Trend Analyses of Total Phosphorus in 6 Acid-rain-sensitive Lakes located in 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 six lakes in British Columbia that have been classified as "acid-rain-sensitive lakes". Of particular interest for these six lakes were the apparent trends observed in the concentrations of total phosphorus.

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

· Maxwell Lake
· Lizard Lake
· Spectacle Lake
· Old Wolf Lake
· Stocking Lake
· Jacobs Lake

Maxwell, Lizard, Spectacle, and Old Wolf Lakes are located on southern Vancouver Island, Stocking Lake on mid-Vancouver Island, and Jacobs Lake in the greater Vancouver region. 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 and box and whisker plots (Tukey, 1977), both blocked by month and by year, were used in the initial data explorations.

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. Four different non-parametric tests, the *seasonal Kendall's Tau*, the *modified seasonal Kendall's Tau*, the *Van Belle and Hughes* test for trends across time and the *Sen slope estimator* were used to detect and determine magnitudes of any trends in the water quality data.

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).

Van Belle and Hughes Statistic

Van Belle and Hughes (1984) presented a non-parametric test for trend across time. The test statistic utilizes the parameters constructed from the Kendall tests described above. This test essentially indicates whether or not a trend exists. It does not indicate the direction or magnitude of any detected trend.

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:

$$D_{ijk} = \frac{Y_{ij} - Y_{kj}}{i - k} \text{ for j = 1,...,12 ; } 1 \le k < i \le n_j \text{ .}$$

The slope estimate is the median of all D_{ijk} values. Hirsch *et al*. (1982) point out that this estimate is robust against extreme outliers and that since the D_{ijk} values are computed on values that are multiples of 12 months apart, confounding effects of serial correlation are unlikely. 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).

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 and seasonality. By accounting for precipitation 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;

$x_{t_{ji}}$ = precipitation 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 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 $(\beta_3 i^2)$. 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

Consultation with the B.C. Ministry of Environment, Lands and Parks concluded that for regression modeling, monthly averaged precipitation data should be used as an explanatory variable (Pommen, 1998). Rainfall data from the Victoria International Airport, Nanaimo City Yard, and the AES station at

Mission were used for those lakes located in southern Vancouver Island, mid-Vancouver Island, and greater Vancouver, respectively.

Consequently, monthly averaged total phosphorus concentrations were used for the statistical analyses described in the preceding sections.

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

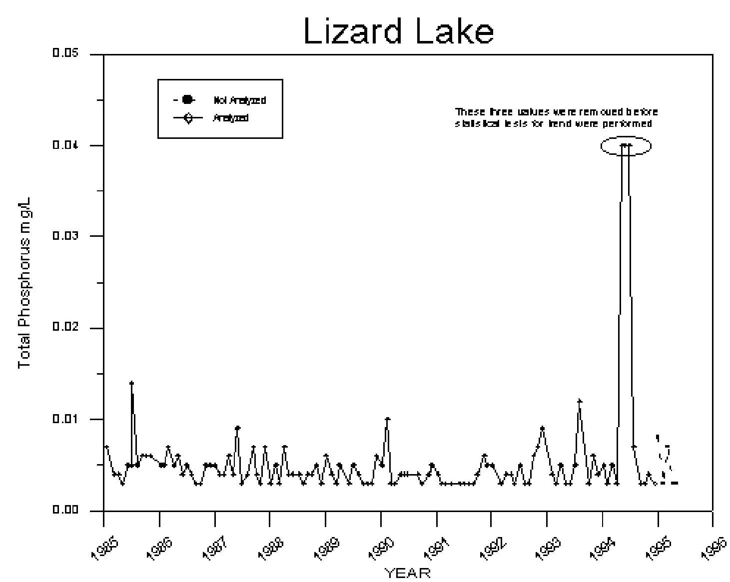
It was also discovered that at five of the lakes, total phosphorus concentrations were correlated better with precipitation recorded several months earlier. The exact number of months precipitation was lagged for each lake are given in the relevant discussions that follow.

The statistical results were tabulated and can be found in the appendix of this report as are the annual and seasonal boxplots for each lake.

Lizard Lake

The data set for total phosphorus at Lizard Lake spanned 10 years, from 1985 to 1995. As Figure 1 shows, data from 1985 to the end of 1994 were selected for further trend analyses. Furthermore, the figure shows three observations that were much different from the bulk of the data and thus were removed from the analysis data. Note that these observations should be investigated as to their possible causes.

