93-03

Environment Canada

Water Science and Technology Directorate

Direction générale des sciences et de la technologie, eau

Environnement Canada

Trend Analysis for Environmental Data

By:
S.R. Esterby

NWRI Contribution # 93-03

93-03

TREND ANALYSES FOR ENVIRONMENTAL DATA

S. R. Esterby
National Water Research Institute
Burlington, Ontario, Canada
L7R 4A6

ABSTRACT

The recently heightened interest in the assessment of the current state of environmental conditions and the detection of any change in environmental conditions has led to a corresponding interest in statistical trend assessment methods. In this paper, the determination of any trend is considered as a part of the larger task of characterizing the variability of an environmental quality indicator. The features of several parametric and nonparametric methods are discussed with respect to their applicability for estimation and detection, the ability to handle changes of different forms, the inclusion of concomitant variables in the analysis and the inherent assumptions of the method. Examples of the analysis of water quality data using these methods are given.

MANAGEMENT PERSPECTIVE

The assessment of trends in environmental quality indicators and the reporting of such assessments are priorities for national and international agencies. Environmental data sets are generally complex and require variability, in addition to changes in long term tendency, to be adequately accounted for in the analysis for any trend. In the present paper, several statistical methods which account for seasonality and trend are compared with respect to the underlying model for trend. It is shown that clear objectives about the types of changes that are important for a particular set of monitoring data are essential, since the methods model the trend and seasonal components in different ways.

INTRODUCTION

Statements about the assessment of trends pervade the publications of national and international agencies concerned with environmental conditions. Consider the example of surface water quality. Monitoring networks have been established in a number of countries, where trend assessment was either the sole objective or one of a number of objectives of the network, for example, the Water Quality Assessment Program, Canada (Kwiatkowski, 1987) and the U.S. Geological Survey's Benchmark and NASQAN networks (Briggs, 1978). Recently, the heightened interest in environmental problems has led to new efforts to convey the information obtained from monitoring networks to decision makers and the public, often called state of the environment reporting, and to define environmental indicators to facilitate the integration of environmental and economic considerations in the decision-making process. Again, trend assessment is a priority.

In the Organization for Economic Cooperation and Development (OECD) report on environmental indicators (OECD, 1991), a preliminary set of 18 indicators of environmental performance was issued. The indicators were chosen to measure environmental performance with respect to level and changes in level of environmental quality and, although they are mainly calculated quantities such as estimates of CO_2 emissions, river quality indicators are measured concentrations of dissolved oxygen and nitrate in river water. In all cases, graphics such as bar graphs, pie charts, or plots, with lines joining yearly values, are used to present trends. The preliminary set of environmental indicators issued by Canada (Environment Canada, 1991), using selection criteria that included support by sufficient data in order to show trends over time, contained a higher proportion of directly measured indicators than the OECD set. For example, as well as calculated CO_2 emissions, atmospheric CO_2 concentration was also taken as an indicator. Other examples of directly measured indicators are phosphorous and nitrogen concentrations in surface water. Graphics similar to those in the OECD report were used to illustrate changes in the indicators.

The question of what is meant by trend arises. The above examples illustrate the expression of objectives in very general terms. It seems reasonable to understand trend as the general direction and tendency (Oxford Dictionary) of the quantity of interest. It then rests with the individuals working with this problem at the scientific level to make the general definition specific so that hypotheses may be tested and the magnitude of change estimated.

Since the primary objective of this paper is to consider statistical methods suitable for environmental trend analysis, and an important contribution that statistical methodology can make is to provide precise and unbiased estimates of trend, it is appropriate to consider trend analysis as part of the larger task of characterizing variability in the data to be used to assess trend. This is particularly important with respect to state of the environment reporting and environmental indicators, since simple indicators or reports are being called for (e.g. Canada's Green Plan in Brief, Department of Supply and Services, 1990). The ideal is a simple and clear presentation of a properly estimated well-defined quantity. Much work has been done towards achieving this goal for monitoring data, and some examples of methods of analysis of atmospheric and water quality variables are considered here. Several of the methods are then applied to a single water quality data set.

In discussing the methods, the major objectives are to show how other sources of variability are separated from trend, and to describe the model for trend underlying the methods. The ability to accommodate covariates in the analysis and the existence of measures of uncertainty are also considered. Methods are not described in detail since this can be found in the references cited. Attention is restricted to a set of observations on one environmental quality variable collected over time at one location under a monitoring program where sampling is done at approximately equal intervals in time, the number of years of data available is short to moderate and a number of samples have been taken within each year.

METHODS OF ACCOUNTING FOR SEASON IN TREND ANALYSIS

BLOCKING BY SEASON

Consider observations on an environmental variable obtained at m identifiable times within a year, over n years. For example, observations could be made daily, monthly or seasonally. It is assumed, for the moment, that identification of the time within the year at the particular interval chosen, for example, identification of the month of observation, is adequate. Let y_{ij} denote the observation of the variable at the j^{th} sampling time within the year, for year i, where $i=1,2,\ldots,n$ and $j=1,2,\ldots,m$.

Many analyses for trend have been based on an additive model of within-year, between-year, and error components. An example of such a model is

$$y_{ij} = \alpha_i + \beta_j + \epsilon_{ij} \qquad (1)$$

where α_i is the component for year i, β_j the component for the j^{th} sampling time within the year and ϵ_{ij} the residual variability. The objective is to determine changes over years free of seasonal effects.

Nonparametric Methods

Recently, nonparametric methods have been widely used in the analysis of water quality data because fewer assumptions must be satisfied than with parametric methods. Hirsch et al. (1982) illustrated the use of a blocked Kendall's τ, the seasonal Kendall test for trend, and extended the nonparametric slope estimator of Thiel (1958) and Sen (1968) to account for seasonality (the seasonal Kendall slope estimator). Gilbert (1987) gave a simple procedure for a confidence interval for the seasonal Kendall slope estimator.

