2012-07-02

Principal Components Analysis, Factor Analysis and Trend Correlations of Twenty-Eight Years of Water Quality Data of Deer Creek Reservoir, Utah

Nicolas Alejandro Gonzalez
Brigham Young University - Provo

Principal Components Analysis, Factor Analysis and Trend Correlations

of Twenty-Eight Years of Water Quality Data

# of Deer Creek Reservoir, Utah

Nicolas Alejandro Gonzalez

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Gustavious P. Williams, Chair
Everett James Nelson
A. Woodruff Miller

Department of Civil and Environmental Engineering

Brigham Young University

August 2012

ABSTRACT

Principal Components Analysis, Factor Analysis and Trend Correlations
of Twenty-Eight Years of Water Quality Data
of Deer Creek Reservoir, Utah

Nicolas Alejandro Gonzalez
Department of Civil and Environmental Engineering, BYU
Master of Science

I evaluated twenty-eight years (1980-2007) of spatial-temporal water quality data from Deer Creek Reservoir in Utah. The data came from three sampling points representing the lotic, transitional and lentic zones. The data included measurements of climatological, hydrological and water quality conditions at four depths; Surface, Above Thermocline, Below Thermocline and Bottom. The time frame spanned dates before and after the completion of the Jordanelle Reservoir (1987-1992), approximately fourteen miles upstream of Deer Creek. I compared temporal groupings and found that a traditional month distribution following standard seasons was not effective in characterizing the measured conditions; I developed a more representative seasonal grouping by performing a Tukey-Kramer multiple comparisons adjustment and a Bonferronian correction of the Student's t comparison. Based on these analyses, I determined the best groupings were Cold (December - April), Semi-Cold (May and November), Semi-Warm (June and October), Warm (July and September) and Transition (August). I performed principal component analysis (PCA) and factor analysis (FA) to determine principal parameters associated with the variability of the water quality of the reservoir. These parameters confirmed our seasonal groups showing the Cold, Transition and Warm seasons as distinct groups. The PCA and FA showed that the variables that drive most of the variability in the reservoir are specific conductivity and variables related with temperature. The PCA and FA showed that the reservoir is highly variable. The first 3 principal components and rotated factors explained a cumulative 59% and 47%, respectively of the variability in Deer Creek. Both parametric and nonparametric approaches provided similar correlations but the evaluations that included censored data (nutrients) were considerably different with the nonparametric approach being preferred.

Keywords: PCA, factor analysis, time trends, nutrient loading, nonparametric, Deer Creek

ACKNOWLEDGMENTS

I would not have been able to complete this thesis without the support and council from Gustavious P. Williams who has always been there willing to help me complete projects and presentations no matter how close I was to the deadline. The United States Bureau of Reclamation, Upper Colorado Region for sponsoring the BYU Deer Creek research group. The Central Utah Water Conservancy District (CUWCD) for providing the data used in this study. I would also like to thank Dr. Oliver Obregon who was my other mentor in the area of environmental engineering and Dr. Eggett for teaching me almost everything I know about statistics. I would like to thank one of my greatest examples, my dad, who is still the smartest person I know. I also want to thank my mom who always believed in me no matter what. I would like to thank my wife Stacey and my daughter Abby who had to put up with the long student hours and always asked me when I was going to start and later finish the thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Deer Creek Reservoir was built in 1938 as part of the Provo River Project (Eckhoff *et al.*, 2002). It is located on the Provo River approximately 20 miles to the north-east of Provo, Utah. Deer Creek is a major source of municipal and agricultural water for Utah and Salt Lake counties with 38% used for irrigation and 62% for culinary use. The reservoir's characteristics are shown in Table 1-1 (UDEQ, 2000).

Table 1-1. Characteristics of Deer Creek Reservoir

Deer Creek Reservoir	
Lake Elevation (meters / feet)	1651/5417
Surface Area (hectares / acres)	1200/2965
Watershed Area (hectares / acres)	187000/462000
Volume (m^3 / acre-feet)	$2.388*10^8$/193614
Retention Time (years)	1.3
Depth (meters / feet)	
maximum	42/137
mean	20/65
Length (km / miles)	9.2/5.7
Width (km / miles)	1.9/1.2

Deer Creek is a focus of water quality research at Brigham Young University because of the potential proliferation of late-summer algae blooms (Eckhoff et al., 2002; Miller, 2008).

Even though the quality of the reservoir has improved since the 1970s, algae have shifted from a predominance of blue-green algae to green algae (UDEQ, 2000). The green algae can cause issues in the treatment plants because they clog treatment filters.