Figure1 Time series plot of total phosphorus in Lizard Lake, 1985 - 1995.

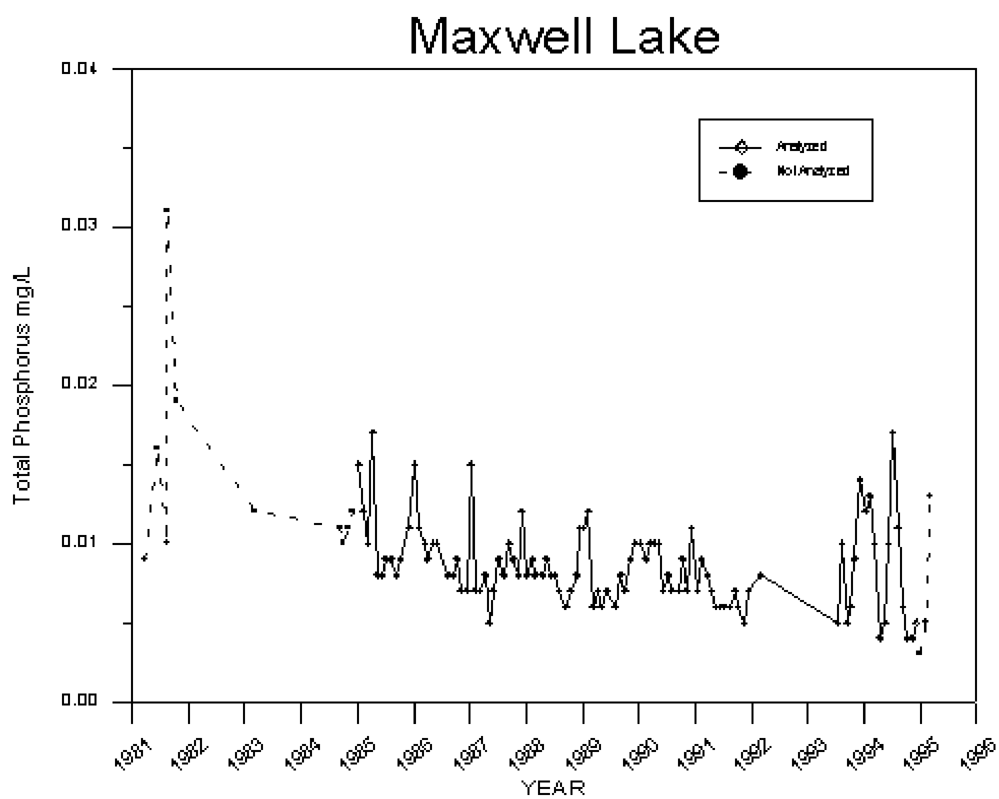


Non-parametric tests for trend indicated that a decreasing linear trend existed over the period of record. Subsequent regression modeling also detected a decreasing trend (although not necessarily of a linear nature due to the log transformations used on the data). Precipitation was insignificant in the model, suggesting that the available rainfall data did not help in providing any insight into the trends detected in the total phosphorus data available. Additional analyses using precipitation lagged 1, 2, 3, and 6 months produced the same results. It is probable that the precipitation at Victoria International Airport is significantly different than that at Lizard Lake. The regression model also indicated that seasonal coefficients were not beneficial in explaining the total phosphorus data at Lizard Lake.

Maxwell Lake

The data set for total phosphorus at Maxwell Lake spanned 15 years, from 1981 to 1995. As Figure 2 shows, data from 1985 to the end of 1994 were selected for further trend analyses.

Figure 2 Time series plot of total phosphorus in Maxwell Lake, 1981 - 1995.

