Correction Factors for Freshwater Dissolved and Total Nutrients for 1998 and 2006 Instrument and Method Changes by the King County Environmental Laboratory

March 2016

King County

Department of Natural Resources and Parks
Water and Land Resources Division
Science and Technical Support Section
King Street Center, KSC-NR-0600
201 South Jackson Street, Suite 600
Seattle, WA 98104
206-477-4800 TTY Relay: 711
www.kingcounty.gov/EnvironmentalScience

Alternate Formats Available

Correction Factors for Freshwater Dissolved and Total Nutrients for 1998 and 2006 Instrument and Method Changes by the King County Environmental Laboratory

Submitted by:

Timothy Clark
Science and Technical Support Section
King County Water and Land Resources Division
Department of Natural Resources and Parks

King County
Department of
Natural Resources and Parks
Water and Land Resources Division

Acknowledgements

I would like to express my gratitude to the current and past employees of the King County Environmental Laboratory; their constant drive to improve their methods of analysis is inspiring and a tremendous asset to the Division. My thanks to Sally Abella, Debra Bouchard, Curtis DeGasperi, and Jim Simmonds for their review of this document.

Citation

King County. 2016. Correction Factors for Dissolved and Total Nutrients for 1998 and 2006 Instrument and Method Changes by the King County Environmental Laboratory. Prepared by Timothy Clark, Water and Land Resources Division. Seattle, Washington.

Table of Contents

Executive Summary vi
Acronyms and Abbreviations vii
1.0 Introduction 1
1.1 Summary of King County's Freshwater Monitoring Program 2
1.2 Summary of Method and Instrument Changes and Subsequent Side-By-Side Analyses 3
1.2.1 1998 Digestion Method Change 3
1.2.2 2007 Instrument and Digestion Method Change 4
2.0 Methods 5
2.1 Statistical Methods 5
2.1.1 Tests for Normality and Paired Difference Tests 5
2.1.2 Segmented Robust Linear Regression 6
2.1.3 Trend Analysis 6
2.2 Impacts of Dilution 6
3.0 1998 Freshwater Total Phosphorus Digestion Change Analysis 8
3.1 Data Used 8
3.2 Normality and Paired Difference 8
3.3 Regression Results 10
3.4 Heteroscedasticity 11
3.5 Dilution of Samples 11
4.0 2006 Freshwater Total Phosphorus Instrument Change Analysis 13
4.1 Data Used 13
4.2 Normality and Paired Difference 13
4.3 Regression Results 15
4.4 Heteroscedasticity 16
4.5 Dilution of Samples 16
5.0 2006 Freshwater Total Nitrogen Instrument Change Analysis 18
5.1 Data Used 18
5.2 Normality and Paired Difference 18
5.3 Regression Results 20

5.4 Heteroscedasticity 21
5.5 Dilution of Samples 21
6.0 2006 Freshwater Orthophosphate Instrument Change Analysis 23
6.1 Data Used 23
6.2 Normality and Paired Difference 23
6.3 Regression Results 25
6.4 Heteroscedasticity 26
6.5 Dilution of Samples 26
7.0 2006 Freshwater Nitrate/Nitrite Instrument Change Analysis 28
7.1 Data Used 28
7.2 Normality and Paired Difference 28
7.3 Regression Results 30
7.4 Heteroscedasticity 30
7.5 Dilution of Samples 31
8.0 2006 Freshwater Ammonia Instrument Change Analysis 32
8.1 Data Used 32
8.2 Normality and Paired Difference 32
8.3 Regression Results 33
9.0 Trend Analysis 34
9.1 Total Phosphorus 35
9.2 Total Nitrogen 36
9.3 Orthophosphate 37
9.4 Nitrate/Nitrite 38
10.0 Summary and Recommendations 39
References 42

Figures

Figure 1. Comparison of TOTP.Single and TOTP.Combined for all freshwater programs. 9
Figure 2. One-breakpoint model for TOTP.Single regressed on TOTP.Combined. 10
Figure 3. Comparison of TOTP.API and TOTP.OI for all data pairs above the MDL threshold. 14

Figure 4. No-breakpoint model for TOTP.OI regressed on TOTP.API. 15
Figure 5. Comparison of TOTN.API and TOTN.OI for all data pairs above the MDL threshold. 20
Figure 6. No-breakpoint model for TOTN.OI regressed on TOTN.API. 21
Figure 7. Comparison of ORTHOP.API and ORTHOP.OI for all data pairs above the MDL threshold. 25
Figure 8. Two-breakpoint model for ORTHOP.OI regressed on ORTHOP.API. 26
Figure 9. Comparison of NO23.API and NO23.OI for all data pairs above the MDL threshold. 30
Figure 10. One-breakpoint linear model for NO23.OI regressed on NO23.API. 31
Figure 11. Comparison of NH3.API and NH3.OI for all data pairs above the MDL threshold. 33

Tables

Table 1. Summary statistics for 2004 total phosphorus side-by-side values. 8
Table 2. 2004 total phosphorus t-test and Wilcoxon sign rank test results Single vs Combined Methods. 9
Table 3. 2004 total phosphorus t-test and Wilcoxon sign rank test results comparing the predictions from the original KCEL equation predictions and the observed values. 10
Table 4. Impact of dilution on the results of applying a correction factor. 12
Table 5. Summary statistics for 2006 total phosphorus values. 13
Table 6. 2006 total phosphorus t-test and Wilcoxon sign rank test results OI vs API Methods. 14
Table 7. 2006 total phosphorus t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values. 15
Table 8. Impact of dilution on the results of applying a correction factor. 17
Table 9. Summary statistics for 2006 total nitrogen values. 18
Table 10. 2006 total nitrogen t-test and Wilcoxon sign rank test results OI vs API Methods. 18
Table 11. 2006 total nitrogen t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values. 19
Table 12. Impact of dilution on the results of applying the correction factor. 22
Table 13. Summary statistics for 2006 orthophosphate values. 23

Table 14. 2006 orthophosphate t-test and Wilcoxon sign rank test results OI vs API Methods. 24
Table 15. t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values. 24
Table 16. Impact of dilution on the results of applying a correction factor. 27
Table 17. Summary statistics for 2006 nitrate/nitrite values. 28
Table 18. 2006 nitrate/nitrite t-test and Wilcoxon sign rank test results OI vs API Methods. 28
Table 19. 2006 nitrate/nitrite t-test and Wilcoxon signed rank test results for original KCEL equation predictions vs observed values. 29
Table 20. Impact of dilution on the results of applying the correction factor. 31
Table 21. Summary statistics for 2006 ammonia values. 32
Table 22. 2006 ammonia t-test and Wilcoxon sign rank test results OI vs API Methods. 33
Table 23. Total phosphorus trend data time period and percent censored. 35
Table 24. Total phosphorus trends comparisons. 35
Table 25. Total nitrogen trend data time period and percent censored 36
Table 26. Total nitrogen trend comparison 36
Table 27. Orthophosphate trend data time period and percent censored 37
Table 28. Orthophosphate trend comparison 37
Table 29. Nitrate/nitrite trend data time period and percent censored 38
Table 30. Nitrate/nitrite trend comparison 38
Table 31. Summary of correction factors for total and dissolved nutrients 41

Appendices

Appendix A: Total Nitrogen & Total Phosphorus Method Changes (1998-2007) – KCEL Memo
Appendix B: Change in Instrumentation for Dissolved Nutrients Analyses – KCEL Memo
Appendix C: Description of Major Statistical Tests and the Iterative R Script

EXECUTIVE SUMMARY

One of the greatest benefits of King County's routine water quality monitoring program is the ability to track how water quality changes over time. This ability is a direct result of the longevity of the program and the commitment to taking samples at specific sites for many years. The purpose of this document is to improve long-term freshwater nutrient (nitrogen and phosphorus) data analysis by providing updated correction factors that increase comparability between data collected with differing methods and instruments. This document provides correction factors for laboratory instrument and method changes for measurements of total and dissolved nutrients in freshwaters that occurred in 1998 and 2007. These factors correct for discontinuities and bias due to method changes in the values reported by King County's long-term monitoring programs and improve confidence in the results of trend analyses.

In 1998 and 2007, the King County Environmental Lab (KCEL) made two important improvements to their methods and instrumentation for examining total and dissolved nutrients in the freshwater matrix. On July 1, 1998, KCEL began using a new sample preparation technique for total phosphorus and total nitrogen analysis. On January 1, 2007, the King County Environmental Lab switched the instrument used for the automated analysis of dissolved nutrients in addition to improving the sample preparation method for total nutrients.

Long-term trend analysis assumes that values remain comparable over the time period examined, and that no bias has been introduced due to changes over time in the collection or measurement methods. Except for ammonia (a form of nitrogen), the 1998 and 2007 instrument and method changes significantly biased results for all nutrients and impacted comparability over time.

The analysis presented in this report improves upon a previous best-fit analysis that was conducted in 2004 and 2006 for the 1998 and 2007 laboratory changes, respectively. To do so, a reanalysis of the impacts of the method changes was performed for orthophosphate, total phosphorus, ammonia, nitrate/nitrite, and total nitrogen using data from side-by-side studies conducted in 2004 (assessing the 1998 change) and 2006 (assessing the 2007 change). Correction factors were produced for all nutrients except ammonia, because no bias was detected for ammonia. Correction factors were produced using robust, segmented linear regression analysis. The bulk of this report describes the details of these analyses.

These correction factors are able to predict values that were statistically insignificant from observed values. Therefore, the application of these correction factors allows for comparability of data collected pre- and post- the 1998 and 2007 changes which allows for long-term trend analysis.

In the future, to provide necessary data to assess potential bias, it is recommended that side-by-side analyses be completed prior to any instrument and method change. It is also recommended that other past laboratory changes be reexamined for bias in the data.

ACRONYMS AND ABBREVIATIONS

API	Astoria Pacific International Astoria2 autoanalyzer
KC	King County
KCEL	King County Environmental Laboratory
L	liter
LIMS	Laboratory Information Management System
MDL	method detection limit
mg	milligram
NH3	Ammonia-N
NO23	Nitrate-nitrite-N
OI	OI Analytical Flow Solution III autoanalyzer
ORTHOP	Orthophosphate phosphorus
ppb	parts per billion
ppm	parts per million
RDL	Reporting Detection Limit
SI	Silica
SOP	Standard Operating Procedure
TOTN	Total Nitrogen
TOTP	Total Phosphorus

This page intentionally left blank.

1.0 INTRODUCTION

One of the greatest benefits of King County's programs that routinely monitor at lake, stream, and marine stations has been the ability to examine long-term trends in water quality parameters because of the longevity and robust nature of the data set. However, long-term trend analysis assumes that values remain comparable over the time period examined, with no bias introduced as the result of changes over time in site locations, collection techniques, or measurement methods. In some cases, instrument and method changes have had an impact on values produced by KCEL by biasing some laboratory results toward higher or lower values relative to those reported prior to the changes. Once a measureable bias has been detected and an appropriate statistical analysis has been carried out, best-fit correction can be applied to the affected data. This will allow trend analyses to exclude shifts unrelated to environmental conditions.

In 1998 and 2007, the King County Environmental Lab (KCEL) made two important improvements to their methods and instrumentation for examining total and dissolved nutrient concentrations in the freshwater matrix. On July 1, 1998, KCEL began using a combined persulfate digestion technique ("combined digestion") for total phosphorus and total nitrogen analysis (SOP SM4500-P-B, F Modified); this method replaced single persulfate digestion ("single digestion") (SM4500-P-B, E). On January 1, 2007, the KCEL switched from using the OI Analytical Flow Solution III instrument (OI) to the Astoria Pacific International Astoria2 instrument (API) for automated analysis of dissolved and total nutrients. Additionally, at the same time, the combined digestion method for total phosphorus and total nitrogen was modified that improved the digestion efficiency for total phosphorus. The Standard Operating Procedure codes (SOPs) for these methods, as well as greater discussions of the method/instrument changes, are available in Appendices A and B.

In 2004 and 2006, side-by-side studies were performed and the data analyzed to determine the degree of impact caused by the instrument/methodology shifts. In 2004, the 1998 digestion method change was examined, and in 2006, the instrument and further digestion method refinement were examined. These analyses were grouped by monitoring program: Major Lakes, Streams, and County Lakes. Statistical regressions, paired parametric t-tests, and tests for normality were run. The results of these analyses can be seen in the attached memos (Appendices A and B). The statistical regressions were run for each monitoring program and included values for results below the method detection limit. The results of these tests were used to produce correction factors recommended for orthophosphate for the Major and County Lakes monitoring programs and total phosphorus for all freshwater programs.

After a reexamination of the distribution of the side-by-side datasets statistical tests applied, there are improvements upon the previous correction factors that will yield better correlated correction factors. These include:

- grouping the data by matrix (i.e., grouping all freshwater monitoring data), applying non-parametric statistical tests,

- excluding values below the method detection limit, and
- performing iterative robust regressions on discrete ranges of concentrations in consideration of instrument sensitivity at high and low concentrations, i.e., segmented, robust linear regression.

This improved analysis was performed for orthophosphate, total phosphorus, ammonia, nitrate/nitrite, and total nitrogen using data from the 2004 and 2006 side-by-side studies. From this, correction factors were produced that could be applied for analyzing long-term data without violating the assumptions inherent to the analysis.

The purpose of this document is to report the results of a re-analysis of the side-by-side comparisons and to improve long-term nutrient data analysis by providing correction factors. This document provides correction factors for instrument and method changes that occurred in 1998 and 2007; these correction factors allow continuity in the values reported by King County's long-term monitoring programs and improve confidence in the results of trend analyses.

1.1 Summary of King County's Freshwater Monitoring Program

King County is committed to monitoring the water quality of the County's lakes and streams to ensure their continued health, as well as the health of the public who utilize the many freshwater resources. The King County Major Lake and Stream Monitoring Program is designed to protect the significant investment in water quality improvement and protection made by the people of King County beginning in the 1950s. At that time wastewater flowed largely untreated into Lake Washington and Puget Sound and many rivers and smaller lakes, fouling water and making local beaches unusable. Shortly after Metro was formed in 1958, construction began on two of the county's regional treatment plants, West Point in Seattle's Magnolia neighborhood and South Treatment Plant in Renton, which were officially up and running by 1966. By the late 1960s, regional water quality began improving dramatically.

The improving water quality was an important milestone and was tracked by this program. Continued ongoing monitoring is important to understand if new problems are appearing that can impact water quality and thereby the health of people who use our regional waters. The long-term environmental impacts of all types of pollution that affect the quality of lakes and streams can only be evaluated by sampling at multiple sites throughout the watershed. In 1994, King County assumed authority of Metro and its legal obligation to treat wastewater for 34 local jurisdictions and local sewer agencies that contract with King County. The County also assumed responsibility for continuing the monitoring of these surface waters. For many of the sites in the Stream and River Monitoring Program routine water quality analysis began as far back as the 1970s. Parameters analyzed in the routine water quality samples include conventionals (flow, temperature, dissolved oxygen, pH, conductivity, turbidity, total suspended solids, nutrients (phosphorus and nitrogen) and fecal coliform bacteria.

1.2 Summary of Method and Instrument Changes and Subsequent Side-By-Side Analyses

1.2.1 1998 Digestion Method Change

Prior to July of 1998, total nitrogen (TOTN) and total phosphorus (TOTP) digestions were performed by two separate persulfate digestions. Subsequent analysis was also separate and performed manually on a spectrophotometer. Both digestion techniques were applied to saline as well as non-saline matrices. Beginning July 1, 1998, a combined persulfate digestion technique (referred to as the "combined digestion") was employed allowing for the simultaneous digestion of both TOTP and TOTN in non-saline matrices. A single persulfate digestion (referred to as the "single digestion"), very similar to that used previous to July 1998, was maintained for saline matrices where only TOTP was reported. Analysis for both digestion techniques was automated via an OI Analytical segmented flow autoanalyzer.

Long term trend analysis for several freshwater monitoring programs conducted in 2003 identified a significant step decrease in TOTP concentrations occurring in 1998. No environmental factors could be identified to account for the stepwise decrease. Based on the information available, it was considered most likely that the discrepancy was associated with the method change that occurred during 1998.

A more extensive method evaluation was undertaken beginning in October 2004 in an attempt to confirm and quantify the effects of the 1998 method change. Side-by-side analyses were conducted using select locators from the Major Lakes, Streams, and County Lakes monitoring programs. Samples were prepared using the single digestion and the combined digestion methods. It is important to note that analysis for both digestion methods were performed in an automated manner using the existing OI Analytical Flow Solution III segmented flow autoanalyzer.

Evaluation of the side-by-side data indicated a clear statistically significant difference between the single and combined digestion methods. This relationship was true for the locators associated with all three monitoring programs. In each case, data collected using the single digestion method consistently provided greater values. These findings were consistent with the 2003 long term trend analysis and strongly suggested that the step decrease in TOTP data occurring in 1998 was directly related to the implementation of the combined digestion method.