I evaluated spatial and temporal water quality trends using data collected from four locations: UpperEnd, MidLake, Wallsburg and NearDam (Figure 1-1) and common methods for data dimensional reduction (Palma *et al.*, 2010; Praus, 2006; Vega *et al.*, 1998; Yang *et al.*, 2010). These locations represent the lotic, transitional and lentic zones. Deer Creek is fed by four main inflows: Provo River, Snake Creek, Daniel's Creek and Main Creek. UpperEnd captures the inflow from Daniels Creek, Provo River and Snake Creek.

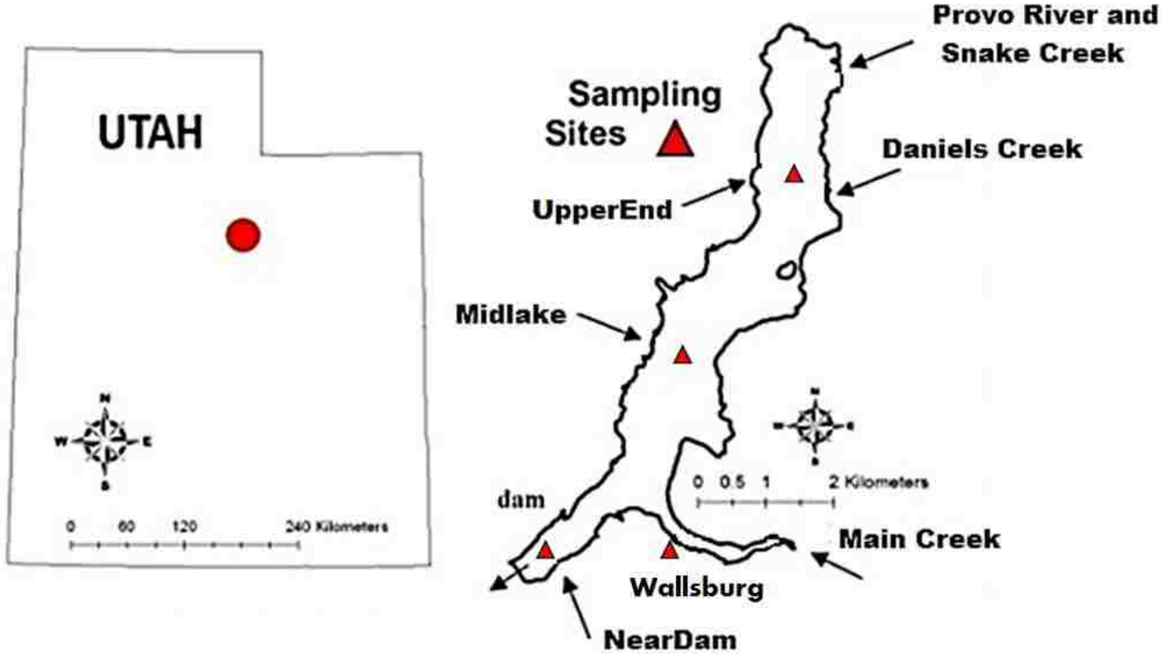

Figure 1-1. Deer Creek Reservoir. Inflows and Sampling Locations

There has been significant effort devoted to understanding the behavior of Deer Creek Reservoir and to identify the main variables that drive the reservoir. One of the purposes of this thesis is to understand the temporal variation in the reservoir and identify the parameters which can be used to describe or define this variation. The main methods used in this thesis to detect and understand the variability in the reservoir were two statistical data dimension reduction methods: Principal Components Analysis (PCA) and Factor Analysis (FA).

2 METHODS

2.1 Data Collection

Data were collected during a period spanning 1980 to 2007. Different subsets of this dataset were used for different portions of the water quality assessment based on what data were available for the desired analysis. The samples were collected at four depths; Surface (between 0 and 1 m), Above Thermocline (top of the thermocline), Below Thermocline (bottom of thermocline), and Bottom (bottom of the reservoir).

Water quality data were provided by the Central Utah Water Conservancy District (CUWCD). CUWCD collects samples from the reservoir on average once a month during the sampling season (April-October) and some years, when the weather allows, they collect samples throughout the whole year. CUWCD water quality data are collected using multi-probe sondes and laboratory results from water samples. During sample analysis where the parameter of interested was not detected, the value was replaced with the respective detection limit of the laboratory method. Nutrients, like nitrogen and phosphorus, were considered in this study because of their direct effect on algae population (Rahman *et al.*, 2005). These parameters had a large number of samples below the detection limit. These samples (below the detection limit) are considered censored data.

Table 2-1 shows the parameters used in this study, though some of them were not included in all of the analysis due to a lack of consistency of measurements.

