Water Quality Trends of Selected Variables in the Salmon River at Salmon Arm

Prepared for:

Environment Canada and
BC Ministry of Environment, Lands and Parks

Prepared by:
Robin Regnier
Central Limit Statistical Consulting
Vancouver, BC

January, 1999

Introduction

This report summarizes the statistical analyses of eight water quality variables sampled at the Salmon River at Salmon Arm monitoring site. The variables analyzed were as follows:

· Fecal Coliforms;
· Dissolved Ortho-phosphorus;
· Total Phosphorus;
· Non-filterable Residue;
· Water Temperature;
· Turbidity;
· Magnesium;
· Potassium; and
· Sodium

Background information for the Salmon River at Salmon Arm monitoring site can be obtained by contacting the B.C. Ministry of Environment, Lands and Parks.

Methods

Exploratory Data Analysis (EDA)

Exploratory data analysis procedures are the `initial look' at a data set, providing a researcher with tools to select appropriate statistical tests and modeling techniques. Apart from computing basic summary statistics (means, medians, minimums, maximums, number of observations), EDA procedures are best represented by graphical displays of the data. Time series plots were used in the initial data explorations in this report.

Non-parametric Analyses
Non-parametric tests to detect trends in water quality have been used by many others in the past (Yu and Zou, 1993; Walker, 1991; Gilbert, 1987; Hirsch and Slack, 1984). The relative simplicity and minimal data assumptions of these tests make them a popular choice for analysis of water quality time series. Three different non-parametric tests, the *seasonal Kendall's Tau,* the *modified seasonal Kendall's Tau,* and the *Sen slope estimator* were used to detect and determine magnitudes of any trends in the water quality data presented here.

Seasonal Kendall's Tau

A rank-order statistic that can be applied to time series exhibiting seasonal cycles, missing and censored data, and indications of non-normality (Yu and Zou, 1993). For computational details, see Gilbert (1987) and Hirsch and Slack (1984).

Modified seasonal Kendall's Tau

The *seasonal Kendall's Tau* assumes that data are serially independent, that is, values are not determined in whole or in part on the previous state in the sequence. To compensate for serial dependence in a data series, Hirsch and Slack (1984) proposed a modification to the *seasonal Kendall's Tau* that takes into account any covariation between seasons in a data set.
Either version of the two *seasonal Kendall* tests are most appropriate if trends are consistent throughout a year. For example, a negative trend for six months followed by a positive trend of six months would yield a test statistic indicating zero trend (the two tests do not measure the size of any trends, only the direction).

Sen slope estimator

This non-parametric statistic calculates the magnitude of any significant trends found. The Sen slope estimator (Sen, 1968) is calculated as follows (Y is the variable of interest; X is the time at which the i^{th} observation was taken) :

$$D_{ij} = \left[\frac{Y_j - Y_i}{X_j - X_i}\right] \text{ for } i < j \text{ , } X_i \neq X_j$$

The slope estimate is the median of all D_{ij} values. Hirsch *et al*. (1982) point out that this estimate is robust against extreme outliers. Confidence bounds for this slope estimator are calculated as a simple percentile of the total number of calculated slopes (Gilbert, 1987).

Parametric Statistics
Non-parametric statistics test for monotonic changes in a data series with minimal assumptions of normality and, in some instances, serial dependence. However, these methods are not very useful in constructing the forms of any detectable trends. Regression analysis has been used for this purpose and

has been applied to water quality data in the past (El-Shaarawi *et al*., 1983, Esterby *et al*., 1989, Helsel & Hirsch, 1995, Shaw & El-Shaarawi, 1995).

Using these methods, many factors can be taken into account for explaining the variation in a water quality constituent over time, factors which include flow rates and seasonality. By accounting for flow and seasonality through functional approximation, their influence on the response constituent can be removed, revealing underlying trends.
The regression model used is as follows:

(1) $$y_{t_{ji}} = \beta_0 + \beta_1 x_{t_{ji}} + \beta_2 i + \alpha_1 \cos \omega t_{ji} + \alpha_2 \sin \omega t_{ji} + \varepsilon_{t_{ji}}$$

where :

$y_{t_{ji}}$ = Observed value of water quality variable at time t_{ji} within year i;

$x_{t_{ji}}$ = Flow rate at time t_{ji} within year i;

α_1, α_2 = Unknown parameters representing the phase of the seasonal cycle;
ω = Unknown parameter representing the frequency of the seasonal cycle;

$\varepsilon_{t_{ji}}$ = Error term assumed to follow a normal distribution with mean 0 and variance σ^2.

This regression technique is an iterative process of parameter estimation and analyses of model residual and quantile plots.

The form in equation (1) above considers only a linear trend with slope β_2. The presence or absence of quadratic (U or upside down U - shaped) trends may be determined by fitting the data to (1) with the addition of a quadratic term $(\beta_3 i^2)$.

ANOVA tables may then be used to determine if the quadratic models significantly improve the linear models. Significance of the model coefficients are tested at the 5 percent level.