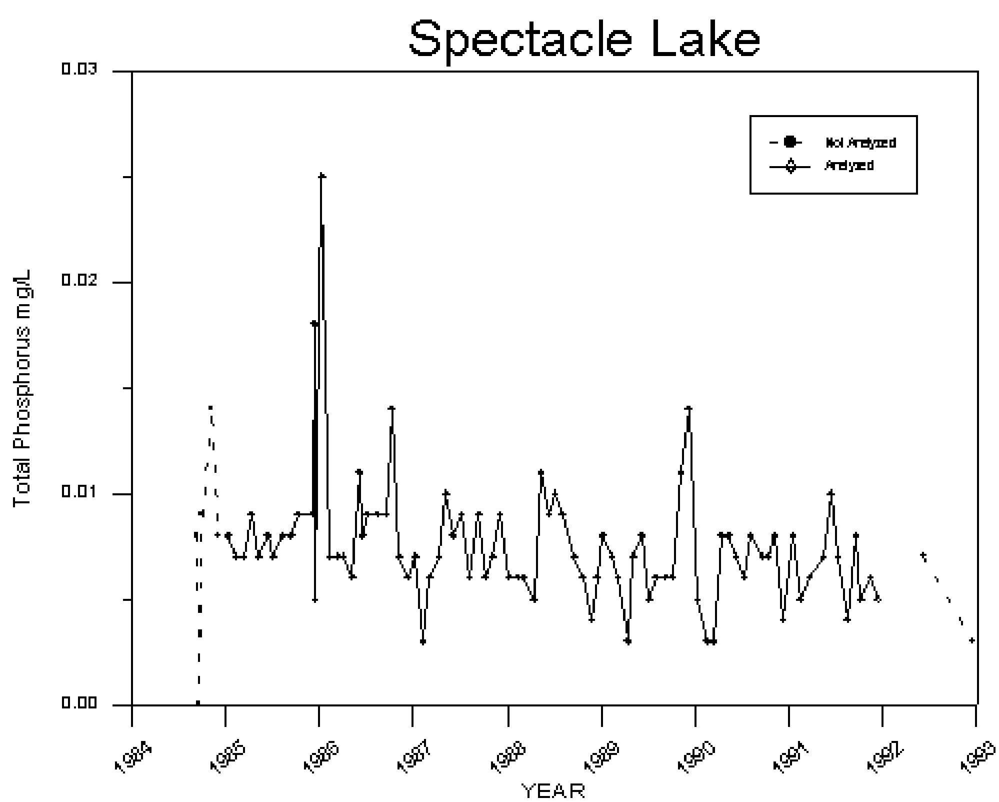


Non-parametric tests for trend indicated a decreasing linear trend existed over the period of record. Subsequent regression modeling also detected a decreasing trend (although not necessarily of a linear nature due to the log transformations used on the data). Precipitation lagged 6 months was significant in the model. Unlike the regression models fit for this constituent at Lizard Lake, the seasonal coefficients were significant in explaining the total phosphorus data at Maxwell Lake.

Spectacle Lake

The data set for total phosphorus at Spectacle Lake spanned 9 years, from 1984 to 1992. As Figure 3 shows, data from 1985 to the end of 1991 were selected for further trend analyses.

Figure 3 Time series plot of total phosphorus in Spectacle Lake, 1984 - 1992.

