Trend Analyses of Copper in the Tsolum River at Farnham, the Tsolum River 500 metres downstream of Murex, Murex Creek at Duncan Main, and Pyrrohotite Creek

February, 1999

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 Tsolum River at Farnham, Murex Creek at Duncan Main, Tsolum River 500 metres downstream of Murex, and Pyrrohotite Creek. Of particular interest were the apparent trends observed in the copper time series.

This report summarizes the statistical analyses of the visually apparent trends seen for copper data at the four sites. Background information for the four water quality sites 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 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. Five different non-parametric tests, the *Kendall Test for Trend*, 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 based on Kendall's Tau that can be applied to time series with indications of non-normality, missing values, censored data, and seasonality. It is applicable to data that exhibit monotonic patterns. For computational details, see Millard (1997).

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. For computational details, see Millard (1997).

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. For computational details, see Millard (1997).

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:

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

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). For computational details, see Millard (1997).

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

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;

$\epsilon_{t_{ji}}$ = 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 $(\beta_3 i^2)$. Residual diagnostics, ANOVA tables, and model goodness-of-fit parameters 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.

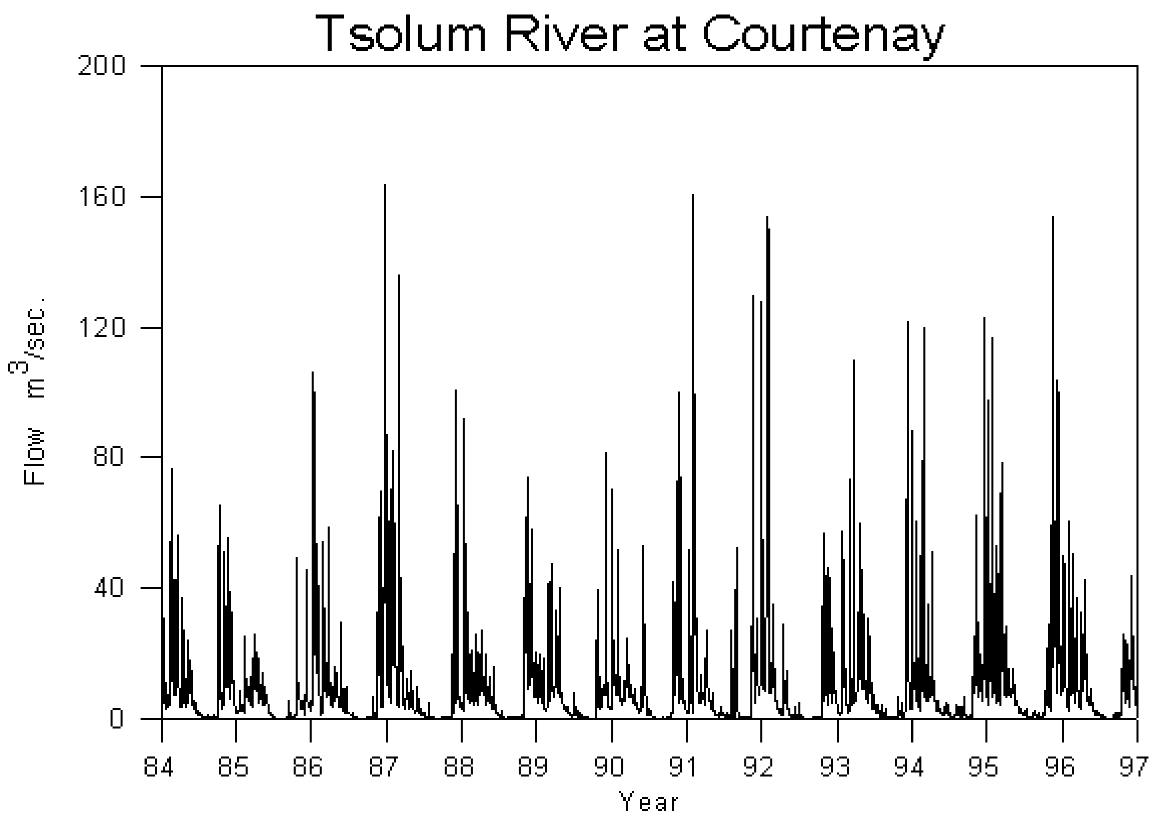
Results and Discussion

To form more complete datasets at all four sites, dissolved and total copper data were merged by using the total cooper data as the base and filling in any missing samples with dissolved copper observations. Initial data analyses on these merged datasets indicated that they were best represented by a lognormal distribution. Subsequently, log transformations were performed on these datasets before regression modeling took place (with the exception of data from Pyrrohotite Creek, which was not analyzed using regression techniques).

The statistical results were tabulated and most can be found in Appendices I-III of this report as are the annual and seasonal boxplots for copper in the two Tsolum River and one Murex Creek sites (Appendix IV).

Consultation with the B.C. Ministry of Environment, Lands, and Parks concluded that for regression modeling, flow data collected from the hydrometric station on the Tsolum River near Courtenay should be used as an explanatory variable (Pommen, 1998). This is the only long-term hydrometric station in the vicinity of the three water quality stations and drains a basin of 258 km^2. Figure 1 is the averaged daily flow at this station.

Figure 1 Time series plot of averaged daily flow rates of the Tsolum River at Courtenay, 1984 - 1996.



Murex Creek at Duncan Main

The data set for copper at Murex Creek spanned portions of 14 years, from 1985 to 1998 (Figure 2). The data for dissolved and total copper were merged to form a more complete time series dataset that was used for the statistical tests employed in this report (Figure 3).

