CHAPTER 23

NONPARAMETRIC TESTS FOR TREND DETECTION

23.1 INTRODUCTION

As demonstrated by the water quality and quantity applications in Sections 22.4 and 19.2 to 19.5, *intervention analysis* constitutes a flexible tool for rigorously ascertaining the effects of interventions upon the mean level of a series. Because the intervention model in [19.5.8] contains parameters which can be conveniently estimated when the innovations are assumed to be normally independently distributed, this model is referred to as a *parametric model*. Even when the data are quite messy and contain many missing values, after employing a technique such as the seasonal adjustment procedure of Section 22.2 to fill in the missing observations, the water quality applications of Section 22.4.2 demonstrate that an appropriate intervention model can then be conveniently calibrated to the estimated series of evenly spaced data. Diagnostic checks of the model residuals of the fitted water quality intervention model in [22.4.5], as well as all the other intervention models fitted to time series in Chapters 19 and 22, confirm that the underlying assumptions of intervention models can be readily satisfied in practical applications. Consequently, the intervention model is a powerful parametric statistical technique for use in *environmental impact assessment.*

Another family of parametric models for employment in environmental impact assessment is the *regression analysis* set of models described in Chapter 24. An advantage of regression analysis is that it can be used with both evenly or unevenly spaced observations.

In order to lessen the number of underlying assumptions required for testing a hypothesis, such as the presence of a specific kind of trend in a data set, researchers developed *nonparametric tests*. Because a nonparametric test is a method for testing a hypothesis whereby the test does not depend upon the form of the underlying distribution of the null hypothesis, a nonparametric test is often referred to as a *distribution free or distribution independent method.* As a matter of fact, some distribution free methods assume there are parameters in the models which form the basis for the tests whereas other distribution free tests do not involve any parameters, either directly or indirectly in the tests. Although the term nonparametric should be confined to describing distribution free tests for which there are no parameters, in practice it has been interpreted as standing for the set of all distribution free methods. Hence, within this text the more commonly used phrase of nonparametric tests will be used even though it is more correct to utilize the expression of distribution free tests.

Nonparametric tests were developed for use in environmental impact assessment because scientists were concerned that the statistical characteristics of *messy environmental data* would make it difficult to use parametric procedures. As noted by Hirsch and Slack (1984), natural time series may contain one or more of a number of properties which are undesirable for use with parametric tests. In particular, hydrologic and water quality data may be nonnormally distributed and follow a distribution which is usually positively skewed. Because the adoption of proper sampling procedures are often not considered, environmental time series are not commonly measured at uniform time intervals. Moreover, data are often *censored* by only listing measurements below a certain level as being "less than" or measurements above a specified level as being "greater than". For instance, concentration values for metals or organic compounds which fall below the *detection limits* for certain chemical tests are reported simply as less than the limits of detection. Fortunately, the foregoing and other characteristics of environmental time series can often be properly accounted for in order to make the data suitable for use with parametric testing. For example, as noted at many locations in this text, invoking a data transformation, such as the Box-Cox transformation in [3.4.30], can often help to alleviate nonnormality, although this is not always the case. Depending upon how much data are missing, an appropriate *data filling* procedure can be selected from Sections 22.2, 19.3, or 18.5.2 to obtain an estimated evenly spaced time series. When observations are given as less than the detection limit, one way to estimate what they should be is to consider them to be missing and estimate them using a suitable data filling procedure.

Nonparametric tests have few underlying assumptions and tend to ignore the magnitude of the observations in favour of the relative values or *ranks* of the data. As a result, a given nonparametric test which is designed, for instance, for checking for the presence of a trend, may only provide a yes or no answer as to whether or not a trend may be contained in the data. The output from the nonparametric test may not give an indication of the type or magnitude of the trend. In order to have a more precise test about what is occurring, many assumptions must be made and as more and more assumptions are formulated, nonparametric tests begin to look more and more like parametric tests. As a matter of fact, as noted by Savage in the encyclopedia edited by Kruskal and Tanur (1978, p. 637), the dividing line between nonparametric and parametric tests is not a sharp one. Finally, Conover and Iman (1981) explain how rank transformations of data sets act as a bridge between parametric and nonparametric statistics.

Because nonparametric tests are usually designed to indicate the presence but not the magnitude of a given statistical characteristic, some authors consider them to be exploratory data analysis procedures. Nonetheless, as is explained for the nonparametric tests described in Section 23.3, and Appendices A23.1 to A23.3, all these tests are designed for specifically testing certain hypotheses. Since they are utilized for *hypothesis testing*, within this text nonparametric tests are deemed to be confirmatory data analysis tools. Of course, after detecting the presence of a trend using a nonparametric test, a more powerful confirmatory data analysis method such as the intervention analysis technique of Section 22.4 and Chapter 19, can be employed for obtaining precise statistical statements about the trends.

Over the years practitioners have argued about whether nonparametric or parametric tests should be employed. An *advantage* of nonparametric tests is that they are distribution free and hence fewer assumptions have to be made about the data. On the other hand, as shown for the intervention model in Section 22.4, often many difficulties with the data which appear to make the series unusable with a parametric technique, can, in fact, be overcome. Cox and Hinkley (1974, Section 6.1) describe a number of *drawbacks* to nonparametric tests which they say limit their practical importance. One of the limitations is that when a parametric test is appropriate, a nonparametric test cannot be as powerful as the most efficient parametric test. Additionally, the results from a nonparametric test often do not adequately describe what is happening with a data set. In order to achieve a reasonable description and understanding of the system under investigation in concise and simple terms, a parsimonious parametric model is required where each parameter describes some important aspect of the system. Keeping in mind the assets and limitations of both parametric and nonparametric tests, a pragmatic approach to data analysis may be to use whatever tests seem to be most appropriate, whether they are nonparametric or parametric tests. For example, when analyzing vast amounts of environmental data for the presence of trends, in conjunction with the exploratory data analysis tools of Section 22.3, nonparametric testing can be used to locate the data which contain trends. Parametric techniques for detecting trends in a time series are referenced in Sections 19.2.3 and 24.2.1. Subsequent to a perusal of the written historical records to find physical causes for trends in the data, intervention analysis (Chapter 19 and Section 22.4) or regression analysis (Chapter 24) can be employed for obtaining rigorous statistical statements about the types and magnitudes of the trends. Bloomfield et al. (1983), Bloomfield (1992) and Bloomfield and Nychka (1992) present frequency domain approaches for analyzing trends. Lettenmaier (1976) compares the ability of various nonparametric and parametric tests for detecting step and linear trends.

From an intuitive point of view, it may be instructive to consider an overall trend to consist of deterministic and stochastic components such that

$$\begin{matrix} \textit{overall} \\ \textit{trend} \end{matrix} = \begin{matrix} \textit{deterministic} \\ \textit{trend} \\ \textit{component} \end{matrix} + \begin{matrix} \textit{stochastic} \\ \textit{trend} \\ \textit{component} \end{matrix}$$

In fact, this kind of interpretation forms the basis of the general intervention model in [19.5.8]. The *stochastic trend* is accounted for by the noise component which contains AR and MA operators to reflect the nonseasonal and seasonal correlation structures of the output and a white noise term for modelling what is left over when all the linear relationships contained in the output have been removed. Additionally, the noise component may require nonseasonal and seasonal differencing operators for modelling the nonstationary characteristics of a stochastic trend. Notice for the intervention model in [19.5.8] that when covariate series are available, they can also be incorporated into the model as extra information which usually increases the accuracy of all the parameter estimates in the overall intervention model and also removes effects upon the output which are not due to one or more interventions. Because it models the effects of one or more known interventions upon the output, the intervention term in [19.5.8] is dependent upon the time of occurrence of an intervention and therefore can be thought of as being a *deterministic component*. However, it should be kept in mind that this component is specifically designed to statistically describe the effects of known physical causes upon the output. Further discussions regarding deterministic and stochastic trends are presented in Sections 23.4.4 and 4.6.

When utilizing a statistical test, such as a specific kind of nonparametric test, to ascertain if there are trends in the data, one should always keep in mind exactly what the test is designed to detect and what are the *underlying assumptions* for the test. For example, due to the theoretical construction of a certain nonparametric test, it may only be designed for discovering the presence of an overall trend and may be incapable of distinguishing between stochastic and deterministic trend components. Further, as is the case for all of the nonparametric tests described in Section 23.3 for finding trends, nonparametric tests can only detect if a trend exists between the beginning and end of a time series and they cannot ascertain when the trends started due to external interventions. As a matter of fact, when only a small amount of data is available, the detection of the presence of trends is often all that one can realistically hope to achieve. Upon the collection of additional data, a more sophisticated procedure, such as intervention analysis, can be employed to describe more precisely the trend effects of known interventions.

When employing a nonparametric or parametric test to check for the presence of trends in a time series, there are different general approaches by which a given statistical test can be designed and executed. Therefore, a brief review of the *statistical testing* procedures consisting of hypothesis testing and significance testing is presented in the next section. Following a general discussion of available nonparametric tests in Section 23.3.1, specific nonparametric tests, which are used in the water resources literature for discovering trends in water quality and quantity time series (Hirsch et al., 1982; Hirsch and Slack, 1984; Van Belle and Hughes, 1984; Hirsch and Gilroy, 1985), are described in detail in Section 23.3.2. Of particular importance is the seasonal Mann-Kendall test (Hirsch et al., 1982) which can be used to test for the presence of a trend in each season of the year for a given data set. When testing for the presence of trends in, say, monthly data, one may wish to know whether each month should be checked separately or perhaps different groups of months should be tested together. Consequently, within Section 23.3.3 procedures are discussed for deciding upon how data should be grouped, and, in particular, the technique suggested by Van Belle and Hughes (1984) is discussed. For combining tests of hypotheses across seasons or groups of seasons, Fisher's (1970) method described in Section 23.3.4 is recommended. When dealing with water quality time series, often the effects of water quantity upon the water quality variables must be properly accounted for and in Section 23.3.5, regression analysis approaches for accomplishing this are described. In Section 24.3.2, the robust locally weighted regression smooth of Section 24.2.2 is utilized to allow for the effects of flow upon water quality when carrying out trend analysis studies of water quality time series measured in rivers. The Spearman partial rank correlation test is presented in Section 23.3.6 as a flexible nonparametric test for discovering trends in, say, a water quality variable measured over time when partialling out the effects of seasonality or riverflows upon the water quality variable. A nonparametric test is described in Section 23.3.7 for checking for the presence of a step trend caused by a known intervention in series measured at multiple stations. This test was devised by Hirsch and Gilroy (1985) and Crawford et al. (1983) and is related to the Mann-Whitney rank-sum test. In the final part of Section 23.3, procedures are presented for handling multiple censored data that are to be subjected to nonparametric trend testing. Within Section 23.4, the ACF (autocorrelation function) at lag one is suggested as a parametric test for finding trends. Using simulation experiments, the power of Kendall's tau (or equivalently the Mann-Kendall statistic) for detecting trends is compared to the power of this parametric statistic. The ACF at lag one is found to be more powerful than Kendall's tau for discovering purely stochastic trends while Kendall's tau is more powerful for finding purely deterministic trends. To demonstrate clearly the efficacy of utilizing various nonparametric tests and also some parametric methods in environmental impact assessment, practical applications are given in Section 23.5. In particular, nonparametric tests are utilized for discovering trends in water quality variables in Lake Erie caused by industrial development at the town of Nanticoke situated on the north shore of Lake Erie in the Canadian province of Ontario.

The nonparametric tests described in Chapter 23 are listed in Table 23.1.1 along with brief descriptions of their main purposes, equation numbers for the test statistics and the reference sources. Notice that the first six nonparametric tests are designed for checking for the presence of trends for a variety of situations and all of the trend tests are explained in Section 23.3. In addition to handling tied data, it is pointed out how the trend tests can be employed with censored time series. For the case of the seasonal Mann-Kendall test, procedures for taking care of correlation are also given. Finally, the last three nonparametric tests given in Table 23.1.1 are described in the three appendices and can be utilized for various useful tasks within a systematic data analysis study.

Table 23.1.1. Nonparametric tests described in Chapter 23.

NAMES	PURPOSES	TEST STATISTIC EQUATIONS	SOURCES
Nonseasonal Mann-Kendall	Determine if a time series contains a monotonic trend over time.	[23.3.1] and [23.3.4]	Mann (1945)
Seasonal Mann-Kendall	Find out if a seasonal time series contains an overall trend component.	[23.3.7], [23.3.8] and [23.3.11]	Hirsch et al. (1982), Hirsch and Slack (1984), Van Belle and Hughes (1984) and Lettenmaier (1988)
Aligned rank	Ascertain if a deseasonalized time series possesses a trend.	[23.3.23]	Sen (1968), Farrel (1980) and Van Belle and Hughes (1984)
Spearman's rho	Check if there is significant correlation between 2 variables X and Y. Can also be used as a trend test if one of the variables is time and the other is a sequence of observations.	[23.3.33] and [23.3.34]	Spearman (1904)
Spearman partial rank correlation	Determine the correlation between variables X and Y after the effects of Z upon X and Y are partialled out. Can be used to check if there is a trend in a series over time after seasonality is partialled out.	[23.3.35] and [23.3.36]	Based upon Spearman (1904)
Step trend	Find out if a known intervention causes significant step trends in series measured at multiple stations. Test is based on Mann-Whitney rank-sum test on grouped data.	[23.3.38] and [23.3.44]	Hirsch and Gilroy (1985) and Hirsch (1988)
Kendall rank correlation	Ascertain if two series X and Y are independent of one another. The Mann-Kendall trend test is a special case when one series is time and the other is sequential observations.	[A23.1.1]	Kendall (1975)
Wilcoxon signed rank	Check if two samples X and Y have the same median.	[A23.2.2] and [A23.2.3]	Wilcoxon (1945)
Kruskal-Wallis	Determine whether or not the distributions across k samples are the same. Can also be used to check if a time series possesses seasonality.	[A23.3.2] and [A23.3.3]	Kruskal and Wallis (1952)

As summarized in Table 1.6.4, three general approaches for carrying out trend analysis studies are presented in Sections 22.4, 23.5 and 24.3. Within each of these methodologies, appropriate exploratory and confirmatory data analysis tools can be employed, including the nonparametric techniques of Table 23.1.1. In fact, for the overall trend assessment procedures explained using environmental applications in Section 23.5 of this chapter and also Section 24.3, nonparametric trend tests have a key role to play. Besides the general approaches given in this book, other methodologies for trend assessment have been devised by researchers. For example, Montgomery and Reckhow (1984) suggest an overall systematic procedure for determining the presence or absence of trends in environmental data. Depending upon the characteristics of the data being analyzed, they suggest various nonparametric and parametric tests which can be used. Berryman et al. (1988) and Harcum et al. (1992) present systematic procedures for deciding upon which nonparametric tests to employ for detecting trends in water quality time series. Hipel and McLeod (1989) explain how both parametric and nonparametric models can be employed in trend assessment within the overall framework of exploratory and confirmatory data analyses. Hirsch et al. (1991) put forward procedures for selecting statistical methods to detect and estimate trends in water quality time series.

23.2 STATISTICAL TESTS

23.2.1 Introduction

Statistical testing can be carried out using nonparametric or parametric tests. However, a given statistical test, either nonparametric or parametric, can be designed for the purpose of hypothesis testing or significance testing. Cox and Hinkley (1974) present detailed descriptions of various kinds of hypothesis and significance tests which are briefly described in this section. The theory of tests of hypotheses was originally developed by Neyman and Pearson (1928, 1933) while significance testing is due largely to Fisher (1973).

23.2.2 Hypothesis Tests

Suppose one would like to determine whether or not a data set possesses a certain property. For example, one may wish to ascertain the existence or nonexistence of a certain kind of trend in a water quality time series. Typically, the *null hypothesis*, H_0, which is sometimes called the hypothesis under test, is that the population from which the sample data set is drawn, does not possess a specified property like a trend. The *alternative hypothesis*, H_1, which specifies a direction of departure from H_0, is that the data set does exhibit the property. In order to choose between H_0 and H_1 a *test statistic*, T, which is a function of the data set $\mathbf{X} = (x_1, x_2, \ldots, x_n)$, is defined. When the hypothesis H_0 is true, the distribution of T must be known, at least approximately, so that the hypothesis test can be executed. By knowing the probability distribution of T, the probability of the sample statistic falling within or outside a given interval can be determined. As shown in Figure 23.2.1, for one sided and two sided tests, let t_r and t_l stand for the right and left values of T, respectively, for the two possible one sided tests, and let t'_r and t'_l define the right and left ends, respectively, of an interval in a two sided test. A chosen *significance level*, α, is the probability that the sample falls outside a specified range of values, given that H_0 is true. In practice, α is selected to have a value of 0.10, 0.05, or 0.01, although any appropriate value can be selected. For the one tailed or *one sided tests*, α represents the area in one of the tails of the distribution. In Figure 23.2.1a, $Pr(t \geq t_r) = \alpha$, and, consequently, if one were checking for the absence of an increasing trend, the hypothesis H_0 would be rejected if $t_x \geq t_r$ and hence H_1 would be accepted, where t_x is the sample or estimated value of T calculated using the sample X. Alternatively, H_0 would be accepted if $t_x < t_r$. The one tailed test in Figure 23.2.1b works in a similar manner. Suppose that H_0 indicates the absence of a decreasing trend. If $t_x \le t_l$, then H_0 would be rejected and H_1 thereby accepted, whereas, when $t_x > t_l$, H_0 would be considered to be correct. A *two sided test* would be used in trend detection when one wishes to test for the presence of a trend which could be increasing or decreasing. For the case of the two sided test in Figure 23.2.1c, H_0 is accepted when $t'_l < t_x < t'_r$ and is rejected when $t_x \ge t'_r$ or $t_x \le t'_l$.

Notice from Figure 23.2.1, that the probability of selecting H_0 when H_0 is true is $1-\alpha$, which is referred to as the *confidence level*. When executing a hypothesis test, two types of errors can arise. The probability of rejecting H_0 when H_0 is true is called a *type 1 error* or error of the first kind. From Figure 23.2.1, the probability of committing a type 1 error is α for both one sided and two sided tests. If, as a result of the same test statistic and a chosen significance level α, the hypothesis H_0 is accepted when it should be rejected, this is called a *type 2 error* or error of the second kind. Letting β represent the probability of committing a type 2 error, the probability of not making a type 2 error is $1-\beta$. The probability of rejecting H_0 when H_1 is true, or equivalently, the probability of not making a type 2 error, is called the *power* of the hypothesis test. If, for example, one were testing for the presence of a trend, the power, given by $1-\beta$, can be interpreted as the probability of detecting a trend when a trend is actually present in the data. In Section 23.4, the powers of two tests for detecting trends are compared using simulation studies for a number of different data generating models. When performing a hypothesis test, one of the four situations given in Table 23.2.1 can arise. Notice that two of the four outcomes to a hypothesis test result in either a type 1 or type 2 error. In Section 23.5, a variety of practical applications are presented using nonparametric statistics for hypothesis testing when checking for the presence of trends in water quality data.

In general, the power, $1-\beta$, and the confidence level, $1-\alpha$, are inversely related. Consequently, increasing the confidence level decreases the power and vice versa. For a specified significance level, α, the power of a test may be made greater by increasing the sample size.

23.2.3 Significance Tests

The type of significance test described here is what Cox and Hinkley (1974, Ch. 3) refer to as a *pure significance test*. As is also the case for hypothesis testing, when designing a significance test the null hypothesis, H_0, must be precisely formulated in terms of a probability distribution for the test statistic T. On the other hand, the major difference between the two tests is that a possible departure from H_0 in the form of an alternative hypothesis, H_1, is not rigorously defined for a significance test whereas it is assumed to be exactly known for a hypothesis test. Nonetheless, for a significance test it is necessary to have some general idea about the type of departure from H_0.

When performing a significance test, the acceptance or rejection of H_0 is decided upon in the same way as it is for a hypothesis test. Hence, as shown in Figure 23.2.1, one can perform either a one sided or two sided significance test. However, when H_0 is rejected only the general kind of departure from H_0 is known because there is no precise statement about an alternative hypothesis.

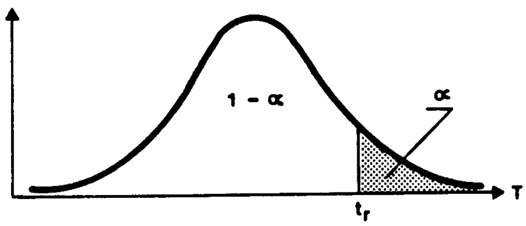

Figure 23.2.1a. One sided hypothesis test on the right.

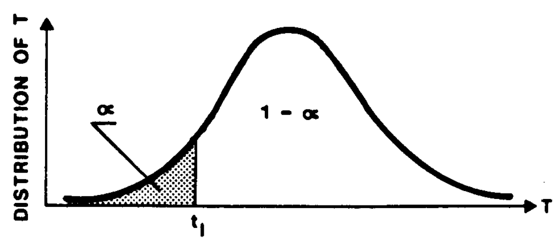

Figure 23.2.1b. One sided hypothesis test on the left.

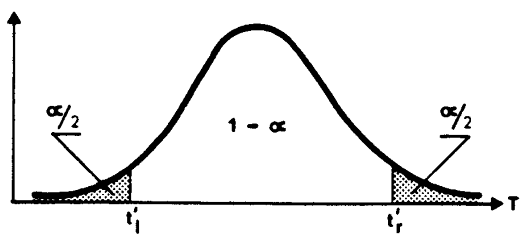

Figure 23.2.1c. Two sided hypothesis test.

Figure 23.2.1. Hypothesis tests.

Most of the diagnostic tests given in Chapter 7 constitute significance tests. For example, when one is checking whether or not the residuals of a stochastic model fitted to a time series are white, the null hypothesis may be that the residuals are white. Based upon an appropriate statistical test such as the Pormanteau statistic in [7.3.6], one can decide if H_0 should be accepted or rejected. When H_0 is rejected because the residuals are not white, the precise type of departure from H_0 is not explicitly defined. The residuals may be correlated for instance, because an ARMA(1,1) model should be fitted to the data instead of a an AR(1) model.

Table 23.2.1 Possible results of an hypothesis test.

	ACTION	
TRUE SITUATION	Accept H_0	Reject H_0
H_0 is true	No error (Pr = 1 - α) Confidence Level	Type 1 error (Pr = α)
H_1 is true	Type 2 error (Pr = β)	No error (Pr = 1 - β) Power

To avoid using too much statistical jargon and to enhance understanding by practitioners, hypothesis and significance tests are often not stated in a very formal manner when applying them to real data. As a matter of fact, statistical tests can be of assistance in the design of informal statistical tools for use in exploratory data analysis (see problem 22.5 in the previous chapter).

23.3 NONPARAMETRIC TESTS

23.3.1 Introduction

As noted in Section 23.1, a nonparametric test is also commonly referred to as a *distribution free or distribution independent method.* This is because no assumptions are made about the specific kind of distribution that the samples follow. The only restriction is that the samples come from the same basic population. Furthermore, a nonparametric test tends to be quite simple in design and easy to understand. In terms of time series analysis, most nonparametric tests can be used with both evenly and unevenly spaced observations.

Different types of data are available for use with statistical tests. The kinds of measurements are usually referred to as *measurement scales* or *systems of measurement.* From weakest to strongest, the four types of measurement scales recognized by Stevens (1946) are the nominal, rank (also called ordinal), interval, and ratio scales. In the *nominal scale* of measurement, numbers or other appropriate symbols are used for classifying objects, properties or elements into categories or sets. For example, when classifying objects according to colour, any number can be selected as a name for a given colour.

In the *rank or ordinal scale,* objects are ranked or ordered on the basis of the relative size of their measurements. A wine taster, for instance, may rank his wines from most to least desirable where a wine having a larger number assigned to it is more preferred than one with a smaller number. However, the amount by which one wine is preferred or not preferred over another is not specified. When analyzing a real world dispute using the conflict analysis approach referred to in Section 1.5.3, only ordinal preference information is assumed and hence each decision maker must rank the possible states or scenarios in the conflict from most to least preferred.

Besides the relative ordering of measurements used in the ordinal scale, the *interval scale* of measurement takes into account the size of the interval between measurements. An informative example of an interval scale is the common scales by which temperatures are usually measured. When using either the Fahrenheit or Celsius scale, a zero point and a unit distance (i.e., one degree of temperature) must be specified. Besides the Fahrenheit or Celsius scales, one could easily define any other temperature scale by stipulating the zero point and the degree unit. In other words, the principle of interval measurement is not violated by a change in scale or location or both.

To convert x degrees Celsius to y degrees Fahrenheit one uses the equation

$$y = \frac{9}{5}x + 32$$

When employing the interval scale, ratios have no meaning and one cannot, for instance, state that 10°C (50°F) is twice as warm as 5°C (41°F). Although the ratio of the two temperatures is $\frac{10}{5} = 2$ when using the Celsius scale, in the Fahrenheit scale the ratio is $\frac{50}{41} = 1.22$. Because there is a true or natural zero point in the Celsius scale, the concept of a ratio makes sense in this scale. When transforming from one *ratio scale* to another, it is only necessary to multiply one of the scales by a constant. Thus, for example, to transform x kilometres to y miles one uses the equation

$$y = 0.62x$$

One can say, for instance, that 10 km (6.2 miles) is twice as far as 5 km (3.1 miles) because $\frac{10}{5} = \frac{6.2}{3.1} = 2$. Other examples of ratio scales include temperature in degrees Kelvin, weight measured in kg or pounds, and time expressed in hours, minutes and seconds.

Most nonparametric methods are designed for use with data expressed in a nominal or ordinal scale. Because each scale of measurement possesses all of the properties of a weaker measurement scale, statistical methods requiring a weaker scale can be used with stronger scales. Consequently, time series observations which are always expressed using either an interval or a ratio scale, can be subjected to nonparametric testing. Most parametric methods can only be used with values given in an interval or ratio scale and cannot be utilized to analyze data belonging to a nominal or ordinal scale. Due to the foregoing and other reasons, Conover (1980, p. 92) defines a statistical method as being *nonparametric* if it satisfies at least one of the following criteria:

1. The method can be used with data possessing a nominal scale of measurement.
2. The method can be employed with data having an ordinal scale of measurement.
3. The method may be used with data having an interval or ratio scale of measurement where the probability distribution function of the random variable generating the data is either unspecified or specified except for an infinite number of unknown parameters.

In an *environmental impact assessment study,* usually the investigators have a fairly clear idea, at least in a general sense, of what they want to accomplish. For example, they may wish to ascertain if increased industrialization has significantly lowered the water quality of a large lake in the industrialized region. If data are not already available, a major task would be to design a suitable data collection scheme (see Sections 1.2.3 and 19.7). Assuming that observations are available for a range of water quality variables at different locations in the lake, a challenging problem is to select the most appropriate set of statistical methods that can be used in an optimal fashion for detecting and modelling trends in the data. Besides uncovering and modelling trends in water quality variables at a single site, statistical methods could be used to model the relationships among variables and trends across sites in the lake. In a large scale environmental impact assessment study, it is often necessary to use a variety of both nonparametric and parametric methods (see the applications in Sections 23.5, 22.3, 22.4 and 24.3).