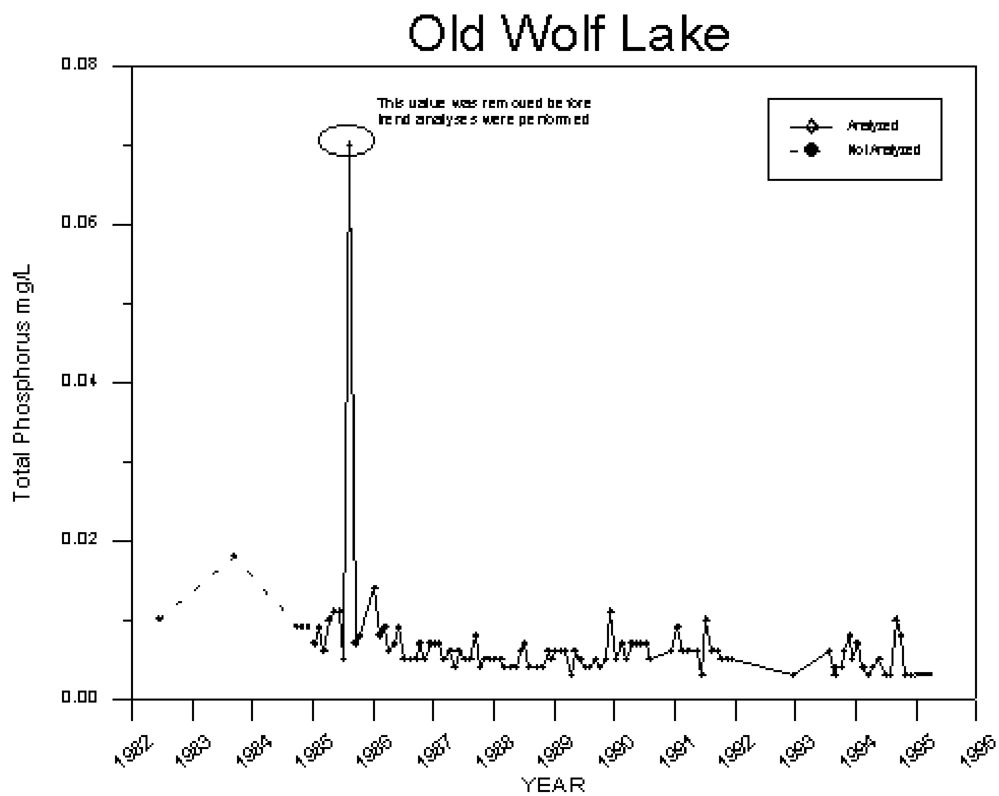


Non-parametric tests for trend indicated a decreasing linear trend existed over the period of record. Subsequent regression modeling also detected a decreasing trend (although not necessarily of a linear nature due to the log transformations used on the data). Precipitation lagged 1 month was significant in the model. The regression model also indicated that seasonal coefficients were not beneficial in explaining the total phosphorus data at Spectacle Lake.

Old Wolf Lake

The data set for total phosphorus at Old Wolf Lake spanned 14 years, from 1982 to 1995. As Figure 4 shows, data from 1985 to the end of 1994 were selected for further trend analyses. Furthermore, the figure shows one observation that was much different from the bulk of the data and thus was removed from the analysis data. Note that this observation should be investigated as to its possible cause.

Figure 4 Time series plot of total phosphorus in Old Wolf Lake, 1982 - 1995.

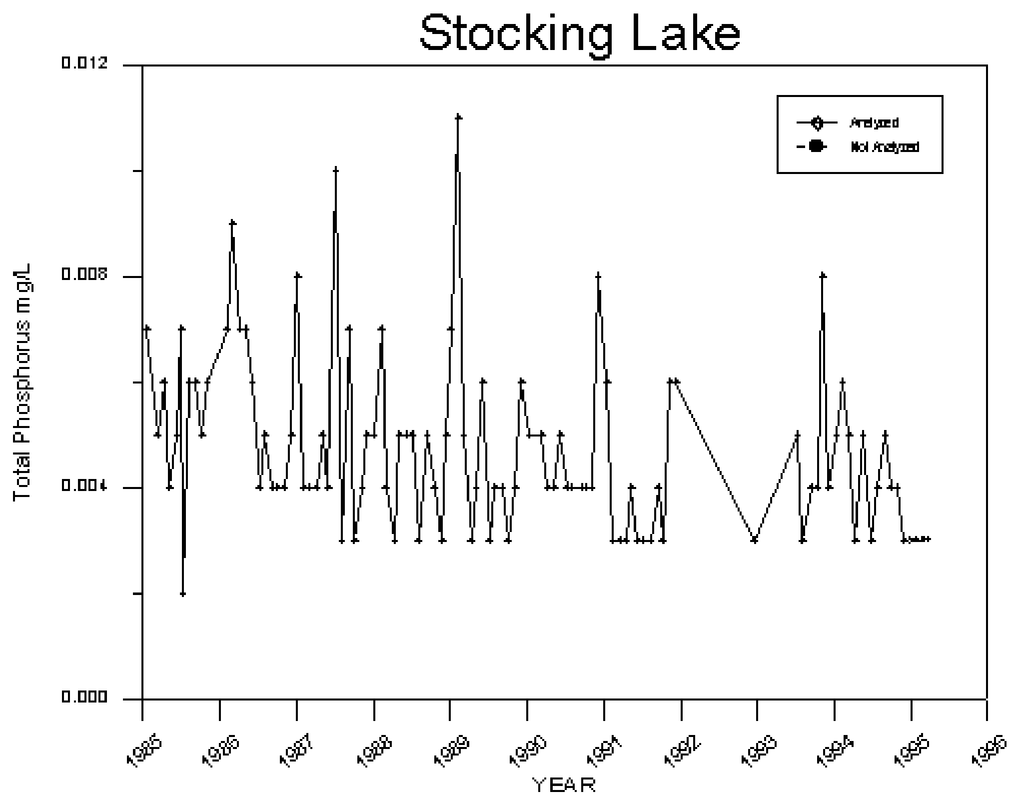


Non-parametric tests for trend indicated a decreasing linear trend existed over the period of record. Subsequent regression modeling at this lake site proved difficult for the full period of record. The non-linear algorithm repeatedly broke down. However, it was found to compute adequately using data for the period 1987 to the end of 1993. The model fit using these data and precipitation lagged 1 month as an explanatory variable found a positive quadratic model (U - shaped). Seasonal coefficients were insignificant in explaining the total phosphorus data at Old Wolf Lake.

