Trend Analyses of Total Phosphorus in the Columbia River at Revelstoke 1986 - 1996

June, 1998

Prepared by
Robin Regnier
Central Limit Statistical Consulting

Introduction

The B.C. 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 water quality data from the Columbia River at Revelstoke. Of particular interest were the apparent trends observed in the total phosphorus concentrations time series.

This report summarizes the statistical analyses of the apparent trend of total phosphorus concentrations at the Columbia River monitoring site. Past reports based on a subset of the available data found no trends (Whitfield and McLeod 1992, Holms and Regnier 1995). The Holms and Regnier report did not include discharge rates of the Columbia River for aiding in the analysis of the total phosphorus data.

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 Statistics

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 test for trend

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:

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

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 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 discharge 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 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 (n - 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.

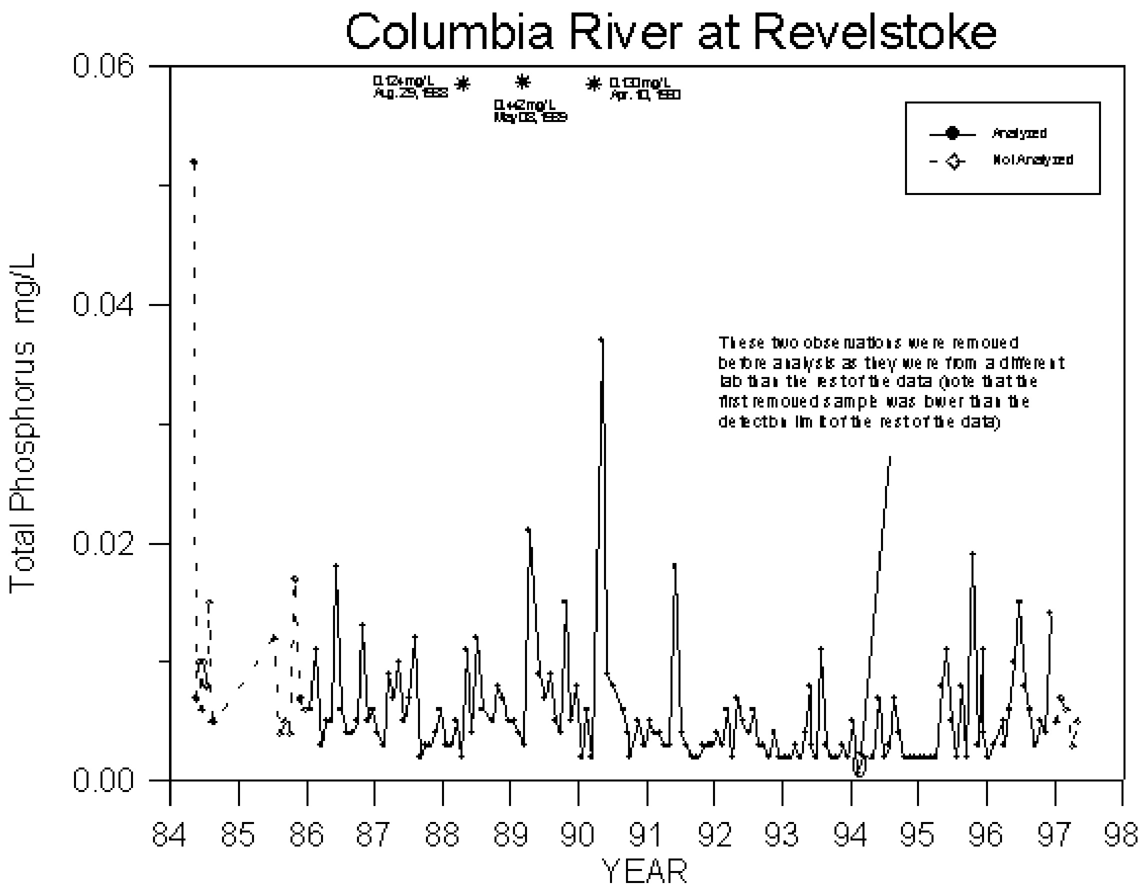
Results and Discussion

Consultation with the B.C. Ministry of Environment, Lands and Parks concluded that for regression modeling, outflow data collected at the Revelstoke Dam, approximately 6 kilometres upstream of the water quality monitoring site, should be used as an explanatory variable (Pommen, 1998). There are no major tributaries between the dam and the water quality monitoring station.

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

The statistical results were tabulated and can be found in the appendix of this report as are the annual and seasonal boxplots for total phosphorus concentrations in the Columbia River at Revelstoke.

Figure 1 Time series plot of total phosphorus concentration in the Columbia River at Revelstoke, 1984 - 1997. Data used for the statistical techniques used were from 1986 to 1996 with three extreme outliers being removed as well as two samples from NLET.



Non-parametric tests indicated that there was strong evidence of a linearly decreasing trend in the total phosphorus concentration data. Note, however, that there does appear to be a recent increasing trend (Figure 1) that may nullify the non-parametric trend finding in the future.

Regression modeling of the data was an extensive process. Initial models fit indicated that the outflows used as an explanatory variable did not adequately explain any variation in the total phosphorus concentrations being modeled. Subsequent attempts at exploring possible dependencies between total phosphorus concentrations and the outflow data (using various lagged outflow datasets) proved fruitless. Hence models to be considered for best fitting the available data were those that did not account for outflow. This analysis indicated that a positive quadratic model best explained the variation in the total phosphorus concentration dataset. Note that the data was transformed on the natural log scale in order to satisfy various normality assumptions required in regression modeling.

Summary

Trend analyses on total phosphorus data collected at the Columbia River at the Revelstoke revealed two items of note:

· Outflow data collected at the Revelstoke Dam were not a good explanatory variable in helping to explain the variation of total phosphorus data. Additional analyses using various lagged values of the outflow data were also inadequate for explanatory purposes; and

· A linear decreasing trend was evident in the total phosphorus data using non-parametric tests. Regression modeling fit a positive quadratic model, suggesting that recent total phosphorus levels have tended to increase.

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.

Holms, G.B. and R.D. Regnier, 1995. Trend assessment of Water Quality over time in the Columbia River [*Draft*]. Water Quality Branch, B.C. Ministry of Environment, Lands and Parks.

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.

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.

Whitfield, P. H., C. MacLeod, 1996. Seasonal and long-term variations in water quality of the Columbia River at Revelstoke, B.C. Northwest Science 70(1): 55-65.

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 I

Non-parametric results

	statistic	p-value
VBT	14.3803	0.000149
SK	-3.7513	0.000176
MSK	-2.1111	0.034761
SSE	-0.0002	NA
LCL	-0.0005	NA
UCL	0	NA

VBT - Van Belle & Hughes test for trend
SK - Seasonal Kendall
MSK - Modified Seasonal Kendall
SEN - Sen Slope Estimator
LCL - Lower Confidence Limit
UCL - Upper Confidence Limit
NA - Not Applicable

Regression modeling results

β_0	β_1	β_2	β_3	α_1	α_2	r^2
8.888*	----	-0.002	9.83e-008	-0.269	0.053**	0.184

* - non-significant at the 5% level but removing the intercept does not make practical sense
** - non-significant at the 5% level but removing only one of the periodic terms produces a model based on an arbitrary shift rather than one determined by the data
---- - term not used in model

Appendix II