Regression data generated from the statistical evaluation provided a correlation between the data collected by the single and combined digestion methods. However, it must be noted that its applicability to correlating data collected prior to and after July 1998 is limited. The potential effects of instrumentation (spectrophotometer vs. autoanalyzer) were not addressed during this evaluation. The evaluation strictly compared the difference between the two digestion steps while applying the same automated determinative step to

both methods. As a result, the correlation may not accurately reflect the true cumulative effect that the change in the digestion and the determinative step may have had.

1.2.2 2007 Instrument and Digestion Method Change

On January 1, 2007, KCEL implemented new instrumentation for analysis of dissolved and total nutrient parameters – NH3, NO23 (singly or combined), ORTHOP, SI, TOTN, and TOTP – on both non-saline and saline matrices. On the same date, KCEL introduced a modified combined persulfate digestion technique (referred to as "modified combined digestion") for total nutrient parameters (TOTN and TOTP) for non-saline matrices. This method improved upon the TOTP digestion efficiency of the previous combined digestion method. In addition, matrix matching was incorporated into the digestion utilized for saline matrices to eliminate instrumentally derived biases. The current LIMS method codes associated with the above mentioned parameters were updated to reflect a revision change in the standard operating procedure (SOP) due to the change in instrumentation. There were no changes to the current method detection limits (MDL) and reporting detection limits (RDL) associated with any of the dissolved nutrient parameters.

Side-by-side analyses of samples from routine monitoring projects were performed in 2006 prior to use of the new instrument so that potential biases between the current instrumentation and the new instrumentation could be determined. Only the dissolved nutrient parameters routinely tested and reported for these projects were evaluated.

2.0 METHODS

The data used for regression analyses were taken from the 2004 side-by-side method comparison for total phosphorus completed for the 1998 digestion method change and the 2006 side-by-side instrument comparison for dissolved and total nutrients completed for the 2007 instrument change. The data consist of paired values for the two evaluated methods (i.e., for each water sample, a value was derived for each method or instrument involved in the change).

All pairs with one or both data points below the method detection limit (MDL) determined by KCEL were dropped from the analysis (i.e., censored). The fractions of censored values for each side-by-side study are discussed within their respective chapters. Analyses were performed on a combined dataset of all freshwater data, including samples from the Major Lakes, County Lakes, and Streams routine monitoring programs. Appendix C further details the major statistical tests applied and the iterative R script employed.

2.1 Statistical Methods

2.1.1 Tests for Normality and Paired Difference Tests

The Shapiro-Wilk test for normality and Quantile-Quantile (Q-Q) plots were used to assess the normality of the data before and after log-transformation. Q-Q plots allow the visual assessment of the normality of distribution and are advantageous over the Shapiro-Wilk test due to the test's bias towards rejecting the null hypothesis for large sample sizes (Rochon et al. 2014). The Shapiro-Wilk test does for a quick comparison and was used as a first-line test of normality between untransformed and log-transformed data. If the data were determined to be normally distributed, the parametric t-test was used to determine if there was a significant difference between pairs in the paired dataset. If the data were determined to be non-normal or log-normally distributed, the paired data were then tested for statistical difference using the non-parametric Wilcoxon signed rank test and log-transformed t-test, where the natural logs of the values were compared. If both of these tests showed no significant difference (p-values>0.10), then no further analyses were performed, and the impacts of the instrument/method change was determined to be insubstantial.

The ability of the original regression equations provided in the KCEL memos (Appendices A and B) to accurately correct the measurement bias was assessed. For nutrient datasets that showed a significant difference between pairs, the regression equations from the KCEL memos were used to correct the post-method/instrument-change data, i.e., predict the result of the other method's measure. The corrected data were then tested for statistical difference against the pre-method/instrument-change data using the log-transformed t-test and the Wilcoxon signed rank test.

2.1.2 Segmented Robust Linear Regression

For significantly different side-by-side nutrient datasets, an R script was run that inserted variable breakpoints into linear and log-log robust, segmented regression models. Robust regression modeling was used because of its ability to handle outliers, such as non-normally distributed measurement errors. The final breakpoints and linear model were selected based on the minimization of the model's robust residual standard error and the statistical significance of the assessed breakpoints. Breakpoints were examined at intervals proportional to the observed values of the nutrient; specific breakpoint intervals are identified in each section.

The heteroscedasticity (i.e., difference in variability between groups) of the segmented linear regression was assessed by comparing the model errors using Bartlett's test for variance homogeneity. While heteroscedasticity does not affect the predictive regression model, this analysis allows the investigation of where there is inherent variability ("noise") after the application of a correction factor. This "noise" may be linked to either method/instrument sensitivity or natural environmental variability; it is not plausible to discern what is the dominant cause of variability.

2.1.3 Trend Analysis

The impact of the method changes and correction factors on long-term trends was assessed using the Seasonal Mann-Kendall test. Trends were assessed at sampling stations A522 (Lake Union), 0826 (Lake Washington), 3106 (Green River), and 0631 (Issaquah Creek). Trend analyses were completed using uncorrected values, values corrected using the previous correction factors, and values corrected using the new correction factors outlined in this document. Mean monthly values were used as input for the tests. To account for intra-annual variability, trends were assessed for individual months and the test statistics summed.

A major issue with trend analysis of long-term datasets is the decreasing detection limits over time; if not handled correctly, trend analysis may track the trend in detection limits rather than in environmental concentrations. To negate the impact of shifting method detection limits over time, all values below the maximum detection limit for the parameter were replaced with the value of the maximum detection limit. While this method lowers the power to detect trends, it is robust and it is the simplest method to handle multiple detection limits for a Mann-Kendall test (Helsel and Hirsch, 2002).

2.2 Impacts of Dilution

In cases where nutrient concentrations exceed the analytical range of the laboratory method, samples are diluted by a factor from 2 up to 100. This dilution allows the projected sample concentration to be within the method analytical range. The value for the diluted sample is multiplied by the dilution factor applied, and this product is reported. It is important to note that if the correction is applied to the reported value rather than the diluted value, the corrected value will be biased. The impacts of applying the correction

factor to the reported value rather than the subsample value are analyzed for each parameter.

3.0 1998 FRESHWATER TOTAL PHOSPHORUS DIGESTION CHANGE ANALYSIS

3.1 Data Used

The data used for this analysis were taken from the original 2004 side-by-side analysis of the 1998 digestion method change. These include 151 pairs from the Major and County Lakes monitoring programs, and 70 pairs from the Streams monitoring program. After combining and censoring the data (MDL=0.005 ppm), 221 pairs (100%) from the freshwater monitoring programs remained; no values in this dataset were below the detection limit.

It must be noted that this analysis only applies to the impact of moving from a single to a combined digestion procedure. On July 1, 1998 when the KCEL switched to combined digestion, the lab also began using the OI Analytical Flow Solution III autoanalyzer, which replaced the manual spectrophotometer instrument, and decreased the sample aliquot size from 100 mL to 30 mL. The side-by-side evaluation was done using only the OI autoanalyzer and 30 mL aliquots, and thus, this analysis compared the digestion method change but not the aliquot or instrument changes. These two other method changes may also have a substantial impact on the true difference between total phosphorus values prior to and after July 1, 1998, but the original spectrophotometer was longer available in 2004 and the comparison could not be made.

3.2 Normality and Paired Difference

The differences between the mean and median of the single- ("TOTP.Single") and combined-digestion ("TOTP.Combined") values for total phosphorus allude to the non-normality of the data (Table 1). The Shapiro-Wilk test for normality found that the data are not normal, rejecting the null hypothesis. Log-transformation of the datasets yielded more normal distributions, according to Q-Q plots, but the Shapiro-Wilk null hypothesis was rejected for each log-transformed dataset as well. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 1. Summary statistics for 2004 total phosphorus side-by-side values.

Monitoring Program	TOTP.Single				TOTP.Combined			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	0.0541	0.0510	0.015	0.150	0.0417	0.0390	0.012	0.105
Lakes	0.0186	0.0170	0.006	0.071	0.0146	0.0140	0.005	0.068
All Programs	0.0299	0.0210	0.006	0.150	0.0232	0.0160	0.005	0.105

All values in ppm.

Table 2 displays the log-transformed t-test and Wilcoxon sign rank test p-values and the sample sizes for the comparison between observed TOTP.Single concentrations and TOTP.Combined concentrations. For all sites, the values of Single and Combined digestion methods were determined to be significantly different with both tests (p-values<0.001).

Table 2. 2004 total phosphorus t-test and Wilcoxon sign rank test results Single vs Combined Methods.

Monitoring Program	Paired t-test p-value	Wilcoxon signed rank p-value
Streams	<0.001** (n=70)	<0.001** (n=70)
Lakes	<0.001** (n=151)	<0.001** (n=151)
All Programs	<0.001** (n=221)	<0.001** (n=221)

* Significant (p<0.05)
** Highly Significant (p<0.01)

Figure 1 displays a plot of the TOTP.Single and TOTP.Combined observations for all freshwater programs. For all observed values, there is an evident deviation from the 1:1 ratio with the single digestion method reading higher values than those digested with the combined method.

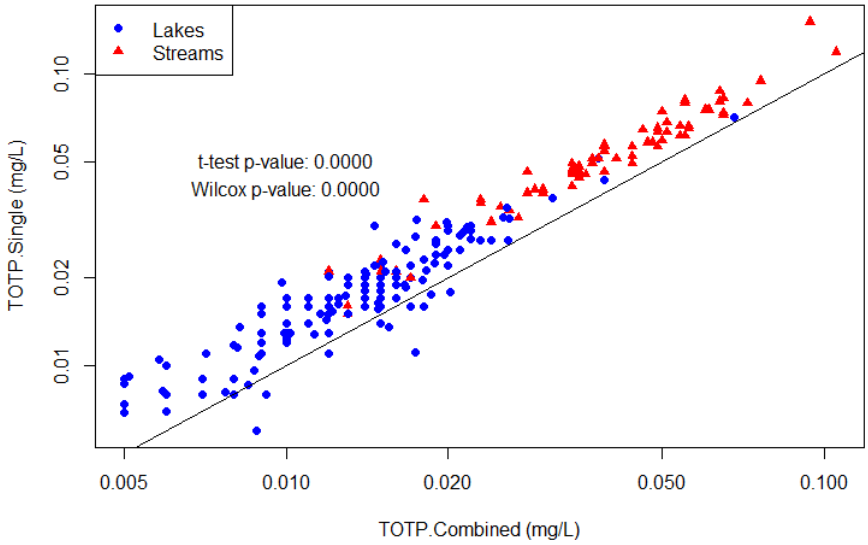

Figure 1. Comparison of TOTP.Single and TOTP.Combined for all freshwater programs.The solid line is 1:1 ratio. Note the log-scale. The two largest values are Newaukum Creek and Springbrook Creek. Note: these are not statistically significant outliers.

The original correction equations provided by KCEL in 2004 for the freshwater monitoring programs are presented in Table 3 (see Appendix A). The Wilcoxon test for the Streams monitoring program dataset yielded a p-value less than 0.10 and a very significant p-value for the t-test, while the Lakes dataset were not found to be significantly different. For the

combined dataset, the values predicted using the original regression and the observed values were found to be significantly different using both tests.

Table 3. 2004 total phosphorus t-test and Wilcoxon sign rank test results comparing the predictions from the original KCEL equation predictions and the observed values.

Monitoring Program	Equation	Paired t-test p-value	Wilcoxon signed rank p-value
Streams	$\boldsymbol{TOTP.Single = 1.262 * TOTP.Combined}$	0.0050** (n=70)	0.0572 (n=70)
Lakes		0.2191 (n=151)	0.1689 (n=151)
All Programs		0.0223* (n=221)	0.0215* (n=221)

* Significant (p<0.05)
** Highly Significant (p<0.01)

3.3 Regression Results

Breakpoints in the iterative R script were examined at 0.0001 ppm intervals. The iterative R script examined robust regression models with no, one, and two breakpoints. The best fitting, with significant breakpoints, was:

When:

TOTP.Combined< 0.018 ppm $\quad TOTP.Single = 0.6203 * TOTP.Combined^{0.8312}$

TOTP.Combined≥ 0.018 ppm $\quad TOTP.Single = 1.066 * TOTP.Combined^{0.9466}$

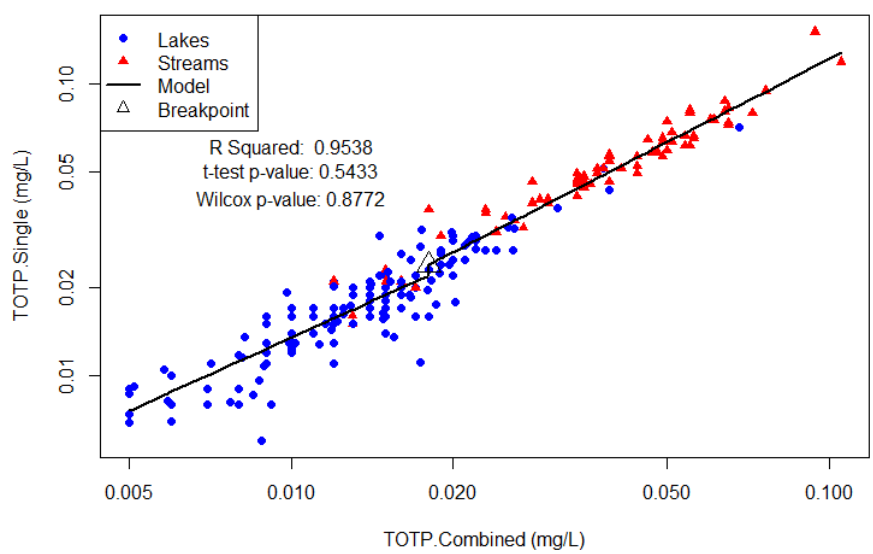

Figure 2. One-breakpoint model for TOTP.Single regressed on TOTP.Combined.Note the log-scale.

Figure 2 displays the plot of the fitted robust segmented model, as well as the R-squared value and paired difference tests' p-values. Both the log-transformed t-test and the Wilcoxon signed rank test failed to find a significant difference between observed and predicted TOTP.Single values. In the dataset, 125 points were below and 96 points were above or equal to the breakpoint.

The single breakpoint segmented regression equation provides a better fit than the original correction factor. This regression yields highly insignificant paired difference test results for the entire freshwater matrix, whereas the original regression yielded significantly different data (Table 2),

3.4 Heteroscedasticity

The high level of variation throughout all observed pairs suggests low precision for one or both of the instruments and/or methodologies. A potential cause may be the aliquot volume used in the digestion process; there may be variation in phosphorus content between 30 mL aliquots, especially when considering likely patchiness and particle size distributions of phytoplankton biomass and other suspended solids. This may also explain the heteroscedasticity (i.e., the difference in variability) seen between lakes and streams.

To better understand the heteroscedasticity between the two monitoring programs, the root mean squared residuals (i.e., the square root of the average squared distance from the best fit line) were investigated. The root mean squared residual in the one-breakpoint model for Streams and Lakes samples were 0.1134 and is 0.1803, respectively. The Bartlett test for variance homogeneity was applied, and the null hypothesis of equal variance between sites was rejected, which means that there is a significant difference in the variation of values between the Streams and Lakes programs. In spite of this, heteroscedasticity does not bias the coefficient estimates; it only biases their standard errors. Therefore, the single breakpoint segmented regression model efficiently predicts the impact of the 1998 method change.

3.5 Dilution of Samples

The current SOP for total phosphorus states that values greater than 0.440 ppm should be diluted and rerun. It is necessary that the value produced for the diluted sample be corrected using the regression model and then scaled up with respect to the amount of dilution. If the correction is applied to the concentration after the dilution factor is applied, the corrected value will be skewed.

For example, a sample is diluted 5x and the resulting analysis using the combined digestion method yields a TOTP concentration of 0.150 ppm. The value reported in LIMS would therefore be 0.75 ppm (0.15 ppm* 5). If the reported value is corrected using the regression model to be analagous to the results of the single digestion method, the corrected value is 0.812 ppm, but if the value from the diluted sample is corrected and then the value is multiplied by 5, the result is 0.888 ppm. Table 4 provides several examples of

this problem. For converting TOTP.Combined to TOTP.Single, correcting reported values rather than correcting diluted values and scaling up will lead to biased-low concentration estimates.

Table 4. Impact of dilution on the results of applying a correction factor.

Dilution	Diluted Result	Reported Value	Reported Value Corrected	Diluted Value Corrected (and Scaled-Up	Relative Percent Difference
1x	0.150	0.15	0.178	0.178	0.0
5x	0.150	0.75	0.812	0.888	8.9
10x	0.150	1.50	1.56	1.78	13.2
20x	0.150	3.00	3.00	3.55	16.8

4.0 2006 FRESHWATER TOTAL PHOSPHORUS INSTRUMENT CHANGE ANALYSIS

4.1 Data Used

The data used for analysis were taken from the original 2006 side-by-side analysis. These include 224 pairs from the Stream monitoring program, 203 pairs from the Major Lakes monitoring program, and 250 pairs from the County Lakes monitoring program. After combining and censoring the data (MDL=0.005 ppm), 444 pairs (65%) from the freshwater monitoring programs remained.

