Trend Analysis of Annual Spring-Overturn Total Phosphorus in 8 Small Lakes in Southern Vancouver Island, British Columbia

June, 1998

Prepared by
Robin Regnier
Central Limit Statistical Consulting

Introduction

The BC Ministry of Environment, Lands and Parks has recently reviewed long-term water quality time series to identify any potential trends visually. Included in these time series were the data for eight lakes located on southern Vancouver Island, British Columbia. Of particular interest for these eight lakes were the apparent trends observed in the concentrations of total phosphorus recorded at the spring overturn.

This report summarizes the statistical analyses of these apparent trends of total phosphorus concentrations at these eight lake monitoring sites which were:

· Cusheon Lake;
· Elk Lake;
· Glen Lake;
· Langford Lake;
· Prospect Lake;
· Quamichan Lake;
· St. Mary Lake; and
· Shawnigan Lake

Background information on each of these sites can be obtained by contacting the BC Ministry of Environment, Lands and Parks.

Methods

Exploratory Data Analysis (EDA)

Exploratory data analysis procedures are the `initial look' at a dataset, 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.

Non-Parametric Tests

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. Two different non-parametric tests, the *Kendall Test for Trend* and the *Sen slope estimator* were used to detect and determine magnitudes of any trends in the water quality data.

Kendall Test for Trend

To perform this non-parametric test, Kendall's S statistic is computed from the data (see Millard, 1997, or any good introductory non-parametric statistics text for details). The null hypothesis of no trend is rejected when S is significantly different from zero. Hirsch et al. (1982) note that this test is appropriate even in the presence of missing observations, and censored values.

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) :

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

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 Modeling

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).

Using these methods, many factors can be taken into account for explaining the variation in a water quality constituent over time, factors which include precipitation. By accounting for precipitation, it's influence on the response constituent can be removed, revealing underlying trends.

The regression model used in this report is as follows:

(1) $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2 i + \epsilon_t$

where :

y_t = Observed value of water quality variable in year t;

x_t = precipitation in year t;

ε_t = Error term assumed to follow a normal distribution with mean 0 and variance σ^2. .

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

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

Consultation with the B.C. Ministry of Environment, Lands and Parks concluded that for regression modeling, total precipitation data collected in the months October to March, preceding the spring overturn measurements of total phosphorus concentrations, should be used as an explanatory variable (Pommen, 1998). Thus rainfall data from the Victoria International Airport were used for these lakes located in southern Vancouver Island (Figure 1).

Initial data analyses indicated that total phosphorus data at all eight lakes were best represented by a lognormal distribution. Subsequently, log transformations were performed on all total phosphorus data before regression modeling took place.

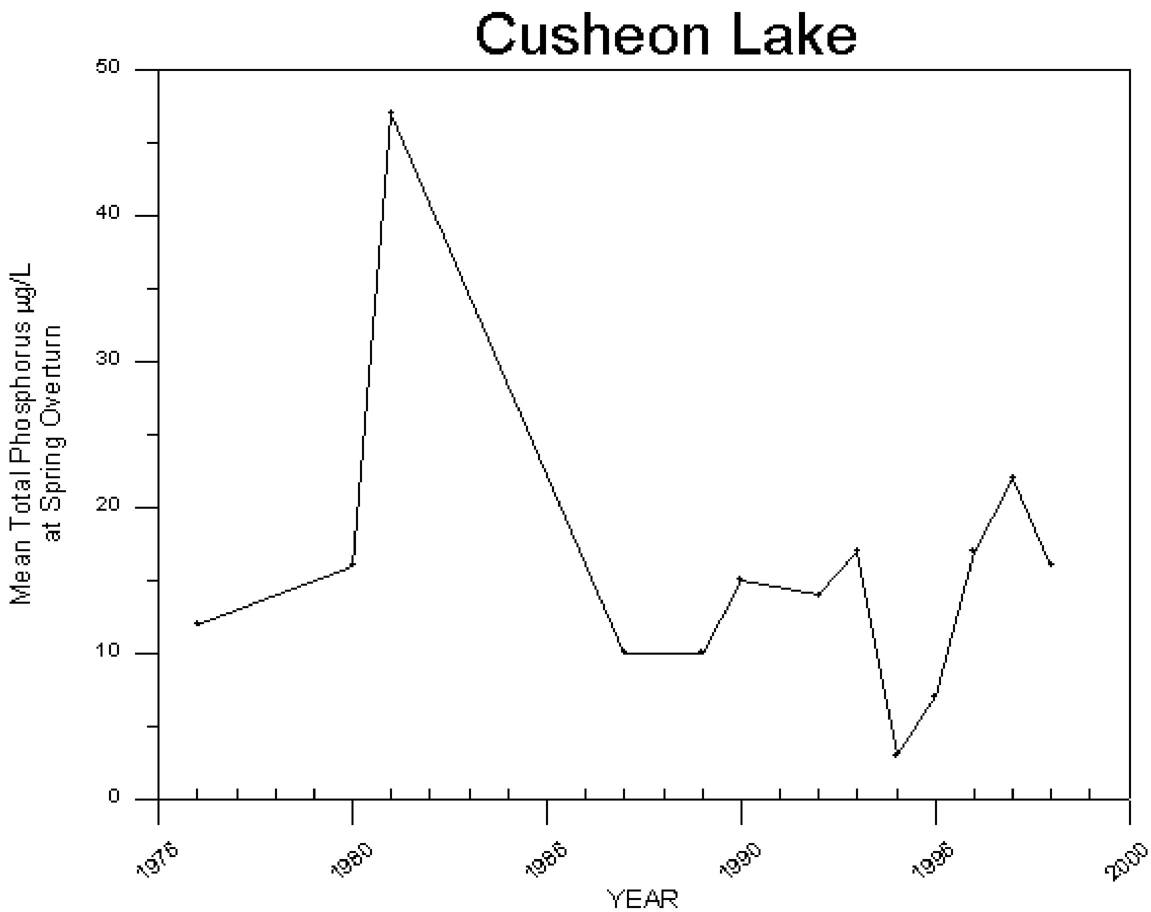
The statistical results were tabulated and can be found in the appendix of this report.

