WATER RESOURCES RESEARCH, VOL. 18, NO. 1, PAGES 107–121, FEBRUARY 1982

Techniques of Trend Analysis for Monthly Water Quality Data

ROBERT M. HIRSCH, JAMES R. SLACK, AND RICHARD A. SMITH

United States Department of the Interior, U.S. Geological Survey, Reston, Virginia 22092

Some of the characteristics that complicate the analysis of water quality time series are non-normal distributions, seasonality, flow relatedness, missing values, values below the limit of detection, and serial correlation. Presented here are techniques that are suitable in the face of the complications listed above for the exploratory analysis of monthly water quality data for monotonic trends. The first procedure described is a nonparametric test for trend applicable to data sets with seasonality, missing values, or values reported as 'less than': the seasonal Kendall test. Under realistic stochastic processes (exhibiting seasonality, skewness, and serial correlation), it is robust in comparison to parametric alternatives, although neither the seasonal Kendall test nor the alternatives can be considered an exact test in the presence of serial correlation. The second procedure, the seasonal Kendall slope estimator, is an estimator of trend magnitude. It is an unbiased estimator of the slope of a linear trend and has considerably higher precision than a regression estimator where data are highly skewed but somewhat lower precision where the data are normal. The third procedure provides a means for testing for change over time in the relationship between constituent concentration and flow, thus avoiding the problem of identifying trends in water quality that are artifacts of the particular sequence of discharges observed (e.g., drought effects). In this method a flow-adjusted concentration is defined as the residual (actual minus conditional expectation) based on a regression of concentration on some function of discharge. These flow-adjusted concentrations, which may also be seasonal and non-normal, can then be tested for trend by using the seasonal Kendall test.

INTRODUCTION

The problem of testing water quality monitoring data for trend in time has received considerable attention in the last decade (see, for example, *Wolman* [1971], *Steele et al.* [1974], *Lettenmaier* [1977], and *Liebetrau* [1979]). Recent interest in methods of water quality trend analysis arises for two reasons. The first is the intrinsic interest in the question of changing water quality arising out of environmental concern and activity. Given legislation that has resulted in the expenditure of large sums of public and private money for the purpose of water quality improvement, there is considerable interest in evaluating the consequences of these expenditures. The second reason for this interest is that only recently has there been a substantial amount of data that is amenable to such analysis. It is clear that in order to detect or assess trends it is necessary that the data be collected at a given location, by using consistent collection and measurement techniques on a regular schedule and over a substantial number of years. Establishment of large networks of water quality stations has occurred mainly since 1970. Some examples of national water quality networks are the U.S. Geological Survey's Benchmark and NASQAN networks [see *Briggs*, 1978].

In this paper we describe procedures suitable for analyzing a large data base to identify stations where water quality characteristics appear to be changing monotonically over time and to estimate the rates of change. The techniques are not intended for exploring the hypothesis that change has occurred at some prespecified time (as a result of known human action, for example), but rather for detecting monotonic trend or change (gradual or sudden) during some interval of time. The techniques do not require complete records. The existence of missing values (a common feature of water quality monthly time series) presents no computational or theoretical problem for applying the techniques. Similarly, the presence of values reported as less than the limit of detection presents no problems for the first of the three techniques.

Meaningful interpretation of the results of these analyses depends on the data collection practices. These techniques are only appropriate for data collected by systematic sampling at a monthly frequency, although stratified random sampling data (with monthly strata) would also be suitable. If the results are to be interpreted as applying to the entire cross section at the station, the water sample must be vertically and horizontally integrated. It is also most important that consistent field and laboratory procedures be used at all times. The achievement of this goal depends on documentation of procedures, training of personnel, and a vigorous program of quality assurance in all phases of the data collection process. Another highly desirable feature is the collection of ancillary data such as time of day, water temperature, and discharge at the time of sample collection. These data provide a basis for explaining a portion of the observed variation in the concentration data. This can enable the analyst to distinguish effects of drought or storms, weather conditions, or effects of solar radiation from possible anthropogenic effects. The use of flow data is illustrated in this report in the discussion of the flow adjustment procedure.

Although presented in terms of hypothesis testing, the procedures presented here are best viewed as exploratory. They are most appropriately used to identify stations where changes are significant or of large magnitude and to quantify these findings. In many cases one may wish to go on from using these techniques to explore the data in graphical form and formulate and test specific hypotheses about the timing, magnitude, or mechanism of change.

The methodology presented here includes three distinct procedures, which can be used alone or in combination. Examples of their combined use are presented by *Smith et al.* [1982].

The three procedures are as follows.

1. A modified form of *Kendall*'s [1938, 1975] tau used as

Paper number 1W1744.

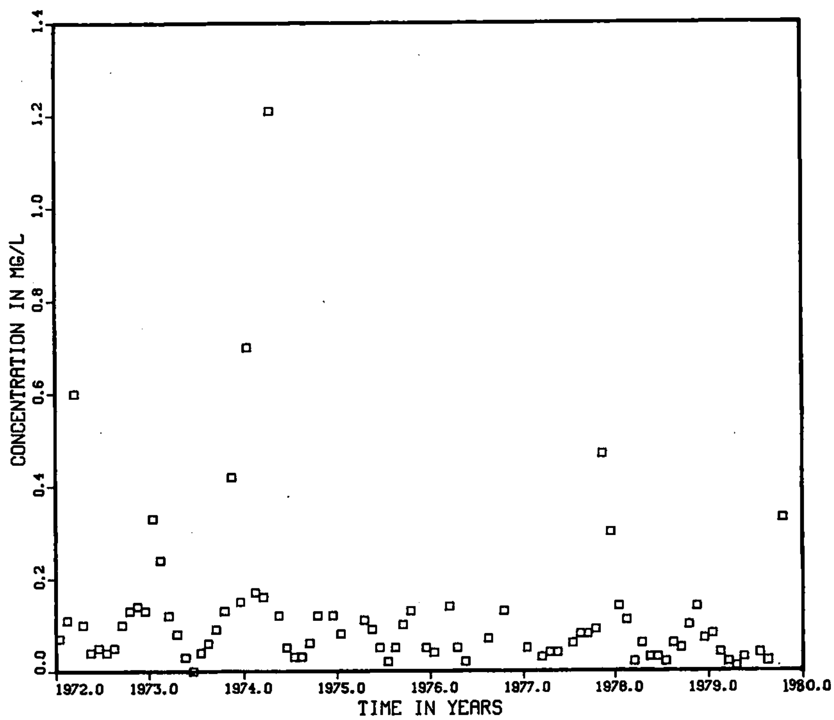

Fig. 1. Concentration of total phosphorus, Klamath River near Klamath, California: $p = 0.007$, slope $= -0.005$ mg/L per year.

a test for trend. This modification is called the seasonal Kendall test for trend.

2. A method of estimating trend magnitude that is closely related to the seasonal Kendall test procedure. This is called the seasonal Kendall slope estimator.

3. A method for computing a time series of flow-adjusted concentrations (FAC). This FAC time series may then be used to examine, graphically or by a formal test (such as the seasonal Kendall test), the question of whether there has been a change in the relationship between flow and concentration over the period of record.

The three techniques are each described, an example of their application is shown, and Monte Carlo experiments exploring the characteristics of the first two procedures are reported. The data used to illustrate the three techniques are the total phosphorus record from a Geological Survey national stream quality accounting network (NASQAN) station on the Klamath River near Klamath, California (station 11-5305.00). NASQAN samples are collected on a systemic sampling schedule (one per month). Collection methods described by *Guy and Norman* [1970] are used to assure that they are cross-sectionally integrated samples. The samples are chilled to 4°C and shipped to the nearest of the two U.S. Geological Survey Central Laboratories for analysis. The Klamath record covers the period from January 1972 through October 1979 and has a total of 80 monthly values. The average concentration over this record is 0.12 mg/L, the standard deviation is 0.17 mg/L, and the coefficient of skewness is 4.0 (see Figure 1).