4.2 Normality and Paired Difference

The differences between the mean and median of the OI (TOTP.OI) and API instrument (TOTP.API) values for total phosphorus allude to the non-normality of the data with one exception (Table 5). The major Lakes' mean and median values are very similar, suggesting a normal distribution, but the Shapiro-Wilk test for normality found that the data are not normal, rejecting the null hypothesis. Log-transformation of the datasets, including the Major Lakes, yielded more normal distributions according to Q-Q plots. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 5. Summary statistics for 2006 total phosphorus values.

Monitoring Program	TOTP.OI				TOTP.API			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	0.0551	0.0483	0.0306	0.2260	0.0600	0.0509	0.0080	0.2599
Major Lakes	0.0153	0.0151	0.0059	0.0607	0.0161	0.0161	0.0072	0.0711
County Lakes	0.0216	0.0155	0.0050	0.3453	0.0275	0.0188	0.0057	0.4633
All Programs	0.0309	0.0184	0.0050	0.3453	0.0349	0.0200	0.0057	0.4633

All values in ppm.

Table 6 displays the t-test and Wilcoxon sign rank test p-values and sample sizes for the comparison between observed TOTP.OI values and TOTP.API values. For all sites, the values from API and OI methods were determined to be significantly different with both tests (p-values<0.001).

Table 6. 2006 total phosphorus t-test and Wilcoxon sign rank test results OI vs API Methods.

	Uncensored		Censored	
Monitoring Program	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams	<0.001** (n=224)	<0.001** (n=224)	<0.001** (n=224)	<0.001** (n=224)
Major Lakes	<0.001** (n=203)	<0.001** (n=203)	<0.001** (n=203)	<0.001** (n=203)
County Lakes	<0.001** (n=250)	<0.001** (n=250)	<0.001** (n=241)	<0.001** (n=241)
All Programs	<0.001** (n=677)	<0.001** (n=677)	<0.001** (n=668)	<0.001** (n=668)

* Significant (p<0.05)
** Highly Significant (p<0.01)

Figure 3 displays a plot of the TOTP.API and TOTP.OI observations for all freshwater programs after censoring the data. As seen in Figure 3, for all observed values there is an apparent departure from the 1:1 ratio with the API instrument reading higher values than OI. Beginning at TOTP.API ≅ 0.08 ppm, a departure is visibly apparent. In Figure 3, high lake values correspond to Cottage, Echo, Star, Allen, Pine, and Haller lakes; all of these samples were taken during wet weather except Pine Lake. High stream values correspond primarily to Fairweather, Newaukum, Springbrook, Juanita, Forbes, and Thornton.

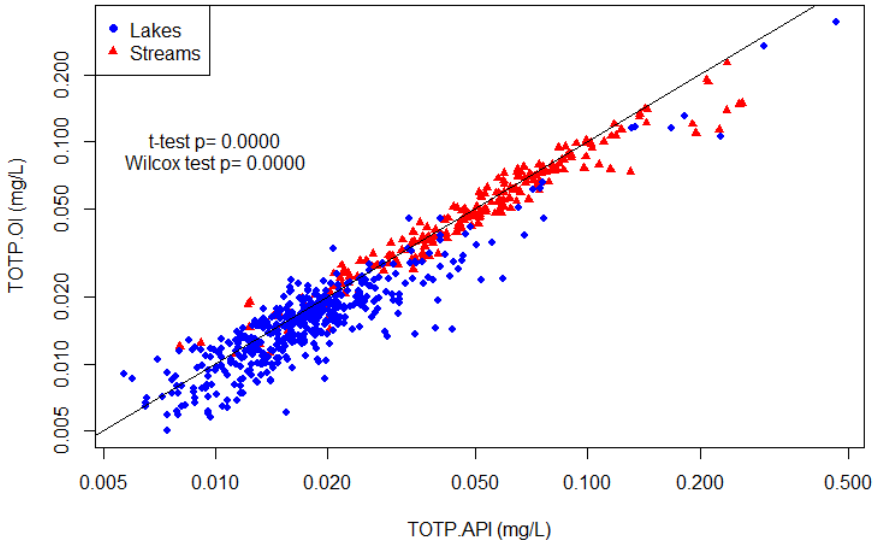

Figure 3. Comparison of TOTP.API and TOTP.OI for all data pairs above the MDL threshold. The solid line is 1:1 ratio. Note the log-scale.

The original correction equations provided by KCEL for the freshwater monitoring programs are presented in Table 7 (see Appendix B), in addition to the results of the paired difference tests comparing observed and predicted values. Censoring of data below the detection limit did not have an apparent impact on any of the sampling programs. The observed and predicted values were not found to be statistically different with any test; the

t-test for Streams and the Wilcoxon signed rank test for County lakes yielded p-values less than 0.10.

Table 7. 2006 total phosphorus t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values.

		Uncensored		Censored	
Monitoring Program	**Equation**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams	$\boldsymbol{TOTP.OI = 0.785 * TOTP.API^{0.927}}$	0.0943 (n=224)	0.6523 (n=224)	0.0943 (n=224)	0.6523 (n=224)
Major Lakes	$\boldsymbol{TOTP.OI = 0.00258 + 0.788 * TOTP.API}$	0.2439 (n=203)	0.8165 (n=203)	0.2439 (n=203)	0.8165 (n=203)
County Lakes	$\boldsymbol{TOTP.OI = 0.509 * TOTP.API^{0.884}}$	0.9286 (n=250)	0.1516 (n=250)	0.4683 (n=241)	0.0641 (n=241)
All Programs		0.2107 (n=677)	0.6740 (n=677)	0.6735 (n=668)	0.5015 (n=668)

* Significant (p<0.05)
** Highly Significant (p<0.01)

4.3 Regression Results

Breakpoints in the iterative R script were examined at 0.0001 ppm intervals. The R script found that the inclusion of one or two breakpoints in the regression model was not statistically significant. The R script yielded a regression model as follows:

$$TOTP.OI = 0.8217 * TOTP.API^{0.9697}$$

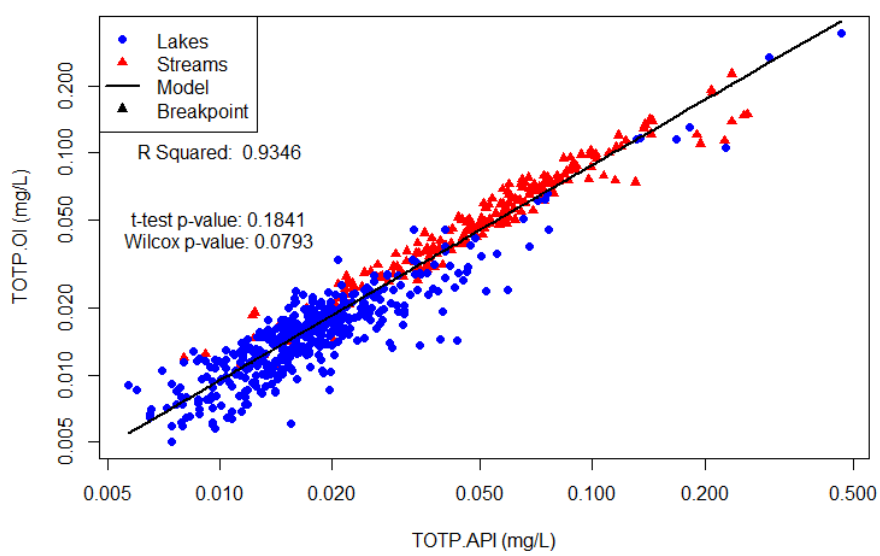

Figure 4. No-breakpoint model for TOTP.OI regressed on TOTP.API.Note the log-scale.

Figure 4 displays the plot of the fitted robust segmented model with no breakpoints, as well as the R-squared value and paired difference tests' p-values. Both tests failed to find a significant difference between observed and predicted TOTP. OI values at the α=0.05 level of significance; although, the Wilcoxon test p-value is less than 0.10 and may be considered too low. Since the breakpoints in the one- and two-breakpoint models were not determined to be significant, the no-breakpoint model is the prefered model. Even though the two-breakpoint was apparently better fitting, the regression is over-fitting, i.e., it describes random error instead of the underlying relationship between the two instruments' results.

4.4 Heteroscedasticity

The high level of variation throughout all observed pairs suggests low precision for one or both of the instruments and/or methodologies. A potential cause may be the aliquot volume used in the digestion process; there may be a large amount of variation in phosphorus content between 30 mL aliquots. The samples used in the side-by-side analysis were from late spring and summer, which coincides with low flow in streams and high algal biomass in lakes. Because there could be greater heterogeneity in lake samples than stream samples (i.e., aliquots may not equally capture phytoplankton), this may explain the heteroscedasticity (i.e., the difference in variability) seen between lakes and streams.

To better understand the heteroscedasticity between the Streams and Lakes programs, the root mean squared residuals were investigated. The root mean squared residual in the no-breakpoint model for Streams and Lakes samples are 0.161 and is 0.240, respectively. The Bartlett test for variance homogeneity was used, and the null hypothesis of equal variance between sites was rejected, which means that there is a significant difference in the variation of Streams and Lakes program values. In spite of this, heteroscedasticity does not bias the regression coefficient estimates; it only biases their standard errors. Therefore, the segmented regression model efficiently predicts the impact of the 2006 instrument change.

4.5 Dilution of Samples

The current SOP for total phosphorus states that values greater than 0.440 ppm should be diluted. It is imperative that the value produced for the diluted sample be corrected using the regression model and then scaled up with respect to its dilution. If the correction is applied to the concentration after the dilution factor is applied, the corrected value will be skewed.

For example, a sample is diluted 5x and the resulting analysis using the API instrutment yields a TOTP concentration of 0.15 ppm. The value reported in LIMS would therefor be 0.75 ppm (0.15 ppm* 5). If the reported value is corrected using the no-breakpoint model to be analagous to the results of the API instrument, the corrected value is 0.622 ppm, but if the value from the diluted sample is corrected and then the value is multiplied by 5, the result is 0.653 ppm. Table 8 provides several examples of this problem. For converting TOTP.API to TOTP.OI, correcting reported values rather than correcting diluted values and scaling up will lead to biased-low concentration estimates.

Table 8. Impact of dilution on the results of applying a correction factor.

Dilution	Diluted Result	Reported Value	Reported Value Corrected	Diluted Value Corrected and Scaled-Up	Relative Percent Difference
1x	0.15	0.15	0.131	0.131	0.0
5x	0.15	0.75	0.622	0.653	4.9
10x	0.15	1.50	1.22	1.31	7.1
20x	0.15	3.00	2.38	2.62	9.6

5.0 2006 FRESHWATER TOTAL NITROGEN INSTRUMENT CHANGE ANALYSIS

5.1 Data Used

The data used for this analysis were taken from the datasets used in the original 2006 side-by-side analysis. These include 224 pairs from the Stream monitoring program, 174 pairs from the Major Lakes monitoring program, and 250 pairs from the County Lakes monitoring program. After combining and censoring the data (MDL=0.05 ppm), 648 pairs (100%) from the freshwater monitoring programs remained; no values were below the detection limit.

5.2 Normality and Paired Difference

The summary statistics for the OI (TOTN.OI) and API method (TOTN.API) values for total nitrogen allude to the non-normality of the data when considering the mean and median values (Table 9). The Shapiro-Wilk test for normality found that the data were not normal, rejecting the null hypothesis. Log-transformation of the datasets yielded more normal distributions, according to Q-Q plots, but the Shapiro-Wilk null hypothesis was rejected for the Major Lakes and Streams log-transformed datasets. The County Lakes dataset was statistically log-normally distributed. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 9. Summary statistics for 2006 total nitrogen values.

Monitoring Program	TOTN.OI				TOTN.API			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	1.070	0.9987	0.1279	2.722	1.080	1.034	0.2004	2.708
Major Lakes	0.4539	0.4164	0.2620	0.8380	0.4483	0.4140	0.2984	0.9279
County Lakes	0.5719	0.5258	0.1919	1.759	0.5766	0.5260	0.2156	2.022
All Programs	0.7125	0.5670	0.1279	2.722	0.7161	0.5642	0.2004	2.708

All values in ppm.

Table 10 displays the t-test and Wilcoxon sign rank test p-values and sample sizes for the comparison between observed TOTN.OI concentrations and TOTN.API concentrations. For all sites, the values of API and OI methods were determined to be significantly different with both tests when using a conservative threshold of $p<0.10$.

Table 10. 2006 total nitrogen t-test and Wilcoxon sign rank test results OI vs API Methods.

Monitoring Program	Paired t-test p-value	Wilcoxon signed rank p-value
Streams	<0.001** (n=224)	0.0196* (n=224)

Monitoring Program	Paired t-test p-value	Wilcoxon signed rank p-value
Major Lakes	0.0810 (n=174)	0.0473* (n=174)
County Lakes	0.0352* (n=250)	0.0945 (n=250)
All Programs	0.0029** (n=648)	0.0894 (n=648)

* Significant (p<0.05)
** Highly Significant (p<0.01)

The original correction equations provided by KCEL for the freshwater monitoring programs are presented in Table 11 (see Appendix B), in addition to the results of the paired difference tests comparing observed and predicted values. The observed and predicted values were not found to be statistically different with any test.

Table 11. 2006 total nitrogen t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values.

Monitoring Program	Equation	Paired t-test p-value	Wilcoxon signed rank p-value
Streams	$\boldsymbol{TOTN.OI = 0.980 * TOTN.API^{1.077}}$	0.9173 (n=224)	0.7554 (n=224)
Major Lakes	$\boldsymbol{TOTN.OI = 1.01 * TOTN.API}$	0.8391 (n=174)	0.7902 (n=174)
County Lakes	$\boldsymbol{TOTN.OI = 0.994 * TOTN.API^{1.01}}$	0.8781 (n=250)	0.7524 (n=250)
All Programs		0.9536 (n=648	0.5575 (n=648)

* Significant (p<0.05)
** Highly Significant (p<0.01)

Figure 5 displays a plot of the TOTN.API and TOTN.OI observations for all freshwater programs after censoring the data. As seen in Figure 5, for all observed values there is an apparent, although subtle, departure from the 1:1 ratio with the API instrument reading higher values than OI. The results in Table 10 concur with this distinction.

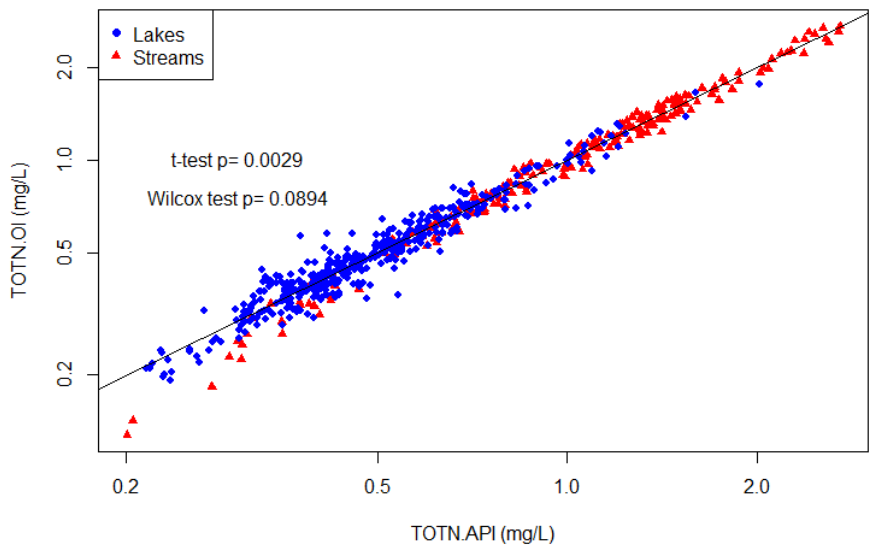

Figure 5. Comparison of TOTN.API and TOTN.OI for all data pairs above the MDL threshold. The solid line is 1:1 ratio. Note the log-scale.

5.3 Regression Results

Breakpoints in the iterative R script were examined at 0.001 ppm intervals. The R script found that there are not any significant breakpoints. The resulting regression equation is:

$$TOTN.OI = 0.9954 * TOTN.API^{1.008}$$

Figure 6 displays the plot of the fitted robust segmented model with no breakpoints, as well as the R-squared value and paired difference tests' p-values. Both tests failed to find a significant difference between observed and predicted TOTN.OI values at the α=0.05 level of significance, but the Wilcoxon test p-value is less than 0.10 and may be considered too low. Since the breakpoints found by the one- and two-breakpoint modeling exercise were determined not to be significant, the no-breakpoint model is the prefered model.

