APPENDIX A

ENSR-NEIWPCC STATISTICAL ANALYSIS RESULTS

TECHNOLOGY PLANNING AND MANAGEMENT CORPORATION (TPMC)

To: David Mitchell
Cc: Jeff Rosen, Isabelle Morin; Jose Sobrinho
From: Kristyn Stevens
Date: August 31, 2001
Subject: ***ENSR – NEIWPCC Statistical Analysis Results***

INTRODUCTION/BACKGROUND

This memorandum summarizes the results and conclusions by TPMC in an effort to support ENSR in providing a technical basis and support for the United States Environmental Protection Agency, New England ("EPA") development of nutrient criteria for lakes, ponds, and reservoirs ("L/P/R") in New England. The statistical analyses described by this memo indicate support for development of sub-classifications of L/P/R that may justify development of categorized nutrient criteria, dependent on the physical and chemical characteristics.

Ecoregions were used as a basis for some data segregations prior to analysis. Three of the four major New England Non-aggregated Level 3 ecoregions were included in these analyses, including: North Eastern Coastal Zone (NECZ); Laurentian Plains and Hills (LPH); and North Eastern Highland (NEH). Variables included in the analyses were morphometric, chemical, and land use categories. Morphometric variables included mean depth, maximum depth, area, Osgood Index (defined as mean depth / square root of the surface area), and latitude (coordinates of estimated lake center). Land use categories included residential, commercial, forested, agricultural, recreational, shrubland, wetland, and barren. Chemical variables included total nitrogen (TN), total phosphorus (TP), secchi disk transparency (SDT), color, chlorophyll *a* (chl a), alkalinity, and pH.

All statistics were performed on the New England Nutrient Database provided by ENSR. This database consisted of the geometric means of trophic parameters for over 1000 waterbodies during the summer index period. The database was produced by averaging duplicate samples taken at a particular station, date, time and depth. Then, the geometric mean was calculated for each waterbody based on the "averaged" samples collected during the summer index period, in the upper 5 meters of the water column. This resulted in a unique value for each waterbody.

ENSR proposed the following two questions which characterized the statistical investigation:

1. *Using reference lake data only, what variables create the greatest "separation" (i.e., clustering in groups along morphological or water quality-based axes) for the trophic state indicators (TP, SDT, chl a TN) ?* As *a priori* assumptions, we expected mean lake depth and alkalinity or color within an ecoregion to be important.

2. *What factors allow the greatest "separation" of reference and impacted lakes in terms of trophic variables ?*

Based on discussion between TPMC, ENSR, and EPA on September 28th, 2000, it was decided to break this analysis into a phased approach, due to budgetary and time constraints. The first recommended step was to identify whether or not there were distinct categories of reference lakes. If distinct categories could be discerned then each category must be considered separately since the between group variability may possibly mask detectable differences between reference and impaired lakes.

METHODS AND RESULTS

Cluster analyses, correlations, and prinicpal component analysis (PCA) were selected as the most appropriate methods to properly classify impaired, general, and reference conditions in each ecoregion. Since these suggested analyses are predicated on a normal distribution, exploratory analyses and tests for normality were conducted to test this assumption. These methods are described below.

Exploratory Analyses

Descriptive Statistics

Three waterbody assessment classes were identified from the New England Nutrient Database. These were reference waterbodies (identified by States through watershed analyses or best professional judgment (BPJ)), impaired (identified by presence on Clean Water Act Section 303(d) lists or state-specific evaluations), and test lakes. Test lakes refer to waterbodies in the New England Nutrient Database that were not identified as reference nor listed on the State's 303(d) lists. Waterbodies identified as "Test2" (subset of Maine lakes with additional comments), and "None" (no comments from States) were put in Test category for statistical analyses.

Preliminary and exploratory analyses of the data were performed for the overall dataset including all assessment types, by assessment type (reference, test, and impaired) and by a combination of EPA ecoregion and assessment type. Sample size, mean, standard deviation, minimum and maximum values were calculated for all numeric variables surveyed in this study. These results are presented in Tables 1 through 5.

Table 1 provides a summary of selected parameters of the New England Nutrient Database used as the master data source for all subsequent statistical analyses. Table 2 provides summary statistics of the entire dataset broken down into the three assessment categories (reference, impaired, test). Tables 3 through5 provide statistical summaries of each assessment category, broken down by ecoregion (i.e., NECZ, LPH, and NEH).

TPMC noted that the majority of the lakes included in the reference assessment class were from Maine. This is likely to introduce some bias in the dataset, both geographically and from a north/south gradient perspective. It may also introduce bias from any specific differences between Maine and the other states in hydrology.

Testing for Normality of Variables and Optimal Data Transformation Analyses

Parametric statistical analyses are often more powerful than non-parametric analyses. However, this is not true when the assumptions of parametric statistical analyses are violated. One of the most critical assumptions that characterize parametric analyses is that a sample taken from a population should be normally distributed. In the case that this assumption fails, data transformations can often result in data that do meet this assumption. Data transformation analyses typically achieve or improve normality, as well as simplify the structure of the model and stabilize the variance.

Tests were conducted to determine if the variables were normally distributed. The Shapiro-Wilk test was used to test the normality of samples with fewer than 2000 observations and the Kolmogorov test for larger sample sizes. None of the numeric variables examined had normal distributions at the 95% confidence level ($p < .05$). Since the majority of data from this study showed a highly skewed and non-normal distribution with p-values of less than 0.01 for most variables, data transformation (power analyses from Box and Cox (1964) analyses were employed to determine what transformation would result in approximate normal distributions.

Two transformations of the morphological and chemical data were attempted - the square root and the log10 transformation. Few of the variables examined had normal or near normal distributions at the 95% confidence levels after the recommended data transformations were performed. These variables included lake area, Osgood Index, color, and chlorophyll *a* for the log 10 transformation and SDT for the square root transformation.

Transformation analyses were also performed on the proportional data (e.g., land use) using the (variable) +1 to account for zeros in the data set. Since the Arc sine inverse of the data is used for proportional data, this transformation was also applied (Krebs, 1989). All measurements were $p<0.01$ which meant that the transformed values were still not normal, most likely due to zeros. None of the transformations helped to reach normality, therefore non-parametric tests were used.

Verification of Classifications through Wilcoxon Rank Sum Tests

The next set of analyses was performed to evaluate whether the classifications made by the State experts in this study defined a statistically distinguishable waterbody population. Table 6 presents the results of non-parametric statistical tests between the impaired and the reference data for each variable measured for both the impaired and reference waterbodies. Although t-tests are somewhat tolerant of a lack of normality, because of the data's extensive departure from normality, non-parametric tests (e.g., Wilcoxon Rank Sum Test) were used instead.

In general, the tests indicated that there were significant statistical differences between most parameters for the impaired lakes versus the reference lakes. The exceptions to this were lake area, Osgood index and color. All other parameters showed statistically significant difference between the impaired and the reference lakes. This suggests that the State experts identified a statistically distinct set of lakes and that there are, in fact, differences between the impaired and the reference lakes, which can be measured statistically with a sample size on the order of a few hundred lakes.

Note: Cluster analyses were also originally proposed to determine whether the data support the impaired versus reference categories over the entire data set. This would present a quantitative perspective on whether or not the a priori classifications are reasonable. Since the Wilcoxon Rank Sum tests indicated that the classifications done by experts are supported by the data, this step was not necessary.

Based on these statistical differences between impaired and reference lakes, the parameters for the reference lakes were summarized by ecoregions. Table 7 shows the results of non-parametric Wilcoxon Rank Sum tests for each variable from the reference lakes when a pair-wise comparison is done between ecoregions. This table indicates that from the perspective of morphometrics (lake area and max depth), ecoregion NECZ is significantly different from both LPH and NEH. However, it is important to note (as shown in Table 3) that the sample sizes for NECZ are much smaller than they are for LPH, which in general is slightly smaller than NEH.

The differences in the sample sizes may account to some degree for some of the consistent differences that we have observed between NECZ, LPH, and NEH.

When looking at the land use categories, the ecoregions are different in all of the following land use classifications: forested, recreational, and wetlands. We do see some differences in residential, shrubland, and commercial land use percentages between NECZ and LPH and NECZ and NEH, while LPH and NEH are the same. To some degree this may be a function of the smaller sample size for NECZ. In addition, NECZ is different from LPH and NEH for secchi depth, color, and chlorophyll *a*.

These results indicate that the ecoregions are sufficiently different and there are consistent, observable, and detectable differences that can be identified. This would suggest that the ecoregion designation is a pertinent one and is likely to be appropriate for use in development of water quality criteria and indicators.

Correlation analyses

To establish a set of parameters that can most effectively be used to separate classes of reference lakes from one another with minimum redundancy, correlation analyses (Pearson or Spearman) were performed to identify parameters that co-vary strongly. If two variables were thought to be correlated, when one changed the other did so in a related manner. The correlation coefficient is a number ranging between –1 (perfect negative correlation) and 1 (perfect positive correlation) that acts as a measure of association between the two variables of interest. Correlations were run on morphometric, land use, and chemistry variables. For purposes of this analysis, a correlation coefficient >0.50 was considered indicative of a strong correlative relationship. Normality tests were used to identify whether Spearman (non-parametric) or Pearson (parametric) correlation methods should be used.