THE SEASONAL KENDALL TEST FOR TREND

Mann [1945] described a nonparametric test for randomness against trend. The test he described is a particular application of Kendall's test for correlation [*Kendall,* 1975] commonly known as Kendall's tau. According to Mann the null hypothesis of randomness H_0 states that the data $(x_1, \cdots, x_n)$ are a sample of n independent and identically distributed random variables. The alternative hypothesis (H_1) of a two-sided test is that the distribution of x_k and x_j are not identical for all $k, j \leq n$ with $k \neq j$. The test statistic S is defined as

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

where

$$\operatorname{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases} \qquad (2)$$

Note that the statistic T which *Mann* [1945] discussed is a linear function of the statistic S used by *Kendall* [1975] and used in the present paper. In particular, $S = 2T - n(n-1)/2$. Mann shows that under H_0 the distribution of T and hence S is symmetrical and is normal in the limit as $n \to \infty$. Kendall gives the mean and variance of S under H_0 given the possibility that there may be ties in the x values.

$$E[S] = 0 \qquad (3a)$$

$$\operatorname{Var}[S] = n(n-1)(2n+5) - \Sigma_t\, t(t-1)(2t+5)/18 \qquad (3b)$$

where t is the extent of any given tie (number of x's involved in a given tie) and Σ_t denotes the summation over all ties. (For example, if there were four ties of two and one tie of three, then $\Sigma_t\, t(t-1)(2t+5) = 4 \cdot 18 + 1 \cdot 66 = 138$). *Mann* [1945] did not consider ties, but his results correspond exactly with Kendall's for the no-tie case. Both Mann and Kendall derive the exact distribution of S for $n \leq 10$ and show that even for $n = 10$ the normal approximation is excellent, provided one uses a continuity correction of one

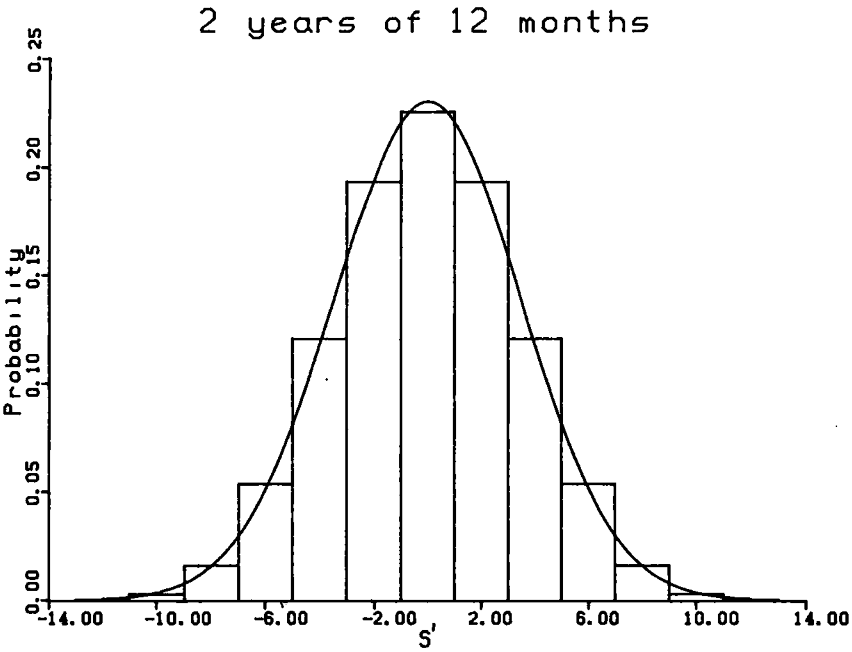

Fig. 2. Histogram of exact distribution of S' under H_0' for $n_i = 2$, $i = 1, 2, \cdots, 12$. The curve is a normal distribution with mean zero and variance = Var [S'].

unit. That is, one computes the standard normal variate Z by

$$Z = \begin{cases} \dfrac{S-1}{(\mathrm{Var}\,(S))^{1/2}} & \text{if} \quad S > 0 \\ 0 & \text{if} \quad S = 0 \\ \dfrac{S+1}{(\mathrm{Var}\,(S))^{1/2}} & \text{if} \quad S < 0 \end{cases} \tag{4}$$

Thus in a two-sided test for trend, the H_0 should be accepted if $|Z| \leq z_{\alpha/2}$, where $F_N(z_{\alpha/2}) = \alpha/2$, F_N being the standard normal cumulative distribution function and α being the size of the significance level for the test. A positive value of S indicates an 'upward trend' (increasing values with time), and a negative value of S indicates a 'downward trend.'

Bradley [1968, p. 288] notes that when this test is used as a test of randomness against normal regression alternatives, this test has an asymptotic relative efficiency of 0.98 relative to the parametric test based on the regression slope coefficient.

If the time series data of interest are monthly water quality data, then the null hypothesis (H_0) given above may be too restrictive. Examination of monthly water quality time series (such as shown in Figure 1) suggests very strongly the presence of seasonality. In fact, 127 of 308 series of monthly total phosphorus concentration data from the U.S. Geological Survey NASQAN program (records of 5–8 years in length) show significant ($\alpha = 0.05$) seasonality, as determined by Kruskal-Wallis multisample test for identical populations [*Bradley*, 1968, p. 129]. Similarly, 139 of 308 series of dissolved solids (residue on evaporation at 180°C) show significant seasonality. These results suggest that seasonality (the existence of different distributions for different times of year) is a common phenomenon.

We propose a test, the 'Seasonal Kendall' test for trend, which is insensitive to the existence of seasonality. The null hypothesis H_0' for this test is a relaxed form of H_0 (which any seasonal but otherwise trend-free process will not violate). Let

$$X = (X_1, X_2, \cdots, X_{12})$$

and

$$X_i = (x_{i1}, x_{i2}, \cdots, x_{in_i})$$

That is, X is the entire sample, made up of subsamples X_1 through X_{12} (one for each month), and each subsample X_i contains the n_i annual values from month i. Note that there is no restriction that $n_i = n_l$, $i \neq l$, or that there be a value for every year and month combination in the sampling period. However, there may be no more than one for each year and month. (A variation on the test will be discussed below in which multiple values are possible.) The null hypothesis H_0' for the seasonal Kendall test is that X is a sample of independent random variables (x_{ij}) and that X_i is a subsample of independent and identically distributed random variables $i = 1, 2, \cdots, 12$. The alternative hypothesis is that for one or more months the subsample is not distributed identically. We define the statistic S_i

$$S_i = \sum_{k=1}^{n_i-1} \sum_{j=k+1}^{n_i} \mathrm{sgn}\,(x_{ij} - x_{ik}) \tag{5}$$

Now, under H_0' the subsample X_i satisfies the null hypothesis H_0 of Mann's test. Therefore relying on Mann and Kendall we have

$$E[S_i] = 0 \tag{6a}$$

$$\mathrm{Var}\,[S_i] = \frac{n_i(n_i-1)(2n_i+5) - \sum_{t_i} t_i\,(t_i-1)(2t_i+5)}{18} \tag{6b}$$

and the distribution of S_i is normal in the limit as $n_i \to \infty$ (t_i is the extent of a given tie in month i). We then define $S' =$

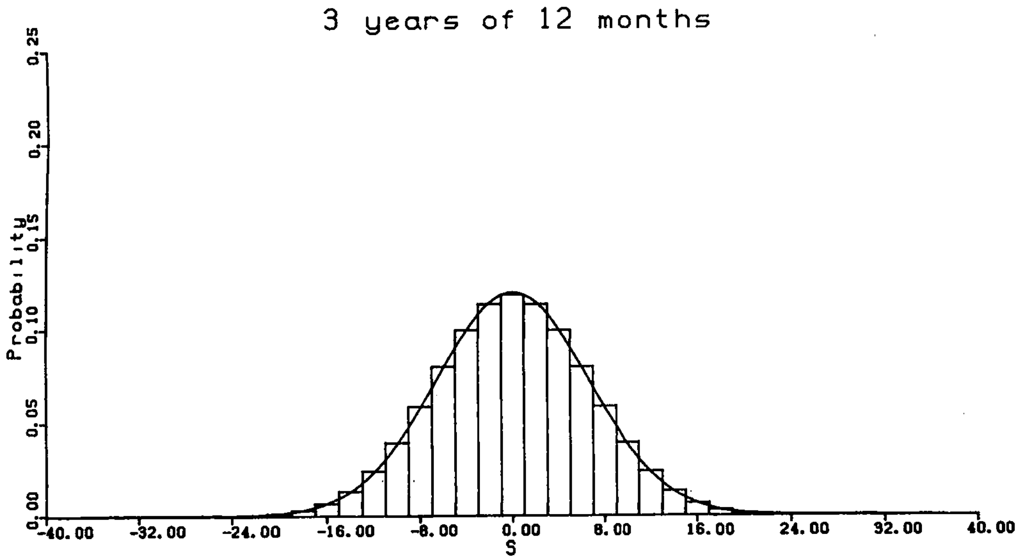

Fig. 3. Histogram of exact distribution of S' under H_0' for $n_i = 3, i = 1, 2, \cdots, 12$. The curve is a normal distribution with mean zero and variance = Var $[S']$.

$\Sigma_{i=1}^{12} S_i$ and can derive its expectation, variance, and limit distribution.

$$E[S'] = \sum_{i=1}^{12} E[S_i] = 0 \tag{7a}$$

$$\mathrm{Var}\,[S'] = \sum_{i=1}^{12} \mathrm{Var}[S_i] + \sum_{i=1}^{12} \sum_{\substack{l=1 \\ i \neq l}}^{12} \mathrm{cov}\,(S_i S_l) \tag{7b}$$

Now S_i and S_l $(i \neq l)$ are functions of independent random variables ($S_i = f(X_i)$, $S_l = f(X_l)$ and $X_i \cap X_l = \phi$ because X_i and X_l are the data from months i and l, respectively, and all elements of X are independent) so cov $(S_i S_l) = 0$. Thus $E[S']$ and Var $[S']$ are known simply from the n_i and t_i values. In addition, S' must be normal in the limit as $n_i \rightarrow \infty$, $i = 1, 2, \cdots, 12$ being the sum of 12 distributions which are normal in the limit.