Figure 2 Time series plot of dissolved and total copper in Murex Creek at Duncan Main, 1985 - 1998.

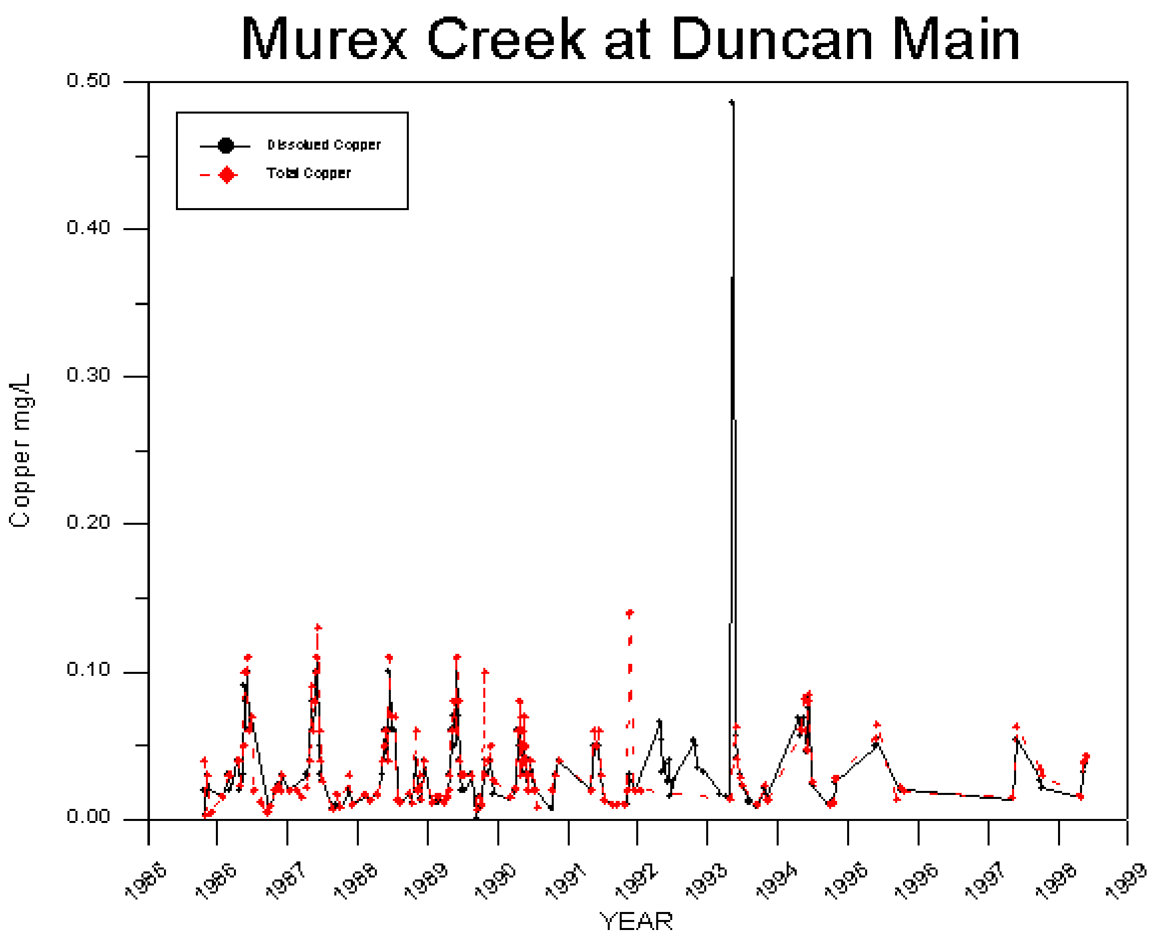
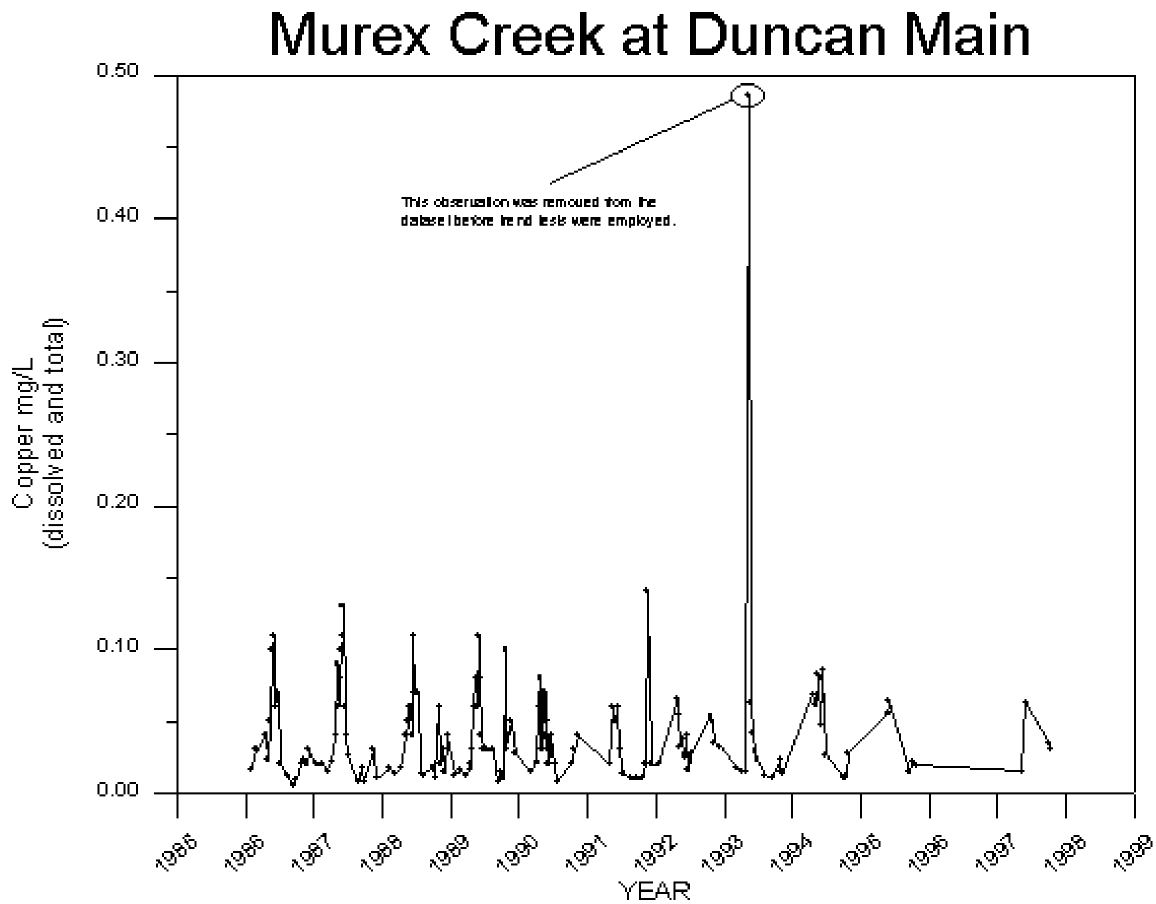