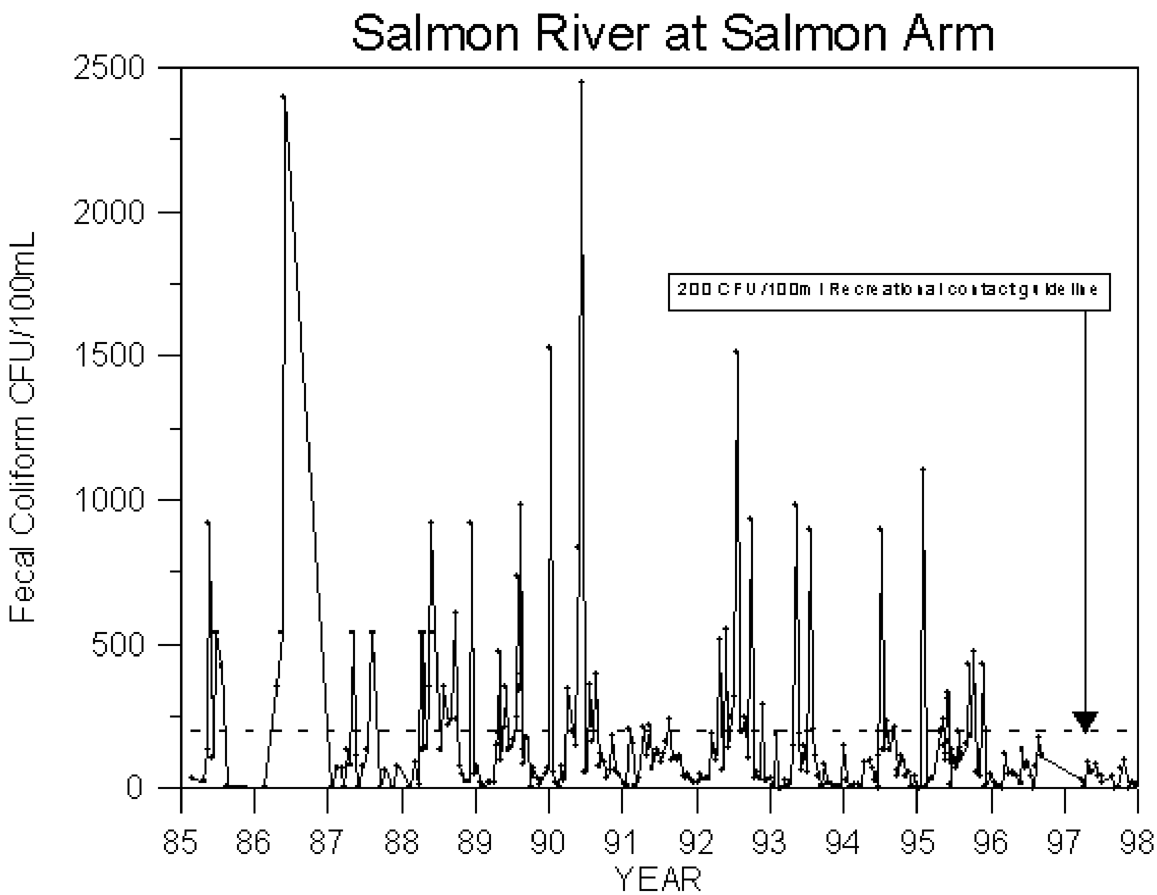
Results and Discussion

The results of the statistical analyses are presented on the following pages, including the time series plots and tables of the non-parametric tests and regression modeling performed.

It should be noted that apparent conflicting results between non-parametric tests and regression model findings are not necessarily a reason to be alarmed. One must remember that regression models take into account additional explanatory variables that may or may not help explain variation in the data being modeled.

Fecal Coliforms
Fecal Coliform data collected from 1985 through 1997 were used for statistical analyses (Figure 1).

Figure 1 Time series plot of fecal coliforms in the Salmon River at Salmon Arm, 1985 - 1997.



Due to missing values in the data set, analyses were conducted on data from the start of 1987 to the end of 1997. Non-parametric tests for trend indicated evidence of a significant linear decreasing trend. Subsequent regression modeling fit a negative-quadraticmodel, suggesting that levels are indeed decreasing, at least for more recent data. Table 1 presents the statistics for the non-parametric results while Table 2 contains the regression findings.

Table 1 Non-parametric results for fecal coliforms in the Salmon River at Salmon Arm.

	Statistic	***P-value***
SK	-3.978	0.0001
MSK	-2.302	0.0214
Sen	-5.817	NA
LCL	-8.949	NA

UCL	-2.814	NA

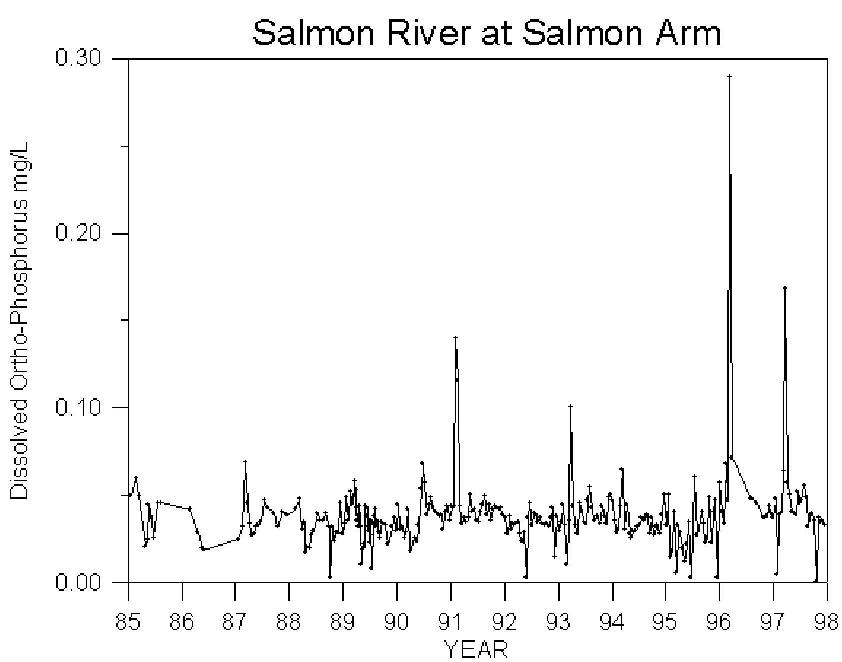
Table 2 Regression modeling results for fecal coliforms in the Salmon River at Salmon Arm. Note that both fecal coliform and flow data were log transformed before modeling.

Coefficient	*Statistic*	*p-value*
b_0	3.240	0.0000
b_1	0.408	0.0003
b_2	0.347	0.0035
b_3	-0.031	0.0001
a1	-0.551	0.0002
a2	-0.821	0.0000
r^2	0.34	NA

Dissolved Ortho-phosphorus

Dissolved ortho-phosphorus data collected from 1985 through 1997 were used for statistical analyses (Figure 2).

Figure 2 Time series plot of dissolved ortho-phosphorus in the Salmon River at Salmon Arm, 1985 - 1997.



Due to missing values in the data set, analyses were conducted on data from the start of 1987 to the end of 1997. Non-parametric tests for trend indicated no evidence of any trends. Subsequent regression modeling fit an increasing-trend model. The overall fit of the regression model was very low however ($r^2 < 0.1$), and suggests that the model is not adequate. Further analyses, which are beyond the scope of this report, would be beneficial to explore different models used for the regression fit. Table 3 presents the statistics for the non-parametric results, while Table 4 contains the regression findings.