The seasonal Kendall test for trend is an intrablock method (van Belle and Hughes, 1984). The effect of season is eliminated by calculating the test statistic separately for each season and summing these statistics to obtain an overall test statistic. In terms of the set of observations, (y_{ij}), this becomes clear if the data are arranged as in Table 1, where, for ease of expression, season j is used instead of the j^{th} sampling time within the year. The possibility of missing data is included through the subscript on the number of observations in a season, n_j for $j=1,2,\ldots,m$. The test statistic calculated for season j is

$$S_j = \sum_{i<k} \mathrm{sgn}(y_{kj} - y_{ij}) \qquad (2)$$

where

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x>0 \\ 0 & \text{if } x=0 \\ -1 & \text{if } x<0 \end{cases} \qquad (3)$$

If this statistic is used for a trend test on the data for month j, it involves testing the hypothesis, H_0, that y_{1j}, y_{2j}, ..., $y_{n_j j}$ are a sample of n_j independent and identically distributed random variables. A two sided alternative, H_1, is the hypothesis that the distribution is not the same for all y_{ij}, y_{kj}, with $i,k\leq n$ and $i\neq k$. It is used as a test for trend because it is powerful for the alternative of a monotonic trend. If (1) is taken to represent all of the observations, the hypotheses for the test within season j are H_0: $\alpha_i=0$ $(i=1,2,\ldots,n)$ and H_1: $\alpha_1\leq\alpha_2\leq\ldots\leq\alpha_n$ or $\alpha_1\geq\alpha_2\geq\ldots\geq\alpha_n$ with at least one strict inequality. There is an explicit assumption of independence between seasonal values at least one year apart under both H_0 and H_1. Further, under the assumption of independence, it is a test of trend alone only if the additional assumption, that the effect for season j is constant over all years, holds (i.e. a model such as (1) above with no interaction terms). If this is true, taking differences, $y_{kj}-y_{ij}$, will remove the seasonal component.

The seasonal Kendall test statistic is

$$S = \sum_{j=1}^{m} S_j \tag{4}$$

and it is used to test the hypothesis of independent and identical distributions within season against the alternative that this does not hold in at least one season. It is inappropriate to combine the S_j if the trend is not homogeneous over seasons and van Belle and Hughes (1984) showed how to test for homogeneity, provided the observations within a year can be assumed independent.

The analogous nonparametric estimate of slope is the median of all possible slopes based on pairs of observations which are in the same season, and thus, is clearly satisfactory only for linear changes. Let B denote this slope estimator, then B is the median of

$$(y_{kj} - y_{ij})/(k-i) \qquad (5)$$

for $i<k$ and j fixed, taken over $j=1,2,\ldots,m$.

Nonparametric estimation of the magnitude of a step change in a water quality variable has been considered by Hirsch (1988) and the seasonal Hodges-Lehman estimator, Δ_{SHL}, which is the analogue of B, was found to perform well. Suppose that a step change occurred after year ℓ. Then the objective is to estimate the change in level between the two periods $i=1,2,\ldots,\ell$ and $i=\ell+1,\ldots,n$. For season j, all possible differences $(y_{kj} - y_{ij})$ are calculated where $i=1,2,\ldots,\ell$ and $k=\ell+1,\ldots,n$, and this is done for $j=1,2,\ldots,m$. Δ_{SHL} is the median of the $m\ell(n-\ell)$ differences. Helsel and Hirsch (1992) give methods for calculating the confidence interval for the Hodges-Lehman estimator.

A corresponding nonparametric method for testing for the existence of a step change and estimating the point of this change is given by Pettitt (1979). Although this has not been applied to water quality variables it has been used to look for changes in diatom concentrations in sediment cores (Esterby et al., 1986).

The seasonal Kendall test is not robust to serial dependence (Hirsch et al., 1982) and the variance of S under different assumptions about serial dependence has been obtained by Zetterqvist (1991) and El-Shaarawi and Niculescu (1992). Hirsch and Slack (1984) have proposed a modification in the presence of serial dependence, and Lettenmaier (1988) has given a similar test.

The Spearman rank correlation coefficient, r_s, can replace Kendall's τ when blocking is present (Taylor, 1987). The coefficient r_s has been used to test for trend under blocking by month (El-Shaarawi et al., 1983). The Spearman partial rank correlation test has been described by McLeod et al. (1991). Lettenmaier (1976) has considered the effect of serial dependence on the Spearman coefficient, the t-test and Mann-Whitney test for step change.

Least Squares Regression

The regression analysis under blocking is also based on model (1) and thus, changes over years are examined separately within each season first and then, over all seasons, if the changes are shown to be homogeneous. The differences between the nonparametric and the regression methods are that, in the latter, the form of the change will be an explicit part of the model, and more general forms can readily be included. For season j, a linear trend, an example of a curvilinear trend, and a step change at year k can be modelled respectively as

$$
\begin{aligned}
y_{ij} &= \alpha_j + \beta_j i + \epsilon_{ij} \\
y_{ij} &= \alpha_j + \beta_{j1} i + \beta_{j2} i^2 + \epsilon_{ij} \qquad (6) \\
\text{and} \quad y_{ij} &= \alpha_{j1} + \beta_{j1} + \epsilon_{ij} \qquad i \le k \\
&= \alpha_{j2} + \beta_{j2} + \epsilon_{ij} \qquad i > k
\end{aligned}
$$

Regression methods have been used to examine changes over time separately for each month, for example, by El-Shaarawi et al.

(1991). Tests of significance of a change over time of a specified form and estimation of the change are part of the regression methodology. Similarly, a test for homogeneity of change can be made prior to estimating an overall change. A method for estimating the point of change in a regression relationship (Esterby and El-Shaarawi, 1981) has been shown to be useful for environmental variables.

The assumptions of independence and the adequacy of the method of blocking to account for seasonality, which were required for the intrablock nonparametric methods, are also required for the regression methods. The additional assumptions required for regression are homogeneity of error variances and normality of the errors. Techniques which help to meet these assumptions are weighted regression and transformations. Serial correlation can be accommodated by generalized least squares, but as with the nonparametric methods, is more problematic than other data features. This aspect is discussed further under cyclical seasonal components.

CYCLICAL SEASONAL COMPONENTS

Some degree of smoothness in the seasonal component is often more realistic. Harmonic components have provided adequate description of seasonal variability in environmental variables which are influenced by cyclical physical processes. Examples are models of physical water measurements such as temperature (McMichael and Hunter, 1972; Neilson and Hsieh, 1982) and chemical water quality parameters (El-Shaarawi et al., 1983). Such deterministic seasonal components seem more appropriate for a time series which shows a consistent strong annual cycle (McMichael and Hunter, 1972) and, evidence for this may actually come from the values of parameters obtained in trying to fit a purely stochastic model (Reinsel and Tiao, 1987).

An additive model for monthly data, which assumes a stable annual cycle and a linear trend, could be represented by

$$y_{ij} = \mu + \beta_1 \sin\frac{2\pi j}{12} + \beta_2 \cos\frac{2\pi j}{12} + \omega(i+j/12) + \eta_{ij} \qquad (7)$$

where y_{ij} is the value of the environmental variable in month j of year i. Additional harmonic terms, other forms for non-seasonal deterministic terms and independent or dependent errors can be accommodated in models of this type. Reinsel and Tiao (1987) found that a model with annual and semiannual seasonal components, a two-phase trend component, and η_{ij} modelled as an autoregressive process, provided a good description of the variability in monthly averages of stratospheric total ozone. An iteratively reweighted least squares procedure was used to obtain parameter estimates and the order of the autoregressive process was determined from the sample autocorrelation and partial autocorrelation functions of the residual series.