Table 2-1. Parameter Names and Descriptions

	Model Parameter Name	Parameter Description	Units
Temporal-Spatial Data	MONTH	Month of reading	-
	YEAR	Year of reading	-
	DEPTH	Depth of Reading (Numerical and Categorical)	m
	LOCATION	Sampling site location	-
Water Quality Data	CHL-a	Chlorophyll-*a*	µg/L
	DO	Dissolved oxygen	mg/L
	pH	pH	-
	SP COND	Specific Conductance	mS/cm
	NO2-NO3-N	Inorganic Nitrogen	mg/L
	NH3	Ammonia	mg/L
	PO4	Total phosphorus	mg/L
	WATER T	Water temperature	ºC
Climate Data	AIR T	Air temperature	ºC
	TD	Dew point temperature	ºC
	PRECIP	Annual Precipitation	mm
Hydraulic Data	VOL	Reservoir volume	M m^3
	INF	Inflow calculated from outflow and reservoir volume	m^3/s
	OUT	Outflow measured at the dam	m^3/s
	PROVO UP	Provo river inflow	m^3/s
	PROVO DOWN	Provo river flow downstream of dam	m^3/s
	SNAKE	Snake Creek inflow	m^3/s

Climatological data (air temperature and dew point temperature) were obtained from the National Oceanic and Atmospheric Administration (NOAA) via the National Climatic Data Center. The collection site used to approximate weather conditions at Deer Creek was the Salt

Lake City International Airport. Although there are other sites closer to our study area, they did not have consistent hourly data for the range of dates needed for our study. Precipitation data were also obtained from NOAA, but in this case there is a station located at the dam for Deer Creek Reservoir.

2.2 Statistical Analyses

The statistical software used to conduct the statistical analyses was JMP version 9.0 (SAS Institute Inc, 2007).

2.2.1 Seasonal Distribution

Researchers have used several methods to characterize temporal data into seasons (or similar time periods). Once these similar time periods have been determined, they can be analyzed using data dimension reduction (i.e., PCA-type) approaches. For example Rangel-Peraza *et al.* (2009) used three seasons called Warm dry (March to June), Rainy (November to February) and Cold dry (July to October). Fan *et al.* (2012) determined seasonality in their data by using a continuous wavelet transform and assigned categories of dry season or wet season for computed values of less than 1 and more than 1, respectively. Reservoirs do not necessarily follow the traditional seasonal pattern established by a calendar, and in order to study them properly it is useful to impose some sort of a structure based on statistical similarity in time.

2.2.1.1 Tukey-Kramer Multiple Comparisons

I performed a Tukey-Kramer multiple comparisons adjustment to evaluate the relationship between water temperature and month. I grouped the months according to whether their differences computed by the analysis were statistically significant or not. I stopped the iterative grouping when there was a significant difference between the group means.

2.2.1.2 Bonferronian Correction of Student's t-Multiplier

A Bonferronian correction finds a critical *p*-value that is more fitted to represent statistical significance than methods like Student's t-multiplier or Tukey-Kramer multiple comparisons (Ramsey *et al.*, 2002). The critical *p*-value is found by equation 2-1.

$$p - value = 100\left(1 - \frac{\alpha}{k}\right)\% \qquad (2\text{-}1)$$

where α is the confidence level and *k* is the number of comparisons (in the case of comparing 12 months it would be 66 comparisons). The calculated critical *p*-value is compared to the *p*-value obtained from the Student's t-test. When the Student's *p*-value is lower than the critical *p*-value, the comparison is considered significantly different (Ramsey & Schafer, 2002).

2.2.2 Correlation Between Variables

Two correlation matrices of all the variables used in the study were created. One with a parametric correlation and the other with a nonparametric correlation. The correlations were measured in R^2 values (Fan et al., 2012; Vega et al., 1998). There are several methods to estimate the correlations, according to the JMP manual (2007), the recommended estimation method for large datasets with several missing data, like the one in this study (See Appendix B), is the Maximum Likelihood or ML method. Therefore ML was used as the estimation method.

2.2.2.1 Parametric Correlation

JMP uses Pearson's *r* as the correlation coefficient to calculate the correlation between variables (SAS Institute Inc, 2007) as defined in equation 2-2 (D. R. Helsel, 2012). The same author warns of the uses of parametric approaches in environmental data because of the skewness of the data and the presence of non-detects which could lead to finding correlations or signals when there is none in reality or the failure to detect correlations where they are present in the data. Equation 2-2 computes *r* from calculated means and standard deviations.

$$Pearson's\ r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\times\left(\frac{y_i-\bar{y}}{s_y}\right) \qquad (2\text{-}2)$$

2.2.2.2 Nonparametric Correlation

The nonparametric version of Pearson's *r* is called Spearman's *rho* (ρ) or Spearman rank correlation which is also computed using equation 2-2 but applying it to the ranks of the data (D. R. Helsel, 2012; Higgins, 2003). Higgins (2003) in his book *Introduction to Modern Nonparametric Statistics* explains how some correlations between two variables can be misleading if parametric approaches are used.