Due to the high non-normality of the environmental variables, the non-parametric Spearman Rank correlations were used. In addition, for the few variables that were found to be normally distributed after transformation (e.g., lake area, Osgood index, color, chl a, and SDT), Pearson correlations were also conducted. These results concurred with the Spearman correlations. Correlation analyses for all data in the New England Nutrient Database (regardless of assessment type) are presented in Table 8. Note that all shaded values in the correlation tables represent coefficients approximately 0.5 or above.

Few highly significant correlations were observed for the overall data including pH and alkalinity, lake mean depth and lake mean depth (which would be expected). However, several environmental factors displayed correlation coefficients approximately at or above 0.50. This included positive relationships between chl *a* and TP, color and TP, lake area with both mean and maximum ("max") depth, SDT and mean and max depth, TN and TP, as well as color and TN. This also included inverse relationships between color and SDT, TN and mean depth, TP and max depth, Osgood Index and lake area, SDT and TN, as well as SDT and TP.

Correlation analyses results for the reference data only are shown in Table 9. Results for the reference data were similar to that of the overall data with few exceptions. The inverse correlation with TP and max depth was not evident in the reference data nor was the inverse relationship between both color and chl *a* with TP. In addition the relationship between lake area and mean depth was also not evident.

Correlations were also run by ecoregion for NECZ (presented in Table 10) and the LPH and NEH ecoregions combined (presented in Table 11). An inverse relationship between SDT and chl *a*, as well as color and max depth was evident in the NECZ. Interestingly, relationships between mean or max depth were not strongly correlated with lake area. In addition, Osgood Index and lake area were not strongly correlated. The relationships between TN and mean depth and TP and max depth were not as strong as they were in the overall dataset.

Results for the LPH and NEH ecoregions combined were very similar to that of the overall dataset with few exceptions. An inverse relationship between chl *a* and SDT was evident in this dataset but it was not as strong in the overall dataset. In addition, an inverse relationship was evident between TN and max depth but was not evident between TP and max depth.

In summary, few correlations were observed between parameters across the major three data categories (morphometric, land use and physiochemical). The main correlations among parameters (across the three data categories) were TN and TP correlated with mean and max depth respectively. In addition, SDT was correlated with both lake mean and max depth and color was correlated with max depth.

Cluster Analyses (Reference Data Only)

Cluster analyses were conducted to determine if clearly identifiable groups could be identified within the reference lakes. Both mean depth and surface area were removed when running a cluster analysis with Osgood Index (see definition earlier) included in the morphological variables.

Preliminary cluster analyses were run on the morphometrics for the reference lakes to see if general classes could be distinguished based on size, depth etc. The cluster analysis separated out into size classes dominated by the lake area. In order to reduce the overwhelming impact of surface area, a square root transformation was performed on the raw value of lake surface area and the cluster analysis was repeated. It was expected that this would increase the influence of the max and mean depths into consideration for the clustering. However, lake surface area continued to dominate the clusters.

Lake surface area clusters were also analyzed to see if they were related to land use, water quality and ecoregions. No apparent trends were observed. To verify that no trends exist with regard to nutrients, we also plotted TN and TP concentrations against lake surface areas and did not observe any apparent relationships (Figures 1 and 2). Based on these results, TPMC concluded that there are no apparent clusters of lakes that would justify consideration of separate criteria based on the morphometrics of the lakes.

Principal Components Analysis (PCA)

PCA reduces the number of parameters required to define and delineate water conditions by combining those parameters into a linear combination that accounts for the greatest amount of variance in the data. This typically provides a clear grouping of parameters that co-vary, resulting in regressions that best define the relationships in nature. A correlation matrix was used to identify the relevant parameters. The analysis results in equations that can be used to linearly combine multiple co-varying parameters, resulting in a new set of parameters that carry the maximum variability for the original parameters while eliminating covarying data.

PCA was conducted on the overall data as well as the overall dataset split into two geographic regions (i.e., NECZ vs. LPH + NEH), and by assessment type. Parameters included in the analysis consisted of the morphometric and chemistry variables only. [Note: all land use variables were excluded from the PCA].

Specifically PCA was conducted on seven dataset-parameter combinations:

1) All data (all parameters, excluding TN and TP)
2) All data (all parameters excluding TN, TP, mean depth and surface area)
3) NECZ (all parameters excluding TN, TP, mean depth and surface area)
4) LPH and NEH combined (all parameters excluding TN, TP, mean depth and surface area)
5) Reference waterbodies (all parameters excluding TN, TP, mean depth and surface area)
6) Test waterbodies (all parameters excluding TN, TP, mean depth and surface area)
7) Impaired waterbodies (all parameters excluding TN, TP, mean depth and surface area)

We initially began with 1,125 waterbody records. However, PCA needs a data point for each field in each record. Due to the low sample size (22 to 44 records), which resulted from using all of the data we removed TN from the analyses (since it only had a limited number of values). Once TN was removed; the sample size was approximately half the actual number of records. We also removed surface area and mean depth from the majority of the analyses since Osgood index is an integrated measure of both surface area and mean depth. Once these analyses were completed we also tried a set of analyses without TP to see if that would change any of the results. Since much is known about TN and TP in lake systems, we did not want to mask other important factors. The results for the analyses without TN and TP are provided below. See Appendix A and Tables A1-A7 for the PCA results that exclude only TN.

TPMC identified the most influential parameters with respect to discriminating/classifying the trophic indicator data. Since the goal of PCA is to identify a new set of reduced variables (principal components) to account for the variance of the dataset we also included variance information. PCA can be described using eigenvalues and eigenvectors. The eigenvalue indicates the amount of variance explained by an eigenvector (or principal component) out of the total variance. The way to determine the number of eigenvectors is to follow some general guidance or stopping rules. The two criteria that were used in the PCA were:

1. The percentage of variance – the specification that factors are to be extracted until some percentage of the total variance has been explained (our experience with environmental data is to use 85%).
2. "Kaiser's stopping rule" –extract (or retain) only eigenvector with eigenvalues of at least 1.

Since this study involves many moderately correlated variables and a sufficient sample size, a large number of factors would be retained if we use the 85% rule. This led us to reduce the percentage of variance criterion to about 70% coupled with the Kaiser stopping rule.

PCA Results

Results from the PCA have been divided by assessment type/ecoregion and also by the inclusion of parameters. Results from each of the seven analyses are presented in Tables 12 through 18, respectively. Each table includes information on the eigenvalues, the cumulative percentage of the total variance contributed by each principal component, and the individual principal components (X_1-X_n) (or eigenvectors).

PCA #1 - All data (all parameters, excluding TN and TP) – Table 12

The eigenvalue for a principal component indicates the percentage of the variance that the Principal component accounts for. In PCA #1. The first principal component accounts for (2.459/9) *100% = 27%, the second for (1.647/9) or 17%, the third for (1.166/9) or 13% and the fourth for (1.113) or 12%. Cumulatively, the second component accounts for approximately 45%, the third for 58% the fourth for 70% and the fifth for 81%, making the last five components account for the remainder of the variance (approximately 19%)

Principal Component 1 (PC1) was dominated by SDT and mean depth with inverse values for chl *a* and color. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by lake area with an inverse value for Osgood index.

PCA #2 - All (all parameters excluding TN, TP, mean depth and surface area) - Table 13

PC1 was dominated by chl a and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #3 - NECZ (all parameters excluding TN, TP, mean depth and surface area) – Table 14

PC1 was dominated by chl a and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with inverse values for SDT and color. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #4 - LPH and NEH combined (all parameters excluding TN, TP, mean depth and surface area) – Table 15

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #5 - Reference (all parameters excluding TN, TP, mean depth and surface area) – Table 16

PC1 was dominated by SDT with inverse values for chl *a* and color. PC2 was dominated by pH and alkalinity with an inverse value for SDT. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #6 - Test (all parameters excluding TN, TP, mean depth and surface area) – Table 17

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for color. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #7 - Impaired (all parameters excluding TN, TP, mean depth and surface area) – Table 18

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

In general, for the seven dataset-parameter combinations, PCA has given us a fairly consistent set of principal components. Principal Component 1 is dominated by chl *a* and color exhibiting positive factors while SDT has a negative factor. Principal Component 2 is dominated by pH and alkalinity with positive factors and color with a negative factor. Principal Component 3 is dominated by Osgood Index with a positive factor and pH with a negative factor. Principal Component 4 is dominated by max depth with a positive factor and Osgood index with a negative factor. The overall assessment (excluding TN and TP only) and the reference PCA display similar components, except Principal Component 1 is dominated by SDT (+ factor) and both chl *a* and color (- factors). Principal Component 2 for the reference lakes only is also dominated by alkalinity (+) and pH (+) and SDT (-).

The results of this PCA suggest that there are consistent parameters that seem to influence the characteristics of lakes and ponds. These same parameters emerge as influential in the principal component analyses regardless of how the data are aggregated and classified. This consistency suggests that these parameters contain the signals of the responses of lakes to their natural and anthropogenic influences. These parameters also probably contain the signals of other less influential parameters that are correlated with the parameters that most influence the principal components.

Finally, it should be noted that the subset of reference lake results indicate the same dominant parameters in each of the principal components as the impacted and test lakes, but with opposite signs from those analyses. This supports the contention that these parameters are fundamentally indicative of the lakes' conditions. More specifically, the factors' reverse sign indicates that when the data are aggregated, the same parameters that dominate for impacted lakes (with positive sign) are negative sign for reference (on unimpacted) lakes. This inverse relationship indicates that if parameters are fundamentally connected to the condition of impaired lakes (i.e., show high factors), we would expect those same parameters to show very low number for reference lakes.