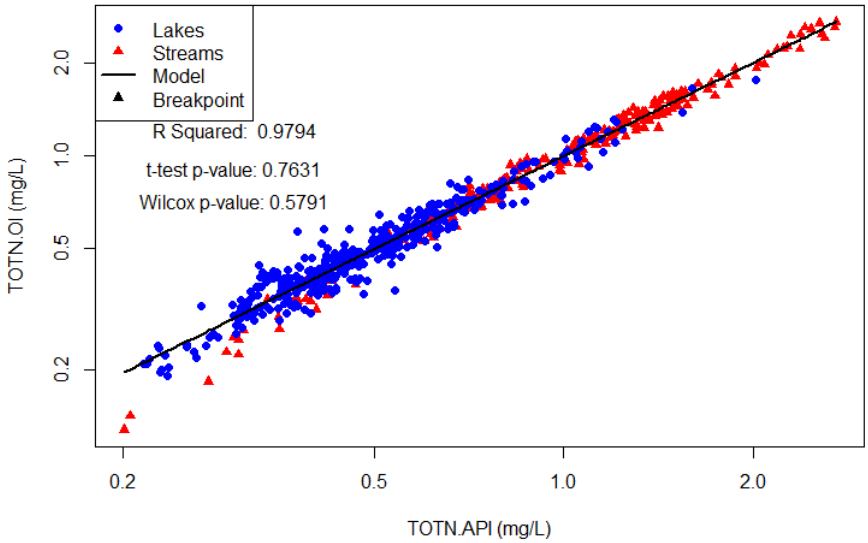

Figure 6. No-breakpoint model for TOTN.OI regressed on TOTN.API.Note the log-scale.

5.4 Heteroscedasticity

The high level of variation throughout observed pairs suggests low precision for one or both of the instruments and/or methodologies. A potential cause may be the aliquot volume used in the digestion process; there may be variation in nitrogen content between 30 mL aliquots, especially when considering likely patchiness and particle size distributions of phytoplankton biomass and other suspended solids. This may also explain the heteroscedasticity (i.e., the difference in variability) seen between lakes and streams.

To better understand the variability of the two monitoring programs, the root mean squared residuals were investigated. The root mean squared residual in the one-breakpoint model for Streams and Lakes samples were 0.084 and is 0.095, respectively. The Bartlett test for variance homogeneity was applied, and the null hypothesis of equal variance between sites was rejected, which means that there is a significant difference in the variation of Streams and Lakes program values. In spite of this, heteroscedasticity does not bias the regression coefficient estimates; it only biases their standard errors. Therefore, the no-breakpoint regression model efficiently predicts the impact of the 2006 instrument change.

5.5 Dilution of Samples

Because the wide analytical range for total nitrogen concentrations, dilution is rarely found to be necessary. The current SOP states that dilution is necessary at concentrations greater than 4.4 ppm. Additionaly, due to the combined digestion process, dilution because of high total phosphorus concentrations will also dilute the total nitrogen sample, even when the value for nitrogen is within the range. The value produced for the diluted sample should be corrected using the regression model and then scaled up with respect to its dilution. If the

correction is applied to the concentration after the dilution factor is applied, the corrected value will be slightly skewed.

For example, a sample is diluted 5x and the resulting analysis using the API instrutment yields a TOTN concentration of 0.50 ppm. The value reported in LIMS would therefor be 2.50 ppm (0.50 ppm* 5). If the reported value is corrected using the no-breakpoint model to be analagous to the results of the API instrument, the corrected value is 2.51 ppm, but if the value from the diluted sample is corrected and then the value is multiplied by 5, the result is 2.41 ppm. Table 12 provides several examples of this issue. The impact of this scaling issue does not appear to be substantial.

Table 12. Impact of dilution on the results of applying the correction factor.

Dilution	Diluted Result	Reported Value	Reported Value Corrected	Diluted Value Corrected and Scaled-Up	Relative Percent Difference
1x	0.50	0.50	0.49	0.49	0.0
5x	0.50	2.50	2.51	2.47	1.6
10x	0.50	5.00	5.04	4.95	1.8
20x	0.50	10.00	10.13	9.90	2.3

6.0 2006 FRESHWATER ORTHOPHOSPHATE INSTRUMENT CHANGE ANALYSIS

6.1 Data Used

The data used for analysis were taken from the original 2006 side-by-side analysis. These included 180 pairs from the Stream monitoring program, 183 pairs from the Major Lakes monitoring program, and 126 pairs from the County Lakes monitoring program. After combining and censoring the data limit (MDL=0.002 ppm), 275 pairs (56%) from the freshwater monitoring programs remained, including nearly all the Streams samples (179) and 32% of Lakes samples.

Since orthophosphate is likely to be found at substantial concentrations (i.e., above the detection limit) in lakes during periods of low productivity (e.g., winter) or in deep waters, the censored lake orthophosphate dataset does not represent all seasons and all depths. This will not influence the results of the regression model.

6.2 Normality and Paired Difference

The summary statistics for the OI (ORTHOP.OI) and API method (ORTHOP.API) values for orthophosphate allude to the non-normality of the data when considering the mean and median values (Table 13). The Shapiro-Wilk test for normality found that the County Lakes and Streams data were not normal, rejecting the null hypothesis. Log-transformation of the datasets yielded more normal distributions, according to Q-Q plots, but the Shapiro-Wilk null hypothesis was rejected for the Major Lakes log-transformed dataset. The Streams and County Lakes datasets are statistically log-normally distributed, whereas the Major Lakes datasets is statistically normally distributed. When all datasets are combined, the distribution is statistically log-normal. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 13. Summary statistics for 2006 orthophosphate values.

Monitoring Program	ORTHP.OI				OTHOP.API			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	0.0257	0.0208	0.0023	0.1350	0.0265	0.0224	0.0043	0.1222
Major Lakes	0.0053	0.0053	0.0020	0.0096	0.0074	0.0075	0.0021	0.0138
County Lakes	0.0245	0.0060	0.0024	0.1921	0.0280	0.0103	0.0021	0.2332
All Programs	0.0206	0.0114	0.0020	0.1921	0.0220	0.0141	0.0021	0.2332

All values in ppm.

Table 14 displays the log-transformed t-test and Wilcoxon sign rank test p-values and sample sizes for the comparison between observed TOTP.OI concentrations and TOTP.API

concentrations; due to its normal distribution, the t-test for Major Lakes was applied to non-transformed data. Due to the log-transformation, some negative concentrations in the uncensored datasets were dropped. For all sites, the values of API and OI methods were determined to be significantly different with both tests.

Table 14. 2006 orthophosphate t-test and Wilcoxon sign rank test results OI vs API Methods.

	Uncensored		Censored	
Monitoring Program	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams	<0.001** (n=180)	<0.001** (n=180)	<0.001** (n=179)	<0.001** (n=179)
Major Lakes	<0.001** (n=183)	<0.001** (n=183)	<0.001** (n=67)	<0.001** (n=67)
County Lakes	<0.001** (n=111)	<0.001** (n=111)	0.0021** (n=29)	0.0044** (n=29)
All Programs	<0.001** (n=473)	<0.001** (n=473)	<0.001** (n=275)	<0.001** (n=275)

* Significant (p<0.05)
** Highly Significant (p<0.01)

The original correction equations provided by KCEL for the freshwater monitoring programs are presented in Table 15 (see Appendix B), in addition to the results of the paired difference tests comparing observed and predicted values. Due to its normal distribution, t-tests for the Major Lakes were performed on non-transformed data.

Table 15. t-test and Wilcoxon sign rank test results for original KCEL equation predictions vs observed values.

		Uncensored		Censored	
Monitoring Program	**Equation**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams[1]	$\boldsymbol{ORTHOP.OI = 1.492 * ORTHOP.API^{1.13}}$	0.8293 (n=180)	0.3360 (n=180)	0.7363 (n=179)	0.3263 (n=179)
Major Lakes	$\boldsymbol{ORTHOP.OI = -0.00018 + 0.706 * ORTHOP.API}$	0.9589 (n=183)	0.5569 (n=183)	0.0286* (n=67)	0.0015** (n=67)
County Lakes	$\boldsymbol{ORTHOP.OI = -0.0013 + 0.884 * ORTHOP.API}$	<0.0001** (n=93)	0.5411 (n=111)	0.8351 (n=29)	0.5504 (n=29)
All Programs		<0.0001** (n=456)	0.1855 (n=456)	0.1998 (n=274)	0.0323* (n=274)

* Significant (p<0.05)
** Highly Significant (p<0.01)
1. KCEL Stream equation in December 15, 2006 memo has a typo, reading:
ln(ORTHOP-OI) = 0.4 = 1.31 * ln(ORTHOP-API)

The censoring of data below the detection limit had a tremendous impact on both lakes programs; Major and County Lakes had 80.3% and 88.9% of observations censored, respectively. The observed and predicted values above the MDL were found to be statistically different for the Major Lakes program using both tests, and the censoring of the

County Lakes increased the t-test p-value, which is likely due to the presence of some substantial outliers.

Figure 7 displays a plot of the ORTHOP.API and ORTHOP.OI observations for all freshwater programs after censoring the data. As seen in Figure 7, beginning at ORTHOP.API ≅ 0.010 ppm, there is an apparent departure from the 1:1 ratio with decreasing values. Additionally at these lower values, there is much more variability between the values produced by the two instruments, suggesting sensitivity issues with one or both of the instruments.

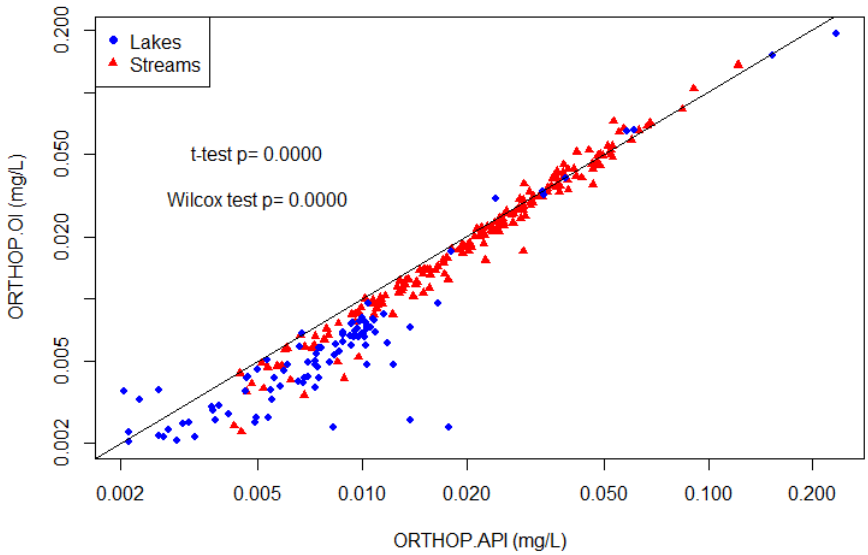

Figure 7. Comparison of ORTHOP.API and ORTHOP.OI for all data pairs above the MDL threshold.The solid line is 1:1 ratio. Note the log-scale.

6.3 Regression Results

Breakpoints in the iterative R script were examined at 0.005 ppm intervals. The R script found that the insertion of breakpoints at 0.011 and 0.042 ppm allowed for regression predictions that resulted in the best fitting model. The breakpoints yielded a linear model as follows:

When:

ORTHOP.API< 0.012 ppm $$ORTHOP.OI = 0.5043 * ORTHOP.API^{0.9172}$$

ORTHOP.API≥ 0.012 ppm
ORTHOP.API< 0.042 ppm $$ORTHOP.OI = 1.691 * ORTHOP.API^{1.160}$$

ORTHOP.API≥ 0.042 ppm $$ORTHOP.OI = 1.069 * ORTHOP.API^{1.018}$$

Figure 8 displays the plot of the fitted robust segmented model with two breakpoints, as well as the R-squared value and paired difference tests' p-values. Both tests failed to find a significant difference between observed and predicted TOTP. OI values at the α=0.05 level

of significance. Therefore, the model adequately corrects the bias introduced via the instrument change.

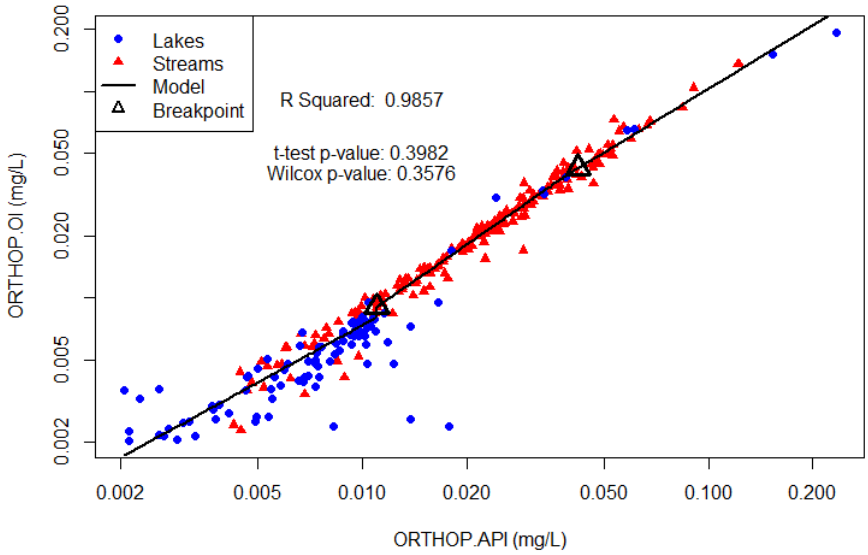

Figure 8. Two-breakpoint model for ORTHOP.OI regressed on ORTHOP.API. Note the log-scale.

6.4 Heteroscedasticity

To better understand the heteroscedasticity between the Streams and Lakes programs, the mean squared residuals. The mean squared residual in the two-breakpoint model for Streams and Lakes samples are 0.130 and is 0.345, respectively. The Bartlett test for variance homogeneity was applied, and the null hypothesis of equal variance between sites was rejected, which means that there is a significant difference in the variation of Streams and Lakes program values. In spite of this, heteroscedasticity does not bias the regression coefficient estimates; it only biases their standard errors. Therefore, the segmented regression model efficiently predicts the impact of the 2006 instrument change.

6.5 Dilution of Samples

The current SOP for orthophosphate states that values greater than 0.176 ppm should be diluted. It may be necessary to correct the measurement produced for the diluted sample using the regression model and then scaled up with respect to its dilution. If the correction is applied to the concentration after the dilution factor is applied, the corrected value will be inacurately skewed.

For example, a sample is diluted 5x and the resulting analysis using the API instrument yields an ORTHOP concentration of 0.150 ppm. The value reported in LIMS would therefore be 0.750 ppm (0.150 ppm* 5). If the reported value is corrected using the two-breakpoint model to be analogous to the results of the API instrument, the corrected value is 0.797 ppm, but if the value from the diluted sample is corrected and then the value is multiplied by 5, the result is 0.774 ppm. Table 16 provides several examples of this problem. For

converting ORTHOP.API to ORTHOP.OI, correcting reported values rather than correcting diluted values and scaling up will lead to exaggerated concentration estimates.

Table 16. Impact of dilution on the results of applying a correction factor. (2-BP: two-breakpoint)

Dilution	Diluted Result	Reported Value	Reported Value Corrected (2-BP Model)	Diluted Value Corrected (2-BP Model) and Scaled-Up	Relative Percent Difference
1x	0.150	0.150	0.155	0.155	0.0
5x	0.150	0.750	0.797	0.774	2.9
10x	0.150	1.50	1.62	1.55	4.4
20x	0.150	3.00	3.27	3.10	5.3
1x	0.015	0.015	0.013	0.013	0.0
5x	0.015	0.075	0.076	0.065	15.6
10x	0.015	0.150	0.155	0.130	17.5
20x	0.015	0.300	0.314	0.259	19.2

7.0 2006 FRESHWATER NITRATE/NITRITE INSTRUMENT CHANGE ANALYSIS

7.1 Data Used

The data used for analysis were taken from the original 2006 side-by-side analysis. These included 224 pairs from the Stream monitoring program, 203 pairs from the Major Lakes monitoring program, and 250 pairs from the County Lakes monitoring program. After combining and censoring the data (MDL=0.020 ppm), 483 pairs (98%) from freshwater monitoring programs remained.

7.2 Normality and Paired Difference

The summary statistics for the OI (NO23.OI) and API method (NO23.API) values for nitrate/nitrite illustrate the non-normality of the data when considering the mean and median values (Table 17). The Shapiro-Wilk test for normality found that the datasets are not normal, rejecting the null hypothesis. Log-transformation of the datasets yielded more normal distributions, according to Q-Q plots, but the Shapiro-Wilk null hypothesis was rejected for the log-transformed datasets. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 17. Summary statistics for 2006 nitrate/nitrite values.