A Spearman's ρ nonparametric test was performed to compare the variables used in the study. Nonparametric correlations have greater power when dealing with data that is censored and not normally distributed (D. Helsel *et al.*, 2002). The data set for this thesis is heavily censored in data that describe the nutrients like phosphate, ammonia and inorganic nitrogen. Also most of the variables are not normally distributed (See Appendix B), and therefore a nonparametric correlation was performed. As explained in the parametric correlation, the estimation method used was Maximum Likelihood.

2.2.3 PCA and FA

2.2.3.1 Principal Components Analysis

The main purpose of PCA is to capture the variability of a dataset while reducing the dimensionality of the same (Jolliffe, 2002). According to Wong *et al.* (2002), PCA uses a linear transformation to accomplish the data reduction. The linear transformation is represented by a set of vectors (commonly called eigenvectors) that maximize the variance of the dataset.

A PCA was used to find linear combinations of our data that captured most of the variability of Deer Creek, these linear combinations are called principal components (PCs). The eigenvalues were used to determine which components were significant in explaining the variability and the eigenvectors were used to group the variables into a subset of common factors. I labeled the dataset using categorical data to assess the temporal and spatial variation by observing how the categories grouped in a plot using the different PCs (Praus, 2006). I used the Varimax method which is the recommended rotational method for orthogonal rotations (SAS Institute Inc, 2007).

Jolliffe (2002) in *Principal Components Analysis* provides an in-depth description of the uses of PCA in regression analysis and other multivariate techniques. PCs can be helpful in these areas of statistics because they solve the problems of multicollinearities between variables that many times it are present but ignored or not accounted for (Jolliffe, 2002). Costello *et al.* (2005) explains that generally PCA overestimates the amount of variance explained by its components as opposed to other methods like FA.

2.2.3.2 Factor Analysis

I used FA to further simplify the dataset by determining the most significant variables that describe the behavior of Deer Creek (Wunderlin *et al.*, 2001). I used the recommended Quartimin method for oblique rotations (SAS Institute Inc, 2007). These rotations are not necessarily orthogonal like the rotations of the components in a PCA (Jolliffe, 2002).

2.2.3.3 Principal Components and Rotated Factors Retained

There are two widely used methods to determine the number of PCs and rotated factors (RFs) to be retained in the analysis. The first one and least accurate according to Costello and Osborne (2005) is to retain the factors which have eigenvalues greater than 1. The same authors performed a Monte Carlo simulation and found that with this method 36% of the populations retained more factors than necessary, showing that picking the eigenvectors with an eigenvalue greater than 1 can retain vectors that do not significantly contribute to explaining the variability of the dataset.

The second method is commonly called the Scree test. The objective of this test is to find the point where there is a break in the plot of eigenvalues versus the number of components. In other words to pick the number of components where the curve stops being steep and it becomes flatter (StatSoft, 2004). See Figure 2-1.

In Figure 2-1 the researcher could chose 4 or 5 components. In the analysis performed for this thesis, I chose to retain the number of components right before the break as recommended by Costello and Osborne (2005), therefore for the scree plot shown in Figure 2-1 I chose to retain 4 PCs.

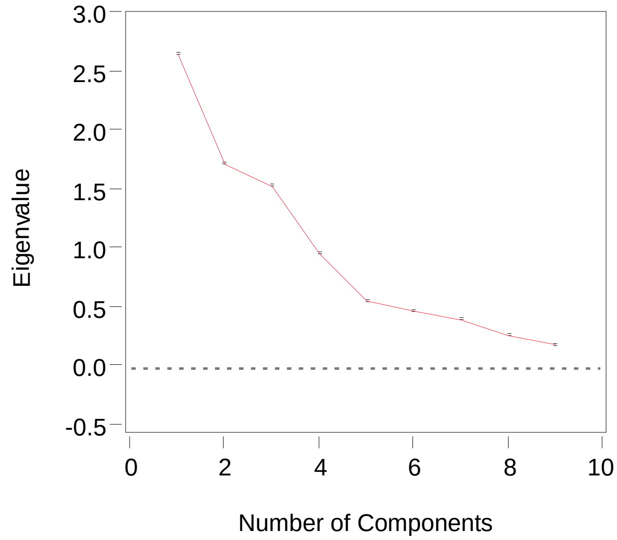

Figure 2-1. Example Scree Plot. Eigenvalue Versus Components

The number of RFs retained has more influence on the results than the number of PCs retained. For example if the researcher choses three PCs and later decides to include a fourth PC, the first three PCs will be unaffected. This is not true for FA, changing from three to four factors may change the distributions and loading of all three previous factors (Jolliffe, 2002).