As discussed in the introduction to Part X and also Sections 22.1, 22.3, as well as 1.2.4, when executing a data analysis study it is recommended to carry out *exploratory data analysis* followed by confirmatory data analysis. Usually simple graphical methods are employed at the exploratory data analysis stage for visually detecting characteristics in the data such as trends and missing values (see Section 22.3). Both nonparametric and parametric tests can be employed for hypothesis and significance testing during *confirmatory data analyses*. Because of the preponderance and proliferation of statistical methods, it is not surprising that a great number of statistical textbooks have been published and a significant number of papers have been printed in journals regarding the development and application of statistical methods (see Section 1.6.3). In fact, at least two major encyclopediae on statistics are now available (Kruskal and Tanner, 1978; Kotz et al., 1988) and a number of informative handbooks (see, for instance, Sachs (1984)) and dictionaries (Kendall and Buckland, 1971) on statistics have been written.

To assist in choosing the best statistical methods to use in a given study, the techniques can be classified according to different criteria. In an introductory paper to an edited monograph on time series analysis in water resources, Hipel (1985), for example, classifies time series models according to specified criteria. For the case of nonparametric tests, Conover (1980) presents a useful chart at the start of his book for *categorizing nonparametric tests* according to the kind of sample, hypothesis being tested, and type of measurement involved (nominal, ordinal and interval). Keep in mind that a test designed for a weaker type of measurement can also be used with stronger measurements. Consequently, all of the tests listed under nominal measurements can be used with both ordinal and interval data. Furthermore, the tests given below ordinal measurements can also be used to analyze interval measurements.

In the remainder of Section 23.3, nonparametric tests which are especially useful for detecting trends in water quality time series are described in detail. More specifically, the first six nonparametric trend tests listed in Table 23.1.1 are defined in Section 23.3 and other useful nonparametric procedures are also discussed. Fortunately, these tests can be modified for handling data sets having tied values as well as censored observations. Other topics in Section 23.3 include grouping seasons in a meaningful way in a trend detection study, procedures for combining trend tests across groups of seasons and adjusting water quality for riverflows.

The reader should keep in mind that in a trend assessment study, it is often necessary to employ a wide range of statistical methods. Although this text describes many useful parametric and nonparametric methods that are frequently used by environmental and water resources engineers, sometimes it may be necessary to refer to other texts and papers for a description of other methods. Besides the book of Conover (1980), other texts on nonparametric testing include contributions by Siegel (1956), Fraser (1957), Bradley (1968), Gibbons (1971, 1976), Hollander and Wolfe (1973), Puri and Sen (1971), Kendall (1975), and Lehman (1975). Gilbert (1987) describes a range of both nonparametric and parametric methods for use in environmental pollution monitoring. Because of the great importance of nonparametric methods in water resources and environmental engineering, the American Water Resources Association published a special monograph on this topic (Hipel, 1988). In the water quality applications in Section 23.5, ways in which a variety of nonparametric and parametric tests can be used for trend detection in a complex water quality study are explained. Other general methodologies for trend assessment are listed in Table 1.6.1 and also referenced at the end of Section 23.1.

23.3.2 Nonparametric Tests for Trend Detection

Introduction

In their paper, Van Belle and Hughes (1984) categorize nonparametric tests for detecting trends into two main classes. The one class is referred to as *intrablock methods* which are procedures that compute a statistic such as Kendall's tau for each block or season and then sum these to produce a single overall statistic (Hirsch et al., 1982; Hirsch and Slack, 1984). The second set of nonparametric tests are called *aligned rank methods*. These techniques remove the block effect from each datum, sum the data over the blocks and then create a statistic from these blocks (Van Belle and Hughes, 1984). The foregoing two classes of techniques are designed for detecting monotonic trends or changes (gradual or sudden) during some specified time interval but unlike the parametric technique of intervention analysis in Chapter 19 and Section 22.4 they are not intended for exploring the hypothesis that a certain type of change has occurred at some prespecified time due to a known external intervention.

Intrablock Methods

Because the tests of Hirsch et al. (1982) and Hirsch and Slack (1984) are based upon earlier work of Mann (1945) and Kendall (1975), the initial research is described first.

Mann-Kendall Test: Mann (1945) presented a nonparametric test for randomness against time which constitutes a particular application of Kendall's test for correlation (Kendall, 1975) commonly known as the *Mann-Kendall or the Kendall t test*. Letting $x_1, x_2, \ldots, x_n$, be a sequence of measurements over time, Mann (1945) proposed to test the null hypothesis, H_0, that the data come from a population where the random variables are independent and identically distributed. The alternative hypothesis, H_1, is that the data follow a monotonic trend over time. Under H_0, the *Mann-Kendall test statistic* is

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_k) \qquad [23.3.1]$$

where

$$sgn(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

Kendall (1975) showed that S is asymptotically normally distributed and gave the mean and variance of S, for the situation where there may be ties in the x values, as

$$E[S] = 0$$

$$Var[S] = \left\{ n(n-1)(2n+5) - \sum_{j=1}^{p} t_j(t_j-1)(2t_j+5) \right\} / 18 \qquad [23.3.2]$$

where p is the number of tied groups in the data set and t_j is the number of data points in the jth tied group.

When using [23.3.1], a positive value of S indicates that there is an upward trend in which the observations increase with time. On the other hand, a negative value of S means that there is a downward trend. Because it is known that S is asymptotically normally distributed and has a mean of zero and variance given by [23.3.2], one can check whether or not an upward or downward trend is significantly different from zero. If the S is significantly different from zero, based upon the available information H_0 can be rejected at a chosen significance level and the presence of a monotonic trend, H_1, can be accepted. For a general review of how to execute a hypothesis test, the reader can refer to Section 23.2.2.

The exact distribution of S for $n \le 10$ was derived by both Mann (1945) and Kendall (1975). They showed that even for small values of n, the normality approximation is good provided one employs the standard normal variate Z given by

$$Z = \begin{cases} \dfrac{S-1}{[Var(S)]^{1/2}} & , \text{if } S > 0 \\ 0 & , \text{if } S = 0 \\ \dfrac{S+1}{[Var(S)]^{1/2}} & , \text{if } S < 0 \end{cases} \qquad [23.3.3]$$

The statistic S in [23.3.1] is a count of the number of times x_j exceeds x_k, for $j>k$, more than x_k exceeds x_j. The maximum possible value of S occurs when $x_1<x_2< \cdots <x_n$. Let this number be called D. A statistic which is closely related to S in [23.3.1] is *Kendall's tau* defined by

$$\tau = \frac{S}{D} \qquad [23.3.4]$$

where

$$D = \left[\frac{1}{2}n(n-1) - \frac{1}{2}\sum_{j=1}^{p} t_j(t_j-1)\right]^{1/2} \left[\frac{1}{2}n(n-1)\right]^{1/2}$$

When there are no ties in the data, [23.3.4] collapses to

$$\tau = \frac{S}{\frac{1}{2}n(n-1)} = \frac{S}{\binom{n}{2}} \qquad [23.3.5]$$

Due to the relationship between τ and S in [23.3.4], the distribution of τ can be easily obtained from the distribution of S. If there are no ties in the data, the algorithm of Best and Gipps (1974) can be employed to obtain the exact upper tail probabilities of Kendall's tau, or equivalently S, for $n \geq 2$.

The Mann-Kendall trend test is, in fact, a special case of the *Kendall rank correlation test.* Kendall (1975) developed the Kendall rank correlation test for ascertaining if two series are independent of one another. This test is described in Appendix A23.1.

Seasonal Mann-Kendall Test: Hirsch et al. (1982) defined a multivariate extension of the Mann-Kendall statistic in [23.3.1] for use with seasonal data, and, as noted by Van Belle and Hughes (1984), their test possesses some similarities to tests proposed by Jonckheere (1954) and Page (1963). Although the statistic of Hirsch et al. (1982), is valid for use with data where there may be missing values and also ties, assume for the present that the time series **X** consists of a complete record sampled over n years where there are m seasons per year such that **X** is given by

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

The null hypothesis, H_0, is that for each of the m seasons the n observations are independent and identically distributed while the alternative hypothesis is there is a monotonic trend. Let the matrix of ranks be denoted by

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix}$$

where the n observations for each season or column in R are ranked among themselves. Hence, the rank of x_{jg}, which is the jth data point in the gth season, is

$$R_{jg} = [n + 1 + \sum_{i=1}^{n} sgn(x_{jg} - x_{ig})]/2 \qquad [23.3.6]$$

and each column of R is a permutation of $(1,2,\ldots,n)$. The *Mann-Kendall test statistic for the gth season* is (Hirsch et al., 1982)

$$S_g = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_{jg} - x_{ig}), \quad g = 1,2,\ldots,m \qquad [23.3.7]$$

Similar to the situation of S in [23.3.1], S_g is asymptotically normally distributed where

$$E[S_g] = 0$$

$$Var[S_g] = \sigma_g^2 = n(n-1)(2n+5)/18 \qquad [23.3.8]$$

Kendall's tau statistic for the gth season is defined as

$$\tau_g = \frac{S_g}{\frac{1}{2}n(n-1)} = \frac{S_g}{\binom{n}{2}} \qquad [23.3.9]$$

Because τ_g is simply a multiple of S_g, the distribution of τ_g can be obtained from the distribution of S_g. In particular, τ_g is asymptotically normally distributed where

$$E[\tau_g] = 0$$

$$Var[\tau_g] = \frac{1}{\left[\frac{1}{2}n(n-1)\right]^2}\left[\frac{n(n-1)(2n+5)}{18}\right] \qquad [23.3.10]$$

Since it is arithmetically more convenient to deal with S_g rather than τ_g and also Hirsch et al. (1982) use mainly S_g rather than τ_g in their research, the statistic S_g is utilized in the rest of this section.

Following Hirsch et al. (1982), the *seasonal Mann-Kendall test statistic* is

$$S' = \sum_{g=1}^{m} S_g \qquad [23.3.11]$$

which is asymptotically normally distributed where

$$E[S'] = 0$$

$$Var[S'] = \sum_{g=1}^{m} \sigma_g^2 + \sum_{\substack{g,h \\ g \neq h}} \sigma_{gh} \qquad [23.3.12]$$

Using [23.3.8], $\sigma_g^2 = Var[S_g]$ can be calculated as well as $\sigma_{gh} = cov(S_g S_h)$. For the situation where each season is independent of each of the other seasons, the second summation in [23.3.12] is zero and

$$Var[S'] = \sum_{g=1}^{m} \sigma_g^2 \qquad [23.3.13]$$

As is done in [23.3.3] for the Mann-Kendall test, for $n \leq 10$ the standard normal deviate Z' should be calculated as

$$Z' = \begin{cases} \dfrac{S'-1}{[Var(S')]^{1/2}} & , \text{ if } S' > 0 \\ 0 & , \text{ if } S' = 0 \\ \dfrac{S'+1}{[Var(S')]^{1/2}} & , \text{ if } S' < 0 \end{cases} \quad [23.3.14]$$

When utilizing [23.3.14], Hirsch et al. (1982) demonstrate that the normal approximation is quite accurate even for data as small as $n = 2$ and $m = 12$.

To handle *missing values*, $sgn(x_{jg} - x_{ig})$ is defined to be zero if either x_{jg} or x_{ig} is missing. Letting n_g be the number of nonmissing observations for season g, equation [23.3.6] is modified as

$$R_{jg} = [n_g + 1 + \sum_{i=1}^{n} sgn(x_{jg} - x_{ig})]/2 \quad [23.3.15]$$

Consequently, the ranks of the known observations remain unchanged and each missing observation is assigned the average or midrank $(n_g + 1)/2$. As is the case when there are no missing observations, equation [23.3.7] is used to calculate S_g and following [23.3.8] the variance of S_g is determined using

$$\sigma_g^2 = n_g(n_g - 1)(2n_g + 5)/18 \quad [23.3.16]$$

S' and its variance are determined using [23.3.11] and [23.3.12], respectively, where each σ_{gh} is zero.

For a *censored time series*, in which some data are reported to be less than a *detection limit*, arbitrarily fix the affected data at some constant value which is less than the limit of detection. Because nonparametric tests are based upon ranks instead of magnitudes, all censored values are interpreted as sharing the same rank which is less than the rank of all uncensored observations. Additionally, this means that handling censored data is equivalent to dealing with *ties*. Assuming, for the moment, that there are no missing values, the ranked data containing ties can be calculated using [23.3.6], which automatically assigns to each of t tied values the average of the next t ranks. Following this, S_g can be determined utilizing [23.3.7] where, similar to the situation in [23.3.2], the variance is

$$Var[S_g] = \sigma_g^2 = [n(n-1)(2n+5) - \sum_{j=1}^{p} t_j(t_j - 1)(2t_j + 5)]/18 \quad [23.3.17]$$

where n is the number of years of data, p is the number of tied groups for the data x_{ig}, $i = 1,2,\ldots,n$, in season g, and t_j is the size of the jth tied group. The seasonal Mann-Kendall statistic is calculated using [23.3.7] while its variance is determined by utilizing [23.3.12] where all the σ_{gh} are zero. When there are both tied data (due to "tied" censored data and ties of actual observations) and missing values, the modifications described in this and the previous paragraph must be combined. Finally, a general description of censored data is presented in Section 23.3.8.

Another problem which can arise when using any of the nonparametric tests in this section, is how to summarize information when there are *several values for a specified season* in a given year. Van Belle and Hughes (1984, p. 135) suggest four possible approaches for accomplishing this. One method is to adjust the seasonal length so that there is not more than one observation per season. This, of course, would mean having a smaller seasonal length, such as biweekly instead of monthly. A second approach is to take a single value which is closest to the center of the season in a given year and to ignore the other information. In many applications, a third, and more reasonable method to follow, is to simply replace the set of values by the median or mean before calculating the test statistic. As a matter of fact, many water quantity and quality records which are collected on a daily basis are often released in reports as average monthly values. Fourthly, an alternative to calculating a mean or median within a given season and year, is to consider these values as tied in the time index and then to compute the test statistic along with a modified variance. Van Belle and Hughes (1984, p. 135) present the formulae for carrying this out.

Recall that for the intervention analysis applications of seasonal data in Chapter 19 and Section 22.4, it is suggested that an intervention may affect each season or groups of seasons in different manners. For example, when modelling the impacts of reservoir operation upon the average monthly flows of the South Saskatchewan River in Section 19.2.5, it is suspected that the seasonal mean levels would increase during certain months and decrease at other times. To accurately model an upward or downward step trend as well as the change in magnitude of the mean for each season, a separate intervention term is incorporated into the intervention model for each month. In a similar manner, one should examine carefully how S_g in [23.3.7] behaves for each season. Only if the same type of trend, such as an upward trend, is detected in each season, will the overall seasonal Mann-Kendall test statistic in [23.3.11] have any meaning. In other words, S' should only be calculated for a group of seasons which are expected to behave in a certain manner where hypothesis testing is done separately for this group. A more detailed discussion of this problem is presented in Section 23.3.3 where approaches are presented for *combining tests of hypotheses across seasons.*

Correlated Seasonal Mann-Kendall Test: In practice, environmental data are usually correlated and not independently distributed as is assumed in the previous section. For instance, when dealing with average monthly phosphorous levels, the phosphorous observations in one month may be significantly correlated with values in the preceding one or more months. This means that in order to employ the seasonal Mann-Kendall tests in [23.3.7] and [23.3.11], one must be able to estimate all the σ_{gh} in [23.3.12].

Based upon research by Dietz and Killeen (1981), Hirsch and Slack (1984) explain how σ_{gh} can be estimated. Assuming, for the moment, that there are no ties or missing values, a consistent estimation for σ_{gh} is

$$\hat{\sigma}_{gh} = K_{gh}/3 + (n^3 - n)r_{gh}/9 \qquad [23.3.18]$$

where

$$K_{gh} = \sum_{i=1}^{n-1}\sum_{j=1}^{n} sgn[(x_{jg} - x_{ig})(x_{jh} - x_{ih})] \qquad [23.3.19]$$

$$r_{gh} = \frac{3}{n^3 - n}\sum_{i,j,k} sgn[(x_{jg} - x_{ig})(x_{jg} - x_{kh})] \qquad [23.3.20]$$

For the situation where there are no ties and no missing values, the statistic r_{gh} is *Spearman's* *correlation coefficient* for seasons g and h (Conover, 1980; Lehman, 1975). If there are no missing observations, equation [23.3.18] can be written as

$$\hat{\sigma}_{gh} = [K_{gh} + 4\sum_{j=1}^{n} R_{jg}R_{jh} - n(n+1)^2]/3 \qquad [23.3.21]$$

where each R_{jg} is determined using [23.3.6].

To employ the correlated seasonal Mann-Kendall test when there are no ties or missing data, S_g, $g = 1,2, \cdots ,m$, and S' are determined using [23.3.7] and [23.3.11], respectively. Following the estimation of the variance of each S_g and σ_{gh} using [23.3.8] and [23.3.18], respectively, the variance of S' can be calculated using [23.3.12]. As can be done for the seasonal Mann-Kendall test, a separate hypothesis can be formulated for each season or each group of seasons, and, when appropriate, based upon S' an overall hypothesis can be made. For a given season or group of seasons, seasonal data are thought to be independently distributed since it is assumed that a particular data point is not correlated with data occurring one or more years before during the same season. Consequently, the null hypothesis, H_0, for a given season is the data are independent and identically distributed while H_1 is the existence of a monotonic trend in that season. For the overall correlated seasonal Mann-Kendall statistic, S', the null hypothesis is that the data are correlated and identically distributed while H_1 is the presence of a monotonic trend. Because the mean, variance and distribution of each S_g and also S' are known, hypothesis testing can be executed. As is the case for the seasonal Mann-Kendall test, when employing the correlated seasonal Mann-Kendall test, the standard normal deviate in [23.3.14] should be calculated for $n \le 10$.

To use the correlated seasonal Mann-Kendall test with *missing values*, the procedure described for the seasonal Mann-Kendall test is used. The ranks of the data for season g are determined using [23.3.15], and then S_g and its variance are calculated using [23.3.7] and [23.3.8], respectively. The correlated seasonal Mann-Kendall statistic is determined using [23.3.11]. To estimate the variance for S' using [23.3.12], equation [23.3.18] or equivalently [23.3.21] must be appropriately modified in order to estimate σ_{gh} for substitution into [23.3.12]. The K_{gh} term is determined as before by using [23.3.19]. However, in the presence of missing values, r_{gh} takes a new form which causes the revised version of [23.3.21] to be

$$\hat{\sigma}_{gh} = [K_{gh} + 4\sum_{j=1}^{n} R_{jg}R_{jh} - n(n_g+1)(n_h+1)]/3 \qquad [23.3.22]$$

where n_g is the number of observations for season g and n_h stands for the number of measured values for season h.

As in the previous section, *censored data* are handled as ties where data reported as being less than a limit of detection are assigned a constant value which is less than the limit of detection. Assuming that there are no missing values, the ranks containing ties can be calculated using [23.3.6], which automatically assigns to each of t tied values the average of the next t ranks. Next, S_g can be determined by utilizing [23.3.7], while the variance of S_g can be calculated using [23.3.17]. The correlated seasonal Mann-Kendall statistic, S', is determined using [23.3.11]. To employ [23.3.18] or [23.3.21] for estimating σ_{gh}, midranks are used in assigning the values of R_{jg}. Hence, if there were t_j censored values, they would all have a rank of $t_j(t_j - 1)/2$. Following this, the variance of the correlated seasonal Mann-Kendall statistic can be determined using [23.3.12]. When there are both tied data, which may be due to censored data and ties of actual observations, and missing values, the alterations outlined in this and the previous paragraph must be combined. Lettenmaier (1988) proposes a technique called the covariance eigenvalue method to handle correlation among seasons. He uses simulation experiments to compare the power of his seasonal trend test to those of Hirsch and Slack (1984) as well as Dietz and Killeen (1981). Utilizing simulation experiments, Loftis et al. (1991a,b) find that Lettenmaier's (1988) technique works well with serially correlated data. Zetterqvist (1988) also proposed a seasonal Mann-Kendall trend test to take care of autocorrelation among data in different seasons, where the observations within each season are assumed to be independent.

El-Shaarawi and Niculescu (1992) extend the Kendall Tau test for handling correlation in nonseasonal data when the underlying process is MA(1) or MA(2). Moreover, they develop a test for use with seasonal data having non-zero correlations between successive seasons and years.

Seasonal Kendall Slope Estimator: The intrablock statistics discussed thus far, are designed for detecting the presence but not the location or magnitude of a trend in a time series. To estimate the magnitude of a trend, Hirsch et al. (1982) suggest an extension to the seasonal case of the method proposed by Theil (1950) and Sen (1968). In particular, the seasonal Kendall slope estimator of Hirsch et al. (1982) expresses the magnitude of a trend as a slope which means change of the series per unit time. When sufficient data are available, the technique of intervention analysis discussed in Chapter 19 and Section 22.4, constitutes a much more powerful procedure for accurately estimating the magnitude of trends caused by one or more known interventions. As shown in Figures 19.2.3 and 19.2.4 and explained in Section 19.2.2, the intervention component contained within the overall intervention model for modelling a trend can be designed so that the geometric shape of the trend is correctly modelled. Other trend detection techniques include the exploratory data analysis graphs of Section 22.3, the change-detection statistics referred to in Sections 19.2.3 and 24.2.1, and the robust locally weighted regression smooth of Section 24.2.2.

The seasonal Kendall slope estimator is defined to be the median of the differences, expressed as slopes, of the ordered pairs of data points that are compared in the seasonal Mann-Kendall test. The computational algorithm for defining the seasonal Kendall slope estimator, Sl, is as follows. Calculate

$$d_{ijk} = (x_{ij} - x_{ik})/(j - k)$$

for all (x_{ij}, x_{ik}) pairs $i = 1, 2, \cdots, m$, where $1 \le k < j \le n_i$ and n_i is the number of known values in the ith season. The slope estimator Sl is the median of the d_{ijk} values. As noted by Hirsch et al. (1982), the slope estimator Sl is closely related to S' in [23.3.11]. If $S' > 0$ then $Sl \ge 0$ ($Sl > 0$ if one or no $d_{ijk} = 0$) and if $S' < 0$, then $Sl \le 0$ ($Sl < 0$ if one or no $d_{ijk} = 0$). The reason for these relationships between S' and Sl is that S' is equivalent to the number of positive d_{ijk} values minus the number of negative d_{ijk} values and Sl is the median of the d_{ijk} values.

As pointed out by Hirsch et al. (1982), because the median of the d_{ijk} values is used to define Sl, the estimator is quite resistant to the presence of extreme observations in the data. Further, since the slopes are always computed between values that are integer multiples of the seasonal length m, the slope estimator is unaffected by seasonality. If, for example, the magnitude of the slope were thought to be different for each season, the slope could be estimated separately for each season. In situations where groups of seasons are suspected of having the same magnitude for the slope estimator, the slope estimator could be separately applied to each group of seasons. A computer program for calculating the seasonal Mann-Kendall statistic and the seasonal Kendall slope is listed by Smith et al. (1982) and also Crawford et al. (1983).

Aligned Rank Methods

In addition to the intrablock methods, the aligned rank techniques constitute nonparametric approaches for checking for the presence of trends in a data set. However, unlike the intrablock methods which can be used with incomplete records, the aligned rank techniques are designed for use with evenly spaced observations for which there are no missing values. One particular kind of aligned rank method is described in this section.

Suppose that a time series, **X**, consists of a complete record sampled over n years where there are m seasons per year. Hence, following the notation suggested by Van Belle and Hughes (1984), the data can be displayed in the following fashion:

		Season					
		1	2	. . .	m	mean
Year	1	x_{11}	x_{12}	. . .	x_{1m}	$x_{1.}$
	2	x_{21}	x_{22}	. . .	x_{2m}	$x_{2.}$
	.	.	.		.	.
	.	.	.		.	.
	.	.	.		.	.
	n	x_{n1}	x_{n2}	. . .	x_{nm}	$x_{n.}$
	mean	$x_{.1}$	$x_{.2}$	. . .	$x_{.m}$	$x_{..}$

When a dot is used to replace a subscript, this indicates the mean taken over that subscript. Therefore,

$$x_{.j} = \sum_{i=1}^{n} x_{ij}/n$$

is the mean of the jth season and

$$x_{i.} = \sum_{j=1}^{m} x_{ij}/m$$

is the mean for the ith year.

Based upon the work of Sen (1968), Farrell (1980) proposed the following procedure to test for the presence of trends which is also described by Van Belle and Hughes (1984).

1. The data are deseasonalized by subtracting the seasonal average from each data point as in [13.2.2] in Section 13.2.2. However, because each season is suspected of containing a trend, one may question the validity of deseasonalization which assumes a constant mean for each season.

2. The nm deseasonalized data points are ranked from 1 to nm. When t values are tied, simply assign the average of the next t ranks to each of the t tied values. The matrix of ranks for the deseasonalized data can be written as:

		Season					
		1	2	. . .	m	mean
	1	R_{11}	R_{12}	. . .	R_{1m}	$R_{1.}$
	2	R_{21}	R_{22}	. . .	R_{2m}	$R_{2.}$
	.	.	.		.	.
Year	.	.	.		.	.
	.	.	.		.	.
	n	R_{n1}	R_{n2}	. . .	R_{nm}	$R_{n.}$
	mean	$R_{.1}$	$R_{.2}$	. . .	$R_{.m}$	$R_{..}$

3. The average ranks for each season and year are obtained. As is the case for the given set of data, **X**, the mean over a subscript is indicated by a dot. Accordingly,

$$R_{\cdot j} = \sum_{i=1}^{n} R_{ij}/n$$

is the average rank for the jth season while

$$R_{i.} = \sum_{j=1}^{m} R_{ij}/m$$

is the average rank for the ith year.

4. For use in hypothesis testing, the following statistic is calculated.

$$T = \left(\frac{12m^2}{n(n+1)\sum_{j=1}^{m}\sum_{i=1}^{n}(R_{ij} - R_{.j})^2} \right)^{1/2} \cdot \left(\sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) \left(R_{i.} - \frac{nm+1}{2} \right) \right) \qquad [23.3.23]$$

In reality, T is the slope of the regression of $R_{i.}$ against i for $i = 1,2,\ldots,n$, standardized by the square root of the residual error given by

$$\sum_{j=1}^{m}\sum_{i=1}^{n}(R_{ij}-R_{.j})^2/[m(n-1)]$$

The null hypothesis, H_0, is that the data are identically independently distributed and, hence, possess no trends. The alternative hypothesis, H_1, is that the data have trends. For large samples, T approaches normality with a mean of zero under H_0 and a variance of unity. Consequently, the test statistic T is approximately a standardized normal variable and, as explained in Section 23.2.2, either a one sided or two sided statistical test can be executed.

If a data set is incomplete, the *missing observations* must be estimated before the aligned rank method can be employed. Depending upon how many data points are missing, a suitable data filling procedure from Sections 22.2, 19.3 or 18.5.2, can be selected for obtaining an estimated evenly spaced time series. Farrell (1980) suggests estimating missing data using a least squares approach. Subsequent to the data filling, the aligned rank test statistic in [23.3.23] can be calculated in order to carry out an hypothesis test.

An alternative approach to estimating missing data, is to adjust the seasonal length until there is at least one observation per season for each year (Van Belle and Hughes, 1984). For example, suppose that one were examining weekly data for which there are quite a few weeks containing no data points. However, when a monthly series is considered for the same data set, there is at least one observation per month of every year. By calculating the mean value for each month within a given year, a time series of evenly spaced monthly values can be created. The aligned rank method can then be applied to this monthly series.