Figure 1 Time series plot of total precipitation at Victoria International Airport recorded from October to March, 1974 - 1998.

Cusheon Lake

The data set for total phosphorus at Cusheon Lake during spring overturns spanned 22 years, from 1976 to 1998 (Figure 2).

Figure 2 Time series plot of total phosphorus recorded at the spring overturn in Cusheon Lake, 1976 - 1998.

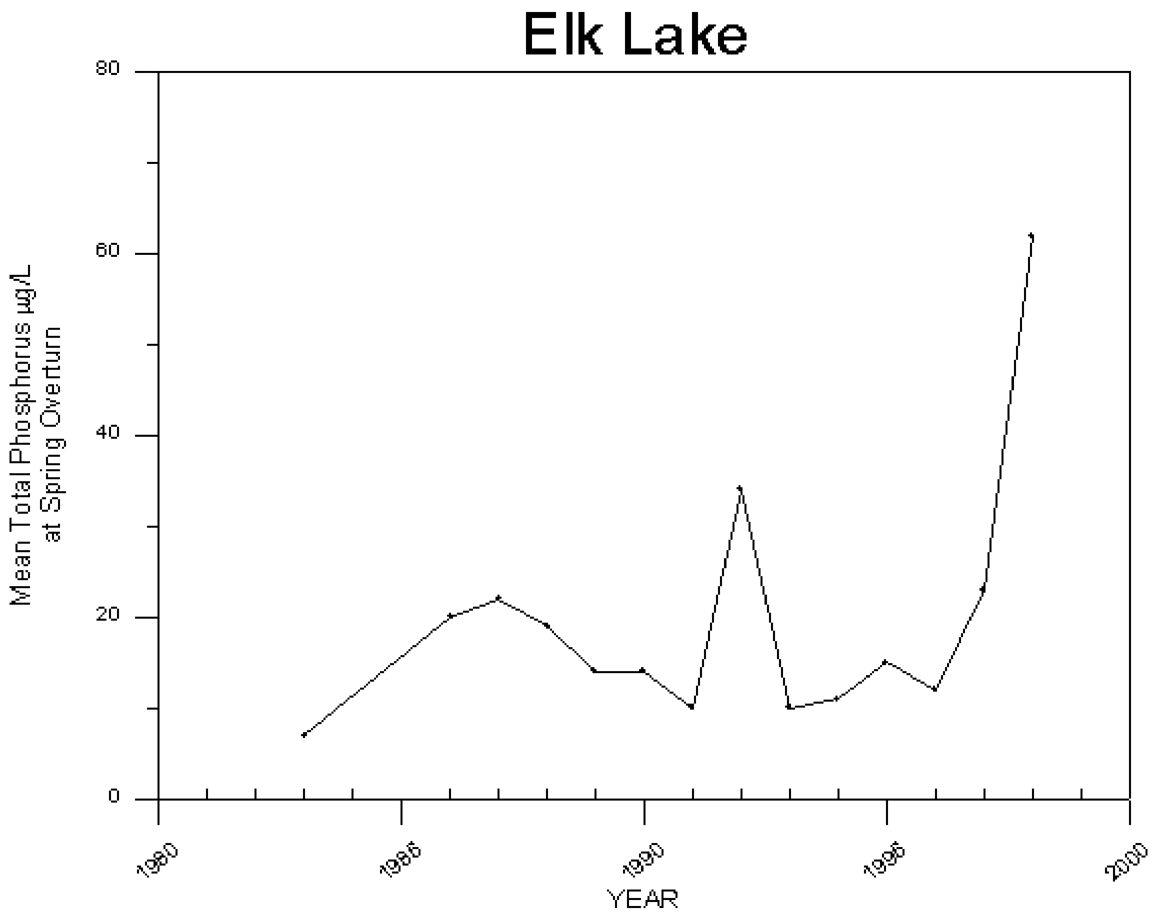


Non-parametric and regression modeling indicated that there was no evidence of any trends in total phosphorus during the spring overturn in Cusheon Lake. Residual diagnostic checking during the regression modeling process indicated that no models could be used to explain the data. The regression model for this site presented in the Appendix was chosen for the simple reason that it came `closest' to adhering to regression assumptions.