2.2.3.4 Variable Loading in Eigenvectors

Once I picked the number of eigenvectors, I evaluate the weights of each variable in the eigenvectors.

The first method was to select the values in the eigenvector that are greater than a specified value, usually 0.4. For example, when JMP presents a table of the eigenvectors, it highlights the values greater than 0.4 in order to show which variables are important or have a heavier weight at the moment of describing a principal component and therefore the variability explained by that component in the dataset (Vega et al., 1998; Wunderlin et al., 2001). This can be seen in Table 2-2 where values greater than 0.4 are presented in bold typeface.

Table 2-2. Example of Variable Loadings in Eigenvectors

Variable	Prin1	Prin2	Prin3	Prin4
Precip	-0.10087	**0.48320**	**0.82122**	-0.00640
Rel	**0.53196**	0.20630	-0.15477	0.18366
Vol	**0.42590**	0.23634	-0.08069	**-0.85303**
Inf	0.33048	**0.59556**	-0.21825	**0.45224**
T Air	**0.47210**	**-0.45324**	0.12754	0.16142
Dew Pt	**0.43973**	-0.32863	**0.48086**	0.08936

By concentrating on the variables that weight the most for each PC or RF, one can understand which variables are the most important in explaining the variability of the dataset or the reservoir.

2.2.3.5 Scatterplots of Principal Components and Rotated Factors

Scatterplot matrices of the three PCs and rotated components were created for analysis. These scatterplots were colored by categorical data in order to visually find indications of segmentation and/or correlations (Praus, 2006).

The categorical data used for this analysis are: Depth, Season, Location, Year and Decade. Plots demonstrating this technique are presented later in Section 3.3.3.

3 RESULTS AND DISCUSSION

3.1 Seasonal Distribution

3.1.1 Tukey-Kramer Multiple Comparisons

I grouped the data into temporal periods using the results from a Tukey-Kramer multiple comparisons. Table 3-1 shows example differences of water temperature between months (the analysis used an exhaustive comparison). Only the comparisons where there was a significant difference were used to distribute the data among seasons. Table 3-2 shows the ordered differences between the months versus water temperature using the Tukey-Kramer multiple comparisons. For a complete table of the Tukey-Kramer multiple comparisons see Table A-1 in the appendix.

Table 3-1. Tukey-Kramer Multiple Comparisons

Month	- Month	Difference	*p*-Value
Dec	Feb	2.22925	0.1621
Mar	Feb	2.09474	0.1407
Apr	Dec	2.01500	0.0772
Dec	Jan	1.99964	0.6739
Mar	Jan	1.86513	0.7008
May	Nov	1.19243	0.1410
Jun	Oct	0.77091	0.3213
Jan	Feb	0.22961	1.0000
Dec	Mar	0.13451	.
Sep	Jul	0.04432	.

Table 3-2. Ordered Differences Report of Tukey-Kramer Multiple Comparisons

Level							Mean
Aug	A						18.532143
Sep		B					17.356339
Jul		B					17.312021
Jun			C				13.881136
Oct			C				13.110229
May				D			10.394030
Nov				D			9.201600
Apr					E		7.152138
Dec					E	F	5.137143
Mar						F	5.002632
Jan						F	3.137500
Feb						F	2.907895

From this grouping I obtained five seasonal groupings. I labeled them Cold (December, January, February, March, and April), Semi-Cold (May and November), Semi-Warm (June and October), Warm (July and October) and Transition (August) shown in Figure 3-1.