As noted earlier, *ties* of the deseasonalized values can easily be handled using the aligned rank method. If there are *censored values* where some data are reported to be below a limit of detection, the affected data can be arbitrarily fixed at some constant value which is less than the limit of detection. A drawback of this approach is that the estimates of the yearly and seasonal means will be biased. Therefore, it should be used only if the relative number of censored values is not too great. Nevertheless, subsequent to determining the ranks of the deseasonalized data, the statistic in [23.3.23] can be calculated.

Comparison of Intrablock and Aligned Rank Methods

As pointed out by Van Belle and Hughes (1984), both the aligned rank method and the seasonal Mann-Kendall test are based upon the same model given by

$$x_{ij}=\mu+a_i+b_j+e_{ij},\quad i=1,2,\ldots,n\text{ , and } j=1,2,\ldots,m \qquad [23.3.24]$$

where μ is the overall mean, $\mathbf{a}=\{a_1,a_2,\ldots,a_n\}$, is the yearly component, $\mathbf{b}=\{b_1,b_2,\ldots,b_m\}$, is the seasonal component and

$$\sum_{i=1}^{n}a_i=\sum_{j=1}^{m}b_j=0.$$

The e_{ij} is the noise term which is independently distributed. For both nonparametric tests, the null hypothesis is that the yearly component is zero and hence

$$H_0:\ \mathbf{a}=\mathbf{0}$$

The alternative hypothesis is

$$H_1: a_1 \le a_2 \le a_3 \quad \cdots \quad \le a_n \text{ and/or}$$

$$a_1 \ge a_2 \ge a_3 \quad \cdots \quad \ge a_n$$

with at least one strict inequality. Accordingly, both approaches are testing for a monotonic trend over the years where the trend is not necessarily linear. A trend within each year could be considered as part of the seasonal trend **b**.

Using the results of Puri and Sen (1971), Van Belle and Hughes (1984) show that the aligned rank test is always more powerful than the seasonal Mann-Kendall test and that the difference is greater for smaller numbers of years of data. As mentioned before, when there are missing data, the intrablock tests can be used without having to estimate the missing observations. However, in order to use the aligned rank method with an unevenly spaced time series, the missing observations must be estimated prior to applying the technique. Finally, using simulation experiments, Taylor and Loftis (1989) find that the correlated seasonal Mann-Kendall test is more powerful than its competitors, including an aligned rank method, for detecting trends.

23.3.3 Grouping Seasons for Trend Detection

As was pointed out in the discussion included with the seasonal Mann-Kendall test, an intervention may affect each season or groups of seasons in different ways. Based upon a physical understanding of the problem and using exploratory data analysis procedures such as the time series plots described in Section 22.3.2 and the box-and-whisker graphs of Section 22.3.3, one can decide how seasons should be grouped together. For instance, as explained in Section 19.5.4, a physical comprehension of the problem and exploratory data analyses make one suspect that a forest fire caused the spring flows of the Pipers Hole River in Newfoundland, Canada, to increase immediately after the fire and to gradually attenuate over the years back to their former levels as the forest recovered. However, during other seasons of the year the fire did not cause any trends in the time series after the fire. By employing the technique of intervention analysis, this behaviour is rigorously confirmed and accurately modelled in Section 19.5.4.

Nonparametric testing can be executed in a fashion similar to the general approach used in intervention analysis studies. In order to *classify seasons into groups* where seasons within each group possess the same kind of trend, one procedure is to rely upon a *physical understanding of the problem* and the output from *exploratory data analyses*. The *Kruskal-Wallis test* (Kruskal and Wallis, 1952) can also be used to test for the presence of seasonality and decide upon which seasons are similar (see Appendix A23.3). The statistic defined for the nonparametric test being used to detect trends can be calculated separately for each group of seasons to ascertain if a certain kind of trend is present within the group. Output from the nonparametric tests may suggest other ways in which the seasons should be grouped and then the statistics can be calculated for the new grouping of the seasons.

Consider, for example, how the seasonal Mann-Kendall test can be used when seasons are grouped according to common patterns recognized in trends. The Mann-Kendall statistic, S_g, for each season can be calculated using [23.3.7] and the variance, σ_g^2, of S_g can be determined using [23.3.8]. Suppose that one of the groups of seasons consists of seasons in the set represented by G. Then the *seasonal Mann-Kendall statistic* for the seasons in group G is calculated as

$$S_G = \sum_{g \in G} S_g \qquad [23.3.25]$$

The variance of S_G is then determined as

$$Var[S_G] = \sum_{g \in G} \sigma_g^2 \qquad [23.3.26]$$

where the expected value of each S_g and also S_G is zero. Because S_G is asymptotically normally distributed, hypothesis testing can be done to see if the seasons in group G possess a common trend. Recall that a significantly large positive value of S_G would indicate an increasing trend, while a significantly large negative value of S_G would mean there is a decreasing trend. Further, for small samples one should employ [23.3.14] to calculate the standard normal deviate where S_G replaces S' in [23.3.14]. In a similar fashion, groups of seasons could be considered when employing the correlated seasonal Mann-Kendall statistic and also the aligned rank method.

For deciding upon how common or homogeneous trends should be grouped, Van Belle and Hughes (1984) suggest employing a *homogeneity test* (Fleiss, 1981, Ch. 10) which is commonly used in the study of cross-classified data. This test is closely related to the seasonal Mann-Kendall test of Section 23.3.2. Van Belle and Hughes (1984) propose that the grouping or homogeneity test should be used as a preliminary test for checking for the homogeneity of trends and, thereby, classifying the seasons into groups where each group possesses a common trend. Subsequent to this, an intrablock statistic such as the seasonal Mann-Kendall statistic or the aligned rank statistic can be calculated for each group to ascertain if there is a significantly large common trend in the group.

For use in the homogeneity test of Van Belle and Hughes (1984), the statistic Z_g^2 is defined as

$$Z_g^2 = S_g^2/Var[S_g] \qquad [23.3.27]$$

where the Mann-Kendall statistic, S_g, for the gth season is given in [23.3.7] and its variance, $Var[S_g]$ is presented in [23.3.8]. Because Z_g is asymptotically normally distributed, Z_g^2 approximately follows a chi-squared distribution with one degree of freedom. As in [23.3.9], Kendall's tau for the gth season is related to S_g by the expression

$$\tau_g = 2S_g/\left(n_g(n_g - 1)\right) \qquad [23.3.28]$$

where n_g is the number of data points in the gth season. The null hypothesis, H_0, is there is no trend in the gth season and it can be written as H_0: $\tau_g = 0$ or, equivalently, H_0: $S_g = 0$.

Notice that because the square of S_g is used in [23.3.27], the sign of S_g is eliminated in the calculation of Z_g^2. Because a positive or negative sign for S_g indicates an increasing or decreasing trend, respectively, one should make sure that different kinds of trends are not being combined when Z_g^2 is summed across seasons. Suppose that a physical appreciation of the problem in conjunction with output from preliminary data analyses indicate that seasons in a set labelled G should be included within one group. This group, for example, may stand for the group of summer seasons where there is an increasing trend in each summer season and, therefore, an increasing trend for the entire group G. For the group G, the overall *homogeneity test statistic* is

$$\chi_G^2 = \sum_{g \in G} Z_g^2 \qquad [23.3.29]$$

which is approximately χ^2 distributed with $|G|$ degrees of freedom where $|G|$ is the number of seasons in group G. The null hypothesis, H_0, is that there is no trend across the seasons in G and, therefore, the Kendall τ or the Mann-Kendall statistic for each season in G is zero. If, at a selected level of significance, the statistic calculated using [23.3.29] is significantly different from zero, one can reject H_0 and conclude that there is an overall trend across the seasons in the group G. The statistic in [23.3.29] is, of course, separately calculated for each grouping of seasons where every season is a member of one of the groups.

The foregoing approach is one way of deciding upon how seasons should be grouped. Other approaches are given by Van Belle and Hughes (1984), Fleiss (1981) and Zar (1974).

In the general situation, one may wish to examine trends in various water quality variables across an entire river basin or other appropriate geographical entities for which there is a set of locations where data are collected. Consequently, for a given variable one must not only ascertain how seasons should be grouped at a single station but also how data can be grouped across stations. For each water quality variable, one procedure is to employ a physical understanding of the problem and exploratory data analyses executed for data collected at each site to decide upon which seasons and site locations should be included in G used for calculating a group statistic such as the Mann-Kendall statistic in [23.3.25] and the homogeneity statistic in [23.3.29]. Based upon the χ^2 statistic in [23.3.29], Van Belle and Hughes (1984) propose a method for *grouping data for a given variable across seasons and sites*.

As noted earlier, one can use a statistical test such as the nonparametric *Kruskal-Wallis test* (see Appendix A23.3) to determine whether or not a given time series contains seasonality. If the data are not seasonal, then, of course, all of the observations fall under one group. One can then employ the nonseasonal Mann-Kendall test in [23.3.1] or [23.3.5] to check for trends. However, one should not employ the seasonal Mann-Kendall trend test when seasonality is not present. This would certainly result in a loss of power in the trend test.

23.3.4 Combining Tests of Hypotheses

One may calculate a test statistic such as the Mann-Kendall statistic S_g in [23.3.7] for each season of the year when examining a seasonal time series. Alternatively, by following one of the procedures described in Section 23.3.3, one may join seasons together and calculate a separate statistic, such as S_G in [23.3.25], for each group of seasons. Whatever the case, following the determination of the test statistic and associated significance level for each season or each group of seasons, one may wish to then combine tests of hypotheses across seasons or groups of seasons. For explanation purposes, suppose that one wants to calculate a separate Mann-Kendall statistic for each of the seasons and then combine tests of hypotheses across seasons in order to arrive at an overall hypothesis test. As noted by Littell and Folks (1971), several authors have considered the problem of combining independent tests of hypotheses. Using the exact Bahadur relative efficiency (Bahadur, 1967), Littell and Folks (1971) compare four methods of combining independent tests of hypotheses. The methods they compare are Fisher's (1970) method, the mean of the normal transforms of the significance levels, the maximum significance level, and the minimum significance level. Although none of the tests is uniformly more powerful than the others, according to the Bahadur relative efficiency, Fisher's method is the most efficient of the four.

Let the observed significance level of a test of hypothesis be denoted by SL_i. For example, because the distribution of S_g in [23.3.7] is known for a given data set, one can calculate SL_g for S_g where $g = 1,2,\ldots,m$. Because of the relationship between Kendall's τ_g and S_g in [23.3.9], SL_g would be the same for both τ_g and S_g in season g. When there are m independent tests, Fisher (1970, p. 99) shows that

$$-2\sum_{i=1}^{m} lnSL_i \approx \chi^2_{2m} \qquad [23.3.30]$$

For the situation where SL_g is the observed significance level for S_g or, equivalently, τ_g in the gth season, the null hypothesis would be that the data for all of the seasons considered in the test come from a population where the random variables are independent and identically distributed. The alternative hypothesis is that the data across the seasons follow a monotonic trend over time. If, for example, the magnitude of the observed chi-squared variable calculated using [23.3.30] were larger than the tabulated χ^2_{2m} value at a chosen significance level, one would reject the null hypothesis. In [10.6.7] within Section 10.6.4, Fisher's combination method is used to demonstrate that ARMA models fitted to geophysical time series statistically preserve the Hurst coefficient and hence provide an explanation for the Hurst phenomenon.

23.3.5 Flow Adjustment of Water Quality Data

In [22.4.5] of Section 22.4.2, an intervention model is presented for describing the effects of cutting down a forest upon a seasonal water quality time series. Notice in [22.4.5] that the response variable, which represents the water quality variable under consideration, is dependent upon a number of different components written on the right hand side of [22.4.5]. Of particular interest is the fact that the riverflows are included as a covariate series in the intervention model and the manner in which the flows stochastically affect the output is modelled by the specific design of the transfer function for the riverflows. Accordingly, the influence of water quantity upon a given water quality variable is realistically and rigorously accounted for by including the flows as an input series to a water quality intervention model.

When employing a nonparametric test for checking for the presence of trends, a more accurate study can be executed if the impacts of water quantity upon water quality are properly accounted for. For a long time, scientists have known that many water quality variables are correlated with river discharge (Hirsch et al., 1982; Langbein and Dawdy, 1964; Johnson et al., 1969; Smith et al., 1982). Consider, for example, the case of total phosphorous which can have a rather complex dependence upon riverflows (Reckhow, 1978; Hobbie and Likens, 1973; Borman et al., 1974). As mentioned by Smith et al. (1982), at base flow conditions in certain watersheds, much of the phosphorous may be due to point-source loadings and, hence, a decrease in flow would cause an increase in phosphorous concentrations. Alternatively, in some river basins the occurrence of a massive rainstorm over a basin may cause the erosion and transport of organic and inorganic materials which carry large amounts of phosphorous and, therefore, the resulting increases in riverflows may be combined with increased phosphorous levels. Consequently, for a given river it is important to have a physical understanding of the type of relationship which exists between a given water quality variable and runoff. In some river basins, more than one physical process may take place where each process is a function of the quantity of riverflow. As noted by Harned et al. (1981), streamflow is the single largest source of variability in water quality data.

Hirsch et al. (1982) and Smith et al. (1982) suggest a general procedure by which the different effects of water quantity upon a water quality variable can be modelled. The purpose of their procedures is to develop a time series of *flow adjusted concentrations (FAC)* for the water quality variable under consideration which can then be tested for trends using appropriate nonparametric tests described in Section 23.3 and elsewhere. Depending upon the physical characteristics of the problem being studied, an appropriate filter can be designed to obtain the FAC series. In general, an equation for determining concentrations may have the form

$$X = f(Q) + \varepsilon \qquad [23.3.31]$$

where Q is the flow, $f(Q)$ gives the functional relationship of the flows upon the water quality variable under consideration and also contains the model parameters, ε is the noise, and X represents the concentration. For the situation where increased flows causes dilution of the water quality variable, $f(Q)$ may have one of the following forms (Hirsch et al., 1982):

$$f(Q) = \lambda_1 + \frac{\lambda_2}{Q}$$

$$f(Q) = \lambda_1 + \frac{\lambda_2}{1 + \lambda_3 Q}$$

where λ_i is the ith parameter. If increased precipitation and hence runoff increase the concentration of a water quality variable, it may be reasonable to model $f(Q)$ as

$$f(Q) = \lambda_1 + \lambda_2 Q + \lambda_3 Q^2$$

As explained by Hirsch et al. (1982) and Smith et al. (1982), regression analysis can be employed to determine which form of $f(Q)$ is most appropriate to use. Given that a significant relationship can be found using regression analysis, the FAC for year i and season j is calculated as

$$w_{ij} = x_{ij} - \hat{x}_{ij}$$

where w_{ij} is the estimated FAC, x_{ij} is the observed concentration and $\hat{x}_{ij}$ is the estimated concentration which is determined using linear regression with the best form of $f(Q)$ in [23.3.31]. The FAC series can then be subjected to nonparametric tests in order to check for trends.

When obtaining the FAC series, one is in fact using a parametric procedure to properly filter the original observations for use with a nonparametric test. An advantage of this approach is that it can be used with unevenly spaced time series. A drawback is that in regression analysis the noise term is assumed to be white. If sufficient data are available so that an estimated evenly spaced series can be obtained, the technique of intervention analysis constitutes a single parametric approach which is a much more flexible and powerful procedure for modelling water quality series. Note only does the intervention model account for the stochastic effects of flows upon the water quality variable but it also rigorously models the forms and magnitudes of trends caused by known interventions. Indeed, if it is suspected that the manner in which flows affect the water quality variable depends upon the season of the year, appropriate dynamic components can be designed to model this behaviour. Further, as explained for the intervention model in [22.4.5] and elsewhere in Chapter 19, any number of input series and trends can be modelled and the noise can be described by a correlated process such as an ARMA model.

In Section 24.3.2 a general methodology is presented for analyzing trends in water quality series measured in rivers. To remove the effects of riverflow upon a given water quality variable, the robust locally weighted regression smooth described in Section 24.2.2 is employed as one of the steps in the overall procedure. Subsequently, the Spearman partial rank correlation trend test described in Section 23.3.6 and other appropriate trend tests are employed for formally testing for the presence of trends in the water quality series. Other approaches for compensating for discharge when evaluating trends in water quality one provided by Harned et al. (1981). Bodo and Unny (1983) explain how stratified sampling can improve the estimation of load-discharge relationships.

23.3.6 Partial Rank Correlation Tests

Introduction

The previous subsection deals with the problem of adjusting water quality data in a river for the impacts of flow before checking for the presence of a trend in the water quality data. To eliminate, hopefully, the effects of flow, various regression models can be used, as described in Section 23.3.5 and also Section 24.3.2 in the next chapter.

The removal of flow effects from a water quality time series when testing for a trend, is part of a more general statistical problem. More specifically, when studying the dependence between two variables X and Y, one may wish to know if the correlation between X and Y is caused by the correlation of both X and Y with a third variable Z. For instance, one may want to find out if a possible trend in a water quality variable, as manifested by the correlation of the water quality variable over time, is independent of riverflows. Hence, one would like to remove or *partial out* the influence of water quantity when testing for a trend in the water quality variable over time. Another example for which eliminating certain effects is desirable, is when one wishes to check for trend in a seasonal water quality variable against time when the seasonality has been partialled out.

The objective of this section is to present a nonparametric trend test in which undesirable effects can be removed. In particular, the Spearman partial rank correlation test (McLeod et al., 1991) is suggested as a useful trend test for employment in environmental engineering. Because this test utilizes some definitions used in the Spearman's rho test, this latter test is first described. Additionally, the Spearman partial rank correlation test is compared to the seasonal Mann-Kendall test of Section 23.3.2 as well as the Kendall partial rank correlation coefficient (Kendall, 1975). Applications of the Spearman partial rank correlation test to a seasonal water quality time series are given in Section 24.3.2.

Spearman's Rho Test

In 1904, Spearman introduced a nonparametric coefficient of rank correlation denoted as ρ_{XY} which is based upon the squared differences of ranks between two variables. Spearman's rho can be employed as a nonparametric test to check whether or not there is significant correlation between two variables X and Y.

Let the sample consist of a bivariate sample (x_i, y_i) for $i = 1, 2, \ldots, n$, where n is the sample size. Suppose that the values of the X variable are ranked from smallest to largest such that the rank of the smallest value is one and that of the largest value is n. Let $R_i^{(X)}$ represent the rank of the X variable measured at time i. Likewise, the values of the Y variable can be ranked and $R_i^{(Y)}$ can represent the value of the rank for the Y variable at time i. The sum of the squared differences of the ranks is

$$S(d^2) = D^2 = \sum_{i=1}^{n} (R_i^{(X)} - R_i^{(Y)})^2 \qquad [23.3.32]$$

Spearman's rho is then defined for the case where there are no ties in X and Y as

$$\rho_{XY} = 1 - \frac{6S(d^2)}{n^3 - n} \qquad [23.3.33]$$

When the two rankings for X and Y are identical, then $\rho_{XY} = 1$ whereas $\rho_{XY} = -1$ when the rankings of X and Y are in reverse order.

If some values of X or Y are tied, these values are simply assigned the average of the ranks to which they would have been assigned. Let p be the number of tied groups in the X data set where t_j is the number of data points in the jth tied group. Likewise, let q be the number of tied groups in the Y sequence where u_j is the number of observations in the jth tied group. Then the formula for calculating ρ_{xy} when there are ties in either or both time series is (Kendall, 1975, p. 38, Equation 3.8)

$$\rho_{XY} = \frac{\frac{1}{6}(n^3 - n) - S(d^2) - \frac{1}{12}\sum_{j=1}^{p}(t_j^3 - t_j) - \frac{1}{12}\sum_{j=1}^{q}(u_j^3 - u_j)}{\left[\left\{\frac{1}{6}(n^3 - n) - \frac{1}{6}\sum_{j=1}^{p}(t_j^3 - t_j)\right\}\left\{\frac{1}{6}(n^3 - n) - \frac{1}{6}\sum_{j=1}^{q}(u_j^3 - u_j)\right\}\right]^{1/2}} \qquad [23.3.34]$$

When using ρ_{XY} in a statistical test to check for the absence or presence of correlation, the null hypothesis, H_0, is that there is no correlation, For large samples, ρ_{XY} is distributed as $N\left(0, \frac{1}{n} - 1\right)$ where n is the sample size. The alternative hypothesis, H_1, is that there is correlation between the X and Y variables.

By letting one of the variables represent time, Spearman's rho test can be interpreted as a *trend test*. In particular, replace (x_i, y_i) by (t, x_t) for which $t = 1, 2, \ldots, n$, and x_t consists of $x_1, x_2, \ldots, x_n$. Equations [23.3.32] to [23.3.34] can then be employed to calculate a statistic for use in a trend test. If, for example, the estimated value of ρ_{XY} is significantly different from zero, then one can argue that time and the X variable are significantly correlated, which in turn means there is a trend.

Spearman Partial Rank Correlation Test

When examining dependence between two variables X and Y, the question arises as to whether or not the correlation between X and Y is due to the correlation of each variable with a third variable Z. For example, one may wish to ascertain if an apparent trend in a water quality variable, as reflected by the correlation of the water quality variable over time, is independent of seasonality. Therefore, one would like to eliminate or *partial out* the effects of seasonality when testing for a trend in the water quality variable over time.

The purpose of the Spearman partial rank correlation test presented in this section is to determine the correlation between variables X and Y after the effects of Z upon X and Y separately are taken into account and are, therefore, removed. The notation used to represent this type of partial correlation is $corr(XY/Z)$ or $\rho_{XY.Z}$.

Let the sample consist of a trivariate sample (x_i,y_i,z_i) for $i = 1,2,\ldots,n$, where n is the sample size. As is also done for the Spearman's rho test, suppose that the values of the X variable are ranked from smallest to largest such that the rank of the smallest value is one and that of the largest value is n. Let $R_i^{(X)}$ represent the rank of the X variable at time i. Likewise, the values of the Y and Z variables can be ranked separately to produce $R_i^{(Y)}$ and $R_i^{(Z)}$, respectively.

The test statistic for the Spearman partial rank correlation test is calculated using

$$\rho_{XY.Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{(1-\rho_{XZ}^2)^{1/2}(1-\rho_{YZ}^2)^{1/2}} \qquad [23.3.35]$$

Each rho term on the right hand side of the above equation is calculated using [23.3.33] when there are no ties in any variable or [23.3.34] where there are ties.

Under the null hypothesis, there is no correlation between X and Y when the effects of Z are partialled out. To test the null hypothesis that $E(\rho_{XY.Z}) = 0$, one calculates

$$t = \frac{(n-2)^{1/2}\rho_{XY.Z}}{(1-\rho_{XY.Z}^2)^{1/2}} \qquad [23.3.36]$$

where $\rho_{XY.Z}$ is defined in [23.3.35]. Under H_0, t follows a student t distribution on $(n-2)$ degrees of freedom (Pitman's approximation), which is the same as ρ_{XY} in [23.3.33] and [23.3.34]. The alternative hypothesis, H_1, is there is correlation between X and Y, after accounting for the influence of Z separately upon X and Y. If, for example, the SL for the test statistic were calculated to be very small and less than 0.05, one could reject the null hypothesis that X and Y are not correlated when Z is partialled out. When the X, Y and Z variables represent a given water quality time series, time and seasonality, respectively, then a significantly large value of the Spearman partial rank correlation coefficient means that there is a trend in the series over time when seasonality is removed. Besides checking for trend after removing seasonality, the Spearman partial rank correlation test can be used for other purposes. For example, it can be extended to take into account correlation when testing for the presence of a trend.

The partial Spearman correlation test can be used with data for which there are missing values, ties and one level of censoring, either on the left or right. If the data are multiply censored, one can use the expected rank vector approach of Hughes and Millard (1988) before applying the test, which is discussed in Section 23.3.8. Some theoretical developments and simulation experiments for the partial Spearman correlation test are reported by Valz (1990).

Comparison to the Seasonal Mann-Kendall Test

Besides the Spearman partial rank correlation test of this section, recall that the seasonal Mann-Kendall test of Section 23.3.2 can also be used to check for trend in a seasonal time series. However, the Spearman partial rank correlation test has advantages over the seasonal Mann-Kendall test including:

1. simulation experiments demonstrate that it has more power,
2. it is more flexible and can be extended, for example, to take into account correlation,
3. and it provides an estimate of the magnitude of the trend through the coefficient in [23.3.35]. For example, when one is using the Spearman partial rank correlation coefficient, to check for a trend in a variable X over time when seasonality is partialled out (see explanation in last subsection), larger positive and negative values indicate bigger upward and downward trends over time, respectively.

Kendall Partial Rank Correlation Coefficient

Another coefficient for determining the correlation between two variables X and Y when the effects of Z upon each of these variables is taken into account is the Kendall partial rank correlation coefficient (Kendall, 1975, Ch. 8). However, because of the way the statistic is defined, it possesses some serious theoretical drawbacks. The end result is the distribution of the statistic is not known so that it cannot be used in hypothesis testing (Valz, 1990).

Before describing in more detail specific disadvantages, some notation is required. Let τ_{XY}, τ_{ZY} and τ_{XZ} represent the Kendall rank correlation coefficient between X and Y, Z and Y and X and Z, respectively. Also, let $\tau_{XY.Z}$ be the Kendall partial rank correlation statistic to determine the rank correlation between X and Y, taking into account the effects of Z upon these variables. Finally, let $\rho'_{XY.Z}$ be the Pearson partial rank correlation coefficient. For equations defining these statistics, the reader can refer to Kendall (1975).

The disadvantages of using the Kendall partial rank correlation coefficient include:

1. If the X, Y and Z variables are multivariate normally distributed and the Pearson partial rank correlation coefficient $\rho'_{XY.Z} = 0$, it can be shown using simulation and also from the relationship

$$E(\tau_{XY.Z}) = \frac{2}{\pi}\sin^{-1}\rho'_{XY.Z} \qquad [23.3.37]$$

that $\tau_{XY.Z}$ is different from zero. This does not occur with the Spearman partial rank correlation coefficient.

2. The variance of the estimate for $\tau_{XY.Z}$ depends on τ_{XY}, τ_{ZY} and τ_{XZ}. The analogous undesirable property does not hold for the Pearson or Spearman partial rank correlation coefficient.
3. On intuitive grounds, it appears that the Kendall partial rank correlation coefficient does not eliminate in a linear way the effects of Z.

Because of the foregoing disadvantages, the Kendall partial rank correlation coefficient is not used in applications by the authors.

23.3.7 Nonparametric Test for Step Trends

The statistical tests described in Sections 23.3.2 to 23.3.4 as well as 23.3.6 are designed solely for discovering trends in a data set. These techniques do not take into account when trends may have started due to external interventions. In this section, the nonparametric test of Hirsch and Gilroy (1985) and Crawford et al. (1983) is described for ascertaining if a known intervention causes a significantly large step trend for series measured at multiple stations. This test is closely related to the Mann-Whitney rank-sum test. As noted by Hirsch and Gilroy (1985), their test does not depend on parametric model assumptions, does not require complete data sets and is resistant to the effects of outliers. Nevertheless, if an evenly spaced data set can be estimated from incomplete records, the technique of intervention analysis could be used for accurately modelling trends. Recall from Chapter 19 and Section 22.4 that intervention analysis can be employed to estimate the exact magnitude of a step trend or, for that matter, many other types of trend created by known interventions.