For any data set it is possible to determine the exact distribution under H_0 of S' based on the n_i and t_i values. This can be done as a straightforward extension of the procedure for computing the exact distribution of S (equivalent to S_i in the seasonal Kendall test) as described by *Kendall* [1975]. The exact distribution of S' is arrived at by enumerating all possible permutations and combinations of S_i for the 12 months, summing the S_i's, multiplying the independent probabilities, and adding the probabilities of all of the S_i sequences that sum to each particular value of S'. Figure 2 shows the exact probabilities for the case of 2 years of monthly data ($n_i = 2$, $i = 1, 2, \cdots, 12$) with no ties, and Figure 3 shows it for 3 years of monthly data ($n_i = 3$, $i = 1, 2, \cdots, 12$) with no ties. For both of these cases S' may only take on even values, and the probability of a given S' value is depicted by a histogram class ranging from $S' - 1$ to $S' + 1$.

Superimposed on each figure is the normal distribution with a mean of zero and variance of Var $[S']$ where

$$\mathrm{Var}\,[S'] = \sum_{i=1}^{12} \mathrm{Var}\,[S_i] = \sum_{i=1}^{12} \frac{n_i(n_i - 1)(2n_i + 5)}{18} \tag{8}$$

On the basis of visual inspection, one can see that even for records as short as 3 years the normal approximation will work quite well for estimating p = Prob $[|S'| \geq s]$ (the probability that S' will depart from zero by the amount s or more) provided that a continuity correction of one unit (toward zero) is made. For using the normal approximation we define the standard normal deviate Z' as

$$Z = \begin{cases} \dfrac{S' - 1}{(\mathrm{Var}\,(S'))^{1/2}} & \text{if } S' > 0 \\ 0 & \text{if } S' = 0 \\ \dfrac{S' + 1}{(\mathrm{Var}\,(S'))^{1/2}} & \text{if } S' < 0 \end{cases} \tag{9}$$

The approximation is certainly adequate for $n_i = 3$ for all i. The worst disagreement between the exact two-sided probability and the approximate probability occurs at $|S'| = 6$ where the exact probability is 0.4530 and the normal approximation is 0.4510. Even for $n_i = 2$ for all i, where the exact distribution could easily be used, the worst disagreement occurs at $|S'| = 8$ where the exact probability is 0.0386 and the approximation is 0.0433. Without the continuity correction the approximate probability would be 0.0209. For either the $n_i = 2$ or $n_i = 3$ case, at any significance level greater than or equal to 0.01, the relative error incurred by using the approximation in place of the exact distribution is less than 14%.

A possible modification of the seasonal Kendall test would involve using multiple observations for each month rather than limiting the time series to one observation per month. The observations occurring in the same month of the same year would be treated as tied observations with respect to their time of occurrence. In the former version of the seasonal Kendall test, ties are only possible in the magnitudes but not in the time index; in this modified version ties may occur in both. *Kendall* [1975] describes the modifications to his test necessary when both kinds of ties are possible. We have not explored the use of this modified test, and it is not clear whether it would be preferable to use all available data or to take, say, the medians of the multiple observations in each of the months and use them in the former version of the test.

One feature of water quality data, particularly metals and organic compounds, is the reporting of 'less than' values. For any analytic technique used in the laboratory, a limit of detection (LD) is defined and all measurements below the LD are reported as being less than that limit. When a time series contains any 'less than' values, then parametric methods of trend detection become unusable. These 'less than' values present no difficulty for nonparametric methods such as the seasonal Kendall test because nonparametric tests require making comparisons of values to determine which is the larger. The 'less than' data can all be considered to be smaller than any numerical value equal to or greater than the LD and tied with any other 'less than' value. In cases where the LD has changed over time as more sensitive instruments are developed, it is necessary to take all data reported below the highest LD (including those reported as less than any lower LD) and consider them all to be tied at the highest LD.

It should be recognized that there may be instances in which some months exhibit strong evidence of upward trends and others exhibit strong evidence of downward trends (that is, some S_i's are large positive values and others are large negative values), and yet the test result indicates no trend (S' close to zero). If one is interested in trends in specific months, then it would be appropriate to use the Mann-Kendall test for each of the months and report the test results for each. The seasonal Kendall test is specifically designed to provide a single summary statistic for the entire record and will not indicate when there are trends in opposing directions in different months.

For the example of the Klamath River (see Figure 1), the statistic $S' = -62$ and the Var $[S']$ under the null hypothesis is 514. Thus the Z' value is -2.69, and the p value or two-sided significance level of the trend is 0.0072. That is, $p = \text{Prob}\ [|S'| \geq 62] = 0.0072$.

MONTE CARLO EXPERIMENT ON THE SEASONAL KENDALL TEST

We have defined the seasonal Kendall test for trend, derived the mean and variance of the test statistic S' under the null hypothesis H_0', and verified that the normal distribution provides a good approximation to the exact distribution of S' when the continuity correction is used for records as short as three years. We now proceed to address some questions about the significance and power of the test as compared to other reasonable alternative tests for trend. In particular, we explore the impacts of underlying distributions, of seasonality, and of serial dependence on the significance and power of the test. This is done through a Monte Carlo experiment. The purpose of this experiment is not to precisely quantify these effects but rather to provide insight on their general character. Thus no attempt is made to describe relationships between population characteristics, sample size, and the significance or power of the test. Given the problems of estimating the relevant population characteristics from the small samples that are typically available, it is unclear that knowledge of such relationships would be particularly useful.

Before describing the Monte Carlo experiments some definitions should be given. The actual significance level of a test under some particular trend-free stochastic process is the probability that the test would indicate trend (fail to accept the null hypothesis of no trend) at the preselected nominal significance level α.

We call a process trend-free if, for each i, the distributions of x_{ij} and x_{ik} are identical for all j and k. The seasonal Kendall test was designed to be particularly powerful against the alternative of trend. One of the purposes of the Monte Carlo experiment is to evaluate the power of the test against other departures from H_0' that are trend free, specifically serial dependence. One would prefer to use a test with minimal power against serial dependence such that the probability of rejecting the null hypothesis is close to α when there is serial dependence but no trend.

Power is the probability that the test would indicate trend (fail to accept the null hypothesis) when the generating process did, in fact, have trend. Clearly, the power of a test will be a function of the stochastic process, trend magnitude, as well as record length.

The objectives to consider in selecting a test for use in an exploratory study (assuming that α has already been selected) are these. (1) The actual significance should be relatively close to α under stochastic processes thought to be relatively similar to the time series one expects to be testing, and (2) the power for detecting trends should be relatively high compared to some alternative tests for processes in which trend exists and which are thought to be similar to the time series one expects to be testing.

The first of the two alternative tests for trend is based on linear regression. In this test the parameters of the regression equation (10) are estimated by ordinary least squares.

$$\hat{x}_{ij} = a + b \cdot \left(j + \frac{i}{12}\right) \tag{10}$$

The null hypothesis (H_0'') is that the x_{ij} are normal independent and identically distributed in time, which implies that $b = 0$. The test statistic used is T where

$$T = \frac{r \cdot (m - 2)^{1/2}}{(1 - r^2)^{1/2}} \tag{11}$$

where m is the total number of observations and r is the product moment correlation coefficient between x_{ij} and time $(j + i/12)$. The probability distribution of T under H_0'' is the Student t distribution with $m - 2$ degrees of freedom [*Kendall and Stuart,* 1969, p. 387]. This statistical test is denoted LR (linear regression).

The other test for trend considered is performed by deseasonalizing the data before regressing them against time. In this procedure, called seasonal regression (SR), the sample mean ($\bar{x}_i$) and sample standard deviation (s_{xi}) are computed for each of the 12 months. The deseasonalized data are denoted u_{ij} where

$$u_{ij} = \frac{x_{ij} - \bar{x}_i}{s_{xi}} \tag{12}$$

The parameters of the equation

$$\hat{u}_{ij} = a' + b' \cdot j \tag{13}$$

are estimated by using ordinary least squares. The null hypothesis H_0''' for the test is that the x_{ij} are normal, independent, and that for all i, x_{ij} and x_{ik} are identically distributed for all j and k. Thus H_0''' implies that $b' = 0$. The

TABLE 1. Mean Value of Sample Statistics of Deseasonalized u_{ij} Data

Data Source	Coefficient of Skews	Lag 1 Correlation Coefficient	Lag 2 Correlation Coefficient	Number of Series	Mean Series Length, months
Process					
NI	0.01	−0.01	−0.01	100	72
LNI	0.70	−0.00	−0.01	100	72
NIS	0.01	−0.01	−0.01	100	72
NAR	0.01	0.18	0.03	100	72
NARMA	0.02	0.19	0.14	100	72
LNARS	0.70	0.22	0.04	100	72
NASQAN					
Dissolved solids	0.07	0.26	0.17	260	57
concentration	(31)	(55)	(28)		
Total phosphorus	0.46	0.17	0.10	281	59
concentration	(65)	(43)	(38)		

Numbers in parentheses under skew coefficients are the percent of stations at which normality is rejected by chi-squared goodness of fit test ($\alpha = 0.05$). Numbers in parentheses under Lag 1 and Lag 2 correlation coefficients are percent of stations at which the correlation coefficient is positive and significantly different form zero ($\alpha = 0.05$).

test statistic is T' where $T' = (n(n-1)(n+1))^{1/2}b'$, where n is the number of years of record. T' is distributed approximately as a Student t random variable with $2n$ degrees of freedom. This approximation has been found to be adequate for $n \geq 3$, and this is seen in the Monte Carlo experiment for $n = 5$, 10, and 20. The reason that the usual T statistic may not be used in this case is that the u_{ij} are not independent, even when the x_{ij} are independent (thereby reducing the number of degrees of freedom).