The general form of the mean plus seasonal component in (7) is equivalent to using monthly indicator variables, i.e. blocking, but now a more parsimonious representation of the monthly mean is given by the fewer sinusoidal terms. Here the trend over the whole time period can be thought of as being determined from the original observations minus the value at month j given by the harmonic seasonal component, which itself is estimated over all observations. Under blocking, the trend term is estimated in the regression analysis from the original observations minus the appropriate monthly mean and in the seasonal nonparametric trend estimators (slope or step change) by differences between observations within the same month. Thus numerical differences in trend estimates using regression methods under the same error model would result from the different ways of modelling the seasonal component, with the further difference of method when comparing with nonparametric results.

SMOOTHED SEASONAL COMPONENTS

In some cases the seasonal variation may be poorly represented by sinusoidal terms and a general smoothing procedure may be more appropriate. An example of a robust smoother is the locally-weighted regression smoother, LOESS (Cleveland and Grosse, 1991), which has been used to show the general tendency in large sets of water quality data (Bodo, 1989). A seasonal-trend decomposition procedure based on LOESS, STL, (Cleveland et al., 1990) has been applied to monthly averages of atmospheric CO_2. STL iteratively smooths the detrended seasonal component and then the deseasonalized trend component. The series is decomposed into seasonal, trend and residual components. The major purpose of STL is to estimate the seasonal component so seasonal adjustment may be performed, but post-smoothing of the sum of the trend and residual components may provide an adequate description of the low-frequency variation in the data. The algorithm is designed to reduce the competition for the same variation by the high frequency (seasonal) and low frequency (trend) terms.

Of particular interest to the present paper, is the method of smoothing the seasonal component. It is assumed that change over time for data collected in a particular season will be smooth, whereas changes from season to season within a year will be irregular. The seasonal component is constructed by fitting a slowly changing smooth curve to the measurements over years within a season (called cycle subseries by Cleveland et al., (1990) and measurements within a block here) separately for each season. The estimates $\tilde{y}_{1j}, \tilde{y}_{2j}, \ldots, \tilde{y}_{nj}$ obtained for $j=1,2,\ldots,m$ are then rearranged in the order of observation, to give $\tilde{y}_{11}, \tilde{y}_{12}, \ldots, \tilde{y}_{1m}, \tilde{y}_{21}, \ldots, \tilde{y}_{n1}, \tilde{y}_{n2}, \ldots, \tilde{y}_{nm}$, which provides the temporary seasonal component. The result of a low-pass filtering is subtracted from the temporary seasonal component to give the final seasonal component at the current loop of the procedure.

This low frequency smoothing within season, which generates the seasonal component in STL, is providing a picture of gradual change in exactly the same subset of the data used to test for and estimate trend in the blocking methods above. In fact, this form of plot, without the smoothed curve, has been recommended by Esterby et al. (1991) for use with the seasonal Kendall τ and nonparametric slope estimator so that a plot which corresponds to the method of analysis can be examined in the course of the analysis. Despite the way the individual terms of the seasonal component, $\tilde{y}_{ij}$, are obtained, it is the collection of terms in order of observation that provides the seasonal component and thus this method is more similar to the use of a cyclical seasonal component than blocking methods. The STL seasonal component is really a global estimate of the seasonal component (i.e. over all data) and not one derived from values within a particular season.

An important difference between STL and the previous methods are that there is no underlying assumption of homogeneity of the within-season change over all seasons, as can be seen from the cycle sub-series plots of Cleveland et al. (1990). This allows more flexibility in modelling the seasonal variation in individual years and results from the underlying assumptions of separating total variation into low frequency, high frequency and residual components. The assumption, that the value, y_{ij}, is an adequate representation for season j in year i, is necessary, but none of the other assumptions considered for the previous methods are necessary since STL is providing primarily graphical summaries of the variability in the data. However, the authors note that by using STL in combination with a standard ARIMA model, confidence intervals for the seasonal component could be obtained.

OTHER CONSIDERATIONS IN THE ANALYSIS FOR TREND

Fixed interval versus time of observation

For all environmental series the interval of observation and/or interval used to provide the data for analysis is important. Environmental observations collected over time exhibit variation at each level of division of the time scale. If daily data is to be used, then either at the sampling or data analysis stage, decisions have had to be made to make the observation for a day representative of the day, under some assumptions. For example, if diurnal variation occurs, two possibilities are that sampling be done at the same time each day, or an average taken of samples collected over the day. Similarly, using one value of the water quality variable to represent the month requires careful consideration, since the hydrological cycle differs from year to year due to changes in weather conditions.

An alternative to using fixed monthly intervals is to use the day of observation and fit a smooth curve to the seasonal variation within year. Between year differences can then be looked at as step changes or smooth changes. Esterby et al. (1991) fitted, to river water quality data, regression models of the form

$$y_{ij} = \mu + \beta_{i1} \cos \omega t_{ij} + \beta_{i2} \sin \omega t_{ij} + f(i,j) + \epsilon_{ij} \qquad (8)$$

where the pair (y_{ij}, t_{ij}) is the value of the water quality variable and day of measurement within the year for the j^{th} observation in year i, $\omega=2\pi/365$, f(i,j) represents the change over years, either linear trend as αi or yearly mean α_i for i=1,2,...,n, and the subscripts for y and t are as shown in Table 1. Curvilinear change, additional harmonic terms, time variable as (i+j/12) and dependent errors are all possible modifications. The results of the regression methods were compared with results using the seasonal Kendall trend test and seasonal nonparametric slope estimate for specific conductance measurements at locations along a river system.

Robust smoothing methods could also be used to obtain seasonal, trend and residual components when day of measurement is used. However, the algorithm in STL is not applicable because cycle sub-series are not available. Either observations are not taken on the same day each year or the hydrological cycle is not at the same point on a given day each year. The procedures of smoothing detrended seasonal components and deseasonalized trend components could be applied, however the smoothing step for season is done over the entire data set, as for trend, but with a frequency different from the trend frequency.

<u>Inclusion of covariates</u>

The question of whether covariates should be included in the analysis for trend is complex and depends upon the interconnected matters of the definition of trend, form of environmental quality variable, reason for the analysis and method of analysis. With respect to water quality, Zetterqvist (1991) states what seems to be the motivation when covariates are included. That is, if the objective is to test for or estimate trends in environmental variables due to human activities, then covariates that will remove variation produced by natural phenomena should be included. If the change over time of the variable, irrespective of what may be the underlying reason, is wanted, such covariates are not needed. However, when covariates are included, the objective will be met only if the relationship between environmental variable and covariate is adequately modelled.