Stocking Lake

The data set for total phosphorus at Stocking Lake spanned 11 years, from 1985 to 1995. As Figure 5 shows, data from 1985 to the end of 1994 were selected for further trend analyses.

Figure 5 Time series plot of total phosphorus in Stocking Lake, 1985 - 1995.

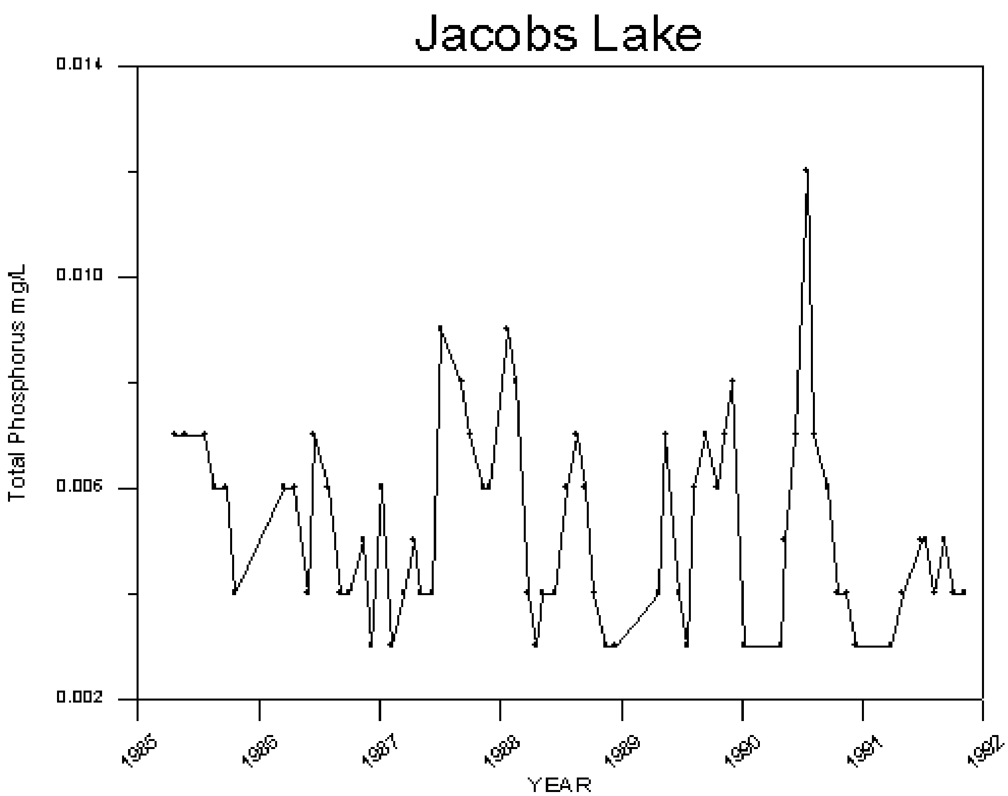


Non-parametric tests for trend indicated a decreasing linear trend existed over the period of record. Subsequent regression modeling also detected a decreasing trend (although not necessarily of a linear nature due to the log transformations used on the data). Precipitation lagged 1 month was significant in the model. The regression model also indicated that seasonal coefficients were not beneficial in explaining the total phosphorus data at Stocking Lake.

Jacobs Lake

The data set for total phosphorus at Jacobs Lake spanned 7 years, from 1985 to 1991. As Figure 5 shows, all available data were selected for further trend analyses.

Figure 6 Time series plot of total phosphorus in Jacobs Lake, 1985 - 1991.



Non-parametric tests for trend and regression modeling indicated that no trends existed in the total phosphorus data. Precipitation was found to be significant in the regression model. Of the six lakes analyzed in this report, this was the only site in which precipitation was not lagged. The regression model also indicated that seasonal coefficients were not beneficial in explaining the total phosphorus data at Jacobs Lake.

