portion, the maximum number of possible comparisons per season is N(N-1)/2.

4) For each possible seasonal designation (i.e., 12, 6, 4, 3, 2, or 1), the ratio of the observed number of seasonal, pairwise comparisons to the maximum number of possible seasonal comparisons is computed separately for each season (e.g, each of 12 months or each of 6 bimonths) for the beginning and ending portions of the record and for the middle portion of the record.

5) In general, the suggested seasonal designation for a record should have more than 50 percent of the maximum possible number of comparisons present in more than 50 percent of the seasons. For each season, table gives the suggested minimum number of seasons that should have at least 50 percent of the possible comparisons present. With less restrictive requirements, there is an increased risk that the data were actually collected according to a less frequent sampling schedule than selected by the user. In particular, the failure to choose the smallest observed sampling frequency in the beginning and ending portion of the record will increase the likelihood that the results of the trend test are temporally biased because certain seasons will have significantly less data in the tails of the record. In cases where the above rule is satisfied by several seasonal designations, the largest seasonal designation (e.g., 12 rather than 4) is a better estimate of the actual sampling frequency. The ratios resulting from comparisons of the beginning and ending portions of the record should generally receive more emphasis than comparisons of the middle portion of the record because gaps in the middle portion of the record have less effect on the performance of the statistical procedures.

Table 2.--*Suggested minimum number of seasons that should have at least 50 percent of the possible comparisons present*

Season	Minimum number of seasons with at least 50 percent of comparison present
Monthly	7
Bimonthly	5
Quarterly	3
Biannual	2

Comparison of Seasonal Trends

For each water-quality record, the degree of similarity in the seasonal trend tests is determined through the use of a statistical test proposed by van Belle and Hughes *(*1984*)*. In cases where the direction of the seasonal trends differs markedly (e.g., half of seasons show increasing trend and hall show decreasing trend), the overall Seasonal Kendall trend may be somewhat misleading because no significant trend would be found. Thus, a test for differences in seasonal trends is useful to identify such cases. The test statistic X^2_{homog} is defined as:

$$X^2_{homog} = X^2_{total} - X^2_{trend} = \sum_{i=1}^{m} Z_i^2 - m(z_{mean})^2 \quad (1)$$

where, $$Z_i = \frac{S_1}{(Var(S_i))^{1/2}}$$

$$Z_{mean} = \frac{1}{m} \sum_{i=1}^{m} Z_i$$

m = total number of seasons, and

S_i is the Mann-Kendall test statistic for the ith season as given in Hirsch and others (1982).

The null hypothesis of homogeneous (or similar) seasonal trend results is rejected if X^2_{homog} exceeds the alpha critical level for the chi-square distribution with n-1 degrees of freedom. Thus, a value of X^2_{homog} that is large and exceeds its critical value would suggest that differences exist among seasonal trends with respect to their direction and statistical significance.

Trend Slope Estimator

The rate of change over time (trend slope) is computed according to Sen (1968). The trend slope, expressed as change in original units per year, is the median slope of all pairwise comparisons (each pairwise difference is divided by the number of years separating the pair of observations). The trend slope is also expressed as a percent of the mean water-quality concentration by dividing the slope (in original units per year) by the mean and multiplying by 100. For water-quality constituents that are log transformed, the slope, expressed as change in original units per year, is computed as:

$$\text{Slope} = (e^b - 1)\ C \quad (2)$$

where b is the Seasonal Kendall slope estimate in log units, and C is mean concentration.

The rate of change in percent per year for log transformed constituents is computed as:

$$\text{Slope} = (e^b - 1)\ 100 \quad (3)$$

Both equations 2 and 3 provide an exponential rather than a linear estimate of the rate of change in the water-quality constituent.

For water-quality records with a large number of censored values, the estimate of the slope will be influenced by the value selected to represent nondetected observations (for example, zero or one half of the reporting limit). The sign of the slope is resistant to the presence of nondetected values in the water-quality record. However, the estimated magnitude of the slope should not be reported if more than about 5 percent of the water-quality record is censored.

Interpretation of Trend Slopes

The trend slopes represent the median rate of change in constituent concentrations or values for the selected period. They assist the user in comparing the magnitudes of trends that represent the same period for stations in a study. The trend slope is a measure of monotonic trend during the selected period, and serves as an approximation to actual temporal variations in the data. The actual data may change linearly, may change in steps, or many show reversals during portions of the selected period.

Any interpretation of trends in water quality should give careful consideration to the magnitude of trend slopes. This is especially true in investigations of constituent trends and their relation to water-quality standards. The magnitude of statistically significant trend slopes may not always be environmentally important when compared with water-quality standards. Thus, investigations that emphasize the environmental relevance of trend slopes should consider how both trend slope magnitudes and mean water-quality concentrations compare with water-quality standards or criteria.

Trends in Flow-Adjusted Concentration

The concentrations of water-quality constituents are often correlated with streamflow. The causes of the relation may vary from constituent to constituent. For example, the input of many constituents to a stream is from a point source that remains relatively constant. Increases in streamflow will result in decreases in concentration (dilution) for these constituents. Concentrations of other constituents that are contributed to a stream from overland flow will increase in concentration as streamflow increases (wash-off). Some constituents may exhibit combinations of both effects (Hirsch and others, 1982).

The detection of water-quality trends may be complicated by the presence of flow-related variability in water-quality records. Flow-related variability may be large relative to the magnitude of change in water quality resulting from human activity in the stream basin. Investigations of the effect of changes in human activities in stream basins on water-quality trends should focus on tests for trend in flow independent water-quality concentrations (or flow-adjusted concentrations). The reduction of flow-related variability in a water-quality record will reduce total variability in water quality and, thus, increase the power of the test or the chance of detecting a trend that is a result of some influence other than streamflow (see example in fig. 9). If trends in the actual ambient concentrations of water quality are of interest, however, for possible comparison with water-quality criteria or standards, trends in raw or non-flow adjusted concentrations should be examined.

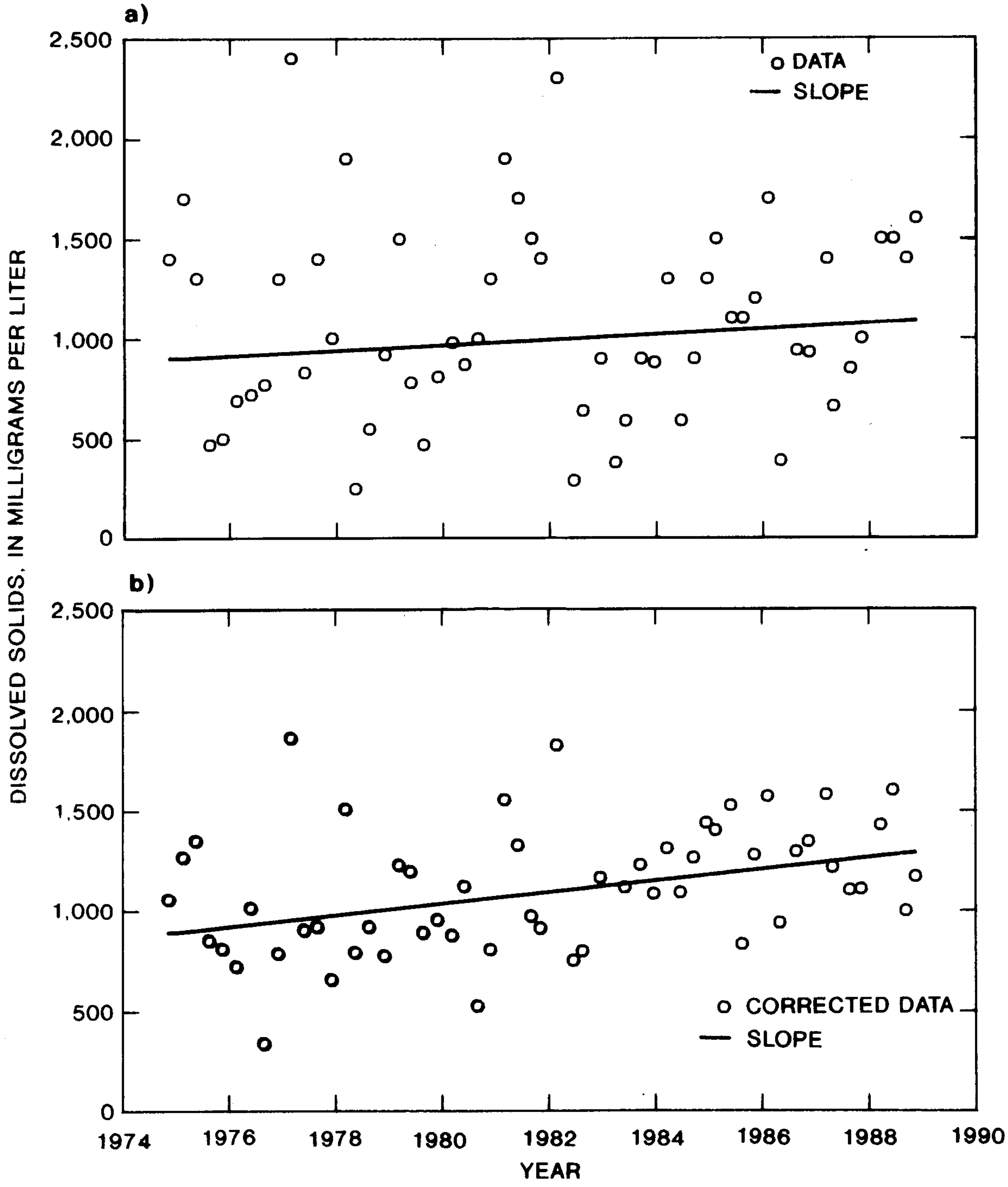

Figure 9.--Reduction in variability in dissolved solids for the James River near Scotland, South Dakota, resulting from a regression-based adjustment for discharge; (a) time series of unadjusted concentrations, slope = 13.8 milligrams per liter per year, p = 0.47; (b) time series of flow-adjusted concentrations (corrected by addition of the mean), slope = 29 milligrams per liter per year, p = 0.0001.

In conjunction with the Seasonal Kendall test for uncensored data, trend testing of water-quality records is performed according to a stagewise method (Hirsch and others, 1982; Alley, 1988). In this method, the reduction of flow variability in a water quality data record is achieved prior to trend testing by first describing the relation between concentration and flow either through a linear regression fit or nonlinear smooth such as LOWESS. Residuals are computed as the difference between the observed concentration and the concentration predicted by the mathematical relation. All available concentration flow data pairs are used in fitting these relations. The residual or flow independent values are then tested for trend with the Seasonal Kendall test. Because the Seasonal Kendall test is a univariate statistical test, an adjustment for flow variability cannot be performed simultaneously while testing for trend as in conventional regression procedures, but must be handled as a separate procedure. This method of flow adjustment should not be used if more than about 5 percent of the observations in a record are censored. Alternatively, the parametric trend procedure, Tobit, may be used to flow adjust highly censored water-quality records (see section 5.2).

One of two methods (consisting of a total of 14 possible models) may be used to obtain a mathematical description of the relation between concentration and flow. The first method regresses concentration on various functional forms of flow. The second method fits a smoothed LOWESS line to concentration and discharge or log transformations of these variables.

Regression Flow Adjustment

The regression method of flow adjustment may use one of two general models to describe the relation between concentration and flow (Smith and Alexander, 1983). The first, a simple linear model, is recommended for dissolved water-quality constituents that are typically not subject to adsorption on sediment particles. The model is of the form:

$$\hat{C} = \hat{b}_o + b_1 f(Q) \qquad (4)$$

where $\hat{b}0$ and $\hat{b}1$ are the parameters (intercept and slope, respectively) estimated in the regression procedure;

$\hat{C}$ = the estimated concentration,

Q = the instantaneous discharge, and

f(Q) has one of the following four functional forms:

f(Q) = Q (linear) [Model 1],

f(Q) = ln Q (log) [Model 2],

where In is the natural logarithm;

f(Q) = 1/(1 + BQ) (hyperbolic) [Models 3-10]

where B is one of eight possible coefficients scaled according to the observed range of discharge (see Smith and others, 1982); and

f(Q) = 1/Q (inverse) [Model 11].