CONCLUSIONS

On the basis of the statistical analyses conducted, the following conclusions were reached. However, it should be noted that there may be underlying uncertainty associated with these conclusions due to different sample size among ecoregions and parameters.

- The classifications of assessment type (reference versus impacted) by experts are supported by the data as seen in the descriptive statistics and the Wilcoxon Rank Sum tests.
- The NECZ ecoregion is significantly different from the LPH and the NEH ecoregions for morphometric parameters. This supports the conclusion that the ecoregions do provide statistical segregation for many important lake parameters
- Based on the results of the cluster analyses, there are no apparent clusters of lakes that would justify consideration of separate criteria based on the morphometrics of the lakes.
- PCA provides tentative identification of the four most influential principal components. Principal Components 1 and 2 are a function of water quality. Principal component 3 is a function of morphometrics with inverse relationships to water quality. Principal Component 4 is a function of morphometrics alone.
- PCA identified clear but not strong principal components. Each component cannot be easily and strongly named and described. This may be attributed to the fact that most of the variables are moderately correlated with one another in some manner which may account for the smaller percentages represented by each component.

RECOMMENDATIONS

Results from the PCA show promise that the parameters measured can be recombined into principal components that account for the majority of the variability observed in all the parameters measured. This would reduce the complexity of further exploratory analyses and might indicate that only the dominant parameters need to be measured in the future in order to properly classify lakes as impacted or not impacted. The best way to determine the efficacy of the dominant parameters would be to calculate the principal components for each observation and then use these data in a step wise discriminant analysis. This would establish if discriminant functions could be developed based on the principal components that can properly classify lakes with a high enough classification efficiency.

The approach would be as follows:

1. Subset out the assessment type = test. These records will not be included in the discriminant analysis.

2. Calculate the principal components for each record using the factors calculated in this study. Keep only the categorization parameters (assessment class, the ecoregion, the sampling date, etc.) and the principal components.

3. Partition the remaining data set into two equal subsets by randomly selecting equal numbers of reference and impaired lakes from the entire data set.

4. Use one set of the data (we'll call it the training set since it will be used to train the discriminant analysis to identify reference and impaired) for the stepwise discriminant analysis and the second set of data for the validation of the discriminant function.

5. Predefine what an acceptable rate of correct classification of lakes will be. This rate will be considered following step 8 below.

6. With the first set, run a stepwise discriminant analysis to identify the principal components that best segregate between impaired and reference. Define stopping rules similar to the ones employed for the Principal Component Analysis.

7. Once the discriminating principal components are identified, run a regular discriminant analysis. This will output a discriminant function (based on the principal components) and a decision rule indicating what values of the discriminant function classify into the reference category and which values classify into the impaired category.

8. Using the outputs from step 6 above, the second set of data (the validation set) should be used to calculate the appropriate discriminant scores and the resulting classifications. These classifications should be compared to the a-priori classification in the attribute assessment class, and a percent correct classification calculated.

9. If the percent correct classification exceeds the percentage set in step 5 above then you have a useful disciminant function, which you can apply and use in resource management. If not, then you can try to redo your stepwise discriminant function or consider if the lower classification efficiency is marginally acceptable. The other alternative is that the data still contain too much variability to develop a strict mathematical approach to identifying lakes at risk.

10. More complicated assessments of the efficacy of classification can be performed by separately considering the percent correct classification for the known impaired systems and the known reference systems. This gets quite complicated but can be used to determine if a particular discriminant function classifies "protectively". An example would be if the impaired lakes are properly classified 80% of the time and the reference lakes are only properly classified 50% of the time then the approach would be protective. This means that you might be requiring monitoring or reduction of inputs in some lakes that don't really need mitigation. This is still classifying protectively in that 80% of the impaired lakes will have action taken (20% will not) and 50% of the lakes that do not need any action will have actions taken that improve the environment. If the approach needs to balance cost with protection then the percentages of misclassification of reference lakes need to be looked at very carefully.

11. If the determination of classification efficiencies are acceptable based on the criteria in either step 9 or 10 above then the discriminant function development should be repeated, but for this iteration no principal components should be calculated. Instead those parameters which emerged as dominant in the principal components should be used directly with no additional transformation or calculations. If the classification efficiency is still acceptable then only the dominant parameters need to be measured. The discriminant function developed for the parameters is applied and the result gives you a classification (impaired or reference) with a confidence interval that can be used to make management decisions.

REFERENCES

Box, G.E.P., and Cox, D.R. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society., Ser. B 26:211-243.

Kaiser, H.F. 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Krebs, C.J. 1989. Chapter 12 – Survival rates In: Ecological Methodology pp. 446-451.

Table 1. Descriptive Statistics for Overall Dataset (all Assessment Types)

Variable	Statistics				
	N	Min	Max	Mean	Std Dev
Lake Mean Depth (m)	996	0.2	42.7	5.2	4.09
Lake Max Depth (m)	963	0.5	5468.0	18.8	176.15
Lake Area (ha)	1073	0.40	30876.00	327.92	1348.608
Osgood Index	992	0.578	134.397	6.181	6.9661
Land Use Residential (%)	1029	0	0.6452	0.0300	0.0800
Land Use Commercial (%)	1029	0	0.2187	0.0113	0.0248
Land Use Forested (%)	1029	0	0.9986	0.8179	0.1525
Land Use Agricultural (%)	1029	0	0.6375	0.0612	0.0743
Land Use Recreational (%)	1029	0	0.1939	0.0076	0.0205
Land Use Shrubland (%)	1029	0	0.6044	0.0034	0.0316
Land Use Wetland (%)	1029	0	0.4097	0.0539	0.0454
Land Use Barren (%)	1029	0	0.2794	0.0119	0.0274
Total Nitrogen (μg/L)	321	20.00	3796.77	484.38	400.385
Total Phosphorus (μg/L)	937	0.95	376.25	15.16	22.284
Secchi Disk Transparency (m)	1111	0.39	13.92	4.21	2.106
Color (color units)	724	1.00	315.00	26.89	23.901
Chlorophyll *a (μg/L)*	926	0.10	172.25	5.48	8.369
Alkalinity (mg/L as $CaCO_3$)	786	0.11	1897.53	36.74	129.675
pH (S.U.)	710	4.27	8.94	6.81	0.545

Notes:
Includes REF, TEST, TEST2, NONE, and IMP

REF = waterbodies identified as reference by State experts
IMP = waterbodies listed on a State 303(d) list
TEST = waterbodies identified as neither reference nor impacted
TEST2 = Maine waterbodies identified as neither reference nor impacted
NONE = waterbodies with none comments from States, assumed to be neither impacted nor reference

Table 2. Descriptive Statistics by Assessment Types

Variable	Reference					Test					Impaired				
	N	Min	Max	Mean	Std Dev	N	Min	Max	Mean	Std Dev	N	Min	Max	Mean	Std Dev
Lake Mean Depth (m)	353	0.2	24.7	5.9	3.98	493	0.2	42.7	4.7	4.21	150	0.8	33.3	5.2	3.71
Lake Max Depth (m)	330	0.5	67.7	15.6	10.55	485	0.5	5468.0	22.5	248.01	148	1.0	107.0	13.4	13.88
Lake Area (ha)	375	0.80	18043.40	426.35	1257.431	539	0.40	30876.00	247.73	1478.872	159	2.00	11949.00	367.60	1050.494
Osgood Index	353	0.578	28.520	5.900	4.5490	489	0.657	134.397	6.537	8.6104	150	0.707	45.255	5.681	5.4951
Land Use Residential (%)	383	0	0.4336	0.0112	0.0426	492	0	0.6452	0.0343	0.0808	154	0	0.6452	0.0632	0.1245
Land Use Commercial (%)	383	0	0.1045	0.0042	0.0130	492	0	0.2187	0.0140	0.0265	154	0	0.2062	0.0202	0.0352
Land Use Forested (%)	383	0	0.9986	0.8833	0.1054	492	0	0.9967	0.7922	0.1492	154	0	0.9915	0.7374	0.1941
Land Use Agricultural (%)	383	0	0.3601	0.0305	0.0515	492	0	0.5524	0.0744	0.0724	154	0	0.6375	0.0955	0.0975
Land Use Recreational (%)	383	0	0.1939	0.0038	0.0147	492	0	0.1837	0.0089	0.0221	154	0	0.1661	0.0126	0.0253
Land Use Shrubland (%)	383	0	0.0532	0.0027	0.0062	492	0	0.6044	0.0048	0.0453	154	0	0.0272	0.0009	0.0033
Land Use Wetland (%)	383	0	0.4097	0.0438	0.0424	492	0	0.3063	0.0613	0.0473	154	0	0.3106	0.0555	0.0420
Land Use Barren (%)	383	0	0.2794	0.0176	0.0351	492	0	0.2577	0.0090	0.0226	154	0	0.1170	0.0070	0.0147
Total Nitrogen (µg/L)	80	50.00	1770.00	334.28	215.213	202	20.00	3175.55	494.92	364.395	39	198.00	3796.77	737.68	661.398
Total Phosphorus (µg/L)	287	1.00	67.00	9.32	5.697	495	0.95	290.25	15.85	21.440	155	3.32	376.25	23.78	36.653
Secchi Disk Transparency (m)	386	0.75	13.92	5.07	2.069	557	0.39	13.29	3.69	1.898	168	0.55	10.92	3.95	2.216
Color (color units)	247	1.00	129.61	25.36	18.299	350	2.00	315.00	28.94	28.852	127	5.00	89.15	24.22	17.173
Chlorophyll *a* (µg/L)	283	0.20	11.70	3.30	1.993	494	0.10	51.65	5.39	5.795	149	1.50	172.25	9.90	17.015
Alkalinity (mg/L as $CaCO_3$)	253	0.11	408.64	23.60	53.848	390	0.20	1897.53	48.61	175.540	143	1.83	357.64	27.63	51.195
pH (S.U.)	213	4.60	8.36	6.74	0.482	348	4.27	8.94	6.77	0.594	149	5.37	8.90	7.02	0.455