Table 3 Non-parametric results for dissolved ortho-phosphorus in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	0.724	0.469
MSK	0.397	0.691
Sen	0.0001	NA
LCL	-0.0003	NA
UCL	0.0005	NA

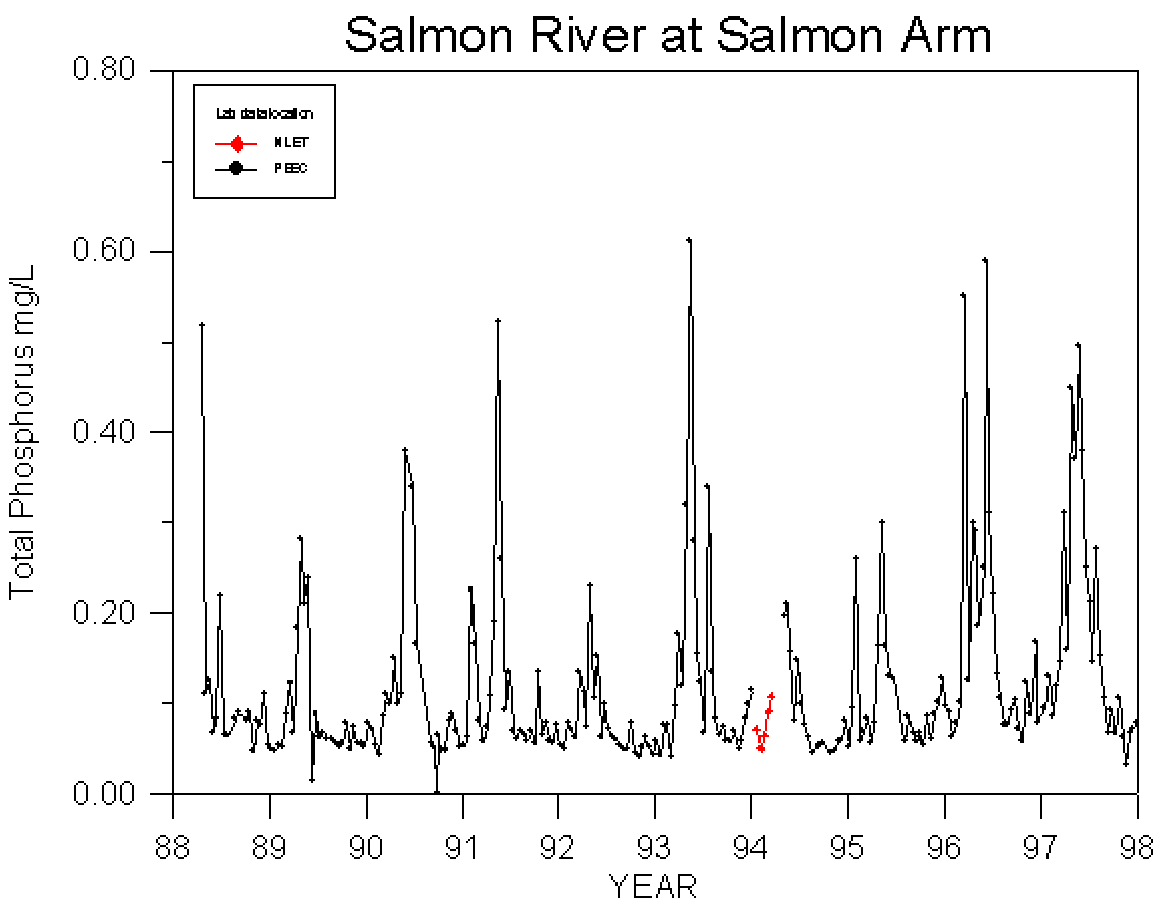
Table 4 Regression modeling results for dissolved ortho-phosphorus in the Salmon River at Salmon Arm. Note that flow data only were log transformed before modeling.

Coefficient	*Statistic*	*p-value*
b_0	0.0328	0.0000
b_1	-0.0004	0.7889
b_2	0.0009	0.0243
b_3	----	----
a1	0.0046	0.0264
a2	0.0009	0.7440
r^2	0.04	NA

Total Phosphorus

Total phosphorus data collected from 1988 through 1997 were used for statistical analyses (Figure 3).

Figure 3 Time series plot of total phosphorus in the Salmon River at Salmon Arm, 1988 - 1997.



Under the assumption that the data is not serially correlated, the seasonal Kendall test for trend found strong evidence of a linear increasing trend. Using the modified version of this test, where the data is assumed to be serially correlated, there is some evidence of a linear increasing trend, although much weaker than the unmodified version (0.05 < p-value < 0.1; see table 5).
Subsequent regression modeling fit a trend-less model. This may indicate that the non-parametric evidence of an increasing trend may be explained by historical flow patterns. Table 5 presents the statistics for the non-parametric results while Table 6 contains the regression findings.

Table 5 Non-parametric results for total phosphorus in the Salmon River at Salmon Arm.

	Statistic	***P-value***
SK	3.643	0.0003
MSK	1.741	0.0817
Sen	0.004	NA
LCL	0.0015	NA

UCL	0.0063	NA

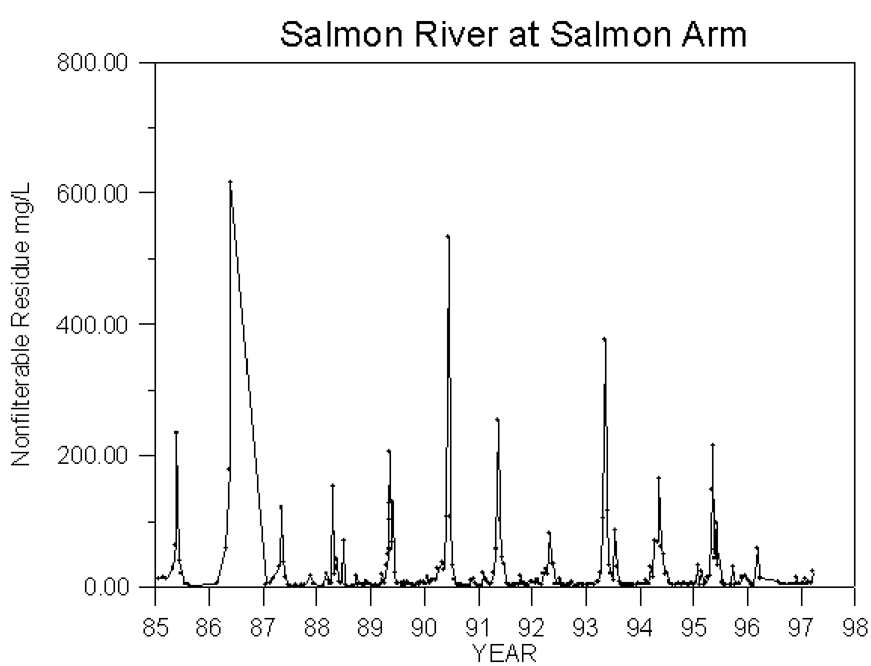
Table 6 Regression modeling results for total phosphorus in the Salmon River at Salmon Arm. Note that both total phosphorus and flow data were log transformed before modeling.