The residuals (flow-adjusted concentrations) associated with the "best" model are selected for use in tests for trend. The "best" model may be specified by the user or selected according to an automated procedure (Model 0 is used to invoke the automated selection procedure). The automated procedure computes a PRESS (prediction sum of squares) statistic (Myers, 1986) for each of the 11 simple linear models and selects the regression model with the minimum value of PRESS. This model has the lowest prediction error among the models compared.

Use of the PRESS statistic as a model selection method is closely associated with model validation procedures that involve the splitting of data sets into calibration samples (for regression fitting) and validation samples (for evaluating the predictive power and fit of the regression). Computation of the PRESS statistic involves the execution of N (number of samples) validations in which the calibration data set is composed of N-1 observations. Each observation is withheld one at a time from a regression of N-1 observations. The difference between the actual value of the withheld observation and its associated predicted value from a regression of the other N-1 observations is the prediction error or residual for that observation. For each regression model, the PRESS statistic is computed as the sum of the squares of the prediction residuals for all observations.

The second general model (Model 12) used to describe the relation between concentration and discharge involves a log-log multiple regression of the form:

$$\ln \hat{C} = \hat{b}_0 + \hat{b}_1 \ln Q + \hat{b}_2 (\ln Q)^2 \quad (5)$$

where $\hat{C}$ is the estimated concentration,
Q is instantaneous discharge,
ln is the natural logarithm, and
$\hat{b}_0$, $\hat{b}_1$, and $\hat{b}_2$ are the parameters estimated in the regression procedure.

This model is often used for water-quality constituents that are typically non-conservative in their transport. Use of log transformations of all model terms and use of the log flow squared term in particular allows more precise fitting of concentration-flow relations that have a large degree of curvature in their tails. A previous investigation (Smith and Alexander, 1983) found this model form to be appropriate for the following constituents: suspended sediment, organic carbon, total phosphorus, most nitrogen species, fecal bacteria, turbidity, and phytoplankton.

For a given constituent, it is recommended that only one of the two general model forms be chosen. If the intent of the trend investigation is to compare results among stations, the selection of a single form of the model (log or unlogged concentration) eliminates the difficulty that would arise in making comparisons of slope estimates computed in log space with those computed in real space.

If feasible, the user is advised to visually inspect residual plots for the selected model to insure that a good fit is obtained. In particular, checks for homoscedasticity (constant variance) and approximate normality of residuals should be made to assure that regression assumptions have been met

Trend in flow-adjusted concentrations can be reported when the regression of concentration on flow is statistically significant, and the regression model has no significant deficiencies (i.e., residuals are generally homoscedastic and approximately normal). The user should only report trend in concentrations when the regression of concentration on flow shows only weak statistical significance. However, if there are no other significant deficiences in the regression model, then the removal of even small amounts of flow related variability in water-quality concentrations (as indicated by moderately significant regressions) may improve the chances of detecting trend. Thus relatively large alpha levels (0.10 or greater) may be acceptable in making decisions about whether flow adjustment is necessary.

LOWESS Flow Adjustment

Regression methods are sometimes difficult to use either because a good fit cannot be obtained (regression assumptions violated) or because the large number of possible regressions (transformations) that must be examined prevents a thorough evaluation of model fit. Therefore, a second, more robust method of flow adjustment is available for use. A LOWESS smooth (Cleveland, 1979) may be fit to one of the following concentration and discharge plots:

1) concentration and discharge data [Model 13]
2) log transformed concentration and log transformed discharge data [Model 14]

The log transformation of water quality made in the third case may be appropriate for many of the non-conservative water-quality constituents described above in the section on linear regression flow adjustment. As indicated above, the choice of a transformation of the data should not involve a mixture of raw concentration with log transformed concentration smooths because of the difficulty that arises in comparing logged and unlogged estimates of trend slopes.

The robust characteristics of the LOWESS procedure are due to the use of distance and residual weighting functions with weighted least squares which minimize the influence of outliers in fitting a smooth line to the data. The concentration-flow relation fit with LOWESS may be highly nonlinear because the method involves fitting N (number of observations) weighted regressions to the data (see fig. 10). For each observation in the data set, a predicted (smoothed) value is obtained from a re-iterated weighted regression. Smaller weights are assigned for observations with large residuals and at greater distance from the predicted observation. The number of observations used in the LOWESS regressions may be selected by the user by specifying a value for the parameter f. The value f is the fraction of the total observations to be used in the LOWESS regressions. Typically values between 0.3 and 0.7 (i.e., about 1/3 to 2/3 of the observations) provide a good fit to the data without removing important features of the relation or provide a good fit to the data without removing important features of the relation of producing very abrupt changes in local slope. Recent inspections of LOWESS fits for numerous water quality discharge plots for USGS stations (Lanfear and Alexander, 1990) suggested that an f value of about 0.5 gives a reasonably good fit to the data for many water-quality constituents.

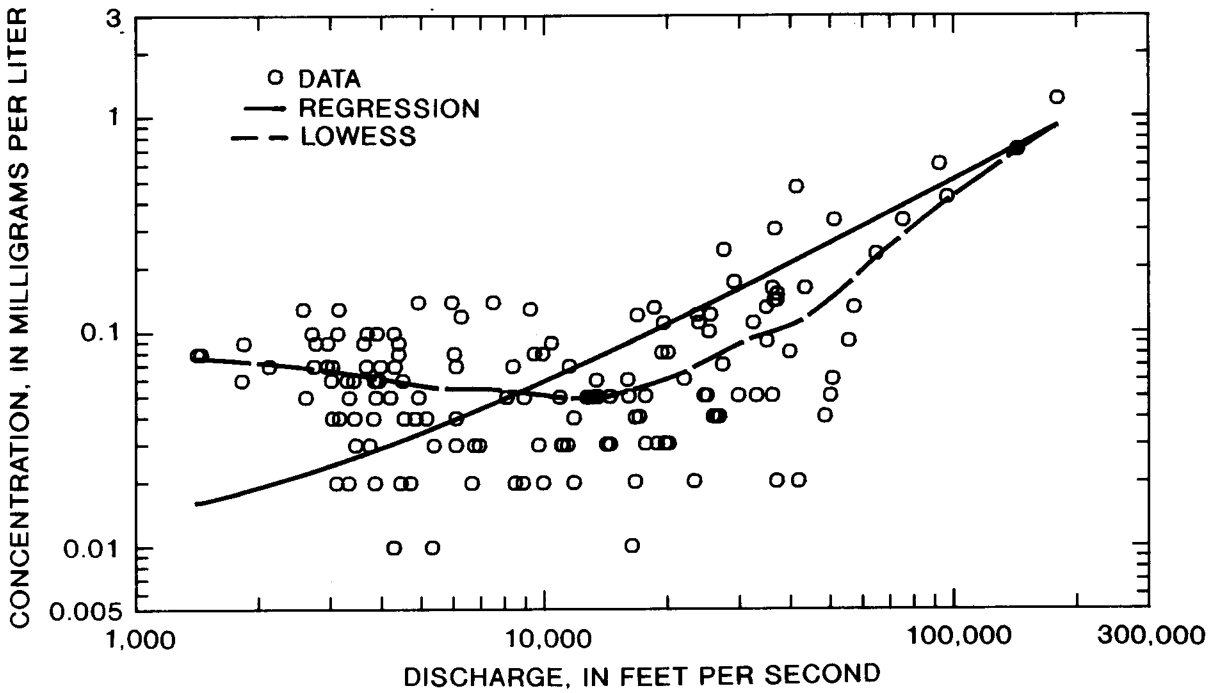

Figure 10.--A comparison of a LOWESS smooth and linear regression fit of total phosphorus and discharge for Klamath River near Klamath, California.

Interpretation of Flow-Adjusted Trends

The analysis of trend in flow-adjusted concentrations (residuals) assumes that the time series of flows is stationary (has undergone no change with time). For stationary flow, a trend in flow-adjusted concentrations is viewed as a change in the intercept of the flow-concentration relation but not in its slope. Thus, trend in flow-adjusted concentration is consistently in one direction at all levels of flow. Flows are nonstationary if there is actually a change in the slope of the flow-concentration relation. Under these circumstances, the interpretation of trend in flow-adjusted concentrations becomes difficult because the direction of trend may differ depending upon the magnitude of flow. If the user is aware of changes in flow resulting from changes in basin activities such as reservoir regulation or diversions, then an analysis of trends in flow-adjusted concentration should not be attempted. Trend analysis of flow observations may be used to determine whether the time series of flow is stationary or nonstationary if the history of the flow record is unknown.

Criteria for Trend Analysis of Uncensored Data

The period of record for each set of trend results is selected by the user. The criteria listed in table 3 are used to determine which data records have sufficient data for trend analysis for the selected period. The first criterion insures that each data record spans a minimum of 5 years. The second criterion insures that a minimum number of observations are in each data record. The third criterion insures that the data records represent the same period so that the trend results are comparable. The third criterion is based on a minimum percentage of total possible observations that must be present in the first fifth and last fifth of the selected period. The minima percentage, which is established by the user, has generally been set to 50 percent by previous trend studies.

Table 3.--*Criteria for use of the Seasonal Kendall test for uncensored data*

1)	The record must span a minimum of 5 years as determined by the difference in years between the beginning and ending observations.
2)	The minimum number of observations in the record must be at least three times the number of designated annual seasons, and must be greater than or equal to 10.
3)	A minimum percentage (as specified by the user) of the total possible number of seasonal water-quality values in the beginning and ending fifths of the record must be present in the record. The beginning and ending fifths of the record are determined as described in section 4.2.

Estimates of Moments and Percentiles

Estimates of the mean, variance, and percentiles (5th, 25th, 50th, 75th, and 95th) for uncensored data are computed using all water-quality values (no seasonal selection of the data is involved) for the selected time period. These computations are made for records with 10 or more observations. Rank ordered correlations (Kendall Tau as described in Kendall, 1975) between concentration and flow are reported using all concentration-flow data pairs. A test for statistical significance of these correlations is based on probability values from a standard normal distribution.

Seasonal Kendall Test

Water-quality records with more than about 5 percent of the observations censored at a single reporting limit may be analyzed with the Seasonal Kendall test. The reporting limit used in applying the test must be specified by the user. All values, detected and nondetected, that are less than the selected reporting limit are recoded to zero and considered tied (nondetected values are assigned a value of zero in the computer code to distinguish them from detected values that are rounded to the reporting limit).

Multiple reporting limits periodically occur in water-quality records when analytical methods are improved or constraints of funding require the use of a less accurate method (see example in fig. 11). The Seasonal Kendall test can also be used for data that has been censored at multiple reporting limits,

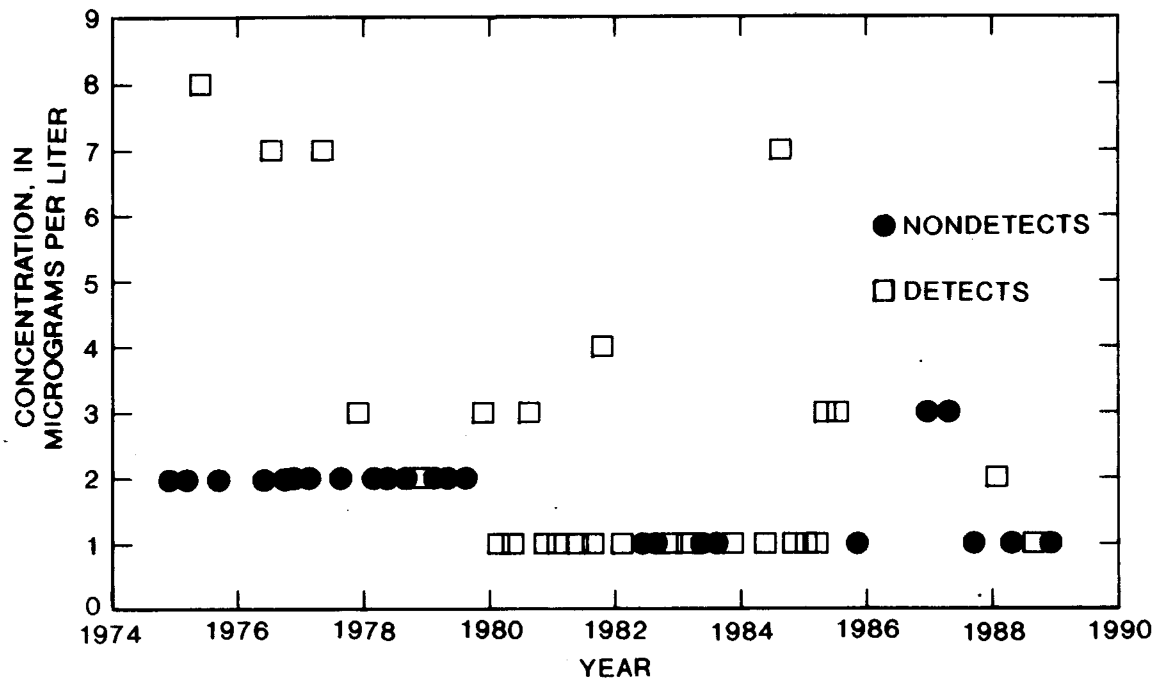

Figure 11.--The occurrence of multiple reporting limits in the record of dissolved cadmium for the North Fork Red River near Headrick, Oklahoma.