Six different trend-free stochastic processes are considered in this experiment. They are defined as follows:

1. Normal independent (NI)

$$x_{ij} = \varepsilon_{ij} \tag{14a}$$

2. Log normal independent (LNI)

$$x_{ij} = \exp\,[0.83 \cdot \varepsilon_{ij} - 0.35] - 1.0 \tag{14b}$$

3. Normal independent with seasonal cycle (NIS)

$$x_{ij} = (0.5)\ \varepsilon_{ij} + \sin\left(\frac{\pi}{3} + \frac{\pi}{6}\cdot i\right) \tag{14c}$$

4. Normal autoregressive (NAR)

$$x_{ij} = 0.2\,[x_{ij}]_L + 0.98 \cdot \varepsilon_{ij} \tag{14d}$$

5. Normal autoregressive-moving average (1, 1) (NARMA)

$$x_{ij} = 0.75\,[x_{ij}]_L + 0.97 \cdot \varepsilon_{ij} - 0.57\,[\varepsilon_{ij}]_L \tag{14e}$$

6. Log normal, autoregressive with seasonal cycle (LNARS)

$$x_{ij} = (0.5) \cdot \exp\,[0.22[x_{ij}]_L + 0.80\ \varepsilon_{ij} - 0.35] - 0.71 + \sin\left(\frac{\pi}{3} + \frac{\pi}{6}\cdot i\right) \tag{14f}$$

The series are generated for $i = 1, 2, \cdots, 12; j = 1, 2, \cdots, n$, for $n = 5$, 10, 20. The notation $[\ \ _{ij}]_L = \ \ _{i-1,j}$ for $i = 2, 3, \cdots, 12$, $[\ \ _{ij}]_L = \ \ _{12,j-1}$ for $i = 1$ is used. The variable ε_{ij} is a normal random variable with zero mean and unit variance. For all six processes, the x_{ij} have zero mean and unit variance over all months taken together. However, for the NIS and LNARS processes in any given month, the variance is 0.5 and the mean takes on various values between −0.71 and +0.71.

In order to illustrate the characteristics of these processes, the average values of some sample statistics are given in Table 1 for 100 repetitions of samples of 6 years in length. Also presented are average statistics for two historical time series of water quality data from the U.S. Geological Survey NASQAN program. The record lengths for these series are 5–8 years. The historical data used here are 'detrended' in order to make them comparable to the computer-generated data. The detrending is accomplished by using the seasonal Kendall slope estimate described in the next section of this paper. All of the data, historical or generated, is then deseasonalized before the skewness and correlation coefficients are computed.

Each series generated was tested by each of the three tests (SK, LR, and SR) by using the nominal significance level α of 0.05. The number of repetitions for each process and record length was 500. The same series were modified by adding linear trend to create a new series v_{ij}

$$v_{ij} = x_{ij} + \beta\left(\frac{i}{12} + j\right) \tag{15}$$

For each record length, eight different β values (trend slope in units of standard deviations per year) were used; for $n = 5$ the β values were 0.05 (0.05) 0.4, for $n = 10$ they were 0.02 (0.02) 0.16, and for $n = 20$ they were 0.0065 (0.0065) 0.05. For each of these v_{ij} series, the three trend tests were applied. The results expressed in terms of frequency of detecting trends are shown in Figures 4–6.

The Monte Carlo experiment is used to test the null hypothesis that the true frequency of rejection is α against the alternative that it is not equal to α. The results in Table 2 give $\hat{\alpha}$, the ratio of rejections to the number of trials, for those cases where no trend exists. The probability distribution of the number of rejections in 500 is binomial, and the 95% confidence band, expressed in terms of $\hat{\alpha}$, is [0.032, 0.068]. Those cases in which $\hat{\alpha}$ falls outside this confidence band are noted in Table 2. Also, the power of the three tests are compared for each process, record length, and true

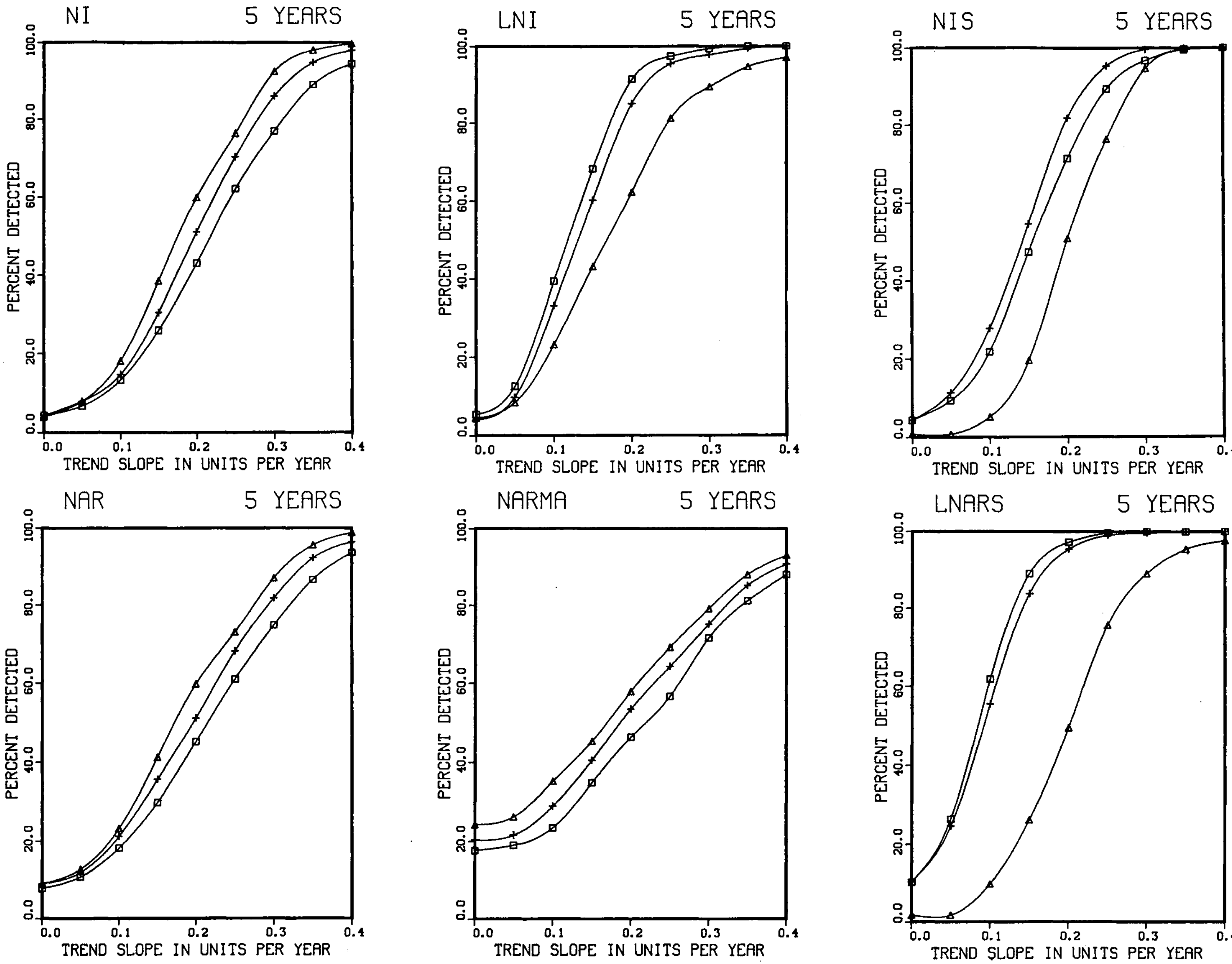

Fig. 4. Power of the three tests for sequences of length 5. Symbols are as follows: seasonal Kendall (square), linear regression (triangle), seasonal regression (cross). Significance level of the tests, α, is 0.05.