Figure 3 Time series plot of merged dissolved and total copper in Murex Creek at Duncan Main used for statistical trend tests, 1986 - 1997.



Non-parametric tests indicated that there was no evidence in the data that suggested any type of trends during 1986-97 (Appendix I). Subsequent regression modeling also suggested no evidence of any trends and indicated that flow and seasonality accounted for much of the variation in the merged copper dataset at Murex Creek.

Tsolum River at Farnham

The data set for copper in the Tsolum River at Farnham spanned portions of 28 years, from 1971 to 1998 (Figure 4). The data for dissolved and total copper were merged to form a more complete time series dataset that was used for the statistical tests employed in this report (Figure 5). There were insufficient data outside of the period of 1987-94 to do statistical testing.

Figure 4 Time series plot of dissolved and total copper in the Tsolum River at Farnham, 1971 - 1998.

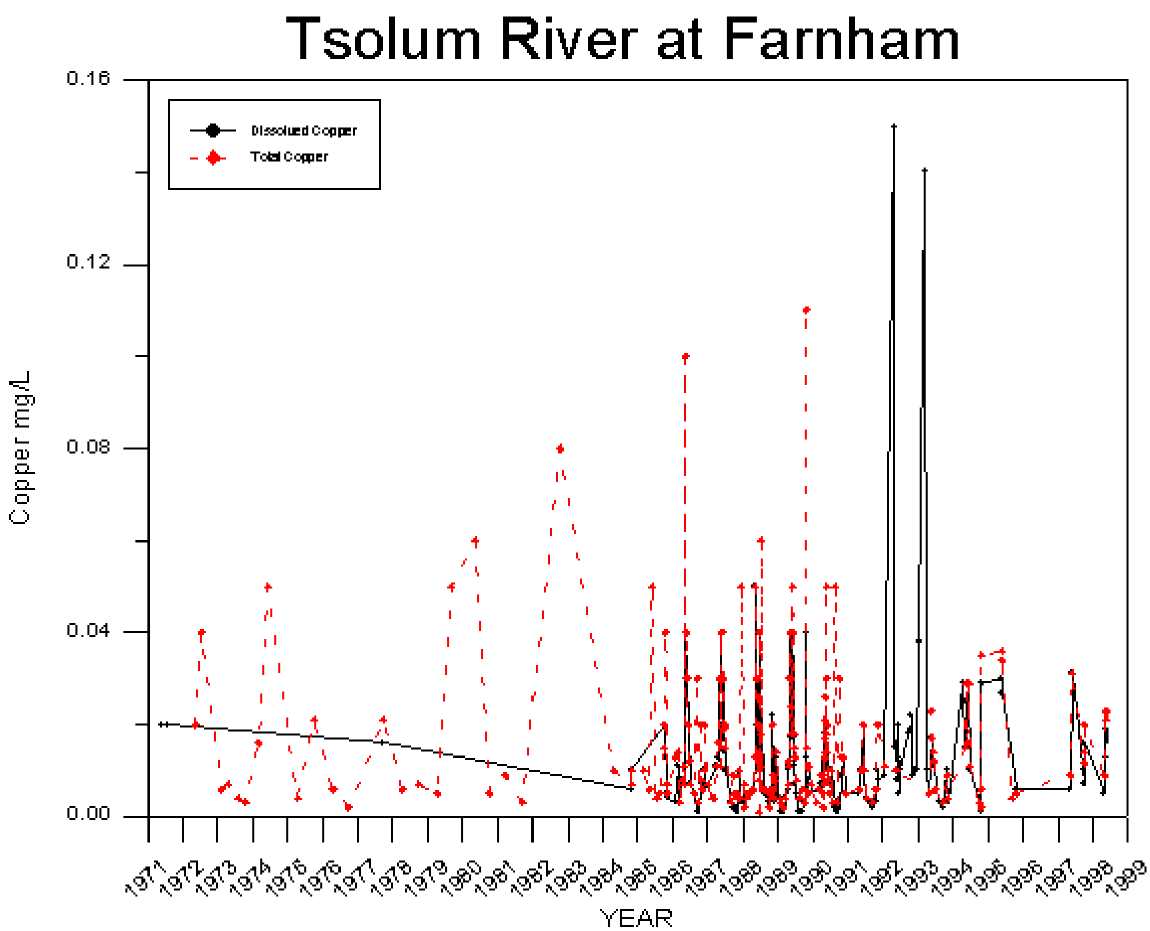