but the procedure is best used for multiply-censored records that contain relatively few detected values below the largest reporting limit. If a large number of detected values fall below the largest reporting limit, the recoding of data below this reporting limit may result in a significant loss of information. For example, the application of the largest reporting limit to the USGS water-quality data base in Texas resulted in a substantial loss of detected values for several constituents (table 4). In addition, such recoding of the data may restrict the evaluation of trend to a range of concentrations that are rarely observed. To overcome the possible loss of information the user may alternatively, but cautiously, choose to apply the test with a reporting limit that is less than the largest reporting limit in the record. ESTREND automatically discards any nondetected values that are censored at a reporting limit that is larger than the one selected by the user. A message is printed by ESTREND to notify the user of the number of values deleted. Although the user is permitted to select a reporting limit less than the largest reporting limit that occurs in the record, this should only be done if the number of values censored at the larger reporting limit are extremely few. The elimination of data is not recommended as it can bias the results.

The application of the Seasonal Kendall test to data records with a large number of censored observations (more than about 5 percent of the observations) prevents the use of the flow-adjustment procedure, and is likely to result in an inaccurate estimate of the trend slope. Flow adjustment of the water-quality data through ordinary regression or a LOWESS smooth cannot be made with highly censored data records because of the uncertainty in the censored water-quality values. Similarly, estimates of the Seasonal Kendall trend slope are not reliable because of uncertainty in the censored water-quality values. The slope values are calculated and stored but should not be reported for water-quality records with more than about 5 percent of the data censored.

Criteria for Trend Analysis of Censored Data with the Seasonal Kendall Test

The period of record for each set of trend results is selected by the user. The criteria listed in table 5 are used to determine which data records have sufficient data for trend analysis for the selected period. The first criterion insures that each data record spans a minimum of 5 years. The second criterion insures that a minimum number of observations are in each data record. The third criterion insures that the observations are representative of the selected period by requiring that a minimum number of observations be located in the beginning and ending fifths of the record.

Tobit Test

Water-quality records that are highly censored by one or more reporting limits may be tested for monotonic trend with a censored regression technique, Tobit (Cohen, 1976; Cohn, 1988). Conventional regression techniques for trend detection do not perform well with

Table 4.--*Occurrence of observations below the reporting limits of selected water-quality constituents for the 9169-86 water years for Texas surface water-quality stations*

[obs., observations]

Water-quality constituent (dissolved, mg/L)	Obs.	Reporting limit	Date Range	Obs. below the reporting limit	Percent of obs. below the reporting limit
Chromium	3,383	1	1981-86	409	12.1
		2	1975-77	17	0.5
		10	1981-86	424	12.5
		20	1972-79	241	7.1
Copper	3,437	1	1981-86	175	5.1
		2	1972-79	336	9.8
		10	1980-82	114	3.3
		20	1972-79	14	0.4
Lead	3,440	1	1981-86	405	11.8
		2	1973-83	213	6.2
		5	1982-86	56	1.6
		10	1980-82	81	2.4
Zinc	3,434	3	1978-86	261	7.6
		10	1983-86	14	0.4
		12	1982	16	0.5
		20	1972-79	536	15.6

Table 5.—*Criteria for use of the Seasonal Kendall test for censored data*

1) The record must span a minimum of five years as determined by the difference in years between the beginning and ending observations.
2) The minimum number of detected observations in the record must be at least three times the number of designated annual seasons, and must be greater than or equal to 10.
3) A minimum of one observation per year must be present in the beginning and ending fifths of the record.

censored data because estimates of regression parameters (intercept and slope) and the outcome of hypothesis tests can be influenced by the user's choice of a value (e.g., zero or the reporting limit) to represent nondetected observations. A maximum likelihood estimation (MLE) procedure (Cohn, 1988) in Tobit is used to estimate the parameters of a regression model relating concentration and time, which may be described as

$$\ln(\hat{C}) = \hat{b}_0 + \hat{b}_1 T \qquad (6)$$

where $\hat{C}$ is the estimated concentration,
T is time,
1n is the natural logarithm, and
$\hat{b}_0$ and $\hat{b}_1$, are the model parameters of intercept and slope, respectively.

The statistical significance of trend in water quality is evaluated in a likelihood ratio test (Ho: $\hat{b}_1$=0) on the MLE estimate of the slope (Hosmer and Lemeshow, 1989). The p-value (probability of incorrectly rejecting the null hypothesis) for the likelihood ratio test is reported as part of the ESTREND application of the MLE procedure.

Because the MLE estimation procedure assumes a linear model with normally distributed residuals, the water-quality variable is log transformed in equation 6 to improve the linearity of the data and make the data more nearly normal in its distribution. Failure of the data to conform to these assumptions will tend to lower the statistical power of the test, thereby reducing the chance that the test will detect a trend that actually occurred (Hirsch, et al., 1982). Unreliable estimates of the model parameters may also result from violations of the normality assumption (Montgomery and Peck, 1982). However, the Type I error of the test (probability of incorrectly rejecting the null hypothesis of no trend) as given by the p value is relatively insensitive to violations of the normality assumption.

Tobit Trend Slope Estimate

The trend slope, expressed as change in original units per year, is computed according to equation 2, and the trend slope expressed in percent per year is computed according to equation 3. Both computations give an exponential rather than a linear estimate of the rate of change in the water-quality constituent.

Estimates of Reporting Limits in Tobit

To use the TOBIT procedure, the reporting limits for all detected water-quality observations must be determined. The reporting limits for values in the water-quality record are selected in one of the following ways.

1) The closest nondetected value in time is chosen.
2) If no nondetected value is available for a given record, then the highest reporting limit for a constituent that is less than the smallest detected value in the record is chosen.

3) The determination of the reporting limit for zero values (nondetected values with unrecorded reporting limit) is similarly done as described above. If all nondetected values in the record of a station are zero values, then the lowest reporting limit for the constituent is chosen.

Criteria for Trend Analysis of Uncensored Data with the Tobit Test

The period of record for each set of trend results is selected by the user. The criteria listed in table 6 are used to determine which data records have sufficient data for trend analysis for the selected period. The first criterion insures that each data record spans a minimum of 5 years. The second criterion insures that a minimum number of detected observations are in each data record. The third criterion insures that a minimum percentage of the observations are detected. The minimum percentage is set by the user (see section 7.3) and a minimum of 20 percent is suggested. The fourth criterion insures that the observations are representative of the selected period by requiring that a minimum number of observations be located in the beginning and ending fifths of the record.

Table 6.—*Criteria for use of the Tobit test for censored data*

1)	The record must span a minimum of five years as determined by the difference in years between the beginning and ending observations.
2)	A minimum of 10 detected observations must be present in the data record.
3	A minimum percentage (as specified by the user) of the total number of observations in the record must be detected observations.
4)	A minimum of one observation per year must be present in the beginning and ending fifths of the record.

Estimates of Moments and Percentiles

Estimates of the mean and variance of water-quality records classified as censored are obtained by log regression procedures described by Helsel and Cohn (1988). A regression of log concentration values on normal quantiles is used to predict the magnitude of censored values in the record. The mean, variance, and percentiles (5th, 25th, 50th, 75th, and 95th) are then estimated from the record with predicted values in place of censored values. In general, the log regression method provides the most accurate (lowest root mean square error) estimates of the mean and variance among several possible statistical techniques (Helsel and Cohn, 1988). Moreover, the log regression method tends to be more robust than other techniques even when the statistical distribution of the data is unknown.

Alternatively, the actual percentiles of the water-quality data present in a station record can be also reported in the ESTREND procedures. This is achieved by first performing a separate ranking of detected and censored values. A combined ranking is then determined by assigning the lowest ranks to the group of censored values and the highest ranks to the group of detected values. The percentiles actually observed in the water-quality records are computed from these combined values.

Estimates of moments and percentiles are made for water-quality records with more than a specified percentage of the observations detected. The required percentage must be specified by the user.

DATA SELECTION AND MANAGEMENT

The ESTREND procedures allow for the screening of data so that statistical outliers (values greater than 5 standard deviations from the mean; less than a 1/1000 chance of being observed) are eliminated. An acceptable range on observed data may be established for each water-quality constituent to eliminate data with obvious coding errors. Water-quality values from the U.S. Geological Survey's National Water Information System (NWIS) with the remarks '2', '3', '4', '>', 'M', or 'N' are typically not interpretable in statistical trend procedures, and are automatically deleted from the records. Commonly, only a very small fraction of observed water-quality records contain such remarks. However, a water-quality record that has a large number of values with these remarks should not be analyzed for trend because the deletion of such values may introduce significant bias in the analysis. Unremarked values, estimated values (remark codes of 'E', 'O', 'K', 'B'), "less-than" values (remark codes of 'l', 'U', or '<'), and zero values are all used in the statistical procedures.

In the ESTREND package, complete records for any specified water-quality constituent(s) are first retrieved from NWIS water-quality files. The records for each water-quality constituent or for instantaneous flow are then selected, screened, and finally stored as a separate and permanent datafile on the computer that may be accessed for trend analysis. The maximum number of water-quality values to be selected and stored in these datafiles for each month may be specified. Each month is divided into a number of intervals that corresponds to the specified maximum number of water-quality values per month. Within each interval, the concentration value closest in time to the middle of the interval and having instantaneous flow (or daily mean flow if instantaneous flow is unavailable) is selected for storage. If neither instantaneous flow nor daily mean flow is available, then the concentration value closest in time to the middle of the interval is selected for storage.

ESTREND SYSTEM

Overview of Programs

The programs in the EStimate TREND (ESTREND) software system are designed primarily for multiple-station investigations of trends in water-quality constituents. The data manipulation tools and the statistical and graphical procedures provide the ability to compare trend results and summary statistics among stations of interest. Although the intended use of ESTREND is for large regional investigations, the facilities to examine and manipulate the analysis of individual station records are also provided. The instructions in this section correspond to revision 91.0 of ESTREND.

Capabilities

The ESTREND software package is a complete system which allows the user many options in examining, manipulating, and analyzing the data. Capabilities include:

1) defining the specific month and day to begin and end each season,
2) determining the most appropriate number of seasons per year to apply to the data for each constituent for each station,
3) creating files that contain the number of seasons per year and the flow-adjustment model to apply to the data for each constituent for each station,
4) creating data files,
5) running trend procedures on any selected period within the retrieved data base,
6) producing a variety of data distribution plots and X-Y plots of data and results for a selected station,
7) producing tables of summary statistics and trend results, and
8) producing maps of trend results.

Processes

The data used by the ESTREND package is initially stored in files created by using the National Water Information System (NWIS) software installed on the U.S. Geological Survey Water Resources Division's PRIME minicomputers. The data is read from these files and stored in unformatted files by constituent. A separate file containing a summary of the reporting limits encountered for each constituent is created concurrently. These two files form the database for each constituent and are used by the remaining programs in the ESTREND software system.

A program assists the user in creating a file that defines the seasonality of the water-quality data for use in the Seasonal Kendall test. The completed file contains date ranges for each of twelve, six, four, three and two seasons per year. The period of a year covered by each season is completely up to the user and need not be equal divisions of a year.