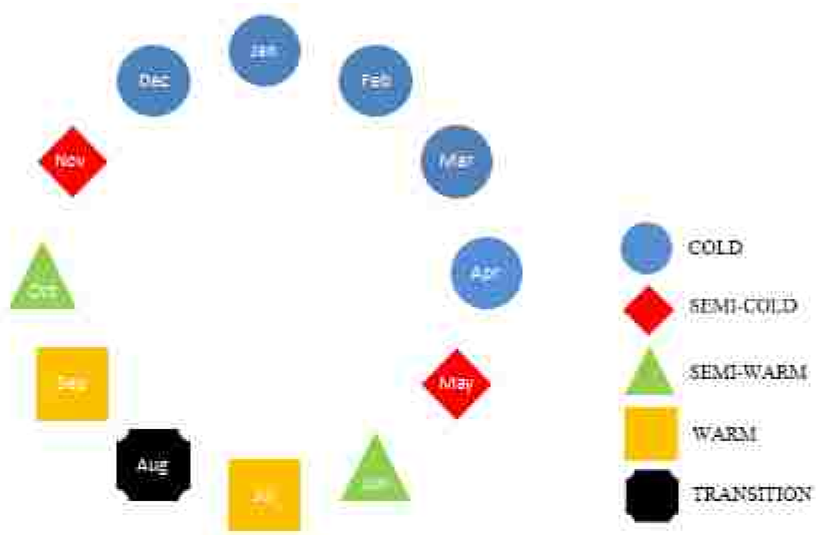

Figure 3-1. Seasonal Distribution by Month

As shown in Figure 3-1, the distribution of the months is not continuous like in the traditional approach but it does not have to be as long as the months show no differences between their means. The data show a lag in the water conditions compared to air conditions. The data indicate that the water does not start to get colder until approximately two months into the fall after air temperatures drop and it does not get warmer until two months into the spring after air temperatures rise. This approach developed groups better suited to statistical analysis.

3.1.2 Bonferronian Correction of Student's t-Multiplier

The Student's *p*-values for each pairwise comparison are shown in Table 3-3. As explained in the methods section these *p*-values are not representative and need to be adjusted by using a Bonferronian correction. By using equation 2-1 the critical *p*-value was found to be 0.0008 (instead of the usual *p*-value of 0.05). All the *p*-values higher than the critical value are highlighted in Table 3-3. See Table A-2 in the appendix for the complete table of Student's *p*-values. This seasonal distribution matches the one in section 3.1.1 (Figure 3-1).

Table 3-3. Bonferronian Correction of Student's t-Multiplier

Month	- Month	Difference	*p*-value
Dec	Feb	2.22925	**0.0045**
Mar	Feb	2.09474	**0.0037**
Apr	Dec	2.015	**0.0018**
Dec	Jan	1.99964	**0.0427**
Mar	Jan	1.86513	**0.0468**
May	Nov	1.19243	**0.0038**
Jun	Oct	0.77091	**0.0115**
Jan	Feb	0.22961	**0.8066**
Dec	Mar	0.13451	**0.8637**
Sep	Jul	0.04432	**0.874**

3.2 Correlation Between Variables

The parametric correlation matrix for the variables used in the study is shown in Table 3-4. The nonparametric correlation matrix is shown in Table 3-5. The correlations with a value around 0.4 or higher are underlined and bolded.

3.2.1 Parametric Correlation

The highest correlation was found between air temperature and water temperature, this correlation was expected as were correlations between release and volume and inflow or DO and pH. Correlations I did not expect were found between PO4-P and NH3-N and between depth and NO2-NO3-N. The correlation between PO4-P and NH3-N may be due to point sources like fertilizer and livestock from nearby farms. The correlation between NH3-N and depth may be due to chemical processes at different reservoir levels influencing redox reactions.

3.2.2 Nonparametric Correlation

By looking at both correlation tables one can see that some correlations are similar, some stronger and some weaker. For example the correlation between PO4-P and NH3-N is weaker (0.39 versus 0.32), while the correlation between NO3-NO2-N and pH is stronger in the nonparametric approach (-0.15 versus -0.47). Interesting correlations are found between NO3-NO2-N and PO4-P versus depth, which predicted higher nutrient concentrations at deeper levels. A study of the reservoir conducted by the Utah Department of Environmental Quality explains that the stratification develops anoxic conditions in the bottom of the reservoir, these anoxic conditions release phosphorus from the sediment loads deposited at the bottom of the reservoir (UDEQ, 2000). My statistical analysis supports this hypothesis and helps understand nutrient distributions in the reservoir.

Table 3-4. Parametric Correlation Between Variables

Variable	Rel	Vol	Inf	pH	Sp Cond	DO	Water T	Chl-a	PO4-P	NO2-NO3-N	NH3-N	Secchi	Air T	Precip	Depth
Rel	1.00														
Vol	**0.40**	1.00													
Inf	**0.44**	0.27	1.00												
pH	-0.05	-0.11	0.06	1.00											
Sp Cond	**-0.41**	**-0.46**	-0.18	0.12	1.00										
DO	-0.02	0.12	0.12	**0.54**	0.12	1.00									
Water T	0.13	-0.11	-0.13	0.22	-0.34	-0.17	1.00								
Chl-a	-0.03	-0.04	-0.12	0.00	-0.05	0.16	0.00	1.00							
PO4-P	0.02	0.00	-0.08	-0.28	-0.10	-0.30	-0.04	0.10	1.00						
NO2-NO3-N	0.04	0.11	0.02	-0.37	0.08	-0.16	-0.51	-0.11	0.10	1.00					
NH3-N	0.00	-0.01	-0.08	-0.15	-0.02	-0.15	-0.02	0.09	**0.39**	0.01	1.00				
Secchi	0.02	0.15	0.11	-0.08	0.09	-0.10	-0.17	-0.29	-0.07	0.04	-0.02	1.00			
Air T	0.37	0.17	-0.03	-0.05	**-0.43**	**-0.42**	**0.73**	0.00	0.05	-0.29	-0.02	-0.09	1.00		
Precip	0.26	0.38	0.00	-0.20	**-0.39**	0.01	-0.05	0.01	0.22	0.24	0.17	-0.06	-0.03	1.00	
Depth	0.05	0.12	0.02	**-0.53**	0.09	**-0.57**	**-0.43**	-0.06	0.26	**0.40**	0.14	0.16	0.00	0.07	1.00