Elk Lake

The data set for total phosphorus at Elk Lake during spring overturns spanned 15 years, from 1983 to 1998 (Figure 3).

Figure 3 Time series plot of total phosphorus recorded at the spring overturn in Elk Lake, 1983 - 1998.

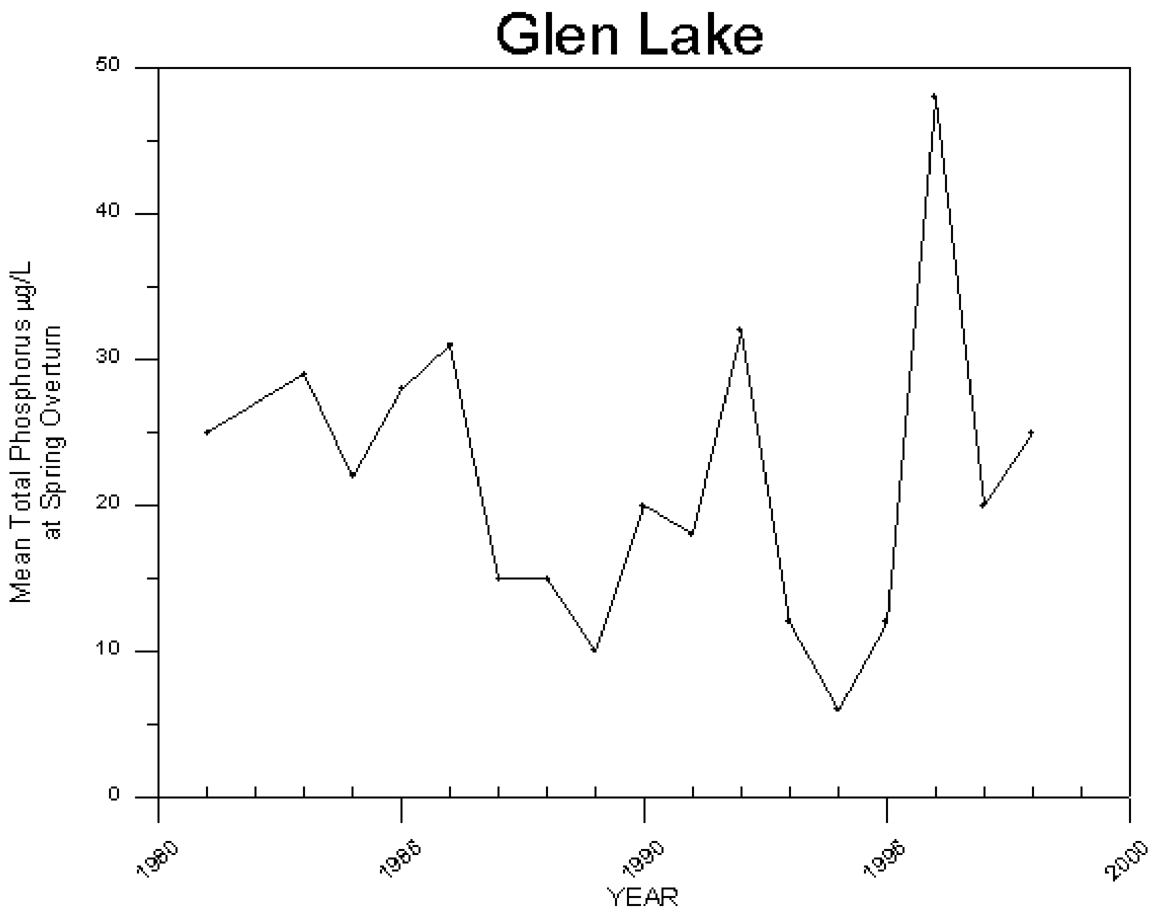


Non-parametric and regression modeling indicated that there was no evidence of any trends in total phosphorus during the spring overturn in Elk Lake. Residual diagnostic checking during the regression modeling process indicated that no models could be used to explain the data. The regression model for this site presented in the Appendix was chosen for the simple reason that it came `closest' to adhering to regression assumptions.