Monitoring Program	NO23.OI				NO23.API			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	0.8819	0.7353	0.0935	2.759	0.8572	0.7089	0.0941	2.691
Major Lakes	0.2801	0.2395	0.0200	4.267	0.2743	0.2330	0.0220	3.981
County Lakes	0.2386	0.2090	0.0240	0.7400	0.2276	0.1960	0.0220	0.7460
All Programs	0.5412	0.3090	0.0200	4.267	0.5263	0.3070	0.0220	3.981

All values in ppm.

Table 18 displays the t-test and Wilcoxon sign rank test p-values and sample sizes for the comparison between observed NO23.OI values and NO23.API values. For all sites, the values of API and OI methods were determined to be significantly different with one or both of the tests.

Table 18. 2006 nitrate/nitrite t-test and Wilcoxon sign rank test results OI vs API Methods.

	Uncensored		Censored	
Monitoring Program	Paired t-test p-value	Wilcoxon signed rank p-value	Paired t-test p-value	Wilcoxon signed rank p-value
Streams	<0.001** (n=214)	<0.001** (n=214)	<0.001** (n=214)	<0.001** (n=214)
Major Lakes	<0.001** (n=219)	<0.001** (n=219)	0.2286 (n=204)	<0.001** (n=204)

	Uncensored		Censored	
Monitoring Program	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
County Lakes	<0.001** (n=65)	<0.001** (n=65)	<0.001** (n=65)	<0.001** (n=65)
All Programs	0.2587 (n=494)	<0.001** (n=494)	<0.001** (n=483)	<0.001** (n=483)

* Significant (p<0.05)
** Highly Significant (p<0.01)

The original correction equations provided by KCEL for the freshwater monitoring programs are presented in Table 19 (see Appendix B), in addition to the results of the paired difference tests comparing observed and predicted values. The censoring of data below the detection limit did not have an apparent impact on any of the sampling programs. The observed and predicted values were found to be statistically different for the Major Lakes and all programs with the log-transformed t-test.

Table 19. 2006 nitrate/nitrite t-test and Wilcoxon signed rank test results for original KCEL equation predictions vs observed values.

		Uncensored		Censored	
Monitoring Program	**Equation**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams	$\boldsymbol{NO23.OI = 1.03 * NO23.API^{1.004}}$	0.6091 (n=214)	0.2048 (n=214)	0.6091 (n=214)	0.2048 (n=214)
Major Lakes	$\boldsymbol{NO23.OI = -0.011 + 1.06 * NO23.API}$	0.0137* (n=212)	0.0506* (n=212)	0.0002** (n=204)	0.1902 (n=204)
County Lakes	$\boldsymbol{NO23.OI = 1.051 * NO23.API^{1.012}}$	0.3108 (n=65)	0.6710 (n=65)	0.3108 (n=65)	0.6710 (n=65)
All Programs		0.0173* (n=491)	0.0347* (n=491)	0.0024** (n=483)	0.1006 (n=483)

* Significant (p<0.05)
** Highly Significant (p<0.01)
Note: Linear regressions reported in December 15, 2006 memo were incorrectly reported. Correct regressions using KCEL methods are reported here.

Figure 9 displays a plot of the NO23.API and NO23.OI observations for all freshwater programs after censoring 2% of the data pairs. As seen in Figure 9, for all observed values there is an apparent departure from the 1:1 ratio with the OI instrument reading slightly higher values than API.

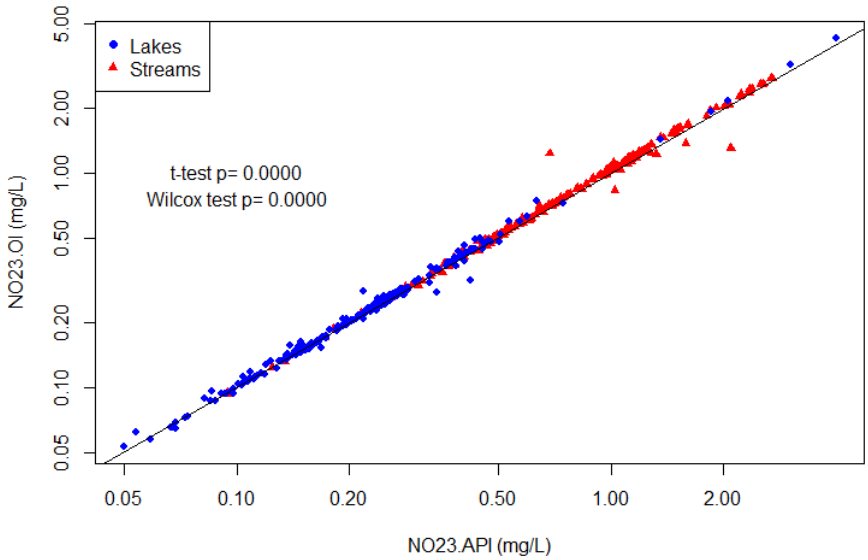

Figure 9. Comparison of NO23.API and NO23.OI for all data pairs above the MDL threshold. The solid line is 1:1 ratio. Note the log-scale.

7.3 Regression Results

Breakpoints for the regression model were examined at 0.001 ppm intervals. The R script found that a one-breakpoint, level linear model best fits the data. The breakpoint was located at 0.66 ppm. Log -log regressions failed to accurately predict NO23.OI concentrations for all breakpoint iterations. The breakpoints yielded a level linear model as follows:

When:

NO23.API< 0.66 ppm $$NO23.OI = -0.002 + 1.026 * NO23.API$$

NO23.API≥ 0.66 ppm $$NO23.OI = -0.026 + 1.066 * NO23.API$$

Figure 10 displays the plot of the fitted robust segmented model with no breakpoints, as well as the R-squared value and paired difference tests' p-values. Both tests failed to find a significant difference between observed and predicted NO23.OI values at the α=0.05 level of significance.

7.4 Heteroscedasticity

To better understand the heteroscedasticity between the two monitoring programs, the root mean squared residuals were investigated. The root mean squared residual in the two-breakpoint model for Streams and Lakes samples were 0.080 and is 0.015, respectively. The Bartlett test for variance homogeneity was applied, and the null hypothesis of equal variance between sites was rejected, which means that there is a significant difference in the variation of Streams and Lakes program values. In spite of this, heteroscedasticity does

not bias the coefficient estimates; it only biases their standard errors. Therefore, the segmented regression model efficiently predicts the impact of the 2006 instrument change.

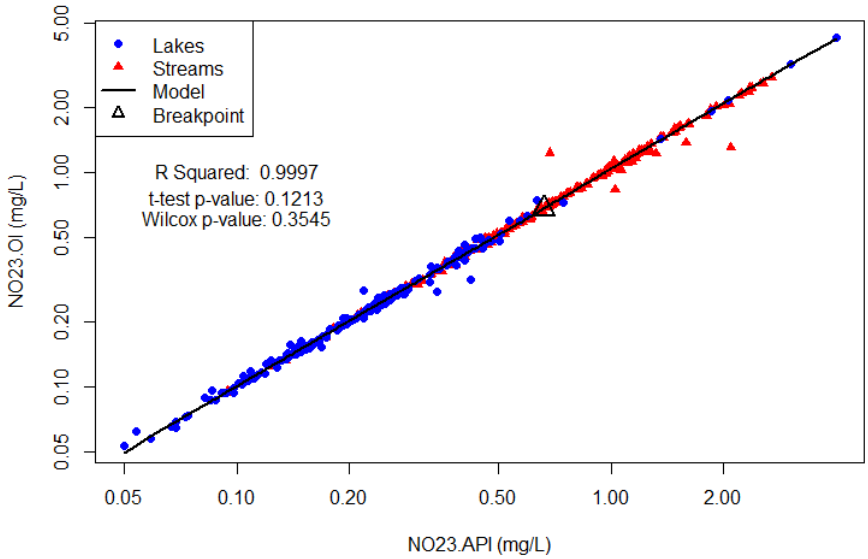

Figure 10. One-breakpoint linear model for NO23.OI regressed on NO23.API.Note the log-scale.

7.5 Dilution of Samples

The current SOP for nitrate/nitrite states that values greater than 1.76 ppm should be diluted. As discussed in section 2.2, the value produced for the diluted sample should be corrected using the regression model and then scaled up with respect to its dilution. If the correction is applied to the concentration after the dilution factor is applied, the corrected value will be inaccurately skewed.

For example, a sample is diluted 5x and the resulting analysis using the API instrument yields a NO23 concentration of 0.25 ppm. The value reported in LIMS would therefore be 1.25 ppm (0.25 ppm * 5). If the reported value is corrected using the one-breakpoint model to be analogous to the results of the API instrument, the corrected value is 1.31 ppm, but if the value from the diluted sample is corrected and then the value is multiplied by 5, the result is 1.27 ppm. Table 20 provides several examples of this issue. The impact of this scaling issue does not appear to be substantial.

Table 20. Impact of dilution on the results of applying the correction factor.

Dilution	Diluted Result	Reported Value	Reported Value Corrected	Diluted Value Corrected and Scaled-Up	Relative Percent Difference
1x	0.25	0.25	0.25	0.25	0.0
5x	0.25	1.25	1.31	1.27	3.1
10x	0.25	2.5	2.64	2.54	3.9
20x	0.25	5.0	5.30	5.09	4.0

8.0 2006 FRESHWATER AMMONIA INSTRUMENT CHANGE ANALYSIS

8.1 Data Used

The data used for analysis were taken from the datasets used in the original 2006 side-by-side analysis. These included 128 pairs from the Stream monitoring program, 96 pairs from the Major Lakes monitoring program, and 72 pairs from the County Lakes monitoring program. After combining and censoring the data (MDL=0.010 ppm), 157 pairs (80%) from the freshwater monitoring programs remained.

8.2 Normality and Paired Difference

The summary statistics for the OI (NH3.OI) and API method (NH3.API) values for ammonia allude to the non-normality of the data when considering the mean and median values (Table 21). The Shapiro-Wilk test for normality found that the datasets are not normal, rejecting the null hypothesis. Log-transformation of the datasets yielded more normal distributions, according to Q-Q plots, but the Shapiro-Wilk null hypothesis was rejected for the log-transformed datasets. Based on these results, it was determined that non-parametric statistical tests must be applied or that, to satisfy the assumptions inherent to parametric tests, log-transformation of the data was necessary.

Table 21. Summary statistics for 2006 ammonia values.

Monitoring Program	NH3.OI				NH3.API			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Streams	0.0323	0.0205	0.0106	0.1455	0.0332	0.0223	0.0102	0.1392
Major Lakes	0.0502	0.0170	0.0105	0.5658	0.0413	0.0181	0.0102	0.4221
County Lakes	0.6587	0.2302	0.0105	3.127	0.5760	0.2112	0.0103	2.894
All Programs	0.1652	0.0216	0.0105	3.127	0.1462	0.0223	0.0102	2.894

All values in ppm.

Table 22 displays the t-test and Wilcoxon sign rank test p-values and sample sizes for the comparison between observed NH3.OI concentrations and NH3.API concentrations. For nearly all sites, the values of API and OI methods were determined to not be significantly different with both tests.

The censoring of data below the detection limit had a substantial impact on some of the sampling programs. After censoring, the two methods were no longer statistically different when examining all freshwater programs. Statistical difference was detected for censored streams data using the paired t-test and for censored county lakes using the Wilcoxon signed rank test. Both tests found p-values less than 0.10 for censored major lakes data. Figure 11 displays a plot of the natural log of NH3.API and NH3.OI observations for all freshwater programs after censoring the data. As seen in Figure 11, the observed values

appear to follow the 1:1 ratio until NH3.API≅ 0.135 ppm, where the OI instrument appears to read higher values than the API.

Table 22. 2006 ammonia t-test and Wilcoxon sign rank test results OI vs API Methods.

	Uncensored		Censored	
Monitoring Program	**Paired t-test p-value**	**Wilcoxon signed rank p-value**	**Paired t-test p-value**	**Wilcoxon signed rank p-value**
Streams	0.0002** (n=128)	0.0008** (n=128)	0.0056** (n=79)	0.1255 (n=79)
Major Lakes	0.6138 (n=96)	0.8922 (n=96)	0.0718 (n=46)	0.0602 (n=46)
County Lakes	0.0126* (n=72)	0.6446 (n=72)	0.3169 (n=32)	0.0065** (n=32)
All Programs	0.0118* (n=296)	0.0655 (n=296)	0.4510 (n=157)	0.1574 (n=157)

* Significant (p<0.05)
** Highly Significant (p<0.01)

8.3 Regression Results

Since the paired difference tests failed to reject the null hypothesis for the freshwater matrix, it is not recommended that a correction factor be implemented for ammonia.

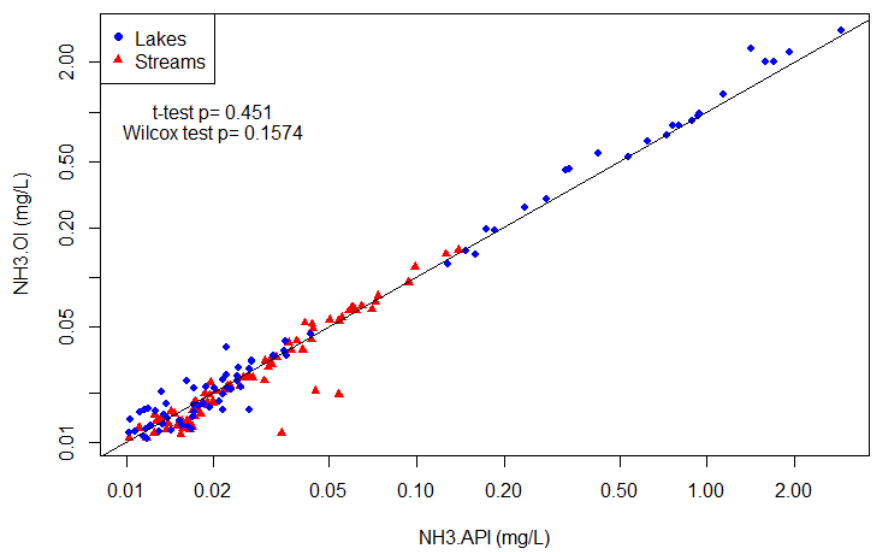

Figure 11. Comparison of NH3.API and NH3.OI for all data pairs above the MDL threshold. The solid line is 1:1 ratio. Note the log-scale.

9.0 TREND ANALYSIS

For each parameter assessed in this analysis, long-term trends were evaluated using uncorrected data, data corrected using the original correction factors, and data corrected using the newly calculated correction factors. Trend analysis was completed for A522 (Lake Union), 0826 (Lake Washington), 3106 (Green River), 0631 (Issaquah Creek), and A708 (Pine Lake). These sites were chosen to represent the major lakes, small lakes, large rivers, and streams. Long-term trends were analyzed using the Seasonal Mann-Kendall trends test on an annual basis, unless noted otherwise.

There were some limitations in which trends could be computed due to data gaps or a high frequency of non-detects. Ammonium trends were not analyzed because no correction factors were recommended and the high rate of non-detects prevent robust trend analysis. Long-term monitoring data for dissolved nutrients (orthophosphate, nitrate-nitrite, and ammonium) were not available for Pine Lake or any other small lake monitored by King County. Due to the high presence of non-detects for dissolved nutrients at the lakes stations (A522 and 0826), only total nitrogen and total phosphorus trends were calculated for the epilimnion (surface 10 m) over an entire year; orthophosphate trends were calculated for summer hypolimnetic concentrations (14 m for A522 and below 40 m for 0826), and nitrate/nitrite trends were calculated for winter epilimnetic concentrations. For site 3106, it should be noted that the removal of the Renton Wastewater Treatment Plant outfall in 1987 caused a step decrease in the concentrations of nutrients detected downstream. The results of the trend analyses are presented in the following section and summarized below.

The trend analyses using the different correction factors resulted in different slopes and levels of significance. It is important to note that no applications of a correction factor yielded a significant trend when the uncorrected data produced no trend, and all but two did not diminish the significance (at $\alpha=0.05$) of a trend. The two exceptions are the total phosphorus trends at sites 0631 and 3106, where the trends lost significance after the application of the original ("old") correction factor (the new factors did not cause this loss of significance).

The sampling locations used as test cases for this analysis typically had either very strong trends ($p<0.01$) or none at all ($p>0.10$) regardless of the correction factor applied. This does not rule out the possibility that these correction factors may play an important role in borderline cases, where trends are less evident or more subtle, i.e., the p-value is near the selected value of α.

9.1 Total Phosphorus

A stepwise decrease in total phosphorus at site 3106 was observed from 1986 to 1987, where values dropped from medians of ~0.20 to 0.05 ppm. This is due to the 1987 removal of the Renton Wastewater Treatment Plant outfall from the Green River just upstream of site 3106. The time periods evaluated are presented in Table 23.

Table 23. Total phosphorus trend data time period and percent censored.