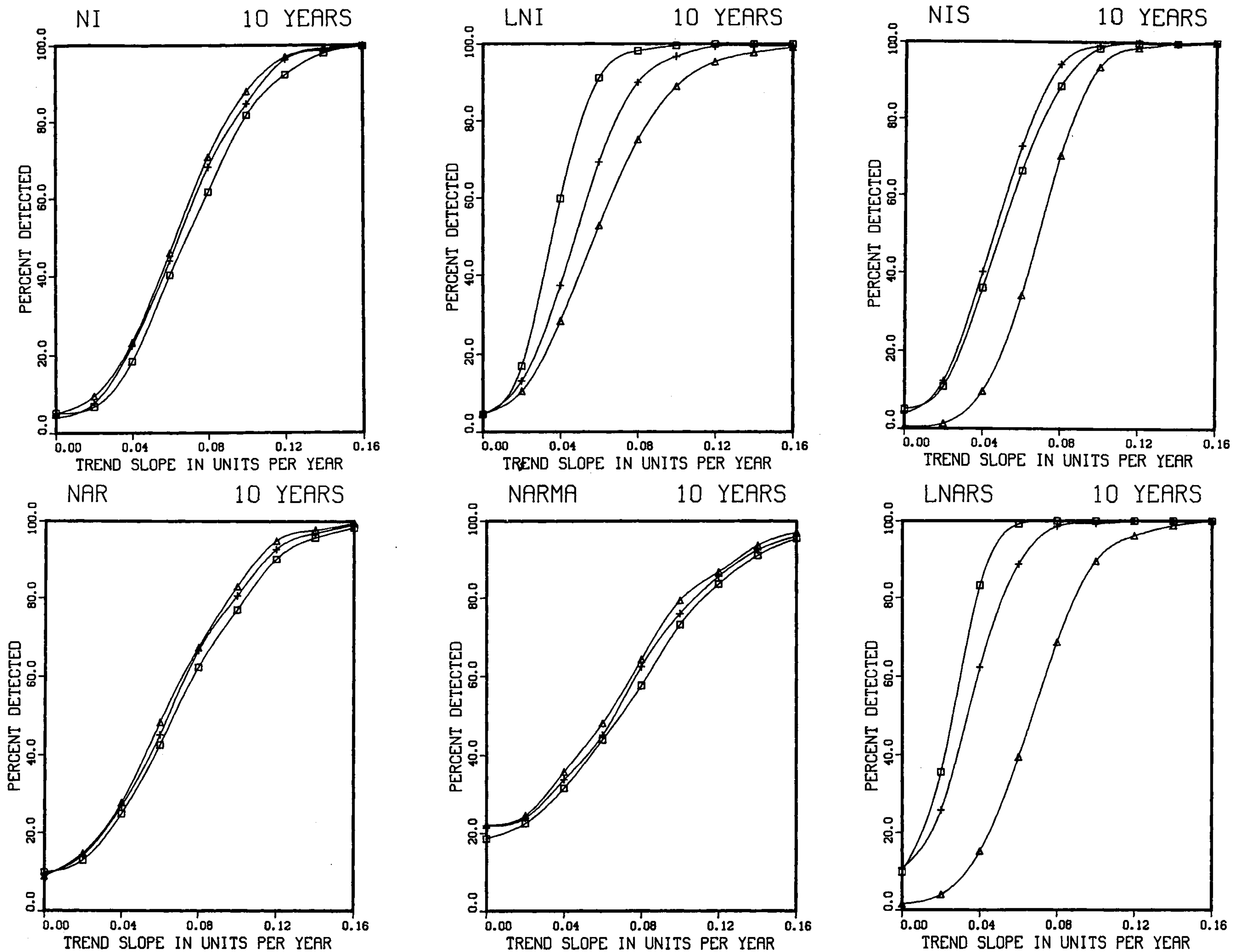

Fig. 5. Power of the three tests for sequences of length 10. Symbols are as follows: seasonal Kendall (square), linear regression (triangle), seasonal regression (cross). Significance level of the tests, α, is 0.05.

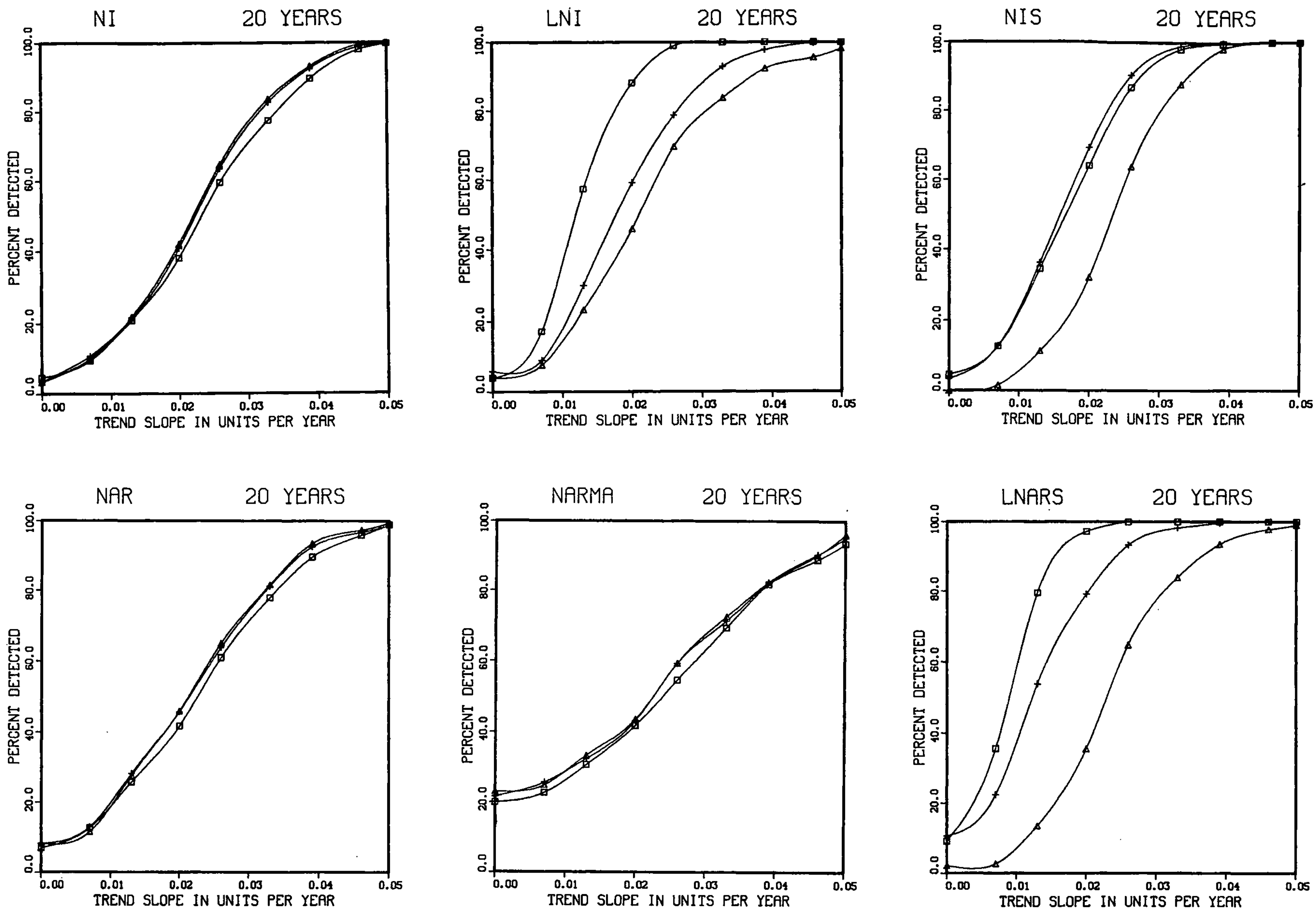

Fig. 6. Power of the three tests for sequences of length 20. Symbols are as follows: seasonal Kendall (square), linear regression (triangle), seasonal regression (cross). Significance level of the tests, α, is 0.05.

TABLE 2. Observed α values where α is defined as the frequency with which trends are indicated where none exists.

Process	n	SK	LR	SR
NI	5	0.044	0.040	0.044
NI	10	0.052	0.050	0.040
NI	20	0.046	0.034	0.034
LNI	5	0.054	0.044	0.040
LNI	10	0.046	0.048	0.046
LNI	20	0.040	0.040	0.058
NIS	5	0.044	0.008^-	0.044
NIS	10	0.052	0.006^-	0.040
NIS	20	0.032	0.004^-	0.044
NAR	5	0.078^+	0.090^+	0.090^+
NAR	10	0.098^+	0.088^+	0.094^+
NAR	20	0.070^+	0.082^+	0.082^+
NARMA	5	0.176^+	0.242^+	0.204^+
NARMA	10	0.186^+	0.222^+	0.220^+
NARMA	20	0.200^+	0.223^+	0.216^+
LNARS	5	0.102^+	0.018^-	0.104^+
LNARS	10	0.098^+	0.016^-	0.108^+
LNARS	20	0.092^+	0.022^-	0.106^+

The 'plus' indicates that α is above the 95% confidence band around the nominal α (0.05), and the "minus" indicates that α is below this confidence band.

slope. To simplify the comparison of tests, Table 3 presents the minimum relative power $\omega(K)$ of each test K for each record length n and process. Here $\omega(K)$ is defined as

$$\omega(K) = 100 \cdot \max_{\beta,\kappa} (D(\kappa, \beta) - D(K, \beta)) \tag{16}$$

where $D(\kappa, \beta)$ is the frequency of trend detection by test κ, where $\kappa \varepsilon \{SK, LR, SR\}$, given the slope β, record length n, and the process. The following observations about the three tests may be made from the figures and two tables. In the succeeding discussion, 'significant' should be taken to mean significant at the 5% level.

