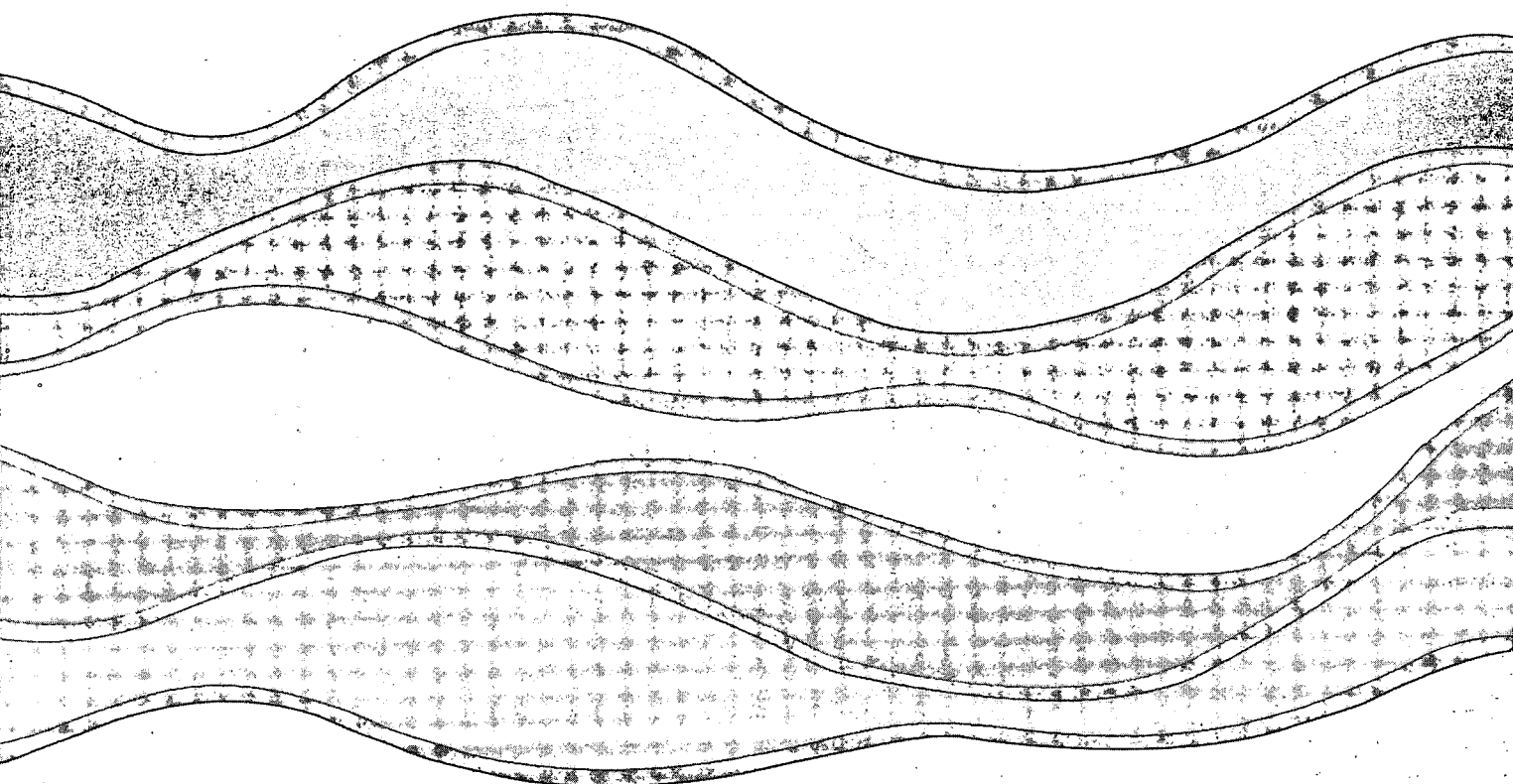


CCIW
MAR 5 1992
LIBRARY

NATIONAL
WATER
RESEARCH
INSTITUTE

INSTITUT
NATIONAL
de RECHERCHE
sur les
EAUX



**THE EVALUATION OF THE EFFECT OF
METHOD AND LABORATORY CHANGES
USING MONITORING NETWORK DATA**

**S.R. Esterby, A.H. El-Shaarawi
and H.O. Block**

NWRI Contribution No. 90-44

TD
226
N87
No. 90-
44
c. 1

**THE EVALUATION OF THE EFFECT OF
METHOD AND LABORATORY CHANGES
USING MONITORING NETWORK DATA**

**S.R. Esterby¹, A.H. El-Shaarawi²
and H.O. Block³**

¹ Lakes Research Branch

**² Rivers Research Branch
National Water Research Institute
Burlington, Ontario L7R 4A6**

**³ Water Quality Branch
Western and Northern Region
Calgary, Alberta T2P 2M7**

NWRI Contribution No. 90-44

MANAGEMENT PERSPECTIVE

The effect of changes in analytical methodology, sampling methodology or laboratory on water quality measurements may need to be evaluated from monitoring data collected over time due to either inadequate evaluation of compatibility prior to the change or to other circumstances. The possibility of doing this is considered for sodium and nitrate plus nitrite concentrations measured at locations on the Bow and South Saskatchewan Rivers between 1977 and 1987. Models were fitted to account for seasonal and yearly variation plus methodological changes. Sodium concentrations were found to be unaffected by a single laboratory change but inconclusive results were obtained for nitrate plus nitrite. Although such a procedure may be used when no better alternative is available, the present results underline the need for proper evaluation of the effect of methodological changes prior to the implementation of such changes.

Useful insights into modelling the inherent variability in the monitoring data were also obtained. Variation introduced by the yearly hydrological regime is the largest source of variability in the data sets and to obtain precise estimates of other changes, seasonal terms must be included in models. To illustrate, the percentage of variation explained did not exceed 5% when only method or laboratory terms were in the model, compared with 72 to 87% for $\text{NO}_3 + \text{NO}_2$ and 38 to 76% for Na when yearly mean and seasonal terms were also included. At the monthly sampling frequency, it was difficult to characterize seasonality if it was more complex than a summer

minimum. The nature of the seasonal variation did not consist only of differences in level between seasons and thus a comparison of seasonal means will not be appropriate in some cases. For example, the $\text{NO}_3 + \text{NO}_2$ concentrations in the first four to five months of the year consists of a rapid decline. The report is to be incorporated into the report of the departmental task force on design and assessment of water quality data collection programs.

PERSPECTIVES DE LA DIRECTION

Il pourrait être nécessaire d'évaluer l'effet des changements de méthodes d'analyse, de méthodes d'échantillonnage ou de laboratoire sur les mesures de la qualité de l'eau d'après les données de surveillance recueillies pendant un certain temps, parce que l'évaluation de la compatibilité faite avant le changement s'est révélé inadéquate ou pour d'autres raisons. On envisage une telle évaluation pour la concentration de sodium et la concentration de nitrate et de nitrite mesurées à divers emplacements des rivières Bow et Saskatchewan-Sud entre 1977 et 1987. On a ajusté des modèles pour tenir compte de la variation saisonnière et annuelle en plus des changements méthodologiques. Un seul changement de laboratoire n'a pas eu d'influence sur les concentrations de sodium, mais les résultats ne sont pas concluants dans le cas concentrations de nitrates et de nitrites. Bien que la méthode décrite ici s'applique à défaut d'une meilleure solution, il ressort à l'examen des résultats qu'il faut faire une évaluation dans les règles de l'effet des changements méthodologiques avant de réaliser ceux-ci.

Par ailleurs, on a appris des choses utiles sur la modélisation de la variabilité inhérente des données de surveillance. La variation liée au régime hydrologique annuel est la plus grande source de variabilité dans les ensembles de données; pour faire une estimation précise des autres changements, il faut introduire des termes saisonniers dans les modèles. Par exemple, le taux de variation justifiée n'a pas dépassé 5 % lorsque le modèle ne comportait que des termes de méthodologie ou de laboratoire, tandis qu'on a obtenu des valeurs de 72 à 87 % pour le $\text{NO}_3 + \text{NO}_2$ et de 38 à 76 % pour le Na lorsqu'on a introduit en plus un terme de moyenne annuelle et un terme saisonnier. Avec un échantillonnage mensuel, il était difficile de caractériser l'aspect saisonnier lorsque celui-ci était plus complexe qu'un minimum estival. La variation saisonnière n'était pas simplement une différence de valeur d'une

saison à l'autre; la comparaison des moyennes saisonnières n'est donc pas toujours indiquée. C'est le cas par exemple des concentrations de $\text{NO}_3 + \text{NO}_2$, qui baissent rapidement au cours des quatre ou cinq premiers mois de l'année. Ce rapport doit être intégré au rapport du groupe de travail ministériel sur l'élaboration et l'évaluation des programmes d'acquisition de données sur la qualité de l'eau.

ABSTRACT

When data is collected over many years within a monitoring program, it is often impractical or impossible for analyses to be done by the same method or laboratory. Ideally, the compatibility of the methods or laboratories should be evaluated before the change is made. If this is not done, or sometimes even if it is done, it may be necessary to assess the effects of such changes from the monitoring data collected over time. The possibility of doing this is investigated using two water quality parameters, nitrate plus nitrite ($\text{NO}_3 + \text{NO}_2$) and sodium (Na), measured on samples collected approximately monthly within the period 1977 to 1987 at five locations on the Bow and South Saskatchewan Rivers. The statistical analysis consists of fitting a model with mean, seasonal, method and laboratory terms by regression methods. As would be expected at this sampling frequency and from a reasonable model for the year and seasonal components, the residuals did not exhibit positive serial correlation. Inconclusive results about the effects of method and laboratory changes were obtained for $\text{NO}_3 + \text{NO}_2$, the parameter with more changes. However, Na concentrations were not found to be affected by the single laboratory change.

Although not the primary objective, the analyses showed that at this sampling frequency it is difficult to characterize the seasonality in the data, unless it consists of a simple summer minimum. A higher percentage of the variation was accounted for in modelling the $\text{NO}_3 + \text{NO}_2$ concentrations (72 to 87 percent) than for the Na concentration (38 to 76 percent). The consensus over the five locations and eleven years

is that seasonality is important for both parameters and all locations. This does not consist only of changes in mean level between seasons. For example, the $\text{NO}_3 + \text{NO}_2$ concentration is characterized as a rapid decline from early in the year until a summer minimum is reached. These observations are of importance to questions of time trends and spatial differences.

RÉSUMÉ

Lorsqu'on recueille des données pendant plusieurs années dans le cadre d'un programme de surveillance, il est souvent difficile, sinon impossible, que les analyses soient toujours faites selon la même méthode ou dans le même laboratoire. Idéalement, il faudrait évaluer la compatibilité des méthodes ou des laboratoires avant de faire un changement. Si l'on néglige de le faire, et parfois, même dans le cas contraire, il peut être nécessaire d'évaluer les effets des changements d'après les données de surveillance recueillies. On examine cette possibilité en étudiant deux paramètres de la qualité de l'eau, les concentrations de nitrate et de nitrite ($\text{NO}_3 + \text{NO}_2$) et de sodium (Na), qui ont été mesurées dans des échantillons prélevés à peu près tous les mois de 1977 à 1987 dans cinq emplacements des rivières Bow et Saskatchewan-Sud. L'analyse statistique a consisté à ajuster un modèle avec des termes de moyenne, de variabilité saisonnière, de méthodes et de laboratoires par régression. Comme on pouvait s'y attendre à une telle fréquence d'échantillonnage et avec des termes annuels et saisonniers raisonnables, les variances résiduelles n'étaient pas en corrélation sériale positive. Dans le cas de la concentration de $\text{NO}_3 + \text{NO}_2$, paramètre pour lequel il y a eu le plus de changements, l'évaluation des effets du changement des méthodes et de laboratoire n'a pas donné de résultats concluants. Par contre, le seul changement de laboratoire qui a été fait dans le cas du Na n'a pas eu d'influence.

Même si ce n'était pas le principal objectif, les analyses ont montré qu'à cette fréquence d'échantillonnage, il est difficile de caractériser la variation saisonnière d'une donnée, à moins qu'il s'agisse d'un simple minimum estival. Avec le modèle des concentrations de $\text{NO}_3 + \text{NO}_2$, on a pu rendre compte de la variation dans une proportion plus élevée (de 72 à 87 %) qu'avec la concentration de Na (de 38 à 76 %). La concordance observée entre

les cinq emplacements et les onze années examinées est importante pour les deux paramètres ainsi que pour tous les emplacements. Il ne s'agit pas simplement de changements de la valeur moyenne d'une saison à l'autre. Par exemple, la concentration de $\text{NO}_3 + \text{NO}_2$ se caractérise par une baisse rapide commençant au début de l'année et se poursuivant jusqu'à ce qu'un minimum estival soit atteint. Ces observations ont une incidence en ce qui concerne les tendances en fonction du temps et les différences en fonction du lieu.

INTRODUCTION

Water quality sampling networks are providing long-term data sets, where samples have been collected at selected locations within a watershed at some specified frequency. It is expected that such a data set can be used to determine whether long-term trends in water quality exist and to estimate differences in water quality between locations, where, for example, sampling locations are upstream and downstream from a city or agricultural region. However, sometimes there are features of the data set that may limit its usefulness for such purposes. An example is the change in method or laboratory during the period of interest without first showing that the different methods or laboratories will produce compatible results. The latter question is addressed here, using data at the locations on the Bow and South Saskatchewan Rivers which were described in the preceding chapter. This was considered to be a necessary first step, even though the questions of primary interest concern the existence of trends and spatial differences, since estimates of the latter should be free of methodological biases.

Objectives

On the basis of a model, which includes the features of the data set, determine whether differences in the concentration of a water quality variable, attributable to analytical or laboratory changes, can be detected.

Description of the Data

A set of 16 water quality variables were available for analysis. To choose variables to use as examples, these 16 variables were examined for completeness and features related to the quality of the data, such as precision. Calculated variables were excluded due to the difficulty of method changes at different times for variables entering the calculation. Several variables, such as pH and chloride, showed little variation at the headwater location BA0011. Some variables had more changes in method or laboratory than others. The two variables chosen were nitrate plus nitrite ($\text{NO}_3 + \text{NO}_2$) and sodium (Na). The former is a nutrient, with considerable seasonal variability, and several method and laboratory changes. The latter is a parameter expected to reflect urban and agricultural inputs, and for which only a change in laboratory occurred.