In their paper, Hirsch and Gilroy (1985) examine the detectability of step trends in the monthly rate of atmospheric deposition of sulphate. A downward step trend in the sulphate levels would be due to a sulphate emission control program which came into effect at a known date. Prior to employing the statistical test, it is recommended that the original data set be appropriately filtered to remove unwanted sources of variation. When dealing with a water quality variable such as phosphorous, the technique described in Section 23.3.4 is a procedure for filtering the phosphorous data so the effects of water quantity upon water quality are taken into account. For the case of removing unwanted variation from a time series describing the rate of atmospheric deposition of sulphate, Hirsch and Gilroy (1985) present a specific type of filtering to remove the portion of the variance in sulphate loading rates which is due to the variance in precipitation rates and also to the variance in the seasonally varying mean values. In particular, the filtered sulphate loading series consists of the residuals obtained from the regression of logarithmic sulphate loadings on the logarithmic precipitation series.

Hirsch and Gilroy (1985) apply the nonparametric test for step trends to filtered sulphate loading series available at n_s measuring stations. The nonparametric test is the Mann-Whitney rank-sum test on grouped data which is described by Bradley (1968, p. 105). To apply the test to the same physical variable measured across seasons at n_s sites, the following steps are adhered to:

1. If it is not advisable to test the original series, obtain a filtered series by utilizing an appropriate filter. As just noted, for the case of sulphate loading series, one may wish to employ the filter presented by Hirsch and Gilroy (1985) while one may wish to use one of the filters described in Section 23.3.5 when dealing with certain kinds of water quality data.

2. Calculate the Mann-Whitney rank-sum statistic and other related statistics for data which are grouped according to season and station. In particular, by letting the subscripts i, j, and k represent the year, season and station, respectively, for group jk (season j and station k), the *Mann-Whitney rank-sum statistic* is

$$W_{jk} = \sum_{i=1}^{n_1} R_{ijk} \qquad [23.3.38]$$

where n_1 and n_2 are the number of years before and after the known intervention, respectively, for which data are collected, and R_{ijk} is the rank over the entire n years of the filtered data, X_{ijk}, $i = 1,2, \cdots ,n = n_1 + n_2$, which are ranked for the jth season and kth station. The statistic in [23.3.38] is determined for every season at each of the stations. If one assumes the null hypothesis that there is no trend in group jk, the statistic W_{jk} has the expectation of

$$\mu = n_1(n_1 + n_2 + 1)/2 \qquad [23.3.39]$$

and variance of

$$\sigma^2 = n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)/m \qquad [23.3.40]$$

where m is the number of seasons. Note that the previous three equations assume that all groups have the same record length. However, the test described in this section can be used with data sets where the numbers of data points are different across groups. Although extra notation could be used to allow for varying record lengths across the groups, for simplicity of explanation the foregoing and upcoming equations are explained for the situation where each group has the same record length.

If the data are independent, the mean and variance of the sum of the Mann-Whitney rank-sum statistics across all the groups can be easily determined. In particular, the mean and variance of $\sum_{j=1}^{m} \sum_{k=1}^{n_s} W_{jk}$ are given by $m \cdot n_s \cdot \mu$ and $m \cdot n_s \cdot \sigma^2$, respectively.

The variance in [23.3.40] is based upon the assumption that the data and, hence, the W_{jk} are independently distributed. Hirsch and Gilroy (1985) describe the following approach for estimating σ_{jk}^2 when the data are correlated. The covariance between W_{jk} and W_{gh} is given by

$$C(W_{jk}, W_{gh}) = \sigma^2 \rho(X_{ijk}, X_{igh}) \qquad [23.3.41]$$

where $\rho(X_{ijk}, X_{igh})$ is the rank correlation between data in season j station k and data in season g station h. The variance of the sum of the W_{jk}'s is determined using

$$Var\left[\sum_{j=1}^{m} \sum_{k=1}^{n_s} W_{jk}\right] = \sum_{j=1}^{m} \sum_{g=1}^{m} \sum_{k=1}^{n_s} \sum_{h=1}^{n_s} C(W_{jk}, W_{gh}) \qquad [23.3.42]$$

By assuming that the serial correlation of the ranks is lag one autoregressive and the same correlation coefficient can be used at all the stations, the estimation of the covariances can be greatly simplified. Based upon these assumptions, the estimated correlation coefficient, r_1, can be easily calculated. All of the ranks, R_{ijk}, except the last one, R_{nmk}, for each station $k = 1,2, \ldots, n_s$, are paired with the rank, $R_{i,j+1,k}$, of the succeeding filtered observations except when $j = m$ it is $R_{i+1,1,k}$. The product moment correlation coefficient of all of the pairs determines r_1. The covariances $C(W_{jk}, W_{gk})$ between different seasons at the same station are estimated using

$$\hat{C}(W_{ik},W_{gk}) = \sigma^2 r_1 |g-i|$$

The covariances $C(W_{jk},W_{jh})$, $k \neq h$, for different stations and the same season are estimated as

$$\hat{C}(W_{jk},W_{jh}) = \sigma^2 r_0(k,h)$$

where $r_0(k,h)$ is the product moment correlation coefficient of the concurrent ranks

$$(R_{ijk},R_{ijh}),\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,m$$

for stations k and h. For different stations and different seasons, the covariances $C(W_{jk},W_{gh})$, $j \neq g$, $k \neq h$, are estimated using

$$\hat{C}(W_{jk},W_{gh}) = \sigma^2 r_0(k,h) r_1 |g-i|$$

When considering the same season and station, the covariance $C(W_{jk},W_{jk})$ is simply the variance σ^2 given in [23.3.40]. Based upon the foregoing, to estimate the variance of the sum of the W_{jk} statistics in [23.3.42], the following expression is utilized

$$Var\left[\sum_{j=1}^{m}\sum_{k=1}^{n_s} W_{jk}\right] = \sum_{j=1}^{m}\sum_{g=1}^{m}\sum_{k=1}^{n_s}\sum_{h=1}^{n_s}\hat{C}(W_{jk},W_{gh}) \qquad [23.3.43]$$

3. Perform a hypothesis test to ascertain if there is a significantly large step trend in the time series due to a known intervention. The null hypothesis, H_0, is that there is no step trend while the alternative hypothesis is that there is a step trend. This test could be restricted to a certain group of seasons of the year if it were expected that the intervention only affected the seasons within that group. For example, a pollution spill may only influence certain physical variables when the temperature is above a certain level and, therefore, the data from the winter months may be excluded from the group. For the purpose of the test described here, it is assumed that a step trend may be formed for data in each season across all of the stations due to a single known intervention.

   The test statistic for checking the validity of the null hypothesis is

$$Z' = \left(\sum_{j=1}^{m}\sum_{k=1}^{n_s} W_{jk} - \mu_{jk}\right)/\sigma_{jk} \qquad [23.3.44]$$

where μ_{jk} and σ_{jk}^2 are the mean and variance, respectively, of

$$\sum_{j=1}^{m}\sum_{k=1}^{n_s} W_{jk}$$

For the situation where the filtered series and hence the W_{jk} are independent, the mean and variance are given by $m \cdot n_s \cdot \mu$ and $m \cdot n_s \cdot \sigma^2$, respectively, where μ and σ^2 are presented in [23.3.39] and [23.3.40], respectively. When the data are correlated, the mean is still given as $m \cdot n_s \cdot \mu$ but the variance is calculated using [23.3.43]. Because Z' is asymptotically normally distributed one can compare the estimated value of Z' to the value of a standard normal distribution at a selected significance level. If, for example, the estimated value of Z' were significantly different from zero, based upon the available evidence one could reject H_0 and thereby conclude that there is a significant step trend in the data. Because the W_{jk} in [23.3.38] are calculated for the ranks before the intervention, a negative value of Z' would indicate an upward step trend after the intervention whereas a positive value of Z' would indicate a downward step trend after the intervention. To demonstrate the efficacy of the aforesaid test for detecting a step trend caused by a known intervention, Hirsch and Gilroy (1985) perform simulation studies. For the situation where the data are independently distributed, Crawford et al. (1983) present a computer program to calculate the test statistic. Research on comparing statistical methods for estimating step trends and their use in sampling design is presented by Hirsch (1988).

23.3.8 Multiple Censored Data

Introduction

As noted in Section 23.3.2, often water quality data are reported as being less than a *detection level*. These observations are referred to as *censored data*. If a single limit of detection is used for a specified time series, the data is said to be *singly censored*. When there is more than one detection limit, the observations are *multiple censored.*

To apply a nonparametric test to singly censored data, the version of the test modified for use with ties can be employed. In Section 23.3.2, for instance, it is explained how the seasonal Mann-Kendall trend test can be applied to a time series with one detection limit by simply treating the censored observations as being tied. If, however, the detection limits vary within a time series and, therefore, the data are multiple censored, then one should follow other approaches in order to employ nonparametric tests.

Because multiple detection levels occur frequently in practice, the purpose of this section is to put this problem into perspective and point out procedures for handling multiple censored observations so that one can apply a given nonparametric test to the data set. In this way, practitioners will be able to make the most efficient use of the data available to them for estimating test statistics or parameters, even though the observations may possess the undesirable property of being multiple censored.

There are a variety of reasons as to why water quality and other kinds of data have multiple detection levels. As noted by authors such as Millard and Deverel (1988) and Helsel and Cohn (1988), these include:

1. The detection level changes because different methods are used to measure water quality samples at various time periods, either in the field or in the laboratory. For example, over time analytical methods may improve so that the detection levels are lowered.
2. To reduce costs, management may at different points in time request the use of cheaper measurement techniques which have higher detection levels.
3. A range of methods may be available for measuring a given water quality variable at any given time. However, each technique may have a range of the concentration of the variable for which it can provide the optimal measurement. Hence, each method has a different detection level.

4. Multiple detection levels can be caused by the process of dilution. For example, because of time constraints, a laboratory technician may adhere to a procedure whereby he can only have a specified maximum number of dilutions for any single sample. Since the detection limit is dependent upon the amount of dilution, this procedure may create multiple detection limits.

5. When data are sent to several agencies or laboratories for analyses, these organizations may have different reporting levels. Often environmental bodies such as the Environmental Protection Agency in the United States and Environment Canada are obligated to send their samples to many different private and government laboratories for analyses in order to treat everyone fairly. However, this may result in having multiple censored data.

As mentioned by Millard and Deverel (1988) as well as other authors in the field of water resources, an impressive array of techniques for handling censored data was originally developed within the areas of *survival analysis and life testing* (see, for example, Kalbfleisch and Prentice (1980)). Consequently, the basic censoring definitions and methods developed in these areas are outlined and then the censoring techniques that are suitable for use with water quality and other types of environmental data are pointed out in the next section. An attractive procedure to use with multiple censored data is the expected rank vector method first suggested for use in environmental engineering by Hughes and Millard (1988).

Censoring Definitions in Survival Analysis

Before defining censoring, first consider the meaning of *truncation*. A sample of data is said to be *truncated on the left* if only observations above a specified truncation point are reported. Likewise, a data set is *truncated on the right* when only measurements below a given truncation level are used. If, for example, a phosphorous sample is left truncated at 5 mg/l, then only the measurements that are greater than 5 mg/l would be reported.

A sample consisting of n observations is *singly censored on the left* if n_c of these measurements, where $n_c \geq 1$, are known only to fall below a *censoring level* c. The remaining $(n-n_c)$ uncensored observations would thus lie above the censoring or detection level and would be fully reported. A sample of n measurements is *multiple censored on the left* with m censoring levels if n_{c1}, $n_{c2}, \ldots,$ and n_{cm} observations are censored on the left at levels $c_1, c_2, \ldots,$ and c_m, respectively.

In a similar fashion, one can also define singly or multiple censored observations on the right. For instance, a sample of n observations is *singly censored on the right* if $n_{c'}$ of these observations are known only to fall above a specified censoring level c' while the remaining $(n-n_{c'})$ observations are reported exactly.

One can further characterize censoring according to type I and type II censoring. A singly censored sample of n measurements constitutes *type I censoring on the left* if a given censoring level c_1 is specified in advance and values below c_1 are only reported as less than c_1. Likewise, a singly censored sample of n observations arises from *type I censoring on the right* when a specified censoring level is fixed in advance and observations lying above c_1 are simply reported as being greater than c_1.

When there is *type II censoring on the left*, only the r largest observations of a sample of size n, where $1 \le r < n$, are reported, and the remaining $(n - r)$ measurements are known to lie below the rth largest values.

For *type II censoring on the right*, only the r smallest measurements of a sample of size n, where $1 \le r < n$, are reported, while the remaining $(n-r)$ observations are known to lie above the rth smallest value. As an example of type II censoring, consider a situation where one is determining the failure times of n electronic components which are started at the same time. This experiment is stopped under type II censoring after r of the components have failed.

When dealing with environmental data, only some of the definitions developed in survival analysis and life testing are required for practical purposes. In particular, *environmental time series having detection limits almost always fall under the category of type I left censoring for either single or multiple censoring*. Within the field of survival analysis, usually right censored data are encountered. Fortunately, many statistical techniques developed for use with right censored data can be converted for use with data censored on the left. The reader may wish to refer to texts by authors such as Kalbfleisch and Prentice (1980), Lee (1980) and Miller (1981) for a description of statistical censoring techniques used in survival analysis and life testing.

Multiple Censoring in Environmental Engineering

In the area of environmental research, work has been carried out for estimating parameters when the data sets are singly censored (Kushner, 1976; Owen and DeRouen, 1980; Gilbert and Kennison, 1981; Gilliom et al., 1984; Gleit, 1985; Gilliom and Helsel, 1986; Gilliom and Helsel, 1986; El Shaarawi, 1989; Porter and Ward, 1991). For the case of the seasonal Mann-Kendall trend test of Section 23.3.2, Gilliom et al. (1984) demonstrate the effects of censoring with one detection limit upon the power of the test.

Although less research has been carried out in the environmental area for handling multiple censored data, some valuable contributions have been made. Helsel and Cohn (1988) use Monte Carlo methods to compare eight procedures for estimating descriptive statistics when the data are multiple censored. They show that the adjusted maximum likelihood technique (Cohn, 1988) and the plotting position method (Hirsch and Stedinger, 1987) perform substantially better then what are called simple substitution methods. Millard and Deverel (1988) discuss nonparametric tests for comparing medians from two samples, explain how multiple censored data can be handled when using these tests, and then employ Monte Carlo studies to compare the tests.

An innovative approach to extend the nonseasonal and seasonal Mann-Kendall trend tests of Section 23.3.2 for use with multiple censored data is the method proposed by Hughes and Millard (1988) which is referred to as a *tau-like test for trend in the presence of multiple censoring points*. The first step is to assign an average rank for each observation and thereby obtain a rank vector, by taking into account all permissible combinations of ranks in the presence of multiple censoring. Second, after the expected ranks are obtained for each observation in order to get the overall *expected rank vector*, a standard linear rank test can be applied to the expected rank vector. Hughes and Millard (1988) show in detail how this approach is carried out with the Mann-Kendall trend test.

To explain how the procedure of Hughes and Millard (1988) works in practice, consider the situation presented below in Table 23.3.2 where measurements are available at four points in time. The symbol X^- indicates a left censored observation at detection level X. For this simple example, notice that there are the two detection levels: 10 and 4. Below the vector of observations are the possible three rank vectors that could occur for the data set. In each rank vector, the observations are ranked from 1 to 4, where the smallest value is assigned a 1 and the largest observation a 4. Notice that the observation at time 4 is always the largest value and, therefore, is assigned a rank of 4 in each of the three rank vectors. However, depending upon how far below a detection level the unknown actual observation may fall, one can obtain the possible rankings as shown in the table. The expected rank vector listed in the last row of Table 23.3.1 simply gives the average rank across the three possible rank vectors at each point in time.

Table 23.3.1. Hypothetical example for calculating the expected rank vector.

	Time t			
	1	2	3	4
Observation X_t	10^-	9	4^-	18
Possible Rank Vectors	3 2 1	2 3 3	1 1 2	4 4 4
Expected Rank Vector	2	2.7	1.3	4

To apply a nonparametric test to data having multiple censoring levels, one simply calculates the test statistic or parameters using the expected rank vector. As pointed out by Hughes and Millard (1988), the expected rank vector method furnishes the justification for employing the commonly accepted technique of splitting ranks when there are tied data (see Section 23.3.2). However, the conditional test statistic calculated using expected ranks does not have the same null distribution as in the case where there are no ties. In particular, the variance of the test statistic is smaller when ties are present (Lehmann, 1975). Consequently, a variance correction is usually required for test statistics when dealing with multiple censored data and data having ties.

Hughes and Millard (1988) present formulae for calculating expected rank vectors when there are two or more censoring levels. Additionally, they explain how to calculate the statistic required in the Mann-Kendall trend test of Section 23.3.2 and how to determine the expected value and variance of the test statistic. More specifically, for the case of nonseasonal data, one can use [23.3.1] or [23.3.4] to calculate the Mann-Kendall test statistic S or τ, respectively, using the expected rank vector. As would be expected, the Mann-Kendall test statistic is asymptotically normally distributed. Assuming that the method of censoring is independent of time, one can calculate the expected value and variance of S or τ. The expected value of S when there is multiple censoring is a function of the true value of τ, the sample size n and pattern of censoring. This expected value is determined as

$$E(S) = \tau[n(n-1)/2 - \sum_{j=1}^{p} t_j(t_j-1)/2] \qquad [23.3.45]$$

where p is the number of tied groups in the data set and t_j is the number of data points in the jth tied group. Usually, p is the same as the number of censoring levels while the number of observations t_j in the jth tied group is the same as the corresponding number of censored observations. One can employ [23.3.2] to determine the variance of S. By computing expected ranks separately within each season, one can utilize the expected rank vector method with the seasonal Mann-Kendall test.

In some applications, the censoring levels may not be independent of time. For example, censoring levels may decrease over time due to better laboratory methods. Hughes and Millard (1988) explain how simulation can be used to determine the approximate distribution of the test statistic for this situation. Further research is still required in order to obtain theoretical results.

23.4 POWER COMPARISONS OF PARAMETRIC AND NONPARAMETRIC TREND TESTS

23.4.1 Introduction

The objective of this section is to employ *Monte Carlo experiments* to compare the powers of a specific parametric and nonparametric test for detecting trends. In particular, the ACF at lag one given in [2.5.4] and Kendall's tau in [23.3.5] constitute the parametric and nonparametric tests, respectively, which are utilized in the simulation studies. Following a brief review of these two statistics in the next two subsections, the six models that are used for generating data containing trends are described. In Section 23.4.5, the abilities in terms of power of the ACF at lag one and Kendall's tau for detecting trends are rigorously compared. Simulation experiments demonstrate that the ACF at lag one is more powerful than Kendall's tau for discovering purely stochastic trends. On the other hand, Kendall's tau is more powerful when deterministic trends are present. The results of these experiments were originally presented by Hipel et al. (1986).

23.4.2 Autocorrelation Function at Lag One

Although the ACF test at lag one could perhaps be considered to be a nonparametric test, it could also be thought of as a parametric test since according to the Yule-Walker equations in [3.2.12] the ACF at lag one is the same as the AR parameter in an AR(1) process. Nevertheless, it is presented here, because, like the nonparametric tests described in Section 23.3, it is only used for discovering the presence of trends. Unlike the intervention model, for example, the ACF test is not designed for modelling the shapes and magnitudes of trends caused by known interventions.

The theoretical definition for the ACF at lag k is given in [2.5.4] while the formula, r_k, for estimating the ACF at lag k is presented in [2.5.9]. In Section 22.3.6, the ACF at lag k is suggested as an exploratory data analysis tool and the statistical properties of r_k are discussed. Of particular interest in this section is the *ACF at lag one*, denoted by r_1, which can be used for significance testing in trend detection at the confirmatory data analysis stage. The ACF at lag one is often referred to as the *serial correlation coefficient at lag one* or the *first serial correlation coefficient.*

The estimate for the *ACF at lag k* for an evenly spaced annual series, x_t, $t = 1, 2, \ldots, n$, can be calculated using [2.5.9] (Jenkins and Watts, 1968) as

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \quad k > 0 \qquad [23.4.1]$$

where $\bar{x}$ is the estimated mean of the x_t series. When $k = 1$ in [23.4.1], one obtains the value of r_1. For samples as small as 10, Cox (1966) observed that r_1 has an approximate normal null distribution for both normal and some nonnormal parent distributions. Knoke (1977) found empirically that the normal distribution provides an adequate approximation for determining the critical regions (the subset of the sample space for which the null hypothesis is rejected if the data fall there). The asymptotic distribution of r_1 was established by Wald and Wolfowitz (1943) and Noether (1950). In this section, critical regions for r_1 are determined by the normal approximations with the following moments (Kendall et al., 1983; Dufour and Roy, 1985)

$$mean = -1/n$$

and

$$variance = \frac{(n-2)^2}{n^2(n-1)} \qquad [23.4.2]$$

Knoke (1975) noted that r_1 is a powerful test for detecting nonrandomness for first order autoregression alternatives and that it performs reasonably well for a wider class of alternatives including the first order moving average model.

When dealing with seasonal data, a separate ACF can be estimated for each season. In Section 14.3.2, this is referred to as the *periodic ACF*. Let x_{ij} be a time series value for the ith year and jth season where there are n years of data for each of the m seasons. As in [14.3.4], for the jth season, the ACF is estimated using

$$r_k^{(j)} = \frac{\sum_{i=1}^{n-k}(x_{ij} - x_{.j})(x_{ij-k} - x_{.j-k})}{\left[\sum_{i=1}^{n}(x_{ij} - x_{.j})^2\right]\left[\sum_{i=1}^{n}(x_{ij-k} - x_{.j-k})^2\right]}, \quad k = 1, 2, \cdots \qquad [23.4.3]$$

where $x_{.j}$ is the mean of the jth season.

The ACF for season j is asymptotically normally distributed with a mean of zero and a variance of $1/n$. To be more accurate, the formula in [23.4.2] could be used for calculating the mean and variance of $r_k^{(j)}$ for each season. To check if there is a trend in the jth season, one can calculate $r_l^{(j)}$ using [23.4.3] and then perform a significance test to ascertain if $r_l^{(j)}$ is significantly different from zero at the chosen level of significance. If $r_l^{(j)}$ were significantly large, this may indicate the presence of a trend. This kind of test can be done separately for each of the seasons in order to check for a trend in each season of the year. Note that when $r_j(1)$ is determined using [23.4.3], only the data within seasons j and $j-1$ are employed in the calculation. To produce an overall test for trends across seasons, Fisher's formula in [23.3.30] can be utilized.

23.4.3 Kendall's Tau

For the case of nonseasonal data, Kendall's tau, denoted by τ, is defined in [23.3.5] in terms of the Mann-Kendall statistic S given in [23.3.1]. When dealing with seasonal time series, one can use Kendall's tau for the gth season which is presented in [23.3.9]. However, for the purpose of this discussion, the nonseasonal version of Kendall's tau is entertained. Rather than defining τ in terms of a test statistic as in [23.3.5], an interesting interpretation is to express τ using probabilities. Specifically, for any two pairs of random variables (X_i,Y_i) and (X_j,Y_j), *Kendall's tau* is defined as the difference (Gibbons, 1971)

$$\tau = \pi_c - \pi_d \qquad [23.4.4]$$

where

$$\pi_c = Pr[(X_i < X_j)\cap(Y_i < Y_j)] + Pr[(X_i > X_j)\cap(Y_i > Y_j)]$$

and

$$\pi_d = Pr[(X_i < X_j)\cap(Y_i > Y_j)] + Pr[(X_i > X_j)\cap(Y_i < Y_j)]$$

In the case of no possibility of ties in either the X's or the Y's, the τ can be further expressed as

$$\tau = 2\pi_c - 1 = 1 - 2\pi_d$$

As in Section 23.4.2, let an evenly spaced yearly time series be denoted as x_t, $t = 1,2,\ldots,n$. For this sequence of observations, Kendall's τ is estimated by (Gibbons, 1971, 1976; Conover, 1980; Kendall, 1975; Hollander and Wolfe, 1973)

$$\tau = \frac{S}{\binom{n}{2}} = \frac{N_c - N_d}{\binom{n}{2}} \qquad [23.4.5]$$

where N_c, N_d and S are given by

$$N_c = \sum_{i<j}^{n} \theta'_i$$

for which

$$\theta'_i = \begin{cases} 1, & \text{if } x_i < x_j \\ 0, & \text{otherwise} \end{cases}$$

and

$$N_d = \sum_{i<j}^{n} \delta'_i$$

for which

$$\delta'_i = \begin{cases} 1, & \text{if } x_i > x_j \\ 0, & \text{otherwise} \end{cases}$$

and

$$S = \sum_{i<j}^{n} \phi'_i$$

for which

$$\phi'_i = \begin{cases} 1, & \text{if } x_i < x_j \\ 0, & \text{if } x_i = x_j \\ -1, & \text{otherwise} . \end{cases}$$

The statistic S can also be expressed as

$$S = \binom{n}{2} - 2N_d = N_c - N_d = \binom{n}{2}\tau$$

Under the assumption that the x_i's are IID (identically independently distributed), the means and variances for S and τ, respectively, are given in [23.3.2] and [23.3.10], respectively. Kendall (1975) and Mann (1945) derive the exact distribution of S for $n \leq 10$, and, for samples as small as 10, show that the normal assumption is adequate. However, for use with the normal approximation, Kendall (1975) suggests a continuity correction which is the standard normal variate given in [23.3.3].

23.4.4 Alternative Generating Models

The six models defined in this section are used for simulating the nonseasonal data employed for comparing the powers of the ACF at lag 1 which is r_1 in [23.4.1] and Kendall's tau in [23.4.5]. The first three models contain only *deterministic trends* while the last three have purely *stochastic trends*. Furthermore, under the null hypothesis it is assumed that the time series x_t, $t = 1,2,\ldots,n$, consists of IID random variables. As noted in Section 23.3.2, the Mann-Kendall statistic S, equivalently defined in both Section 23.4.3 and [23.3.1], is often used in place of τ which is defined in [23.4.4], [23.4.5] and [23.3.5] (Kendall, 1975; Hirsch et al., 1982; Hirsch and Slack, 1984; Van Belle and Hughes, 1984). In fact, τ and S are statistically equivalent.

For the case of a *purely deterministic trend component*, the time series, x_t, may be written as

$$x_t = f(t) + a_t \qquad [23.4.6]$$

where $f(t)$ is a function of time only and hence is a purely deterministic trend, while a_t is an IID sequence. On the other hand, a time series having a purely stochastic trend may be defined as

$$x_t = f(x_{t-1}, x_{t-2}, \cdots) + a_t \qquad [23.4.7]$$

where $f(x_{t-1}, x_{t-2}, \ldots)$ is a function of the past data and a_t is an innovation series assumed to be IID and with the property

$$E[a_t \cdot x_{t-k}] = 0, \quad k = 1,2, \cdots . \qquad [23.4.8]$$

In actual practice, it may be difficult to distinguish between deterministic and stochastic trends. For example, the series plotted in Figure 23.4.1 was simulated from the model

$$(1-B)^3 x_t = a_t \qquad [23.4.9]$$

where B is the backward shift operator, $a_t \approx NID(0,1)$ and the starting values are $x_1 = 100$, $x_2 = 101$ and $x_3 = 102$. The model in [23.4.9] is an ARIMA(0,3,0) model and the procedure for simulating with any type of ARIMA model is described in detail in Chapter 9. Based upon the shape of the graph in Figure 23.4.1, the series could probably be adequately described using a purely deterministic trend even though the correct model is purely stochastic. The same comments are also valid for the simulated sequences in Figures 4.2.1, 4.2.2, 4.3.4, and 4.3.5 of Chapter 4. Moreover, a discussion regarding deterministic and stochastic trends is provided in Section 4.6.