RESULTS OF THE MONTE CARLO EXPERIMENT

1. In all cases where a process satisfies the null hypothesis of a given test, the $\hat{\alpha}$ values fall within the 95% confidence band around the nominal significance level α (0.05). These cases are NI, LNI, and NIS for the SK test, NI for the LR test, and NI and NIS for the SR test.

2. In addition to these cases the two regression-based tests appear to be robust (in terms of significance) against a departure from normality. That is, for the LNI process (population skewness = 4.0), both the LR and SR test have $\hat{\alpha}$ values within the 95% confidence band around α.

3. For the nonseasonal dependent processes (NAR, NARMA) the actual significance level of each of the three tests against trend appears to depart from the nominal significance level α. Specifically, the probability that the null hypothesis will be rejected is greater than the preselected probability α. In the case of the NAR process no definitive statement can be made about the relative magnitude of this discrepancy. However, for the NARMA process without trend, at n = 5, 10, and 20, in those trials where the SK and LR tests reached different conclusions (rejection or nonrejection), LR rejected the null hypothesis significantly more frequently than did SK. To illustrate this, consider $n = 5$; in 364 trials, both SK and LR failed to reject their null hypothesis, in 73 trials both reject their null hypothesis. Of the remaining 63 trials LR rejected its null hypothesis 48 times and SK rejected its null hypothesis 15 times. The hypothesis that rejections are equally likely where disagreements occur is rejected ($p \leq 0.001$, two-sided). The results are similar for $n = 10$, 20 with $p = 0.010$ and $p = 0.024$, respectively. Similarly, SR rejected significantly more than SK at $n = 5$ and $n = 10$ ($p = 0.044$, $p = 0.008$, respectively). Thus for the NARMA process the SK test has a slight advantage over the other two tests in the sense that it indicates the existence of trend less frequently when no trend exists. For the NAR process none of the three tests can be shown to be superior to the others in this sense.

4. For the LNARS (where both seasonality and dependence exist) the results show $\hat{\alpha}$ values for the SK and SR tests lying above the 95% confidence band about α as a result of the dependence. For the LR test $\hat{\alpha}$ is below the 95% confidence band, suggesting that the seasonality has a more profound effect on the significance than does the dependence.

5. Where the x_{ij}'s satisfy the null hypothesis for all three tests, then the LR test appears to be most powerful, followed by SR, followed by SK. However, in the two cases where H_0'' is violated but H_0' is not (LNI and NIS), the violation is sufficient to make the LR test become a less powerful test for trend than the SK test. Similarly, in the one case where H_0''' is violated (LNI), the violation is sufficient to make the SR test become a less powerful test for trend than the SK test.

6. For the NI process with trend, the difference in the power of the three tests decreases with increasing record length. In fact, when $n = 20$ the difference in frequency of trend detection is no more than 6.2% over all β values considered. For the NIS process the difference in power between the SK and SR tests also decline with increasing record length. However, for the LNI process the difference

TABLE 3. Minimum Relative Power, $\omega(K)$, Expressed in Percent

Process	n	SK	LR	SR
NI	5	16.8	0.0	8.8
NI	10	9.0	0.0	3.2
NI	20	6.2	0.8	1.2
LNI	5	0.0	29.2	8.0
LNI	10	0.0	38.4	22.4
LNI	20	0.0	42.2	29.0
NIS	5	10.6	35.0	0.0
NIS	10	6.4	38.6	0.0
NIS	20	5.2	37.2	0.0
NAR	5	14.6	0.0	8.6
NAR	10	6.0	0.0	3.2
NAR	20	4.2	1.4	1.2
NARMA	5	12.4	0.0	6.2
NARMA	10	6.6	0.0	3.4
NARMA	20	4.8	0.8	1.2
LNARS	5	0.0	62.8	6.2
LNARS	10	0.0	68.0	21.0
LNARS	20	0.0	66.0	25.8

An entry of 0.0 indicates test is most powerful at all values of β.

TABLE 4. The Average, μ, Standard Deviation, σ, and Relative Standard Deviation ψ, of the Three Estimators B, b, and b^* (Associated With Seasonal Kendall Procedure, Linear Regression, and Seasonal Regression, Respectively)

Process	n	$\mu(B-\beta)\times 10^3$	$\mu(b-\beta)\times 10^3$	$\mu(b^*-\beta)\times 10^3$	$\sigma(B)\times 10^3$	$\sigma(b)\times 10^3$	$\sigma(b^*)\times 10^3$	$\psi(B)$	$\psi(b)$	$\psi(b^*)$
NI	5	−1.4	−1.4	−9.9*	102.0	85.1	91.4	1.20	1.00	1.07
NI	10	−0.6	−0.4	−0.6	35.2	31.2	31.5	1.13	1.00	1.01
NI	20	−0.0	3.3	0.4	11.1	10.6	10.7	1.05	1.00	1.01
LNI	5	−0.1	3.9	13.8*	53.9	91.6	95.6	1.00	1.70	1.77
LNI	10	−0.3	−0.8	1.1	16.4	29.6	31.2	1.00	1.80	1.90
LNI	20	−0.5	−0.2	−0.1	5.9	10.4	10.4	1.00	1.77	1.77
NIS	5	−1.0	−21.0*	−2.7	72.1	60.1	63.1	1.20	1.00	1.05
NIS	10	−0.4	−5.3*	−0.2	24.9	22.0	22.2	1.13	1.00	1.01
NIS	20	−0.0	−1.0	0.3	7.9	7.5	7.6	1.05	1.00	1.01
NAR	5	0.5	−1.8	−9.4	118.4	103.1	110.3	1.15	1.00	1.07
NAR	10	0.0	−0.5	−0.6	41.6	37.9	38.3	1.10	1.00	1.01
NAR	20	0.3	0.4	0.6	13.4	13.1	13.2	1.03	1.00	1.01
NARMA	5	2.6	0.1	−9.2	155.3	145.5	150.8	1.07	1.00	1.03
NARMA	10	−0.0	−0.2	−0.7	52.3	51.2	51.5	1.02	1.00	1.01
NARMA	20	−0.5	−0.4	−0.3	16.8	16.9	16.8	1.00	1.01	1.00
LNARS	5	−1.8	−2.6*	−2.3	45.5	74.0	76.3	1.00	1.63	1.68
LNARS	10	0.1	−5.4*	0.3	14.6	26.1	26.9	1.00	1.79	1.84
LNARS	20	−0.4	−1.6*	−0.4	5.3	9.3	9.3	1.00	1.77	1.77

*Estimate is biased at the 5% significance level.

in power between the SK and LR tests (or the SK and SR tests) increases with increasing record length.

7. For the two processes with dependence and no seasonality (NAR and NARMA) the relative power of the tests is not unlike what is observed for the NI process. The power curves are simply shifted upward, reflecting the inflated actual significance of all three tests when dependence exists.

8. For the LNARS process the SK tests is most powerful. The power of LR is severely reduced (in comparison) by seasonality and skewness, and the power of SR is reduced by skewness.

In summary, if one knows that the data to be examined for trends are normal and nonseasonal, then LR is clearly the best of the three tests. If one knows that the data are normal but seasonal, then SR may be best (depending on the magnitude of the seasonality). In general, we do not know much about these characteristics of the data. In general, 'No . . . obvious indication advises the experimenter that a parametric assumption has been violated. Of course he may apply time-consuming tests for normality or homogeneity to the obtained data, but such tests are rather unsatisfactory. They are unlikely to detect any but the most extreme violations when samples are small, and they are almost certain to detect the most trivially slight violations when samples are very large,' [*Bradley*, 1968, p. 23]. Given that our data analysis has shown departure from normality and the presence of seasonality to be common features of water quality data, coupled with the rather small loss of power associated with using the SK test where the LR test would be most powerful, we would argue for the use of the seasonal Kendall test as an exploratory test for trend.

None of the three tests considered here is designed to distinguish between trend and the long-term variations typical of trend-free dependent processes, although the SK test appears to have a slightly better performance than do the other tests under a long-memory ARMA process. There is certainly a need for a trend test that is robust against seasonal behavior, departures from normality, and dependence. The seasonal Kendall test can only be regarded as robust against the former two and not the latter.

THE SEASONAL KENDALL SLOPE ESTIMATOR

In addition to identifying time series that exhibit trend, it may be desirable for some applications to estimate the magnitude of such a trend. We have chosen to express this magnitude as a slope (change per unit time), but this does not imply any belief that the trend takes the form of a linear trend in the process mean. In an overview of many stations, one may wish to identify those stations for which trend slope is large with respect to the mean value. One may also want to identify those stations where extrapolation of an existing trend would suggest that frequent violations of some relevant water quality criterion might occur in the near future. The estimator we define is an extension (to account for seasonality) of one proposed by *Theil* [1950] and by *Sen* [1968].