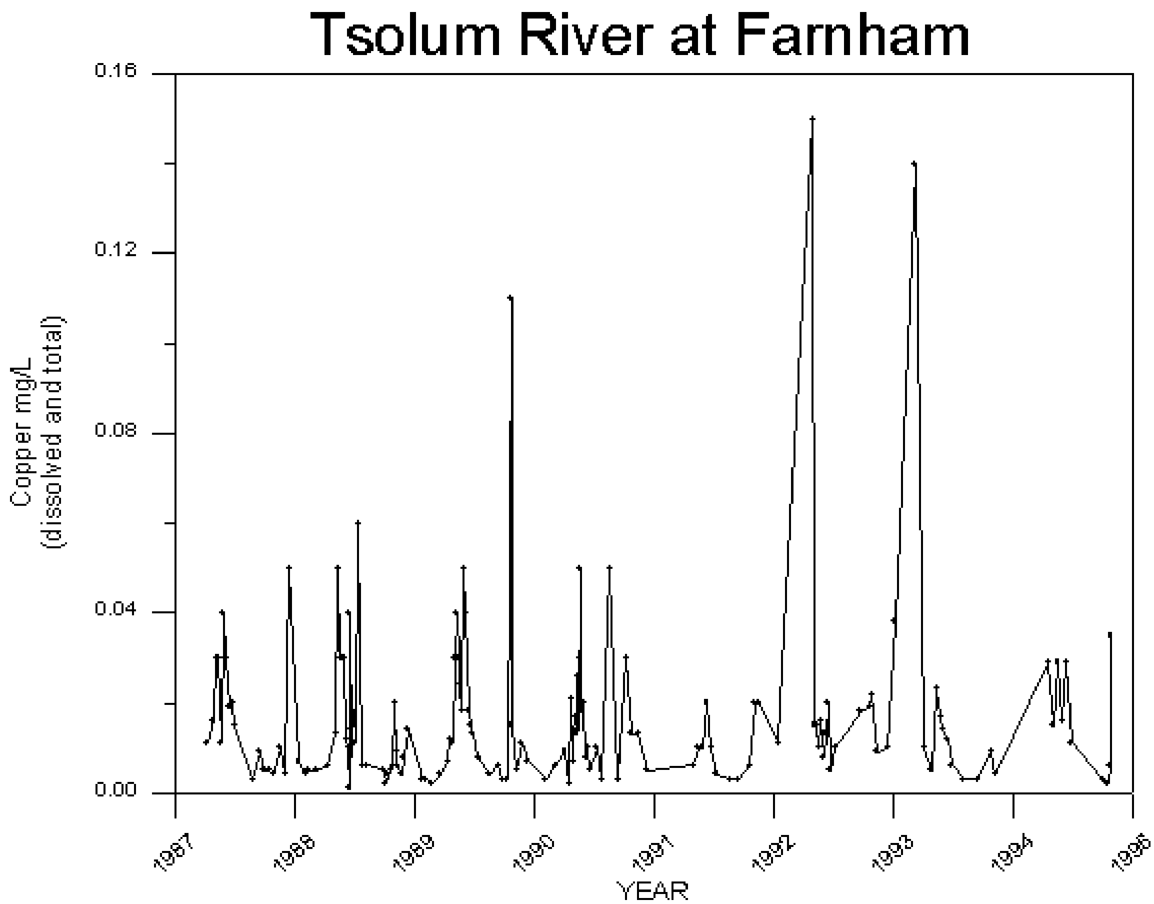


Figure 5 Time series plot of merged dissolved and total copper in the Tsolum River at Farnham used for statistical trend tests, 1987 - 1994.



Non-parametric tests indicated that there was no evidence in the data that suggested any type of trend during 1987-94 (Appendix II). Subsequent regression modeling also suggested no evidence of any trends and indicated that flow and seasonality accounted for much of the variation in the merged copper dataset at the Tsolum River at Farnham.

Tsolum River 500 metres downstream of Murex

The data set for copper in the Tsolum River 500 metres downstream of Murex spanned 10 years, from 1989 to 1998 (Figure 6). The data for dissolved and total copper were merged to form a more complete time series dataset that was used for the statistical tests employed in this report (Figure 7).

Figure 6 Time series plot of dissolved and total copper in the Tsolum River 500 metres downstream of Murex, 1989 - 1998.