Notes:
Test - includes TEST, TEST2, and NONE
TEST = waterbodies identified as neither reference nor impacted
TEST2 = Maine waterbodies identified as neither reference nor impacted
NONE = waterbodies with none comments from States, assumed to be neither impacted nor reference

Table 3. Descriptive Statistics by Ecoregion and Assessment Type - Reference Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	15	1.1	16.3	4.6	5.1	3.46	152	0.9	24.2	4.5	5.6	3.63	186	0.2	24.7	5.1	6.2	4.28
Lake Max Depth (m)	29	2.7	19.7	7.0	8.1	4.55	139	3.4	67.7	12.8	15.9	10.79	162	0.5	54.9	13.7	16.6	10.63
Lake Area (ha)	34	3.00	1673.40	38.82	139.15	337.825	152	4.00	7405.00	164.50	515.80	1056.20 3	189	0.80	18043.4 0	77.58	406.08	1485.80 0
Osgood	15	1.518	17.898	6.225	7.336	5.1194	152	0.578	28.520	3.765	5.097	4.3251	186	0.742	25.425	5.243	6.439	4.5970
Land Use Residential (%)	31	0	0.4336	0.0408	0.0895	0.1206	162	0	0.0725	0.0003	0.0039	0.0087	190	0	0.1284	0.0001	0.0047	0.0145
Land Use Commercial (%)	31	0	0.0923	0.0093	0.0210	0.0273	162	0	0.0489	0.0003	0.0022	0.0056	190	0	0.1045	0.0003	0.0031	0.0120
Land Use Forested (%)	31	0	0.9441	0.8152	0.7548	0.1837	162	0	0.9986	0.8923	0.8782	0.0949	190	1	0.9974	0.9252	0.9085	0.0777
Land Use Agricultural (%)	31	0	0.1724	0.0257	0.0323	0.0365	162	0	0.3377	0.0013	0.0351	0.0579	190	0	0.3601	0.0016	0.0263	0.0475
Land Use Recreational (%)	31	0	0.1939	0.0072	0.0280	0.0420	162	0	0.0260	0.0001	0.0020	0.0038	190	0	0.0758	0.0000	0.0014	0.0066
Land Use Shrubland (%)	31	0	0.0069	0.0000	0.0006	0.0017	162	0	0.0331	0.0003	0.0025	0.0058	190	0	0.0532	0.0001	0.0032	0.0069
Land Use Wetland (%)	31	0	0.1348	0.0460	0.0632	0.0290	162	0	0.4097	0.0475	0.0569	0.0543	190	0	0.1327	0.0248	0.0296	0.0241
Land Use Barren (%)	31	0	0.1062	0.0014	0.0103	0.0240	162	0	0.2794	0.0039	0.0160	0.0356	190	0	0.2756	0.0044	0.0202	0.0361
Total Nitrogen (µg/L)	32	208.56	1000.00	303.43	369.09	162.331	14	153.00	1770.00	236.00	362.71	411.610	34	50.00	602.46	269.50	289.82	125.809
Total Phosphorus (µg/L)	37	4.80	46.63	8.52	11.33	8.085	115	2.00	18.33	9.00	9.62	3.645	135	1.00	67.00	7.13	8.52	6.211
Secchi Disk Transparency (m)	38	1.21	7.50	4.30	3.93	1.483	160	1.00	13.92	4.84	5.32	2.068	188	0.75	13.20	4.89	5.09	2.102
Color (color units)	10	3.74	35.52	8.50	11.54	9.678	114	4.47	129.61	22.18	28.21	19.838	123	1.00	99.00	18.34	23.84	16.680
Chlorphyll *a* (µg/L)	37	0.60	7.53	1.90	2.25	1.407	115	0.80	11.70	3.13	3.72	2.062	131	0.20	10.59	2.70	3.23	1.967
Alkalinity (mg/L as $CaCO_3$)	13	2.00	402.82	9.00	44.59	108.780	117	2.45	408.64	8.73	21.08	44.290	123	0.11	367.56	7.74	23.78	53.918
pH (S.U.)	11	6.00	8.30	6.70	6.73	0.712	111	4.60	8.36	6.81	6.78	0.458	91	5.30	8.00	6.74	6.69	0.479

Table 4. Descriptive Statistics by Ecoregion and Assessment Type - Test Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	102	0.7	14.6	2.7	3.2	2.34	119	0.6	18.0	3.9	4.9	3.66	272	0.2	42.7	3.7	5.2	4.82
Lake Max Depth (m)	110	0.8	24.1	4.5	6.6	5.23	108	1.8	5468.0	9.8	63.0	525.05	267	0.5	93.9	9.1	12.8	12.08
Lake Area (ha)	134	1.00	469.72	28.30	52.19	69.279	119	1.00	7257.60	84.00	362.58	920.577	286	0.40	30876.00	48.77	291.56	1936.500
Osgood	102	1.293	66.000	4.526	6.378	7.4065	119	0.657	71.000	3.939	6.027	8.4298	268	0.667	134.397	5.198	6.824	9.1181
Land Use Residential (%)	96	0	0.6452	0.0676	0.1248	0.1478	121	0	0.0844	0.0034	0.0085	0.0148	275	0	0.1201	0.0063	0.0140	0.0213
Land Use Commercial (%)	96	0	0.2187	0.0205	0.0364	0.0438	121	0	0.0483	0.0008	0.0051	0.0101	275	0	0.1524	0.0036	0.0101	0.0179
Land Use Forested (%)	96	0	0.9591	0.6941	0.6431	0.1931	121	0	0.9943	0.8402	0.8203	0.1165	275	0	0.9967	0.8515	0.8318	0.1064
Land Use Agricultural (%)	96	0	0.3317	0.0528	0.0636	0.0646	121	0	0.5524	0.0725	0.0785	0.0811	275	0	0.4073	0.0614	0.0764	0.0709
Land Use Recreational (%)	96	0	0.1837	0.0206	0.0345	0.0383	121	0	0.0563	0.0016	0.0045	0.0082	275	0	0.0683	0.0001	0.0019	0.0074
Land Use Shrubland (%)	96	0	0.0119	0.0000	0.0007	0.0024	121	0	0.6044	0.0000	0.0120	0.0751	275	0	0.5664	0.0000	0.0030	0.0343
Land Use Wetland (%)	96	0	0.2329	0.0828	0.0878	0.0433	121	0	0.3063	0.0458	0.0601	0.0553	275	0	0.2889	0.0415	0.0527	0.0413
Land Use Barren (%)	96	0	0.1810	0.0023	0.0084	0.0207	121	0	0.1170	0.0022	0.0076	0.0157	275	0	0.2577	0.0038	0.0099	0.0256
Total Nitrogen (µg/L)	116	122.92	3175.55	452.04	579.67	412.029	11	211.00	654.00	333.00	387.91	136.392	75	20.00	1877.39	312.33	379.52	260.513
Total Phosphorus (µg/L)	144	0.95	290.25	15.80	24.48	35.720	103	3.00	87.12	10.94	14.02	10.932	248	1.26	77.94	9.53	11.60	8.267
Secchi Disk Transparency (m)	145	0.39	9.14	2.15	2.50	1.407	122	0.71	11.36	4.32	4.47	2.035	290	0.63	13.29	3.67	3.96	1.781
Color (color units)	62	4.00	315.00	24.15	37.20	46.322	96	5.00	155.00	23.00	32.65	28.099	192	2.00	150.00	20.10	24.43	19.838
Chlorphyll *a* (µg/L)	144	0.20	40.50	4.07	6.27	7.133	98	1.60	50.57	3.98	5.51	6.005	252	0.10	51.65	3.95	4.84	4.720
Alkalinity (mg/L as $CaCO_3$)	93	1.00	1680.98	11.00	82.67	255.374	99	1.90	332.71	10.00	23.72	48.155	198	0.20	1897.53	7.08	45.06	168.420
pH (S.U.)	77	4.50	8.94	6.67	6.72	0.733	97	5.47	8.38	6.93	6.93	0.471	174	4.27	8.80	6.80	6.71	0.574