A detailed description of the analysis is given for $\text{NO}_3 + \text{NO}_2$ concentrations at the BA0011 location, and the results of the analyses for $\text{NO}_3 + \text{NO}_2$ at the other four locations and for Na at all locations are given. The nitrate plus nitrite data for the headwater station of the Bow River (BA0011) was chosen as the example for several reasons. The existence of a pronounced seasonality for $\text{NO}_3 + \text{NO}_2$ concentrations plus changes in both method and laboratory were expected to lead to a model of sufficient complexity that similar steps applied to other parameters and stations would provide an adequate analysis. The regularity of the seasonal cycle of $\text{NO}_3 + \text{NO}_2$ over years for station BA0011, as judged by a plot such as Figure 1, suggested that there would be less unexplainable variability than at the other stations, and hence, a better chance of detecting any

effect due to changes in method or laboratory. Thus, the analysis for the station most likely to be used for drawing conclusions about effects of method and laboratory changes, would be given in detail.

The sampling program was begun at the BH0017 location in 1977 and, thus, to use the period in which all stations can be compared, the statistical analyses here used data within the interval 1977 to 1987. During these years the analytical method for $\text{NO}_3 + \text{NO}_2$ was the same (NAQUADAT code 07110, Water Quality Branch, 1984) except for alternate periods of laboratory filtration (07110L) and field filtration (07110D) indicated as 10L and 10D in Figure 1, respectively. The analyses were performed in the Calgary laboratory prior to December 1, 1983 and in the Saskatoon laboratory thereafter. The analytical method for Na was constant over the entire period (NAQUADAT code 11103, Water Quality Branch, 1984) with a change from the Calgary laboratory to the Burlington laboratory on October 24, 1983.

Results in the NAQUADAT data files which are the means of several replicate samples are indicated by a code, but the number of replicates is not given. These concentrations have been used as if they were single analyses. To be consistent, means of multiple samples taken on one day at location AK0001 have been calculated and used in the analyses.

MODEL WHICH FORMS THE BASIS OF THE ANALYSIS

Although the objective is to determine whether differences in concentration have occurred due to method or laboratory

changes, it is clear from Figure 1 that there is a seasonal cycle in the nitrogen concentrations. Further, differences in yearly mean level and seasonal cycle might be expected. Since the periods of stable methods and laboratory combinations are at least two years, yearly effects as well as those due to method and laboratory can be included in the model.

A model taking all these effects into account is

$$Y_{ijks} = \mu + \alpha_i + m_j + l_k + ml_{jk} + (\beta_{11} \cos \omega t_{is} + \beta_{12} \sin \omega t_{is}) \\ + (\gamma_{j1} \cos \omega t_{is} + \gamma_{j2} \sin \omega t_{is}) + \epsilon_{ijks} \quad (1)$$

where Y_{ijks} = nitrate + nitrite concentration on day s of year i where the analysis was performed by laboratory k using method j

μ = general mean level

α_i = shift in the mean due to year i

m_j = shift in the mean due to method j

l_k = shift in the mean due to laboratory k

ml_{jk} = shift in the mean due to interaction between method j and laboratory k

t_{is} = day s of year i where t goes from 1 to 365, or 366 in a leap year

ω = frequency of the seasonal cycle

ϵ_{ijks} = error term

The sum $(\beta_{11} \cos \omega t_{is} + \beta_{12} \sin \omega t_{is})$ gives the seasonal cycle due to year i and the sum $(\gamma_{j1} \cos \omega t_{is} + \gamma_{j2} \sin \omega t_{is})$ gives the adjustment of the seasonal cycle due to method j . Sums such as $(\beta_1 \cos \omega t + \beta_2 \sin \omega t)$ can be shown to be equivalent to $R \cos (\omega t + \phi)$ where the amplitude R , and phase, ϕ , are given by

$$R = (\beta_1^2 + \beta_2^2)^{1/2} \quad \text{and} \quad \phi = \arctan - \beta_2/\beta_1.$$

The form involving β_1 and β_2 is used since it is linear in the parameters (El-Shaarawi et al., 1983).

STATISTICAL METHODS

A model such as (1) above can be fitted by multiple regression methods. The terms α_i , m_j , l_k and ml_{ijk} are handled by the method of dummy variables (Draper and Smith, 1981). Consider the general form of the regression model in terms of matrices,

$$\underline{Y} = X \underline{\beta} + \underline{\epsilon} \quad (2)$$

and the specific situation of our case study (Figure 1). To put model (1) into the form of model (2), the elements of $\underline{\beta}$ were chosen as follows

- μ = general mean
- $\alpha_1, \alpha_2, \dots, \alpha_{10}$ = effect due to ten of the eleven years relative to some standard
- m = difference between the two methods
- l = difference between the two laboratories
- ml = modification to the difference between methods when the second laboratory performed the analyses
- $(\beta_{11}, \beta_{12}), (\beta_{21}, \beta_{22}), \dots, (\beta_{11,1}, \beta_{11,2})$ = coefficients of the yearly seasonal component

(γ_1, γ_2) = coefficients of the adjustment to the seasonal component when the second method was used.

The first column of the matrix X is a column of ones. Different parameterizations of the dummy variables corresponding to the next thirteen elements of β are possible. The simplest, consisting of columns of zeros and ones, was used here. Since the columns of X cannot be dependent, there is one parameter fewer than the number of years, methods and laboratories. Thus, for example, if the yearly means are determined as $\mu_1 = \mu$ and $\mu_i = \mu + \alpha_i$ for $i = 2, \dots, 11$, as was done here, columns 2 to 11 of X are

$x_{rc} = \begin{cases} 1 & \text{if the sample corresponding} \\ & \text{to row } r \text{ was taken in year } c \\ 0 & \text{otherwise} \end{cases}$

where $c = 2, 3, \dots, 11$. Columns 12, 13 and 14 also consist of zeros and ones where

$x_{r,12} = \begin{cases} 1 & \text{if row } r \text{ corresponds to a} \\ & \text{sample analyzed by method 2} \\ 0 & \text{otherwise} \end{cases}$

$x_{r,13} = \begin{cases} 1 & \text{if row } r \text{ corresponds to a result} \\ & \text{produced by laboratory 2} \\ 0 & \text{otherwise} \end{cases}$

$x_{r,14} = \begin{cases} 1 & \text{if the sample of row } r \text{ was analyzed} \\ & \text{by method 2 in laboratory 2} \\ 0 & \text{otherwise} \end{cases}$

The columns of X corresponding to the seasonal terms are columns 15 to 36. For example, the first pair of columns in this block of columns is given by

$$x_{r,15} = \cos \omega t_r$$

$$x_{r,16} = \sin \omega t_r$$

if row r corresponds to a sample collected in year 1 and t_r is the day of sample collection, for t defined as above. Otherwise, $x_{r,15} = x_{r,16} = 0$. The last two columns of X are given by

$$x_{r,37} = \cos \omega t_r$$

$$x_{r,38} = \sin \omega t_r$$

if row r corresponds to a sample analyzed by method 2 and $x_{r,37} = x_{r,38} = 0$, otherwise.

Due to the non-orthogonality of the data, the importance of the sets of components is determined by fitting a model with all the parameters, a model with a reduced set of parameters and then comparing the mean square obtained from the difference between the residual sums of squares of the two models with the estimate of σ^2 , given by the residual mean square of the model with more parameters. Clearly an automatic selection procedure cannot be used here because both terms in a seasonal component must be left in the model. Similarly, if an interaction term is included in the model, the individual terms,

which are components of the interaction, should appear in the model.

Diagnostic procedures to assess whether the assumptions underlying the analysis are met were also performed. This includes examination of the residuals for normality, equality of variance, serial correlation, and adequacy of the form of the model (Draper and Smith, 1981). The analyses were performed using programs based on IMSL subroutines (IMSL, 1987) and plotting programs written previously, for similar applications.

ANALYSIS OF $\text{NO}_3 + \text{NO}_2$ CONCENTRATIONS AT BAO011

The number of samples for which $\text{NO}_3 + \text{NO}_2$ concentrations are available, are shown (Table 1) for each year and for the successive periods of consistent method and laboratory. Figure 1 and Table 1 are part of the preliminary examination of the data. The figure led to the postulation of the terms in model (1), including the two terms with coefficients (γ_{j1} , γ_{j2}) which represent the apparent larger amplitude of the seasonal cycle under method 1 (10L). The table shows the number of samples available for estimating the parameters and the unequal numbers of samples in the periods of consistent method and laboratory combinations. Figure 1 also indicates that the seasonality can be represented by a single cosine function and thus the frequency was taken as $\omega = 0.0172$.

Steps in the Regression Analysis

The objective of this analysis was to look for evidence of changes in concentrations associated with method or laboratory changes, having accounted for yearly differences in mean level and seasonal cycle. In the regression analysis, this is done by fitting a model with only yearly means and yearly seasonal terms, and another with method or laboratory terms as well as the yearly means and yearly seasonal terms. The mean square, corresponding to the reduction in the residual sum of squares obtained by adding method or laboratory terms, is then compared to a residual mean square which provides a good estimate of σ^2 , the variance of ϵ_{ijk} in model (1). The analysis of variance table for the complete model is shown in the top of Table 2, and the sequential tests in the lower part of the table. The sequential test of any set of terms occurring as a row in the lower part of the table gives a test of significance for inclusion in the model of the terms in that particular row, given that all terms in the rows above have also been included in the model.

The complete model explains 86.8 percent of the variation and the regression is highly significant (Table 2). The fitted values are plotted with the data in Figure 2. In this plot and all of the other plots of fitted values, the curves are not smooth because fitted values were calculated only at times when samples were taken. Also, discontinuities between years are present in plots of models with yearly mean and seasonal terms. It is clear that both yearly mean and seasonal terms should be included in the model since the F value is 22.2 for seasonal terms, alone in the model, and 3.7 for yearly means,

with yearly season also in the model. However, method and laboratory explain a much smaller proportion of the variance, as does the method by season interaction. The order of entry of terms in the model may affect the conclusions drawn from sequential tests, and hence, for our stated purpose, the order given here should be used. For example, if yearly means are followed by the method and laboratory terms, then method and method by laboratory interaction are significant at a level of 0.003 or less because these terms remove some of the residual variation, which is considerable since the seasonal cycle terms are not in the model.

To see the importance of the different types of terms, the residual mean squares and the C_p statistics (Draper and Smith, 1981) for each set of terms and some selected combinations are given in Table 3. The residual mean square for a total of p parameters is obtained with only the indicated parameters in the model. The expected value of C_p is approximately equal to p and thus values of C_p near p are taken to indicate models which do not suffer from lack of fit. The last model in the table has the lowest residual mean square and this was used as s^2 in the calculation of C_p . Comparing the values of the mean square and C_p over all rows shows that both the model with only yearly mean and seasonal terms, and the model with these terms plus method and laboratory effects are nearly as good as the last model. The high C_p values for the models without seasonal terms indicate a large degree of lack of fit and the seasonal terms are seen to explain much more variability than any of the others.

The conclusion is that changes in the mean level or the seasonal cycle associated with method and laboratory changes, were not detected. The different results obtained by alternate ordering of the terms entering the regression show the difficulty of trying to answer the question of the effect of methodological changes from data such as this, with so many factors confounded. At best the conclusions are weak.

Checking the Assumptions of the Analysis

The examination of the residuals, $\hat{\epsilon}_{ijks}$, given by the difference between the observations Y_{ijks} , and the fitted values, \hat{Y}_{ijks} , is the primary method for checking the validity of the assumptions. The small differences in the residual mean squares for models with seasonal terms (Table 3) suggest that the residual mean square from model (1) will provide a reliable unbiased estimate of σ^2 , the variance of ϵ_{ijks} , if no important terms have been omitted from the model. The plot of $\hat{\epsilon}_{ijks}$, obtained from model (1), versus day within year over all years (Figure 3) gives no indication of omitted terms. This is clearer if contrasted with the same plot of the residuals from the model without year terms (Figure 4), in which the residuals are predominantly positive for some years and negative for others.

The assumptions of normality and constancy of variance (homoscedasticity) for ϵ_{ijks} also appear reasonable. Figure 3 has the form of an approximately horizontal band, consistent with homoscedasticity, except that there are a few more residuals of larger magnitude in the years 1977 to 1979, the period

in which the Calgary laboratory used method 10L. The assumption of normality was checked by means of a probability plot called a Q-Q plot. If the ϵ_{ijks} are normally distributed, the points should lie approximately on a straight line. This is seen to be the case for the residuals of model (1) shown in Figure 5, except for several residuals of larger magnitude at each end, which tend to shift the slope of this general linear trend away from the 45 degree line. However, minor variations in the extremes of a few points are common even in samples from a truly normal distribution.

It is the general picture of the residual plot that is to be assessed with respect to the assumptions of constancy of variance and normality, and on this basis the conclusion is that these assumptions are met. It is often suggested that a logarithmic transformation should be applied if the assumption of normality is to be appropriate. However, the assumption ^{is that} ~~of~~ ^{being to a more normal distribution,} ~~normality here is of~~ the error terms, ϵ_{ijks} , not the original concentration. In fact, after transforming to logarithms and using $\log Y_{ijks}$ in the analysis, the Q-Q plot showed marked departure from a straight line (Figure 6) and, although the residuals of large magnitude in the years 1977 to 1979 were reduced, non constancy of variance was induced due to smaller variance in 1984 and 1985 (Figure 7).

The assumption of independence was checked by means of the runs test (Draper and Smith, 1981) applied to the residuals. A run consists of residuals of like sign which are contiguous in time. Too few runs indicate positive serial correlation, the type of serial correlation to be expected for the present type of data if measurements are close in time. There

was no evidence of positive serial correlation in the residuals from models fitted to this data, even when the residual plots indicated additional terms were needed, for example, yearly means.

Estimates of Seasonal Terms, Means and \hat{Y}_{ijks}

To show how the various terms in model (1) contribute to the fitted value, \hat{Y}_{ijks} , the terms affecting mean level (Table 4) and those contributing to the seasonal cycle (Table 5) in terms of amplitude, R , and phase shift, ϕ , are given. As an example, the seasonal components for 1986 are plotted in Figure 8a. Finally, \hat{Y}_{ijks} is plotted in Figure 8b together with the final mean for 1986 (last column of Table 4) and the data. Note that the estimate of the day at which the $\text{NO}_3 + \text{NO}_2$ concentrations peak can be obtained from the amplitude and phase form of the seasonal component and this is given for each year in Table 5.

Method Differences within the Saskatoon Laboratory

The analysis was also performed for the period 1984 to 1987, during which both methods, 10L and 10D, were used in the Saskatoon laboratory. There is no evidence of differences in concentration due to method for this subset of the data (Table 6). Further, the same seasonal cycle applies to all four years, although the mean levels differ from year to year.