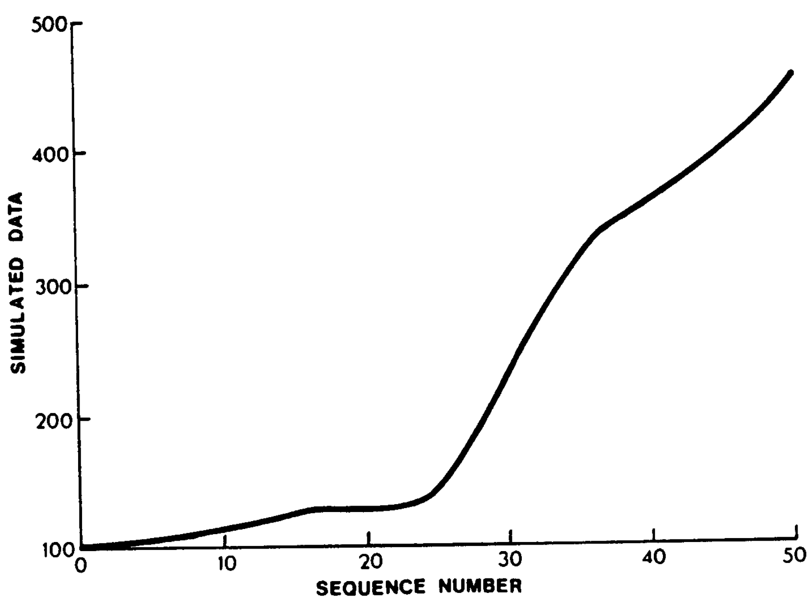

Figure 23.4.1. Simulated sequence from an ARIMA(0,3,0) model.

Box and Jenkins (1976) suggest that for forecasting purposes it is usually better to use a purely stochastic trend model provided that such a model appears to be reasonable a priori for fitting to a given time series and also provides an adequate fit. However, in water quality studies it is often of interest to test if the level of the series has changed in some way and in this case a model with a possible deterministic trend component may seem more suitable beforehand. Three deterministic models, followed by three purely stochastic models are now defined.

Linear Model

In the water resources literature, using linear regression models as alternative hypotheses is quite common (Lettenmaier, 1976; Hirsch et al., 1982; Hirsch and Slack, 1984; van Belle and Hughes, 1984). Assume x_t is given by the linear model which is also written in [4.5.2] as

$$x_t = c + bt + a_t\,,\ t = 1,2,\ldots,n \qquad [23.4.10]$$

where $a_t \approx NID(0,\sigma_a^2)$, and c and b are constants. Without loss of generality, let $c = 0.0$.

Logistic Model

Because it is possible for a series to change rapidly at the start and then gradually approach a limit, a *logistic model* constitutes a reasonable choice for an alternative model. This model is defined as (Cleary and Levenbach, 1982)

$$x_t = M/[1 - c\left\{\exp(-bt)\right\}] + a_t\,,\ t = 1,2,\ldots,n \qquad [23.4.11]$$

where $a_t \approx NID(0,1)$, M is the limit of x_t as t tends to infinity, and b and c are constants.

Step Function Model

Following [4.5.1], the *step function model* is defined as

$$x_t = \begin{cases} a_t, & \text{if } 0 \le t \le n/2 \\ c + a_t, & \text{if } n/2 < t \le n \end{cases} \qquad [23.4.12]$$

where $a_t \approx NID(0,\sigma_a^2)$ and c is the average change in the level of the series after time $t = n/2$. The step function model is a special type of intervention model. The unit step function is defined in [19.2.3] while an intervention model that can handle step interventions is given in [19.2.9]. Besides Chapter 19, applications of intervention models to water quality and quantity time series are presented in Section 22.4.

Barnard's Model

The *Barnard model* is defined as (Barnard, 1959)

$$x_t = x_{t-1} + \sum_{i=1}^{n_t}\delta_i + a_t\,,\ t = 1,2,\ldots,n \qquad [23.4.13]$$

where n_t follows a Poisson distribution with parameter λ, $\delta_i \approx NID(0,\sigma_i^2)$ and $a_t \approx NID(0,1)$.

Without loss of generality, let $x_1 = a_1$. Barnard (1959) developed this model for the use in quality control where there may be a series of n_t correctional jumps between measurements.

Second Order Autoregressive Model

From [3.2.4] or [3.2.5], an *AR(2) model* may be written as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t\ ,\ t = 1,2,\ldots,n \qquad [23.4.14]$$

where $a_t \approx NID(0,\sigma_a^2)$. For the simulation studies executed in Section 23.4.5, $\sigma_a^2 = 1.0$ and $E(x_t) = 0.0$. In Section 3.2.2, the general expression for the theoretical ACF of an AR(p) process is given in [3.2.10] and the approach for solving for the theoretical ACF using the Yule-Walker equations in [3.2.12] is explained. Because of the correlation structure present in an AR(p) model or the AR(2) model in [23.4.14], simulated data from an AR model will contain stochastic trends. In the graph of a series simulated using an AR model such as the one in [23.4.14], a sequence of high values will often be grouped together and low values will often follow other small data points.

Threshold Autoregressive Model

The development of the *threshold autoregressive (TAR) model* is due to Tong (1977, 1978, 1983), Tong and Lim (1980) and Tong et al. (1985). Tong (1983), Tong and Lim (1980), and Tong et al. (1985) found TAR models to be suitable for modelling and forecasting riverflows.

The particular model considered here (Tong, 1983; Tong et al., 1985) is given by

$$x_t = \begin{cases} 1.79 + 0.76x_{t-1} - 0.05x_{t-2} + a_t^{(1)}, & \text{if } J_t < -1 \\ \\ 0.87 + 1.3x_{t-1} - 0.71x_{t-2} + 0.34x_{t-3} + a_t^{(2)}, & \text{otherwise} \end{cases} \qquad [23.4.15]$$

where x_t is the volume of riverflow in cubic metres per second per day, J_t is the temperature in degrees centigrade, and $a_t^{(1)} \approx NID(0,0.69)$ and $a_t^{(2)} \approx NID(0,7.18)$. The above model was estimated for the Vatnsdalsa River in Iceland for the period from 1972 to 1974.

23.4.5 Simulation Experiments

The procedures and algorithms for simulating with ARMA and ARIMA models are described in detail in Chapter 9. When a *trend* component is present, which is the case for the first three models of Section 23.4.4, the noise component is simulated separately and the deterministic component at each point in time is added to this. Note that for the first three models of Section 23.4.4, the noise component is white (i.e., it is independently distributed). Because of the white noise component, the first three models are deemed to have purely deterministic trends. However, if the noise component were correlated and were, for example, an ARMA model, a model containing a deterministic trend component plus the correlated or stochastic noise component would no longer possess purely deterministic trends. This is because the AR and MA components of the ARMA model would create a *stochastic trend* component and when this is added to the deterministic trend part of the model, the overall result would be a *mixed-deterministic-stochastic trend*. In fact, as already pointed out in Section 23.1, the general intervention model in [19.5.8] contains both deterministic trend components (i.e., the intervention terms) and a stochastic trend component (i.e., the correlated noise term). In order to be able to clearly discriminate between the powers of r_1 and Kendall's tau for detecting trends, only purely deterministic trends (the first three models in Section 23.4.4) and purely stochastic trends (the last three models in Section 23.4.4) are entertained in the simulation experiments of this section.

For each of the six generating models of the previous section, sample sizes of length 10, 20, 50 and 100 are considered. For each sample size or length of series, 1000 sequences of the same length are simulated. The null hypothesis is that each replication of a given length is IID while the alternative hypothesis is the replication contains a deterministic or stochastic trend component. For both the ACF at lag one, r_1, in [23.4.1] and Kendall's tau, τ, in [23.3.4] or [23.4.5] and a specified sample size, power functions are estimated for a significance level of 5% by the proportions of rejection of the null hypothesis from 1000 replications. As explained in Section 23.2.2 and Table 23.2.1, the proportion of rejections can be interpreted as the probability of accepting the alternative hypothesis which is the *power*.

For the estimated significance level, the test is said to be *conservative* if the estimated level is clearly less than the nominal level (in this case 0.05). On the other hand, if the estimated level is clearly greater than the 0.05, the test is said to be *optimistic*. Otherwise, the test is said to be *adequately approximated*.

Empirical significance levels and powers are given in Tables 23.4.1 to 23.4.6 for the six models defined in Section 23.4.4, respectively. Notice that except for the TAR model, for each model a range of values is used for each of the parameters and the estimated powers are given for τ and r_1 for sample sizes or series having lengths of 10, 20, 50 and 100. The standard error of any entry in the tables is $\sqrt{\pi(1-\pi)/N}$ (Cochran, 1977), where N is the number of replications and π is the true rejection rate. For example, for the estimated significance level of 5%, the standard error is $\left(\frac{0.05(1-0.05)}{1000}\right)^{1/2} = 0.0069$. The entries in the tables suggest that the critical regions are adequately determined by the null approximate distribution. The results of the simulation experiments are discussed separately for each model.

Linear Model

The findings of the simulation study for the linear model in [23.4.10] are presented in Table 23.4.1. Notice that the two tests, consisting of Kendall's tau and r_1 perform better when the standard deviation, σ_a, for the white noise term, a_t, is smaller. This implies that the better the fit of a linear regression to a time series, the greater the chance of detection of nonrandomness. For instance, for samples as small as 10, the tests are very powerful for small standard deviations. An encouraging aspect of this model is that both tests attain asymptotic efficiency quite rapidly. For example, there is considerable improvement in the power functions from $n = 10$ to $n = 20$. A noteworthy point is that τ is generally more powerful compared to r_1, even though the difference is almost negligible for $n = 50$ and $n = 100$ when both tests approach asymptotic efficiency.

Table 23.4.1. Power comparisons for the linear models with an empirical rejection rate at the 5 percent level of significance.

Parameter Values		n							
		10		20		50		100	
b	σ_a	τ	r_1	τ	r_1	τ	r_1	τ	r_1
0.00	0.05	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.01	0.05	0.335	0.138	0.995	0.749	1.000	1.000	1.000	1.000
0.01	0.50	0.057	0.039	0.072	0.043	0.487	0.080	1.000	0.690
0.01	1.00	0.053	0.040	0.050	0.043	0.146	0.042	0.769	0.131
0.01	2.00	0.050	0.041	0.036	0.047	0.073	0.042	0.268	0.065
0.05	0.05	1.000	0.990	1.000	1.000	1.000	1.000	1.000	1.000
0.05	0.50	0.121	0.063	0.632	0.194	1.000	1.000	1.000	1.000
0.05	1.00	0.066	0.049	0.208	0.064	1.000	0.655	1.000	1.000
0.05	2.00	0.050	0.040	0.084	0.043	0.669	0.121	1.000	0.927
0.10	0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	0.50	0.335	0.138	0.995	0.749	1.000	1.000	1.000	1.000
0.10	1.00	0.121	0.063	0.632	0.194	1.000	1.000	1.000	1.000
0.10	2.00	0.066	0.049	0.208	0.064	1.000	0.655	1.000	1.000

Logistic Model

The results for the logistic model in [23.4.11] are given in Table 23.4.2. As is the case for the linear model, the tests perform better for the logistic model when there is a good fit, indicating nonrandomness. When $M < 1.0$, obviously the standard deviation of 1.0 used in the simulation studies tends to have a greater impact on the simulated data than the other parameters of the model. Hence, a substantial component of x_t is determined by a_t, which is random. For $M \geq 1.0$, the two tests (especially τ) prove effective for detecting the presence of trends. Finally, it can be seen that τ is more powerful than r_1, especially for cases where the logistic model describes the data fairly well ($M \geq 1.0$). There is, however, not much difference between the two tests when $n = 100$.

Step Function Model

As can be seen in Table 23.4.3, the output for the step function model in [23.4.12] indicates greater power for relatively small standard deviations (and hence fairly good fits). What is remarkable about this model is the great power of both tests even for samples as small as 10. For example, for $c = 5$, the power is at least 50% for all sample sizes. The power functions also improve as c increases. Both tests are very effective in detecting trends for even a slight shift of 0.5 in the mean level of the series. For a change of 5 in the mean level, both tests are very powerful. Even though τ is more powerful than r_1, both tests are almost equally powerful for $n \geq 50$.

Table 23.4.2. Power comparisons for the logistic models.

Parameter Values			n							
			10		20		50		100	
b	c	M	τ	r_1	τ	r_1	τ	r_1	τ	r_1
0.01	0.01	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.01	0.01	0.1	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.01	0.01	1.0	0.052	0.041	0.035	0.044	0.041	0.045	0.047	0.050
0.01	0.01	5.0	0.052	0.041	0.037	0.045	0.041	0.045	0.047	0.050
0.01	0.50	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.01	0.50	0.1	0.053	0.042	0.037	0.047	0.043	0.047	0.050	0.048
0.01	0.50	1.0	0.051	0.042	0.049	0.040	0.166	0.057	0.514	0.075
0.01	0.50	5.0	0.096	0.046	0.375	0.078	1.000	0.771	1.000	0.999
0.01	0.90	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.01	0.90	0.1	0.057	0.045	0.072	0.045	0.174	0.060	0.279	0.057
0.01	0.90	1.0	0.825	0.431	1.000	0.950	1.000	1.000	1.000	1.000
0.01	0.90	5.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	0.01	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.10	0.01	0.1	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.10	0.01	1.0	0.053	0.041	0.037	0.047	0.040	0.044	0.047	0.050
0.10	0.01	5.0	0.053	0.041	0.037	0.047	0.043	0.045	0.046	0.050
0.10	0.50	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.10	0.50	0.1	0.054	0.046	0.040	0.047	0.045	0.048	0.047	0.049
0.10	0.50	1.0	0.076	0.044	0.092	0.045	0.155	0.063	0.153	0.053
0.10	0.50	5.0	0.622	0.272	0.934	0.579	0.983	0.893	0.951	0.905
0.10	0.90	0.0	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.10	0.90	0.1	0.052	0.045	0.047	0.045	0.058	0.047	0.053	0.045
0.10	0.90	1.0	0.603	0.239	0.745	0.366	0.733	0.490	0.583	0.464
0.10	0.90	5.0	1.000	0.976	1.000	1.000	1.000	1.000	0.999	1.000

Barnard's Model

The results in Table 23.4.4 for the model in [23.4.13], are very consistent and easily comprehensible. For all sample sizes and all combinations of lambda (λ) and standard deviation (σ_a), r_1 has greater power than τ. For n as small as 50, r_1 attains asymptotic efficiency, while τ is only about 80% efficient. The power of the two tests can be well appreciated by considering the results for $n = 10$. While the power of τ is about 50% that of r_1 is always greater than 50%.

Second Order Autoregressive Model

From Table 23.4.5, one can see that the findings for the AR(2) model in [23.4.14] parallel fairly closely those of Barnard's model. The main difference between the two models is that the results here are not as dramatic as in Table 23.4.4. Here too, r_1 is more powerful than τ. As n increases, the power of r_1 increases faster than that of τ. For $n = 100$, r_1 attains almost 100% of efficiency while τ performs fairly poorly in some cases. For example, for $\phi_1 = -0.2$ and $\phi_2 = 0.5$

Table 23.4.3. Power comparisons for the step function models.

Parameter Values		n							
		10		20		50		100	
c	σ_a	τ	r_1	τ	r_1	τ	r_1	τ	r_1
0.00	0.05	0.052	0.041	0.035	0.044	0.040	0.045	0.048	0.051
0.05	0.05	0.199	0.121	0.382	0.152	0.803	0.281	0.983	0.520
0.05	0.50	0.058	0.040	0.038	0.048	0.051	0.039	0.069	0.051
0.05	1.00	0.053	0.040	0.036	0.047	0.046	0.044	0.055	0.049
0.05	2.00	0.053	0.040	0.035	0.045	0.045	0.045	0.052	0.050
0.50	0.05	0.711	1.000	0.994	1.000	1.000	1.000	1.000	1.000
0.50	0.50	0.199	0.121	0.382	0.152	0.803	0.281	0.983	0.520
0.50	1.00	0.090	0.057	0.131	0.064	0.283	0.070	0.525	0.113
0.50	2.00	0.055	0.048	0.060	0.042	0.103	0.043	0.160	0.059
1.00	0.05	0.711	1.000	0.994	1.000	1.000	1.000	1.000	1.000
1.00	0.50	0.490	0.352	0.887	0.594	1.000	0.958	1.000	1.000
1.00	1.00	0.199	0.121	0.382	0.152	0.803	0.281	0.983	0.520
1.00	2.00	0.090	0.057	0.131	0.064	0.283	0.070	0.525	0.113
5.00	0.05	0.711	1.000	0.994	1.000	1.000	1.000	1.000	1.000
5.00	0.50	0.711	1.000	0.994	1.000	1.000	1.000	1.000	1.000
5.00	1.00	0.711	0.943	0.944	1.000	1.000	1.000	1.000	1.000
5.00	2.00	0.596	0.507	0.960	0.821	1.000	0.998	1.000	1.000

the power of r_1 is 0.881 while that of τ is only 0.061 for $n = 100$.

Threshold Autoregressive Model

The specific TAR model used in the simulation experiments is given in [23.4.15]. From Table 23.4.6, one can see that the results for this TAR model are very similar to those for Barnard's model and the AR(2) model in Tables 23.4.4 and 23.4.5, respectively. The r_1 test is obviously more powerful than τ. Moreover, while the power of r_1 increases very rapidly with increasing n, the power of τ only makes a slow progression. Finally, the r_1 test is about 90% efficient for $n = 20$ and it attains 100% efficiency at $n = 50$. On the other hand, the power of τ is less than 50% even for $n = 100$.

23.4.6 Conclusions

The general deductions from the simulation experiments presented in Section 23.4.5 are that the nonparametric test, using τ, is more powerful for detecting trends for data generated from the first three models while the parametric test, utilizing r_1, is more powerful for discovering trends in synthetic sequences from the last three models. As noted in Section 23.4.4, the first three models contain deterministic trends and the last three have stochastic trends. Therefore, it is reasonable to conclude that τ is more powerful for detecting deterministic trends while r_1 is more powerful for discovering stochastic trends.

Table 23.4.4. Power comparisons for Barnard's models.

Parameter Values		n					
		10		20		50	
λ	σ_a	τ	r_1	τ	r_1	τ	r_1
1.0	0.05	0.484	0.579	0.681	0.933	0.809	1.000
1.0	0.50	0.459	0.579	0.667	0.944	0.819	1.000
1.0	1.00	0.482	0.571	0.680	0.944	0.817	1.000
1.0	2.00	0.486	0.579	0.655	0.955	0.800	1.000
2.0	0.05	0.477	0.573	0.688	0.937	0.802	1.000
2.0	0.50	0.575	0.586	0.690	0.946	0.789	1.000
2.0	1.00	0.467	0.586	0.683	0.954	0.798	1.000
2.0	2.00	0.460	0.571	0.669	0.958	0.784	1.000
5.0	0.05	0.485	0.560	0.689	0.935	0.812	1.000
5.0	0.50	0.469	0.562	0.670	0.937	0.802	1.000
5.0	1.00	0.491	0.577	0.680	0.948	0.802	1.000
5.0	2.00	0.478	0.591	0.674	0.952	0.783	1.000
10.0	0.05	0.488	0.565	0.690	0.938	0.796	1.000
10.0	0.50	0.501	0.628	0.677	0.961	0.802	1.000
10.0	1.00	0.472	0.603	0.665	0.946	0.790	1.000
10.0	2.00	0.480	0.586	0.665	0.960	0.796	1.000
20.0	0.05	0.473	0.569	0.689	0.941	0.801	1.000
20.0	0.50	0.501	0.612	0.658	0.949	0.804	1.000
20.0	1.00	0.498	0.593	0.654	0.950	0.811	1.000
20.0	2.00	0.506	0.607	0.666	0.946	0.819	1.000

In practice, it is advantageous to have both a sound physical and statistical understanding of the time series being analyzed. This will allow one to decide whether one should employ models possessing deterministic trends or whether one should use models having stochastic trends. For example, it may be better to describe certain kinds of water quality measurements using models having deterministic trends. On the other hand, for modelling seasonal riverflows, models having stochastic trends, such as a TAR model, may work well (Tong, 1983; Tong et al., 1985). In other cases, one may wish to use a model which possesses both deterministic and stochastic trends. As a matter of fact, most of the intervention models used in the water quality and quantity applications of Chapter 19 and Section 22.4 have components to model both deterministic and stochastic trends.

23.5 WATER QUALITY APPLICATIONS

23.5.1 Introduction

When executing a complex *environmental impact assessment study*, usually a wide variety of statistical tests are required in order to check a range of hypotheses regarding the statistical properties of the data. The main objective of this section is to clearly explain how both nonparametric and parametric tests can be employed in an optimal fashion to extract

Table 23.4.5. Power comparisons for the AR(2) models.

Parameter Values		n							
		10		20		50		100	
ϕ_1	ϕ_2	τ	r_1	τ	r_1	τ	r_1	τ	r_1
-1.40	-0.80	0.000	0.731	0.000	1.000	0.000	1.000	0.000	1.000
-0.70	-0.80	0.002	0.009	0.000	0.154	0.000	0.975	0.000	1.000
0.70	-0.80	0.031	0.170	0.009	0.485	0.002	0.991	0.000	1.000
1.40	-0.80	0.247	0.881	0.132	0.998	0.067	1.000	0.045	1.000
-1.20	-0.50	0.000	0.660	0.000	0.989	0.000	1.000	0.000	1.000
-0.60	-0.50	0.007	0.064	0.002	0.279	0.000	0.900	0.000	0.998
0.60	-0.50	0.072	0.223	0.050	0.470	0.036	0.932	0.031	1.000
1.20	-0.50	0.305	0.745	0.264	0.985	0.285	1.000	0.230	1.000
-0.80	-0.20	0.001	0.422	0.001	0.866	0.000	1.000	0.000	1.000
-0.40	-0.20	0.009	0.082	0.004	0.210	0.002	0.645	0.002	0.939
0.40	-0.20	0.082	0.137	0.091	0.267	0.089	0.684	0.103	0.951
0.80	-0.20	0.260	0.456	0.290	0.850	0.264	1.000	0.268	1.000
-0.40	0.10	0.010	0.207	0.010	0.470	0.008	0.840	0.007	0.991
0.40	0.10	0.180	0.180	0.187	0.386	0.248	0.814	0.245	0.985
0.80	0.10	0.390	0.452	0.536	0.872	0.633	1.000	0.625	1.000
-0.50	0.30	0.013	0.500	0.005	0.788	0.002	0.990	0.005	1.000
-0.30	0.30	0.024	0.239	0.023	0.466	0.023	0.776	0.028	1.000
0.30	0.30	0.193	0.138	0.297	0.306	0.335	0.701	0.348	0.934
0.50	0.30	0.309	0.243	0.401	0.585	0.524	0.975	0.527	0.999
-0.40	0.50	0.017	0.615	0.013	0.824	0.005	0.984	0.008	1.000
-0.20	0.50	0.046	0.319	0.059	0.470	0.088	0.682	0.061	0.881
0.20	0.50	0.175	0.143	0.302	0.250	0.377	0.578	0.421	0.831
0.40	0.50	0.288	0.211	0.470	0.513	0.610	0.933	0.686	0.999

Table 23.4.6. Power comparisons for the TAR model.

n							
10		20		50		100	
τ	r_1	τ	r_1	τ	r_1	τ	r_1
0.320	0.499	0.435	0.904	0.320	1.000	0.338	1.000

systematically relevant information from a set of water quality time series. In particular, the effects of industrial development at Nanticoke, Ontario, upon the nearshore Lake Erie water chemistry are examined in a comprehensive statistical study. This undertaking was originally carried out by the authors in conjunction with Acres International Limited of Niagara Falls, Ontario, for the Ministry of the Environment in the Canadian province of Ontario. Some of the statistical findings for the Lake Erie study given in Section 23.5.2 are also presented by Hipel et al. (1988).

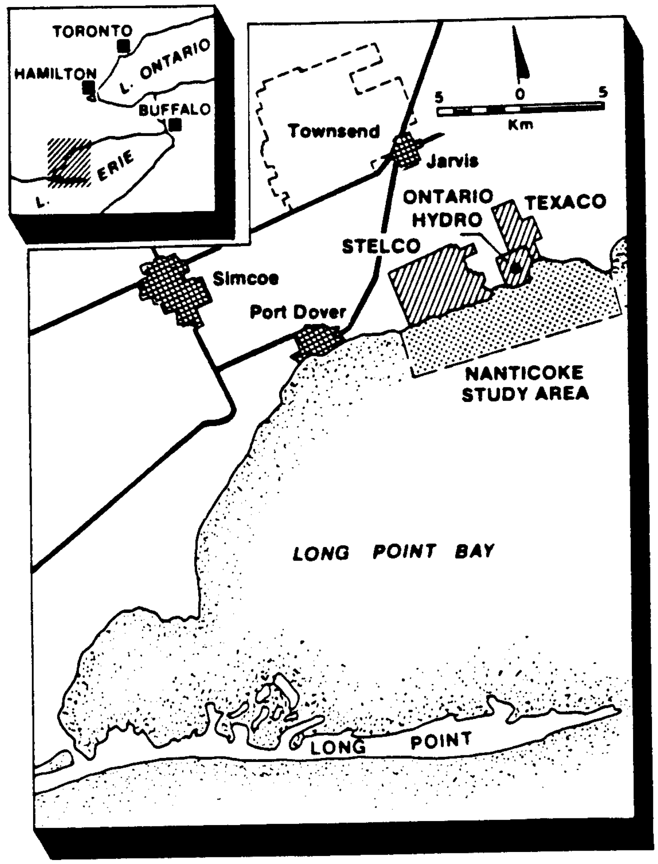

Figure 23.5.1. Location of the Lake Erie water quality study.

A map of the Long Point Bay region of Lake Erie which contains the Nanticoke study area is displayed in Figure 23.5.1. Notice that Nanticoke is situated on the north shore of Long Point Bay in Lake Erie. From the late 1960's and onwards, major industrial development took place at the town of Nanticoke. In January, 1972, Ontario Hydro's (the provincial company that generates almost all of the electrical power in Ontario) 4000 MW fossil-fueled thermal generating station commenced operations. Texaco Canada Inc. constructed an oil refinery which came into operation in November, 1978. Finally, the Steel Company of Canada (Stelco) built a steel mill that started to produce steel in May, 1980. Because the foregoing industrial projects could adversely affect the water quality of Lake Erie, the goal of the Nanticoke study was to detect trends in various water quality variables.