Summary

Trend analyses on total phosphorus data collected at the six lakes revealed three things:

· Monthly average precipitation data were a good explanatory variable in helping to explain the variation of total phosphorus data at five of the six lakes, although lagging the monthly average precipitation by one to six months was needed to produce stronger correlations at all lakes but one (Jacobs);
· Seasonality was found to be very weak in total phosphorus data at all but two of the six lakes (Maxwell and Old Wolf); and
· Decreasing trends were evident in the available data at four of the lakes, with only Jacobs Lake total phosphorus data not exhibiting any trends of any kind. Regression modeling on total phosphorus data at Old Wolf Lake was found to be best fit with a positive quadratic model (U - shaped).

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.

Pommen, L., 1998. Water Quality Section, Water Management Branch, BC Environment, Victoria, B.C. 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.

Tukey, J. W., 1977. Exploratory data analysis. Addison-Wesley Publishing.

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

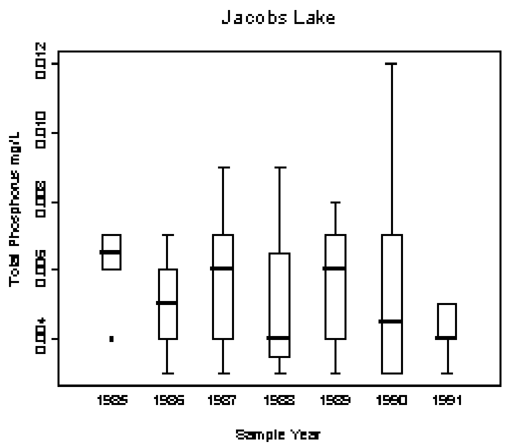
Walker, W., 1991. Water quality trends at inflows to Everglades National Park. Water Resources Bulletin 27 (1): 59-72.