Tsolum River 500 m d/s from Murex Creek

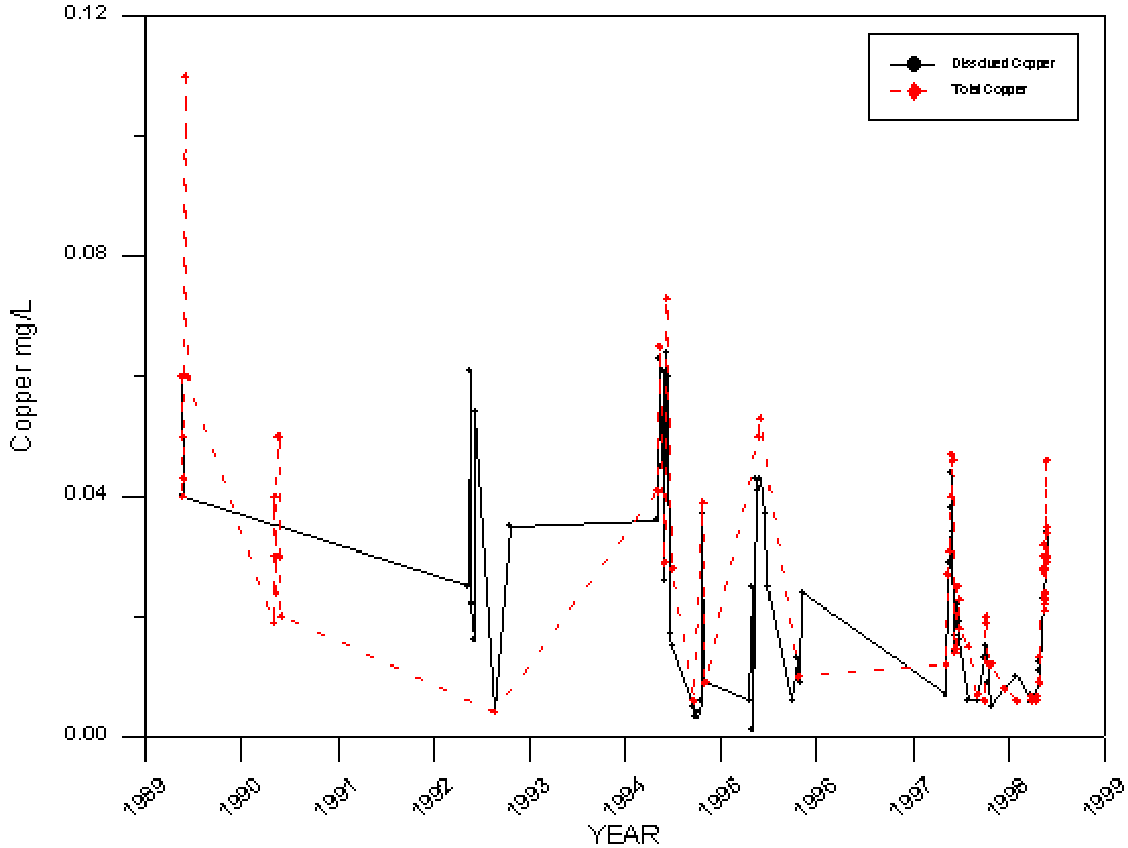
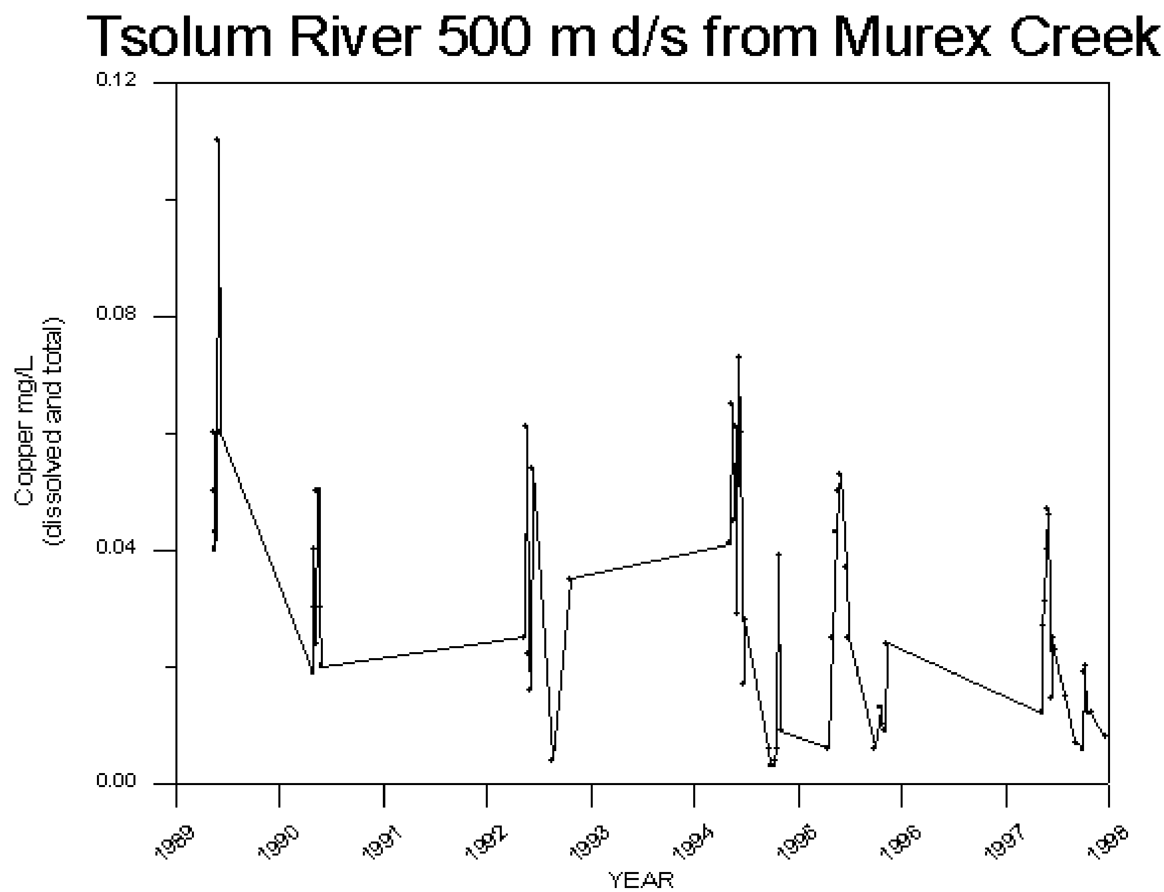


Figure 7 Time series plot of merged dissolved and total copper in the Tsolum River 500 metres downstream of Murex used for statistical trend tests, 1989 - 1997.