Of the methods considered here, regression methods allow the inclusion of covariates directly as independent variables and this includes the case of cyclical seasonal components. Adjustment of variables is required before nonparametric methods can be used. Smith and Rose (1991) compare multiple regression with the Kendall correlation coefficient τ where adjustment for the covariate was to a) the dependent variable only, and b) both dependent variable and time. They show that method a) can produce misleading results and method b) is generally less powerful than multiple regression. The LOESS algorithm (Cleveland and Grosse, 1991) handles vector valued predictors but this feature has not been used in STL.

Transfer function models have also been used to include flow, as well as other covariates, in the analyses of water quality series (McLeod et al., 1983, and Zetterqvist, 1991). McLeod et al. (1991) propose a methodology for inclusion of covariates when water quality series are not amenable to analyses by parametric time series methods.

Flow, as a covariate of water quality parameters, provides an example where careful modelling is required. Zetterqvist (1991) found transfer function models adequately accounted for intervention-type effects on phosphorous concentrations in the Ljungbyån River but were not able to model the complicated relationship with covariates such as flow. Esterby et al. (1990) noted that the relationship between a water quality variable and flow may change with season and differ by location on a river system. Teti (1984) gives another example of time-variant patterns due to different sources of water.

Missing data

In the presence of missing data, the nonparametric and regression methods, excluding methods where serial dependence is modelled, can be applied. However, the interpretation of the results should always be done in view of how much information is missing due to gaps in the record. The regression methods with serially dependent errors require equal spacing of observations for the estimation of the error component.

EXAMPLE

Water quality monitoring for major ions and nutrients at locations in the South Saskatchewan River basin, in the provinces of Alberta and Saskatchewan, has been conducted in some form since the early 1950's (Munro, 1987). The water quality issues in the basin, which are relevant to major ions and nutrient monitoring, result from deterioration of water quality due to eutrophication, increased salinity, and industrial and urban discharge of pollutants. Specific conductance data at a South Saskatchewan River location, below major urban centres and the irrigation district of southern Alberta, are considered here. Sampling was done at approximately monthly intervals and the data record between 1975 to 1986 was used since there was a sample within each month over this period. The results from the nonparametric analysis of the 1978 to 1985 data blocked by month and the regression analysis using a sinusoidal seasonal component and day of sampling (Esterby et al., 1991) will be compared. The longer period, 1975 to 1986, has been used for the run of the STL program and, for comparison, the model with a sinusoidal seasonal component and day of sampling is also shown.

Under blocking, the change over time being considered is that within each month and Figure 1 allows an initial visual impression. The test of the hypothesis of randomness using the Kendall statistic separately for each month provides a significant result at the 0.05 level only for April (Table 2). Since the test for homogeneity is not significant, the seasonal Kendall statistic can be used to draw a conclusion for all months, which is that there is no evidence against the hypothesis of randomness. The confidence interval based on the nonparametric slope estimator gives the same conclusion.

The regression analysis with a sinusoidal seasonal component and a linear trend over years provides an estimate of the slope comparable to that from the nonparametric method (Table 2). Examination of the plot of the residuals versus time (not shown here) suggests why the change from year to year in the mean level is not well modelled by a linear term. The change that occurs in this period is one of step changes, with the mean level of 436 in 1979 and 1980 being higher than the mean level of 394 in the other years (Esterby et al., 1992). Such a conclusion would not appear to hold for individual months although there are generally higher values in the early years for most months. The regression model however allows for yearly differences in the seasonal cycle in the form of different phase shifts and amplitudes, and the change over years that is looked at is in the mean level about which this yearly seasonal cycle fluctuates.

Other stations on the same river system gave similar conclusions about step changes and the estimated yearly levels are shown in Figure 2, where the progression from headwater site to location of greatest impact corresponds to going from top to bottom in the figure, and AK0001 is the location considered in detail here. The regression model fitted to the period 1974 to 1987 is shown in Figure 6, with the residuals from the model as the second plot.

Analysis under blocking, either by regression or nonparametric methods, allows assessment of changes in individual seasons or months to be made directly. This may be necessary when sampling does not occur throughout the year, e.g. it may be done only in certain seasons, or when one expects or wants to know about changes which correspond to different events, such as varying sources of water over the year.

The components of the decomposition of specific conductance between 1975 and 1986 are shown in Figure 3. The form of the seasonal cycle is, in a general way, comparable to the 2 component sinusoidal (i.e. two frequencies) model fitted by regression (Figure 6). The shorter period used in the nonparametric and regression analyses above was 1978 to 1985, which corresponds to month 37 to 120 in the Figure 3. Months 49 to 71 are in years 1979 and 1980 and an increase in concentration is shown here, in agreement with the regression results. Other plots which are provided by the STL program are the cycle sub-series plots of the seasonal components, which correspond to Figure 1, but are now estimated quantities, and the line plots of these components by month (Figure 5). The latter plot dramatically shows that the decomposition model allows both different form and direction of change within a month from month to month.

ACKNOWLEDGEMENT

The STL program used here was the Fortran implementation available as public domain software (Cleveland et al., 1990) and I would like to thank the individuals involved for making STL available.

REFERENCES

Bodo, B.A., 1989. Robust graphical methods for diagnosing trend in irregularly spaced water quality time series. Environmental Monitoring and Assessment, 12, 407-428.

Briggs, J.C., 1978. Nationwide surface water quality monitoring networks of the U.S. Geological Survey, in Establishment of Water Quality Monitoring Programs, American Water Resources Association, San Francisco, California, pp. 49-57.

Cleveland, W.S. and Grosse, E., 1991. Computational methods for local regression. Statistics and Computing, 1, 47-62.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I., 1990. STL: A seasonal-trend decomposition procedure based on LOESS. Journal of Official Statistics, 6, 3-73.

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. J. Great Lakes Research, 9, 234-240.

El-Shaarawi, A.H., Esterby S.R. and Yip, J., 1991. Trend Assessment Techniques: Application to Prairie Provinces Water Board Water Quality Data Set. Prairie Provinces Water Board Report No. 113, Regina, Saskatchewan, Canada.

El-Shaarawi, A.H. and Niculescu, S., 1992. On Kendall's τ as a trend test for auto correlated data. National Water Research Institute Contribution No. 92.

Environment Canada, 1991. A Report on Canada's Progress Towards a National Set of Environmental Indicators. State of the Environment Report No. 91-1, Supply and Services Canada, Ottawa.

Esterby, S.R. and El-Shaarawi, A.H., 1981. Inference about the point of change in a regression model. Applied Statistics, 30, 277-285.

Esterby, S.R., Delorme, L.D., Duthie, H. and Harper, N.S., 1986. Analysis of multiple depth profiles in sediment cores: an application to pollen and diatom data from lakes sensitive to acidification. Hydrobiologia, 141, 207-235.