Coefficient	*Statistic*	*p-value*
b_0	-3.0541	0.0000
b_1	0.5611	0.0000
b_2	0.0112	0.2504
b_3	----	----
a1	-0.0989	0.0111
a2	0.1296	0.0018
r^2	0.61	NA

Non-filterable Residue

Non-filterable residue data collected from 1987 through 1996 were used for statistical analyses (Figure 4).

Figure 4 Time series plot of non-filterable residue in the Salmon River at Salmon Arm, 1985 - 1998.



Due to missing values in the data set, analyses were conducted on data from the start of 1987 to the end of 1996. Non-parametric tests for trend indicated strong evidence of a linear increasing trend. Subsequent regression modeling fit a trend-less model. Note that the regression modeling used, as in throughout this report, are not robust to outliers nor to multiple detection limits (see figure 5 for clear evidence of multiple detection limits in these data). This allows non-parametric techniques to be much better suited to analyzing these data. Other regression techniques are available to better handle multiple detection limit data, but are beyond the scope of this analysis. Table 7 presents the statistics for the non-parametric results while Table 8 contains the regression findings.

Figure 5 Semi-Log time series plot of non-filterable residue in the Salmon River at Salmon Arm, 1987 - 1996.

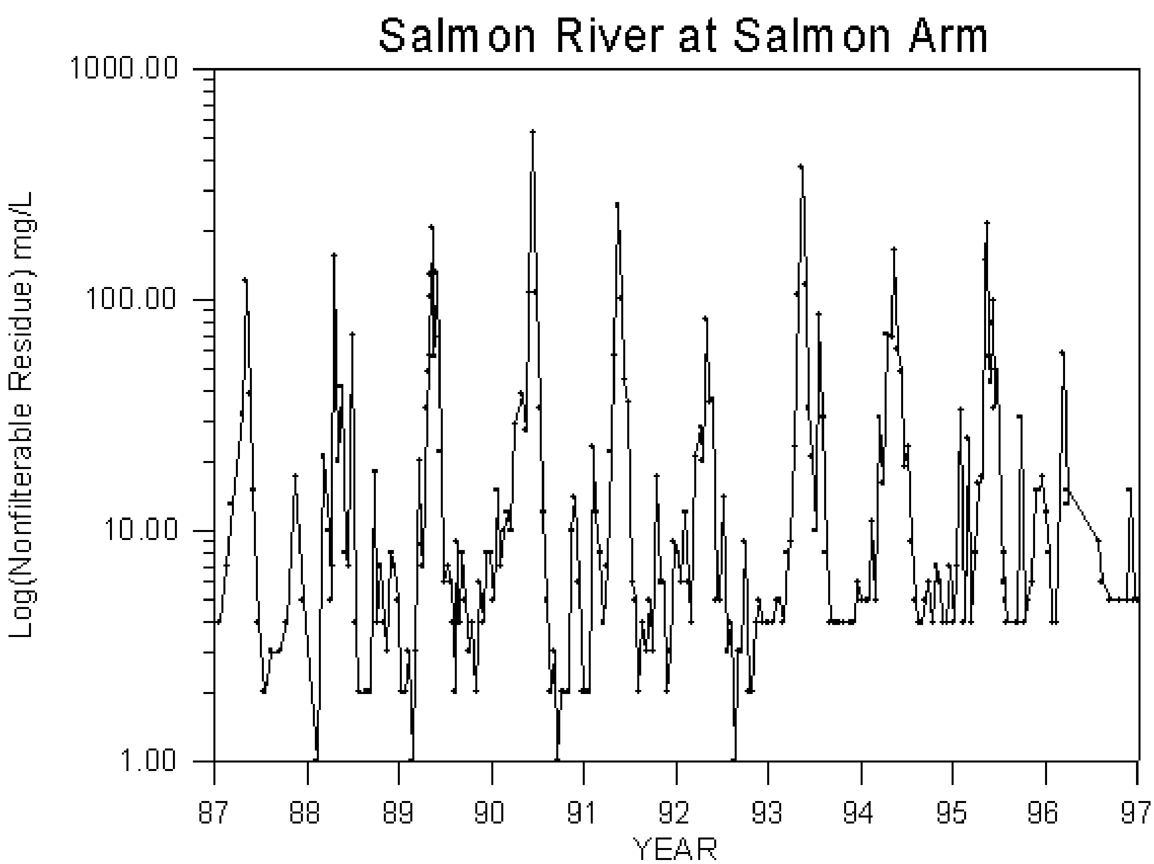


Table 7 Non-parametric results for non-filterable residue in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	2.6743	0.0075
MSK	2.3093	0.0209
Sen	0.3452	NA