The software package has two programs which assist the user in determining which seasonal definition to use for each station and constituent. The first program uses the constituent data and the seasonal definitions to produce a table which lists the number of valid seasonal comparisons versus the total number of possible comparisons for each seasonal definition for all stations (see section 4.2 for a description of these comparisons). This program also produces an unformatted file for use by the second program to create a table containing the selected seasonal definition for a station based upon user-defined criteria. The selected seasonal definitions are output in a table.

Files that contain the selected seasonality and flow-adjustment model type for each station are created by a second program in the software package. These two files must be created for every constituent before the trend programs can be run.

The program which computes the summary statistics and the trend estimates uses the datafile, reporting limit file, the seasonal selection file and the flow-adjustment model file as input. The results are stored in a file that is used as the input file to the programs in the ESTREND package which produce the tables, plots and graphs of the data and the results. The process is outlined in figure 12.

Data Files

The data, reporting limits, and results of the statistical summary and trend estimates are stored in unformatted files. This type of file can not be listed at a terminal or a printer. There is one of each of these files in the separate subdirectories for each constituent. Each datafile contains the data for all stations for the entire period of record for a single constituent. The reporting limit file contains the reporting limits encountered in the data and an assigned code for each of these values. The file containing the results of the tests holds the statistical summary for each station, the trend test results for each station for the concentrations and the flow-adjusted concentrations and the flow-adjustment parameters for each model.

Support Files

There are four information files and one insert file that all programs in the ESTREND package access. The information files contain station header descriptions, season definitions, pathnames to datafiles and support files and pertinent constituent information. The insert file provides specific parameters from the study to the ESTREND programs. With the exception of the season definition file, these files are created by the user before running any of the programs in the package. There is a program that assists the user in creating the season definition file.

Organization

The programs and files which make up the ESTREND software package are contained in a series of directories and subdirectories. The main, or root directory contains 2 files and 12 subdirectories. The subdirectories and descriptions of their contents are listed in table 7. The files contained in each subdirectory are listed in Appendix I. A separate directory contains the data files, support files and the constituent subdirectories. The data directory is a subdirectory of the root directory (fig. 13).

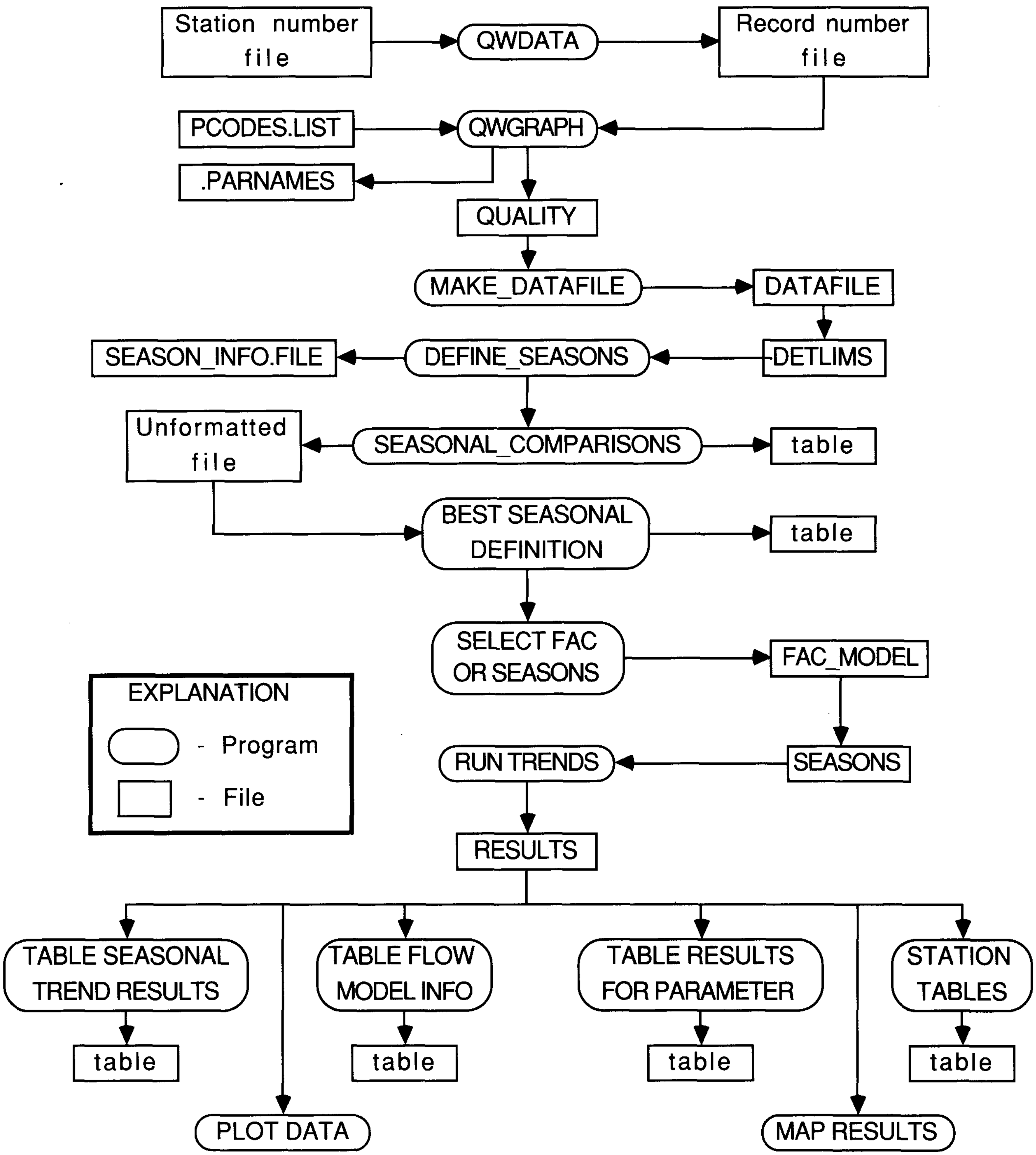

Figure12.—Procedures of the ESTREND system.

Table 7.—*Description of subdirectories used in the* ESTREND *software system*

Subdirectory name	Contents
SOURCE	FORTRAN source code for the main programs in the ESTREND system
SOURCE>INSERT	Insert files used for constant and common block definitions
SOURCE>LIB	FORTRAN source code for the ESTREND subroutine library
COMPILE_CPLS	Command Procedure Language (CPL) programs which compile and link source code into executable files
RUN	Executable program files
DOC	Documentation files for the ESTREND system
MAKE_DATA	Command output (como) files which capture data sent to the terminal during the preparation of the constituent datafiles for the ESTREND system
TABLES	Output tables from the ESTREND system
PLOTS	Plot output files
MAPS	Map output files
SEASON	Output tables which help the user determine the seasonality of the data
DATA	Information files and constituent subdirectories

Retrieval of Water-Quality Data

Water-quality data for the selected stations and constituents are retrieved from the water-quality database on the PRIME. This involves several steps and two programs included in the National Water Information System (NWIS) software package. These programs that retrieve the data from the distributed database are initiated through the QWDATA and QWGRAPH option menus.

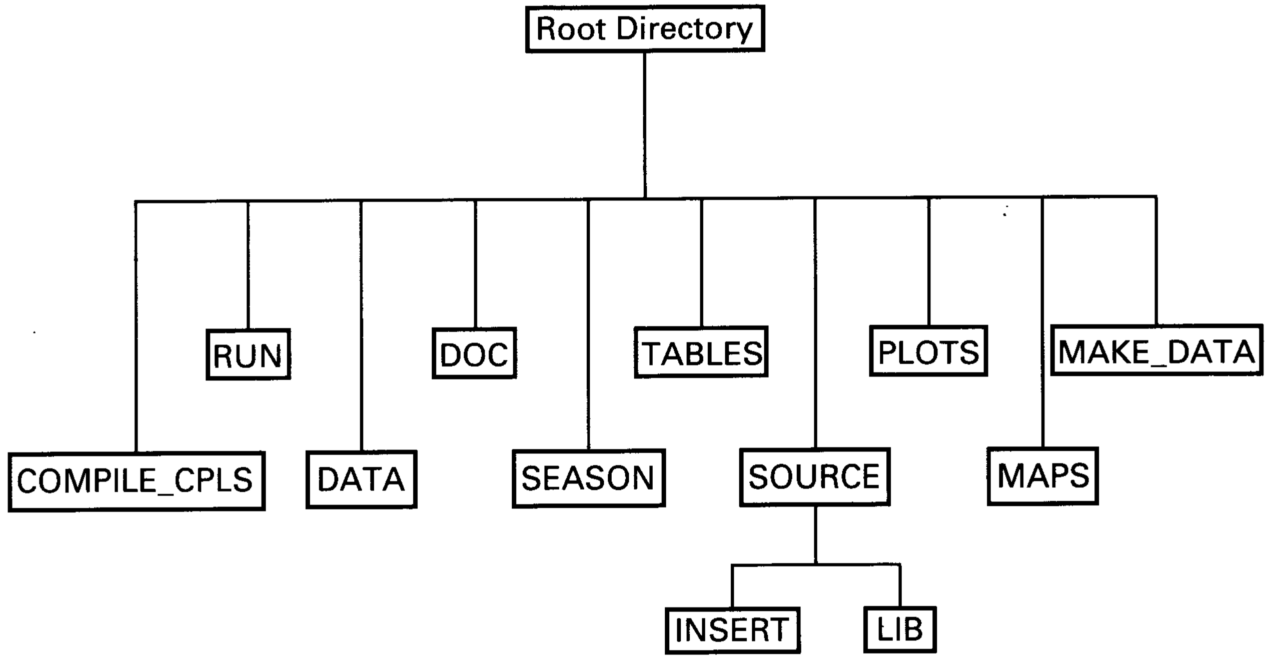

Figure 13.--Directory structure of the ESTREND system.

There is one input file which must be created prior to using the QWDATA program. This file contains records consisting of station numbers of the selected stations. The format of the record is:

```
                 1         2
Column:  12345679012345678 90
         USGS XXXXXXX
```

with the agency code in columns 1-4 and the station number starting in column 6. The file should be sorted by the station number in ascending order and reside in the data directory.

The following steps should be executed to retrieve the data using the QWDATA program.

1. Attach to the data directory, and invoke the data retrieval program by entering QWDATA at the operating system level. Select OPTION 3, RETRIEVE SAMPLES, from the QWDATA menu, then select Option 3, LOCATE SITES AND QW SAMPLES.

2. The program will ask if the user wants to locate sites. Answer "NO" to this question because the site number file has already been created.

3. The program will ask if the user wants to locate samples for specific sites. Answer "YES" to this query.

4. The program will ask if the stations will be entered from the terminal. Answer "NO". The program will then prompt for the name of the file containing the specific sites. Enter the name of the file containing the site numbers.

5. A menu to specify how to select the QW records will be displayed. Select a search by dates by entering an "X" in the "DATE" field of the menu and then quit by entering a "Q" in the next field.

6. The program will prompt for the earliest and latest dates to search for. After these are entered in the requested format, the program may pause for a long time while displaying the message: "SEARCHING FOR QW RECORDS...". This period of time will vary according to the period of record and number of stations requested.

7. Once the program has completed it's search, it will ask if the records are to be sorted. Answer "YES" to this query. A menu with the possible sort keys will be displayed. Choose as the primary sort key the field "STATION ID" (selection B) and as the secondary sort key the field "DATES AND TIMES" (selection C). NOTE: It is very important that the records be sorted in this order. Not doing so will generate an error in subsequent programs.

8. The program will prompt for a filename to hold the record numbers. After this is entered) the program will then ask if a list of sites with water-quality data should be saved. This is not necessary but may be requested if the user would find it useful.

9. The program will return to the OPTION menu. Select the options which will return to the operating system. The first step of the data retrieval is then complete.

Option 3 of the QWDATA program does not retrieve the actual water-quality data. The program creates a file which contains a list of the record numbers which correspond to the records in the water-quality database which do contain the requested data. The file of record numbers is used as input to the next step of the retrieval process.