Site	Description	Time Period	% Censored
A522	Lake Union	1985-2013	0.6
0826	Lake Washington	1994-2013	3.8
3106	Green River	1975-2013	0.0
0631	Issaquah Creek	1979-2013	0.0
A708	Pine Lake	1983-2013	3.4

Total phosphorus trends were decreasing at all sites for all correction factors (Table 24). These trends were all significant at the α=0.05 level, except for the original correction at 0631 and 3106, which yielded p-values of 0.0756 and 0.2010, respectively.

Generally, p-values were highest and the absolute values of Tau and the slope were lowest for trend analysis using the original correction. The new correction factor also curbed the magnitude of the uncorrected trend but not as substantially as the original correction factor. At sites 3106 and 0631, the old correction factor caused the trend to lose statistical significance and lowered the slope nearly fivefold relative to the uncorrected trend.

Table 24. Total phosphorus trends comparisons

Site	Correction Factor	Tau	p-value	Sen Slope (ppb/year)
A522	Uncorrected	-0.304	<0.0001	-0.260
	Old correction	-0.136	0.0003	-0.108
	New correction	-0.146	0.0001	-0.111
0826	Uncorrected	-0.299	<0.0001	-0.236
	Old correction	-0.140	0.0016	-0.108
	New correction	-0.180	<0.0001	-0.122
3106	Uncorrected	-0.543	<0.0001	-3.85
	Old correction	-0.455	<0.0001	-3.20
	New correction	-0.494	<0.0001	-3.64
0631	Uncorrected	-0.279	<0.0001	-0.359
	Old correction	-0.062	0.0756	-0.071
	New correction	-0.115	0.0009	-0.134
A708	Uncorrected	-0.494	<0.0001	-0.323
	Old correction	-0.450	<0.0001	-0.289
	New correction	-0.372	<0.0001	-0.200

9.2 Total Nitrogen

Total nitrogen joined the suite of nutrients routinely measured in freshwater in April 1993. The time periods evaluated are presented in Table 25. Total nitrogen trends were significantly decreasing at all sites, except 3106 (Table 26). With the introduction of the proposed correction factor, the absolute values of Tau and the trend slope increased. This is due to the back-correction increasing values prior to 2007 and thereby increasing the relative change.

Table 25. Total nitrogen trend data time period and percent censored

Site	Description	Time Period	% Censored
A522	Lake Union	1993-2013	3.2
0826	Lake Washington	1994-2013	0.0
3106	Green River	1993-2013	0.0
0631	Issaquah Creek	1993-2013	0.0
A708	Pine Lake	1993-2013	0.0

Table 26. Total nitrogen trend comparison

Site	Correction Factor	Tau	p-value	Sen Slope (ppb/year)
A522	Uncorrected	-0.187	<0.0001	-1.95
	Old correction	No correction factor recommended		
	New correction	-0.212	<0.0001	-2.22
0826	Uncorrected	-0.177	<0.0001	-2.05
	Old correction	No correction factor recommended		
	New correction	-0.213	<0.0001	-2.34
3106	Uncorrected	-0.035	0.4996	-1.07
	Old correction	No correction factor recommended		
	New correction	-0.046	0.3287	-1.51
0631	Uncorrected	-0.513	<0.0001	-18.4
	Old correction	No correction factor recommended		
	New correction	-0.520	<0.0001	-18.6
A708	Uncorrected	-0.417	<0.0001	-3.35
	Old correction	No correction factor recommended		
	New correction	-0.441	<0.0001	-3.68

9.3 Orthophosphate

A stepwise decrease in orthophosphate at site 3106 was observed from 1986 to 1987, where values dropped medians of ~0.20 to <0.05 ppm. This is due to the 1987 removal of the Renton Wastewater Treatment Plant outfall from the Green River just upstream of site 3106. Trends for A522 and 0826 were evaluated for summer hypolimnetic concentrations due to the high number of non-detects in the remainder of the year and near the surface. The time periods evaluated are presented in Table 27.

Table 27. Orthophosphate trend data time period and percent censored

Site	Description	Time Period	% Censored
A522	Lake Union Summer Hypolimnion	1985-2013	5.4
0826	Lake Washington Summer Hypolimnion	1994-2013	9.3
3106	Green River	1980-2013	0.3
0631	Issaquah Creek	1979-2013	2.5

Orthophosphate trends were significantly decreasing at all sites, but at A522 the slope magnitude was slight (Table 28). With the introduction of the original and the proposed correction factor, the absolute values of Tau and the trend slope increased. This is due the back-correction increasing values prior to 2007. Generally, the new correction factor did not increase the magnitude of the trend as greatly as the original correction factor.

Table 28. Orthophosphate trend comparison

Site	Correction Factor	Tau	p-value	Sen Slope (ppb/year)
A522 Summer Hypolimnion	Uncorrected	-0.268	0.0005	-2.47
	Old correction	-0.314	<0.0001	-4.40
	New correction	-0.264	0.0006	-2.36
0826 Summer Hypolimnion	Uncorrected	-0.373	0.0001	-0.569
	Old correction	-0.535	<0.0001	-1.13
	New correction	-0.493	<0.0001	-0.725
3106	Uncorrected	-0.488	<0.0001	-1.05
	Old correction	-0.530	<0.0001	-1.10
	New correction	-0.530	<0.0001	-1.08
0631	Uncorrected	-0.393	<0.0001	-0.257
	Old correction	-0.465	<0.0001	-0.326
	New correction	-0.473	<0.0001	-0.321

9.4 Nitrate/Nitrite

Trends for A522 and 0826 were evaluated for winter epilimnetic concentrations due to the high number of non-detects in the remainder of the year and at depth. The time periods evaluated are presented in Table 29. Nitrate/nitrite trends were significantly decreasing at all sites, except 3106 (Table 30). With the introduction of the proposed correction factor, the absolute values of Tau and the trend slope decreased. This is due the back-correction decreasing values prior to 2007.

Table 29. Nitrate/nitrite trend data time period and percent censored

Site	Description	Time Period	% Censored
A522	Lake Union Winter Epilimnion	1985-2013	0
0826	Lake Washington Winter Epilimnion	1994-2013	0
3106	Green River	1970-2013	0.5
0631	Issaquah Creek	1979-2013	0.2

Table 30. Nitrate/nitrite trend comparison

Site	Correction Factor	Tau	p-value	Sen Slope (ppb/year)
A522 Winter Epilimnion	Uncorrected	-0.259	0.0008	-2.13
	Old correction	No correction factor recommended		
	New correction	-0.244	0.0016	-1.95
0826 Winter Epilimnion	Uncorrected	-0.367	0.0001	-4.46
	Old correction	No correction factor recommended		
	New correction	-0.354	0.0002	-4.16
3106	Uncorrected	-0.042	0.1736	-0.500
	Old correction	No correction factor recommended		
	New correction	-0.027	0.3836	-0.334
0631	Uncorrected	-0.325	<0.0001	-7.17
	Old correction	No correction factor recommended		
	New correction	-0.306	<0.0001	-6.25

10.0 SUMMARY AND RECOMMENDATIONS

In 1998 and 2007, KCEL made two substantial improvements in their methodology and instrumentation for analyzing freshwater nutrients. This document's analysis inspected the influence of these improvements on the values reported by KCEL and adequacy of the previously provided correction factors.

Previous analyses of these changes were revisited and an improved set of correction factors developed for comparing datasets before and after the changes. Through nonparametric statistical testing and segmented linear regression, a better fitting model for correcting nutrient method changes was created for most parameters. These correction factors yield improved comparison between data produced through different methods. The correction factors are summarized in Table 31.

It is recommended that, when possible, the correction factors calculated in this report be applied to the diluted value result rather than the reported value and then scaled up after correction. This is most important for total phosphorus and orthophosphate, while total nitrogen and nitrate/nitrite conversion will have some bias. Although dilution is relatively rare, phosphorus (total and orthophosphate) and ammonia samples from lake water taken below a thermocline, stormwater, and wastewater are occasionally (or frequently) diluted. The dilution factor used by KCEL may be found by comparing the sample MDL to the SOP MDL, e.g., if a sample MDL was 0.05 mg/L and the SOP MDL is 0.005 mg/L, then the dilution factor was 10. Additionally since 2009, KCEL has reported the dilution factor for every nutrient sample; these can be matched using the lab sample number.

A high level of variation was found throughout all observed pairs of total phosphorus in lake samples. A potential cause may be variation in the aliquot volume used. There may be patchiness in nutrient content between 30 mL aliquots, especially from phytoplankton biomass and other suspended solids throughout the year. The impacts of aliquot volume could be further investigated in future side-by-side analysis.

Trend analysis comparing uncorrected data and data adjusted with a correction factor determined that the correction factors can influence the significance and magnitude of trend statistics. It is therefore crucial that an accurate correction factor be applied when performing trend analysis and when comparing data produced by differing method. The correction factors produced in this study for total phosphorus, total nitrogen, orthophosphate, and nitrate/nitrite provide a means for such analysis.

Recommendations:

- These correction factors be used in the future when comparing to historic data.
- The lakes and streams data available to the public through the King County website be corrected based on these correction factors.

- An analysis similar to the analysis presented in this document be applied to the saline matrix for nitrate/nitrite, ammonia, and silica for the 2007 instrument change.
- Side-by-side analyses be performed by KCEL prior to all instrument and method changes. Study design can be coordinated between Science and Technical Support Section and KCEL to optimize sample size to achieve adequate statistical power.
- Method change memoranda continue to be provided by KCEL prior to the change occurring and the Science and Technical Support Section alerted of them.
- An analysis comparing nutrient concentrations based on sample aliquot volumes for lakes be completed using the current 30 mL volume, as well as 10, 50, and 100 mL aliquot volumes to determine the impact of aliquot volume on observed concentrations of digested nutrients.
- In LIMS (Laboratory Information Management System), dilution factors be provided in a column alongside the reported value rather than as a separate entry. Currently, dilution factors are entered as a separate parameter.

Table 31. Summary of correction factors for total and dissolved nutrients

Time Period for Applying Correction Factor	Parameter	Correction Factor
Prior to 7/1/1998	Total Nitrogen	No side-by-side data available
	Total Phosphorus	$When\ TOTP_{KCEL} < 0.024\ ppm$ $TOTP_{Corrected} = 1.776 * TOTP_{KCEL}{}^{1..203}$ $Else\ (TOTP_{KCEL} \geq 0.024\ ppm)$ $TOTP_{Corrected} = 0.9347 * TOTP_{KCEL}{}^{1.056}$
Prior to 1/1/2007	Ammonia	No correction factor necessary
	Nitrate/Nitrite	$When\ NO23_{KCEL} < 0.678\ ppm$ $NO23_{Corrected} = 0.002 + 0.9747 * NO23_{KCEL}$ $Else\ (NO23_{KCEL} \geq 0.678\ ppm)$ $NO23_{Corrected} = 0.024 + 0.9381 * NO23_{KCEL}$
	Orthophosphate	$When\ ORTHOP_{KCEL} < 0.0087\ ppm$ $ORTHOP_{Corrected} = 2.109 * ORTHOP_{KCEL}{}^{1.090}$ $Else\ When\ ORTHOP_{KCEL} < 0.0424\ ppm$ $ORTHOP_{Corrected} =$ $0.6358 * ORTHOP_{KCEL}{}^{0.8621}$ $Else\ (ORTHOP_{KCEL} \geq 0.0424\ ppm)$ $ORTHOP_{Corrected}$ $= 0.9366 * ORTHOP_{KCEL}{}^{0.9823}$
	Total Nitrogen	$TOTN_{Corrected} = 1.005 * TOTN_{KCEL}{}^{0.9921}$
	Total Phosphorus	$TOTP_{Corrected} = 1.224 * TOTP_{KCEL}{}^{1.031}$

$\#\#\#\#_{KCEL}$ is the value reported by the King County Environmental Laboratory
$\#\#\#\#_{Corrected}$ is the value corrected for instrument and method changes and prepared for trend analysis.
ppm- parts per million (milligrams per L)
TOTP- Total Phosphorus
TOTN- Total Nitrogen
N023- Nitrite/nitrate-nitrogen
ORTHOP- Orthophosphate

Because total phosphorus has two correction factors, it becomes necessary to double correct when comparing values from after 1/1/2007 with values before 7/1/1998. To do so, the 1998 correction must be applied first and then followed by the 2007 correction.

REFERENCES

Anderson, R. 2008. *Modern Methods for Robust Regression*. SAGE Publications. Thousand Oaks, California.

Gilbert, RO, 1987. Statistical methods for environmental pollution monitoring. Van Nostrand. Reinhold, New York.

Helsel, DR, and RM Hirsch. 1992. Statistical methods in water resources. Elsevier, New York.

Hirsch, RM, JR Slack, and RA Smith. 1982. Techniques for trend assessment for monthly water quality data. Water Resources Research 18: 107–121.

Koller, M, and WA Stahel. 2011. Sharpening Wald-type inference in robust regression for small samples, Computational Statistics and Data Analysis 55(8), 2504–2515.

Marchetto, A. 2014. rkt: Mann-Kendall test, Seasonal and Regional Kendall Tests. R package version 1.3. http://CRAN.R-project.org/package=rkt

Maronna, RA, RD Martin, VJ Yohai. 2006. *Robust statistics*. John Wiley & Sons, Chichester, West Sussex.

Maronna, RA, and VJ Yohai. 2000. Robust regression with both continuous and categorical predictors. Journal of Statistical Planning and Inference 89, 197–214.

Muggeo, VMR. 2003. Estimating regression models with unknown break-points. Statistics in Medicine 22: 3055–3071.

Rousseeuw, P, C Croux, V Todorov, A Ruckstuhl, M Salibian-Barrera, T Verbeke, M Koller, and M Maechler. 2015. robustbase: Basic Robust Statistics. R package version 0.92-3. URL: http://CRAN.R-project.org/package=robustbase

Rochon J, M Gondan, and M Kieser. 2012. To test or not to test: Preliminary assessment of normality when comparing two independent samples. BMC Medical Research Methodology 12:81.

Shapiro, SS and MB Wilk. 1965. Analysis of variance test for normality (complete samples). Biometrika 52: 591–611.

Todorov, V, and P Filzmoser. 2009. An Object-Oriented Framework for Robust Multivariate Analysis. Journal of Statistical Software, 32(3), 1-47.

Yohai, VJ. 1987. High breakdown-point and high efficiency estimates for regression. The Annals of Statistics 15, 642–65.

This page intentionally left blank.

Appendix A:

Total Nitrogen & Total Phosphorus Method Changes (1998-2007) – KCEL Memo

Subject: **Total Nitrogen & Total Phosphorus Method Changes (1998-2007)**

To: **KCEL Total Nitrogen (TOTN) & Total Phosphorus (TOTP) Data Users**

From: **Conventionals Chemistry Unit, King County Environmental Laboratory**

Date: **January 2007**

Introduction:

This narrative addresses significant changes to the King County Environmental Laboratory's Total Nitrogen (TOTN) and Total Phosphorus (TOTP) digestion and analytical methods. For TOTP and TOTN data from time spans that overlap July 1998 and / or January 2007, comparability may have been impacted by these method changes. The method changes implemented in July 1998 and January 2007 are discussed in detail. To the extent possible, statistical analysis has been utilized in an attempt to quantify the effects these changes may have had relative to the historical database. Samples from the projects listed below (Table 1) were utilized for the side-by-side analysis and subsequent statistical evaluation.

Table 1 – Project Names / Numbers for Side-by-Side Study

Project Name	LIMS Project Number(s)
Marine Ambient	421250
County Lakes	421195
Major Lakes	421235
Streams	421240

Overview:

Prior to July of 1998, TOTN and TOTP digestions were performed by two separate persulfate digestions. Subsequent analysis was also separate and performed manually on a spectrophotometer. Both digestion techniques were applied to saline as well as non-saline matrices.

Beginning July 1, 1998, a combined persulfate digestion technique (referred to as the "combined digestion") was employed allowing for the simultaneous digestion of both TOTP and TOTN in non-saline matrices. A single persulfate digestion (referred to as the "single digestion"), very similar to that used previous to July 1998, was maintained for saline matrices where only TOTP was reported. Analysis for both digestion techniques was automated via an OI Analytical segmented flow autoanalyzer. These method changes were initiated to provide greater batching and analytical efficiency necessary to meet the conventional lab's increasing total nutrients workload.

Beginning January 1, 2007 a modified combined persulfate digestion technique (referred to as "modified combined digestion") was introduced for non-saline matrices. This method improved upon the TOTP digestion efficiency of the previous combined digestion method. In addition, matrix matching was incorporated into the digestion utilized for saline

matrices to eliminate instrumentally derived biases. A new segmented flow autoanalyzer manufactured by Astoria Pacific International (API) was also introduced at this time to replace the aging OI Analytical unit that had been in place since 1998.