LCL	0.1135	NA
UCL	0.6914	NA

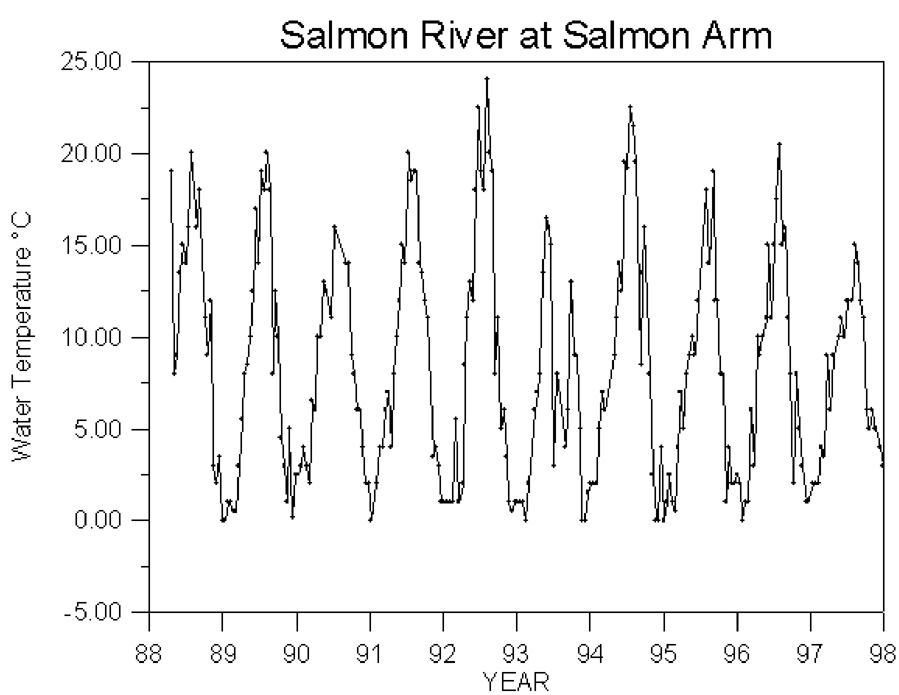
Table 8 Regression modeling results for non-filterable residue in the Salmon River at Salmon Arm. Note that both non-filterable residue and flow data were log transformed before modeling.

Coefficient	***Statistic***	***p-value***
b_0	0.9643	0.0000
b_1	1.1955	0.0000
b_2	0.0224	0.0948
b_3	----	----
a1	-0.0824	0.1315
a2	-0.0233	0.6562
r^2	0.80	NA

Water Temperature

Water temperature data collected from 1988 through 1997 were used for statistical analyses (Figure 6).

Figure 6 Time series plot of water temperature in the Salmon River at Salmon Arm, 1988 - 1997.



Non-parametric tests for trend indicated no evidence of any trend. Subsequent regression modeling also fit a trend-less model, although the model fit is relatively low ($r^2 < 0.2$; see table 10). Table 9 presents the statistics for the non-parametric results while Table 10 contains the regression findings.

Table 9 Non-parametric results for water temperature in the Salmon River at Salmon Arm.

	Statistic	***P-value***
SK	-0.808	0.419
MSK	-0.600	0.549
Sen	-0.035	NA
LCL	-0.154	NA
UCL	0.042	NA

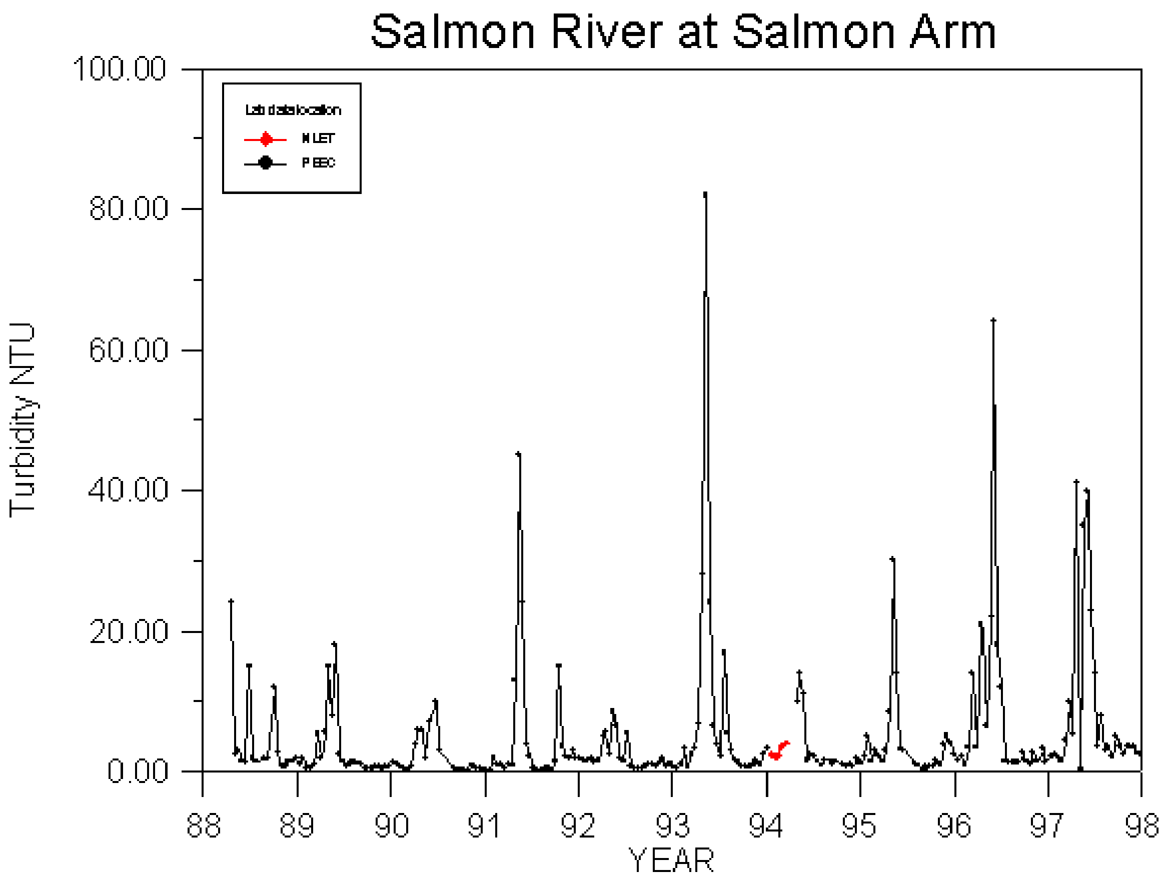
Table 10 Regression modeling results for water temperature in the Salmon River at Salmon Arm. Note that both water temperature and flow data were log transformed before modeling.

Coefficient	***Statistic***	***p-value***
b_0	1.7495	0.0000
b_1	0.2156	0.0016
b_2	-0.0263	0.2235
b_3	----	----
a1	-0.0184	0.8415
a2	-0.4888	0.0000
r^2	0.19	NA

Turbidity

Turbidity data collected from 1988 through 1997 were used for statistical analyses (Figure 7).