ANALYSIS OF NO_3 + NO_2 CONCENTRATIONS AT THE OTHER STATIONS

The NO_3 + NO_2 concentrations plotted against time are given for the four other stations (Figure 9). Comparison with the corresponding plot for BA0011 (Figure 1), shows that the predominant seasonal cycle is present, but that there are considerable differences between stations. Even though the first three stations are upstream from major urban centres and the irrigation districts, and have nitrate plus nitrite concentrations in the same range, there are marked differences. The larger amplitude of the seasonal cycle in years during which the Calgary laboratory used method 07110L relative to the years when the laboratory used method 07110D on samples from BA0011 is not present for BE0013 or BH0017. There appears to be a second peak of lower magnitude at location BH0017. The fairly consistent spike in the trough of the seasonal cycle for BN0001 and AK0001 suggests the same thing, but there are at most two observations defining these spikes for the latter two locations. The range of concentrations is much larger for the two lower locations, which also exhibit periods of constant concentration during minima at a concentration of 0.01 mg/L, higher than the reported detection limit of 0.005 mg/L (Water Quality Branch, 1979). As will be seen below, these features affect the analysis and results.

Results

The single consistent conclusion for all the stations is the existence of seasonality in the NO_3 + NO_2 concentrations. The terms which correspond to a single seasonal cycle

for all years account for the most variability at all locations, as can be seen from the mean squares in Table 7. The additional variability accounted for by fitting a seasonal cycle for each year differs between the stations. Allowing for a second peak in the seasonal cycle at BH0017 accounts for half as much variability as the single peak.

The conclusions about the effects of method and laboratory changes on the $\text{NO}_3 + \text{NO}_2$ concentrations vary between locations. The apparent difference between methods at location BE0013, when data from both laboratories are used, is not present when only data from the Saskatoon laboratory is used in the analysis (Tables 7 and 8). There is no evidence of method or laboratory effects at BH0017 and the interpretation for BN0001 is unclear. The method term is significant when the non-significant yearly mean and seasonal terms are in the model (Table 7), but the significance probability rises to 0.28 when all of the terms with $p > 0.14$ are excluded. Both method and laboratory terms are significant ($p = 0.003$ and 0.02 , respectively) for location AK0001 and these p values are maintained even when interaction terms are removed. The most satisfactory models, given in Table 8, are shown in Figure 10 and as the bottom set of plots in Figure 11.

As with BA0001, no strong conclusions can be drawn about the effects of method and laboratory changes. First, there is an inconsistency about the conclusions from different locations. Second, the effects of method and laboratory changes cannot be separated from the yearly mean and seasonal terms at two of the locations. The significance levels for method at BE0013 and laboratory at BN0001 depend upon the other terms in

the model, namely how the seasonal cycle is modelled. For example, the significance level of 0.04, for method when a seasonal cycle with a single peak is fitted to BE0013 data, rises to 0.62 when a two-peak seasonal cycle is fitted. However, the regression diagnostics suggest overfitting. The runs test indicates too many runs and the probability plots show that some of the data is fitted much better than the remainder by this model. The difficulty of trying to fit a seasonal model with only twelve observations per year is also shown for location BH0017 in Figure 11.

ANALYSIS OF SODIUM CONCENTRATIONS

From the plots of concentration versus time for $\text{NO}_3 + \text{NO}_2$ (Figures 1 and 9), the existence of a seasonal cycle was evident for all five stations. The seasonal pattern is much less clear for Na (Figure 12), except for a reasonably consistent summer minimum. A more consistent seasonal cycle is seen for BA0011 during the years in which the analyses were done in the Burlington laboratory than during the period in which the analyses were done in the Calgary laboratory. Locations BN0001 and AK0001 show more variability and higher concentrations than the other three locations.

For location BN0001, and to some extent BH0017, from the examination of the plot of the concentration of Na alone (Figure 12), it is difficult to see whether a summer minimum is present. The comparison of the Na and discharge plots is useful here. For locations BA0011, BE0013, and AK0001, summer Na minima are quite consistent. In most instances, these minima

correspond to summer discharge maxima, although sometimes there appears to be a lag for Na. Thus, a correspondence between a Na minimum and a discharge maximum would be expected at the other stations. Once the plots for Na and discharge at BN0001 are superimposed (Figure 13), it can be seen that the Na minimum, which is narrow and poorly defined at the monthly sampling frequency, usually corresponds to the flow maximum. From this, it seems reasonable to include seasonal terms in the models for all locations.

Results

Seasonal terms are significant for all locations but the percentage of the variation explained by the regression is generally lower than that of $\text{NO}_3 + \text{NO}_2$. Two peaks are present in the seasonal cycle of BH0017, BN0001 and AK0001. Differences in the Na concentration due to the laboratory change were not detected, with the lowest significance level being 0.23 at location AK0001 (Table 9). The poorly defined seasonal cycle at BA0001, during analyses at the Calgary laboratory, results in a lower $100 R^2$ (41.7) when all the years are included (Table 10), compared with 59.2 when only 1983 to 1987 data are used. Also the estimated residual variance for 1977 to 1982 is two times the estimate (0.01606) for 1983 to 1987. Thus there is higher variability in Na concentrations at BA0011 while analyses were performed in the Calgary laboratory.

Isolated high concentrations had more effect on the fit of the seasonal cycle of Na than on $\text{NO}_3 + \text{NO}_2$, and omission of points (one for each of locations BA0011 and BE0013 and

three for BH0017) resulted in improvements in $100 R^2$ of up to 20 percent. The objective is to test for the effect of the laboratory change, having included yearly mean and seasonal terms. Points which prevent a fit of the seasonal cycle can result in unexpected results because all the observations in that year will contribute large terms to the residual variation.

The most satisfactory models, summarized in Table 10, are plotted in Figure 14. The fitted model for BE0013, although leaving considerable residual variability, shows a single summer minimum. The residual diagnostics are also satisfactory. The other three locations show much more variability not associated with a single seasonal cycle. A model with two peaks in the seasonal cycle explains 76.4 percent of the variation at BH0017 but the residuals exhibit features associated with overfitting. A good fit was not obtained for BN0001 even with a two-peak seasonal cycle. The more regular seasonality resulted in $100 R^2$ equal to 71 percent at AK0001, with the probability plot indicating only variations in the extremes. However, there are again indications of overfitting.

Despite the above comments, the conclusions regarding no detectable differences in Na concentrations due to changes in laboratory are tenable. The many analyses which were done to arrive at Tables 9 and 10 and Figure 14, show that the significance level for the laboratory effect is always high, and generally, those related to seasonality are consistent. These models included single seasonal cycles accounting for a smaller percentage of variation but satisfactory residual diagnostics and models with many parameters which explained more variation but sometimes indicated departures from assumptions of the analysis.

DISCUSSION

The above analyses illustrate the difficulty of trying to draw conclusions about method and laboratory changes from complex data sets not designed for that purpose and emphasize the need for proper evaluation before such changes are made in monitoring networks. For this specific set of data, conclusions can be drawn about how to use the $\text{NO}_3 + \text{NO}_2$ and Na data to estimate differences between locations. For $\text{NO}_3 + \text{NO}_2$, a conservative position is to compare stations within periods in which only one laboratory and method combination occurs. For Na, apart from the modest differences in variability between laboratories for location BA0011, there is no indication of laboratory effects and it is reasonable to use the data in the period 1977 to 1987.

Although not the primary purpose of the analyses, useful results about the ability to characterize the seasonality of the two parameters were obtained. The seasonality of $\text{NO}_3 + \text{NO}_2$ is much more pronounced than that of Na and thus more satisfactorily modelled. However, from the consensus of the five locations and the dilution effect on Na during the summer discharge peak, it is reasonable to assume seasonality in the sodium concentrations as well. The usual reasons for including seasonal terms in a model is to remove this variability from the error and thus permit more precise estimation of trend or spatial differences or to remove bias from estimates. The above analyses indicate that monthly samples are not adequate, at least for stations and parameters with a cycle more complex than a summer minimum. A further important point is the form of the seasonal cycle, since means within season are often used. Consider

$\text{NO}_3 + \text{NO}_2$ at location AK0001, for example. The period before the summer minimum, in most years, consists of a rapid decline in concentrations. To report a mean of these values gives no indication of what is happening.

To show the importance of fitting yearly means and seasonal terms, models without these terms were fitted for $\text{NO}_3 + \text{NO}_2$ and Na concentrations (Table 11). The highest 100 R^2 was 5.3 and the test for runs always indicated too few runs. The probability plots were highly irregular and non-normal.

ACKNOWLEDGEMENTS

The authors would like to thank L.C. Keeler who performed many of the extensive computations required to complete this report.

REFERENCES

- Draper, N.R. and Smith, H. 1981. Applied Regression Analysis. Second Edition. Wiley, New York, 709 pp.
- 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 Res. 9(2): 234-240.

TABLE 1: The number of samples analyzed for $\text{NO}_3 + \text{NO}_2$ by year and method and laboratory combination for station BA0011.

Year i	n_i	Location of Filtration	Laboratory	$n_{m(l)}$	$n(l)$
1977	4	L	C	28	72
1978	10	L	C		
1979	11	L	C		
1980	3 } 12	L	C		
	9 }	D	C	33	
1981	12	D	C		
1982	12	D	C		
1983	11 } 12	L	C	32	
	1 }	L	S		48
1984	11	L	S		
1985	9 } 12	L	S		
	3 }	D	S	27	
1986	12	D	S		
1987	12	D	S		
Totals	120			120	120

n_i is the number of samples in year i, with the number by either different methods or laboratories indicated.
 $n_{m(l)}$ is the number of samples analyzed by method m in laboratory l and n_l is the number of samples analyzed by laboratory l.

L,D indicate filtration in the laboratory and field, respectively.

C,S indicate Calgary and Saskatoon, respectively.

TABLE 2: Analysis of variance tables for NO₃ + NO₂ concentrations at station BA0011 1977 to 1987.

Source	Degrees of Freedom	Sum of Squares	Mean Square	F	p	100 R ²
Regression	37	0.04546	0.001229	14.5	<0.0001	86.8
Residual	82	0.00693	0.000085			
Total, corrected	119	0.05239				

Term(s) Entering the Model	Degrees of Freedom	Sum of Squares	Mean Square	F*	p
Yearly season	22	0.04153	0.00189	22.2	0.0001
Yearly mean	10	0.00315	0.00032	3.7	0.0004
Method, laboratory and interaction	3	0.00036	0.00012	1.4	0.25
Method, season interaction	2	0.00042	0.00021	2.5	0.09

F is the ratio of a regression to a residual mean square and it follows an F-distribution. p is the probability of the value of an F variate exceeding the value of F in the fifth column.

* The F ratio is formed from the mean square in column four of this part of the table and the residual mean square from the complete model, equal to 0.000085.

TABLE 3: Measures of a satisfactory fit for selected sets of parameters in the model for $\text{NO}_3 + \text{NO}_2$, BA0011.

Number of Sets	Sets of Parameters	Residual Mean Square	p	C_p
1	$\{\mu_i\}$	0.00046	11	499
	s	0.00013	3	61
	$\{s_i\}$	0.00011	23	55
	(m, l, mxl)	0.00045	4	502
2	$\{\mu_i\}, \{s_i\}$	0.00009	33	37
	$\{\mu_i\}, (m, l, mxl)$	0.00045	13	472
	s, (m, l, mxl)	0.00013	6	61
3	$\{\mu_i\}, s, (m, l, mxl)$	0.00011	16	48
	$\{\mu_i\}, \{s_i\}, (m, l, mxl)$	0.00009	36	39
	s, (m, l, mxl), mxs	0.00012	8	57
4	$\{\mu_i\}, s, (m, l, mxl), (mxs)$	0.00010	18	41
	$\{\mu_i\}, \{s_i\}, (m, l, mxl), (mxs)$	$s^2 = 0.00008$	38	38*

p is the total number of parameters in the model, including a general mean.

$C_p = (RSS_p/s^2) - (n - 2p)$ where RSS_p is the residual sum of squares from a model with p parameters, s^2 is given in column 3 and $n = 120$.

* $C_p = 38$ because this is the model used to obtain s^2 . Thus C_p cannot be used to judge lack of fit for this model.

$\{\mu_i\}$ = set of yearly means.
s = the same seasonal cycle fit to all years.
 $\{s_i\}$ = different seasonal cycle for each year.
(m, l, mxl) = method and laboratory terms.
(mxs) = adjustment to seasonal component due to method.

TABLE 4: Contribution to the final mean level obtained from fitting model (1) to $\text{NO}_3 + \text{NO}_2$ concentration for BAO011.

Year	Mean for the Year	Adjustment due to			Estimate for the Year with Adjustments
		Method 2	Laboratory 2	Method 2 and Laboratory 2	
1977	0.083	0	0	0	0.083
1978	0.066	0	0	0	0.066
1979	0.066	0	0	0	0.066
1980*	0.071	-0.012	0	0	0.059
1981	0.070	-0.012	0	0	0.057
1982	0.079	-0.012	0	0	0.067
1983*	0.058	0	-0.010	0	0.058
1984	0.069	0	-0.010	0	0.059
1985*	0.074	0	-0.010	0	0.065
1986	0.032	-0.012	-0.010	0.048	0.058
1987	0.033	-0.012	-0.010	0.048	0.059

* In each of these years there were between one and three samples which were analyzed by a different method-laboratory combination than the rest of the samples in that year. The estimate with adjustments has not been shown for these few samples.

TABLE 5: Estimates of the seasonal components obtained from fitting model (1) to the $\text{NO}_3 + \text{NO}_2$ concentration for BA0011.

Year	Component for Year		Adjustment for Method		Component with Adjustment		t at Max
	Amplitude	Phase	Amplitude	Phase	Amplitude	Phase	
1977	0.051	-0.33	0	0	0.051	-0.33	19
1978	0.037	-0.68	0	0	0.037	-0.68	40
1979	0.028	-0.76	0	0	0.028	-0.76	44
1980	0.041	0.20	0.032	0.934	0.027	-0.69	40
1981	0.034	0.53	0.032	0.934	0.013	-0.66	38
1982	0.034	0.16	0.032	0.934	0.025	-0.92	53
1983	0.024	-0.74	0	0	0.024	-0.74	43
1984	0.020	-0.48	0	0	0.020	-0.48	28
1985	0.026	-0.72	0	0	0.026	-0.72	42
1986	0.039	0.23	0.032	0.934	0.025	-0.72	42
1987	0.034	0.34	0.032	0.934	0.019	-0.83	48

The amplitude is in the units of $\text{NO}_3 + \text{NO}_2$ concentration, mg/L, and the phase angle is in radians.

t at Max gives the estimate of the day in the year, with January 1 being day 1, at which the maximum of the fitted seasonal cycle occurs.

TABLE 6: Analysis of variance tables for NO₃ + NO₂ concentrations at BA0011 determined in the Saskatoon laboratory between 1984 and 1987.