Esterby, S.R., El-Shaarawi, A.H. and Block, H.O., 1990. The Estimation of Water Quality Changes between Locations Within a River System. National Water Research Institute Contribution No. 90-45.

Esterby, S.R., El-Shaarawi, A.H., Keeler, L.C. and Block, H.O., 1991. Testing for trend in water quality monitoring data. National Water Research Institute Contribution No. 91-02.

Esterby, S.R., El-Shaarawi, A.H., and Block, H.O., 1992. Detection of water quality changes along a river system. Environmental Monitoring and Assessment (in press).

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

Helsel, D.R. and Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier, Amsterdam.

Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982. Nonparametric tests for trend in water quality. Water Resources Research, 18(1): 107-121.

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., 1988. Statistical methods and sampling design for estimating step trends in surface water quality. Water Resources Bulletin, 24, 493-503.

Kwiatkowski, R.E., 1987. Water Quality Branch monitoring, an historical review. In: Water Quality Monitoring Design. Proceedings of the Technology Transfer Workshop. Technical Workshop Series No. 6, Environment Canada, Ottawa, pp. 5-17.

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

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

McLeod, A.I., Hipel, K.W. and Comancho, F., 1983. Trend assessment in water quality time series. Water Resources Bulletin, 19, 537-547.

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

McMichael, F.C. and Hunter, J.S., 1972. Stochastic modelling of temperature and flow in rivers. Water Resources Research, 8, 87-98.

Munro, D.J., 1987. Water quality monitoring in the South Saskatchewan basin. In: Water Quality Monitoring Design - Proceedings of the Technology Transfer Workshop. Technical Workshop Series No. 6, Environment Canada, Ottawa, pp.103-114.

Neilson, B.J. and Hsieh, B.B., 1982. Analysis of water temperature records using a deterministic-stochastic model. In: Time Series Methods in the Hydrosciences, Editors A.H. El-Shaarawi and S.R. Esterby, Elsevier, Amsterdam, pp.465-473.

OECD, 1991. Environmental Indicators. A Preliminary Set. OECD Publications Service, Paris.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. Applied Statistics, 23, 126-135.

Reinsel, G.C. and Tiao, G.C., 1987. Impact of chlorofluoromethanes on stratospheric ozone. Journal of the American Statistical Association, 82, 20-30.

Sen, P.K., 1968. On a class of aligned rank order test in two-way layouts. Ann. Math. Statist., 39, 1115-1124.

Smith, E.P. and Rose, K.A., 1991. Trend detection in the presence of covariates: stagewise versus multiple regression. Environmetrics, 2, 153-168.

Taylor, J.M.G., 1987. Kendall's and Spearman's correlation coefficients in the presence of a blocking variable. Biometrics, 43, 409-416.

Teti, P., 1984. Time-variant differences in chemistry among four small streams. Water Resources Research, 20, 347-359.

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

van Belle, G. and Hughes, J.P., 1984. Nonparametric tests for trend in water quality. Water Resources Research, 20, 127-136.

Table 1. Data collected in m seasons over n years

	Season						
Year	1	2	3	...	j	...	m
1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1m}
2	y_{21}	y_{22}	y_{23}	...	y_{2j}	...	y_{2m}
3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3m}
⋮	⋮						
i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{im}
⋮	⋮						
n	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{nm}
Number of Observations	n_1	n_2	n_3		n_j		n_m

Table 2. Summary of analysis by nonparametric methods and regression with sinusoidal seasonal component.

Test	Period	Variate Value	
Kendall[a]	Jan.	-0.52	
	Feb.	-0.52	
	Mar.	-1.15	
	Apr.	-2.19*	
	May.	0.73	
	June.	1.15	
	July.	-0.52	
	Aug.	0.31	
	Sept.	0.84	
	Oct.	-0.31	
	Nov.	0.52	
	Dec.	-0.94	
Seasonal Kendall[a]	Year	-0.78	
van Belle and Hughes[b]	Year		
Trend		0.66	
Homogeneity		12.32	
Slope Estimate and 95% Confidence Interval			
Nonparametric	-1.9	-5.1	1.7
Regression	-2.2	-5.6	1.2

* Significant at the 0.05 level.

[a] The entries in the table for the Kendall and seasonal tests are the standard normal deviate form, with continuity correction, as described by Hirsch et al. (1982).

[b] The statistics for test of trend and homogeneity, without continuity corrections, as described by van Belle and Hughes (1984), are distributed approximately as chi squares with 1 and 11 degrees of freedom, respectively.