A program in the QWGRAPH software menu, Option 7--Flat File Output, produces a file containing water-quality data which is used by the MAKE_DATAFILE program in the ESTREND package to create the unformatted data files used by ESTREND. The files are named QUALITY.DATA.#, where # is a sequential number used to identify the data in each file. These files hold the water-quality data for the selected stations and constituents. Before this option can be run a list of parameter codes must be compiled into one or more PCODES.LIST.# files, depending on the number of constituents selected for the project. These files contain the 5-digit numeric parameter codes and the 5-character alpha parameter codes for the selected water-quality constituent. The first four parameter codes in each PCODES.LIST.# file must be STAID, DATES, 00060 (mean daily discharge), and 00061 (instantaneous discharge), in that order. There may be as many as 18 subsequent water-quality parameter codes in the PCODES.LIST.# file, 1 per line. Because the QUALITY.DATA.# files have no headings or means of identifying the columns of data, the only way that the data for a specific constituent can be identified is by the position of the data in the QUALITY.DATA.# files. A file is created concurrently with the QUALITY.DATA.# file which contains a listing of the constituent names associated with each column in the QUALITY.DATA.# file. The name of this file is formed by adding ".PARNAMES" to the end of the QUALITY.DATA.# file name. The constituents are listed in the same order that they are encountered in the QUALITY.DATA.# file.

1. To invoke the program, enter QWGRAPH at the operating system level and select Option 7, Flat File Output. The program will then ask for the name of the file of record numbers that was created by the LOCATE SAMPLES option of QWDATA and will prompt for an output file name. Use a different output name for each PCODES.LIST.# file (QUALITY.DATA.1, QUALITY.DATA.2, etc.).

2. The program will ask if the parameters are to be entered from the terminal. Answer "NO", then the program will prompt for the name of the file containing the parameter codes. Enter the PCODES.LIST.# filename.

3. The program will display a message that the data is being processed. The program may take a long time to run, depending on the number of constituents and the period of record.

4. Once the program has finished processing the data, it will ask if another run is wanted. Answer "NO" to this. The program produces a file containing the parameter codes, constituent names and constituent units. The name of this output file is formed by adding ".PARNAMES" to the name of the QUALITY.DATA.# file. The information in this file is used to prepare support files for the ESTREND package.

5. Repeat steps 2 through 4 for all PCODES.LIST.# files.

For further information of the QWGRAPH and QWDATA systems and related programs, consult the National Water Information System 90.1 User's Manual (Maddy and others, 1989).

Preparation of Support Files

Four information files and one insert file that contain information used in various phases of the programs in the ESTREND package must be created before any of the programs can be run. The information files, PATHNAMES.FILE, HEADER.FILE, PARAM_INFO.FILE, and SEASON_INFO.FILE, are all stored in the data directory. The insert file, CONSTANTS.INS, is stored in a directory under the source directory.

PATHNAMES.FILE

Programs in the ESTREND package use a 'short name' to point to the data for each water-quality constituent. The short name is an arbitrary name assigned to each constituent by the user. The name may get up to 8 characters in length and may contain numbers, but may not begin with numbers (e.g., TEMP for water temperature; S04 for sulfate). The short name is used as the name of the directory where the constituent data is stored and as user input to most of the programs in the ESTREND package.

The PATHNAMES.FILE provides the software with the location of the constituent data directories and the information files. The format of the records in PATHNAMES.FILE is:

Columns 1-8: Short name of key word (left justified)
Columns 10-74: Pathname (left justified)

There is one record per constituent in PATHNAMES.FILE, each of which use the short name to point to the constituent directory. There are four mandatory records in PATH-NAMES.FILE that use key words to point to three information files and the data files. The key words are shown as the first four records in the following example of PATHNAMES.FILE.

Column:
```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890
QUALITY data_dir>QUALITY.DATA.
P_INFO  data_dir>PARAM INFO.FILE
HEADER  data_dir>HEADER.FILE
SEASON  SEASONdata_dir>SEASON INFO.FILE
TEMP    TEMPdata_dir>TEMP>
S04     S04 data_dir>S04>
```

where 'data_dir' is the path to the DATA directory. The QUALITY record must have a '.' at the end so that the number of the QUALITY.DATA.# file can be appended. The P_INFO, HEADER, and SEASON records must end in the filename. All records that point to constituent directories must end in '>' as shown for TEMP and $S0_4$. The maximum number of records in PATHNAMES.FILE is 75.

HEADER.FILE

This file contains information about the individual stations. The records consist of the station number, station name, hydrologic-unit code, drainage area, latitude and longitude, and the three-digit county code. The latitude and longitude are reported as both decimal and degrees, minutes, seconds (DMS). For example, longitude 97 degrees, 30 minutes west would be reported as 97.500 in decimal and 903000 in DMS. Degrees west of the prime meridian are positive and degrees north of the equator are positive. The station name and location in latitude and longitude are used by the tabling and mapping programs. Use the alpha codes STAID, SNAME, and LATLG and the flat file option in the NWIS QWGRAPH system to obtain this information. The format of the records in HEADER.FILE is as follows:

Columns 2-9 :	Station number (left justified)
Columns 18-64 :	Station name (left justified)
Columns 66-73 :	Hydrologic unit code
Columns 75-84 :	Drainage area (two decimal places)
Columns 87-92 :	Latitude in DMS
Columns 95-101 :	Longitude in DMS
Columns 104-106:	County code
Columns 109-114:	Latitude (three decimal places)
Columns 118-123:	Longitude (three decimal places)

This file must be sorted in ascending order on the station number.

PARAM_INFO.FILE

This file contains information about the water-quality constituents selected for the study. The records consist of the constituent short name, parameter code (including leading zeros), position of the constituent in the QUALITY.DATA.# file, number (#) of the QUALITY.DATA.# file, long name, minimum allowable value and maximum allowable value. The parameter code and long name for the constituent can be obtained from the PARNAMES files described in section 7.2. The long name is used by tabling and graphics programs as labels and headings and may be modified to contain any information about the constituent that the user would need to see on output. The position of a constituent in a QUALITY.DATA.# file and the number (#) of the QUALITY.DATA.# file are used by the software to find the data for that constituent. The position is found from the sequential order of the constituent in the PARNAMES file where mean daily discharge (00060) is in position 1 and instantaneous discharge (00061) is in position 2. The format of the records in PARAM_INFO.FILE is as follows:

Columns 1-8:	Constituent short name
Columns 10-14:	Parameter code
Columns 16-17:	Position of the constituent in the QUALITY.DATA.# file
Column 18 :	Number (#) of the QUALITY.DATA.# file that contains the data for the constituent
Columns 19-58:	Forty character constituent long name
Columns 68-76:	Minimum allowable value, in scientific notation
Columns 78-86:	Maximum allowable value, in scientific notation

Use the value 1.000E+30 for the maximum if there is not a **specific maximum** to use. The number of records in PARAM_INFO.FILE may not exceed 71.

SEASON_INFO.FILE

This file contains date ranges which define the season to use in the study. The ranges selected by the user for 12, 6, 4, 3, and 2 seasons are stored in this file for use by the programs in the ESTREND package. The user is assisted with the creation of this file by a program in the ESTREND package. The file and its contents are discussed in more detail in the section covering the execution of this program (DEFINE_SEASONS).

CONSTANTS.INS

Programs in the ESTREND package use values to dimension data structures and define constants for certain equations or conditions that are unique to a study. The values are set in the file CONSTANTS.INS. A copy of this file is provided with the ESTREND package as a template for the user to modify. The file contains the following constants:

NSTAS	-	Number of stations in the study
NOBS	-	Maximum number of observations per station to be stored in the constituent subdirectory DATAFILE
BYEAR	-	Beginning year of record (e.g., 70 for 1970)
NFLOW	-	Parameter short name used for instantaneous flow (must be in uppercase)
PERC	-	A user defined criterion specifying the minimum percent of observations required for trend testing (uncensored test only). Defined as the percentage of the expected number of observations required in the beginning and ending fifths of the record necessary to apply the Seasonal Kendall trend test. (40 percent is recommended).
MAXOBS	-	Maximum number of observations accepted in any month. The month is divided into this number of equal time intervals. In selecting data for storage in DATAFILE, the concentration and flow pair closest to the middle of each interval is chosen. If flow is unavailable, the concentration value closest to the middle of each interval is chosen. This value may be used to reduce the number of values stored in DATAFILE for frequently sampled stations.
MAXSEAS	-	The maximum number of seasons specified in the file SEASON for any uncensored constituent
MAXDL	-	The maximum number of discrete detection limits per parameter
PATHS	-	The full pathname to the file PATHNAMES.FILE
CPERC	-	The minimum percent of detected values (expressed as a percent of the total number of observations in the time period analyzed) required to execute censored statistical summary and regression procedures

The file contains a number of comment lines which describe the constant values. These comment lines begin with a 'C' in the first column. Following the comment lines, the values are declared to be INTEGER, REAL or CHARACTER values. The lines that the user will modify begin with 'PARAMETER'. The format of the PARAMETER statement is

Column:

```
123456789
      PARAMETER(variablel=valuel,variable2=value2,...)
```

where 'variable' is the name of the value in ESTREND and 'value' is the value assigned to that variable. Notice that the values for the variables NFLOW and PATHS are enclosed in single quotes ('). These are character variables and must have character values. The variables which are declared as INTEGER must have whole number values and those declared as REAL must have values which include a decimal point.

Execution

The ESTREND software package contains 12 main programs to assist in preparing the data for trend analysis, running trend analysis and reporting the results. The software package is menu driven, meaning that a program can be invoked by selecting a number from a list of the different programs.

To bring up the ESTREND menu, attach to the root directory and enter the command CPL TREND. This will clear the screen and present the following menu:

ENTER SELECTION:

TRENDS PROGRAMS MENU

ENTER THE NUMBER OF THE PROGRAM TO RUN

1 -- MAKE_DATAFILES
2 -- DEFINE SEASONS
3 -- RUN SEASONAL COMPARISONS
4 -- SELECT BEST SEASONAL BREAKDOWN
5 -- SELECT FLOW MODELS OR SEASONS
6 -- RUN TRENDS
7 -- TABLE RESULTS FOR A CONSTITUENT
8 -- MAP RESULTS FOR A CONSTITUENT
9 -- PLOT RESULTS
10 -- TABLE RESULTS FOR ALL STATIONS
11 --TABLE FLOW MODEL INFORMATION
12 -- TABLE SEASONAL TREND RESULTS

99 -- EXIT TO PRIMOS

ENTER SELECTION:

Once a selected program has completed execution, the following message will appear on the screen:

ENTER NEW SELECTION OR RETURN FOR MENU
ENTER SELECTION:

To run another program, just enter the number of that program or, to bring the menu back to the screen, hit a carriage return. Enter '99' to quit the ESTREND package and return to the operating system.

SELECTION 1 -- MAKE DATAFILE

Purpose: Produces binary datafiles and reporting limit files for each constituent. These files are used by the trend estimation program as well as several of the support programs.

Input files: QUALITY.DATA.# file containing the original water-quality data for the selected constituent.

User input: The constituent short name (sname) in all capital letters.

Output: The program creates a file which resides in the MAKE_DATA subdirectory under the root directory. This captures the tables which are printed to the screen. The program also produces the binary file named DATAFILE which resides in the constituent subdirectory.

The output file in the MAKE_DATA directory (named MAKE.sname) consists of three tables. The first table is a summary of reporting limits encountered or all the data from all the stations. Example:

REPORTING THRESHOLD SUMMARY FOR ALL OBSERVATIONS

REPORTING LIMIT	CODE	N	FRACTION	BEG YEAR	END YEAR
0.0	-888800.0	1434	0.4169	69	81
1.0	-888803.0	405	0.1177	81	86
2.00	-888802.0	213	0.0619	73	83

The first column lists the actual reporting limit value. The second column lists the internal code used by the program. The third column is the number of times the reporting limit was encountered in the entire data set. The fourth column is the fraction of the total number of observations that the reporting limit comprises. The fifth column is the first year that the reporting limit was encountered. The last column is the last year that the reporting limit was encountered.