Glen Lake

The data set for total phosphorus at Glen Lake during spring overturns spanned 17 years, from 1981 to 1998 (Figure 4).

Figure 4 Time series plot of total phosphorus recorded at the spring overturn in Glen Lake, 1981 - 1998.

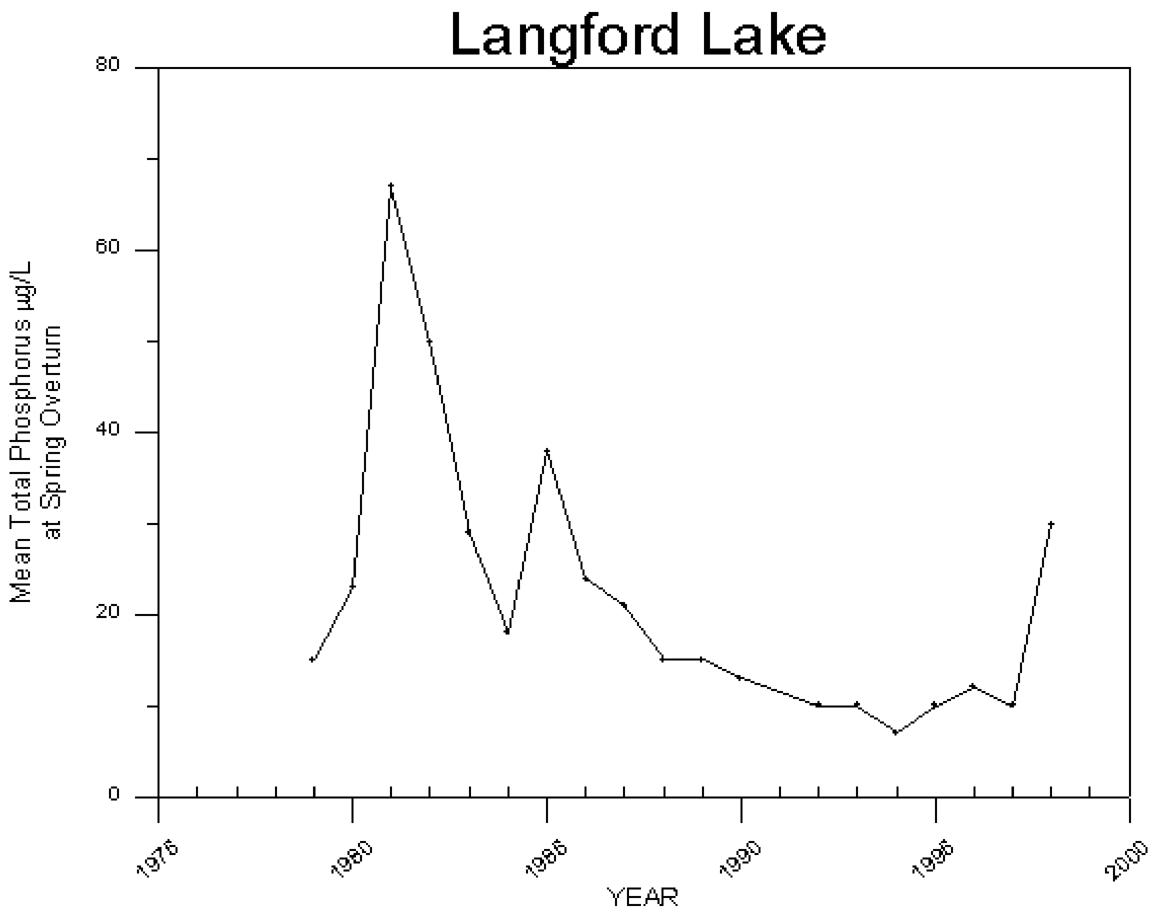


Non-parametric and regression modeling indicated that there was no evidence of any trends in total phosphorus during the spring overturn in Glen Lake. Regression modeling indicated that precipitation, alone, best explained the variation in the available total phosphorus data. However, further analyses using both non-parametric and regression modeling indicated that a decreasing trend existed using data up to 1995. Of interest was that precipitation *did not* help in explaining this abbreviated dataset.

Langford Lake

The data set for total phosphorus at Langford Lake during spring overturns spanned 19 years, from 1979 to 1998 (Figure 5).

Figure 5 Time series plot of total phosphorus recorded at the spring overturn in Langford Lake, 1979 - 1998.