Figure 7 Time series plot of turbidity in the Salmon River at Salmon Arm, 1988 - 1997.



Non-parametric tests for trend indicated evidence of a significant linear increasing trend. Subsequent regression modeling also fit an increasing-trend model. These findings are readily apparent in the semi-log plot of the data analyzed (see figure 8). Table 11 presents the statistics for the non-parametric results while Table 12 contains the regression findings.

Figure 8 Semi-log time series plot of turbidity in the Salmon River at Salmon Arm, 1988 - 1997.

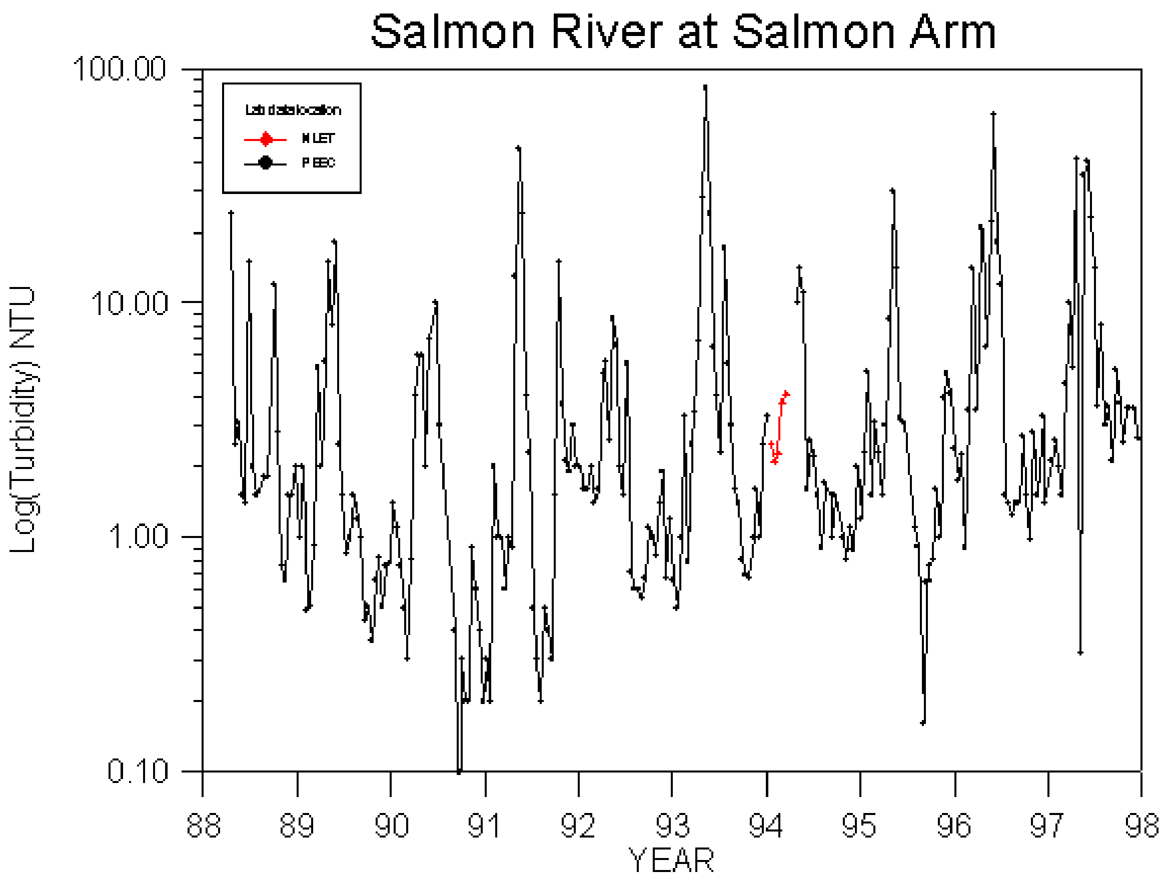


Table 11 Non-parametric results for turbidity in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	5.455	4.9e-08
MSK	2.504	0.012
Sen	0.25	NA
LCL	0.16	NA
UCL	0.35	NA

Table 12 Regression modeling results for turbidity in the Salmon River at Salmon Arm. Note that both turbidity and flow data were log transformed before modeling.

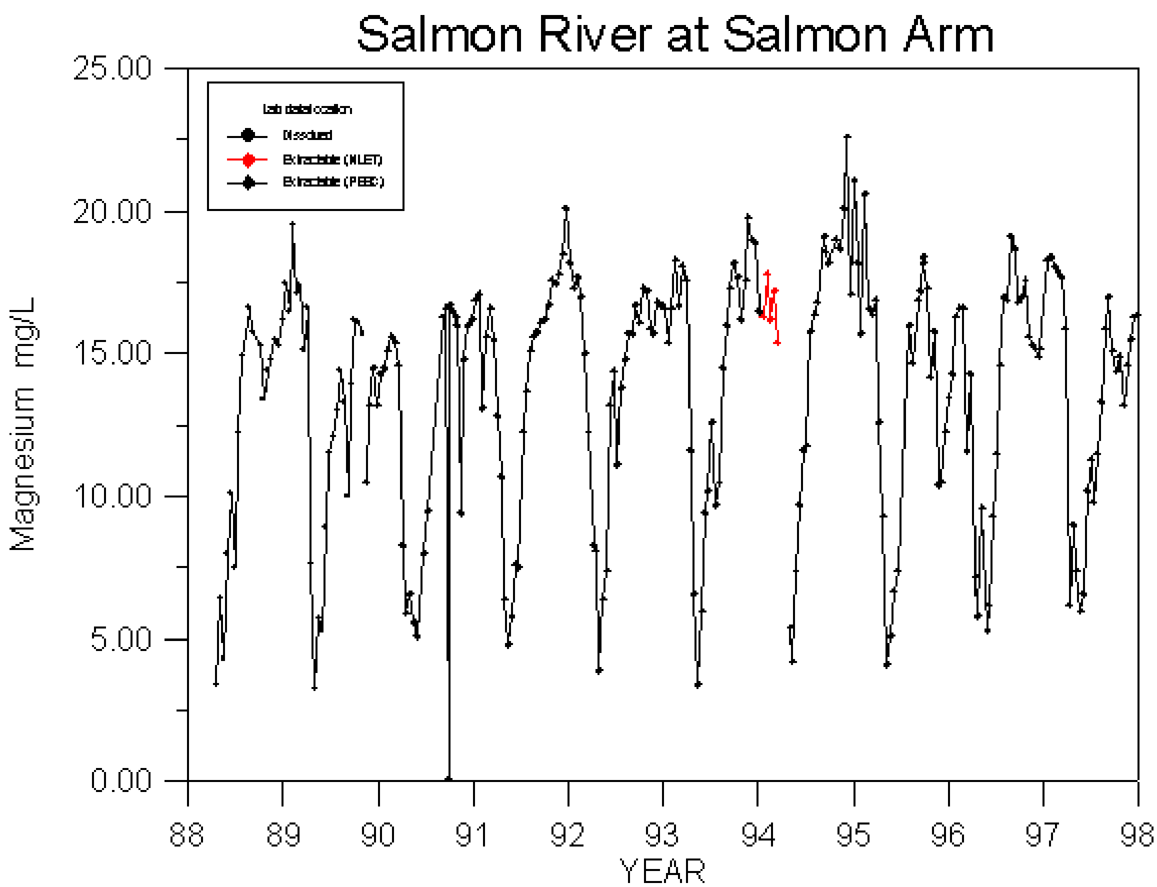
Coefficient	*Statistic*	*p-value*
b_0	-0.625	0.0000