The second table lists specific information for each station. Example:

ISTA	STAID	#AVAILABLE	#SELECTED	#OUT-OF-RANGE	FRAC-ELIMINATED	FRAC-CENS
1	7227500	52	52	0	0.000	0.635
2	7228000	47	47	0	0.000	0.660

The first column is the station sequence number. The second column is the station number. The third column is the total number of observations. The fourth column lists the number of values selected after eliminating multiple monthly values in excess of MAXOBS set in CONSTANTS.INS. The fifth column is the number of values that exceeds the minimum or maximum criteria set in the PARAM_INFO.FILE. The sixth column lists the fraction of the values for that station that was eliminated. The last column is the fraction of the selected values for that station that was below the reporting limits.

The third table is a repeat of the first table except that it lists the information only for selected data values and does not show the first and last years that the reporting limits were detected.

SELECTION 2 -- DEFINE SEASONS

Purpose: To create or edit the seasonal definition file (SEASONAL_INFO.FILE).

Input files: None.

User input:

CREATING THE FILE -- Years may be defined by 12, 6, 4, 3 or 2 seasons. Users will create SEASON_INFO.FILE by entering the beginning and ending month and day of each season for any of the definitions selected. The first season in each seasonal definition is defined as the first season with a beginning and ending date within one calendar year. So, if a calendar year boundary is crossed within a season, this will occur in the last season of a seasonal definition. The beginning and ending dates of the definitions that the user does not define will be set to zeros.

EDITING THE FILE -- Once SEASON_INFO.FILE has been created, any of the definitions (12, 6, 4, 3 or 2) may be modified. The beginning and ending month and day of each season must be entered for each seasonal definition the users selects to edit.

NOTE -- If a season is to end on the last day of February, the day must be entered as 29 to take into account leap years.

Output: The information is stored in the file named and located by the key word SEASON in PATHNAMES.FILE (section 7.3). The file is an ASCII file, but it should only be accessed through this program to insure proper format.

SELECTION 3 -- RUN SEASONAL COMPARISONS

Purpose: Program creates an output table that lists the fraction of valid seasonal comparisons to the total number of possible seasonal comparisons for each seasonal definition.

Input files: The DATAFILE for the selected constituent and the SEASON_INFO.FILE.

User input: Enter the constituent short name (SNAME) in all capital letters. Next enter the beginning and the ending year and month (YYMM) of the selected period of record. The program will ask for the method with which to analyze the constituents.

Output: The program creates two files. The first file will be named sname_SEAS.TBL. The file will reside in the subdirectory SEASON under the root directory. The file has two tables for each station. The first table lists the percentage of valid comparisons for the beginning/ending part of the record and the second table lists the fraction of valid comparisons for the middle part of the record. A valid comparison can be made when a non-missing value is present in a season. Example:

STATION I RECORD LENGTH 17.8 (BEG-END YEAR & SEASON 6810. 8609.)

BEGINNING-ENDING RECORD COMPARISONS ** YEARS COMPARED = 8.

```
12 1.000 0.000 1.000 0.000 1.000 0.000 1.000 0.750 0.250 0.250
 6 1.000 1.000 1.000 1.000 0.250 0.750
 4 1.000 1.000 1.000 0.750
 3 0.750 1.000 1.000
 2 0.750 1.000
```

MIDDLE RECORD COMPARISONS ** YEARS COMPARED = 9.8

```
12 0.333 1.000 0.333 1.000 0.333 0.800 0.467 0.800 0.467 0.778
 6 1.000 1.000 0.800 0.800 1.000 1.000
 4 1.000 1.000 1.000 1.000
 3 1.000 1.000 1.000
 2 1.000 1.000
```

The first line lists the station sequence number, the record length in decimal years, the beginning year and season, and ending year and season of the record for that station. The beginning and ending seasons correspond to the seasons specified by the user in the seasonal definition file for 12 seasons per year. The first season defined for 12 seasons per year will be season 01, the second season defined will be 02, etc.

The first column lists the seasonal definition that corresponds to each row. The remaining columns list the fraction of possible comparisons which can be made for each season in the seasonal definition. If the fraction is 1.000, then all possible comparisons for that season can be made. If the fraction is 0.500, then only 50 percent of the possible comparisons can be made.

The second file created by this program is named sname_SEASON. This file also resides in the subdirectory SEASON under the root directory. The file is in binary format and cannot be listed to the screen or spooled out at a printer. It is used as input to the program in selection 4 of the ESTREND menu.

SELECTION 4 -- SELECT BEST SEASONAL DEFINITION.

Purpose: The program creates a table which lists the "best" seasonal definition for a station based on user-selected criteria.

Input files: File sname_SEASON in the SEASON subdirectory.

User input: Enter the constituent short name. Then the program will prompt for the minimum acceptable ratio of actual comparisons to possible comparisons for each season in a single seasonal definition. For example, in a selected 10 year period, each season of every year is compared to that same season for every other year. The user enters the fraction of possible comparisons that would be adequate to define the trends in each season. Enter a number less than or equal to 1 (suggest 0.5). The program will then ask for the minimum acceptable ratio of seasons in a single seasonal definition that must meet the criterion above to qualify as the seasonal definition for a station. Again, enter a number less than or equal to I (suggest 0.8).

Output: The output table will be in a file named sname_SELSEAS.TBL in the SEASON subdirectory. The table will contain the user-selected criteria at the top of the table for reference and a set of columns across the page. Each set contains the station sequence number, the best seasonal definition for the beginning/ending period and the best seasonal definition for the middle period. Example:

SEASONAL TABLE FOR sname Season ratio .500
BE=beginnig/end selection; MI=middle selection Station ratio .500

STA	BE	MI
1	12	12
2	6	12
3	3	4
4	12	12
5	12	12
6	12	12
7	12	12
8	6	12
9	4	6
10	6	6

For station 1, the table shows that the seasonal definition with the largest number of seasons per year that meets the user's criteria is 12 for both the beginning/end and middle periods. For station 3, the best seasonal definition is 3 seasons per year for the beginning/end period and 4 seasons per year for the middle period. It is up to the user to decide which one to select, but, generally, the best seasonal definition for the beginning/end period would be the most appropriate choice.

SELECTION 5 -- SELECT FLOW MODELS OR SEASONS

Purpose: To create or edit the SEASON and FAC_MODEL files which reside in the constituent subdirectory and are used in the trend analysis programs.

Input files: None.

User input: Enter the constituent short name (sname) in all capital letters. The program will then prompt the user to-select either the SEASON file or the FAC_MODEL file. If the selected file does not exist in the constituent subdirectory, the program will display a message that the file is being created, otherwise the program displays a message stating that the file is being edited. The user then selects one of the following:

1) Set all values in the file to the same number.
2) Set all values in the file to different numbers.
3) Exit without changing the file.

The first choice causes the program to prompt for the one number with which to set the values in the file. The second choice causes the program to prompt for each value by station sequence number. The third choice returns the program to the selection of either the SEASON or FAC_MODEL file. These files contain a value for each station in the study. The SEASON file must have a 0,1,2,3,4,6, or 12 for each station to indicate the number of seasons to use for each station for the selected constituent. The FAC MODEL file must have a 0-14 for each station to indicate the flow adjustment model to use for each station for the selected constituent (see section 4.5 for a description of models 0-14). A table is shown at the top of the screen that lists the flow adjustment models by number. HINT: If several constituents have the same seasonal definitions or model assignments for all stations, it may be easier to create the SEASON file once and the use the operating system's COPY command to copy the file to the other constituent subdirectories.

Output: SEASON or FAC_MODEL files in the constituent subdirectory.

SELECTION 6 -- RUN TRENDS

Purpose: Program does summary statistics on data and runs trend estimation procedures for all stations for a selected constituent.

Input files: DATAFILE, DETLIMS, SEASON and FAC_MODEL files residing in the constituent subdirectory.

User input: Enter the constituent short name (sname) in all capital letters. The program will then prompt for the beginning and ending year and month (YYMM) for the selected period. Next the user will select from a menu shown on the screen the method with which to analyze the constituent.

Output: The program produces a file named RESULTS. This file resides in the constituent subdirectory. The file is in binary format and cannot be listed to the screen or printed. This file is accessed through the tabling, mapping and plotting programs available with this package.

SELECTION 7 -- TABLE RESULTS FOR A CONSTITUENT

Purpose: The program produces tables of the results from the statistical summaries and trend analysis produced from Option 6. It takes the results in binary format and produces an output file with the results in tabular form. The output is for all stations for a selected constituent.

Input file: File RESULTS in the constituent subdirectory.

User input: Enter the constituent short name in all capital letters. Then the program will prompt the user to select either estimated values or actual values and associated 'less-than' flags for the summary statistics for censored data.

Output: The program produces a file named RES_TBLS.sname, located in the subdirectory TABLES under the root directory. The program produces three tables in the output file. The first table contains the statistical summary of the concentration data. The second table reports all observed trends and flow adjustment information where applicable. The third table reports all significant (p=0.1) trends ranked by trend magnitude in units per year.

The elements of the first output table for each station by column are:

Station sequence number
Station number
Station name
Number of total available observations
Number of observations paired with flow
Mean (estimated for censored constituents)
Standard deviation
Skew (not calculated for censored constituents)
Minimum (not reported for censored constituents)
Fifth percentile
Twenty-fifth percentile
Median
Seventy-fifth percentile
Ninety-fifth percentile
Maximum

The elements of the second output table for each station by column are:

Station sequence number
Station number
Number of observations after seasonal selections
Concentration trend in units per year
Concentration trend in percent per year
Significance of concentration trend (p)
Flag for significant concentration trend (p<=0.1)
** Flag for significant Flow-Adjusted Concentration (FAC) trend (p<=0.1)
** Number of observations paired with flow after seasonal selection
** FAC trend in units per year
** FAC trend in percent per year
** Significance of FAC trend (p)
** Flow adjustment model number
** Flow adjustment model significance (p)
** R-square
Rank ordered correlation coefficient (Kendall's Tau) between concentration and flow (XCOR)
Significance of XCOR
** Spearman's Rho correlation coefficient between residuals and flow
** Best flag for choice between Concentration and FAC trends

(**indicates that the value is not reported for censored constituents)

The elements of the third output table for each station by column are:

Station sequence number
Station number
Number of observations
Trend in units per year
Trend in percent per year
Significance of trend

For the third table, the flow-adjusted concentration (FAC) trend, if it is calculated, is chosen if model 13 or 14 is selected. If the model selected is between 0 and 12 inclusive, the FAC trend is selected if the concentration versus flow regression is significant, otherwise the concentration trend is selected. Using these criteria, the significant trends are tabled in order of magnitude of the change in concentration units per year.

When sending the output files to the printer, use the -FTN option to maintain proper paging format.

SELECTION 8 -- MAP RESULTS FOR A CONSTITUENT

Purpose: This program produces a map showing the physical locations of the trends on a representation of the state. The direction of the trends is also shown. Another option in the program maps ranges of the mean concentrations. The program supports five output devices: a Hewlett Packard 7475A and 7550, META file output, a Tektronix 4105 color graphics terminal and a Tektronix 4107 or 4207 color graphics terminal.

Input files: RESULTS file in the constituent subdirectory.

User Input:

1. Enter the constituent short name (sname) in all capital letters.

2. The program will ask the user to select the type of plot desired. The selections are:

    1) MEAN CONCENTRATIONS - ranks mean concentrations across the state and divides them into quartiles. The quartiles are plotted on the map with either different symbols or different colors to distinguish between them.

    2) CONCENTRATION TRENDS - plots significant concentration trends on state map. Increasing trends are represented by an up arrow and decreasing trends are represented by a down arrow. Non-significant trends are represented by a small solid circle.

    3) FAC TRENDS (UNCENSORED STATIONS ONLY) - plots significant flow adjusted concentration (FAC) trends on a state map. Trends are represented the same as concentration trends. FAC trends are not calculated if the censored Seasonal Kendall trend test is used.

    4) TOTAL SIGNIFICANT TRENDS - FAC trend is chosen if model 13 or 14 is selected. If the model selected was between 0 and 12 inclusive, the FAC trend is selected if the concentration versus flow regression is significant, otherwise the concentration trend is selected. Using these criteria, the significant trends are plotted. Non-significant trends are represented by a small solid circle.

3. If selection 2, 3, or 4 is chosen, the user must enter the alpha level to use to determine which trends are significant. This should be entered as a real number less than one. The program will print to the screen a table listing the number of detected trends, the number of significant increasing trends and the number of significant decreasing trends.