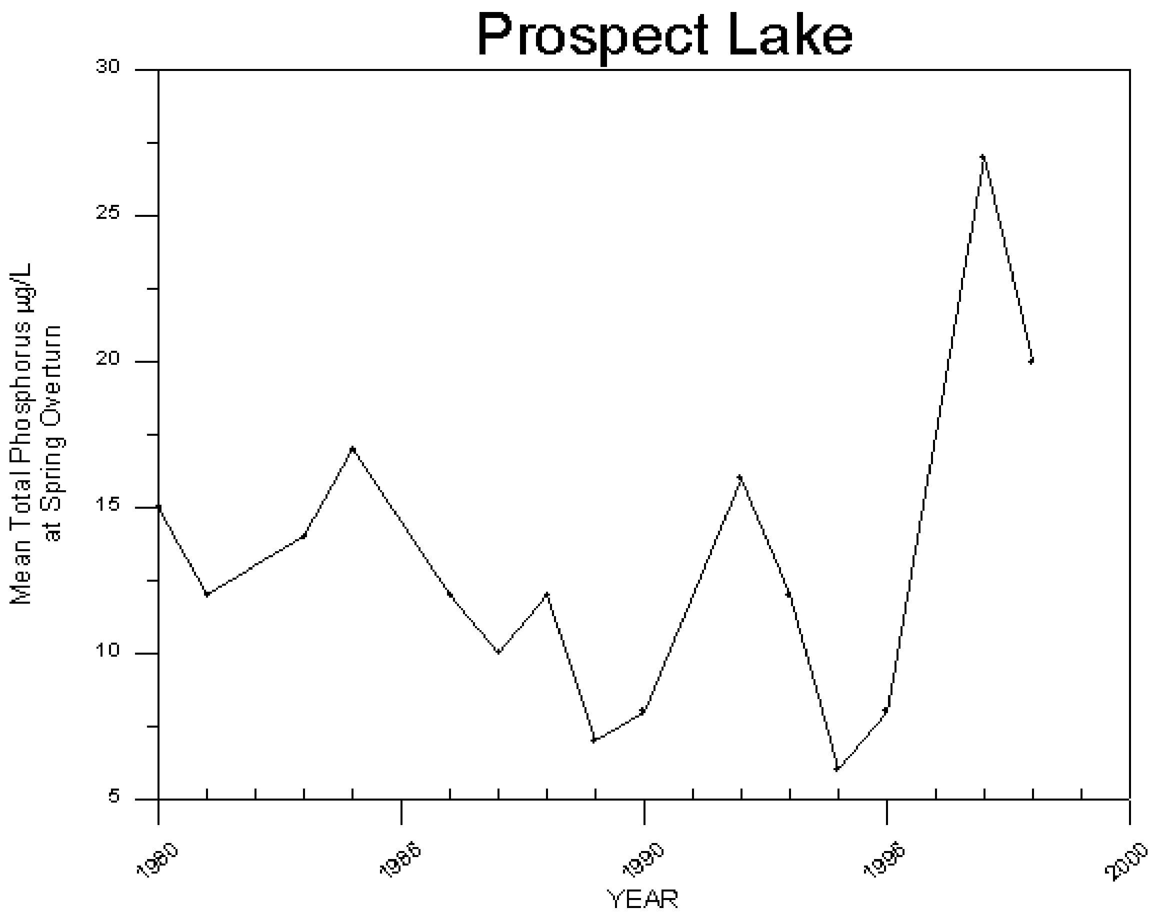


Non-parametric and regression modeling indicated that a decreasing trend was evident in total phosphorus data. The regression model fit revealed that precipitation *did not* help in explaining the total phosphorus data available.

Prospect Lake

The data set for total phosphorus at Prospect Lake during spring overturns spanned 18 years, from 1980 to 1998 (Figure 6).

Figure 6 Time series plot of total phosphorus recorded at the spring overturn in Prospect Lake, 1980 - 1998.