The available dataset contained no samples in any year for the months of January, February, or March. Hence, the non-parametric tests based on seasonality did not include these months (*i.e.*, the tests were based upon nine seasons rather than the usual twelve) and only the seasonal Kendall test for trend and the Sen slope estimator were employed. The results indicated that there was no evidence of a trend in the available data during 1989-97 (Appendix III), contrary to what the time series plot portrays visually. The lack of a more complete data set prohibited the use of regression modeling techniques of the type employed in this report.

Pyrrohotite Creek

The data sets for copper at Pyrrohotite Creek spanned portions of 20 years, from 1979 through 1998 (Figure 8). The data for dissolved and total copper were merged to form a more complete time series data set that was used for the statistical tests employed in this report (Figure 9).

Figure 8 Time series plot of dissolved and total copper in Pyrrohotite Creek, 1979 - 1998.

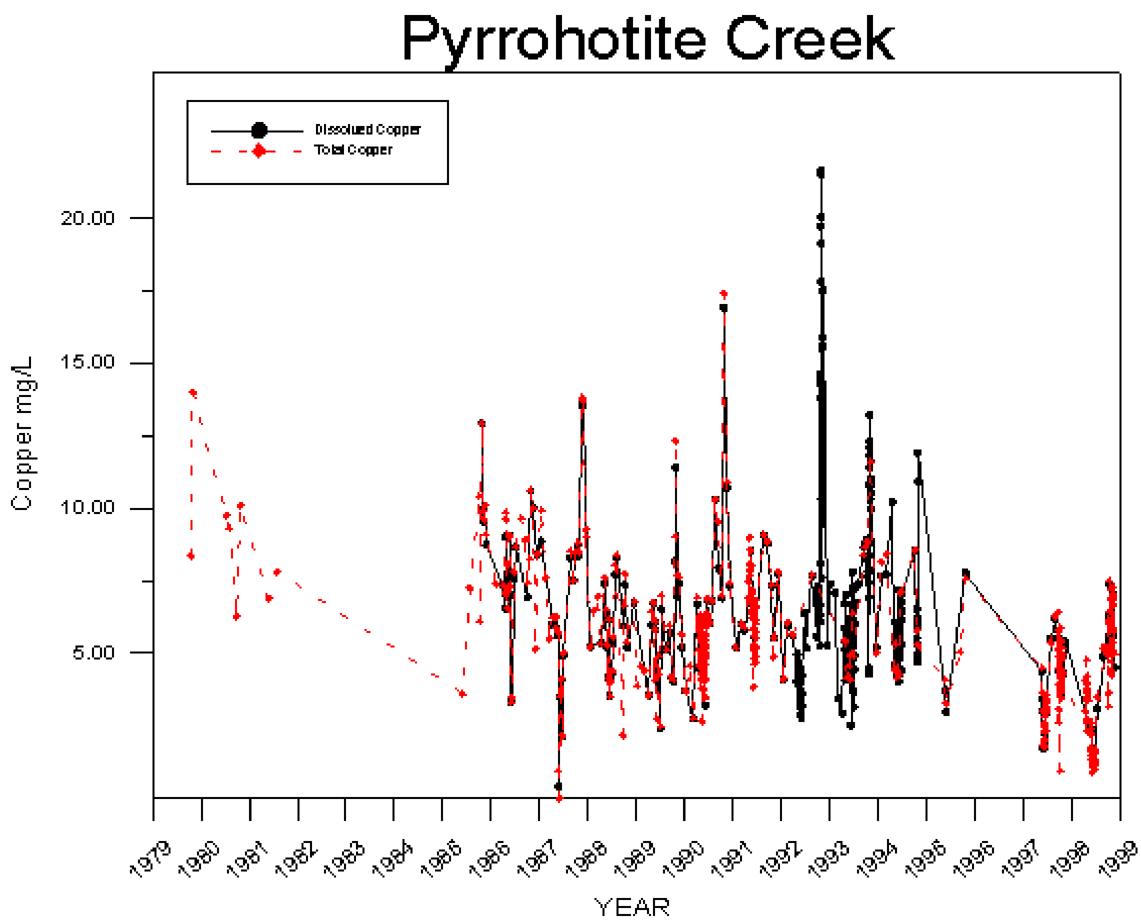
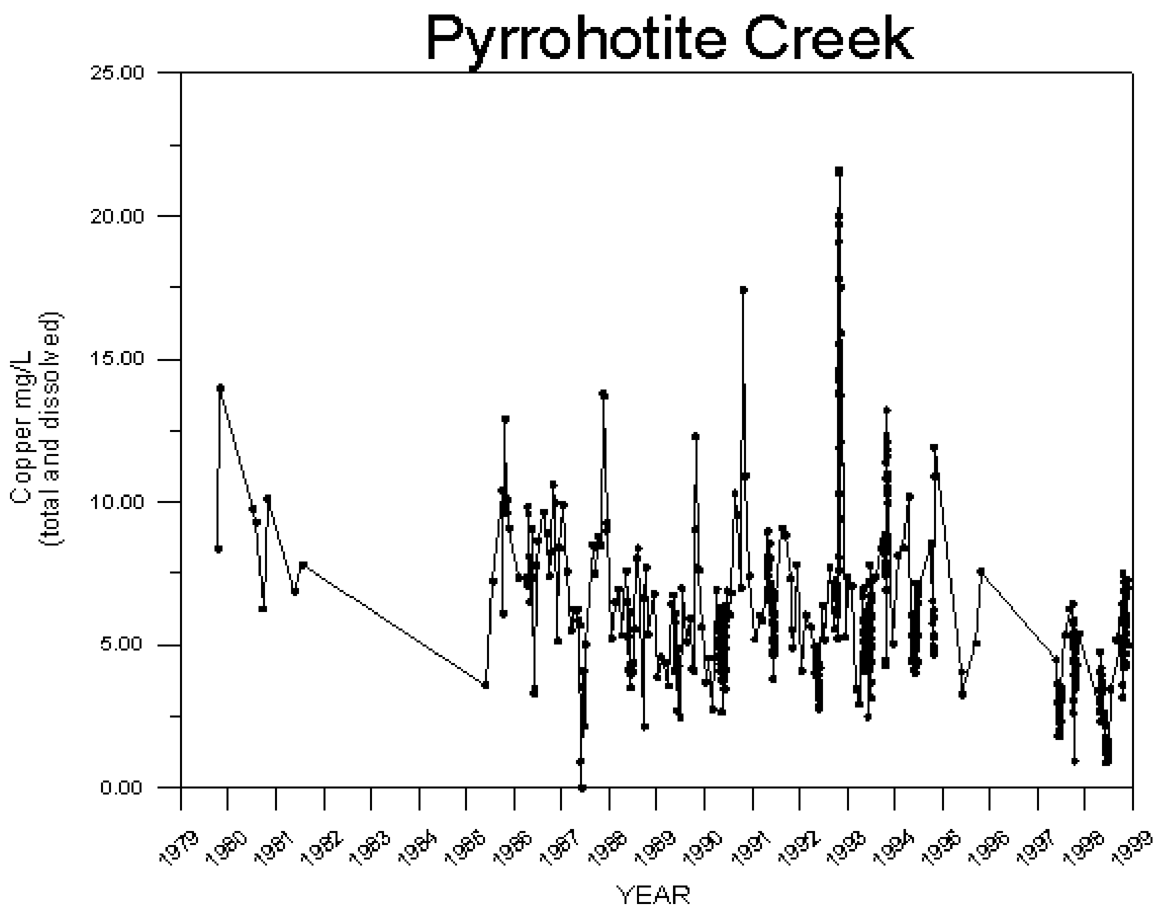