Source	Degrees of Freedom	Sum of Squares	Mean Square	F	p	100 R ²
Regression	14	0.01305	0.00093	18.2	0.0001	88.9
Residual	32	0.00164	0.00005			
Total, corrected	46	0.01468				

Term(s) Entering the Model	Degrees of Freedom	Sum of Squares	Mean Square	F*	p
Yearly mean	3	0.00061	0.00020	4.0	0.02
Season	2	0.01193	0.00597	117	< 0.0001
Yearly season, increment ^a	6	0.00040	0.00007	1.3	0.29
Method	1	0.00007	0.00007	1.3	0.26
Method, season interaction	2	0.00005	0.00003	0.4	0.65

F is the ratio of a regression to a residual mean square and it follows an F-distribution. p is the probability of the value of an F variate exceeding the value of F in the fifth column.

* The F ratio is formed from the mean square in column four of this part of the table and the residual mean square from the complete model, equal to 0.00005.

^a This gives the additional variability due to fitting a seasonal cycle to each year relative to one cycle for all years.

TABLE 7: Sequential statistics for models of NO₃ + NO₂ concentrations at locations BE0013, BH0017, BN0001 and AK0001.

Location, Years, n	s ²	Total Mean Square	Residual ^a Mean Square	100 R ²	Sequential Entry of Terms to Model					
					Terms	Degrees of Freedom	Sum of Squares	Mean Square	F	p
BE0013										
1978-1987	0.00103	0.00037	74.6	{ μ_1 }	9	0.01001	0.00111	2.97	0.004	
n=115				s	2	0.06322	0.03161	84.4	<0.0001	
				{ Δs_1 }	18	0.01113	0.00062	1.65	0.07	
				m	1	0.00164	0.00164	4.39	0.04	
				ℓ	1	0.00003	0.00003	0.09	0.77	
				mx ℓ	1	0.00000	0.00000	0.01	0.91	
				mxs	2	0.00188	0.00094	2.52	0.09	
1983-1987										
n=59	0.00109	0.00037	74.8	{ μ_1 }	4	0.00541	0.00135	3.67	0.01	
				s	2	0.03354	0.01677	45.5	<0.0001	
				{ Δs_1 }	8	0.00776	0.00097	2.63	0.02	
				m	1	0.00039	0.00039	1.05	0.31	
BH0017										
1978-1986	0.00090	0.00034	79.4	{ μ_1 }	8	0.00620	0.00078	2.29	0.03	
n=106 ^b				s ₁	2	0.03321	0.01661	49.2	<0.0001	
				{ Δs_{1f} }	16	0.01098	0.00069	2.03	0.03	
				s ₂	2	0.01850	0.00925	27.4	<0.0001	
				{ Δs_{2f} }	16	0.00553	0.00035	1.02	0.45	
				m	1	0.00008	0.00008	0.24	0.63	
				ℓ	1	0.00030	0.00030	0.88	0.35	
				mx ℓ	1	0.00059	0.00059	1.73	0.19	

cont'd....

TABLE 7 (cont'd): Sequential statistics for models of NO₃ + NO₂ concentrations at locations BE0013, BH0017, BN0001 and AK0001.

Location, Years, n	s ² Total Mean Square	Residual ^a Mean Square	100 R ²	Sequential Entry of Terms to Model					
				Terms	Degrees of Freedom	Sum of Squares	Mean Square	F	p
BN0001 1977-1986 n=113	0.29978	0.08606	80.0	{ μ_1 }	9	0.56052	0.06228	0.72	0.69
				s	2	24.01508	12.00754	139.	< 0.0001
				{ Δs_1 }	18	1.60494	0.08916	1.04	0.43
				m	1	0.00341	0.00341	0.04	0.84
				l	1	0.33442	0.33442	3.89	0.05
				mxl	1	0.00246	0.00246	0.03	0.87
				mxs	2	0.34150	0.17075	1.98	0.14
AK0001 1977-1987 n=140	0.23579	0.04863	84.9	{ μ_1 }	10	1.79408	0.17941	3.59	0.0003
				s	2	23.45672	11.72836	241.	< 0.0001
				{ Δs_1 }	20	1.85617	0.09281	1.91	0.02
				m	1	0.25135	0.25135	5.17	0.03
				l	1	0.25458	0.25458	5.23	0.02
				mxl	1	0.05948	0.05948	1.22	0.27
				mxs	2	0.14169	0.07084	1.46	0.24

a This is the mean square used in the denominator of the F ratio in column nine. For one location, the test for the set of terms in any row is for inclusion of these terms given that terms in all the rows above are in the model.

b The sample of July 12, 1978 was excluded because, in an initial analysis, it had a standardized residual of 4.81. This sample was the cause of a much lower 100R² (43.2) because the fitted seasonal cycle was distorted.

The terms in column 5 are defined in Table 3, with the additional terms { Δs_1 }, { Δs_{11} }, { Δs_{21} } indicating the increment above a single seasonal cycle when a seasonal cycle is fitted for each year.

TABLE 8: The final models for NO_3 + NO_2 concentrations at BE0013, BH0017, BN0001 and AK0001.

Location	Years	100 R^2	Residual Mean Square	Terms in the Model	Runs Test p
BE0013	1978-1987	73.9	0.00037	$\{\mu_1\}, \{s_1\}, m$	0.28 tm
	1983-1987	74.8	0.00037	$\{\mu_1\}, \{s_1\}$	0.14 tm
BH0017 ^a	1978-1986	72.8	0.00034	$\{\mu_1\}, \{s_{11}\}, s_2$	0.12 tm
BN0001	1977-1986	71.9	0.087	s, l	0.25 ^{tf}
AK0001	1977-1987	84.3	0.04916	$\{\mu_1\}, \{s_1\}, m, l$	0.32 ^{tf}

The regression was significant at a level of < 0.0001 for all of the above models.

tm, tf means that the test was for too many runs and too few runs, respectively.

^a Sample of July 12, 1978 was excluded.

TABLE 9: Sequential statistics for models of Na concentrations at the five locations.

Location, Years, n	s^2		100 R^2	Sequential Entry of Terms to Model						
	Total Mean Square	Residual Mean Square		Terms	Degrees of Freedom	Sum of Squares	Mean Square	F	p	
BA0011 1977-1987 n=119 ^a	0.03824	0.02756	48.1	$\{\mu_1\}$	10	0.2424	0.02424	0.88	0.56	
				s	2	0.9354	0.4677	17.0	<0.0001	
				$\{\Delta s_1\}$	20	0.9849	0.0492	1.79	0.04	
				ℓ	1	0.0073	0.0073	0.27	0.61	
BE0013 1978-1987 n=113 ^b	0.1596	0.0772	64.6	$\{\mu_1\}$	9	1.618	0.1798	2.33	0.02	
				s	2	8.603	4.3015	55.7	<0.0001	
				$\{\Delta s_1\}$	18	1.316	0.0731	0.95	0.53	
				ℓ	1	0.002	0.002	0.02	0.88	
BH0017 1978-1986 n=104 ^c	0.3817	0.1601	76.4	$\{\mu_1\}$	8	4.630	0.5788	3.62	0.002	
				s_1	2	1.481	0.7405	4.62	0.01	
				$\{\Delta s_1\}$	16	7.134	0.4459	2.79	0.002	
				s_2	2	6.796	3.3980	21.2	<0.0001	
				$\{\Delta s_2\}$	16	9.981	0.6238	3.90	0.0001	
				ℓ	1	0.002	0.0020	0.01	0.91	

cont'd.../

TABLE 9 (cont'd): Sequential statistics for models of Na concentrations at the five locations.

Location, Years, n	s^2 Total Mean Square	Residual Mean Square	100 R^2	Sequential Entry of Terms to Model				
				Terms	Degrees of Freedom	Sum of Squares	Mean Square	P
BN0001 1978-1986 n=107	14.66	9.99	60.8	$\{\mu_1\}$	8	146.1	18.26	1.83
				s_1	2	112.0	56.00	5.61
				$\{\Delta s_{1f}\}$	16	158.0	9.88	0.99
				s_2	2	329.0	164.5	16.5
				$\{\Delta s_{2f}\}$	16	190.2	11.89	1.19
				ℓ	1	8.9	8.90	0.89
								0.35
AK0001 1978-1987 n=153	35.59	16.81	75.6	$\{\mu_1\}$	9	450.5	50.05	2.97
				s_1	2	594.0	297.0	17.7
				$\{\Delta s_{1f}\}$	18	565.3	31.41	1.87
				s_2	2	843.1	421.5	25.1
				$\{\Delta s_{2f}\}$	18	729.2	40.51	2.41
				ℓ	1	24.7	24.70	1.47
								0.23

a Sample of November 3, 1981 was excluded because on an initial run it had a standardized residual, e_1/s , of 5.44 and its removal resulted in an increase in $100 R^2$ of 18.

b Sample of August 20, 1980 was excluded because \hat{e}_1/s was 3.82 and its removal resulted in an increase in $100 R^2$ of 10.

c Samples of May 18, 1978, May 9, 1978 and May 21, 1986 were excluded since a much better fit of the seasonal cycles was obtained. The significance level for laboratory effect was unaffected by their omission.

See Table 7 for explanation of terms.

TABLE 10: The final models for Na concentrations at the five locations.

Location	Years	100 R ²	Residual Mean Square	Terms in the Model	Test Runs p
BA0011 ^a	1977-1987	41.7	0.02738	{s ₁ }	0.46 ^{tf}
	1983-1987	59.2	0.01606	{s ₁ }	0.47 tm
BE0013 ^b	1978-1987	57.2	0.0757	{u ₁ }, s	0.13 ^{tf}
BH0017 ^c	1978-1986	76.4	0.1574	{u ₁ }, {s ₁₁ }, {s ₂₁ }	0.13 ^{tf}
BN0001	1978-1986	37.8	10.28	{u ₁ }, s ₁ , s ₂	0.07 tm
AK0001	1977-1987	71.0	16.91	{u ₁ }, {s ₁₁ }, {s ₂₁ }	0.07 tm

a, b, c See Table 9 for explanation.

tm, tf means that the test was for too many runs and too few runs, respectively.

TABLE 11: Models with method or method and laboratory terms only.

Variable	Location	100 R ²	Probability					
			Residual Mean Square	Regression		Method		Runs by Lab (tf)
				Square	Method	Lab	Lab	
NO ₃ + NO ₂	BA0011	1.0	0.000045	0.77	0.35	0.85	0.66	<0.0001
	logs ^a	0.2		0.96	0.60	0.95	0.90	<0.0001
	BE0013	1.0	0.00105	0.77	0.57	0.60	0.47	<0.0001
	logs	1.3		0.69	0.45	0.47	0.55	<0.0001
	BH0017	5.3	0.001201	0.13	0.30	0.66	0.04	0.0001
	logs	4.0		0.23	0.59	0.92	0.05	0.0002
	BN0001	1.0	0.3049	0.77	0.77	0.51	0.43	<0.0001
	logs	2.7		0.39	0.56	0.94	0.10	<0.0001
Na	AK0001	0.6	0.2463	0.81	0.75	0.37	0.78	<0.0001
	logs	1.4		0.54	0.70	0.51	0.21	<0.0001
	BA0011	0.3	0.03845	0.56		0.56		<0.0001
	BE0013	2.2	0.1575	0.12		0.12		<0.0001
	BH0017	0.02	0.3853	0.88		0.88		0.001
	BN0001	2.9	14.37	0.08		0.08		0.03
	AK0001	0.7	40.35	0.75		0.75		<0.0001

^a logs indicate that the results of that particular row were obtained from logarithmically transformed data.

FIGURE CAPTIONS

- Figure 1. Dissolved nitrogen concentrations at BA0011 plotted against time, with data points joined by lines, and method and laboratory changes indicated.
- Figure 2. The fitted values for model (1) and $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011, joined by lines, are plotted with the data.
- Figure 3. The residuals, from fitting model (1) to $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011, plotted against day in year.
- Figure 4. The residuals, from fitting a model without year terms to $\text{NO}_3 + \text{NO}_2$ concentration at BA0011, plotted against day within year.
- Figure 5. Q-Q plot of the residuals from fitting model (1) to $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011.
- Figure 6. Q-Q plot of the residuals from fitting model (1) to the logarithm of $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011.
- Figure 7. The residuals, from fitting model (1) to the logarithm of $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011, plotted against day in year.
- Figure 8a. The seasonal components for 1986 as estimated from fitting model (1) to the $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011 between 1977 and 1987.

Figure 8b. The fitted values \hat{Y}_{ijks} and the adjusted mean for 1986, obtained from fitting model (1) to the $\text{NO}_3 + \text{NO}_2$ concentrations at BA0011 between 1977 and 1987, plotted with the data.

Figure 9. Dissolved nitrogen concentration plotted against time, with data points joined by lines, and method and laboratory changes indicated for BE0013, BH0017, BN0001 and AK0001.

Figure 10. Final model plotted with $\text{NO}_3 + \text{NO}_2$ data, standardized residuals plotted against time and the probability plot (Q-Q plot) for locations BE0011, BN0001 and AK0001.

Figure 11. Fitted model and residual plots for three different models fitted to the BH0017 $\text{NO}_3 + \text{NO}_2$ data.

Figure 12. Sodium concentrations plotted against time, with data points joined by lines and the laboratory change indicated.

Figure 13. Sodium concentrations and discharge plotted against time.

Figure 14. Final model plotted with Na data, standardized residuals plotted against time and the probability plot (Q-Q plot) for all five locations.