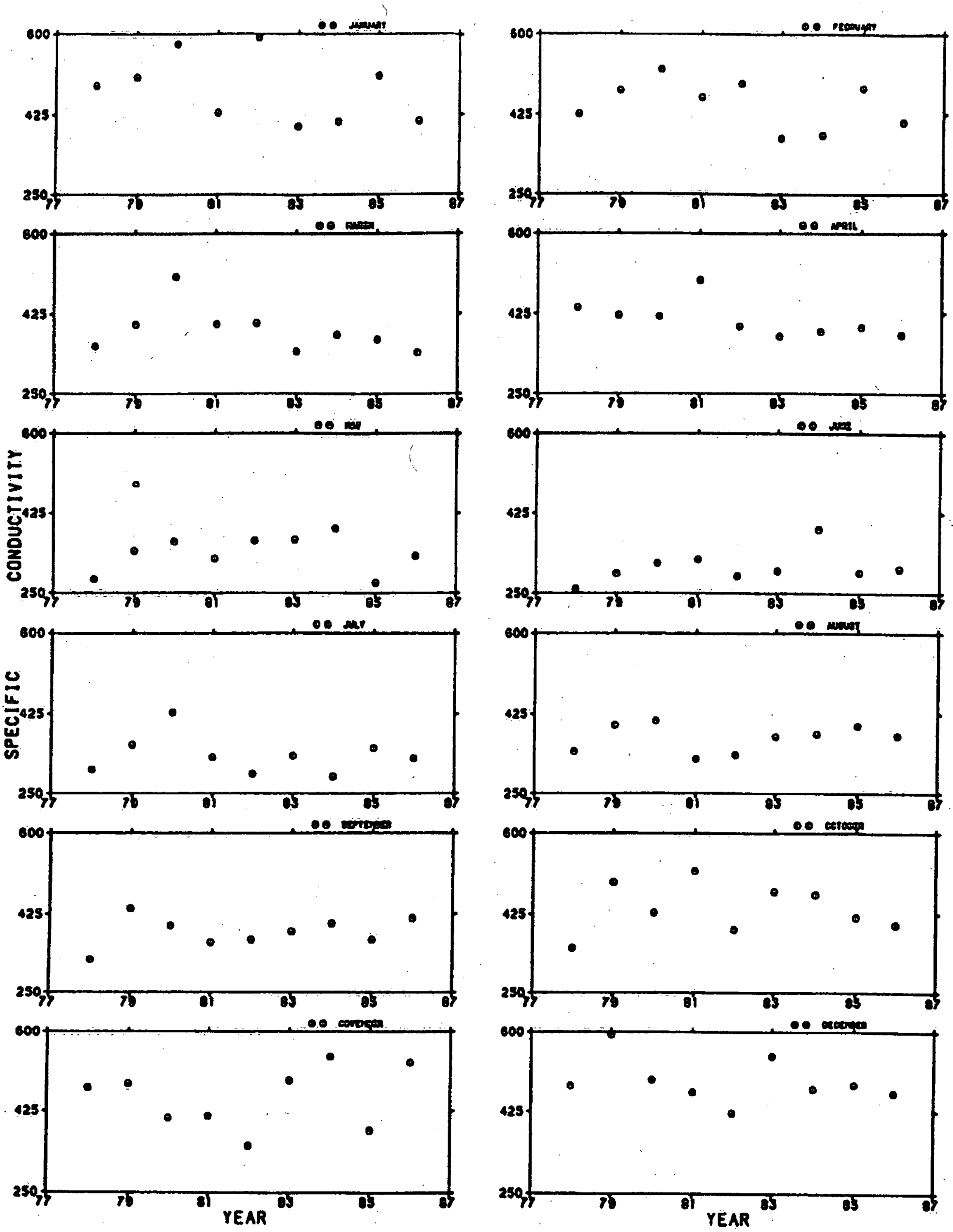

Figure 1 Plot of the specific conductivity series separately by month (block).

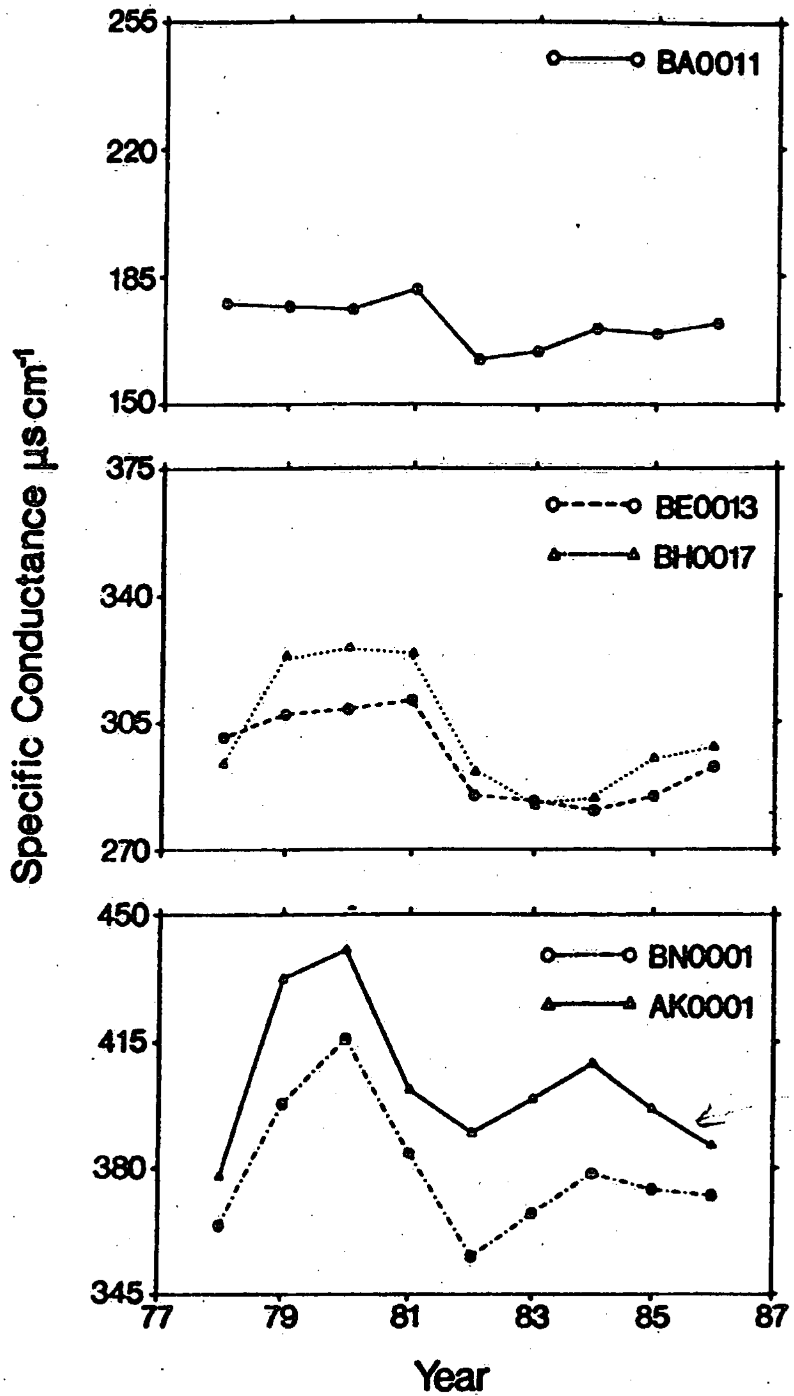

Figure 2 Estimate of yearly mean level from model with sinusoidal seasonal component and day of sampling for 5 locations.