Table 5. Descriptive Statistics by Ecoregion and Assessment Type - Impaired Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	29	0.8	12.5	2.8	3.7	2.79	79	0.8	15.0	4.6	5.1	2.73	42	1.9	33.3	5.3	6.4	5.25
Lake Max Depth (m)	33	1.0	22.9	7.3	7.9	5.66	75	2.7	51.8	10.1	13.4	10.22	40	3.0	107.0	12.7	18.1	21.31
Lake Area (ha)	37	3.80	524.00	23.60	80.93	129.299	79	2.00	2716.00	141.00	383.12	564.823	43	10.10	11949.0 0	156.00	585.76	1850.14 0
Osgood	29	1.473	16.101	4.872	5.832	3.8989	79	0.707	45.255	3.615	5.859	6.6820	42	1.303	21.083	4.129	5.242	3.7601
Land Use Residential (%)	28	0	0.6452	0.1466	0.2386	0.2104	83	0	0.2152	0.0105	0.0202	0.0318	43	0	0.1173	0.0233	0.0319	0.0299
Land Use Commercial (%)	28	0	0.2062	0.0359	0.0623	0.0603	83	0	0.1110	0.0045	0.0109	0.0173	43	0	0.0556	0.0071	0.0108	0.0123
Land Use Forested (%)	28	0	0.8982	0.5765	0.5267	0.2578	83	0	0.9915	0.8110	0.7821	0.1304	43	0	0.9805	0.8222	0.7882	0.1586
Land Use Agricultural (%)	28	0	0.1628	0.0310	0.0498	0.0504	83	0	0.6375	0.0952	0.1205	0.1123	43	0	0.4618	0.0607	0.0769	0.0734
Land Use Recreational (%)	28	0	0.1661	0.0170	0.0374	0.0467	83	0	0.0896	0.0049	0.0088	0.0136	43	0	0.0316	0.0016	0.0036	0.0063
Land Use Shrubland (%)	28	0	0.0206	0.0000	0.0013	0.0046	83	0	0.0125	0.0000	0.0009	0.0023	43	0	0.0272	0.0000	0.0007	0.0041
Land Use Wetland (%)	28	0	0.1017	0.0748	0.0704	0.0206	83	0	0.2724	0.0409	0.0488	0.0405	43	0	0.3106	0.0508	0.0586	0.0521
Land Use Barren (%)	28	0	0.1170	0.0042	0.0133	0.0274	83	0	0.0414	0.0026	0.0053	0.0076	43	0	0.0655	0.0018	0.0061	0.0127
Total Nitrogen (µg/L)	31	200.00	3796.77	629.16	836.66	705.800	4	200.00	708.00	365.50	409.75	219.988	4	198.00	400.00	298.00	298.50	96.341
Total Phosphorus (µg/L)	41	5.47	376.25	25.29	46.25	63.591	71	4.47	72.78	14.00	18.17	12.874	43	3.32	53.57	8.00	11.63	10.064
Secchi Disk Transparency (m)	42	0.55	6.63	1.66	2.28	1.697	83	0.80	9.05	3.81	4.16	1.896	43	1.09	10.92	5.41	5.18	2.296
Color (color units)	17	8.00	68.47	21.00	29.34	20.796	70	6.32	70.87	21.57	24.81	15.032	40	5.00	89.15	13.77	21.03	18.791
Chlorphyll *a* (µg/L)	41	2.40	172.25	6.46	16.68	29.451	67	1.60	35.61	5.41	8.28	7.460	41	1.50	39.38	4.19	5.77	6.560
Alkalinity (mg/L as $CaCO_3$)	29	4.24	357.64	18.12	64.93	97.043	73	2.00	150.58	15.49	23.05	26.783	41	1.83	39.52	8.49	9.40	5.762
pH (S.U.)	27	5.37	8.90	7.10	7.08	0.634	81	5.70	8.25	7.08	7.08	0.434	41	6.08	7.39	6.91	6.85	0.296

Table 6. Non-parametric comparison (Wilcoxon test) between Reference and Impaired data

Variable	P-value	Significant? Yes/No
Lake mean depth (m)	0.032	Yes
Lake max depth (m)	<0.010	Yes
Lake area (ha)	0.431	No
Osgood Index	0.217	No
Land Use Residential (%)	<0.010	Yes
Land Use Commercial (%)	<0.010	Yes
Land Use Forest (%)	<0.010	Yes
Land Use Agricultural (%)	<0.010	Yes
Land Use Recreational (%)	<0.010	Yes
Land Use Shrubland (%)	<0.010	Yes
Land Use Wetland (%)	<0.010	Yes
Land Use Barren (%)	0.043	Yes
Total Nitrogen (µg/L)	<0.010	Yes
Total Phosphorus (µg/L)	<0.010	Yes
SDT (m)	<0.010	Yes
Color (color units)	0.401	No
Chlorophyll *a* (µg/L)	<0.010	Yes
Alkalinity (mg/L as $CaCO_3$)	<0.010	Yes
pH (S.U)	<0.010	Yes

Table 7 - Non-parametric comparison (Wilcoxon test) between EPA Ecoregions (reference data only); Ecoregion 1= NECZ, Ecoregion 2 = LPH, Ecoregion 3 = NEH

Variable	P-values	Significant? Yes/No
Lake mean depth	0.679 for Ecoregions 1 and 2 0.295 for Ecoregions 1 and 3 0.174 for Ecoregions 2 and 3	No
Lake max depth	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.324 for Ecoregions 2 and 3	Yes No
Lake area	<0.010 for all three comparisons	Yes
Osgood	0.073 for Ecoregions 1and 2 0.535 for Ecoregions 1 and 3 <0.010 for Ecoregions 2 and 3	No No Yes
L. Use Residential	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.125 for Ecoregions 2 and 3	Yes No
L. Use Commercial	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.271 for Ecoregions 2 and 3	Yes No
L. Use Forest	<0.010 for all three comparisons	Yes
L. Use Agricultural	0.100 for Ecoregions 1 and 2 0.053 for Ecoregions 1 and 3 0.947 for Ecoregions 2 and 3	No
L. Use Recreational	<0.010 for all three comparisons	Yes
L. Use Shrubland	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.895 for Ecoregions 2 and 3	Yes No
L.Use Wetland	0.039 for Ecoregions 1 and 2 <0.010 for Ecoregions 1 and 3, and 2 and 3	Yes
L. Use Barren	0.082 for Ecoregions 1 and 2 0.020 for Ecoregions 1 and 3 0.227 for Ecoregions 2 and 3	No Yes No
Total Nitrogen	0.023 for Ecoregions 1 and 2 0.026 for Ecoregions 1 and 3 0.626 for Ecoregions 2 and 3	Yes No
Total Phosphorus	0.944 for Eparegions 1 and 2 <0.010 for Eparegions 1 and 3, and 2 and 3	No Yes
Secchi depth	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.525 for Ecoregions 2 and 3	Yes No
Color	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.065 for Ecoregions 2 and 3	Yes No
Chlorophyll *a*	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.055 for Ecoregions and 2 and 3	Yes No
Alkalinity	0.858 for Ecoregions 1 and 2 0.584 for Ecoregions 1 and 3 0.659 for Ecoregions 2 and 3	No
pH	0.287 for Ecoregions 1 and 2 0.627 for Ecoregions 1 and 3 0.239 for Ecoregions 2 and 3	No

Table 8. Spearman Correlations for the overall dataset including all assessment types (REF, IMP, TEST, TEST2, and NONE)

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.898	0.543	0.344	0.264	-0.129	-0.113	0.179	-0.117	-0.147	0.050	-0.214	0.050	-0.574	-0.480	0.662	-0.381	-0.303	-0.050	0.137
Lake Max Depth (m)		1.000	0.590	0.193	0.402	-0.230	-0.200	0.249	-0.191	-0.205	0.120	-0.242	0.078	-0.475	-0.501	0.668	-0.323	-0.273	-0.050	0.099
Lake Area (ha)			1.000	-0.541	0.397	-0.244	-0.224	0.183	-0.206	-0.183	0.124	-0.091	0.069	-0.337	-0.268	0.403	-0.109	-0.201	-0.009	0.118
Osgood Index (OI)				1.000	-0.169	0.126	0.113	-0.016	0.156	0.040	-0.075	-0.129	-0.016	-0.253	-0.196	0.218	-0.257	-0.059	-0.061	0.021
Latitude (dec. degree)					1.000	-0.713	-0.601	0.457	-0.344	-0.460	0.340	-0.407	0.202	-0.347	-0.146	0.347	0.166	-0.033	0.129	0.110
L.U. Residential (%)						1.000	0.845	-0.677	0.526	0.663	-0.349	0.398	-0.089	0.307	0.201	-0.225	-0.099	0.167	-0.014	0.122
L.U. Commercial (%)							1.000	-0.670	0.506	0.562	-0.245	0.352	0.052	0.260	0.189	-0.231	-0.062	0.142	0.039	0.144
LU Forested (%)								1.000	-0.571	-0.537	0.170	-0.570	-0.088	-0.433	-0.352	0.328	-0.103	-0.252	-0.196	-0.272
L.U. Agricultural (%)									1.000	0.303	-0.351	0.128	-0.204	0.116	0.206	-0.143	-0.040	0.244	0.134	0.339
L.U. Recreational (%)										1.000	-0.176	0.342	-0.044	0.336	0.308	-0.222	-0.002	0.143	0.134	0.161
L.U. Shrubland (%)											1.000	-0.073	0.575	-0.077	-0.026	0.061	0.126	-0.040	0.005	-0.130
L.U. Wetland (%)												1.000	0.040	0.353	0.188	-0.275	0.166	0.117	-0.125	-0.114
L.U. Barren (%)													1.000	0.048	0.013	-0.010	0.095	-0.011	0.069	-0.019
T. Nitrogen (µg/L)														1.000	0.643	-0.690	0.612	0.414	-0.073	0.117
T. Phosphorus (µg/L)															1.000	-0.668	0.514	0.581	0.280	0.164
SDT (m)																1.000	-0.583	-0.496	-0.246	0.023
Color (color units)																	1.000	0.439	0.047	-0.081
Chlorophyll *a* (µg/L)																		1.000	0.118	0.128
Alkalinity ($CaCO_3$ mg/L)																			1.000	0.689
pH (S.U.)																				1.000