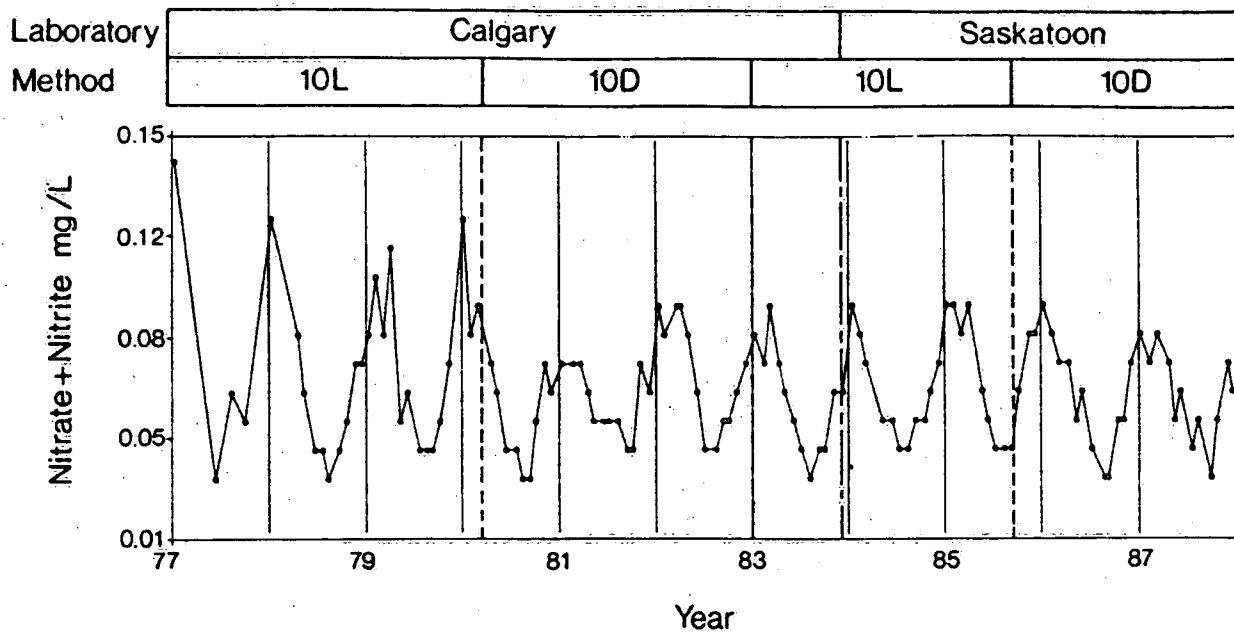


FIGURE 1

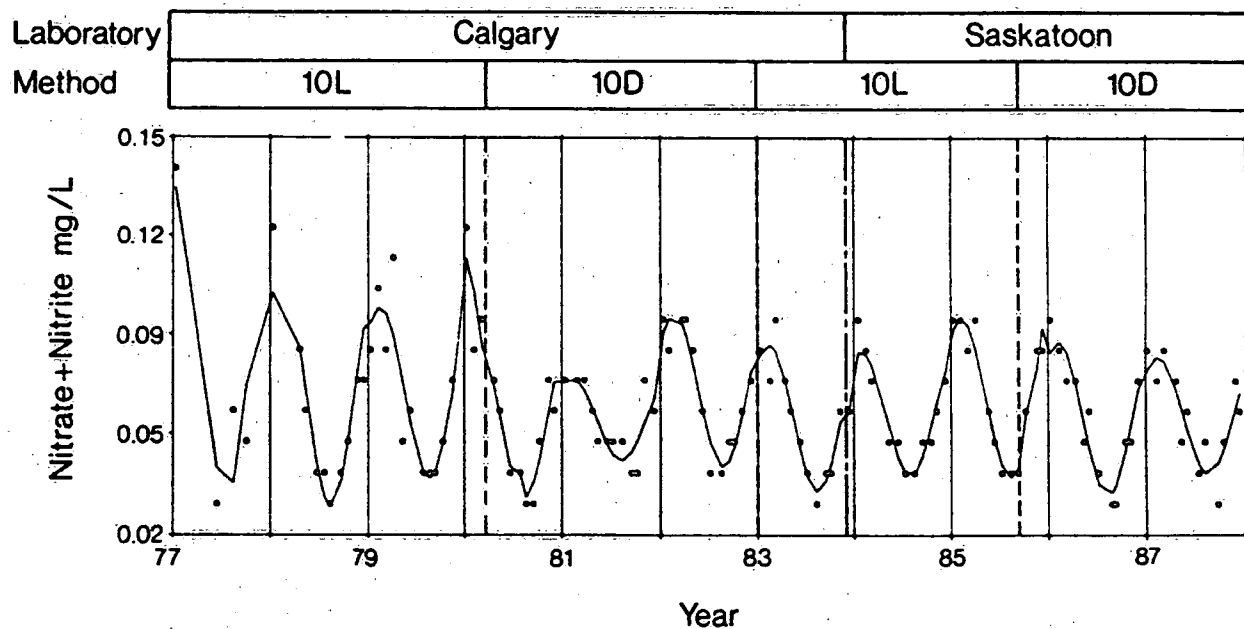


FIGURE 2

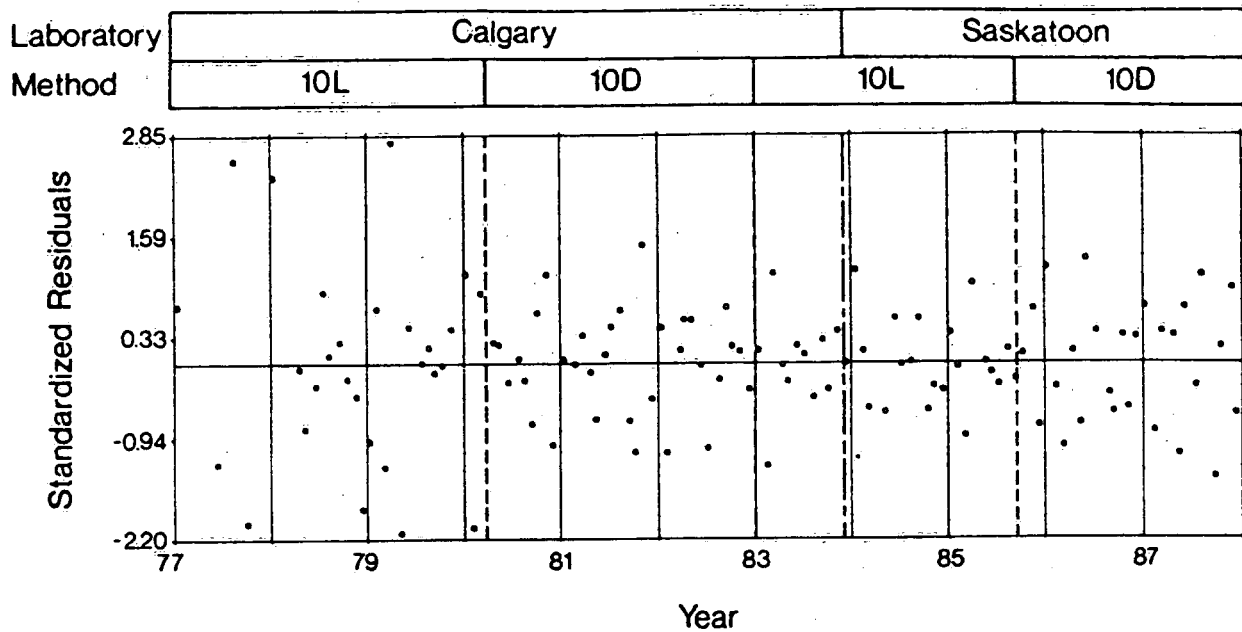


FIGURE 3

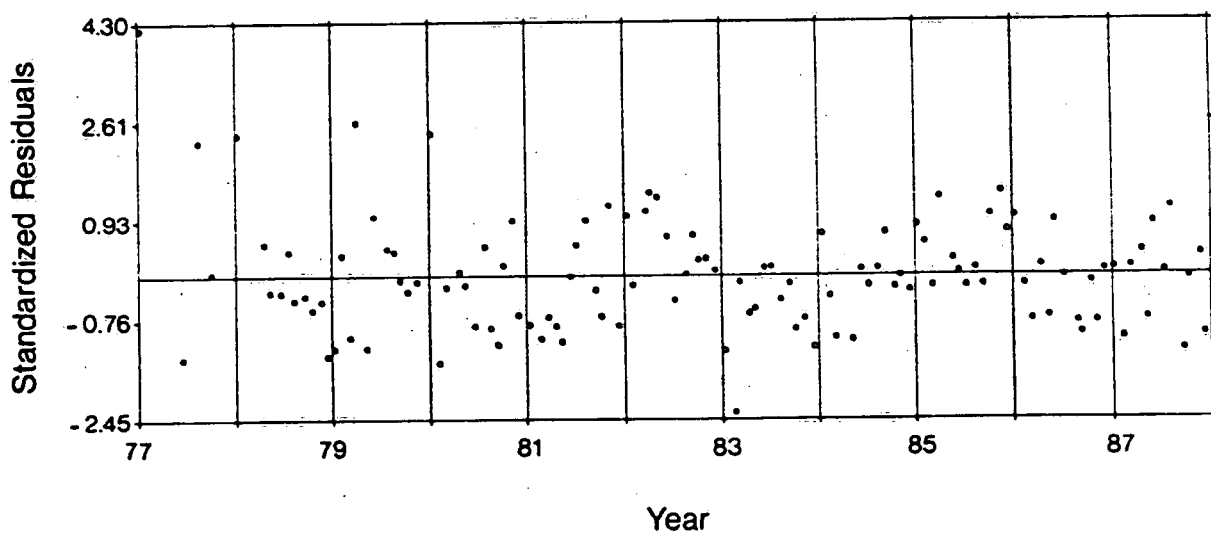


FIGURE 4

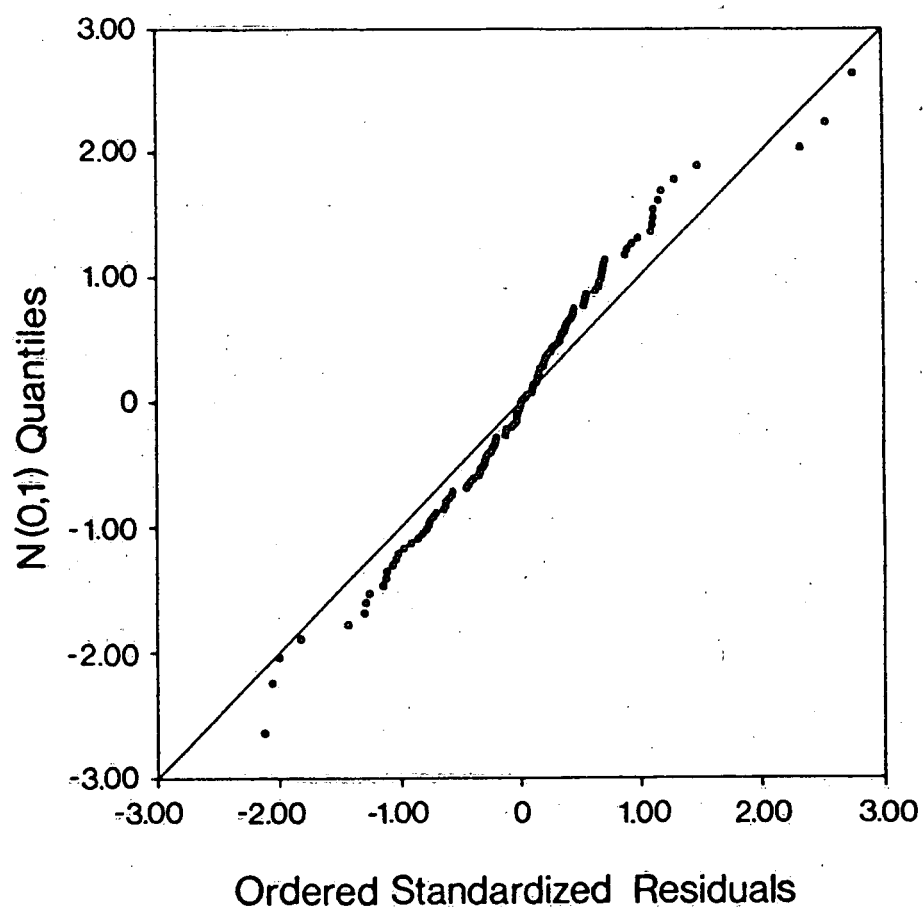


FIGURE 5

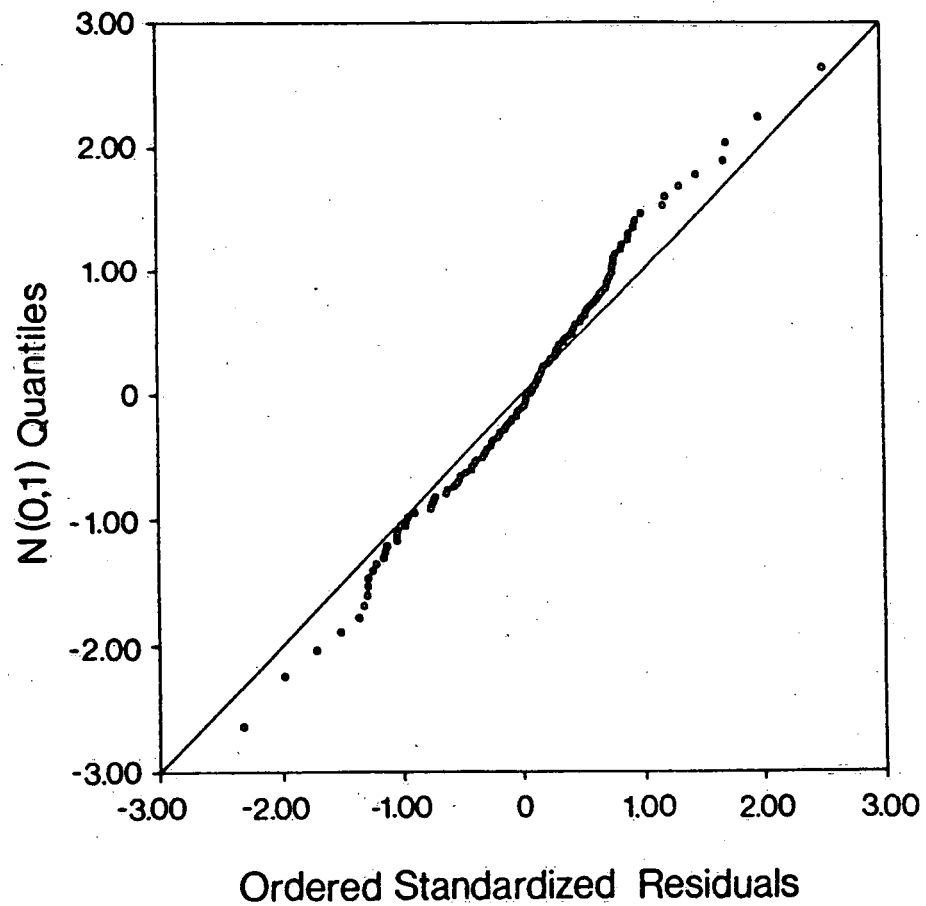


FIGURE 6

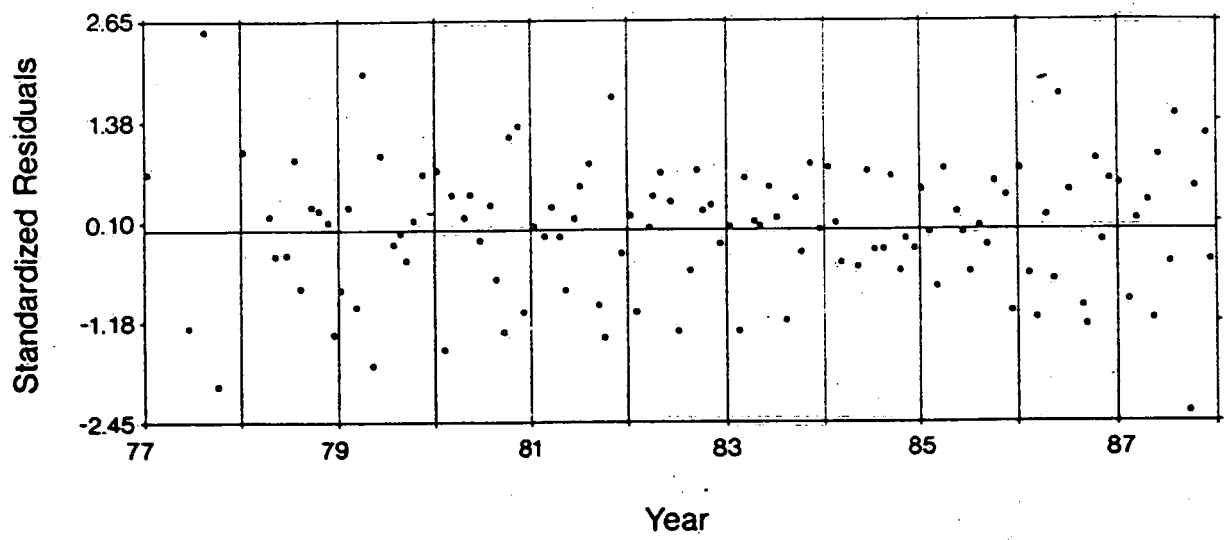
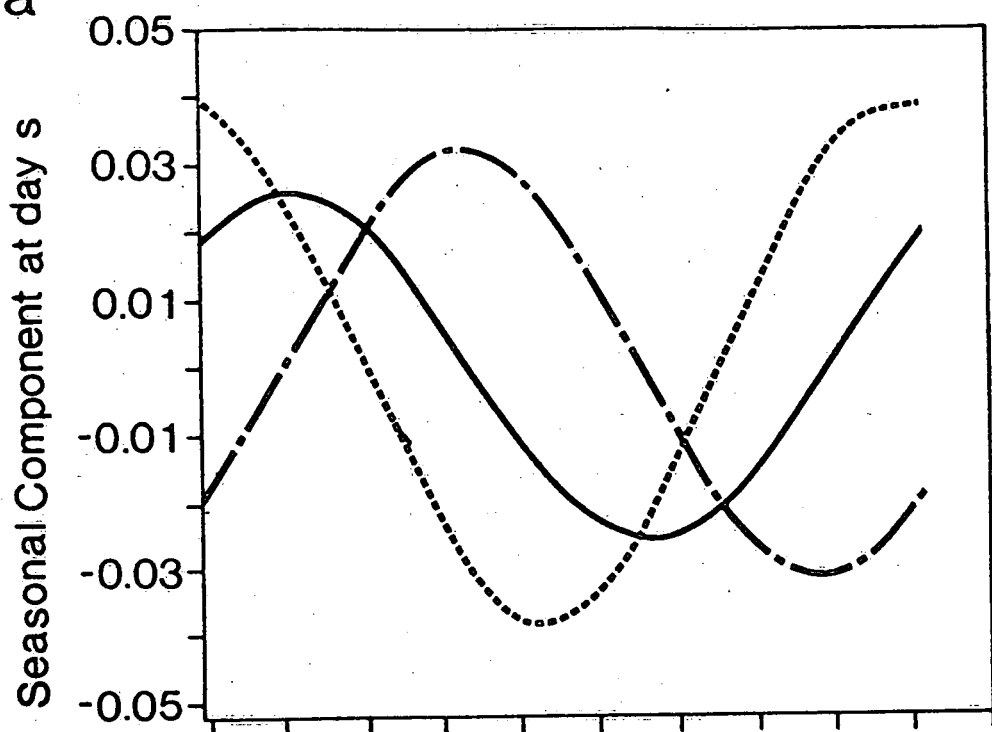


FIGURE 7

a

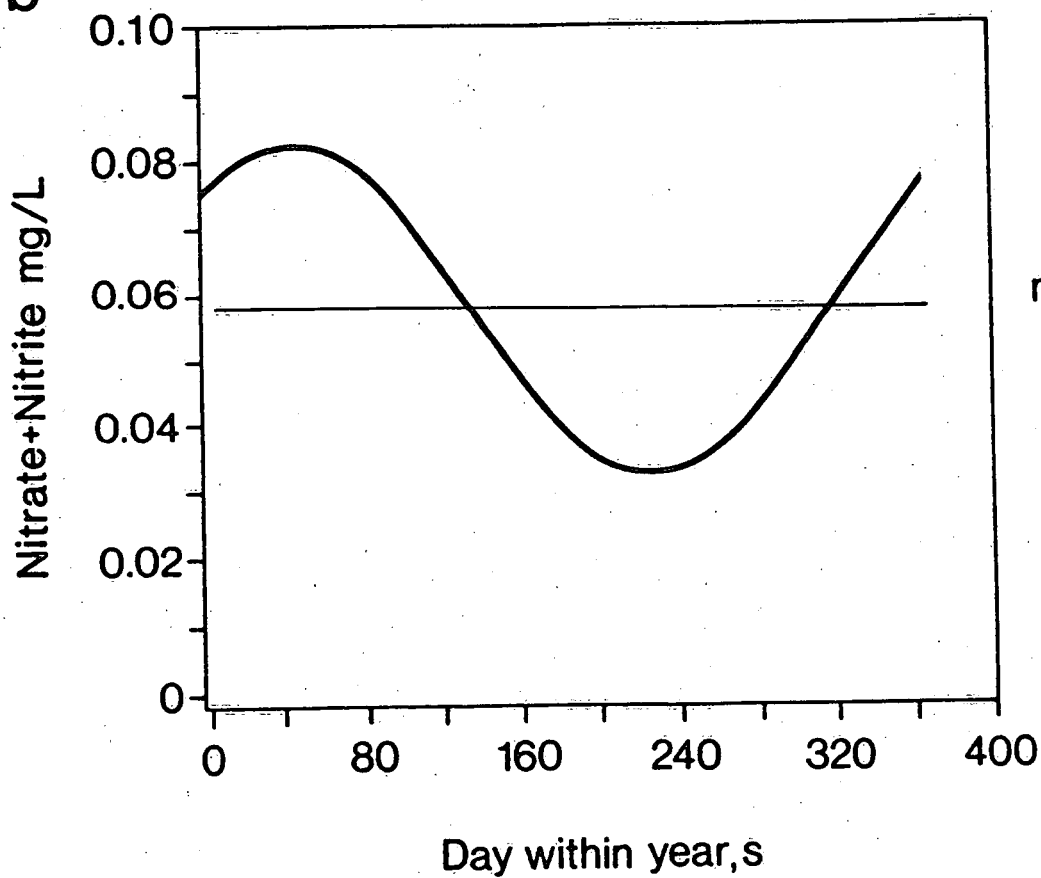