We define the seasonal Kendall slope estimator B by the following computational algorithm. Compute $d_{ijk} = (x_{ij} - x_{ik})/(j-k)$ for all (x_{ij}, x_{ik}) pairs $i = 1, 2, \cdots, 12$; $1 \le k < j \le n_i$. The slope estimator B is the median of these d_{ijk} values. The estimator B is related to the seasonal Kendall test statistic S' such that if $S' > 0$, then $B \ge 0$ ($B > 0$ if one or no $d_{ijk} = 0$), and if $S' < 0$, then $B \le 0$ ($B < 0$ if one or no $d_{ijk} = 0$). This is because S' is equivalent to the number of positive d_{ijk}'s minus the number of negative d_{ijk}'s, and B is the median of these d_{ijk}'s.

By using the median of these individual slope d_{ijk} values, the estimate B is quite resistant to the effect of extreme values in the data. It is also unaffected by seasonality because the slopes are always computed between values that are multiples of 12 months apart.

For the Klamath River example given above, the B value is −0.005 mg/L per year. For comparison, linear regression gives a slope b value of −0.014 mg/L. For skewed data, such as the Klamath, the finding that $|b| > |B|$ is typical. The

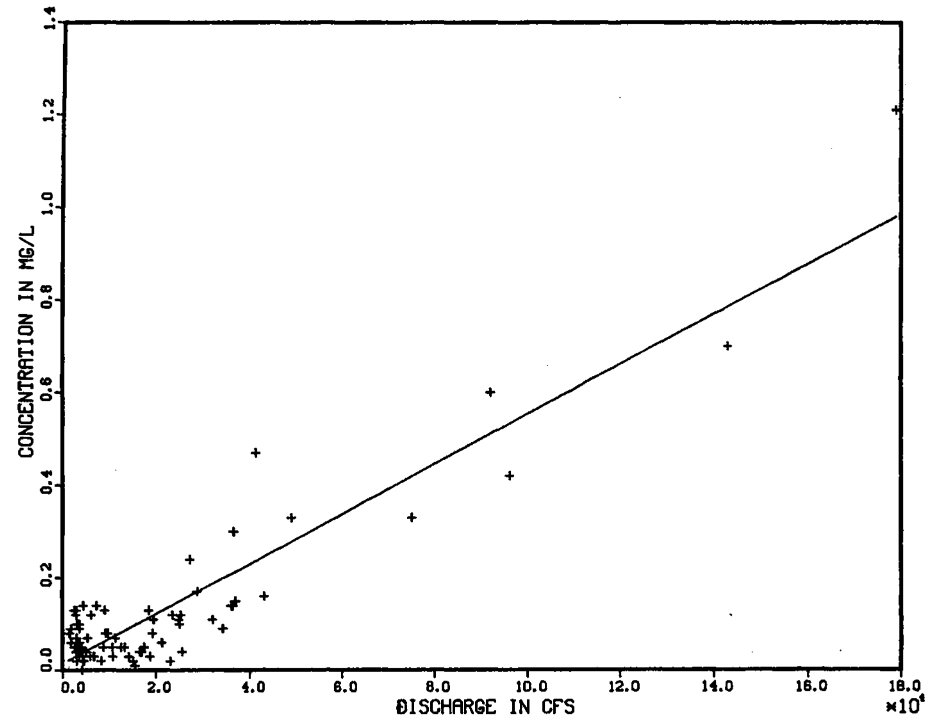

Fig. 7. Relationship between discharge and total phosphorus concentration, Klamath River near Klamath, California. $\hat{C} = 0.14 + 5.4 \times 10^{-6} \cdot Q$, $R^2 = 0.84$.

estimate b is influenced by the extremes of the data much more than is B.

Monte Carlo Experiment on the Seasonal Kendall Slope Estimator

To evaluate the precision and bias of this estimate, B is compared with the estimate b from linear regression (equation (10)) and also an estimate denoted b^* which arises from seasonal regression (equation (13)). Specifically,

$$b^* = b' \left[\frac{1}{12} \cdot \sum_{i=1}^{12} S_{xi} \right] \qquad (17)$$

The estimates B, b, and b^* are computed for 500 series of the six generating processes with a true slope (β) of 1.0 per year for record length of 5 and 10 years and 100 series of 20 years.

Table 4 provides summary statistics of this Monte Carlo experiment: the mean error for the three estimates $\mu(\ )$; the standard deviation $\sigma(\ )$; and the relative standard deviation

$$\psi(\ ) = \frac{\hat{\sigma}(\ )}{\min\{\hat{\sigma}(B), \hat{\sigma}(b), \hat{\sigma}(b^*)\}}$$

The Monte Carlo experiment is used, in part, to test the hypotheses that the various slope estimators are unbiased (that is, $E[B] = \beta$, $E[b] = \beta$, $E[b^*] = \beta$) for all six processes and three record lengths. The estimator B can be shown analytically to be unbiased for all six cases. The distribution

TABLE 5. Characteristics of a NASQAN monthly Data: Concentrations, x_{ij}, and Flow-Adjusted Concentrations, w_{ij}

	Total Phosphorus		Dissolved Solids	
	x_{ij}	w_{ij}	x_{ij}	w_{ij}
Percent of stations with significant seasonality ($\alpha = 0.05$, Kruskal-Wallis test)	49	27	47	29
Average coefficient of skewness	0.49	0.26	0.06	0.05
Percent of stations significantly non-normal ($\alpha = 0.05$, chi-squared goodness of fit test)	68	36	32	26
Average lag 1 correlation coefficient	0.17	0.16	0.26	0.19
Percent of stations with lag 1 correlation coefficient significantly greater than zero ($\alpha = 0.05$)	42	41	55	46
Average lag 2 correlation coefficient	0.10	0.11	0.16	0.11
Percent of stations with lag 2 correlation coefficient significant greater than zero ($\alpha = 0.05$)	28	32	37	27
Number of stations	198	198	223	223
Average record length in months	61	61	57	57

Statistics were computed on deseasonalized data except for Kruskal-Wallis test for seasonality.

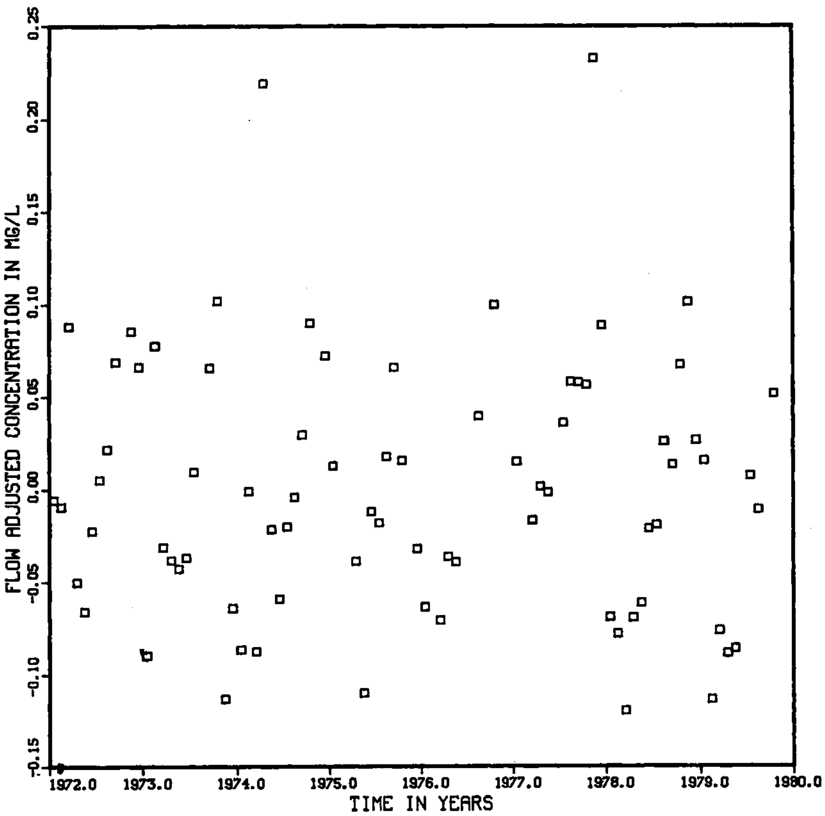

Fig. 8. Flow-adjusted concentration, Klamath River near Klamath, California: $p = 0.434$, slope $= -0.002$ mg/L per year.

of the d_{ijk}'s is symmetric with mean value β. The expectation of the sample median of a symmetric distribution is equal to the mean of the distribution, thus B (the median of the d_{ijk}) has expectation β. This holds for any situation where observations a multiple of 12 months apart are distributed identically except for a shift equal to the product of β and time (in years). The estimate b is significantly biased (at the 5% level) for NIS at $n = 5$ and 10 and for LNARS at $n = 5$, 10, and 20. This bias is related to the phase shift in the generating equations (14*c*) and (14*f*). If the phase shift were $\pi/4$ (or more generally $(\pi/4) \pm (\pi/2)k$ for k integer) rather than $\pi/3$, then b would be unbiased. But for any other phase shift it will be biased and may be much more biased (for example, if the phase shift were zero). The estimator b^* is significantly biased for the NI and LNI processes at $n = 5$. Whether there are conditions in which it is an unbiased estimator is an open question.