Table 9. Spearman Correlations for Reference Data

Parameter	Mean Depth	Max Depth	Lake Area	OI	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	NA	0.895	0.489	0.290	-0.084	-0.069	0.165	-0.114	-0.139	-0.052	-0.172	0.006	-0.551	-0.469	0.582	-0.327	-0.392	0.014	0.085
Lake Max Depth (m)		NA	0.622	0.062	-0.245	-0.192	0.265	-0.223	-0.267	-0.007	-0.194	0.029	-0.471	-0.420	0.562	-0.247	-0.264	0.005	0.088
Lake Area (ha)			NA	-0.629	-0.374	-0.312	0.208	-0.382	-0.297	0.111	-0.079	0.106	-0.355	-0.218	0.256	0.000	-0.148	0.110	0.101
Osgood Index (OI)				NA	0.308	0.234	-0.086	0.308	0.199	-0.098	-0.063	-0.091	-0.267	-0.163	0.211	-0.287	-0.067	-0.124	-0.018
L.U. Residential (%)					NA	0.771	-0.539	0.688	0.674	-0.353	0.420	-0.303	0.073	0.042	-0.042	-0.196	-0.010	-0.265	-0.008
L.U. Commercial (%)						NA	-0.517	0.590	0.578	-0.257	0.343	-0.123	0.118	0.017	-0.059	-0.134	-0.069	-0.106	0.050
LU Forested (%)							NA	-0.502	-0.499	-0.013	-0.671	-0.153	-0.296	-0.164	0.211	-0.072	-0.126	-0.028	-0.159
L.U. Agricultural (%)								NA	0.498	-0.329	0.289	-0.343	0.095	0.107	-0.070	-0.100	0.208	-0.101	0.157
L.U. Recreational (%)									NA	-0.133	0.338	-0.091	0.081	0.207	-0.147	-0.098	0.039	-0.073	0.127
L.U. Shrubland (%)										NA	-0.023	0.652	0.043	0.048	-0.065	0.170	0.090	0.012	-0.068
L.U. Wetland (%)											NA	0.014	0.281	0.091	-0.151	0.106	0.031	-0.230	-0.089
L.U. Barren (%)												NA	0.150	0.014	-0.090	0.133	-0.025	0.052	-0.042
T. Nitrogen (µg/L)													NA	0.541	-0.597	0.672	0.180	-0.216	-0.033
T. Phosphorus (µg/L)														NA	-0.528	0.442	0.471	0.055	0.055
SDT (m)															NA	-0.593	-0.407	-0.255	0.000
Color (color units)																NA	0.394	0.088	-0.003
Chlorophyll *a* (µg/L)																	NA	0.002	-0.029
Alkalinity (mg/L as $CaCO_3$)																		NA	0.625
pH (S.U.)																			NA

Table 10. Spearman Correlations for the North Eastern Coastal Zone (NECZ) dataset

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.823	0.452	0.562	0.249	0.076	-0.064	0.173	-0.071	-0.318	-0.222	-0.027	-0.156	-0.410	-0.373	0.596	-0.447	-0.236	-0.138	0.182
Lake Max Depth (m)		1.000	0.345	0.334	0.130	-0.040	-0.078	0.165	-0.045	-0.256	0.060	-0.026	-0.171	-0.335	-0.420	0.534	-0.505	-0.134	-0.126	0.122
Lake Area (ha)			1.000	-0.430	0.069	-0.111	-0.193	0.168	0.093	-0.232	-0.049	0.130	-0.156	-0.175	-0.226	0.262	-0.230	-0.153	-0.162	0.053
Osgood Index (OI)				1.000	0.128	0.218	0.129	-0.096	-0.112	-0.017	-0.168	-0.130	0.094	-0.242	-0.137	0.288	-0.172	-0.032	0.026	0.132
Latitude (dec. degree)					1.000	-0.002	-0.050	0.204	0.292	-0.628	0.063	-0.203	0.133	-0.001	0.048	0.221	0.247	0.244	-0.299	-0.070
L.U. Residential (%)						1.000	0.783	-0.812	-0.108	0.313	-0.162	0.014	0.373	0.222	0.243	-0.176	-0.076	0.160	0.345	0.375
L.U. Commercial (%)							1.000	-0.820	-0.073	0.399	-0.081	0.149	0.461	0.133	0.241	-0.182	0.055	0.165	0.429	0.455
LU Forested (%)								1.000	-0.015	-0.549	0.117	-0.333	-0.432	-0.319	-0.401	0.339	-0.042	-0.235	-0.499	-0.351
L.U. Agricultural (%)									1.000	-0.371	0.198	-0.038	0.308	0.106	0.044	-0.109	0.085	0.171	0.053	0.187
L.U. Recreational (%)										1.000	-0.139	0.214	-0.012	0.134	0.330	-0.301	-0.098	0.016	0.632	0.079
L.U. Shrubland (%)											1.000	-0.038	0.240	0.015	0.051	-0.050	0.111	0.245	0.115	-0.089
L.U. Wetland (%)												1.000	0.128	0.124	0.038	-0.042	0.104	-0.064	-0.252	-0.204
L.U. Barren (%)													1.000	0.080	0.097	-0.126	0.232	0.171	0.245	0.493
T. Nitrogen (µg/L)														1.000	0.571	-0.612	0.650	0.444	-0.038	0.159
T. Phosphorus (µg/L)															1.000	-0.602	0.547	0.584	0.390	0.347
SDT (m)																1.000	-0.572	-0.540	-0.320	-0.080
Color (color units)																	1.000	0.490	-0.164	-0.020
Chlorophyll *a* (µg/L)																		1.000	0.304	0.320
Alkalinity (mg/L as $CaCO_3$)																			1.000	0.668
pH (S.U.)																				1.000

Table 11. Spearman Correlations for the combined Laurentian Plains and Hills (LPH) and North Eastern Highland (NEH) dataset

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.904	0.519	0.330	0.173	-0.093	-0.079	0.148	-0.115	-0.102	0.063	-0.193	0.068	-0.625	-0.484	0.648	-0.368	-0.310	-0.014	0.119
Lake Max Depth (m)		1.000	0.578	0.190	0.281	-0.175	-0.151	0.206	-0.216	-0.122	0.098	-0.198	0.093	-0.517	-0.484	0.640	-0.298	-0.334	-0.019	0.086
Lake Area (ha)			1.000	-0.575	0.342	-0.182	-0.164	0.116	-0.251	-0.086	0.123	-0.041	0.089	-0.371	-0.208	0.337	-0.078	-0.202	0.056	0.125
Osgood Index (OI)				1.000	-0.218	0.126	0.113	-0.005	0.176	0.036	-0.069	-0.136	-0.028	-0.285	-0.227	0.238	-0.271	-0.070	-0.088	-0.008
Latitude (dec. degree)					1.000	-0.636	-0.525	0.329	-0.475	-0.241	0.358	-0.299	0.245	-0.086	0.078	0.110	0.222	-0.034	0.339	0.154
L.U. Residential (%)						1.000	0.817	-0.599	0.678	0.569	-0.350	0.320	-0.139	-0.064	0.088	-0.067	-0.132	0.172	-0.102	0.148
L.U. Commercial (%)							1.000	-0.587	0.625	0.458	-0.231	0.271	0.010	-0.028	0.090	-0.106	-0.090	0.131	-0.052	0.145
LU Forested (%)								1.000	-0.691	-0.429	0.138	-0.524	-0.063	-0.246	-0.276	0.213	-0.111	-0.254	-0.138	-0.303
L.U. Agricultural (%)									1.000	0.406	-0.414	0.176	-0.275	0.077	0.254	-0.151	-0.060	0.263	0.134	0.355
L.U. Recreational (%)										1.000	-0.136	0.236	-0.041	0.082	0.213	-0.060	0.008	0.164	0.056	0.189
L.U. Shrubland (%)											1.000	-0.052	0.626	0.042	-0.019	0.031	0.125	-0.085	-0.001	-0.134
L.U. Wetland (%)												1.000	0.032	0.215	0.124	-0.197	0.169	0.127	-0.153	-0.085
L.U. Barren (%)													1.000	0.120	0.000	0.001	0.072	-0.055	0.024	-0.059
T. Nitrogen (µg/L)														1.000	0.596	-0.703	0.649	0.388	0.053	0.037
T. Phosphorus (µg/L)															1.000	-0.651	0.517	0.588	0.236	0.144
SDT (m)																1.000	-0.600	-0.511	-0.198	0.013
Color (color units)																	1.000	0.435	0.080	-0.078
Chlorophyll *a* (µg/L)																		1.000	0.059	0.086
Alkalinity (mg/L as $CaCO_3$)																			1.000	0.704
pH (S.U.)																				1.000