seasonal component
due to year only

final seasonal
component

modification due
to method

b



mean=0.058

FIGURE 8

Laboratory Method	Calgary		Saskatoon	
	10L	10D	10L	10D

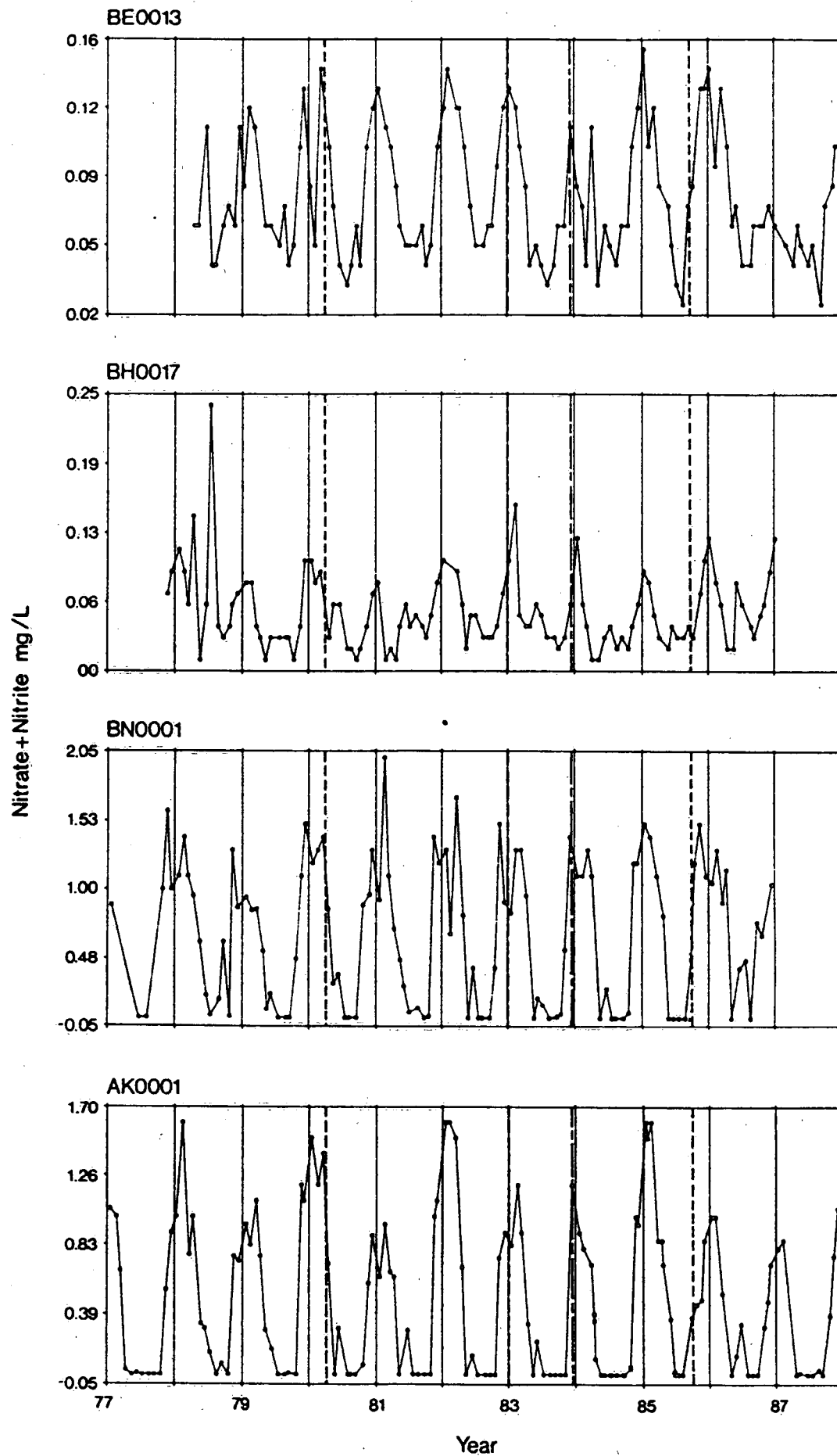


FIGURE 9

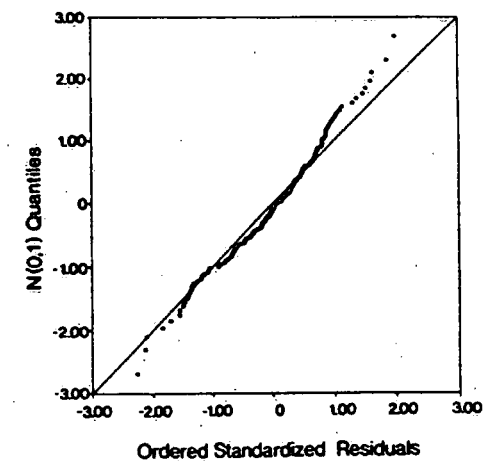
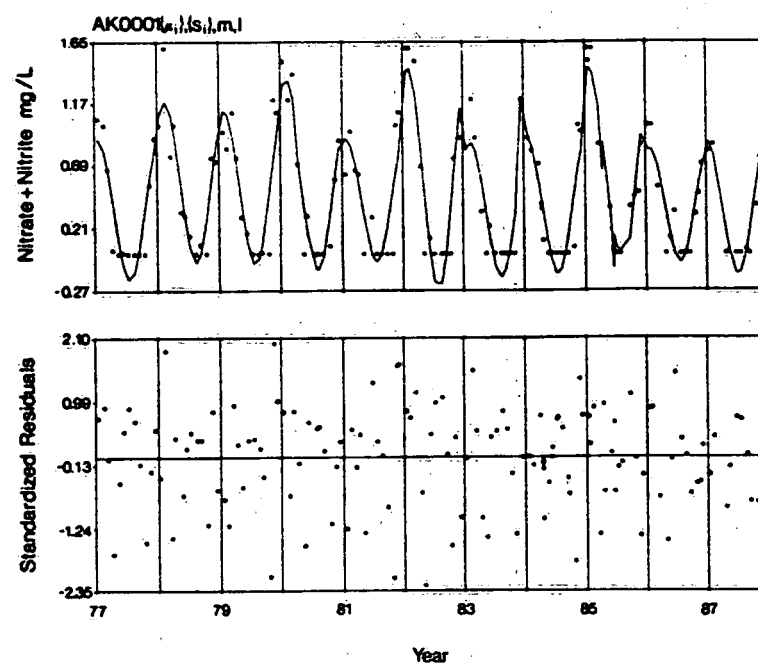
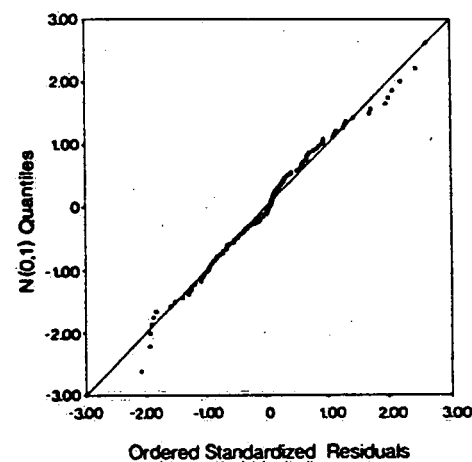
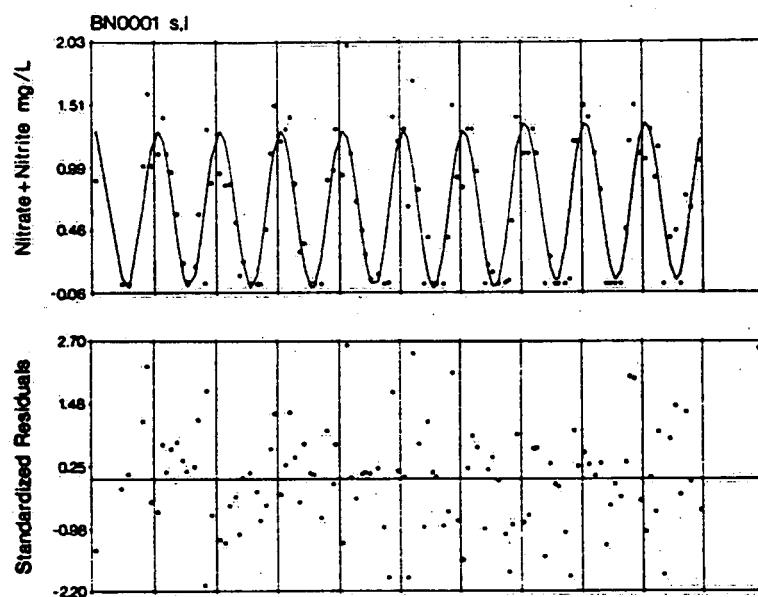
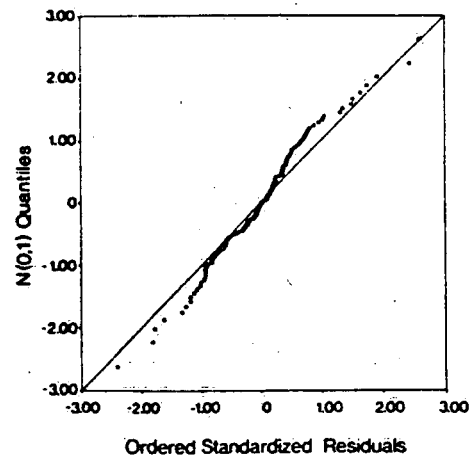
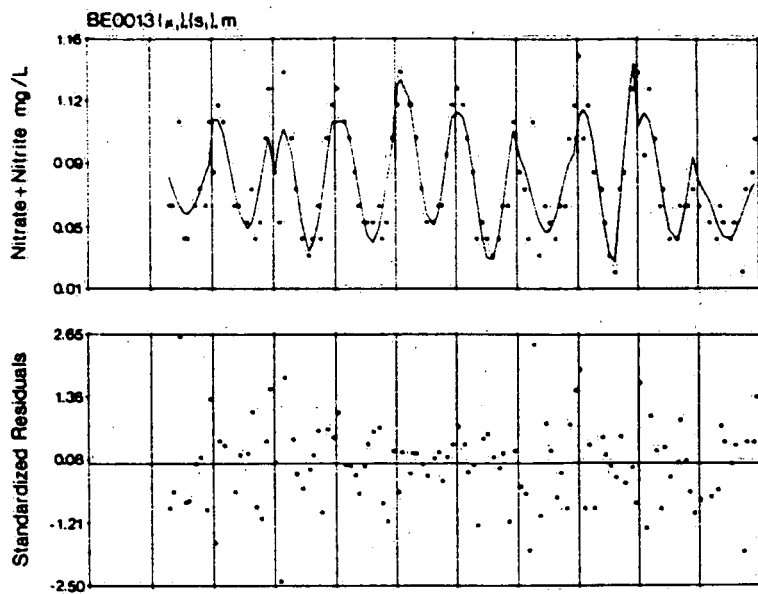