Figure 9 Time series plot of merged dissolved and total copper in Pyrrohotite Creek used for statistical trend tests, 1979 - 1998.



Initial analyses were conducted on the full dissolved copper (1979 - 1998) and total copper (1985 - 1998) data sets. Evidence of strong decreasing trends were found (see Table 1). As a further initial analysis, a new data set that consisted of merged dissolved and total copper values (with totals forming the base and dissolved values filling in the gaps) was used (Figure 9). Again, there was strong evidence of a decreasing trend (Table 1, next page).

As apparent in Figures 8 and 9, the bulk of the data begins in 1986. Further analyses of the three data sets used in the initial tests, beginning in 1986 through 1998, also revealed strong evidence of decreasing trends (Table 2, next page).

Table 1 Nonparametric results for analyses performed on full 1979-98 copper data sets from the Pyrrohotite Creek water quality monitoring station.

	Dissolved	Total	Merged

	Statistic	*P-value*	*Statistic*	*P-value*	*Statistic*	*P-value*
Sen Slope	-4.324	0.00002	-9.0629	1.27e-019	-12.4962	7.83e-036
Upper CI	-0.1493	NA	-0.2733	NA	-0.3286	NA
Lower CI	-0.3480	NA	-0.3769	NA	-0.4325	NA

Table 2 Nonparametric results for analyses performed on 1986-98 copper data from the Pyrrohotite Creek water quality monitoring station.

	Dissolved		**Total**		**Merged**	
	Statistic	*P-value*	*Statistic*	*P-value*	*Statistic*	*P-value*
Sen Slope	-4.2166	0.00002	-8.5416	1.32e-017	-12.2549	1.58e-034
Upper CI	-0.1441	NA	-0.2714	NA	-0.3357	NA
Lower CI	-0.3475	NA	-0.3833	NA	-0.4467	NA

Summary

Trend analyses on copper data collected at the Murex Creek, Tsolum River at Farnham, Tsolum River 500 metres downstream of Murex, and Pyrrohotite Creek water quality monitoring stations revealed the following items of note:

· incomplete data for the Tsolum River 500 metres downstream of Murex site prohibited the use of regression modeling. The dataset for this site also did not contain any data for the months of January, February, or March, which should be taken into consideration when interpreting the data visually;

· Non-parametric tests did not detect any trends at the Murex Creek, Tsolum River at Farnham, or Tsolum River 500 metres downstream of Murex water quality monitoring sites. Subsequent regression modeling at the Murex Creek at Duncan Main and Tsolum River at Farnham sites also did not detect any trends; and

· Regression modeling indicated that flow data collected at the hydrometric station on the Tsolum River at Courtenay explained a good portion of the merged copper data collected at the Murex Creek at Duncan Main and Tsolum River at Farnham water quality monitoring sites.

· Non-parametric tests found strong evidence of a decreasing trend in all analysed copper data sets from the Pyrrohotite Creek water quality monitoring site. Regression modeling was not conducted on the copper data from this site, nor were any annual and seasonal boxplots supplied.

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. Water Quality Section, Water Management Branch, Ministry of Environment, Lands and Parks, Victoria, B.C.

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.

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.