Table 12. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except TN & TP

			Eigenvector, coefficient of								
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	LKMNDEP	LKAREA	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.459	0.273	0.033	0.033	0.470	0.229	0.571	-0.430	-0.419	-0.194	-0.019
PRIN 2	1.647	0.456	-0.046	0.130	0.240	0.149	0.034	-0.160	0.251	0.624	0.650
PRIN 3	1.166	0.586	0.409	0.766	0.299	0.029	-0.015	0.238	0.192	-0.006	-0.250
PRIN 4	1.113	0.709	-0.156	-0.265	0.261	0.784	-0.099	0.388	0.197	-0.079	-0.134
PRIN 5	0.964	0.817	0.893	-0.356	-0.146	0.172	-0.055	-0.075	-0.055	0.052	0.104
PRIN 6	0.568	0.880	0.034	-0.160	0.130	-0.130	0.126	-0.383	0.801	-0.365	-0.043
PRIN 7	0.460	0.931	0.080	-0.354	0.416	-0.473	0.361	0.573	0.044	0.095	0.046
PRIN 8	0.421	0.978	0.026	0.138	0.073	-0.016	-0.197	0.201	-0.120	-0.648	0.679
PRIN 9	0.202	1.000	-0.007	0.171	-0.590	0.197	0.689	0.247	0.160	-0.046	0.145

N= 472

Table 13. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.077	0.297	-0.044	0.050	-0.567	0.467	0.545	0.362	0.168
PRIN 2	1.545	0.517	-0.047	0.039	0.225	-0.424	0.015	0.549	0.681
PRIN 3	1.097	0.674	0.661	0.734	0.066	-0.019	0.046	0.083	-0.098
PRIN 4	0.919	0.806	0.746	-0.646	-0.106	-0.010	-0.019	0.001	0.117
PRIN 5	0.561	0.886	-0.003	0.040	-0.101	0.487	-0.785	0.362	0.060
PRIN 6	0.427	0.947	-0.008	-0.192	0.410	0.029	0.194	0.619	-0.610
PRIN 7	0.373	1.000	0.036	-0.022	0.659	0.603	0.214	-0.214	0.329

N= 472

Table 14. The eigenvalues and eigenvectors of the correlation matrix for NECZ ecoregion, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.683	0.383	-0.435	-0.016	-0.554	0.441	0.451	0.325	0.025
PRIN 2	1.451	0.591	0.146	-0.134	0.106	-0.315	0.148	0.536	0.736
PRIN 3	1.081	0.745	0.169	0.917	-0.149	-0.134	0.228	-0.130	0.146
PRIN 4	0.723	0.848	0.651	-0.228	0.007	0.467	0.400	-0.332	0.189
PRIN 5	0.464	0.914	0.334	0.196	-0.175	0.474	-0.649	0.416	0.025
PRIN 6	0.411	0.973	0.370	0.019	0.146	-0.173	0.353	0.540	-0.629
PRIN 7	0.188	1.000	-0.297	0.225	0.780	0.465	0.129	0.128	0.068

N= 44

Table 15. The eigenvalues and eigenvectors of the correlation matrix for LPH+NEH ecoregions, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.053	0.293	-0.044	0.042	-0.554	0.440	0.540	0.394	0.220
PRIN 2	1.543	0.514	-0.034	0.107	0.271	-0.458	-0.010	0.528	0.652
PRIN 3	1.129	0.675	0.633	0.744	0.092	0.060	0.038	0.090	-0.157
PRIN 4	0.912	0.805	0.770	-0.594	-0.144	-0.094	-0.004	-0.028	0.154
PRIN 5	0.583	0.889	0.013	0.001	-0.132	0.483	-0.801	0.296	0.142
PRIN 6	0.408	0.947	0.010	-0.267	0.354	0.050	0.109	0.657	-0.598
PRIN 7	0.371	1.000	0.053	-0.097	0.669	0.590	0.230	-0.195	0.317

N= 428

Table 16. The eigenvalues and eigenvectors of the correlation matrix for Reference data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.322	0.332	0.470	0.046	0.569	-0.492	-0.421	0.076	0.169
PRIN 2	1.531	0.550	-0.041	-0.096	-0.166	-0.046	0.170	0.687	0.677
PRIN 3	1.131	0.712	0.148	0.876	0.088	-0.045	0.421	0.130	-0.086
PRIN 4	0.727	0.816	0.732	-0.114	0.000	0.669	0.037	0.053	0.011
PRIN 5	0.580	0.899	0.166	-0.318	0.194	-0.246	0.722	-0.434	0.255
PRIN 6	0.427	0.960	-0.042	0.329	-0.202	0.153	-0.304	-0.560	0.648
PRIN 7	0.282	1.000	-0.437	0.012	0.750	0.472	0.039	0.040	0.141

N= 141

Table 17. The eigenvalues and eigenvectors of the correlation matrix for Test data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.041	0.292	-0.044	0.089	-0.571	0.508	0.547	0.325	0.027
PRIN 2	1.586	0.518	-0.115	-0.156	0.189	-0.321	0.132	0.582	0.683
PRIN 3	1.103	0.676	0.661	0.705	0.116	-0.130	0.147	0.047	0.110
PRIN 4	0.909	0.806	0.734	-0.634	-0.111	0.066	-0.107	0.160	-0.074
PRIN 5	0.545	0.884	0.019	0.175	-0.130	0.529	-0.695	0.034	0.433
PRIN 6	0.438	0.946	-0.065	0.088	0.535	0.329	-0.097	0.598	-0.476
PRIN 7	0.377	1.000	0.071	-0.174	0.555	0.479	0.396	-0.411	0.317

N= 236

Table 18. The eigenvalues and eigenvectors of the correlation matrix for Impaired data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.795	0.399	-0.366	0.014	-0.553	0.438	0.463	0.316	0.230
PRIN 2	1.651	0.635	0.275	0.409	0.098	-0.284	-0.087	0.585	0.563
PRIN 3	1.002	0.778	-0.119	0.816	0.015	0.242	-0.123	0.056	-0.492
PRIN 4	0.735	0.883	0.785	-0.010	0.113	0.390	0.452	-0.004	-0.125
PRIN 5	0.428	0.944	-0.127	0.297	0.054	-0.537	0.687	-0.361	0.047
PRIN 6	0.238	0.978	-0.225	-0.267	0.433	-0.146	0.284	0.623	-0.448
PRIN 7	0.151	1.000	-0.306	0.081	0.693	0.455	0.080	-0.192	0.411

N= 95

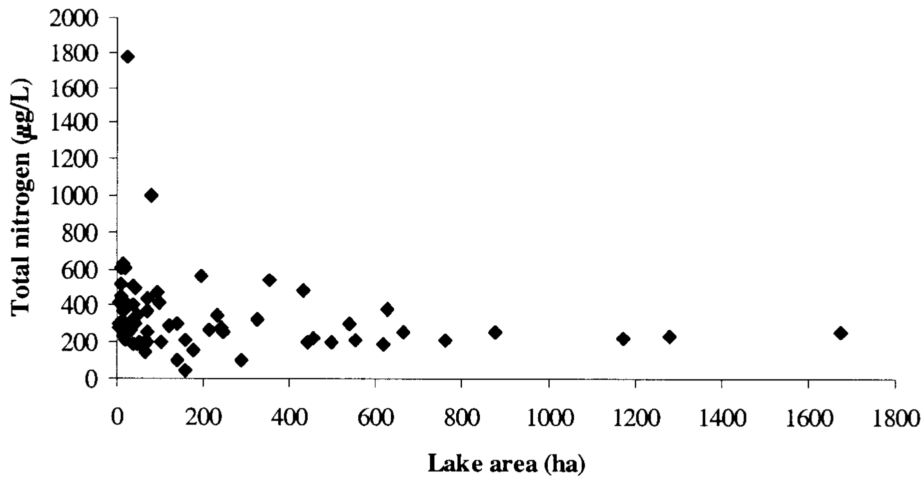

Figure 1. Scatter plot showing the relationship between Total Nitrogen (y- axis) and Lake area (x-axis) for the reference lakes. Note that some datapoints were beyond the scale on the x-axis

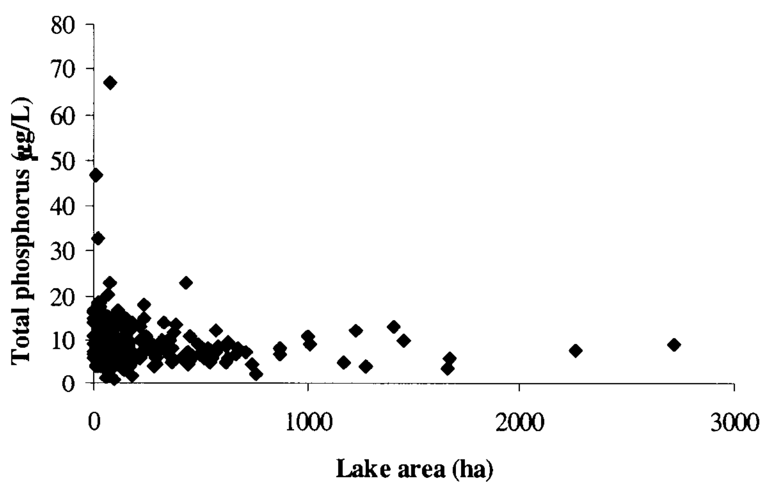

Figure 2. Scatter plot showing the relationship between Total Phosphorus (y- axis) and Lake area (x-axis) for the reference lakes. Note that some datapoints were beyond the scale on the x-axis

Appendix A

PCA was run seven ways:

1) Overall data (all parameters, excluding TN)
2) Overall (all parameters excluding TN, mean depth and surface area)
3) NECZ (all parameters excluding TN, mean depth and surface area)
4) LPH and NEH combined (all parameters excluding TN, mean depth and surface area)
5) Reference (all parameters excluding TN, mean depth and surface area)
6) Test (all parameters excluding TN, mean depth and surface area)
7) Impaired (all parameters excluding TN, mean depth and surface area)

PCA Results

Results from the PCA have been divided by assessment type/ecoregion and also by the inclusion of parameters. Each analysis is presented in Tables A1 through A7. Each table includes information on the eigenvalues, the cumulative percentage of the total variance contributed by each principal component, and the individual principal components (X_1-X_n) (or eigenvectors).