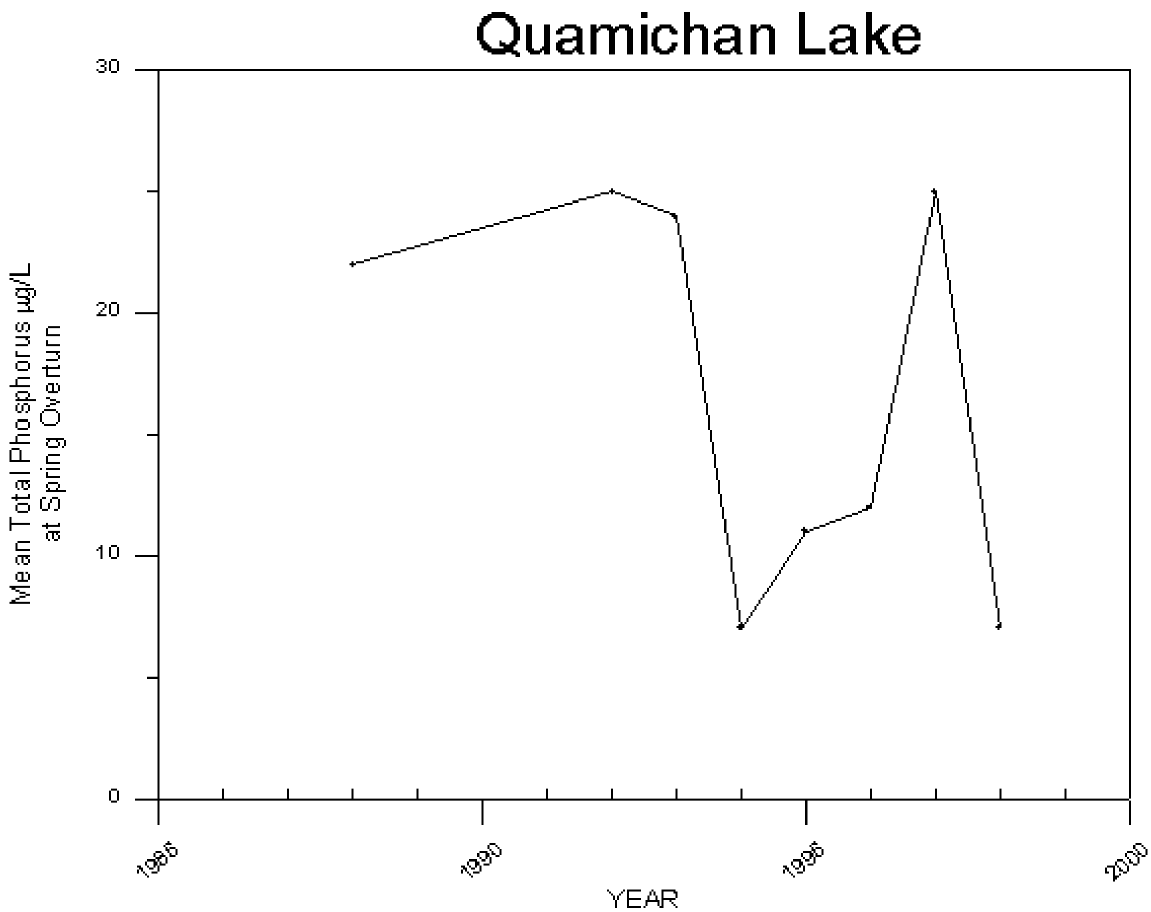


Non-parametric and regression modeling indicated that there was no evidence of any trends in total phosphorus during the spring overturn in Prospect Lake. Regression modeling indicated that precipitation, alone, best explained the variation in the available total phosphorus data. However, like results mentioned previously for Glen Lake, further analyses using both non-parametric and regression modeling indicated that a decreasing trend existed using data up to 1995. Of interest was that precipitation *did not* help in explaining this abbreviated dataset.

Quamichan Lake

The data set for total phosphorus at Quamichan Lake during spring overturns spanned 10 years, from 1988 to 1998 (Figure 7).

Figure 7 Time series plot of total phosphorus recorded at the spring overturn in Quamichan Lake, 1988 - 1998.