b_1	0.9385	0.0000
b_2	0.0547	0.0047
b_3	----	----
a1	-0.0098	0.8978
a2	0.0765	0.3384
r^2	0.58	NA

Magnesium

Magnesium data collected from 1988 through 1997 were used for statistical analyses (Figure 9).

Figure 9 Time series plot of magnesium in the Salmon River at Salmon Arm, 1988 - 1997.



Non-parametric tests for trend indicated no evidence of any trends. Subsequent regression modeling fit an increasing-trend model. Table 13 presents the statistics for the non-parametric results while Table 14 contains the regression findings.

Table 13 Non-parametric results for magnesium in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	1.655	0.098
MSK	1.270	0.204
Sen	0.092	NA
LCL	-0.024	NA
UCL	0.175	NA

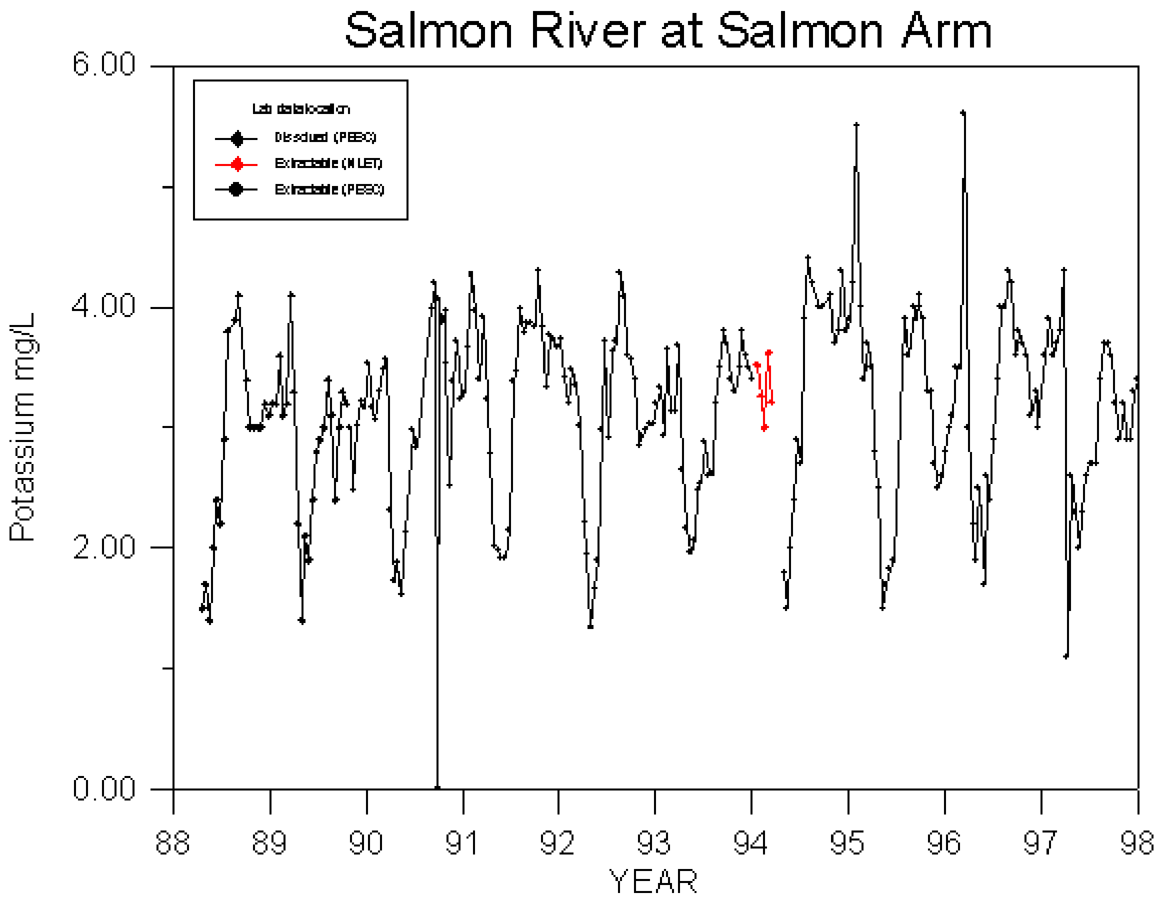
Table 14 Regression modeling results for magnesium in the Salmon River at Salmon Arm. Note that flow data only were log transformed before modeling.

Coefficient	*Statistic*	*p-value*
b_0	15.2316	0.0000
b_1	-3.9932	0.0000
b_2	0.4913	0.0000
b_3	----	----
a1	-0.5601	0.0030
a2	0.4976	0.0092
r^2	0.79	NA

Potassium

Potassium data collected from 1988 through 1997 were used for statistical analyses (Figure 10).

Figure 10 Time series plot of potassium in the Salmon River at Salmon Arm, 1988 - 1997.