Table 3-5. Nonparametric Correlations Between Variables

Variable	Rel	Vol	Inf	pH	Sp Cond	DO	Water T	Chl-a	PO4-P	NO2-NO3-N	NH3-N	Secchi	Air T	Precip	Depth
Rel	1.00														
Vol	**0.45**	1.00													
Inf	0.23	0.27	1.00												
pH	-0.06	-0.12	0.07	1.00											
Sp Cond	**-0.46**	**-0.47**	-0.08	0.13	1.00										
DO	-0.07	0.10	0.15	**0.52**	0.14	1.00									
Water T	0.20	-0.09	-0.18	0.26	-0.37	-0.20	1.00								
Chl-a	0.00	0.01	-0.14	0.02	-0.12	0.14	0.05	1.00							
PO4-P	0.04	0.01	-0.08	-0.35	-0.07	-0.29	-0.17	0.05	1.00						
NO2-NO3-N	-0.03	0.12	0.09	**-0.47**	0.10	-0.12	**-0.59**	-0.05	0.32	1.00					
NH3-N	0.01	0.06	-0.24	-0.25	-0.04	-0.16	-0.14	0.00	0.36	0.16	1.00				
Secchi	0.15	0.20	0.17	-0.12	0.00	-0.13	-0.11	-0.34	-0.09	0.06	-0.03	1.00			
Air T	**0.45**	0.20	-0.13	-0.04	**-0.44**	**-0.45**	**0.70**	0.05	-0.03	-0.33	-0.09	0.03	1.00		
Precip	0.25	0.36	0.00	-0.21	**-0.40**	0.01	-0.06	0.04	0.24	0.32	0.38	0.00	-0.04	1.00	
Depth	0.06	0.11	0.01	**-0.56**	0.07	**-0.53**	**-0.41**	0.10	0.31	**0.43**	0.23	0.18	0.03	0.06	1.00

3.3 PCA and FA

3.3.1 Principal Components Analysis

The variables used to compute the PCA are shown in Table 3-6 which also shows the eigenvectors for the first three principal components.

Table 3-6. Eigenvectors

Parameter	Prin1	Prin2	Prin3
Rel	**0.41**	0.28	0.07
Vol	0.34	0.30	0.28
Inf	0.15	0.38	0.18
pH	-0.19	**0.39**	-0.29
Sp Cond	**-0.48**	-0.13	0.01
DO	-0.23	**0.50**	0.10
Water T	0.27	-0.04	**-0.58**
PO4-P	0.17	**-0.40**	0.25
NH3-N	0.10	-0.32	0.23
Air T	**0.42**	-0.08	**-0.45**
Precip	0.29	0.02	0.37

I chose to use only the first three principal components based on the plot of variance versus component (Scree plot, Figure 3-2). This selection was done by picking the location before the curve flattens out (Costello & Osborne, 2005; Praus, 2006).

The researcher could pick three or five PCs by following the criteria previously explained but for simplicity and to match the procedures performed in the FA section I chose three PCs.

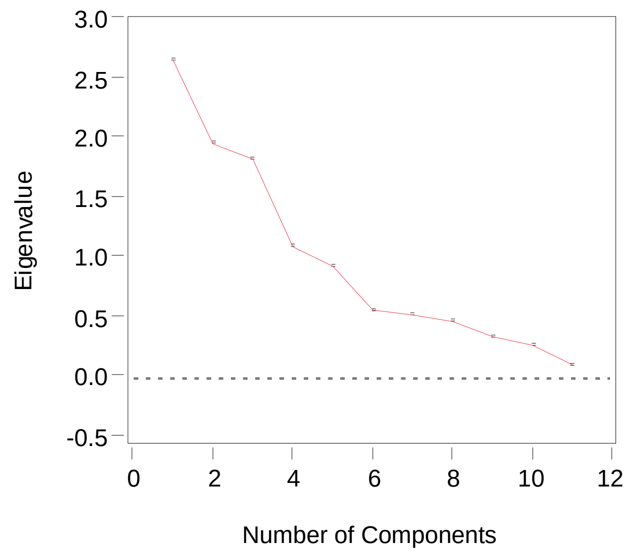

Figure 3-2. Scree Plot

The eigenvalues are shown in Table 3-7. The first three eigenvalues explain 59% of the variation in our data set. With a study of 28 years, higher variations are expected, especially when the study is subject to great spatial distances (Ramsey & Schafer, 2002).