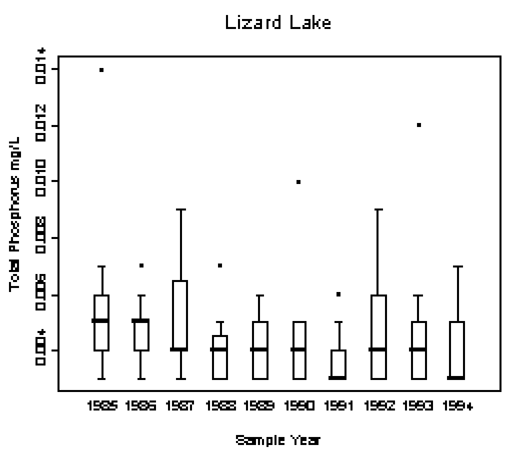
Appendix I

Non-parametric results

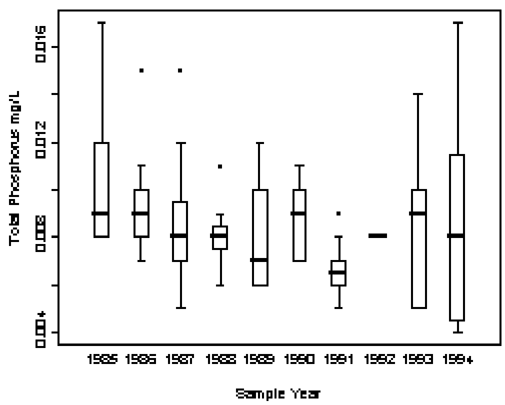
Appendix II

Annual boxplots

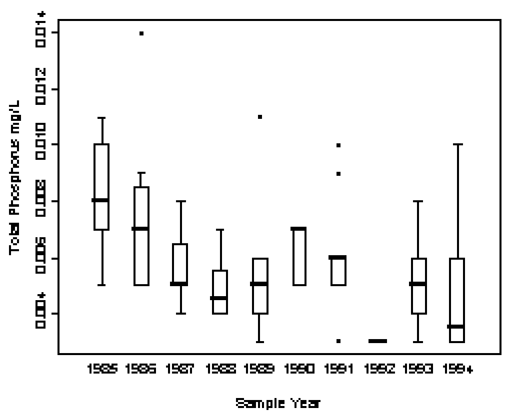


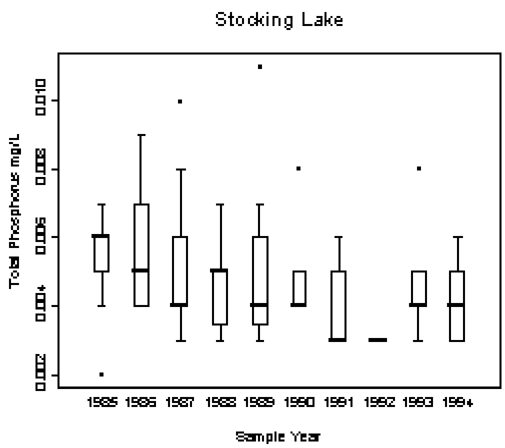


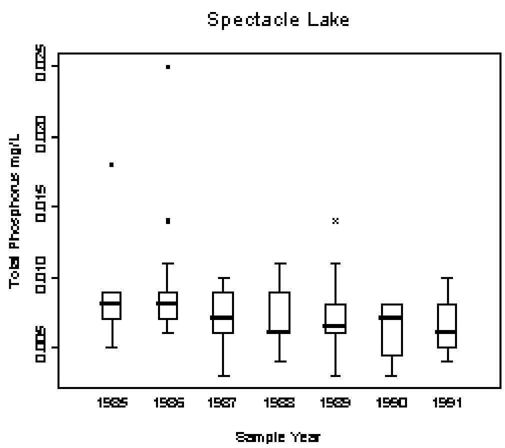
Maxwell Lake



Old Wolf Lake



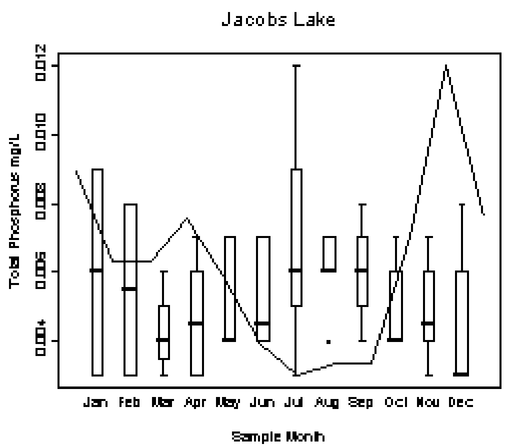


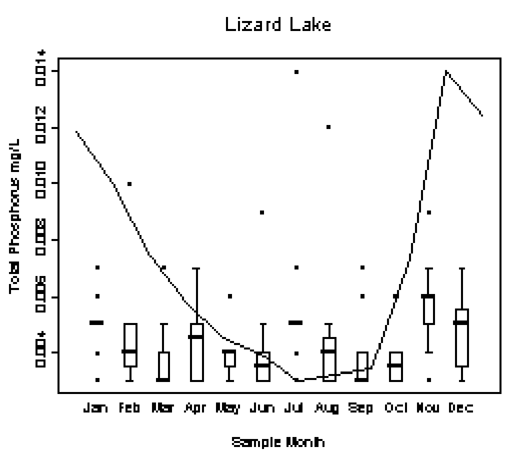


Appendix III

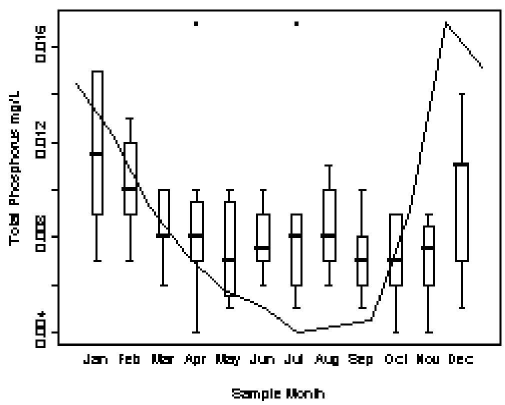
Seasonal Boxplots

note that monthly precipitation is represented by the solid line in each of the graphs

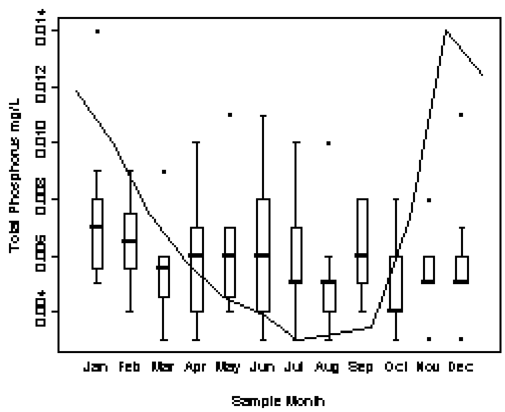

Jacobs Lake
Total Phosphorus mg/L
0.004
0.006
0.008
0.010
0.012
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Sample Month