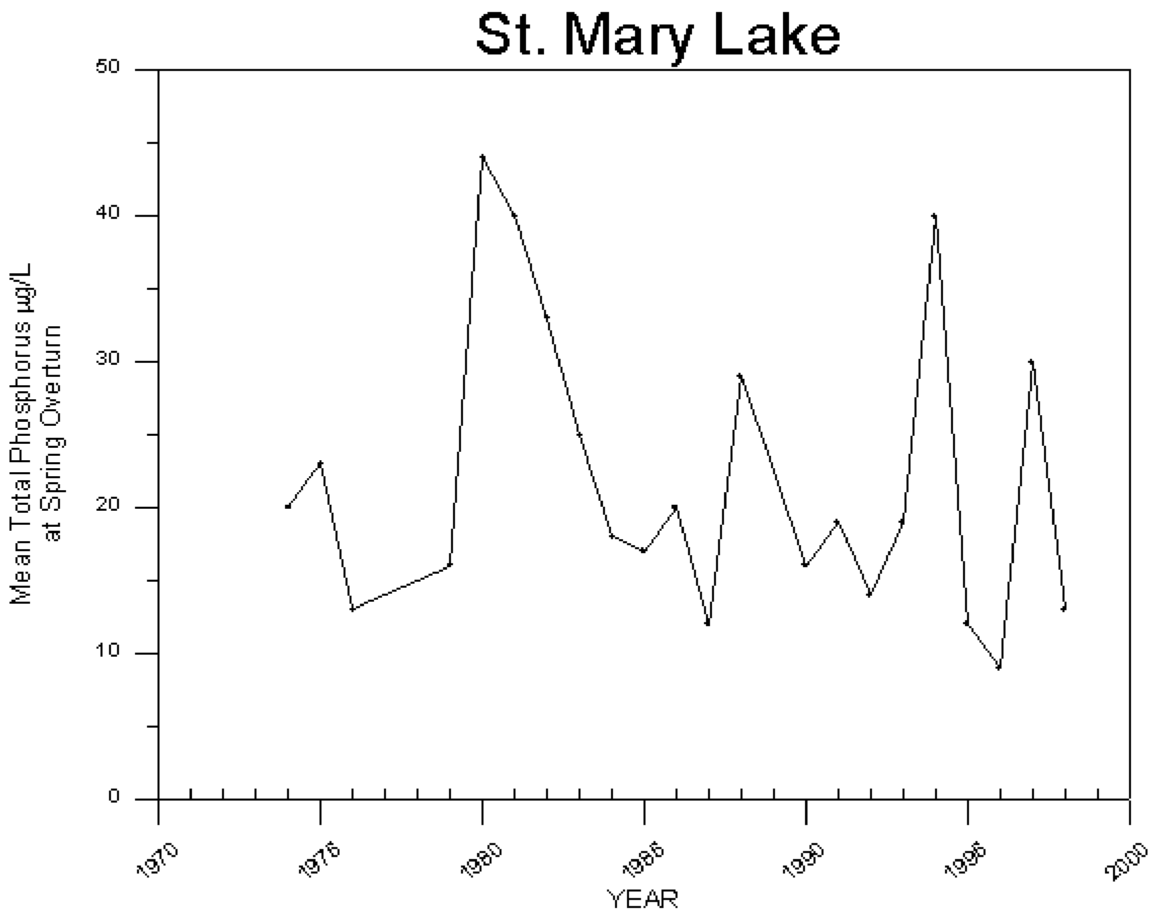


Non-parametric tests indicated that there was no evidence of any trends in total phosphorus during the spring overturn in Quamichan Lake. Residual diagnostic checking during the regression modeling process indicated that no models could be used to explain the data. Subsequently, no regression models for this site are presented in the Appendix.

St. Mary Lake

The data set for total phosphorus at St. Mary Lake during spring overturns spanned 24 years, from 1974 to 1998 (Figure 8).

Figure 8 Time series plot of total phosphorus recorded at the spring overturn in St. Mary Lake, 1974 - 1998.