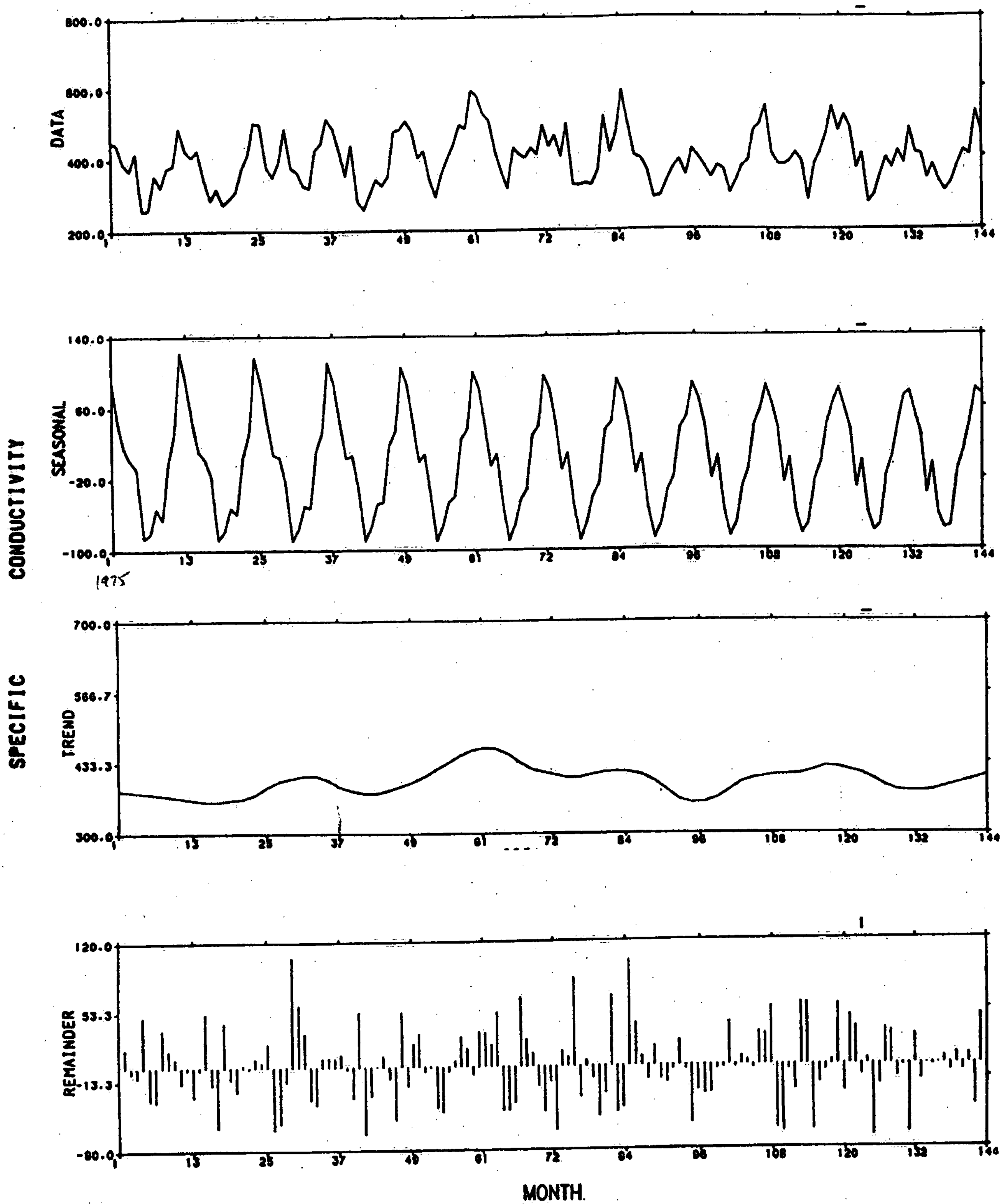

Figure 3 Components of the decomposition of the specific conductivity series 1975 to 1986 by STL.

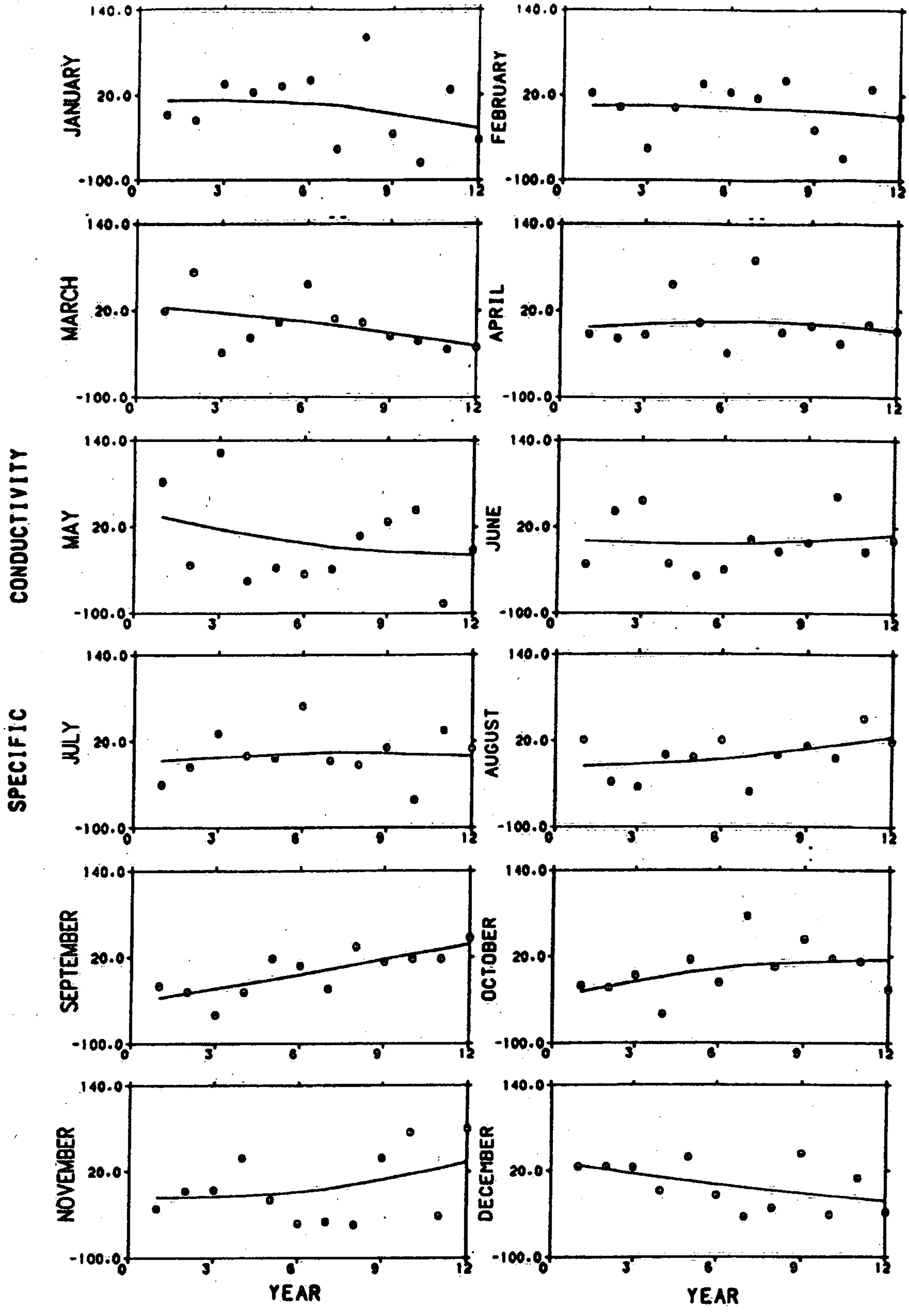

Figure 4 Cycle sub-series of seasonal plus remainder (points) and seasonal (line) determined by STL program.