The table below (Table 2) lists the total nutrients method codes from 1998 through 2007. Please note that the progression of method codes cited from July of 1998 to December 2006 do not represent any significant changes to the digestion or analytical techniques that were applied during this time period. These changes capture updates to the SOP language and content intended to clarify operational details and strengthen QA/QC practices.

Table 2 – Method Code Changes Summary

Date Range	TOTP –saline samples	TOTP – non-saline samples	TOTN – all waters
Prior to 6/1998	SM4500-P-B, E	SM4500-P-B, E	Bachman and Canfield
7/98 to 5/99	SM4500-P-B, F	SM4500-P-B, F Modified	SM4500-N-D + SM4500-NO3-F
6/99 to 6/02	SM4500-P-B, F(03-03-013-000)	SM4500-P-B, F mod (03-03-013-000)	SM4500-N-C (03-03-013-000)
7/02 to 4/05	SM4500-P-B, F(03-03-013-001)	SM4500-P-B, F MOD (03-03-013-001)	SM4500-N-C (03-03-013-001)
4/05 to 12/06	SM4500-P-B, F(03-03-013-002)	SM4500-P-B, F MOD (03-03-013-002)	SM4500-N-C (03-03-013-002)
Starting 1/07	SM4500-P-B, F(03-03-013-003)S	SM4500-P-B, F(03-03-013-003)C	SM4500-N-C (03-03-013-003)C

1998 Method Changes and Implications

The combined digestion coupled with automated analysis via an OI Analytical Flow Solution III segmented flow autoanalyzer was introduced July 1, 1998. The implementation of this method was preceded by a method validation process that included a limited side-by-side evaluation. The results of this evaluation suggested that there was no statistically significant difference between the previous manual methods and the combined digestion method for either TOTN or TOTP.

Long term trend analysis for several freshwater monitoring programs conducted in 2003 identified a significant step decrease in TOTP concentrations occurring in 1998. No environmental factors could be isolated to account for the stepwise decrease. Based on the information available, it was considered most likely that the discrepancy was associated with the method change that occurred during 1998.

A more extensive method evaluation was undertaken beginning in October 2004 in an attempt to confirm and quantify the effects of the 1998 method change. Side-by-side analyses were conducted using select locators from the Major Lakes, Streams, and County Lakes monitoring programs. Samples were prepared using the single digestion and the combined digestion methods. It is important to note that analysis for both digestion methods were performed in an automated manner using the existing OI Analytical Flow Solution III segmented flow autoanalyzer.

Evaluation of the side-by-side data indicated a clear statistically significant difference between the single and combined digestion methods. This relationship was true for the locators associated with all three monitoring programs. In each case, data collected using the single digestion method consistently provided greater values. These findings were consistent with the 2003 long term trend analysis and strongly suggested that the step decrease in TOTP data occurring in 1998 was directly related to the implementation of the combined digestion method.

Regression data generated from the statistical evaluation provided a correlation between the data collected by the single and combined digestion methods. However, it must be noted that its applicability to correlating data collected prior to and after July 1998 is limited. The potential effects of instrumentation (spectrophotometer vs. autoanalyzer) were not addressed during this evaluation. The evaluation strictly compared the difference between the two digestion steps while applying the same automated determinative step to both methods. As a result, the correlation may not accurately reflect the true cumulative effect that the change in the digestion and the determinative step may have had. With these limitations in mind, a summary of the statistical evaluation has been provided in the table below (Table 3).

Table 3 - 2004 Statistical Evaluation Summary

Method	Monitoring Program	Statistical Regression	R^2	Stat. Sig. Difference @ 95% Confidence Level?
TOTP	Major Lakes, Streams	TOTP(single) = 1.262 * TOTP(combined)	0.989	Yes

TOTP(single) = total phosphorus data generated from the single digestion with automated analysis
TOTP(combined) = total phosphorus data generated from the combined digestion with automated analysis

2007 Method Changes and Implications

Continued method development work conducted from June 2005 through February 2006 isolated significant limitations to the existing digestion techniques (combined and single). Most notably, the combined digestion was observed to greatly underestimate inorganic forms of phosphorus. Additionally, the amount of other oxidizable materials (besides nitrogen and phosphorus) present in a sample was found to affect the level of phosphorus oxidation that could be achieved.

Alternatively, the single digestion (employed primarily for saline samples) was confirmed as being quite robust providing a very complete oxidation of the phosphorus present. However, the use of standards prepared in reverse osmosis (RO) water for seawater sample analysis was isolated as a source of bias at the colorimetric step. The dissolved salts present in seawater impede color development and thus lower overall instrument response associated with this matrix. When compared to a calibration curve developed in RO water, the resulting seawater sample concentrations exhibited a low bias.

Investigation of alternative digestion techniques was conducted concurrent to the evaluation of the existing in-house methodologies. The limitations and merits of multiple

approaches were considered in developing a method with improved digestion efficiency. The method that was developed maintains a combined digestion approach for TOTN and TOTP in non-saline matrices. However, the conditions under which the digestion is conducted are considerably more robust and can tolerate greater matrix complexity. A separate but similar digestion for TOTP in saline matrices is used in conjunction with matrix matching to eliminate bias at the instrument. This modified combined digestion coupled with automated analysis on an Astoria Pacific International Astoria2 segmented flow analyzer was introduced January 1, 2007.

This method change was preceded by a validation process that included side-by-side evaluation. Four projects were selected for this evaluation representing ongoing monitoring programs for which an extensive volume of historical data is available. This included the Major Lakes, Streams, County Lakes, and Marine Ambient monitoring programs. A summary of the statistical evaluation associated with this data has been provided in a table on the following page.

In the case of TOTN, no statistically significant difference was observed between the combined and modified combined methods for any of the monitoring programs where TOTN is routinely requested. TOTN is not a component of the Marine Ambient monitoring program and therefore was not included. The combined digestion was considered to be quite robust providing good nitrogen oxidation. Comparison of the data collected via the combined and modified combined digestion methods were in very good agreement confirming this assessment.

Speaking strictly in terms of TOTP t-test results, no statistically significant difference was observed between the combined and modified combined methods for the Streams monitoring program. This indicates that on average, the difference between the two methods is too small to be considered statistically significant. The Streams regression results backed-up the t-test data indicating no statistically significant difference with the exception of an isolated anomaly discussed below. However the Major Lakes regression results used in conjunction with the t-test data suggest the presence of a small low bias associated with the combined digestion method. The TOTP results associated with the County Lakes and Marine Ambient monitoring programs indicates the presence of a clear statistically significant low bias associated with the existing methods (combined digestion and single digestion respectively). The above results as a whole were consistent with expectations for method performance based on reasons discussed previously.

An isolated but reoccurring anomaly was observed for two specific Streams locators (0317: Springbrook Creek & A315: Mill Creek). These locators were observed to consistently provide results with significant differences between the combined and modified combined digestion methods. The number of data points associated with these two locators compared to the entire data set was insufficient to dramatically affect the overall statistical evaluation. However, removing these points from the data set noticeably improved the overall regression. Based on the consistent behavior of these two locators over several months of data collection, removing these data was considered appropriate. Therefore, although only a limited number of data points were available (n=11), a separate regression

was analyzed for these specific locators for which an acceptable correlation was obtained. A summary of the statistical evaluation resulting from the side-by-side study conducted in 2006 has been provided in the table below (Table 4).

Table 4 – 2006 Statistical Evaluation Summary

Method	Monitoring Program	Statistical Regression	R^2	Distribution	Stat. Sig. Difference @ 95% Confidence Level?
TOTN	Major Lakes	TOTN-OI = 1.01*(TOTN-API)	0.994	Unknown**	No
TOTP	Major Lakes	TOTP-OI = 0.00258 + 0.788*(TOTP-API)	0.762	Normal	Yes
TOTN	Streams	*ln*(TOTN-OI) = -0.02 + 1.077**ln*(TOTN-API)	0.987	Log-Normal	No
TOTP	Streams	*ln*(TOTP-OI) = -0.242 + 0.927**ln*(TOTP-API)	0.964	Log-Normal	No
TOTP	Streams: Locators: A315, 0317	*ln*(TOTP-OI) = -1.15 + 0.620**ln*(TOTP-API)	0.891	Log-Normal	Yes
TOTN	County Lakes	*ln*(TOTN-OI) = -0.006 + .1.01**ln*(TOTN-API)	0.946	Log-Normal	No
TOTP	County Lakes	*ln*(TOTP-OI) = -0.676 + 0.884**ln*(TOTP-API)	0.917	Log-Normal	Yes
TOTP	Marine Ambient	TOTP-OI = 0.0011 + 0.9122*(TOTP-API)	0.947	Unknown**	Yes

TOTN-OI = total nitrogen data generated on the OI Analytical Flow Solution III instrument using the combined digestion method
TOTN-API = total nitrogen data generated on the Astoria Pacific International Astoria2 instrument using the modified digestion method
TOTP-OI = total phosphorus data generated on the OI Analytical Flow Solution III instrument using the combined or single digestion method
TOTP-API = total phosphorus data generated on the Astoria Pacific International Astoria2 instrument using the modified digestion method
ln = natural log
** Data were determined to be neither normally nor log-normally distributed.

Appendix B:

Change in Instrumentation for Dissolved Nutrients Analyses – KCEL Memo

Subject: **Change in Instrumentation for Dissolved Nutrients Analyses**
To: **KCEL Ammonia (NH3), Nitrate (NO3), Nitrate-Nitrite (NO23), Nitrite (NO2), Ortho-Phosphate (ORTHOP), and Silica (SI) Data Users**
From: **Conventionals Chemistry Unit, King County Environmental Laboratory**
Date: **January 2007**

INTRODUCTION

This narrative addresses the change in instrumentation to the King County Environmental Laboratory's analyses of dissolved nutrients parameters - NH3, NO23 (singly or combined), ORTHOP, and SI – on both non-saline and saline matrices. The change in instrumentation implemented in January 2007 is discussed in detail. To the extent possible, statistical analysis has been utilized in an attempt to quantify the effects these changes may have relative to the historical database.

OVERVIEW

On January 1, 2007, KCEL implemented new instrumentation for analysis of dissolved nutrient parameters – NH3, NO23 (singly or combined), ORTHOP, and SI – on both non-saline and saline matrices. The current LIMS method codes associated with the above mentioned parameters were updated to reflect a revision change in the standard operating procedure (SOP) due to the change in instrumentation. There were no changes to the current method detection limits (MDL) and reporting detection limits (RDL) associated with any of the dissolved nutrient parameters.

Table 1 – List of Project Names / Numbers for Side-by-Side Study

Project Name	LIMS Project Number (s)
Marine Ambient	421250
County Lakes	421195
Major Lakes	421235
Streams	421240

Side-by-side analyses of samples from routine monitoring projects (see Table 1) were performed so that potential biases between the current instrumentation and the new instrumentation could be determined. Only the dissolved nutrient parameters routinely tested and reported for these projects were evaluated. The results are summarized below. For more detailed comparisons, see **Potential Impacts to Data**.

1. Marine Ambient Project: Results from the side-by-side study showed statistically significant differences between the current and new instrumentation for the NO23 and SI data. On average, the new API instruments yielded higher results for NO23 and SI. No statistical bias was observed on the NH3 data. ORTHOP analysis is currently not performed for this project.

2. County Lakes and Major Lakes Projects: Results from the study showed statistically significant differences between the current and new instrumentation for the ORTHOP data. On average, the new API instruments yielded higher results for ORTHOP. No statistical biases were observed on the NO23 and NH3 data for both projects nor was there any bias observed for the SI data for the Major Lakes project. SI analysis is currently not performed for the County Lakes project.

3. Streams Project: No statistical biases were observed for any of the dissolved nutrient data for this project.

Past Practices

In 1995, analyses for dissolved nutrient parameters – NH3, NO23 (singly or combined), and ORTHOP – on non-saline and saline matrices changed from Technicon auto-analyzers to the OI Flow Solution III auto-analyzer. In 1997, the use of a gas diffusion module with the OI Flow Solution auto-analyzers was implemented for analyses of ammonia for both non-saline and saline water matrices. In 1999, analysis of dissolved nutrients in saline water matrix was transferred to the OI Flow Solution IV auto-analyzer. Analysis of SI on saline matrices, with the Flow Solution IV, was also first implemented in 1999 and was expanded to include non-saline matrices in 2002.

Description of Changes

An extensive evaluation of autoanalyzer technologies begun in 2004 culminated in 2005 with the purchase of four multiple channel segmented flow auto-analyzers manufactured by Astoria Pacific International (API). The API systems replaced the existing OI Flow Solution systems utilized for dissolved nutrients (NH3, NO23 (singly or combined), ORTHOP, SI).

Beginning with samples collected on or after January 1, 2007, the LIMS method codes for each dissolved nutrient parameter reflected a revision change in the SOP #. Table 2 below shows the existing and new method codes for each:

Table 2 – Method Codes Summary

Parameter	**LIMS Method Code before 1/1/07**	**LIMS Method Code after 1/1/07**
NO23	SM4500-NO3-F (03-03-012-003)	SM4500-NO3-F (03-03-012-00**4**)
NO3	Same as above	Same as above
NO2	SM4500-NO2-B (03-03-012-003)	SM4500-NO2-B (03-03-012-00**4**)
NH3	SM4500-NH3-G (03-03-012-003)	SM4500-NH3-G (03-03-012-00**4**)
ORTHOP	SM4500-P-F (03-03-012-003)	SM4500-P-F (03-03-012-00**4**)
SI	Whitledge 1981 (03-03-012-003)	Whitledge 1981 (03-03-012-00**4**)

POTENTIAL IMPACTS TO DATA

To evaluate the potential effects that the new instrumentation may have on the historical database, side-by-side analyses were conducted from March through November 2006. Four projects were selected representing ongoing monitoring programs for which an extensive volume of historical data is available. This included the Major Lakes, Streams, County Lakes, and Marine Ambient projects. All data collected were provided to Tom Georgianna for statistical evaluation. A summary of the statistical evaluation is provided in Table 3 below:

Table 3 – Project Statistical Evaluation Summary

Parameter	Monitoring Program	Statistical Regression	R^2	Distribution	Stat. Sig. Difference @ 95% Confidence Level?
NO23	Marine Ambient	(NO23-OI) = -0.009 + 0.965 * (NO23-API)	0.976	Unknown**	Yes
NH3	Marine Ambient	(NH3-OI) = -0.0004 + 1.13 * (NH3-API)	0.987	Unknown**	No
SI	Marine Ambient	(SI-OI) = 0.082 + 0.904 * (Si-API)	0.923	Unknown**	Yes
NO23	Major Lakes	(NO23-OI) = -0.011 + 1.06 * (NO23-API)	0.999	Unknown**	No
NH3	Major Lakes	(NH3-OI) = -0.004 + 1.35 * (NH3-API)	0.997	Unknown**	No
ORTHOP	Major Lakes	(ORTHOP-OI) = -0.00018 + 0.706 * (ORTHOP-API)	0.896	Unknown**	Yes
SI	Major Lakes	(SI-OI) = 0.018 + 1.003 * (SI-API)	0.993	Unknown**	No
NO23	Streams	*ln*(NO23-OI) = -0.02 + 1.004 * *ln*(NO23-API)	0.990	Log-Normal	No
NH3	Streams	*ln*(NH3-OI) = -0.216 + 0.97 * *ln*(NH3-API)	0.873	Log-Normal	No
ORTHOP	Streams	*ln*(ORTHOP-OI) = 0.4 = 1.31 * *ln*(ORTHOP-API)	0.978	Log-Normal	No
SI	Streams	(SI-OI) = 0.954 + 0.918 * (SI-API)	0.990	Normal	No
NO23	County Lakes	*ln*(NO23-OI) = 0.05 + 1.01 * *ln*(NO23-API)	0.998	Log-Normal	No
NH3	County Lakes	(NH3-OI) = -0.008 + 1.17 * (NH3-API)	0.975	Unknown**	No
ORTHOP	County Lakes	(ORTHOP-OI) = -0.0013 + 0.884 * (ORTHOP-API)	0.959	Unknown**	Yes

NO23-OI = nitrite + nitrate nitrogen data generated on the OI Analytical instrument
NO23-API = nitrite + nitrate nitrogen data generated on the Astoria Pacific International instrument
NH3-OI = ammonia nitrogen data generated on the OI Analytical instrument
NH3-API = ammonia nitrogen data generated on the Astoria Pacific International instrument
ORTHOP-OI = orthophosphate phosphorus data generated on the OI Analytical instrument
ORTHOP-API = orthophosphate phosphorus data generated on the Astoria Pacific International instrument
SI-OI = silica data generated on the OI Analytical instrument
SI-API = silica data generated on the Astoria Pacific International instrument
ln = Natural Log
** Data were determined to be neither normally nor log-normally distributed.