FIGURE 10

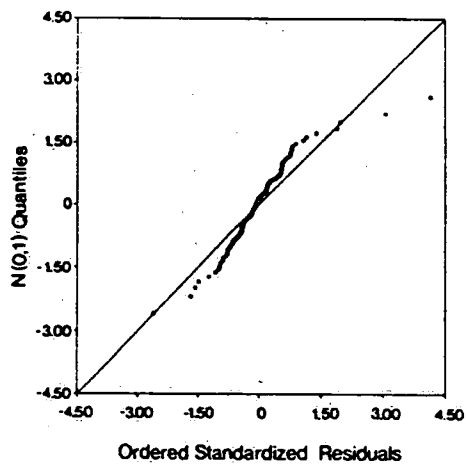
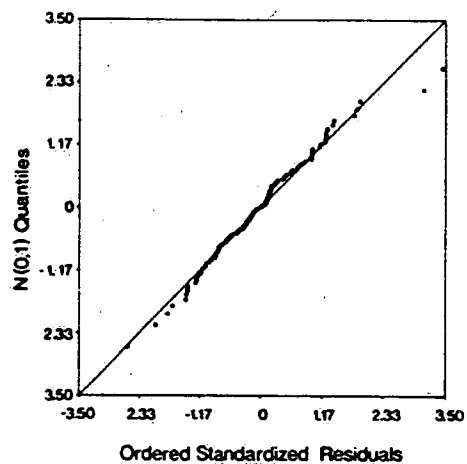
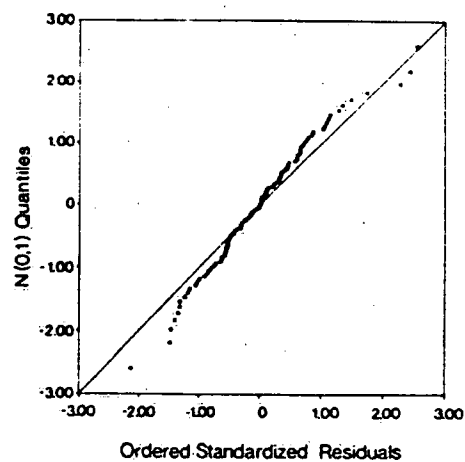
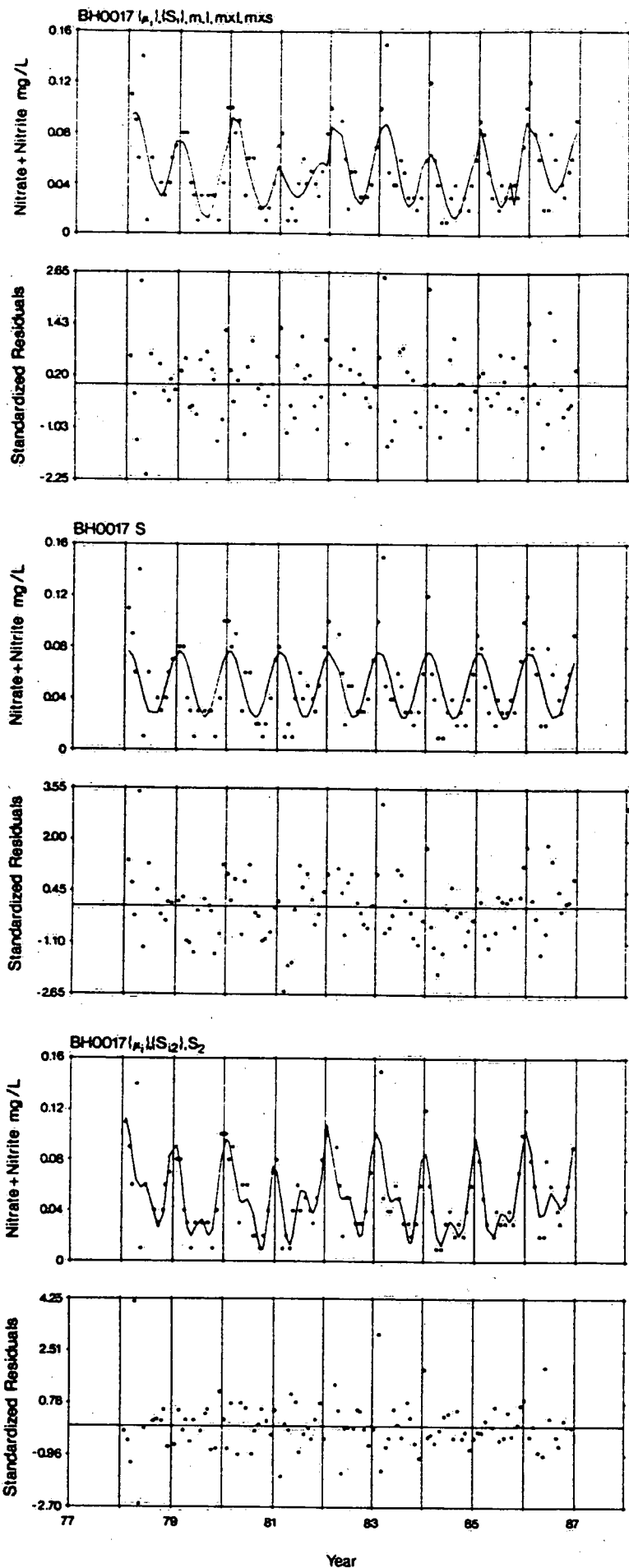