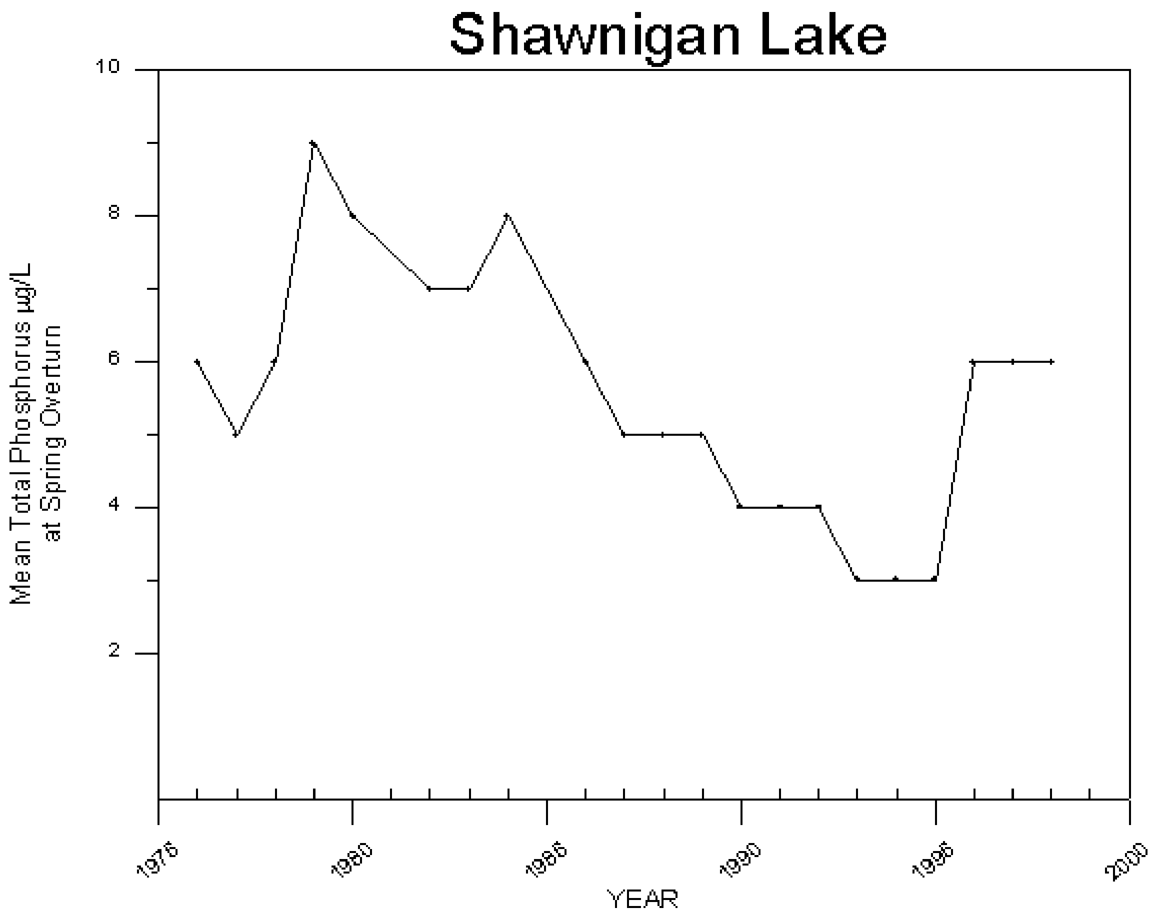


Non-parametric and regression modeling indicated that there was no evidence of any trends in total phosphorus during the spring overturn in St. Mary Lake. Residual diagnostic checking during the regression modeling process indicated that a model without using precipitation data could be used to explain the available total phosphorus data.

Shawnigan Lake

The data set for total phosphorus at Shawnigan Lake during spring overturns spanned 22 years, from 1976 to 1998 (Figure 9).

Figure 9 Time series plot of total phosphorus recorded at the spring overturn in Shawnigan Lake, 1976 - 1998.



Non-parametric and regression modeling indicated that a decreasing trend was evident in total phosphorus data. The regression model fit revealed that precipitation *did not* help in explaining the total phosphorus data available.

Summary

Trend analyses on total phosphorus data collected at the eight lakes during spring overturns revealed two items of note:

· Total precipitation data collected from October to March were not a good explanatory variable in helping to explain the variation of total phosphorus data at all but two of the eight lakes, although including data only up to 1995 at these two sites apparently removed the explanatory significance of the precipitation data; and

· Decreasing trends were evident in the available data at two of the lakes using all available data (Langford and Shawnigan Lakes), while abbreviated data for Glen Lake and Prospect Lake also revealed decreasing trends.

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 (*2nd ed*.). Statistical methods in water resources. Elsevier Science B.V., Amsterdam, 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.

Millard, S.P., 1997. EnvironmentalStats for S-Plus User's Manual, Version 1.0. Probability, Statistics & Information, Seattle, WA.

Pommen, L., 1998. Personal communication.

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

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.

Appendices

Appendix 1

Non-parametric results

	Cusheon		Elk		Glen[1]		Glen[2]		Langford	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
KTT	0.368	0.713	0.769	0.442	-0.703	0.482	-1.977	0.048	-2.996	0.003
SSE	NA	NA	NA	NA	NA	NA	-1.375	NA	-1.1	NA
LCL	NA	NA	NA	NA	NA	NA	-2.188	NA	-2.033	NA
UCL	NA	NA	NA	NA	NA	NA	0	NA	-0.429	NA

	Prospect[1]		Prospect[2]		Quamichan		St. Mary		Shawnigan	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
KTT	-0.351	0.726	-1.988	0.047	-0.754	0.451	-1.47	0.142	-2.53	0.011
SSE	NA	NA	-0.464	NA	NA	NA	NA	NA	-0.154	NA
LCL	NA	NA	-0.848	NA	NA	NA	NA	NA	-0.333	NA
UCL	NA	NA	0	NA	NA	NA	NA	NA	0	NA

KTT - Kendall Test for Trend
SSE - Sen Slope Estimator
LCL - Lower Confidence Limit
UCL - Upper Confidence Limit

1 - analysis over entire data record
2 - analysis over data record up to 1995
NA - not applicable

Regression results

	Cusheon**	Elk**	Glen[1**]	Glen[2**]	Langford*
b_0	NS	NS	NS	120.345	144.29
b_1	NS	NS	1.469	NS	NS
b_2	NS	NS	NS	-0.063	-0.072
b_3	NS	---	---	---	---
r^2	0.274	0.153	0.354	0.516	0.5

	Prospect[1*]	Prospect[2*]	Quamichan	St. Mary*	Shawnigan**
b_0	NS	78.9	---	NS	55.487
b_1	0.002	---	---	---	NS
b_2	NS	-0.038	---	NS	-0.028
b_3	---	---	---	---	---
r^2	0.407	0.033	---	0.066	0.379

1 - analysis over entire data record
2 - analysis over data record up to 1995
* - total phosphorus log transformed
** - total phosphorus and precipitation log transformed
--- - not in regression model
NS - not significant at 5% level