Marine Ambient

The side-by-side data results for the Marine Ambient project showed statistically significant differences for both NO23 and SI. On average, the API system yielded higher results than the OI systems for NO23 and SI. The average concentrations and concentration ranges of the side-by-side data for NO23 and SI from the project are summarized in Table 4 below:

Table 4 – Marine Ambient Data Summary

Parameter	Average Concentration	Concentration Range
NO23-OI	0.282 mg/L	<0.02 – 1.64 mg/L
NO23-API	0.298 mg/L	

SI-OI	2.39 mg/L	<0.05 – 7.8 mg/L
SI-API	2.55 mg/L	

The performance of the API system for the duration of the side-by-side studies was very stable and consistent. The API system was not prone to drifts due to fluctuations in ambient temperature during SI analysis as compared with the OI Flow Solution IV. The cadmium coil used for NO23 analysis in saline matrices provided better reduction efficiency as verified by the recoveries from an efficiency check standard. This suggested that the comparatively higher API NO23 results, on the average, for analysis of saline matrices were potentially related to the improved reduction efficiency observed on the system.

Major Lakes and County Lakes

The side-by-side study data results for the Major Lakes and County Lakes projects showed statistically significant differences for ORTHOP. On average, the API system yielded higher ORTHOP results than the OI system for both projects. The majority of the ORTHOP data for both projects were < MDL or <RDL. The data suggest that the differences in results between the two systems could be related to the low ORTHOP levels associated with both projects. The API system's calibration curve for ORTHOP analysis utilized a lower calibration range and incorporated a second order fit to take advantage of the system's greater sensitivity. This provided for an overall better fit and sensitivity for determination of low level ORTHOP. The average concentrations and concentration ranges of the side-by-side data for ORTHOP from the Major Lakes and County Lakes project are summarized in Table 5 below:

Table 5 – Major Lakes and County Lakes Data Summary

Parameter	Average Concentration	Concentration Range
ORTHOP-OI_1	0.0025 mg/L	<0.002 – 0.0104 mg/L
ORTHOP-API_1	0.0037 mg/L	
ORTHOP-OI_2	0.0067 mg/L	<0.002 – 0.233 mg/L
ORTHOP-API_2	0.009 mg/L	

ORTHOP-OI_1 – Major Lakes

ORTHOP-OI_2 – County Lakes

In contrast, the average ORTHOP concentrations for the Streams data were 0.026 mg/L for both OI and API with a concentration range of <MDL to 0.135 mg/L. Only 16 data points were <MDL or <RDL.

Others

There were no statistically significant differences between the OI and API side-by-side study data with NH3 for all four (4) monitoring projects; for NO23 with Major Lakes, Streams, and County Lakes; for ORTHOP with Streams; and for SI with Streams and Major Lakes. This indicated that on average, the differences between the two systems for the above mentioned parameters and associated programs were too small to be considered statistically significant.

This page intentionally left blank.

Appendix C:

Description of Major Statistical Tests and the Iterative R Script

Paired difference tests

This analysis employed two paired difference tests to determine whether the application of a method/instrument change or of a correction factor resulted in significantly different values, e.g. if the switch from single to combined digestion resulted statistically distinguishable values or if a correction factor adequately nullifies the statistical difference between datasets. The two tests used were the t-test and the Wilcoxon signed rank test.

t-test

The t-test is a commonly used parametric statistical test with a null hypothesis that the means of the two populations are equal. When used for paired data, the null hypothesis is that the difference between pairs is zero. The t-statistic is calculated using the following equation:

$$t = \frac{\overline{X_D}}{s_D/\sqrt{n}}$$

where $\overline{X_D}$ is the average difference between paired samples, s_D is the standard deviation of pair differences, and n is the number of paired samples. The t-statistic is then compared to the Student's t-distribution to determine the probability that the difference between pairs has a population mean of zero.

A major assumption of the t-test is that the population is normally distributed, i.e., the data are distributed symmetrically about a mean value. Normal distribution is not commonly found in environmental phenomena. Environmental data may be log-normally distributed; taking the natural log of log-normally distributed data yield a normal distribution. Thus, the log-transformation of environmental data may allow the assumption of normal distribution to be met.

Wilcoxon signed rank test

The Wilcoxon signed rank test is a commonly used nonparametric statistical test that does not require an assumption for the distribution of a population. The null hypothesis of the Wilcoxon signed rank test is that the median difference between pairs is zero. For the Wilcoxon signed rank test, the difference between pairs is assigned a signed rank based on its distance from zero and whether the difference is positive or negative. The test statistic, W, is the absolute value of the sum of the signed ranks. With a sample size greater than ten, W converges to a normal distribution and can be compared with the distribution to determine the probability that the median difference is zero.

Shapiro-Wilk Test for normality and quantile-quantile plots

The Shapiro-Wilk test tests the null hypothesis that "the samples come from a normal distribution" against the alternative hypothesis "the samples do not come from a normal distribution". Since normality is being tested against, as the sample size increases, the likelihood that the test will determine the data to be non-normal increases. Therefore, the

Shapiro-Wilk test should be used cautiously and accompanied with additional tests for normality.

Quantile-quantile (Q-Q) plotting is a graphical method for comparing two probability distributions by plotting their quantiles against each other. In testing for normality, the quantiles from the samples are plotted against quantiles from a randomly generated, normally-distributed dataset. Figure C-1(a) displays a Q-Q plot for normally distributed dataset, and it can visually assessed that the data are normally distributed, i.e., close to the theoretical quantile line. Figure C-1(b) display a Q-Q plot for log-normally distributed data, and it is evident that data departs greatly from the theoretical quantile line. When the same data are log-transformed and Q-Q plotted, it is visually apparent that the data follow the theoretical quantile line more closely (Figure C-1(c)).

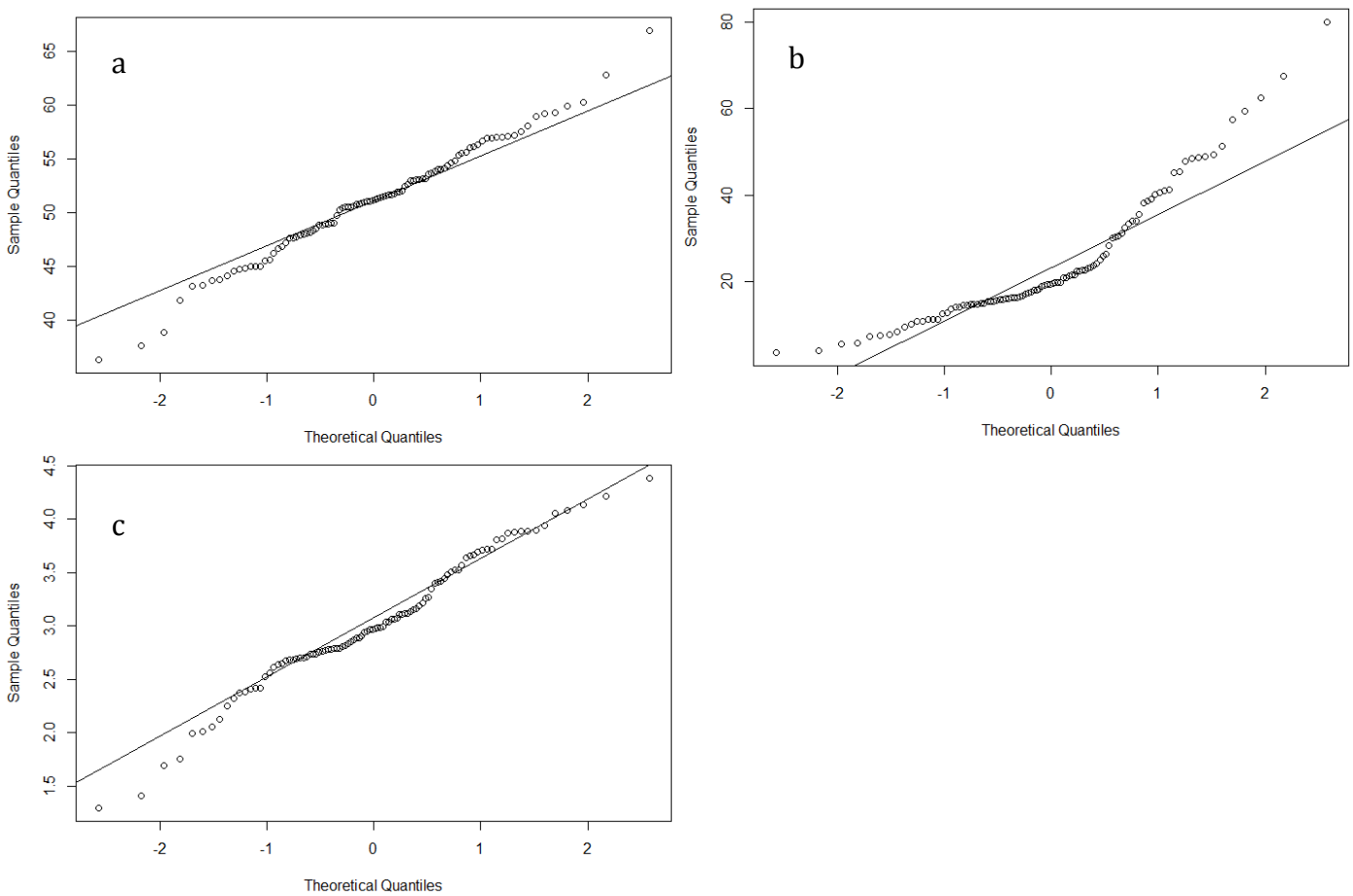

Figure C-1. Q-Q plots of (a) normally distributed data, (b) log-normally distributed data, and (c) log-transformed log-normally distributed data.

Seasonal Mann-Kendall test for trends

The Mann-Kendall test is a non-parametric test for identifying trends in time series data. The test compares the relative magnitudes of sample data rather than the data values

themselves. One benefit of this test is that the data need not conform to any particular distribution. Moreover, data reported as non-detects can be included by assigning them a common value that is smaller than the smallest measured value in the data set.

The data values are evaluated as an ordered time series. Each data value is compared to all subsequent data values. The initial value of the Mann-Kendall statistic, S, is assumed to be 0 (i.e.., no trend). If a data value from a later time period is higher than a data value from an earlier time period, S is incremented by 1. On the other hand, if the data value from a later time period is lower than a data value sampled earlier, S is decremented by 1. The net result of all such increments and decrements yields the final value of S.

This significance of the trend (i.e., the S statistic) is calculated by determining the variance in S. With a sample size greater than ten, S converges to a normal distribution and can be compared with the distribution to determine the probability that there is no trend.

The Seasonal Mann-Kendall is able to account for intra-annual variation. The test computes the Mann-Kendall S statistic for individual block groups; these block groups can be seasons, months, or another factor. The overall statistic is calculated by summing the individual month S statistics, and the overall S variance is calculated by summing the individual month S variances. See Hirsch et al. (1982) for further details. Sen's nonparametric estimator of slope is used to calculate the magnitude of trend when a significant linear trend is detected. Sen's slope is the median of the slopes of all lines through pairs of two-dimensional sample points. The 'rkt' script in the "rkt" R package was used to apply the Seasonal Mann-Kendall test for trends (Marcheto 2014).

To negate the impact of shifting method detection limits over time, all values below the maximum detection limit for the parameter were replaced with the value of the maximum detection limit.

Segmented (breakpoint) linear regression and the iterative R script

Segmented linear regression is method of linear regression where the independent variable is partitioned into intervals and line segments are fitted to each interval (Muggeo 2003). The boundaries between intervals are termed "breakpoints". Segmented regressions are used when there are apparent shifts in the relationship between the independent and dependent variables at discrete values. Figure C-2 provides a visual example of segmented regression versus normal linear regression; it can be seen that the non-segmented linear regression biases the impact of the independent variable on the dependent variable at values less than 40 and greater than 60 and underestimates the impacts at values between 40 and 60.

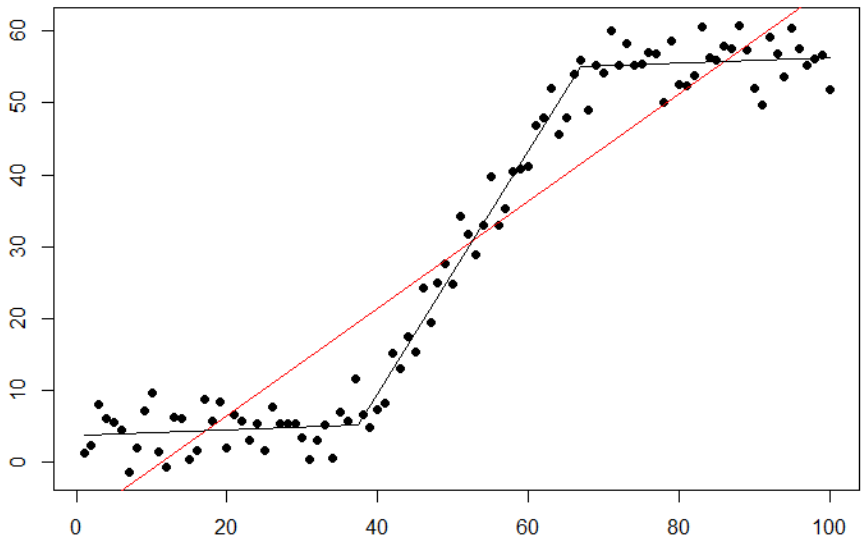

Figure C-2. Example of segmented regression best fit line (black line) and normal linear regression (red line).

Robust linear regression was used instead of ordinary least squares regression. While being the most common regression technique, ordinary least squares regressions are highly sensitive to outliers (Figure C-3). Robust linear regression, an alternative to least squares regression, was used because of its ability to handle outliers, such as non-normally distributed measurement errors (Maronna and Yohaj 2000; Todorov and Filzmoser 2009). The 'lmrob' script in the "robustbase" R package was used for applying robust linear regressions (Rousseeuw et al. 2015). This function computes an MM-type regression estimator as described in Yohai (1987) and Koller and Stahel (2011). See Maronna et al. (2006) and Anderson (2008) for more discussion of robust statistical analysis.

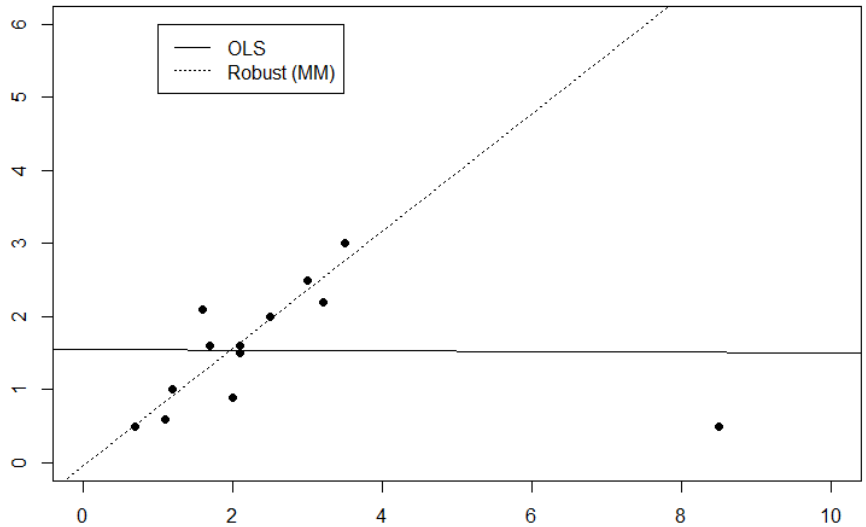

Figure C-3. Comparison of ordinary least squares (OLS) regression and robust, MM-type regression.

For the creation of correction factors, the iterative R script inserts one or two potential breakpoints into the robust linear model and regresses the pre-instrument/method change values on the post-instrument/method change values. For each resulting regression equation, the robust residual standard error and the statistical significance of the breakpoints are assessed. Breakpoints are assessed at intervals relative to the sample values and for ranges determined by the visual assessment of the plotted paired data. Depending on the number of breakpoints and potential breakpoint locations, between 90 and 2500 iterations of segmented regressions are run.

In some cases, the independent and the dependent variables may not have a linear relationship. For these cases, a log-log linear regression may be employed. In log-log linear regression, the natural log of the dependent variable is regressed on the natural log of the independent variable. If a level-level linear regression is employed, the residuals may not be normally distributed, which is a violation of an assumption inherent to linear regressions. The equations for the level-level and log-log regressions are presented below.

Level-level regression

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Log-log regression

$$\log(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 \log(X)$$

$$OR$$

$$\hat{Y} = e^{\hat{\beta}_0} * X^{\hat{\beta}_1}$$

After the completion of the iterative R script, the paired difference test statistics are evaluated of each iteration of the model. The final segmented regression model is determined by the minimization of the model's robust residual standard error and the statistical significance of the assessed breakpoints.

Heteroscedasticity

The Bartlett test for equal variances was used to assess the heteroscedasticity of the errors in the regression model based on which sampling program that data came from (e.g., Lakes vs. Streams). While heteroscedasticity does not affect the predictive regression model, this analysis allows the investigation of where there is inherent variability ("noise") after the application of a correction factor.