Water quality monitoring began in 1969 at eight (stations 112, 501, 648, 518, 810, 1008, 1016, and 994) of the fifteen sampling stations shown in Figure 23.5.2. Since 1969, the remaining seven of the fifteen stations were added to the network. Many of the stations were sampled at more than one depth, such as near the surface and near the bottom. Unfortunately, as is also the case for the water quality data examined in Chapter 22 and Section 24.3.2, the water quality series measured in Lake Erie contain observations separated by unequal time intervals, many of

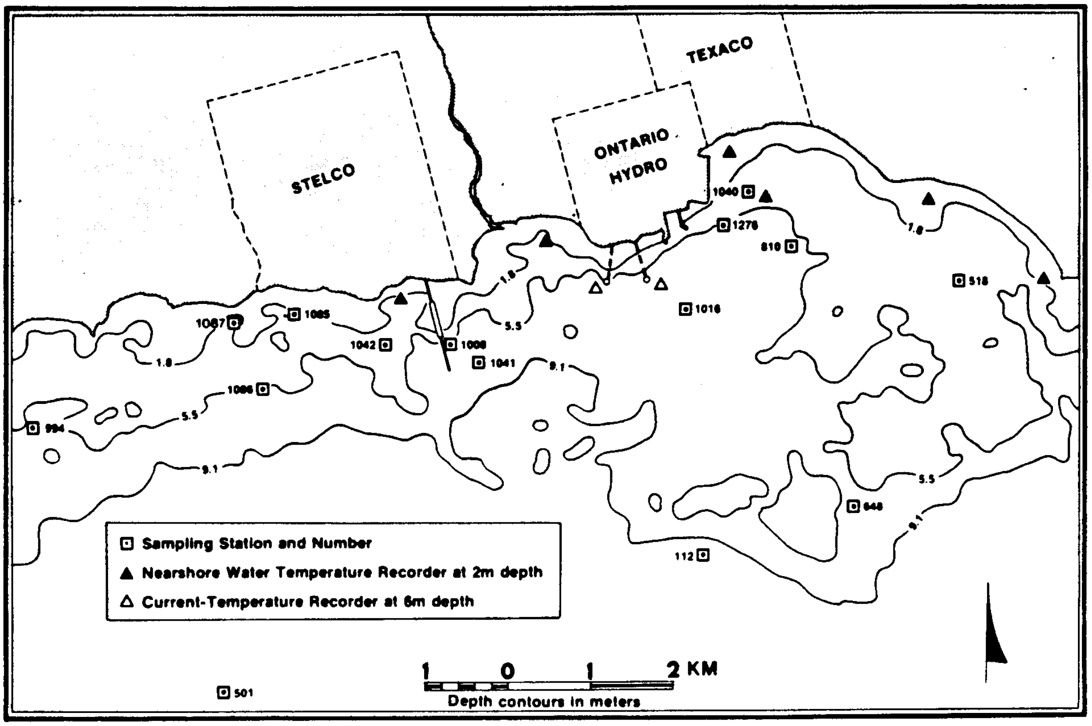

Figure 2. Sampling Stations at Long Point Bay in Lake Erie.

Figure 23.5.2. Sampling stations at Nanticoke in Long Point Bay, Lake Erie.

which are relatively long. Although not all of the water quality variables were assessed for trends, about 50 different water quality variables were sampled across all of the stations and the more important variables were collected at each of the stations. Detailed statistical testing for the presence of trends and other statistical characteristics were carried out for 14 water quality variables (see Table 23.5.6 for a list of these 14 variables) measured at Stations 501, 810, 994, 1085 and 1086. In the next section, the statistical procedures used in the study are outlined and some representative results are given. Finally, the *trend analysis methodology* put forward in the upcoming section constitutes one of the three overall trend procedures described in the book and summarized in Table 1.6.1.

23.5.2 Trend Analysis of the Lake Erie Water Quality Series

Selecting Appropriate Statistical Tests

To detect trends and uncover other statistical properties of the Lake Erie water quality time series, appropriate statistical tests must be employed. In order to have the highest probability of discovering suspected statistical characteristics which may be present in the time series, one must select the set of tests that possess the best capabilities for uncovering the specified statistical properties. To accomplish this, *one must be cognizant of both the general statistical properties of the data and the main attributes of the statistical tests*. For example, with respect to the characteristics of the data, one should be aware of properties such as the quantity of data, large time gaps where no measurements were taken, outliers and data which fall below the *detection limits* (see Section 23.3.2 for discussions about detection limits in water quality time series). As discussed in Sections 19.2.3 and 22.3, often these properties of the data are known in advance or else are revealed using *exploratory data analysis* tools. From the point of view of a statistical test that can be used, one should know key facts which include the specific null and alternative hypotheses that the given statistical test is designed to check, the major distributional assumptions underlying the test, the types of samples with which the test can be used, and the kinds of measurements that can be utilized with the test. By being cognizant of the properties of the data and the main capabilities of a wide range of both *parametric and nonparametric tests*, one can choose the most appropriate statistical tests for testing specified hypotheses such as the presence of trends. As mentioned in Section 23.3.1, summaries and charts regarding the capabilities of both nonparametric and parametric tests are available in many well known statistical texts. The handbook of Sachs (1982), for instance, is very helpful for locating the most appropriate parametric and nonparametric tests to employ in a given study. Table 23.1.1 provides a list of the nonparametric tests described in Chapter 23. The tests that are eventually selected can then be used at the *confirmatory data analysis* stage for hypothesis testing (see Section 23.2 for a general discussion of statistical tests).

The particular statistical methods used in the Lake Erie study are listed in Table 23.5.1. Notice that for each statistical method, the general purpose of the technique is described and the specific reason for using it in the Lake Erie study is explained. Furthermore, if the method is described in the text, the location is cited. Otherwise, an appropriate reference is given. The first four statistical methods in Table 23.5.1 constitute exploratory data analysis tools while the remaining methods are usually employed at the confirmatory data analysis stage. The nonparametric tests given in the table are marked with asterisks. Notice that the last statistical method, regression analysis, is discussed in detail in Section 24.2.3. All of the statistical methods listed in Table 23.5.1 were applied to each of the 14 specified water quality variables at each of the 5 stations, consisting of Stations 501, 810, 994, 1085 and 1086. To clearly explain how an environmental impact assessment project is carried out, some of the informative results of the Lake Erie study are now presented for the methods marked with a cross in Table 23.5.1. For explaining how the techniques are used in practice, the chloride water quality (mg/*l*) and total phosphorous (mg/*l*) variables are used the most. Finally, for a description of water quality processes, the reader can refer to the book of Waite (1984), as well as other authors.

Data Listing

For the chloride variable at Station 501 in Figure 23.5.2, 173 observations are available in mg/*l* from July 13, 1970, to November 19, 1979. The first 25 of these measurements are listed in Table 23.5.2. To calculate the day number, the start of January 1, 1969, is taken as day number 0.0. The component to the right of the decimal for a day number refers to the fraction of a 24 hour period at which the measurement was taken. Consequently, from Table 23.5.2, the first observation was taken on July 13, 1970, at 16 hours and 40 minutes, which has the day number 558.611. The gap, expressed in number of days between adjacent observations, is given

Table 23.5.1 Statistical methods used in the Lake Erie water quality study.

METHOD	GENERAL PURPOSE	SPECIFIC PURPOSE in the Study
Data listing†	For each series, want to know exact values, dates of measurements, depths of measurements, and station number (Section 23.5.2, Table 23.5.2).	Same as under general purpose.
Graphs of the data†	Visually detect main statistical characteristics of a series (Section 22.3.2).	Visually detect trends and see spacing of the observations.
Tukey five number summary†	Describe how the observations in a series are distributed in each season (Section 23.3.3).	Describe how measurements in a water quality series are distributed in each month.
Box and whisker graphs†	See a graphical display of the five number summary for each season in a series (Section 23.3.3).	See a plot of the five numbers for each month in a series.
Seasonal Mann-Kendall test*†	Check for trends in a series for each season of the year (Section 23.3.2).	Check for trends in a series for each month.
Fisher's combination test†	Combining tests of hypotheses (Section 23.3.4).	Test for a trend across all the months in a series by combining the significance levels from the seasonal Mann-Kendall tests for each month into a χ^2 statistic (see [23.3.30]).
Wilcoxon signed rank test*†	Determine whether the medians of two samples are the same (Appendix A23.2).	Find out whether or not measurements taken at two different depths at exactly the same time possess the same median.
Confidence interval for the median*†	Calculate a confidence interval for the difference in the medians between two samples (Appendix A23.2).	Calculate 95% confidence interval for the difference in medians between measurements taken at two different depths at exactly the same time. If zero is not contained in the confidence interval, the two medians are significantly different from one another.
Kendall rank correlation*	Determine whether or not two series are independent of one another (Appendix A23.1).	Ascertain if measurements taken at the same time at alternate depths are correlated with one another.
Cross correlation function†	Determine whether or not two series are independent of one another (see Section 22.3.4 and also Kendall (1975, Ch. 2)).	Find out if measurements taken at the same time at alternate depths are correlated with one another.
Pitman's test for equality of correlated variance	Ascertain whether or not two correlated variances are the same (Pitman, 1939).	Determine if the variances of samples taken at two different depths are the same.
One way analysis of variance	Determine if the means across k samples are significantly different from one another (Sachs, 1984, pp. 501-509). It is assumed that the k populations are normally independently distributed and have equal variances.	Find out whether or not the means among replicated samples are the same.
Kruskal-Wallis test*†	Nonparametric test to check whether or not the distributions or means across k samples are the same (Appendix A23.3). The observations are assumed to be independent of one another and follow the same distribution.	Determine if the means among replicated samples are the same.
Regression analysis†	Parametrically model relationships within a series and among series.	Determine the best data transformation, ascertain the components required in a regression model, and estimate both the average monthly and annual values for a series (see Section 24.2.3).

* - nonparametric test

† - applications for this method are given in the text

in the third column. Each measured value in mg/l along with the depth in meters at which the sample was taken are presented in the fourth and fifth columns, respectively. Notice that the extreme values consisting of the outside and far-out values defined in Section 22.3.3 are marked.

Table 23.5.3 shows the number of available chloride measurements by month and year for Station 501. Notice, for example, that no measurements were taken during January, February and March across all of the years. Additionally, no observations are available for the years 1969, and 1980 to 1983.

Table 23.5.2. Data listing of chloride measurements (mg/*l*) at Station 501, Long Point Bay, Lake Erie.

Day Number	Date	Gap	Measured Value	Depth (m)
558.611	July 13, 1970		27.0*	4.80
588.604	Aug. 12, 1970	30	26.0*	6.20
859.726	May 10, 1971	271	35.0**	6.10
887.632	June 7, 1971	28	25.0	6.20
915.736	July 5, 1971	28	24.0	6.70
944.635	Aug. 3, 1971	29	24.0	6.40
971.615	Aug. 30, 1971	27	23.0	6.20
999.625	Sep. 27, 1971	28	25.0	6.40
1028.604	Oct. 26, 1971	29	25.0	6.00
1197.628	Apr. 12, 1972	169	25.0	6.00
1223.802	May 8, 1972	26	25.0	6.50
1253.660	June 7, 1972	30	24.0	6.20
1280.708	July 4, 1972	27	24.0	6.50
1308.618	Aug. 1, 1972	28	24.0	6.30
1338.646	Aug. 31, 1972	30	25.0	6.00
1365.503	Sep. 27, 1972	27	24.0	6.50
1419.635	Nov. 20, 1972	54	23.0	5.50
1622.521	June 11, 1973	203	23.0	1.00
1622.521	June 11, 1973	0	24.0	10.60
1650.635	July 9, 1973	28	23.0	1.00
1650.635	July 9, 1973	0	23.0	10.60
1679.545	Aug. 7, 1973	29	23.0	1.00
1679.545	Aug. 7, 1973	0	24.0	10.60
1707.597	Sep. 4, 1973	28	24.0	1.00
1707.597	Sept. 4, 1973	0	25.0	11.00
.	.	.	.	.

Remarks:

1. January 1, 1969 at 0:00 AM is taken as Day Number 0.0.
2. Outside values are indicated by *.
3. Far-outside values are indicated by **.
4. For independent and identically distributed normal variables, the expected percentages of outside and far-outside values are 0.76% and 0.000013%, respectively.

The *Tukey 5-number summary* defined in Section 22.3.3 is also listed in Table 23.5.3. In addition, the numbers of observed and expected outside and far-out values are given. The expected number of outside and far-out values are calculated by assuming that the data follow a normal distribution (see Section 22.3.3). To calculate the expected number of outside values, one multiplies 173 (the total number of observations in Table 23.5.3) times 0.0076 (the probability of having outside values if the data are NID) to obtain the expected figure of 1.3148 shown in

Table 23.5.3. Number of chloride measurements (mg/l) at Station 501 according to the month and year.

	Number of measurements available by month and year															
	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	TOTAL
Jan																0
Feb																0
Mar																0
Apr				1		2	2	2	2	2	3					14
May			1	1		4	2	5	4	8	6					31
Jun			1	1	2	2		2	2	4	3					17
Jul		1	1	1	2	4	2	2	4	2	3					22
Aug		1	2	2	2	2	4	7	2	7						29
Sep			1	1	2	2	2	2	4	2	3					19
Oct			1		2	2	2	2		4	3					16
Nov				1	2		2	3	4	6	3					21
Dec						2				2						4
TOTAL	0	2	7	8	12	20	16	25	22	37	24	0	0	0	0	173

Tukey five-number summary: 19.0 20.4 21.0 22.5 35.0

	Observed	Expected
Number of Outside values	2	1.314800
Number of Far-outside values	1	0.000225

Table 23.5.3. Likewise, to determine the expected number of far-outside values one multiplies 13×10^{-8} times 173 to get 0.000225.

Graphs of the Data

Graphs of time series are presented throughout the text for a wide range of series while a discussion regarding the usefulness of graphs as exploratory data analysis tools is given in Sections 5.3.3 and 22.3.2. Figures 1.1.1 and 19.1.1, for example, displays a plot of 72 average monthly phosphorous data points (mg/l) from January, 1972, until December, 1977, for measurements taken downstream from the Guelph sewage treatment plant located on the Speed River in Ontario, Canada. In the figure, it can be seen that conventional phosphorous treatment has dramatically decreased the mean level of the series after the intervention date when the tertiary treatment was implemented. The black dots indicate locations where data are missing and hence had to be estimated. In Section 19.4.5, intervention analysis is used to model the effects of the phosphorous treatment and estimate the missing observations.

A simple approach to display effectively the statistical characteristics of a data set using ordinary output paper from a computer is to employ a *jittered one dimensional plot*. The data are plotted horizontally between the smallest observation on the left and the largest value on the right. The exact magnitudes of the smallest and largest values are given as part of the 5 number summary shown in Table 23.5.3. In the plot, the letters $a,b,c,\ldots,$ denote $1,2,3,\ldots,$ data points, respectively.

Figure 23.5.3a displays the jittered plot for all of the chloride data at Station 501. When the chloride data are plotted according to each month across all years as in Figure 23.5.3b, the manner in which the data are distributed according to each season can be observed. Notice from the jittered plot of chloride according to each year in Figure 23.5.3c, that the chloride level is

```
  b  e e d  a a   a   c  a
  b  e e d  b a a  b   c  a
a b  e e e  b b a  b  ac  a
a b  e e e  b b a aba ac  b   a   a                    a
  b  e e e  b a a  b   c  a
  b  e e d  a a    b   c  a
  a  e d d  a a    a   c  a
   Note:  a,b,c,... etc. denote 1,2,3,... data points, respectively.
```

(a) Jittered one-dimensional data plot of measured values.

```
Jan
Feb
Mar
Apr   d a  d a   a    b     a
May   c ℓ  e a b      b aab  a                          a
Jun     f  d c        a   b  a
Jul   b e  d b    b aba   b        a
Aug        e g c c b  d    c  a   a
Sep   b b  g b   b         b  b
Oct   b a    h b           b  a
Nov b   a  e e d a    a   a  a
Dec          b   b
```

(b) Jittered one-dimensional data plot of measured value by month.

```
1968
1969
1970                              a   a
1971                  a   b   c                         a
1972                  a   d   c
1973                  d   f   b
1974  b b  b b  b b abaaab
1975         a d e b  d
1976    a  e ℓ e b
1977       m g b
1978  d ℓ  ℓ i
1979b g m  b
1980
1981
1982
1983
```

(c) Jittered one-dimensional data plot of measured value by year.

Figure 23.5.3. Jittered one-dimensional plots of the chloride observations (mg/l) at Station 501, Long Point Bay, Lake Erie.

decreasing over the years. A jittered plot of depths at which the chloride observations are taken is presented in Figure 23.5.4.

```
     Tukey five-number summary of all the depths (m) used
         1.00      1.00      1.50     11.00     13.00
               Jittered one-dimensional data plot of depths
i  e                                                          a   d  b
i  e                           aa a                     a   a a   d  b
i  e                        a  aa a               a     a   a b a d  b
i  e                    a   a  aaaa a             a  a  a   a b a da c   a
i  e                           aa a               a     a   a a a d  b
i  e                           aa a                     a     a   d  b
i  d                                                          a   d  b
     Note:  a,b,c,... etc. denote 1,2,3,... data points respectively
```

Figure 23.5.4. Jittered one-dimensional plot of depths at which the chloride observations at Station 501 in Long Point Bay, Lake Erie, are taken.

Box-and-Whisker Graphs

The box-and-whisker graph, which is based upon the 5-number summary (Tukey, 1977) is described in Section 22.3.3 where illustrative plots are also given. When entertaining seasonal data such as monthly or quarterly data, it is instructive to calculate a 5-number summary plus outside and far-outside values for each season. For the given total phosphorous data (mg/l) at Station 501, Figure 23.5.5 depicts box-and-whisker graphs that are commonly referred to as box plots. In this figure, the data have not been transformed using a Box-Cox transformation from [3.4.30]. The *far-out values* are indicated by a circle in Figure 23.5.5, where far-out values are not marked if there are four or less data points for a given month. Below each month is a number which gives the number of data points used to calculate the box-and-whisker graph for that . When there are not many data points used to determine a box-and-whisker plot for a given month, any peculiarities in the plot should be cautiously considered. The total number of observations across all the months is listed on the right below the x axis.

For a given month in a box-and-whisker diagram, symmetric data would cause the median to lie in the middle of the box or rectangle and the lengths of the upper and lower whiskers would be about the same. Notice in Figure 23.5.5, for the total phosphorous data at Station 501, that the whiskers are almost entirely above the rectangles for almost all of the months and there are 8 far-out values above the boxes. This lack of symmetry can at least be partially rectified by transforming the given data using the Box-Cox transformation of $\lambda = 0$ in [3.4.30]. By comparing Figure 23.5.5 to Figure 23.5.6, where natural logarithms are taken of the total phosphorous data, the improvement in symmetry can be clearly seen. Furthermore, the Box-Cox transformation has reduced the number of far-out entries from 8 in Figure 23.5.5 to 5 in Figure 23.5.6.

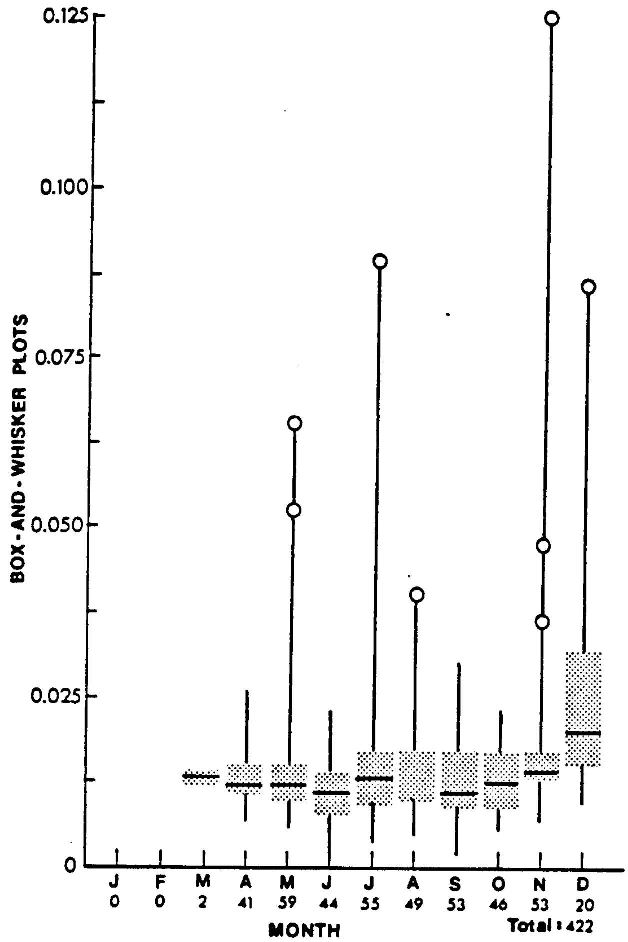

Figure 23.5.5. Box-and-whisker plots of the total phosphorous (mg/l) data at Station 501, Long Point Bay, Lake Erie, from April 22, 1969, to December 13, 1983.

As is also explained in Section 23.3.3, Box-and-whisker plots can be employed as an important exploratory tool in *intervention studies*. If the date of the intervention is known, box-and-whisker diagrams can be constructed for each season for the data before and after the time of the intervention. These two graphs can be compared to ascertain for which seasons the intervention has caused noticeable changes. When there are sufficient data, this type of information is crucial for designing a proper intervention model to fit to the data at the confirmatory data analysis stage (see Section 19.2.3).

For the Nanticoke data, there are two major interventions. First, Ontario Hydro built a fossil-fuelled electrical generating plant which began operating in January, 1972, and came into full operation by about January 1, 1976. Because not much data are available before 1972, January 1, 1976, is taken as the intervention data at which water quality measurements near the

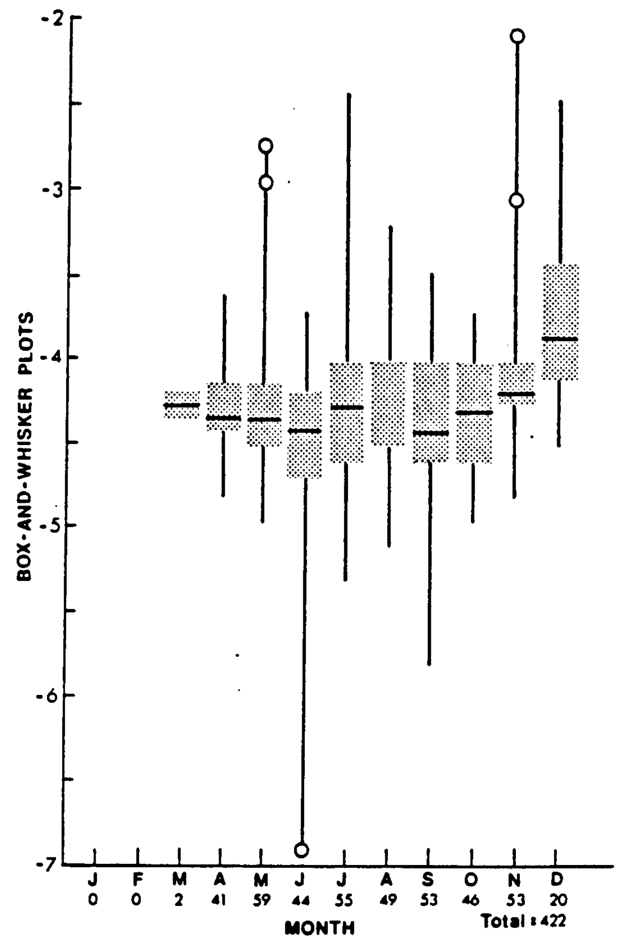

Figure 23.5.6. Box-and-whisker plots of the logarithmic total phosphorous (mg/l) data at Station 501, Long Point Bay, Lake Erie, from April 22, 1969, to December 13, 1983.

Ontario Hydro plant may be affected. Of the 5 stations analyzed, only Station 810 is close to the plant. For each of the water quality variables measured at Station 810, box-and-whisker plots are made before and after the intervention date in order to qualitatively discover any possible statistical impacts of the intervention.

Second, the Steel Company of Canada (Stelco) plant came into operation about April 1, 1980. Because sites 501, 994, 1085, and 1086 are relatively close to the Stelco factory, box-and-whisker graphs are made before and after the intervention for each water quality time series at each station. Figures 23.5.7 and 23.5.8 display the box-and-whisker graphs for the natural logarithms of the total phosphorous data at Station 501 before and after the Stelco intervention, respectively. The dates in brackets in the titles for Figures 23.5.7 and 23.5.8 indicate the intervals of time for which measurements were taken before and after the intervention, respectively.

When these two graphs are compared, it appears that for most of the months there is a slight drop in the median level after the intervention. Using an intervention model based upon a regression analysis design (see Chapter 24 and references therein), a confirmatory data analysis could be executed to ascertain the magnitudes of the changes in the monthly means and if they are significant. Furthermore, one should also take into account overall changes in Lake Erie by considering measurements at locations outside of the Nanticoke region.

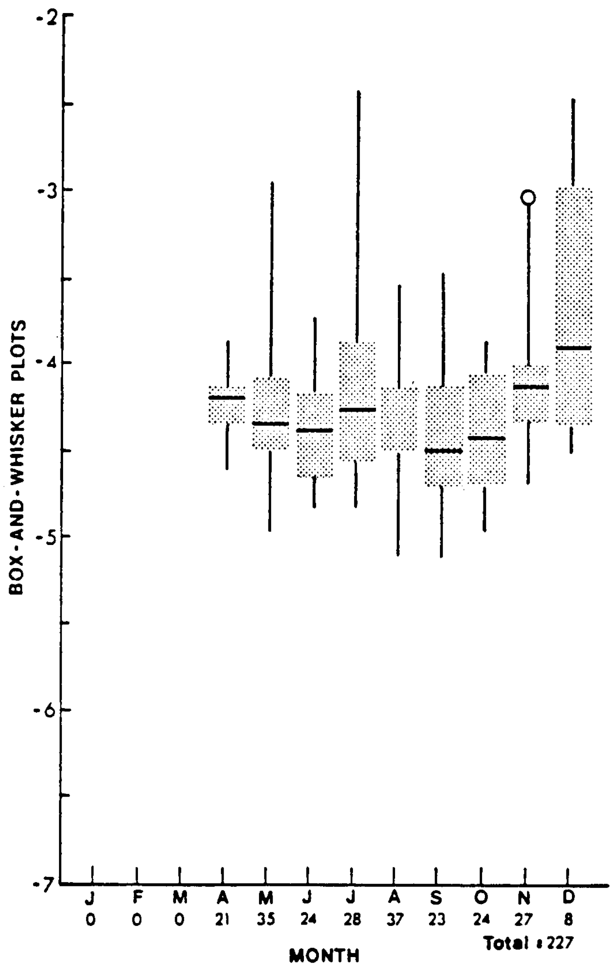

Figure 23.5.7. Box and whisker plots of the logarithmic total phosphorous data (mg/l) at Station 501, Long Point Bay, Lake Erie, before April 1, 1980 (data available from April 22, 1969, to November 19, 1979).