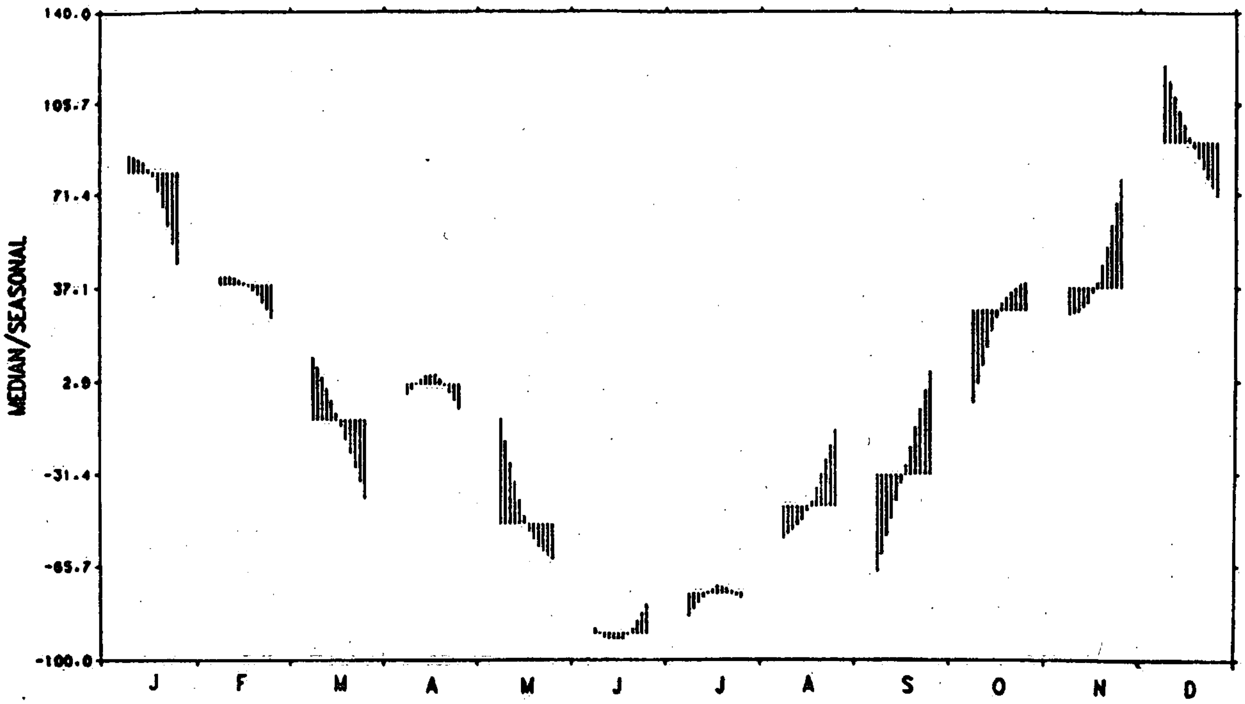

Figure 5 Seasonal component cycle sub-series plotted by month as line above or below median of the components for the month.

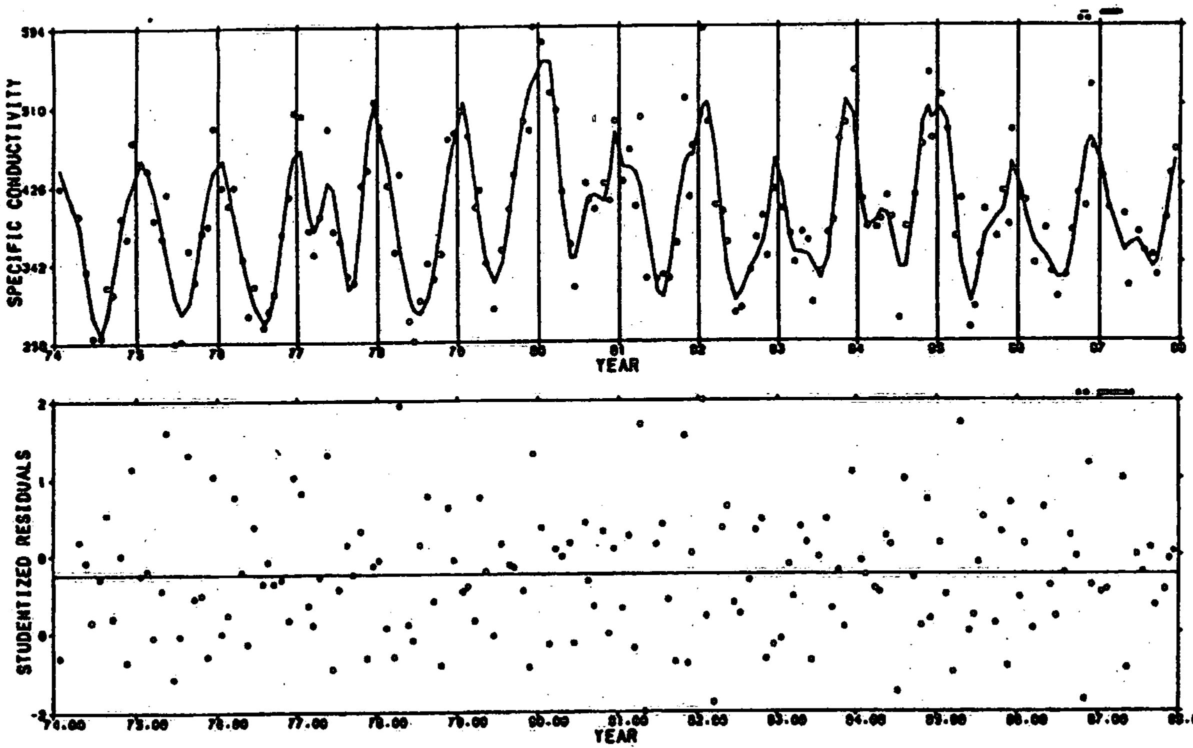

Figure 6 Model with sinusoidal seasonal component and day of sampling (top) and residuals from this model plotted vs. day of sampling.

Environment Canada / Environnement Canada

Canada

Canada Centre for Inland Waters
P.O. Box 5050
867 Lakeshore Road
Burlington, Ontario
L7R 4A6 Canada

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, Saskatchewan
S7N 3H5 Canada

St. Lawrence Centre
105 McGill Street
Montreal, Quebec
H2Y 2E7 Canada

Place Vincent Massey
351 St. Joseph Boulevard
Gatineau, Quebec
K1A 0H3 Canada

Centre canadien des eaux intérieures
Case postale 5050
867, chemin Lakeshore
Burlington (Ontario)
L7R 4A6 Canada

Centre national de recherche en hydrologie
11, boul. Innovation
Saskatoon (Saskatchewan)
S7N 3H5 Canada

Centre Saint-Laurent
105, rue McGill
Montréal (Québec)
H2Y 2E7 Canada

Place Vincent-Massey
351 boul. St-Joseph
Gatineau (Québec)
K1A 0H3 Canada