<u>PCA #A1 - Overall data</u> (all parameters, excluding TN) - Table A1

The eigenvalue for a principal component indicates the percentage of the variance that the Principal component accounts for. The first principal component accounts for (2.861/10) *100% = 28%, the second for (1.703/10) or 17%, the third and fourth for 11% each, and the fifth for about 9%. Cumulatively, the second component accounts for 45%, the third for 58% the fourth for 68% and the fifth for 78%, making the last five components account for the remainder of the variance (approximately 22%)

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by lake area with an inverse value for Osgood index.

<u>PCA #A2 - Overall</u> (all parameters excluding TN, mean depth and surface area) – Table A2

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A3 - NECZ</u> (all parameters excluding TN, mean depth and surface area) – Table A3

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with inverse values for alkalinity, color, and SDT. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A4 - LPH and NEH combined</u> (all parameters excluding TN, mean depth and surface area) – Table A4

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A5 - Reference</u> (all parameters excluding TN, mean depth and surface area) – Table A5

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A6 - Test</u> (all parameters excluding TN, mean depth and surface area) – Table A6

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for color. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A7 - Impaired</u> (all parameters excluding TN, mean depth and surface area) – Table A7

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

Table A1. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except TN

			Eigenvector, coefficient of									
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	LKMNDEP	LKAREA	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.861	0.286	-0.029	-0.019	-0.374	-0.164	0.428	-0.501	0.382	0.441	0.230	0.065
PRIN 2	1.703	0.456	-0.031	0.133	0.353	0.225	0.184	0.152	-0.220	0.196	0.551	0.599
PRIN 3	1.184	0.575	0.309	0.539	0.405	0.347	0.176	0.024	0.323	0.216	-0.138	-0.359
PRIN 4	1.139	0.689	-0.312	-0.600	0.118	0.668	0.133	-0.020	0.194	0.063	-0.137	-0.062
PRIN 5	0.964	0.785	0.893	-0.357	-0.147	0.168	0.003	-0.053	-0.078	-0.056	0.051	0.104
PRIN 6	0.692	0.854	0.033	-0.208	0.114	-0.283	0.529	0.338	-0.420	0.407	-0.319	-0.156
PRIN 7	0.473	0.902	0.049	-0.278	0.277	-0.375	0.219	0.376	0.533	-0.317	0.336	-0.130
PRIN 8	0.435	0.945	0.064	-0.073	0.280	-0.227	-0.200	-0.041	0.363	0.185	-0.550	0.591
PRIN 9	0.349	0.980	-0.022	0.232	-0.179	0.136	0.599	-0.041	-0.009	-0.604	-0.292	0.291
PRIN 10	0.199	1.000	-0.006	0.157	-0.580	0.191	-0.087	0.681	0.254	0.218	-0.028	0.136

N= 471

Table A2. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.574	0.322	-0.033	0.027	0.491	-0.474	0.391	0.510	0.314	0.142
PRIN 2	1.545	0.515	-0.046	0.039	0.008	0.233	-0.431	0.010	0.544	0.679
PRIN 3	1.097	0.652	0.658	0.738	-0.015	0.056	-0.011	0.050	0.089	-0.095
PRIN 4	0.922	0.767	0.749	-0.646	0.073	-0.061	-0.047	-0.021	-0.025	0.098
PRIN 5	0.679	0.852	-0.034	-0.016	0.551	0.410	-0.453	0.437	-0.300	-0.202
PRIN 6	0.465	0.910	-0.009	-0.106	0.337	0.464	0.311	-0.351	0.539	-0.387
PRIN 7	0.388	0.959	0.028	0.095	0.237	0.370	0.513	-0.160	-0.449	0.554
PRIN 8	0.329	1.000	0.025	-0.127	-0.528	0.442	0.307	0.630	0.124	-0.050

N=471

Table A3. The eigenvalues and eigenvectors of the correlation matrix for NECZ ecoregion, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	3.140	0.393	-0.369	-0.027	0.425	-0.475	0.366	0.463	0.322	0.063
PRIN 2	1.518	0.582	0.242	-0.123	0.223	0.223	-0.392	0.123	0.435	0.684
PRIN 3	1.083	0.718	0.187	0.912	0.053	-0.127	-0.135	0.234	-0.169	0.109
PRIN 4	0.813	0.819	0.491	-0.240	0.476	0.251	0.136	0.416	-0.437	-0.173
PRIN 5	0.644	0.900	0.458	-0.064	-0.387	-0.258	0.614	-0.056	-0.053	0.432
PRIN 6	0.431	0.954	0.474	0.148	0.088	0.044	0.135	-0.153	0.669	-0.505
PRIN 7	0.223	0.981	-0.238	0.251	0.405	0.502	0.487	-0.433	-0.018	0.202
PRIN 8	0.149	1.000	-0.196	0.073	-0.468	0.571	0.208	0.573	0.192	-0.057

N = 44

Table A4. The eigenvalues and eigenvectors of the correlation matrix for LPH+NEH ecoregions, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.797	0.350	-0.029	0.025	0.542	-0.454	0.360	0.485	0.324	0.170
PRIN 2	1.544	0.543	-0.035	0.108	-0.006	0.263	-0.452	-0.004	0.533	0.655
PRIN 3	1.129	0.684	0.633	0.743	0.002	0.093	0.058	0.037	0.088	-0.158
PRIN 4	0.913	0.798	0.770	-0.596	0.030	-0.122	-0.113	-0.010	-0.041	0.145
PRIN 5	0.622	0.876	-0.022	-0.018	0.254	0.291	-0.521	0.660	-0.314	-0.213
PRIN 6	0.420	0.928	0.015	-0.266	0.177	0.471	0.138	-0.059	0.603	-0.539
PRIN 7	0.374	0.975	0.050	-0.052	0.085	0.626	0.573	0.131	-0.299	0.400
PRIN 8	0.202	1.000	0.013	-0.076	-0.776	-0.053	0.175	0.555	0.222	-0.030

N= 427

Table A5. The eigenvalues and eigenvectors of the correlation matrix for Reference data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.841	0.355	-0.406	-0.043	0.473	-0.513	0.423	0.399	0.007	-0.090
PRIN 2	1.572	0.552	0.056	-0.072	0.147	-0.052	-0.139	0.098	0.686	0.683
PRIN 3	1.134	0.693	0.171	0.876	0.051	0.117	-0.064	0.400	0.110	-0.105
PRIN 4	0.728	0.784	0.747	-0.123	0.050	0.021	0.649	0.028	0.044	0.000
PRIN 5	0.584	0.857	0.194	-0.335	0.147	0.244	-0.307	0.678	-0.420	0.197
PRIN 6	0.484	0.918	0.153	-0.177	0.619	0.244	-0.283	-0.110	0.348	-0.537
PRIN 7	0.380	0.965	0.135	0.257	0.567	-0.106	-0.095	-0.444	-0.466	0.398
PRIN 8	0.278	1.000	-0.406	0.035	0.152	0.768	0.440	-0.030	-0.009	0.166

N=141

Table A6. The eigenvalues and eigenvectors of the correlation matrix for Test data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.495	0.312	-0.037	0.053	0.486	-0.463	0.404	0.513	0.336	0.075
PRIN 2	1.601	0.512	-0.101	-0.168	0.081	0.278	-0.396	0.055	0.525	0.665
PRIN 3	1.104	0.650	0.664	0.702	0.025	0.128	-0.138	0.139	0.035	0.100
PRIN 4	0.911	0.764	0.731	-0.642	0.068	-0.077	0.040	-0.120	0.132	-0.097
PRIN 5	0.649	0.845	-0.057	-0.120	0.593	0.395	-0.418	0.326	-0.209	-0.388
PRIN 6	0.463	0.903	-0.021	0.109	0.411	0.503	0.558	-0.483	0.139	0.027
PRIN 7	0.425	0.956	-0.075	0.081	-0.311	0.173	-0.022	0.080	0.703	-0.600
PRIN 8	0.353	1.000	0.057	-0.179	-0.368	0.497	0.414	0.595	-0.188	0.147

N=235

Table A7. The eigenvalues and eigenvectors of the correlation matrix for Impaired data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	3.564	0.445	-0.299	-0.008	0.482	-0.478	0.391	0.441	0.260	0.187
PRIN 2	1.655	0.652	0.247	0.413	-0.043	0.067	-0.261	-0.071	0.601	0.573
PRIN 3	1.010	0.779	-0.203	0.801	-0.092	-0.039	0.257	-0.136	0.072	-0.467
PRIN 4	0.800	0.879	0.781	0.108	0.269	0.230	0.251	0.382	-0.050	-0.206
PRIN 5	0.455	0.935	0.225	-0.299	-0.226	-0.133	0.644	-0.516	0.331	0.000
PRIN 6	0.255	0.967	-0.227	-0.269	0.261	0.457	-0.157	0.020	0.610	-0.450
PRIN 7	0.151	0.986	-0.310	0.074	-0.040	0.695	0.456	0.117	-0.176	0.402
PRIN 8	0.110	1.000	0.032	0.096	0.751	0.062	-0.052	-0.596	-0.227	0.108

N=95