Appendix I Statistical Results for Murex Creek at Duncan Main

Non-parametric results

	statistic	p-value
VBT	0.0279	0.867
SK	-0.1858	0.853
MSK	-0.1946	0.846
SSE	0	NA
LCL	-0.0004	NA
UCL	0.0001	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
-4.022	0.261	NS	----	-0.608	0.033*	0.392

NS - non-significant at the 5% level
* - 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

PAGE < TOC PAGE >

PAGE < TOC PAGE >

Appendix II Statistical Results for Tsolum River at Farnham

Non-parametric results

	statistic	p-value
VBT	0.2055	0.650
SK	0.3974	0.691
MSK	0.4187	0.675
SSE	0	NA
LCL	-0.0002	NA
UCL	0.0004	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
-4.900	0.232	NS	----	-0.578	-0.063*	0.195

NS - non-significant at the 5% level
* - 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

PAGE < TOC PAGE >

Appendix III Statistical Results for Tsolum River 500 metres downstream of Murex

Non-parametric results

	statistic	p-value
SK	-0.0772	0.205
SSE	0	NA
LCL	0	NA
UCL	0	NA

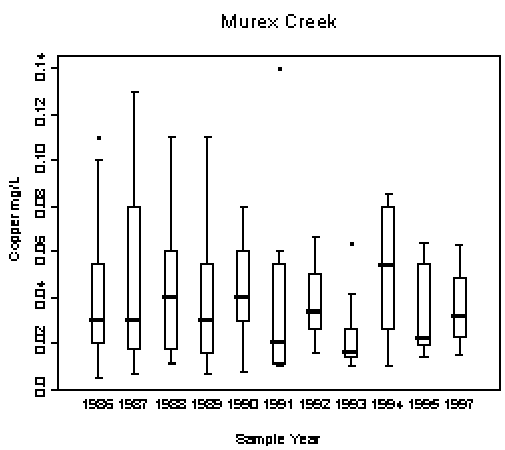
SK - Seasonal Kendall
SEN - Sen Slope Estimator
LCL - Lower Confidence Limit
UCL - Upper Confidence Limit
NA - Not Applicable

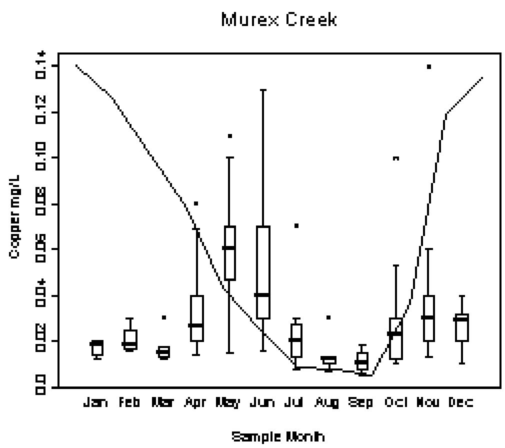
Regression modeling results

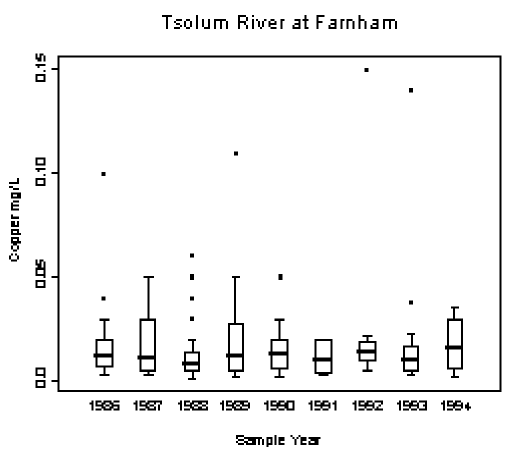
Not Applicable

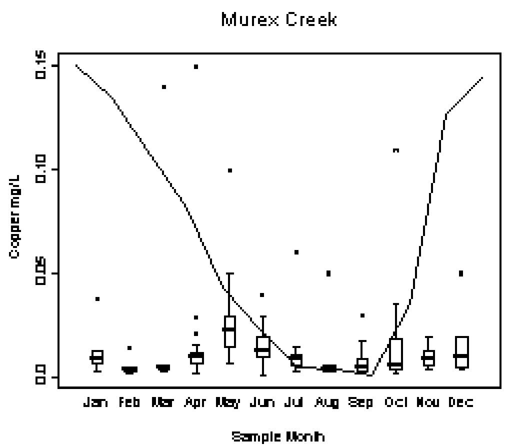
Appendix IV Annual and Seasonal Boxplots

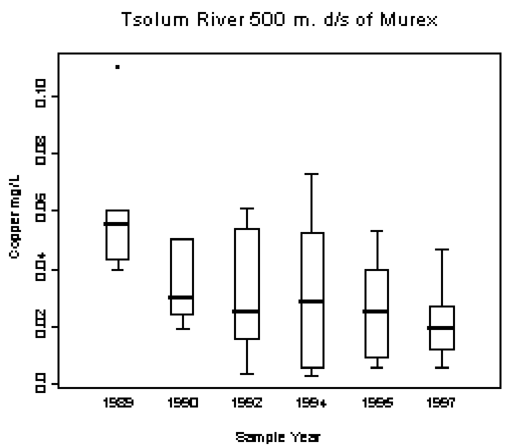
Solid lines in seasonal plots represent seasonal flow rates.




Murex Creek
Copper mg/L
0.0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Sample Month


Tsolum River at Farnham
Copper mg/L
0.0
0.05
0.10
0.15
1986
1987
1988
1989
1990
1991
1992
1993
1994
Sample Year


Murex Creek
Copper mg/L
0.0
0.05
0.10
0.15
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Sample Month


Tsolum River 500 m. d/s of Murex
Copper mg/L
0.0
0.02
0.04
0.06
0.08
0.10
1989 1990 1992 1994 1995 1997
Sample Year