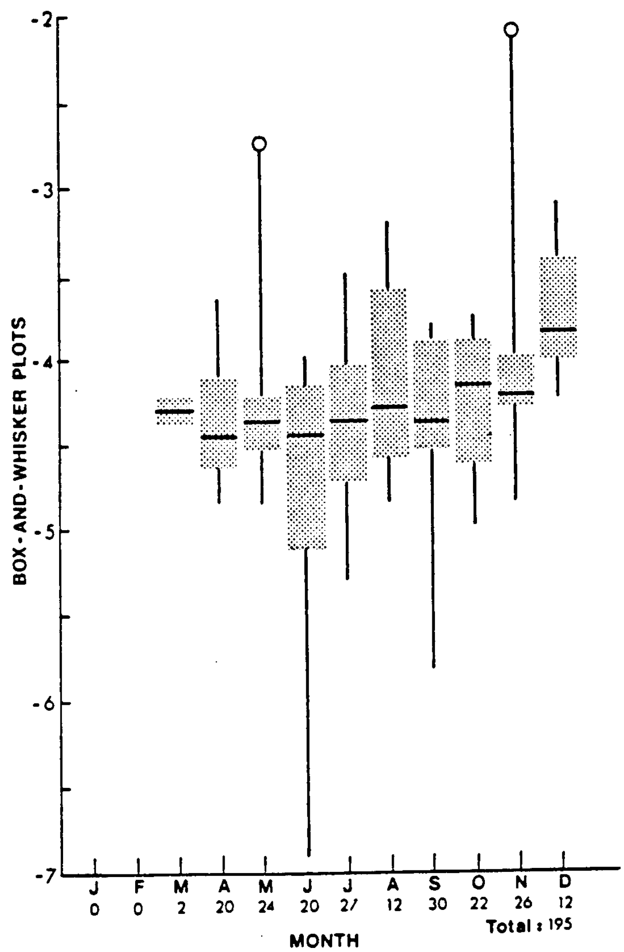

Figure 23.5.8. Box and whisker plots of the logarithmic total phosphorous (mg/l) data at Station 501, Long Point Bay, Lake Erie, after April 1, 1980 (data available from April 13, 1981 to December 13, 1983).

A third intervention in the Nanticoke region is due to the Texaco oil refinery which began production in November, 1978. Because the discharge from this plant is relatively quite small, the possible effects of the Texaco intervention are not considered in this study.

Seasonal Mann-Kendall Tests

For a given water quality variable at a specified station, the seasonal Mann-Kendall test can be used to detect trends in each month of the year. A detailed description of this test is given in Section 23.3.2. For the case of the chloride measurements taken at Station 501, Table 23.5.4 presents results of the seasonal Mann-Kendall and other related tests. Each entry in the table of

Table 23.5.4. Trend analysis of monthly median values using the seasonal Mann-Kendall test for the chloride series (mg/*l*) at Station 501, Long Point Bay, Lake Erie.

Monthly Median Values times 10

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1970							270	260				
1971					350	250	240	235	250	250		
1972				250	250	240	240	245	240		230	
1973						235	230	235	245	240	245	
1974				230	238	200	211	225	205	210		220
1975				215	230		225	225	220	215	215	
1976				205	205	207	210	210	210	210	215	
1977				205	205	210	205	215	205		210	
1978				197	200	200	200	205	202	210	205	210
1979				195	200	205	200		195	195	190	
tau				-0.98	-0.96	-0.62	-0.93	-0.86	-0.82	-0.82	-0.88	
SL(%)				0.39	0.17	4.61	0.03	0.22	0.33	1.87	0.98	

Combination of Scores and their Variances

Sum	Variance	SL
-196.0	5.59000×10^{2}	1.13324×10^{-16}

Fisherian Combination of the Significance Levels

CHI-SQ	DF	SL
87.01	16	0.00000

years versus months is the median value for a given month and year. By utilizing [23.3.4] or [23.3.9] for the data in a specific month across all of the years for which data are available, Kendall's tau can be determined. The observed value of S_g for each month which is calculated using [23.3.7] is not displayed in the table. Because the observed τ value for each month is negative, this indicates that there may be a decreasing trend in each month. Consider, for instance, the month of April for which the calculated τ value is -0.98. Since the SL (significance level) for this month is 0.39%, this strongly suggests that the null hypothesis of having identically independently distributed data should be rejected in favour of accepting the alternative hypothesis of there being a monotonic decreasing trend. Notice that for each month the SL is not greater than 5% and usually less than 1%. Consequently, one would expect that across all of the seasons, a combination test would confirm the presence of an overall decreasing trend. The seasonal Mann-Kendall test statistic in [23.3.11] has a magnitude of -196.0 and a very small significance level. In addition, Table 23.5.4 shows that the significance level of Fisher's combination test in [23.3.30] is also very small. Hence, both of the combination tests indicate that there is an overall trend which is decreasing due to the negative sign of S_g or τ in each season and also S' in [23.3.11] across all of the seasons. As noted earlier, this decreasing trend over the years is also readily apparent in the jittered plot of chloride according to each year in Figure 23.5.3c.

For each of the fourteen water quality variables at each of the five stations where there are sufficient data, the seasonal Mann-Kendall test in [23.3.7] or [23.3.9] is applied. Consider Table 23.5.5 which summarizes the results for chloride. Notice that at Stations 501, 810 and 994, there are obvious decreasing trends (indicated by the negative signs) for all the months for which data are available at all three sites. Except for two cases where the significance level is $a = 10\%$, all

Table 23.5.5. Seasonal Mann-Kendall tests for trend in the chloride series (mg/l) for stations at Long Point Bay, Lake Erie.

MONTHS	STATIONS				
	501	810	994	1085	1086
January					
February					
March					
April	-c	-c	-b		
May	-c	-c	-d		
June	-b	-a	-a		
July	-d	-c	-c		
August	-c	-c	-c		
September	-c	-c	-c		
October	-b	-c	-c		
November	-c	-d	-d		
December					

Note:

1. a, b, c, d denote significance levels of 10, 5, 1, 0.1 percent, respectively.
2. # denotes result not significant at 10 percent.
3. Otherwise, a blank indicates insufficient data.
4. A positive or negative value of tau is indicated by + or -.

of the significance levels are 5% or less. Consequently, these trends are significant. For a given month and station, one should certainly reject the null hypothesis that the chloride data are independently and identically distributed. Table 23.5.5 shows that sufficient data for executing a seasonal Mann-Kendall test are not available for the months of January, February and March at Stations 501, 810 and 994. Also, there are not enough observations for all of the months at Stations 1085 and 1086.

Another water quality variable for which there may be decreasing monthly trends is specific conductance for Stations 501, 810 and 994. However, as is the case for chlorophyll *a*, for most of the water quality variables across most of the months and stations, significant trends are not detected by the seasonal Mann-Kendall test.

As explained earlier, along with the seasonal Mann-Kendall test, for a specified water quality variable and station one can combine the monthly results using the combined score method in [23.3.11] and the Fisherian combination in [23.3.30]. Table 23.5.6 summarizes these two types of combination results for the fourteen water quality variables across all of the stations. When interpreting these results one should keep in mind the limitations of the combination approaches described in Sections 23.3.2 to 23.3.4. Notice in Table 23.5.6 for the chloride variable that the results are highly significant for Stations 501, 810 and 994 for both combination tests (all the significance levels are $d = 0.1\%$). Consequently, there are obvious trends in chloride across the months at all three sites. Due to the negative signs in Table 23.5.5, the trends are decreasing.

Table 23.5.6. Combined score tests from [23.3.11] and Fisher's combination results using [23.3.30] for the 14 water quality variables at Long Point Bay, Lake Erie.

	STATIONS				
	501	810	994	1085	1086
turbidity (FTU)	a,#	#,#	c,b	#,#	#,#
specific conductance (µs/cm)	d,d	d,d	d,d	#,#	#,#
lab pH	#,#	#,#	#,#	#,d	#,#
chloride (mg/*l*)	d,d	d,d	d,d		
ammonia - N (mg/*l*)	d,a	c,b	b,#	#,#	#,#
morganic -N (mg/*l*)	#,#	#,#	#,#	#,d	
filtered total Kjeldahl N	c,#	#,c	a,#	b,#	#,a
Kjeldahl organic N (mg/*l*)	b,#	#,a	#,#	#,#	#,a
chlorophyll a	#,#	b,#	a,#	a,#	#,d
chlorophyll b	c,#	d,a	#,#	#,#	#,#
phytoplankton density					
filtered reactivce phosphate	d,d	b,#	d,c	#,#	#,#
total phosphorous (mg/*l*)	d,a	a,b	d,c	#,#	#,#
iron (mg/*l*)	d,b	#,a	c,b	#,#	#,#

Note:

1. a, b, c, d denote significance levels of 10, 5, 1, 0.1 percent, respectively.
2. # denotes result not significant at the 10 percent level.
3. Otherwise, a blank indicates insufficient data.
4. The combined Kendall test for trend in [23.3.11] is the first entry in each cell while the second entry is the Fisherian combination calculated using [23.3.30].

Wilcoxon Signed Rank Tests

Detailed descriptions of the Wilcoxon signed rank test along with the related confidence interval for the median are given in Appendix A23.2. The purpose of using these tests is to ascertain whether or not paired measurements taken at the same time and alternate depths are significantly different from one another.

Table 23.5.7 displays the statistical analyses of paired measurements for chloride at Station 501. At the top of the table, the availability of paired measurements by month and year is shown. Tukey 5-number summaries for both the depths and measured values at the shallow depths explain how both the depth and measured values are distributed. Likewise, for the deep samples, Tukey 5-number summaries are given for the depths of the deep measurements and the measurements themselves. Notice that the distributions of the measurements appear to be the same at the shallow and deep depths according to the Tukey 5-number summaries. When the paired measurements, denoted by X and Y for the shallow and deep depths, respectively, are subtracted from one another, the Tukey 5-number summary of $X - Y$ shows that the subtracted values are symmetrically distributed about zero. Hence, the depth of the measurement does not appear to affect the distribution of chloride.

Table 23.5.7. Analyses of paired measured chloride (mg/l) values at the same time and alternate depths at Station 501, Long Point Bay, Lake Erie.

Number of paired (shallow, deep) measurements available by month and year

	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	TOTAL
Jan																0
Feb																0
Mar																0
Apr						1	1	1	1	1	1					6
May						2	1	1	2	4	2					12
Jun					1	1		1	1	2						6
Jul					1	2	1	1	2	1						8
Aug					1	1	2	2	1	2						9
Sep					1	1	1	1	2	1						7
Oct					1	1	1	1		2						6
Nov					1		1		2	3						7
Dec						1				1						2
TOTAL	0	0	0	0	6	10	8	8	11	17	3	0	0	0	0	63

X: Shallow Samples

	Depths(m)	Measured Values
Extreme:	1.00	19.4
Hinge:	1.00	20.5
Median:	1.00	21.0
Hinge:	1.00	22.0
Extreme:	1.50	25.0

Y: Deep Samples

	Depths(m)	Measured Values
Extreme:	5.50	19.4
Hinge:	11.00	20.5
Median:	11.50	21.0
Hinge:	11.92	22.0
Extreme:	13.00	25.0

Five-number summary of paired X-Y: -1.0 0.0 0.0 0.0 1.0

Two-sided Wilcoxon signed rank test for paired measured values at the same time and at alternate depths:

Number of pairs is 18. Wilcoxon Statistic is 84.0. SL is 0.96526.
95% confidence interval for the median difference (X-Y) is (0.0000, 0.0000).

The Wilcoxon signed rank test can be employed as a nonparametric test to check whether or not measurements taken at two different depths at exactly the same time have the same median. As noted in Appendix A23.2, the values in a pair are not used if they are equal. As can be seen in Table 23.5.7, there are 18 pairs which do not have equal values. For these 18 pairs of chloride measurements, the value of the Wilcoxon statistic in [A23.2.3] is 84 with a SL of 0.97. Because of this very large SL, one can conclude that the medians or means of the paired chloride measurements taken at alternate depths are not significantly different from one another. Because zero is contained within the 95% confidence interval for $X - Y$, the fact that the medians are the same is further substantiated. Since the Wilcoxon test is only used when the number of paired samples having unequal measurements is greater than 7, results are not shown for some of the water quality variables at different stations.

Table 23.5.8. Summary of the Wilcoxon tests for the equality of medians of paired X and Y, where X is the shallow sample value and Y is the deep sample value.

	Stations				
	501	810	994	1085	1086
turbidity (FTU)	-c	-#	-d		-a
specific conductance (µs/cm)	-d	+#	-a		
lab pH	+c	+#	+#		+b
chloride (mg/*l*)	+#	-#	+b		
ammonia - N (mg/*l*)	-c	+#	-b		+#
inorganic -N (mg/*l*)	-d	-#	-a		+#
filtered total Kjeldahl N	-#	+#	-b		-#
Kjeldahl organic N (mg/*l*)	+#	+#	-b		-#
chlorophyll a	-b	-d			
chlorophyll b	-a	+#			
phytoplankton density					
filtered reactive phosphate	-b	-#	-#		-#
total phosphorous (mg/*l*)	-#	-c	-c		-#
iron (mg/*l*)	-d	+#	-#		-#

Note:

1. a, b, c, d denote significance levels of 10, 5, 1, 0.1 percent, respectively.
2. # denotes result not significant at the 10 percent level.
3. Otherwise, a blank indicates insufficient data.
4. + or - according as the median of X is > or < the median of Y.

Table 23.5.8 presents all of the results of the Wilcoxon tests for all of the variables and stations for which there are sufficient data. Notice that for Stations 501 and 994 the test results for most of the water quality variables are significant. For example, for iron at Station 501 the significance level is $d = 0.1\%$. Hence, one should reject the null hypothesis that the means are the same at the two depths. The negative sign indicates that the mean or median at the shallow depth is less than the mean for samples taken at the deeper depths. At Stations 810 and 1086, the means appear to be the same at the shallow and deep depths for most of the water quality variables.

Kruskal-Wallis Tests

The Kruskal-Wallis test outlined in Appendix A23.3 is a nonparametric test for checking whether or not the means among replicated samples are significantly different from one another. Although the results are not shown here, the one way analysis of variance constitutes a parametric approach for performing the same test but under stricter assumptions.

In Table 23.5.9, the studies for the replicated samples for the total phosphorous data (mg/*l*) at Station 501 are presented. At the top of the table, the number of replicated samples by month and year are given. Below this, the Tukey 5-number summary of all depths is displayed. For a given replicated sample, the range is defined as the largest minus the smallest value. The Tukey 5-number summary for the ranges of all the replicated samples is also given in Table 23.5.9.

Table 23.5.9. Analyses of replicated samples for total phosphorous (mg/l) at Station 501, Long Point Bay, Lake Erie.

Number of replicated samples available by month and year

	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	TOTAL
Jan																0
Feb																0
Mar																0
Apr													4	2		6
May													4	2		6
Jun													4	2		6
Jul													6	1		7
Aug													2			2
Sep													6	2		8
Oct													2	3		5
Nov													4	4		8
Dec													2	1		3
TOTAL	0	0	0	0	0	0	0	0	0	0	0	0	34	17	0	51

Tukey five number summary of all the depths(m) used

1.50	1.50	1.50	11.00	11.50

Tukey five number summary of the ranges of the replicates

0.000	0.002	0.003	0.007	0.110

Kruskal-Wallis Nonparametric Test

KW-Statistic	SL
109.31	7.15×10^{-7}

Note that if all the entries in a given replicated sample were the same and this were true for all the replicated samples, all of the entries in the Tukey 5-number summary would be zero.

The results for the Kruskal-Wallis test are given at the bottom of Table 23.5.9. Using [A23.3.3], the test statistic is found to be 109.31 with a SL of 7.15×10^{-7}. Because of the very small SL, the statistic is significant and hence one could argue that the means for two or more replicated samples are significantly different from one another.

The reader should keep in mind that the differences among the means for the replicated samples could be due to causes such as seasonality, trend and depth. Seasonality may be the main reason for the mean differences but more data would be required to test this hypothesis. The fact that the samples are not independent may also influence the results.

The overall results for when the Kruskal-Wallis test is applied to all the time series having sufficient data across all five sites, are shown in Table 23.5.10. From this table, it can be seen that for iron the means among replicated samples are significantly different from one another for iron across all stations for which there are enough data. However, these differences could be due to causes such as seasonality trend or depth.

Table 23.5.10. Kruskal-Wallis tests for comparing means among replicated samples at Long Point Bay, Lake Erie.

	Stations				
	501	810	994	1085	1086
turbidity (FTU)	a		#		#
specific conductance (μs/cm)	#		#		d
lab pH			#		#
chloride (mg/*l*)					
ammonia - N (mg/*l*)	#		a	#	#
inorganic -N (mg/*l*)			#		#
filtered total Kjeldahl N	c		#	a	#
Kjeldahl organic N (mg/*l*)	#		a	#	b
chlorophyll a	#				
chlorophyll b	#				
phytoplankton density	#	#	#		
filtered reactive phosphate			#		#
total phosphorous (mg/*l*)	d		d	d	c
iron (mg/*l*)	d		d	d	d

Note:

1. a, b, c, d denote significance levels of 10, 5, 1, 0.1 percent, respectively.
2. # denotes result not significant at the 10 percent level.
3. Otherwise, a blank indicates insufficient data.

23.6 CONCLUSIONS

Environmental data, such as water quality time series, are often very messy. For example, water quality time series may possess problems which include having missing observations, following nonnormal distributions, possessing outliers, and being short in length. Because nonparametric tests usually have less restrictive assumptions than their parametric counterparts, nonparametric tests are often ideally suited for detecting characteristics such as trends in environmental data (Helsel, 1987). Furthermore, because of the increasing importance of environmental impact assessment studies in modern day society, the import of both nonparametric and parametric tests will continue to expand.

Following a general discussion of statistical testing in Section 23.2, in Section 23.3 a number of useful nonparametric tests are described for detecting trends in data sets. In particular, the seasonal Mann-Kendall test and the correlated seasonal Mann-Kendall test of Section 23.3.2 constitute important intrablock methods for discovering trends in time series for which there may be missing observations. However, the aligned rank technique discussed in the latter part of Section 23.3.2 must be used with an evenly spaced time series.

When dealing with seasonal data measured at one or more sites, the procedures described in Section 23.3.3 can be used for grouping data together when checking for the presence of trends. For instance, if it is suspected that there is an increasing trend during the summer seasons and a decreasing trend at other times of the year, the data can be subdivided into the summer and non-summer groups. A good way to combine tests of hypotheses across seasons or groups of seasons is to employ Fisher's method given in [23.3.30] in Section 23.3.4. Besides grouping the data, sometimes it is worthwhile to first filter the given time series in order to account for the effect of water quantity upon water quality. One particular approach for filtering or preprocessing data before they are subjected to statistical testing is briefly outlined in Section 23.3.5 while another procedure is presented in Section 24.3.2. Moreover, the Spearman partial rank correlation test of Section 23.3.6 provides a more flexible approach for filtering out undesirable effects when checking for trends over time in a series.

The only nonparametric test that is used to model a step trend due to a known intervention is described in Section 23.3.7. For a given physical variable, this test can be used to confirm the presence of a step trend across seasons at one or more measuring locations. However, it cannot be used to test for the presence of a trend which does not take place as a step change in the mean level after the date of occurrence of a known intervention. Recall from Section 22.4 and Chapter 19 that intervention analysis can be used to model a wide spectrum of trend shapes caused by one or more interventions and also accurately estimate the magnitudes of the trends.

The ACF at lag one is a parametric test which can be used for detecting trends in a data set. On the other hand, Kendall's tau in [23.3.5] or, equivalently, the Mann-Kendall statistic in [23.3.1] or [23.3.7], constitute statistics that can be used in the nonparametric Mann-Kendall test for trend detection. The simulation experiments executed in Section 23.4 demonstrate that the ACF at lag one is more powerful than Kendall's tau for discovering purely stochastic trends while Kendall's tau is more powerful for uncovering purely deterministic trends.

Often water quality and other types of time series are multiple censored. In order to be able to apply nonparametric tests to multiple censored data, one can employ procedures described in Section 23.3.8.

To clearly demonstrate the efficacy of employing nonparametric and also parametric methods in a complex environmental impact assessment study, the effects of industrial development upon water quality in Long Point Bay in Lake Erie, are systematically examined in Section 23.5. The specific statistical methods used in the application for exploratory and confirmatory data analyses are listed in Table 23.5.1. Within Section 23.5.2, the method of application and representative results are given for each of the techniques marked by a cross in Table 23.5.1. Of particular importance is the seasonal Mann-Kendall test that is used to check for the presence of trends in a range of water quality variables at different sites (see Tables 23.5.4 to 23.5.6). Other nonparametric tests utilized in the study include the Wilcoxon signed rank and Kruskal-Wallis tests described in Appendices A23.2 and A23.3, respectively. When deciding upon which tests to employ in a given study, it is informative to refer to tables that characterize statistical methods according to various criteria. Table 23.1.1 summarizes the purpose of all of the nonparametric tests described in Chapter 23.

Beyond applications given in Sections 23.5.2 and 24.3, as well as references already cited in this chapter, there is, of course, other published literature dealing with nonparametric modelling in water resources and environmental engineering. For example, Fox et al. (1990), Potter (1991), and El-Shaarawi and Niculescu (1993) apply nonparametric trend tests to water quantity problems, El-Shaarawi et al. (1983, 1985), Smith et al. (1987), Alexander and Smith (1988), Karlsson et al. (1988), Loftis and Taylor (1989), Lettenmaier et al. (1991), Sanden et al. (1991), Walker (1991), Zetterqvist (1991), and Tsirkunov et al. (1992), employ nonparametric tests for detecting trends in water quality time series, and Harned and Davenport (1990), Wiseman et al.

(1990), Jordan et al. (1991), and Stanley (1993) apply nonparametric trend tests to estuarine data. In fact, assessment of water quality is of national concern to many countries throughout the world, including the United States of America (see, for instance, Cohen et al., 1988). Consequently, the need for further developing flexible nonparametric trend tests, as well as many other kinds of statistical procedures, will continue to expand.

In the last row of Table 23.5.1, it is noted that regression analysis is used to model some problems related to the water quality study on Lake Erie reported in Section 23.5. Regression analysis is, in fact, a very flexible and general tool that has wide applicability in water resources and environmental engineering. Consequently, in the next chapter various types of regression models are put forward for use as exploratory and confirmatory data analysis tools. Additionally, an overall methodology is presented for systematically carrying out a trend assessment study with messy water quality data measured in a river. When dealing with water quality data from a river, the flow levels must be accounted for and regression analysis provides a superb means for doing this. As is shown in Chapter 24, regression analysis techniques, along with some nonparametric trend tests as well as other statistical methods, play a key role in this methodology.

APPENDIX A23.1

KENDALL RANK CORRELATION TEST

The *Kendall rank correlation test* (Kendall, 1975) is a nonparametric test for checking if two series are independent of one another. The null hypothesis, H_0, is that the two series are independent of each other while the alternative hypothesis, H_1, is they are not independent.

Suppose that the data consist of a bivariate random sample if size n, (x_i,y_i), for $i = 1,2,\ldots,n$. Two observations are concordant if both members of one bivariate observation are larger than their respective members of the other observation. For example, the two bivariate observations (3.2,9.6) and (4.7,11.2) are concordant. Out of the $\binom{n}{2}$ total possible pairs, let N_c denote the number of concordant pairs of observations. A pair of bivariate observations, such as (5.2,8.6) and (4.3,12.4), is discordant if it is not concordant. Let N_d be the total number of discordant pairs. Under H_0, the test statistic for the Kendall rank correlation test is

$$\tau = \frac{N_c - N_d}{\frac{1}{2}n(n-1)} \qquad [A23.1.1]$$

If all pairs are concordant, the two series are perfectly correlated and $\tau = 1$. For the case of total discordance, $\tau = -1$. Consequently, τ varies between -1 and +1. Because τ is asymptotically normally distributed and its distribution can be tabulated exactly for small n, one can determine the SL for a computed value of τ. If the calculated τ is greater than or less than 0.05, one can accept or reject, respectively, the null hypothesis. Theoretical results regarding this test are provided by Valz et al. (1994).

Notice that the symbol for tau used in [A23.1.1] is identical to that used in [23.3.5] in Section 23.3.2 for the Mann-Kendall trend test. This is because the test statistic given in [23.3.5] is a special case of the test statistic for the Kendall rank correlation test in [A23.1.1]. To obtain [23.3.5] from [A23.1.1], simply replace (x_i, y_i) by (t, x_t) for which time $t = 1,2, \ldots, n$, and x_t consists of $x_1, x_2, \ldots, x_n$.

APPENDIX A23.2

WILCOXON SIGNED RANK TEST

Wilcoxon Test: In water quality modelling, one may wish to know whether or not measurements taken at two different depths at exactly the same time possess the same median. Moreover, as explained and demonstrated in Sections 8.3 and 15.3, the Wilcoxon signed rank test can be employed for checking whether one time series models forecasts significantly better than another. A test proposed by Wilcoxon (1945) and described in detail by Conover (1980, pp. 280-288) can be used with paired data. Let the data consist of n' pairs of observations $(x_1,y_1),(x_2,y_2), \ldots, (x_{n'},y_{n'})$ generated by their respective bivariate random variables $(X_1,Y_1),(X_2,Y_2), \ldots, (X_{n'},Y_{n'})$. For each of the n' pairs, (X_i,Y_i), the absolute differences can be computed using

$$|D_i| = |Y_i - X_i| \quad i = 1,2, \ldots, n' \qquad [A23.2.1]$$

In the test, pairs for which $X_i = Y_i$ and hence $D_i = 0$ are omitted. Let $n \leq n'$ denote the number of remaining pairs. The n pairs can then be ranked from 1 to n where rank 1 is given to the pair with the smallest $|D_i|$ and rank n is assigned to the pair with the largest $|D_i|$. When there are ties among the absolute differences for a set of paired values, each of the pairs in the set is assigned the average of the ranks that otherwise would have been assigned. As pointed out by Conover (1971, p. 281), the assumptions underlying the Wilcoxon test are each D_i is a continuous random variable, the distribution of each D_i is symmetric, the D_i's are mutually independent, all of the D_i's possess the same median, and the measurement scale of the D_i's is at least interval.

Let d_{50} be the median of the D_i's. For a two-tailed Wilcoxon test, the null hypothesis is

$$H_0: d_{50} = 0$$

This implies that the medians of the X_i's and Y_i's are the same. The alternative hypothesis is that the medians of the X_i's and Y_i's are different. This can be written as

$$H_1: d_{50} \neq 0$$

Because the distribution of each D_i is assumed to be symmetric, the median is identical to the mean and one is also testing whether or not the means of the X_i's and Y_i's are the same.

The test statistic, T, used to decide if H_0 should be accepted or rejected is defined to be the sum of the ranks assigned to those pairs (X_i,Y_i) where Y_i exceeds X_i. Therefore, for each pair (X_i,Y_i)

$$R_i = \begin{cases} 0 & \text{if } X_i > Y_i \\ \text{rank assigned to } (X_i,Y_i) & \text{if } X_i < Y_i \end{cases} \quad \text{[A23.2.2]}$$

and the test statistic is written as

$$T = \sum_{i=1}^{n} R_i \quad \text{[A23.2.3]}$$

Because the exact distribution of T is known, one can easily calculate the SL for T (Conover, 1971, pp. 211-215, p. 383). If, for example, the value of T is either sufficiently large or small enough to cause the significance level to be less than say 5%, one can reject H_0 and thereby assume that the medians of the X_i's and Y_i's are different.