4. If the MEAN CONCENTRATIONS option was selected in step 2, the user is asked if the quartiles are to be represented by different colors or by shades of one color. Choose by entering the number of the selection. If a trends map was selected in step 2, the user may request the trend arrows to be shaded to indicate magnitude. If shading to indicate magnitude is selected, the trends are ranked according to magnitude and the divided into quartiles. The lower quartile is not shaded at all, the second and third quartiles are progressively darker. The upper quartile is filled in.

5. The program will ask the user to specify the output device. Select the output device by entering the number associated with the device.

6. If the ZETA (POST PROCESSOR) device is selected, the program will ask for the META file name to use. If the Hewlett Packard plotter is selected, the program should begin to plot immediately after entering this selection. Make sure that the pens and paper are loaded in the plotter. The suggested pen colors are:

    Pen 1 - Black
    Pen 2 - Red
    Pen 3 - Green
    Pen 4 - Blue

Pens 5 and 6 are not used by the program.

If the Tektronix terminals were selected, the program will draw the map to the screen after entering the device selection. After the bell has sounded, hit return to bring the text back to the screen.

7. Once the plotting has finished, the user will be prompted to either replot the same data, create a new type of plot for the same constituent, begin with a new constituent or quit. Enter the number associated with the selection or hit return to quit the program.

8. The program takes advantage of a special feature on the Tektronix 4207 terminals. If it is selected as the output device, the program will plot what appear to be small dots in the middle of the trend arrows on the screen. Use the zoom feature on this terminal to enlarge areas of interest and the dots enlarge to be the station sequence number associated with the trend. This allows for easy reference between the screen and tables to determine exactly which stations have trends.

SELECTION 9 -- PLOT RESULTS.

Purpose: To produce plots of the data for a single constituent and station in various formats on different media.

Input files: RESULTS, DATA, DETLIMS, SEASON and FAC_MODEL files residing in the constituent subdirectory.

User input:

1. A menu will appear on the screen which lists the output devices available. The available graphics devices are GRAPHON terminal, Tektronix 4105, 4107 or 4207 graphics terminals, Hewlett Packard (HP)7475a and 7550 and a META file for use with the DISSPLA postprocessor. If the HP 7475a or 7550 plotter is selected, the user will select the paper size (8-1/2 X 11 or 11 X 17) and indicate whether color is to be used in the plot. If the META file is selected as the output device, the program will prompt for a file name to use and will default to the name 'META' if no file name is supplied.

2. The program will prompt for the constituent short name and the beginning and ending month and year of the selected period. Enter the short name in all capital letters and the dates in the format YYMM. The user will select the method for analysis of the constituent.

3. The main menu allows the user to select from the following options:

    1) DATA DISTRIBUTION PLOTS - univariate plots for a single station
    2) X - Y PLOTS - bivariate plots for a single station
    3) CHANGE OUTPUT DEVICE - allows the user to select a different output device
    4) SELECT NEW CONSTITUENT
    5) QUIT PROGRAM

4. Once the type of plot has been selected, the user will enter the selected station sequence number. Entering a "0" will cause the program to return to the previous menu.

5. The following two sections describe the elements of menu options 1 (data distribution plots) and 2 (x-y plots) from the main menu listed for selection 9.

Option 1: DATA DISTRIBUTION PLOTS.

A) The user will select either concentration or flow for the variable to be plotted for the selected station. Only values above the detection limit are used in this option.

B) The DATA DISTRIBUTION PLOT menu contains the following options:

    1) DISPLAY SAMPLE STATISTICS - lists the sample size, mean, standard deviation, variance, coefficient of variance, skew, kurtosis, median, range, first quartile value, third quartile value, interquartile range, number of outside values and number of detached values for the selected constituent at the selected station.

    2) PLOT BOXPLOT

    3) PLOT HISTOGRAM - allows the user to select between a Frequency or Relative Frequency histogram. The user is provided with a suggested number of intervals and interval size but may elect to change them.

    4) PLOT EMPIRICAL DISTRIBUTION FUNCTION

5) PLOT LILLIEFOR TEST FOR NORMALITY

6) PLOT SYMMETRY PLOT

7) TRANSFORM DATA - allows the following transformations on the entire data set:

   a) Cube, power of 3
   b) Square, power of 2
   c) Original units
   d) Square root, power of 1/2
   e) Logarithm (base 10), power of 0
   f) Reciprocal root, power of -1/2
   g) Reciprocal, power of -1
   h) Reciprocal square, power of -2

[Sequential transformations of the data set are not allowed. To return the data to the original units after a transformation, it is necessary to reselect the data for that station by choosing menu selection 9.]

8) SELECT FLOW/CONCENTRATION - may switch between concentration and flow as the selected plotting parameter for the selected station.

9) NEW STATION - allows the selection of a new station for the same constituent.

10) NEW OUTPUT DEVICE

11) NEW CONSTITUENT - upon selecting the new constituent, the user is returned to the main menu.

12) RETURN TO MAIN MENU

13) STOP - exits program

Option 2: X - Y PLOTS.

A) Select either all the available data or data that has been seasonally screened.

B) If uncensored methods were selected when choosing a constituent, the X - Y PLOTTING ROUTINES menu allows the following options:

   1) CONCENTRATION VS. FLOW. - allows the user to plot concentration versus flow values and to superimpose on the plot the results of 14 different flow-adjustment models. The models include linear, natural log, 8 different hyperbolic models, inverse and exponential models, and LOWESS smooths on raw and log values.

   2) RESIDUALS VS. FLOW

   3) CONCENTRATION VS.TIME

   4) RESIDUALS VS. TIME

5) FLOW VS. TIME

6) NEW STATION

7) NEW OUTPUT DEVICE

8) NEW CONSTITUENT

9) RETURN TO MAIN MENU

10) STOP

C) If censored methods were selected when choosing a constituent, the following X - Y plotting routine are available:

1) CONCENTRATION VS. FLOW
2) CONCENTRATION VS. TIME
3) FLOW VS. TIME
4) NEW STATION
5) NEW OUTPUT DEVICE
6) NEW CONSTITUENT
7) RETURN TO MAIN MENU
8) STOP

[Options I and 2 plot flagged less-than data as the flagged value and use a different symbol for the data point. The flagged value is also plotted in a different color on a color output device.]

D) For options 1,2,4, 5, and 6 using the uncensored methods and for all plots using the censored methods, a smooth of the data may be plotted on the scatterplot of the data. The smoothing technique used is LOWESS (Cleveland, 1979), which does a robust weighted least-squares smooth with bisquare weights. The user enters the proportion of differences of the x-axis used in smoothing, which is a number between 0 and 1.

E) Select the plot axis style from the following:

1) ARITHMETIC X-Y AXIS (both axis linear)
2) LOGARITHMIC X AXIS
3) LOGARITHMIC Y AXIS
4) LOG - LOG PLOT

F) Default titles and axis labels are supplied for the plots, but the user can elect to change any or all of the titles or labels.

G) The axis may be changed from the default lengths by entering a minimum and maximum axis value and the number of axis divisions required by the user.

Output: Selected plots.

SELECTION 10 -- TABLE RESULTS FOR ALL STATIONS

Purpose: Produces tables of the statistical summaries and trend test results for all stations for all constituents.

Input files: RESULTS file residing in the constituent subdirectory.

User input: Beginning and ending water year of selected period of record and the name of output file.

Output: Produces a file in the TABLES directory which contains separate results tables for all stations. Each table contains a header with station information including station number, station name, latitude, longitude and drainage area. The table also includes the constituent long name, sample size, mean, the 25th percentile value, the median and the 75th percentile value for each constituent with enough data to compute this information. If the trend test was executed for the constituent, the results of the test are listed including the number of observations used for the test, the magnitude and direction of the best trend in units per year and percent per year, the significance of the trend test and a code to indicate which trend was used as the best trend.

SELECTION 11 -- TABLE FLOW MODEL INFORMATION

Purpose: Tables the parameters of each flow model for each station for the selected constituent.

Input files: RESULTS file residing in the constituent subdirectory.

Output: The program produces a table that contains the statistical parameters of each regression (MODEL (1-11), INTERCEPT, SLOPE, SIGNIFICANCE(P), R-SQUARE, SPEARMAN'S RHO between residuals and flow, BETA for the hyperbolic models, the PRESS statistic, and a star for the MODEL SELECTED by the automated procedure), between flow and concentration. A second table contains the information for the log-log model. The information is produced for each station for the selected constituent and resides in the tables subdirectory in a file named MODE_INFO.sname.

SELECTION 12 -- TABLE SEASONAL TREND RESULTS

Purpose: Tables the parameters of the Seasonal Kendall trend test for the period selected as well as for each individual seasons in the selected seasonal definition for each station.

Input files: RESULTS file residing in the constituent subdirectory.

User input: The constituent short name in all capital letters.

Output: The program produces a file that contains the station number and period of trend test at the top of the table. The estimated trend magnitude for concentration and flow-adjusted concentration in units per year. The significance of the Chi-square test for homogeneity of seasonal trends in concentration is listed. A table of the number of observations for each season in the selected time period and the significance of the concentration trend test for that season is listed. The sign of the significance indicates the direction of the trend. Following this table is the significance of the Chi-square test for homogeneity of seasonal trends in flow-adjusted concentration. Finally there is another table which lists the number of observations for each season and the significance of the trend test for each season for flow-adjusted concentrations with the direction of the trend again indicated by the sign of the significance. This file resides in the TABLES directory under the name SEAS_KEN_TBL.sname.

REFERENCES

Alexander, R.B. and Smith, R.A., 1988, Trends in lead concentrations in major U.S. rivers and their relation to historical changes in gasoline-lead consumption: Water Resources Bulletin, v. 24, no. 3, p. 557-569.

Alley, W.M., 1988, Using exogenous variables in testing for monotonic trends in hydrologic time series: Water Resources Research, v. 24, no. 11, p. 1955-1961.

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: Journal of American Statistical Association, v. 74, no. 368, p. 829-836.

Cohen, A.C., Jr., 1950, Estimating the mean and variance of normal populations for singly truncated and doubly truncated samples: Annals of Mathematical Statistics, v. 21, p. 557-69.

Cohen, A.C., Jr., 1976, Progressively censored sampling in the three parameter log-normal distribution: Technometrics, v. 18, no. 1, p. 99-103.

Cohn, T.A., 1988, Adjusted maximum likelihood estimation of the moments of lognormal populations from type I censored samples: U.S. Geological Survey Open-File Report 88-350, 34 p.

Hald, A., 1949, Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point: Skandinavisk Aktuarietidskrift, v. 32, p. 119-134.

Hay, L.E..and Campbell, J.P., 1991, Water-quality trend in New Jersey rivers: U.S. Geological Survey Water Resources Investigations Report 90-4046.

Helsel, D.R., 1987, Advantages of nonparametric procedures for analysis of water quality data: Journal of Hydrological Sciences, v. 32, no. 2, p. 179-189.

Helsel, D.R. and Cohn, T.A., 1988, Estimation of descriptive statistics for multiply censored water quality data: Water Resources Research, v. 24, no. 12, p. 1997-2004.

Hirsch, R.M., Alexander, R.B., and Smith, R.A., 1991, Selection of methods for the detection and estimation of trends in water quality: Water Resources Research, v. 29, no. 5, p. 803-813.

Hirsch, R.M. and Slack, J.R., 1984, A nonparametric trend test for seasonal data with serial dependence: Water Resources Research, v. 20, no. 1, p. 727-732.

Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of trend analysis for monthly water-quality data: Water Resources Research, v. 18, no. 6, p. 107-121.

Hosmer, D.W., Jr. and Lemeshow, S., 1989, Applied Logistic Regression: New York, John Wiley and Sons, p. 11-18.

Kendall, M.G., 1975, Rank correlation methods (4th ed.): London, Charles Griffin.

Lanfear, K.J. and Alexander, R.B., 1990, Methodology to derive water-quality trends for use by the national water summary program of the U.S. Geological Survey: U.S. Geological Survey Open-File Report 90-359, 10 p. 56

Maddy, D.V., Lopp, L.E., Jackson, D.L., Coupe, R.H., and Schertz, T.L., 1989, National Water Information System User's Manual: vol. 2, chap. 2., WaterQuality System: U.S. Geological Survey Open File Report 89-617.