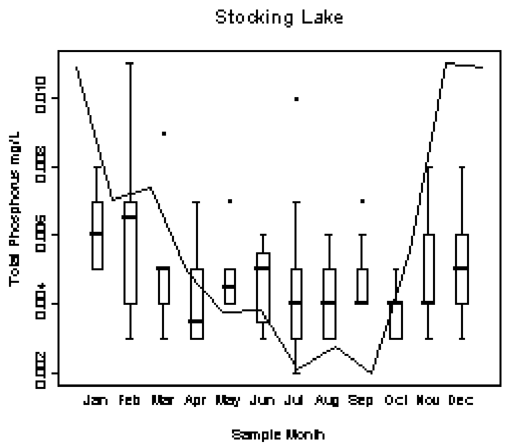

Lizard Lake
Total Phosphorus mg/L
0.004
0.006
0.008
0.010
0.012
0.014
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Sample Month

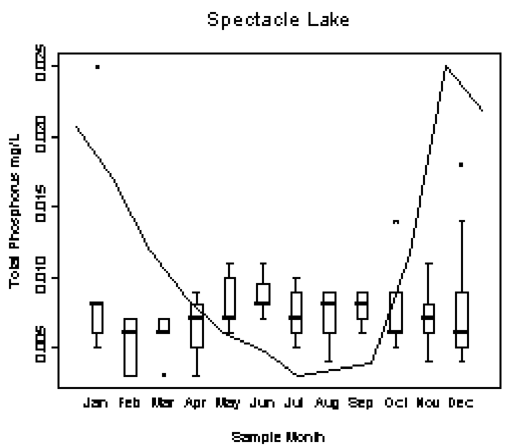
Maxwell Lake



Old Wolf Lake







	Jacobs		**Lizard**		**Maxwell**		**Old Wolf**		**Stocking**		**Spectacle**	
	stat	p-value	stat	p-value	stat	p-value	stat	p-value	stat	p-value	stat	p-value
VBT	2.14165	0.14335	8.53508	0.00348	7.73922	0.00540	6.46211	0.01102	5.04266	0.02473	8.69664	0.00319
SK	-1.47452	0.14034	-2.88407	0.00393	-2.75438	0.00588	-2.51867	0.01178	-2.21717	0.02661	-2.90012	0.00373
MSK	-1.35770	0.17456	-2.05188	0.04018	-1.62295	0.10460	-1.34548	0.17847	-1.60007	0.10958	-2.16359	0.03050
Sen	0.00000	NA	0.00000	NA	-0.00014	NA	-0.00010	NA	0.00000	NA	-0.00023	NA
LCL	-0.00020	NA	-0.00020	NA	-0.00040	NA	-0.00033	NA	-0.00016	NA	-0.00050	NA
UCL	0.00000	NA	0.00000	NA	0.00000	NA	0.00000	NA	0.00000	NA	0.00000	NA
NA -	not applicable											

Regression results

	Jacobs		Lizard		Maxwell		Old Wolf		Stocking		Spectacle	
	stat	p-value	stat	p-value	stat	p-value	stat	p-value	stat	p-value	stat	p-value
b_0	-4.76805	0.00000	-5.27795	0.00000	-4.36487	0.00000	-5.30478	0.00000	-5.48323	0.00000	-4.49958	0.00000
b_1	-0.09483	0.00868	-0.01594	0.55896	-0.08559	0.00021	0.08922	0.00051	0.05920	0.01427	-0.07939	0.01765
b_2	-0.01835	0.26141	-0.02097	0.03420	-0.02084	0.01291	-0.12512	0.01325	-0.02377	0.00782	-0.04299	0.01771
b_3	****	****	****	****	****	****	0.01345	0.01268	****	****	****	****
α_1	0.05722	0.19706	0.01216	0.75153	0.03699	0.26724	0.04273	0.21432	0.00852	0.80046	0.09747	0.09558
α_2	-0.04189	0.37942	0.08133	0.05289	0.07894	0.02501	-0.02456	0.50520	0.02615	0.50608	-0.01519	0.77794
ω	0.14142	NA	0.08080	NA	0.07644	NA	0.02718	NA	0.08304	NA	0.10857	NA
r^2	0.12476	NA	0.07890	NA	0.21148	NA	0.17584	NA	0.11441	NA	0.18253	NA

**** - term not included in model fit

NA - not applicable