Concerning the precision of the estimator, the results show that, for all of the normal processes (NI, NIS, NAR, NARMA) and all record lengths ($n = 5$, 10, and 20), $\hat{\sigma}(b) < \hat{\sigma}(b^*) < \hat{\sigma}(B)$. (With one exception: NARMA, $n = 20$, in which case all three are nearly equal.) For these processes, as the record length increases, the ratio of $\hat{\sigma}(B)/\hat{\sigma}(b^*)$ decreases. For processes with skewness (LNI and LNARS), $\hat{\sigma}(B) < \hat{\sigma}(b) \leq \hat{\sigma}(b^*)$ for all record lengths.

Given a desire to find a method of slope estimation that is unbiased and has a lower variance in situations where seasonality, skewness, and serial correlation may be present, the seasonal Kendall slope estimator B appears to be an appropriate choice. In no case does it perform much worse than the alternative methods, and in some cases it performs a great deal better.

FLOW ADJUSTMENT

It is well known that in many cases constituent concentrations are correlated with river discharge (see, for example, *Langbein and Dawdy* [1964], *Johnson et al.* [1969], *Borman et al.* [1974], *Smith et al.* [1982]). The causes of the relationship and the particular functional form that might be used to characterize it vary from site to site and constituent to constituent.

In some instances constituent loading rates are relatively constant because the main source of the constituent is a point-source discharge or the natural base flow supplied by soil moisture or aquifer storage. In such cases the effect of increased discharge (due to precipitation, snowmelt, or reservoir release) is a dilution effect. The resulting relationship may be characterized by relationships such as

$$X = \lambda_1 + \lambda_2 \frac{1}{Q} + \varepsilon$$

$$X = \lambda_1 + \lambda_2 \frac{1}{1 + \lambda_3 Q} + \varepsilon$$

where X is concentration, Q is discharge, ε is an error term with zero mean and λ_1, λ_2, and λ_3 are coefficients ($\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\lambda_3 > 0$).

In other instances the constituent load may increase dramatically with an increase in discharge because of wash-off during storm runoff or because the constituent is primarily transported in a suspended state (adsorbed to particles) and suspended sediment loads (and concentrations) increase with discharge. The resulting relationship may be characterized by equations such as

$$X = \lambda_1 + \lambda_2 Q + \lambda_3 Q^2 + \varepsilon$$

where $\lambda_3 \geq 0$.

Analysis of the flow versus concentration relationship can be a useful addition to the examination of trends in water quality. Consider, for example, the Klamath River data. For the period of record (1972–1979) many of the samples of phosphorus in the earlier years occurred at higher discharges than many of those in the later part of the record. In fact a seasonal Kendall test on these discharge values indicates a downward trend with $p = 0.14$. Regressing phosphorus concentration versus discharge shows that they are highly correlated (see Figure 7). Given these three findings, (1) the seasonal Kendall test on the concentration data indicated a highly significant downward trend, (2) the discharges at which the observations were made show a downward trend as well (although not significant), and (3) the discharge and concentration are strongly and positively correlated; then it may be that the perceived trend in concentration may be a result of the particular history of discharge and not a consequence of any underlying change in the process by which phosphorus enters and is carried by the river.

In general, one would like to explore the two possibilities that (1) the perceived trend in concentration is an artifact of the particular record of discharges observed or (2) the perceived trend indicates that some change has taken place in the river basin such that the discharge versus concentration relationship has changed over time (that is, $E[X|Q]$ has changed). Conversely, one may also wish to explore the possibility that a 'no trend' result may have occurred (1) even though the relationship has changed but the flow record has masked the effect or (2) because the relationship itself has not changed.

To explore these possibilities we have applied the following residuals analysis procedure.

1. Use regression to find the 'best fit' relationship trying various functional forms $\hat{x} = f(Q)$ where $\hat{x}$ is the estimated concentration and $f(Q)$ is some function of discharge Q.

2. Given that a significant relationship exists, compute the time series of flow-adjusted concentration (FAC).

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

where w_{ij} is the FAC month i, year j, x_{ij} is the actual concentration, month i, year j.

3. Then, apply the seasonal Kendall test for trend and slope estimator to the time series of FAC values w_{ij}.

Table 5 provides some comparisons of the characteristics of concentration data x_{ij} and flow-adjusted concentration data w_{ij}. The stations used here were NASQAN stations with the 36 or more (flow, concentration) data pairs and for which the flow-adjusted relationship was significant ($\alpha = 0.10$). These comparisons show that the flow adjustment has a tendency to reduce seasonality and skewness but it cannot be expected to eliminate them at all stations. Thus the use of the seasonal Kendall procedures, as opposed to regression, is appropriate for both the x_{ij} and the w_{ij} series. The flow adjustment procedure has no consistent effect on serial correlation.

The use of the parametric procedure, linear regression, in this process is not entirely satisfactory for the reasons suggested in the previous sections of the paper. However, in this case linear regression is not being used for statisical testing but rather for the removal of variance that can be explained by an exogenous variable, discharge. It would, perhaps, be desirable to find a more resistant estimate of this relationship of concentration and discharge. There are some other methods for identifying time trends in the concentration-discharge relationship. These include multiple regression with time-varying coefficients (interaction models), and the recursive residuals approach suggested by *Brown et al.* [1975]. Exploration of the robustness and resistance of the flow-adjustment method proposed here or the other methods suggested above is beyond the scope of this paper.

For the Klamath River phosphorus data the w_{ij} series had a sample standard deviation of 0.07 mg/L and a sample coefficient of skew of 0.74. The seasonal Kendall test results are ($S = -18, Z = -0.78, p = 0.43$) and the seasonal Kendall slope estimate B is -0.002 mg/L per year (see Figure 8). Thus one may conclude that although there was a highly significant ($p = 0.007$) downward trend in concentrations over this 8-year period, there is no real indication of a change in the relationship between concentration and discharge or, stated more broadly, no indication of a change in the processes by which phosphorus is supplied to or transported by the river.

CONCLUSION

The methods presented in this paper (the seasonal Kendall test for trend, the seasonal Kendall slope estimator, and flow adjustment coupled with the seasonal Kendall test) are intended to be exploratory methods for identifying and quantifying changes in water quality time series. Together they provide means of identifying data sets where significant monotonic changes are occurring in the water quality variables of interest or where changes are occurring in the relationship between the variable and discharge. In addition, they provide an estimate of the magnitude of the trend over the period of record. These techniques are not a substitute for individualized analysis of the processes occurring at a station and in its basin. They are also not a substitute for visual examination of plots of the time series and other associated time series. It should be noted, however, that where considerable seasonality and/or skewness is present, it is not uncommon for the conclusions of subjective examination of the data to be substantially different from the conclusions arrived at from these procedures. This is probably due to the tendency for the observer to concentrate more on the extreme values of the series rather than on more subtle but regular trends in the bulk of the values nearer to the mean. As the number and length of water quality time series grows, it is desirable to have a set of objective automatic procedures that are reasonably powerful over a wide range of situations for identifying trends. We believe that the methods presented are useful and appropriate for this purpose.

REFERENCES

Borman, F. H., G. E. Likens, T. G. Siccama, R. S. Pierce, and J. S. Easton, The export of nutrients and recovery of stable conditions following deforestation at Hubbard Brook, *Ecol. Monogr., 44*, 255–277, 1974.

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

Briggs, J. C., Nationwide surface water quality monitoring networks of the U.S. Geological Survey, in *Establishment of Water Quality*

Monitoring Programs, pp. 49–57, American Water Resources Association, San Francisco, Calif., June 1978.

Brown, R. L., J. Durbin, and J. M. Evans, Techniques for testing the constancy of regression relationship over time, *J. R. Statist. Soc. Ser. B*, 149–163, 1975.

Guy, H. P., and V. W. Norman, Field methods for measurement of fluvial sediment, in *Techiques of Water-Resources Investigations of the United States Geological Survey*, book 3, chap. C2, 59 pp., U.S. Geol. Surv., Reston, Va., 1970.

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

Kendall, M. G., A new measure of rank correlation, *Biometrika, 30*, 81–93, 1938.

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

Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, vol. 1, 3rd ed., Hafner, New York, 1969.

Langbein, W. B., and D. R. Dawdy, Occurrence of dissolved solids in surface waters in the United States, *U.S. Geol. Surv. Prof. Pap. 501-D*, D115–D117, 1964.

Lettenmaier, D. P., Detection of trends in stream quality: Monitoring network design and data analysis, *Tech. Rep. 51*, Harris Hydraul. Lab., Dept. of Civil Eng., Univ. of Wash., Seattle, 1977.

Liebetrau, A. M., Water quality sampling: Some statistical considerations, *Water Resour. Res.*, *15*(6), 1717–1725, 1979.

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

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

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

Steele, T. D., E. J. Gilroy, and R. O. Hawkinson, An assessment of areal and temporal variations in streamflow quality using selected data from the National Stream Quality Accounting Network, *Open File Rep. 74-217*, U.S. Geol. Surv., Reston, Va., 1974.

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

Wolman, M. G., The Nation's rivers, *Science*, *174*(4012), 905–918, 1971.

(Received April 14, 1981;
revised September 15, 1981;
accepted November 2, 1981.)