Mann, H.B., 1945, Non-parametric test against trend: Econometrica, v. 13, p. 245-259.

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

Myers, R.H ., 1986, Classical and Modern Regression with Applications: Boston, Massachusetts, Duxbury Press.

Sen, P.K., 1968, Estimates of the regression coefficient based on Kendall's Tau: Journal of American Statistical Association, v. 63, p. 1379-1389.

Schertz, T.L. and Hirsch, R.M., 1985, Trend analysis of weekly acid rain data: U.S. Geological Survey Water Resources Investigations Report 85-4211.

Schertz, T.L., 1990, Trends in water-quality data in Texas: U.S. Geological Survey Water Resources Investigations Report 89-4178, 177 p.

Smith, R.A., Hirsch, R.M., and Slack, J.R., 1982, A study of trends in total phosphorus measurements at NASQAN stations: U. S. Geological Survey WaterSupply Paper 2190.

Smith, R.A., and Alexander, R.B., 1983, A statistical summary of data from the U.S. Geological Survey's national water quality networks: U.S. Geological Survey Open-File Report 83-533, 30 p.

Smith, R.A., Alexander, R. B., and Wolman, M.G., 1987, Water quality trends in the Nation's rivers: Science, v. 235, p. 1607-1615.

U.S. Environmental Protection Agency, 1986, Quality criteria for water 1986: Office of Regulations and Standards, Washington, D.C., EPA 440/5-86-001.

Van Belle, G. and Hughes, J.P., 1984, Nonparametric tests for trends in water quality: Water Resources Research, v. 20, no. 1, p. 127-136. 57

Appendix I: Subdirectory contents in the ESTREND,

Root Directory contents.

TREND.CPL
    Menu interface which runs all of the programs in the package
INSTALL.CPL
    Copies distribution files to user's directory and builds ESTREND system

COMPILE_CPLS directory contents.

COMPILE_ALL.CPL
    Initiates CPL programs which compile and link all Fortran programs in the package
COMPILE_DEFINE_SEAS.CPL
    Compiles and links program DEFINE_SEAS.F77
COMPILE_ESTREND.CPL
    Compiles and links program ESTREND.F77
COMPILE_LIB.CPL
    Compiles and links TRENDLIB EPF library
COMPILE_MAKE_DATAFILE.CPL
    Compiles and links program MAKE_DATAFILE.F77
COMPILE_MAP.CPL
    Compiles and links program MAP.F77
COMPILE_PLOT.CPL
    Compiles and links program PLOT.F77
COMPILE_RES_TBLS.CPL
    Compiles and links program RES_TBLS.F77
COMPILE_SEASON.CPL
    Compiles and links programs SEAS_COMPS.F77 and SELSEAS.F77
COMPILE_SEAS_KEN_TBL.CPL
    Compiles and links program SEAS_KEN_TBL.F77
COMPILE_SEL_FLWSSN.CPL
    Compiles and links program SEL_FLWSSN.F77
COMPILE_STATION_TABLES.CPL
    Compiles and links program STATION_TABLES.F77
COMPILE_TBL_MODEL_INFO.CPL
    Compiles and links program TBL_MODEL_INFO.F77
DBG_ALL.CPL
    Initiates CPL programs which compile and link all Fortran programs in the package with the -DEBUG option
DBG_DEFINE_SEAS.CPL
    Compiles and links program DEFINE_SEAS.F77 with the -DEBUG option
DBG_ESTREND.CPL
    Compiles and links program ESTREND.F77 with the -DEBUG option
DBG_LIB.CPL
    Compiles and links TRENDLIB EPF library with the -DEBUG option
DBG_MAKE_DATAFILE.CPL
    Compiles and links program MAKE_DATAFILE.F77 with the -DEBUG option
DBG_MAP.CPL
    Compiles and links program MAP.F77 with the -DEBUG option

DBG_PLOT.CPL

Compiles and links program PLOT.F77 with the -DEBUG option

DBG_RES_TBLS.CPL

Compiles and links program RES_TBLS.F77 with the -DEBUG option 58

DBG_SEASON.CPL

Compiles and links programs SEAS_COMPS.F77 and SELSEAS.F77 with the -DEBUG option

DBG_SEAS_KEN_TBL.CPL

Compiles and links program SEAS_KEN_TBL.F77 with the -DEBUG option

DBG_SEL_FLWSSN.CPL

Compiles and links program SEL_FLWSSN.F77 with the -DEBUG option

DBG_STATION_TABLES.CPL

Compiles and links program STATION_TABLES.F77 with the -DEBUG option

DBG_TBL_MODEL_INFO.CPL

Compiles and links program TBL_MODEL_INFO.F77 with the -DEBUG option

SOURCE directory contents.

DEFINE_SEAS.F77

Program assists the user in creating the file which contains the definitions of each seasonal choice (2,3,4,6, and 12 seasons per year)

ESTREND.F77

Driver for the routines which estimate the trends and do the summary statistics

MAKE_DATAFILE.F77

Creates the binary datafiles for each constituent from the flat ASCII files produced from the retreival from the QW database

MAP.F77

Program produces map of trend results and station mean values plotted on an outline of the state. Output is either to a terminal, metafile or plotter.

PLOT.F77

Produces X-Y plots and statistics plots of the data for a constituent. Output is to a terminal, plotter or meta file.

RES_TBLS.F77

Program produces a dump of summary statistics and trend results for all stations for a single constituent. Results and statistics are in tabular form.

SEAS_COMPS.F77

Assists the user in determining which seasonal choice to use in defining the seasonal selection for a station. The program uses the seasons defined by program DEFINE-SEAS and checks for the highest ratio of seasonal comparisons.

SEAS_KEN_TBL.F77

Outputs a table of the trend results for each season in the selected seasonal definition for all stations for a single parameter.

SELSEAS.F77

Produces a table of the best seasonal selections for each station. Table is based on the results of the program SEL_SEAS.

SEL_FLWSSN.F77

Program assists the user in creating the files FAC_MODEL and SEASON which reside in the constituent subdirectories.

STATION_TABLES.F77

Produces a near publication ready set of tables for each station for all constituents and properties. The tables contain headers with station information and summary statistics and trend results for all constituents and properties at that station. Tables are paginated with FORTRAN paging controls and may contain more than one station per page. If no trend was found at a station, that station is ignored.

TBL_MODEL_INFO.F77

Tables the flow-adjustment model information (significance, Spearman rho, etc.) for all flow-adjustment models for all stations for a single constituent or property.

SOURCE>LIB directory contents.

AMLEREG.F77

Censored regression routines BEST.F77 Selects between Flow-Adjusted trend and Concentration trend based on trend significance and flow-adjustment model information for selected flow model

CDFN.F77

Cumulative distribution function for the normal zero-one distribution

CHECK_CENS_ENDS.F77

Checks for enough data in the ends of the selected period to justify attempting the trend test on censored data

CHECK_ENDS.F77

Checks for enough data in the ends of the selected period to justify attempting the trend test on noncensored data

CLS.F77

Function to clear terminal screen

CROSS.F77

Computes the cross correlation coefficient

CS_SEVENVS.F77

Computes percentiles for censored data. Percentiles are computed on actual data and corresponding less-than flags are associated with data below the reporting limit(s)

DTIME.F77

Converts date in MMYYDD format to decimal time with beginning of the full period of record as base time (begin time = 0.0)

FLOW_ADJ.F77

Does flow-adjustment on noncensored data according to different models and stores all pertinent information concerning the models

GET_FNAME.F77

Utility to return a pathname for a selected constituent

GET_UNIT.F77

Utility to return an available (open) FORTRAN file unit

LOWESS.F77

Performs LOcally WEighted Scatterplot Smoothing

MOM4.F77

Returns mean, standard deviation and skew

PACKER.F77

Removes all missing data from a vector and packs the data into the top of the vector

PRCTLS.F77

Computes percentiles for noncensored data

RANIK.F77

Ranks the values in a vector

REGRCDD.F77

Performs regression on censored data with multiple reporting limits REGRES.F77 Standard simple all-purpose linear regression

REGRESS2 F77

Returns standard regression statistics for censored data

RMED.F77

Returns the median value of a vector

SEAS_OBS.F77

Subroutine does seasonal selection on the raw data. For each defined season, the middle value in time in that season is returned from the traw data vector.

SEL_DATA.F77

Selects the data for a specified period from the binary DATAFILE which resides in each constituent sub-directory

SEL_HEAD.F77

Reads in all header information from the header file

SEL_RES.F77

Reads in and stores results for all stations for a single constituent or property. Used in all tabling, mapping and plotting programs.

SET_KEN.F77

Subroutine sets-up two variables to perform Kendall rank correlation test SKEND.F77 Performs Seasonal Kendall test for trend on noncensored data

SORTDN.F77

Sorts a vector in descending order

SORTDNP.F77

Sorts a vector in descending order and returns the permutation

SORTUP.F77

Sorts a vector in ascending order

SORTUPP.F77

Sorts a vector in ascending order and returns the permutation

STREND.F77

Subroutine provides summary statistics for concentration and flow including means, standard deviations, skew, and percentile estimates. Correlation statistics for flow and concentration are made. This routine calls the routines that do the trend estimation and summary statistics on both censored and noncensored data.

STUDENT.F77

Subroutine calls an IMSL routine to do Student’s t distribution

<u>SOURCE>INSERT directory contents.</u>

CONSTANTS.INS

Contains constants for array dimensioning, censored levels and pathnames

COMMON.DATA.INS

Contains the variable declarations for the common block DATA in the source code

COMMON.DLIMITS.INS

Contains the variable declarations for the common block DLIMITS in the source code

COMMON.MDATA.INS
: Contains the variable declarations for the common block MDATA in the source code

COMMON.RFILE.INS
: Contains the variable declarations for the common block RFILE in the source code

COMMON.SEASON DATA.INS
: Contains the variable declarations for the common block SEASON-DATA in the source code

COMMON.STATIONS.INS
: Contains the variable declarations for the common block STATIONS in the source code

COMMON.USER SEL.INS
: Contains the variable declarations for the common block USER_SEL in the source code

<u>RUN directory contents.</u>

MDL.LIB
: EPF library containing routines which do estimation of censored summary statistics

SMOOTHLIB
: EPF library containing routines which do line smoothing in plotting program (called by routine LOWESS.F77)

STATPLOT.UTIL.BIN
: EPF library containing utility routines used by the plotting program

PLOTLIB
: EPF library which contain routines used by the plotting program

TRENDLIB
: EPF library containing compiled and linked versions of all the FORTRAM outines contained in the subdirectory *>SOURCE>LIB which is used by all of the programs in the package

DEFINE SEAS.RUN
: EPF runfile for program DEFINE-SEAS

ESTREND.RUN
: EPF runfile for program ESTREND

MAKE.CPL
: Runs program MAKE_DATAFILE and records information in a command output (como) file contained in the MAKE_DATA subdirectory

MAKE_DATAFILE.RUN
: EPF.runfile for program MAKE_DATAFILE

MAP.RUN
: EPF runfile for program MAP

PLOT.RUN
: EPF runfile for program PLOT

RES_TBLS.RUN
: EPF runfile for program RES_TBLS

SEAS_COMPS.RUN
: EPF runfile for program SEAS_COMPS

SEAS_KEN_TBL.RUN
: EPF funrile for program SEAS_KEN_TBL

SELSEAS.RUN
: EPF runfile for program SELSEAS

SEL_FLWSSN.RUN

EPF runfile for program SEL_FLWSSN

STATION_TABLES.RUN

EPF runfile for program STATION_TABLES

TBL_MODEL_INFO.RUN

EPF runfile for program TBL_NODEL_INFO

DOC directory contents.

SETUP.INSTRUCTIONS

Instructions on preparing and running the ESTREND package

PARAM_INFO.FILE.EXAMPLE

Example of the file PARAM_INFO.FILE

PATHNAMES.FILE.EXAMPLE

Example of the file PATHNAMES.FILE

RFILE.DEFS

Definitions of all results variables stored in the file RESULTS in the constituent subdirectories

*U.S. GOVERNMENT PRINTING OFFICE 1991-0-573-207/50001