Non-parametric tests for trend indicated no evidence of any trend. Subsequent regression modeling fit an increasing-trend model. Table 15 presents the statistics for the non-parametric results while Table 16 contains the regression findings.

Table 15 Non-parametric results for potassium in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	1.215	0.224
MSK	0.976	0.329
Sen	0.016	NA
LCL	-0.012	NA
UCL	0.038	NA

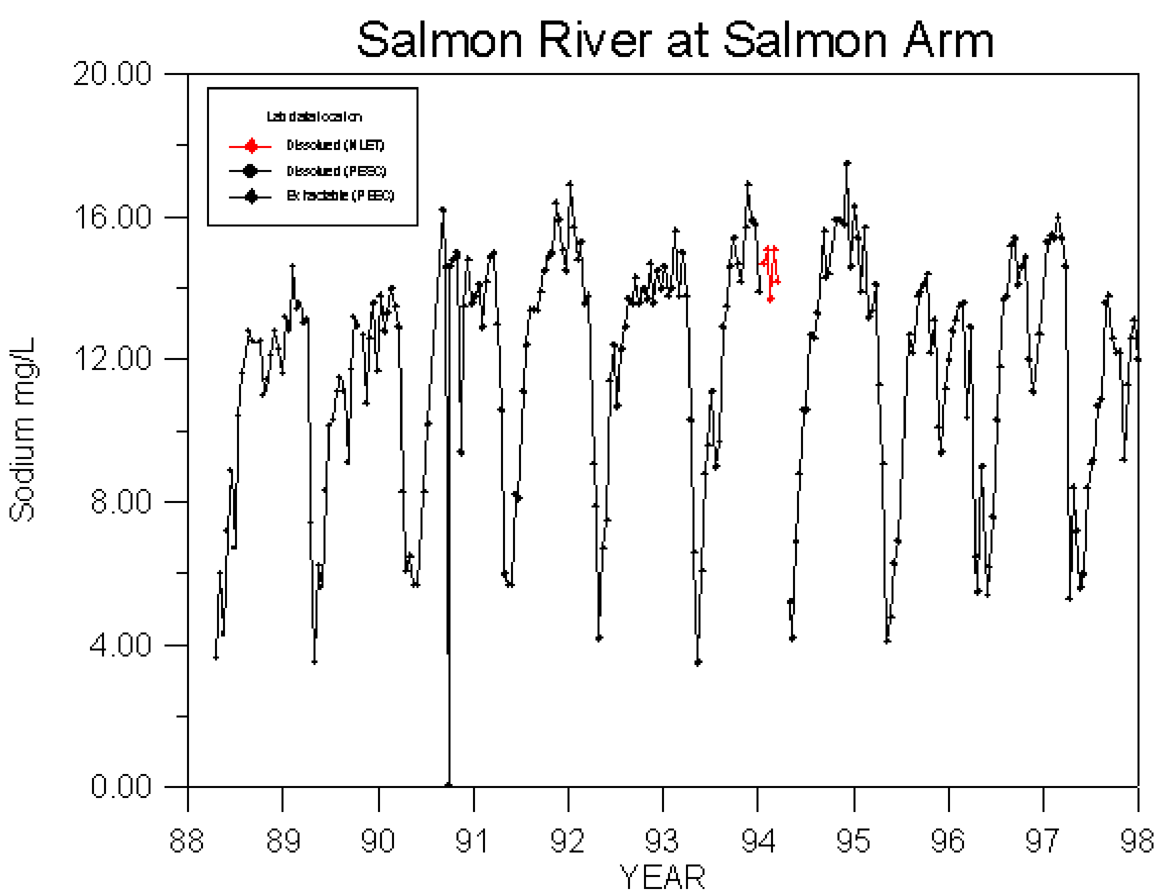
Table 16 Regression modeling results for potassium in the Salmon River at Salmon Arm. Note that flow data only were log transformed before modeling.

Coefficient	Statistic	p-value
b_0	3.3145	0.0000
b_1	-0.5442	0.0000
b_2	0.0754	0.0000
b_3	----	----
a1	-0.0523	0.2503
a2	0.2012	0.0001
r^2	0.60	NA

Sodium

Sodium data collected from 1988 through 1997 were used for statistical analyses (Figure 11).

Figure 11 Time series plot of sodium in the Salmon River at Salmon Arm, 1988 - 1997.



Non-parametric tests for trend indicated no evidence of any trend. Subsequent regression modeling fit an increasing-trend model. Table 17 presents the statistics for the non-parametric results while Table 18 contains the regression findings.

Table 17 Non-parametric results for sodium in the Salmon River at Salmon Arm.

	Statistic	*P-value*
SK	0.828	0.408
MSK	0.511	0.609
Sen	0.049	NA
LCL	-0.05	NA
UCL	0.145	NA

Table 18 Regression modeling results for sodium in the Salmon River at Salmon Arm. Note that flow data only were log transformed before modeling.

Coefficient	*Statistic*	*p-value*
b_0	13.1191	0.0000
b_1	-2.9095	0.0000
b_2	0.325	0.0000
b_3	----	----
a1	-0.5952	0.0003
a2	0.3888	0.0170
r^2	0.73	NA

References

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

Esterby, S.R., A.H. El-Shaarawi, L.C. Keeler and H.O. Block, 1989. Testing for trend in water quality monitoring data. National Water Research Institute, Contribution No. 91-02.

Gilbert, R.O., 1987. Statistical methods for environmental pollution monitoring. Van Nostrand-Reinhold, New York.

Helsel, D.R. and R.M., Hirsch, 1995. Statistical methods in water resources. 2^{nd} edition Elsevier Science, The Netherlands.

Hirsh, R.M. and J.R, Slack, 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20: 727-732.

Hirsh, R.M., J.R, Slack, and R. Smith, 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18 (1): 107:121.

Sen, P.K., 1968. On a class of aligned rank order tests in two-way layouts. Annals of Mathematical Statistics 39, 1115-1124.

Shaw, P. and A.H. El-Shaarawi, 1995. Patterns in water quality at selected stations in the Fraser River Basin (1985-1991). DOE FRAP 1995-20.

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

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

Yu, Y.S. and S. Zou, 1993. Research trends of principal components to trends of water-quality constituents. Water Resources Bulletin 29(5): 797-806.