Confidence Interval for the Median: The Wilcoxon signed rank test is employed to check whether or not the median of the X_i's and Y_i's are significantly different from one another. In the test, one actually checks whether or not the median of D_i is significantly different from zero. To obtain an estimate of the magnitude of the unknown median of the D_i's, one can calculate a confidence interval for this median using the method given by Tukey (1949) and also described by Walker and Lev (1953, p. 445) and Conover (1980, pp. 288-290). Moreover, the confidence interval for the median is also a confidence interval for the mean difference

$$E(D_i) = E(Y_i) - E(X_i)$$

if (X_i,Y_i) or D_i, $i = 1,2,\ldots,n$, is a random sample and if the mean difference exists.

To calculate a confidence limit for D_i first one must select a significance level α which means the confidence interval is $1-\alpha$. From the tables for the exact distribution of D_i (Conover, 1980, p. 460-461) one can obtain the $\alpha/2$ quantile denoted by $\omega_{\alpha/2}$. Next, determine the $n(n+1)/2$ possible averages $(D_i + D_j)/2$ for all i and j, including $i = j$. The upper and lower bounds for the $1-\alpha$ confidence interval are given by the $w_{\alpha/2}$th largest of the averages and the $w_{\alpha/2}$th smallest of the averages, respectively. Because of this, one only has to compute the averages near the largest and smallest D_i's and not the entire $n(n+1)/2$ averages.

APPENDIX A23.3

KRUSKAL-WALLIS TEST

The Kruskal-Wallis test constitutes a nonparametric approach for checking whether or not the distributions or means across k samples are the same (Kruskal and Wallis, 1952). In contrast to the normality assumption for the one way analysis of variance (ANOVA), the k population distributions are only assumed to be identical for the Kruskal-Wallis test. However, as is the case for the one way ANOVA, the observations are assumed to be independent of another.

Suppose that there are k samples of sizes n_i, $i = 1,2, \ldots, k$, having a combined sample size n defined by

$$\sum_{i=1}^{k} n_k = n \qquad [A23.3.1]$$

Let x_{ij} stand for the jth value in the ith sample so that $i = 1,2, \ldots, k$, and $j = 1,2, \ldots, n_i$. Denote the ith random sample of size n_i by $x_{i1}, x_{i2}, \ldots, x_{in_i}$. Rank the n observations from 1 to n where ranks 1 and n are assigned to the smallest and largest observations, respectively. For tied observations, assign each observation the average of the ranks that would be assigned to the observations. Let $R(x_{ij})$ represent the rank assigned to x_{ij}. For the ith sample, the sum of the ranks is given by

$$R_i = \sum_{j=1}^{n_i} R(x_{ij}) \quad i = 1,2, \ldots, k \qquad [A23.3.2]$$

The null hypothesis is that all of the k population distribution functions are identical. This implies that the k means are the same. The alternative hypothesis is that at least one of the populations yields larger observations than at least one of the other populations. Therefore, the k populations do not all have identical means.

The test statistic for the Kruskal-Wallis test is

$$T = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{[R_i - (1/2)n(n+1)]^2}{n_i} \qquad [A23.3.3]$$

where n and R_i are defined in [A23.2.1] and [A23.2.2], respectively. The exact distribution of T is known (Kruskal and Wallis, 1952) and for small samples (say $k = 3$ and $n_i \leq 5$, $i = 1,2,3$) one can obtain the SL for the observed T from tables. For larger samples, T is approximately χ^2 distributed on $k - 1$ degrees of freedom.

Compared to the usual parametric F test used in the one way ANOVA, the Kruskal-Wallis test is very efficient. For example, when the assumptions of the F test are satisfied, the asymptotic relative efficiency of the Kruskal-Wallis test compared to the F test is 0.955. A detailed description of the Kruskal-Wallis test is presented by Conover (1980, pp. 229-237).

The Kruskal-Wallis test can be employed for testing whether or not a time series is seasonal. To apply the test, the data contained within a given season is considered as one separate sample. For example, when dealing with monthly observations, each of the twelve samples would consist of the monthly data from that month across all of the years. A significantly large value of the Kruskal-Wallis statistic would mean that the means and perhaps other distributional parameters vary across the seasons. On the other hand, if the Kruskal-Wallis statistic were not significant, this would mean that the data are not seasonal.

PROBLEMS

23.1 In Section 23.3.2, the correlated seasonal Mann-Kendall statistic is presented as a nonparametric approach for use in trend detection. Describe some research projects that you think would improve this nonparametric test from both theoretical and practical viewpoints when it is used to discover trends in seasonal water quality data.

23.2 Compare the relative advantages and disadvantages of using parametric and nonparametric tests in trend detection and modelling. If you address points that are raised in Chapter 23, give more depth to your explanations than those presented in this chapter. Also, explain some new points of comparison which are not addressed in Chapter 23. Properly reference the sources of your information.

23.3 In Section 23.3.3, procedures are given for grouping seasons for use in trend detection. Compare the relative advantages and disadvantages of the different approaches. Outline a method of grouping which is not given in Section 23.3.3.

23.4 Without referring to the published literature, prove that the Mann-Kendall test statistic S in [23.3.1] is asymptotically normally distributed with the mean and variance given in [23.3.2]. Why is it necessary to know the distribution of S?

23.5 How can one calculate the exact distribution of the Mann-Kendall test statistic, S, in small samples?

23.6 Select a nonseasonal time series which you suspect may contain a trend. Using the Mann-Kendall test, carry out a formal hypothesis test to ascertain if your suspicions are correct. Comment upon your results.

23.7 Outline how the covariance eigenvalue method of Lettenmaier (1988) works for handling correlation among seasons when employing the correlated seasonal Mann-Kendall test. Describe the advantages and drawbacks of this technique when compared to its competitors.

23.8 Choose a seasonal environmental time series that is of interest to you. Employ the seasonal Mann-Kendall test to check for the presence of trends. Be sure to determine how a trend behaves separately within each season and only group seasons together in a meaningful way when carrying out the trend test across seasons. Be certain to emphasize the most interesting results when explaining your findings.

23.9 Using equations, outline the four methods of combining independent tests of hypothesis which are compared by Littell and Folks (1971). Summarize the advantages and drawbacks of each of the four approaches.

23.10 By referring to the research of Hirsch et al. (1982) and Smith et al. (1982), use equations to outline the procedure for determining the flow adjusted concentration (FAC) for a water quality variable. Appraise their approach for calculating the FAC's.

23.11 By employing a seasonal environmental time series, demonstrate how the Spearman partial rank correlation test is used for trend detection when the effects of seasonality are partialled out. Clearly, explain how your calculations are carried out when applying this test and comment upon any interesting results.

23.12 Explain how the Spearman partial rank correlation test can take into account the effects of correlation when checking for the presence of a trend in a time series.

23.13 Design comprehensive simulation experiments to compare the powers of the seasonal Mann-Kendall and Spearman partial rank correlation tests for trend detection.

23.14 Carry out a sensible portion of the simulation experiments designed in the previous problem.

23.15 Define the Kendall partial rank correlation statistic for three variables labelled as X, Y and Z. Explain how this statistic could be utilized in trend tests. Discuss the advantages and disadvantages of using the Kendall partial rank correlation coefficient for discovering trends.

23.16 Define the Pearson partial rank correlation coefficient for three variables denoted as X, Y and Z. Explain various ways in which this statistic could be used for trend tests. Comment upon the advantages and drawbacks of employing the Pearson partial rank correlation coefficient for trend detection.

23.17 In Section 23.3.7 it is noted that Hirsch and Gilroy (1985) employ a filter for determining a filtered sulphate loading series which can then be tested for the presence of a step trend which started at a known intervention date. Using equations, outline how this filter works and explain its advantages and limitations.

23.18 Using equations, explain how the nonseasonal Mann-Kendall test is applied to a multiple censured data set by employing the expected rank vector approach of Hughes and Millard (1988). Using either an actual or hypothetical time series having at best three levels of censoring on the left, demonstrate how the trend test is carried out.

23.19 Execute the instructions of the previous problem for the case of a seasonal time series.

23.20 Define one more deterministic trend model and one more stochastic trend model which could have been used in the simulation studies for trend detection in Section 23.4. Explain the reasons for your choice of models.

23.21 Define a mixed deterministic-stochastic trend model which could be employed in the simulation studies for trend detection presented in Section 23.4. Justify the reasons for your choice of a mixed model and explain why the authors did not use a mixed model in their simulation experiments.

23.22 In Section 23.4, simulation experiments are used to ascertain the ability of the ACF at lag one, r_1, in [23.4.1], and Kendall's tau in [23.3.5] to detect deterministic and stochastic trends. Define one other parametric statistic and another nonparametric statistic which could have been used in the simulation studies. Explain why you selected these statistics and compare them to those used in Section 23.4.

23.23 Prove that the ACF at lag one, r_1, in [23.4.1] is asymptotically NID$(0,\frac{1}{n})$ where n is the sample size.

23.24 A threshold autoregressive model is defined in [23.4.15]. Explain how this model is fitted to a given time series by following the three stages of model construction.

23.25 As noted in Section 23.5.2, both the Kruskal-Wallis test and one way analysis of variance can be used for checking whether or not the means among replicated samples are significantly different from one another. After briefly outlining how each test is designed, compare the merits and drawbacks of the two methods.

23.26 Outline the approach of Montgomery and Reckhow (1984) for carrying out a trend assessment study. Compare their procedure to methodology employed in Section 23.5.

23.27 Summarize the procedure of Hirsch et al. (1991) for selecting methods to employ for detecting and estimating trends in water quality time series.

23.28 Suppose that a government agency gives you a set of environmental time series to examine for the presence of trends. Briefly describe how you would decide upon what to do.

23.29 Obtain some environmental time series which are suspected of possessing trends. List the exploratory and confirmatory data analysis tools which you plan to use and then execute a comprehensive data analysis study.

23.30 Conover and Iman (1981) describe a valuable procedure for linking parametric and nonparametric statistics. Outline how this connection is carried out and summarize the advantages of the approach. Be sure to explain how their procedure can enhance regression analysis, which is the topic of the next chapter.

REFERENCES

BOOKS ON NONPARAMETRIC STATISTICAL METHODS

Bradley, J. V. (1968). *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, New Jersey.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. John Wiley, New York, second edition.

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. Wiley, New York.

Gibbons, J. D. (1971). *Nonparametric Statistical Inference*. McGraw-Hill, New York.

Gibbons, J. D. (1976). *Nonparametric Methods for Quantitative Analysis*. Holt, Rinehart and Winston, New York.

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

Hipel, K. W., editor (1988). *Nonparametric Approaches to Environmental Impact Assessment*. American Water Resources Association (AWRA), AWRA Monograph, Series No. 10, Bethesda, Maryland.

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

Kendall, M. G. (1975). *Rank Correlation Methods*. Charles Griffin, London, fourth edition.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, Oakland, California.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Siegel, S. (1956). *Nonparametric Statistics for Behavioral Sciences*. McGraw-Hill, New York.

CENSORED DATA

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, pages 88-350.

El-Shaarawi, A. H. (1989). Inferences about the mean from censored water quality data. *Water Resources Research*, 25(4):685-690.

Gilbert, R. O. and Kinnison, R. R. (1981). Statistical methods for estimating the mean and variance from radionuclide data sets containing negative, unreported or less-than values. *Health Physics*, 40:377-390.

Gilliom, R. J. and Helsel, D. R. (1986). Estimation of distributional parameters for censored trace level water quality data, 1. Estimation techniques. *Water Resources Research*, 22(2):135-146.

Gilliom, R. J., Hirsch, R. M., and Gillroy, E. J. (1984). Effect of censoring trace-level water-quality data on trend-detection capability. *Environmental Science Technology*, 18:530-535.

Gleit, A. (1985). Estimation for small normal data sets with detection limits. *Environmental Science Technology*, 19:1201-1206.

Helsel, D. R. and Cohn, T. A. (1988). Estimation of descriptive statistics for multiple censored water quality data. *Water Resources Research*, 24(12):1997-2004.

Helsel, D. R. and Gilliom, R. J. (1986). Estimation of distributional parameters for censored trace level water quality data, 2. Verification and applications. *Water Resources Research*, 22(2):147-155.

Hirsch, R. M. and Stedinger, J. R. (1987). Plotting positions for historical floods and their precision. *Water Resources Research*, 23(4):715-727.

Hughes, J. and Millard, S. P. (1988). A tau-like test for trend in the presence of multiple censoring points. *Water Resources Bulletin*, 24(3):521-531.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley, New York.

Kushner, E. J. (1976). On determining the statistical parameters for pollution concentration from a truncated data set. *Atmospheric Environment*, 10:975-979.

Lee, E. T. (1980). *Statistical Methods for Survival Data Analysis*. Lifetime Learning, Belmont, California.

Millard, S. P. and Deverel, S. J. (1988). Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. *Water Resources Research*, 24(12):2087-2098.

Miller, R. G. (1981). Survival Analysis. John Wiley, New York.

Owen, W. J. and DeRouen, T. A. (1980). Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. *Biometrics*, 36:707-719.

Porter, P. S. and Ward, R. C. (1991). Estimating central tendency from uncensored trace level measurements. *Water Resources Bulletin*, 27(4):687-700.

ENCYCLOPEDIAE, HANDBOOK AND DICTIONARY ON STATISTICS

Kendall, M. G. and Buckland, W. R. (1971). *A Dictionary of Statistical Terms*. Longman Group Limited, Thetford, Norfolk, Great Britain, third edition.

Kotz, S. and Johnson, N. L., editors (1988). *Encyclopedia of Statistical Sciences, Volumes 1 to 9*. Wiley, New York.

Kruskal, W. H. and Tanur, J. M. (1978). *International Encyclopedia of Statistics, Volumes 1 and 2*. The Free Press, New York.

Sachs, L. (1984). *Applied Statistics, A Handbook of Techniques*. Springer-Verlag, New York, second edition.

FREQUENCY DOMAIN TREND TESTS

Bloomfield, P. (1992). Trends in global temperature. *Climatic Change*, 21:1-16.

Bloomfield, P. and Nychka, D. (1992). Climate spectra and detecting climate change. *Climatic Change*, 21:275-287.

Bloomfield, P., Oehlert, G., Thompson, M. L. and Zeger, S. (1983). A frequency domain analysis of trends in Dobson total ozone records. *Journal of Geophysical Research*, 88(C13):8512-8522.

NONPARAMETRIC TREND TESTS

Best, D. J. and Gipps, P. G. (1974). The upper tail probabilities of Kendall's tau. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 23(1):98-100.

Crawford, C. G., Hirsch, R. M., and Slack, J. R. (1983). Nonparametric tests for trends in water quality data using the statistical analysis system (SAS). Technical Report U. S. Geological Survey Open-File Report 83-550, U. S. Government.

Dietz, E. J. and Killeen, T. J. (1981). A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*, 76(373):169-174.

El-Shaarawi, A. H. and Niculescu, S. P. (1992). On Kendall's Tau as a test of trend in time series data. *Enviornmetrics*, 3(4):385-411.

El-Shaarawi, A. H. and Niculescu, S. P. (1993). A simple test for detecting non-linear trend. *Environmetrics*, 4(2):233-242.

Farrell, R. (1980). Methods for classifying changes in environmental conditions. Technical Report VRF-EP A7.4-FR80-1, Vector Research Inc., Ann Arbor, Michigan.

Hipel, K. W., McLeod, A. I., and Fosu, P. K. (1986). Empirical power comparisons of some tests for trend. In El-Shaarawi, A. H. and Kwiatkowski, R. E., editors, *Statistical Aspects of Water Quality Monitoring*. Elsevier, Amsterdam.

Hirsch, R. M. (1988). Statistical methods and sampling design for estimating step trends in surface-water quality. *Water Resources Bulletin*, 24(3):493-503.

Hirsch, R. M. and Gilroy, E. J. (1985). Detectability of step trends in the rate of atmospheric deposition of sulphate. *Water Resources Bulletin*, 21(5):773-784.

Hirsch, R. M. and Slack, J. R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20(6):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*, 18(1):107-121.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41:133-145.

Lettenmaier, D. P. (1976). Detection of trends in water quality data from records with dependent observations. *Water Resources Research*, 12(5):1037-1046.

Lettenmaier, D. P. (1988). Multivariate nonparametric tests for trend in water quality. *Water Resources Bulletin*, 24(3):505-512.

Loftis, J. C., Taylor, C. H., and Chapman, P. L. (1991a). Multivariate tests for trend in water quality. *Water Resources Research*, 27(7):1419-1429.

Loftis, J. C., Taylor, C. H., Newell, A. D. and Chapman, P. L. (1991b). Multivariate trend testing of lake water quality. *Water Resources Research*, 27(3):461-473.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13:245-259.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, 58:216-230.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63:1379-1389.

Smith, R. A., Hirsch, R. M., and Slack, J. R. (1982). A study of trends in total phosphorous measurements at stations in the NASQAN network. Technical Report Water Supply Paper 2190, U. S. Geological Survey, Reston, Virginia.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15:88-97.

Taylor, C. H. and Loftis, J. C. (1989). Testing for trend in lake and ground water quality time series. *Water Resources Bulletin*, 25(4):715-726.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, 1, 2, and 3. *Ned. Akad. Wetensch Proc.*, 53:386-392, 521-525, 1397-1412.

Valz, P. (1990). *Developments in Rank Correlation Procedures*. PhD thesis, Department of Statistical and Actuarial Sciences. The University of Western Ontario, London, Ontario, Canada.

Valz, P. D., McLeod, A. I. and Thompson, M. E. (1994). Cumulant generating function and tail probability approximations for Kendall's score with tied rankings. *The Annals of Statistics*.

Van Belle, G. and Hughes, J. P. (1984). Nonparametric tests for trend in water quality. *Water Resources Research*, 20(1):127-136.

Zetterqvist, L. (1988). Asymptotic distribution of Mann's test for trend for m-dependent seasonal observations. *Scandinavian Journal of Statistics*, 15:81-95.

OTHER NONPARAMETRIC TESTS

Conover, W. J. and Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3):124-129.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks on one-criterion variance analysis. *Journal of the American Statistical Association*, 47:583-621.

Tukey, J. W. (1949). The simplest signed-rank tests. Mimeographed report No. 17, Statistical Research Group, Princeton University.

Walker, H. M. and Lev, J. (1953). *Statistical Inference*. Holt, New York.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1:80-83.

SERIAL CORRELATION COEFFICIENT AT LAG ONE

Cox, D. R. (1966). The null distribution of the first serial correlation coefficient. *Biometrika*, 53:623-626.

Dufuor, J. M. and Roy, R. (1985). Some robust exact results on sample autocorrelation and tests of randomness. Technical report, Department of Data Processing and Operations Research, University of Montreal, Montreal, Quebec, Canada.

Kendall, M. G., Stuart, A., and Ord, J. K. (1983). *The Advanced Theory of Statistics, Volume 3*. Griffin, London.

Knoke, J. D. (1975). Testing for randomness against autocorrelated alternatives: The parametric case. *Biometrika*, 62:571-575.

Knoke, J. D. (1977). Testing for randomness against autocorrelation: Alternative tests. *Biometrika*, 64:523-529.

Noether, G. E. (1950). Asymptotic properties of the Wald-Wolfowicz test of randomness. *Annals of Mathematical Statistics*, 21:231-246.

Wald, A. and Wolfowicz, J. (1943). An exact test for randomness in the nonparametric case based on serial correlation. *Annals of Mathematical Statistics*, 14:378-388.

SIMULATION MODELS

Barnard, G. A. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society, Series B*, 21:239-244.

Cleary, J. A. and Levenbach, H. (1982). *The Professional Forecaster*. Lifetime Learning Publications, Belmont, California.

Cochran, W. G. (1977). *Sampling Techniques*. Third edition, Wiley, New York.

Tong, H. (1977). Discussion of paper by A. J. Lawrence and N. T. Kottegoda. *Journal of the Royal Statistical Society, Series A*, 140:34-35.

Tong, H. (1978). On a threshold model. In Chen, C. H., editor, *Pattern Recognition and Signal Processing*. Sijthoff and Noordhoff, The Netherlands.

Tong, H. (1983). Threshold models in non-linear time series analysis. *Lecture Notes in Statistics*, No. 21, Springer-Verlag, New York.

Tong, H. and Lim, S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42:245-292.

Tong, H., Thanoon, B., and Gudmundsson, G. (1985). Threshold time series modelling of two Icelandic riverflow systems. *Water Resources Bulletin*, 21(4):651-661.

STATISTICS

Bahadur, R. R. (1967). An optimal property of the likelihood ratio statistic. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, I:13-26.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburg, England.

Fisher, R. A. (1973). *Statistical Methods and Scientific Inference*. 3rd edition, Oliver and Boyd, Edinburg.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. Wiley, New York, second edition.

Hipel, K. W. (Editor) (1985). *Time Series Analysis in Water Resources*. American Water Resources Association, Bethesda, Maryland.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral Analysis and its Applications*. Holden-Day, San Francisco.

Littell, R. C. and Folks, J. L. (1971). Asymptotic optimality of Fisher's method for combining independent tests. *Journal of the American Statistical Association*, 66:802-806.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika, Series A*, 20:175-240, 263-294.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A*, 231:289-337.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika*, 31:9-12.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103:677-680.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.

Zar, J. H. (1974). *Biostatistical Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

TREND ASSESSMENT METHODOLOGIES

Berryman, D., Bobee, B., Cluis, D., and Haemmerli, J. (1988). Nonparametric tests for trend detection in water quality time series. *Water Resources Bulletin*, 24(3):545-556.

Harcum, J. B., Loftis, J. C. and Ward, R. C. (1992). Selecting trend tests for water quality series with serial correlation and missing values. *Water Resources Bulletin*, 28(3):469-478.

Hipel, K. W., McLeod, A. I., and Weiler, R. R. (1988). Data analysis of water quality time series in Lake Erie. *Water Resources Bulletin*, 24(3):533-544.

Hipel, K. W. and McLeod, A. I. (1989). Intervention analysis in environmental engineering. *Environmental Monitoring and Assessment*, 12:185-201.

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*, 27(5):803-813.

McLeod, A. I., Hipel, K. W., and Bodo, B. A. (1991). Trend analysis methodology for water quality time series. *Environmetrics*, 2(2):169-200.

Montgomery, R. H. and Reckhow, K. H. (1984). Techniques for detecting trends in lake water quality. *Water Resources Bulletin*, 20(1):43-52.

WATER QUALITY AND QUANTITY

Water quality applications are also given in published literature listed under other categories for the references in this chapter.

Alexander, R. B. and Smith, R. A. (1988). Trends in lead concentrations in major U. S. rivers and their relationship to historical changes in gasoline-lead consumption. *Water Resources Bulletin*, 24(3):557-569.

Bodo, B. and Unny, T. E. (1983). Sampling strategies for mass-discharge estimation. *Journal of Environmental Engineering, American Society of Civil Engineers*, 109(4):812-829.

Borman, F. H., Likens, G. E., Siccama, T. G., Pierce, R. S., and Eaton, J. S. (1974). The export of nutrients and recovery of stable conditions following deforestation at Hubbard Brook. *Ecological Monographs*, 44:255-277.

Cohen, P., Alley, W. M., and Wilber, W. G. (1988). National water-quality assessment: Future directions of the U. S. Geological Survey. *Water Resources Bulletin*, 24(6):1147-1151.

El-Shaarawi, A. H., Esterby, S. R., and Kuntz, K. W. (1983). A statistical evaluation of trends in the water quality of the Niagara River. *Journal of Great Lakes Research*, 9(2):234-240.

El-Shaarawi, A. H., Esterby, S. R., Warry, N. D., and Kuntz, K. W. (1985). Evidence of contaminant loading to Lake Ontario from the Niagara River. *Canadian Journal of Fisheries and Aquatic Sciences*, 42:1278-1289.

Fox, J. P., Mongan, T. R., and Miller, W. J. (1990). Trends in freshwater inflow to San Francisco Bay from the Sacramento-San Joaquin Delta. *Water Resources Bulletin*, 26(1):101-116.

Harned, D. A., Daniel III, C. C., and Crawford, J. K. (1981). Methods of compensation as an aid to the evaluation of water quality trends. *Water Resources Research*, 17(5):1389-1400.

Harned, D. A. and Davenport, M. S. (1990). Water-quality trends and basin activities and characteristics for the Albemarle-Pamlico estuarine system, North Carolina and Virginia. *U. S. Geological Survey*, Open File Report, 90-398.

Helsel, D. R. (1987). Advantages of nonparametric procedures for analysis of water quality data. *Journal of Hydrological Sciences*, 32(2):179-190.

Hobbie, J. E. and Likens, G. E. (1973). Output of phosphorous, dissolved organic carbon and fine particulate carbon from Hubbard Brook watersheds. *Limnology Oceanography*, 18(5):734-742.

Johnson, N. M., Likens, G. E., Bormann, F. H., Fisher, D. W., and Pierce, R. S. (1969). A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire. *Water Resources Research*, 5(6):1353-1363.

Jordan, T. E. and Correll, D. L., Miklar, J. and Weller, D. E. (1991). Long-term trends in estuarine nutrients and chlorophyll, and short-term effects of variation in water-shed discharge. *Mar. Ecol. Prog. Ser.*, 75:121-132.

Karlsson, G., Grimvall, A. and Lowgren, M. (1988). Riverbasin perspective on long-term changes in the transport of nitrogen and phosphorous. *Water Research*, 22:139-149.

Langbein, W. B. and Dawdy, D. R. (1964). Occurrence of dissolved solids in surface waters in the United States. Professional Paper 501-D, D115-D117, U. S. Geological Survey.

Lettenmaier, D. P., Hooper, E. R., Wagoner, C., and Faris, K. B. (1991). Trends in stream quality in the continental United States, 1978-1987. *Water Resources Research*, 27(3):327-339.

Loftis, J. C. and Taylor, C. H. (1989). Detecting acid precipitation impacts on lake water quality. *Environmental Management*, 13:529-539.

Potter, K. W. (1991). Hydrological impacts of changing land management practices in a moderate-sized agricultural catchment. *Water Resources Research*, 27(5):845-855.

Reckhow, K. H. (1978). Quantitative techniques for the assessment of lake quality. Technical report, Department of Resource Development, Michigan State University.

Sanden, P., Rahm, L., and Wulff, F. (1991). Non-parametric trend test of Baltic Sea data. *Environmetrics*, 2(3):263-278.

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

Stanley, D. W. (1993). Long-term trends in Pamlico River estuary nutrients, Chlorophpyll, dissolved oxygen, and watershed nutrient production. *Water Resources Research*, 29(8):2651-2662.

Tsirkunov, V. V., Nikanorov, A. M., Laznik, M. M. and Zhu, D. (1992). Analysis of long-term and seasonal river water quality changes in Latvia. *Water Research*, 26(5):1203-1216.

Waite, T. D. (1984). *Principles of Water Quality*. Academic Press, Orlando, Florida.

Walker, W. W. (1991). Water quality trends at inflows to Everglades National Park. *Water Resources Bulletin*, 27(1):59-72.

Wiseman, W. J., Jr., Swenson, E. M. and Power, J. (1990). Salinity trends in Louisiana estuaries. *Estuaries*, 13:265-271.

Zetterqvist, L. (1991). Statistical estimation and interpretation of trends in water quality time series. *Water Resources Research*, 27(7):1637-1648.