FIGURE 11

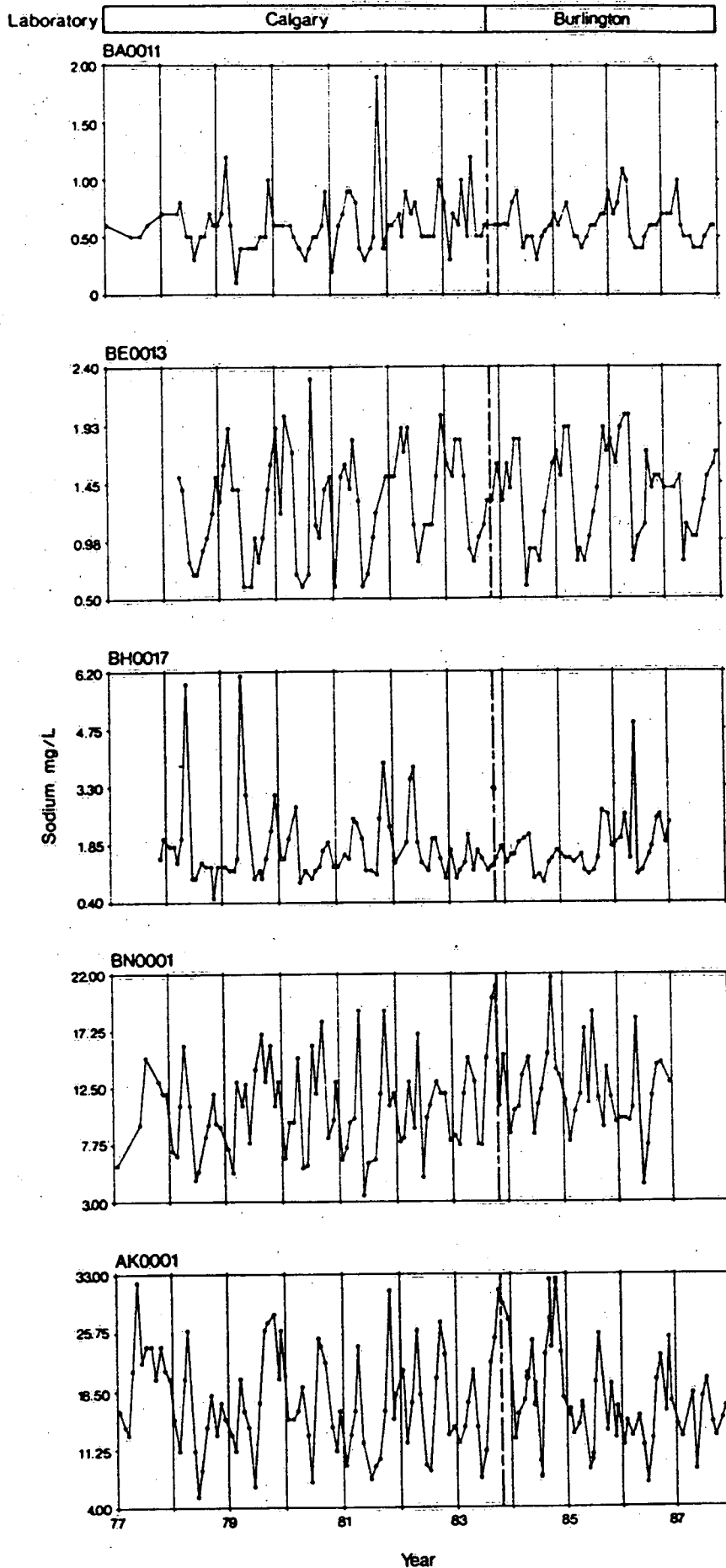


FIGURE 12

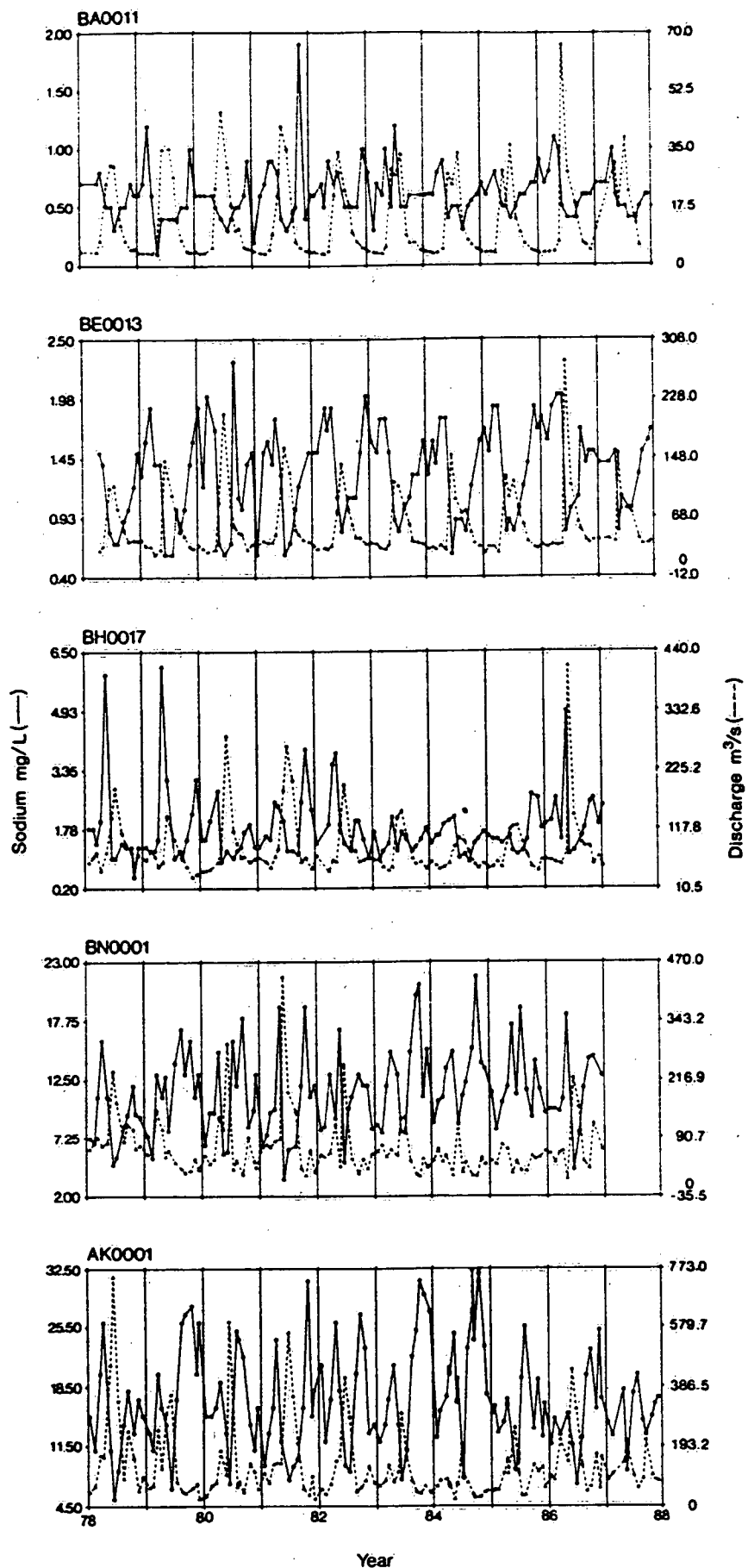


FIGURE 13

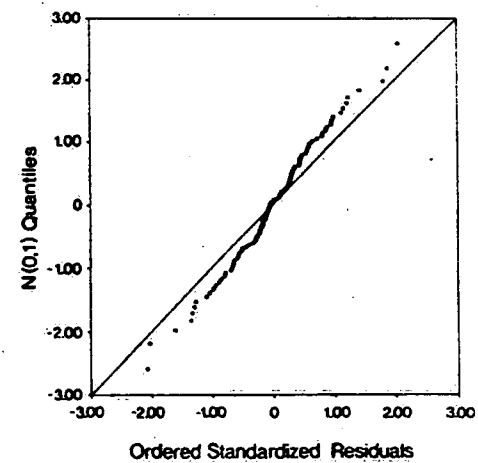
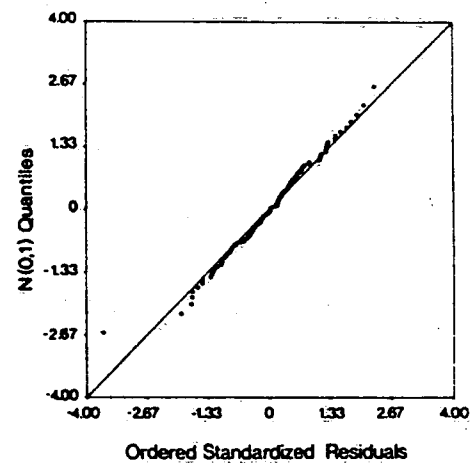
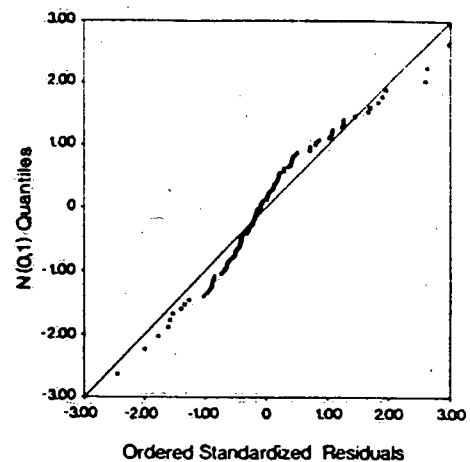
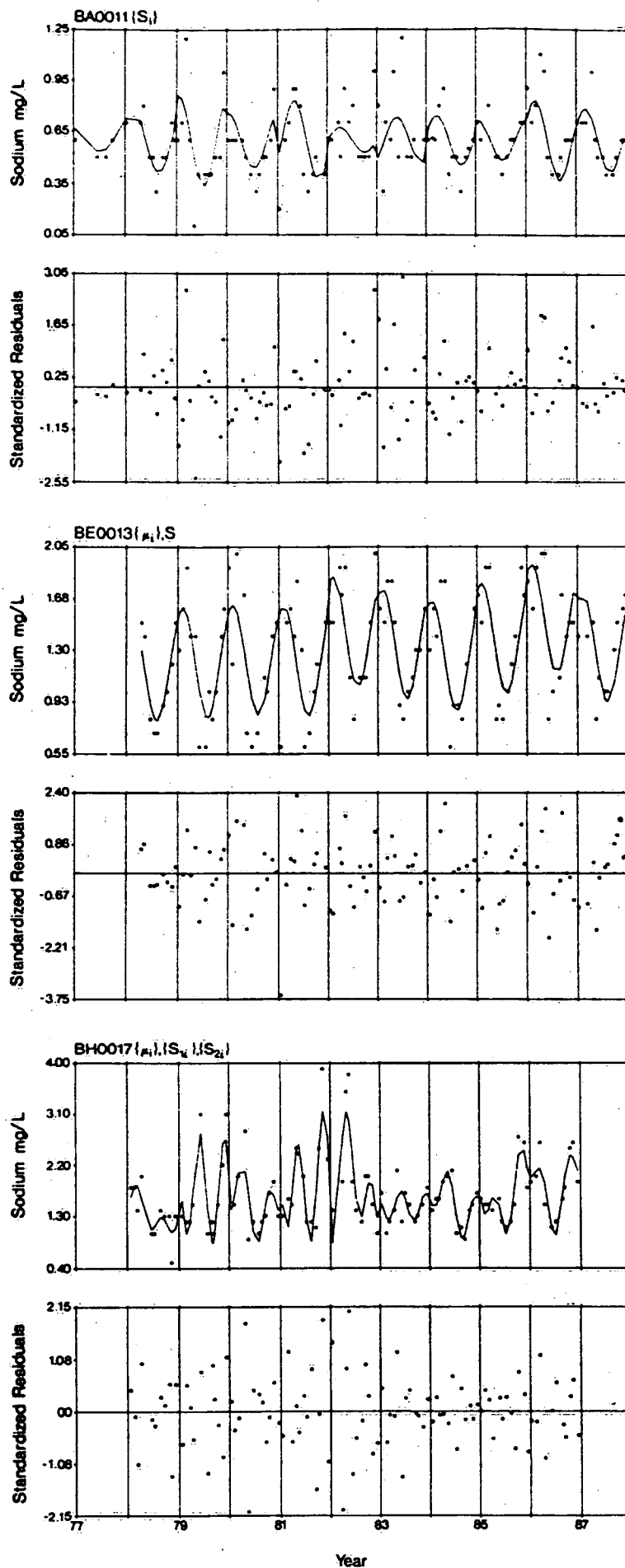


FIGURE 14

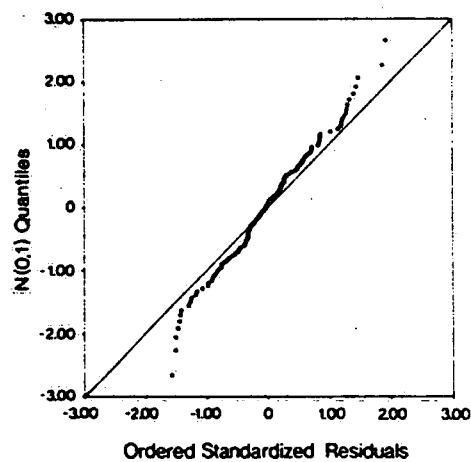
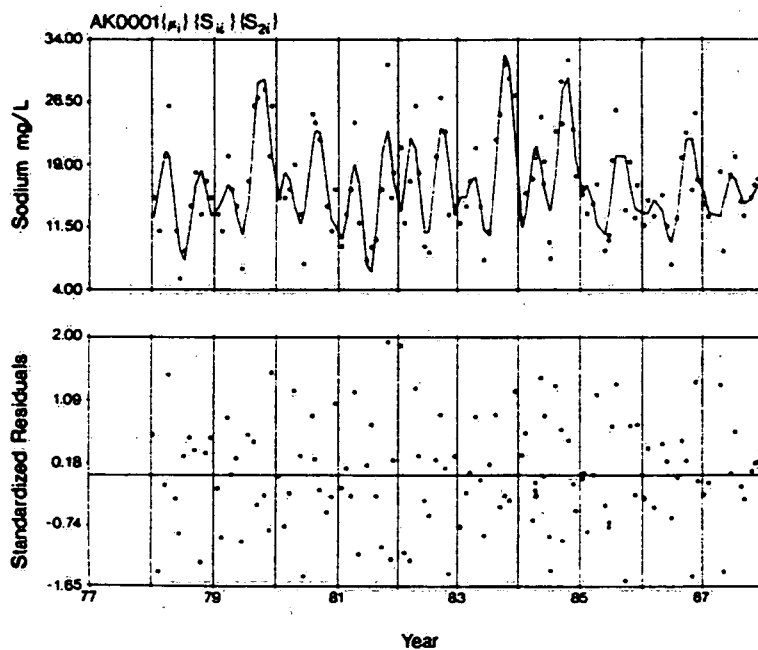
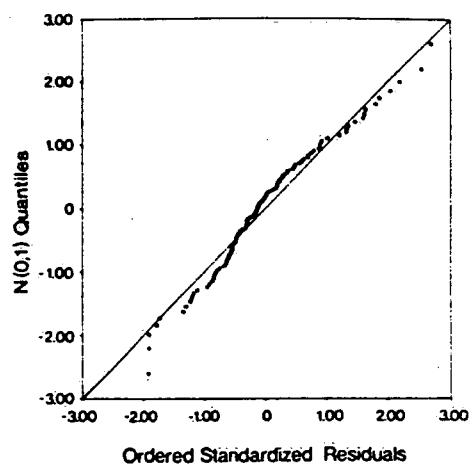
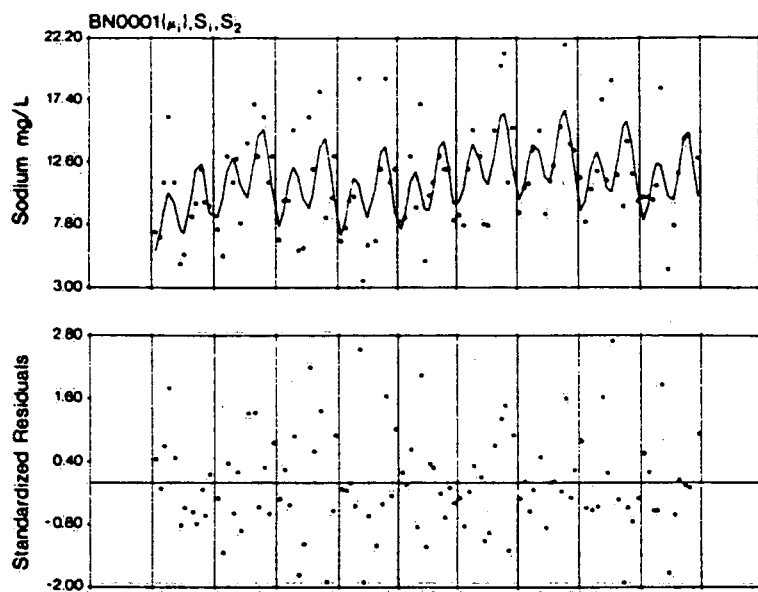
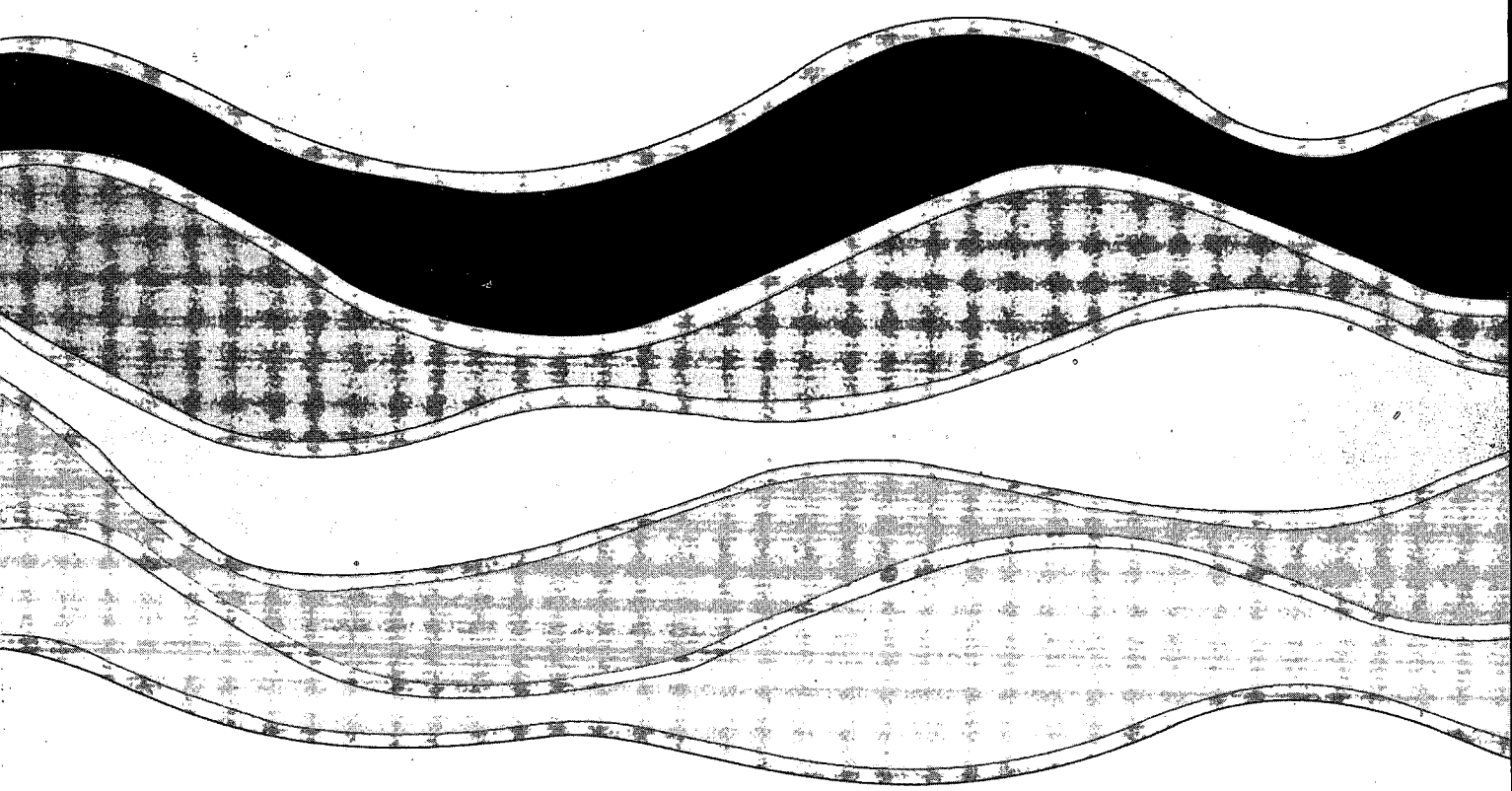


FIGURE 14 (continued)

Environment Canada Library, Burlington



3 9055 1017 0641 3



NATIONAL WATER RESEARCH INSTITUTE
P.O. BOX 5050, BURLINGTON, ONTARIO L7R 4A6



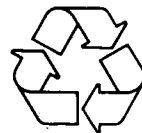
Environnement
Canada

Environnement
Canada

Canada

INSTITUT NATIONAL DE RECHERCHE SUR LES EAUX
C.P. 5050, BURLINGTON (ONTARIO) L7R 4A6

Think Recycling!



Pensez à recycler!