Table 3-7. Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	**2.68**	**24.36**	**24.36**
2	**1.98**	**17.99**	**42.35**
3	**1.85**	**16.82**	**59.17**
4	1.12	10.15	69.33
5	0.96	8.68	78.01
6	0.59	5.36	83.36
7	0.55	4.98	88.34
8	0.49	4.47	92.81
9	0.37	3.33	96.14
10	0.29	2.68	98.82
11	0.13	1.18	100.00

3.3.1.1 Principal Component 1

According to Table 3-7 the first PC explains 24% of the variability. The variables that most heavily load this PC are: Rel, Sp Cond and Air T (See Table 3-6). I called this PC *Dilution* because of the effects of these parameters are in indication of the dilution of other parameters in the reservoir.

According to the box plots of these variables (See Appendix C), Air T and Rel increase in the summer months while Sp Cond decreases in these months. Table 3-6 shows that Sp Cond has a negative value when loading PC1. This suggests that when Rel and Air T have high values and Sp Cond has low values, they load PC1 the highest. On the other hand, the opposite values of these variables would give a low or negative value for PC1.

Sp Cond is the only value in PC1 that is not consistently dominated by month, in other words, the variance of PC1 is almost entirely dominated by seasonality, not by best management practices or other manmade changes in the reservoir.

3.3.1.2 Principal Component 2

The second PC explains 18% of the variability and it is most heavily loaded by pH, DO and PO4-P. I called this PC *Water Quality* because of the relation of the loading variables with water quality. It is believed that the limiting factor of Deer Creek is phosphorus. My analysis supports this statement and shows statistically that changes in nitrogen concentrations are not as important as changes in phosphorus when explaining the variability of the reservoir.

According to Table 3-6, this PC is positively loaded by pH and DO and negatively loaded by PO4-P. As explained in section 3.3.1.1, the magnitude of these variables would determine the level of the PC. This also indicates that lower DO and pH values are associated with higher PO4-P levels as chemistry would suggest.

The box and whisker plots of the selected parameters (Appendix C) show that there is not much variability over months. Therefore seasonality is not a big contributor to this PC. This means that when pH and DO are high and PO4-P levels are low, the PC will load at high levels regardless of what time of the year that happens. This means that the changes in the reservoir explained by PC2 are related to water chemistry, not season.

3.3.1.3 Principal Component 3

I called the third PC *Seasons* because it is most heavily loaded by both Water T and Air T. This PC explains 17% of the variability. The two boxplots of water temperature versus month and air temperature versus month (Appendix C) show that there is a lag of one to two months from the change in air temperature to the change in water temperature. Water chemistry suggests that changes in the quality of the water are far more susceptible to changes in water temperature than in air temperature.

This PC is negatively loaded by these two variables, this means that at higher temperatures (summer months) this PC will load in the low or negative levels. Again a great part of the explained variability is due to seasonality which suggests that most of the changes in the reservoir explained by PC3 are due to the actual month of the year instead of other water quality changes.

3.3.2 Factor Analysis

Table 3-8 shows the RF loading with the first three factors obtained from the FA. Correlated factors were underlined (Wunderlin et al., 2001).

Table 3-8. Rotated Factor Loadings

Parameter	Factor 1	Factor 2	Factor 3
Rel	0.18	**0.59**	-0.02
Vol	-0.06	**0.78**	0.03
Inf	-0.11	0.39	0.04
pH	0.16	-0.09	**0.65**
Sp Cond	**-0.40**	**-0.58**	0.05
DO	-0.25	0.23	**0.94**
Water T	**1.00**	-0.13	0.19
PO4-P	0.00	0.02	-0.33
NH3-N	0.00	-0.01	-0.16
Air T	**0.78**	0.15	-0.19
Precip	-0.01	**0.46**	-0.04

Table 3-9 shows the variance explained by each of the three RF. These explained variances were expected to be lower than the ones of the PCA (Costello & Osborne, 2005).

Table 3-9. Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	1.9393	17.630	17.630
Factor 2	1.7657	16.052	33.682
Factor 3	1.5078	13.707	47.389

3.3.2.1 Rotated Factor 1

Factor 1 explains 18% of the variability of the dataset. It is most heavily loaded by Sp Cond, Water T and Air T. For FA, these three variables explained most of the variability in Deer Creek. For PCA the seasonality appeared in the third PC but in the case of FA the seasonality showed up in the first factor. This suggests that 18% of the variability of the reservoir is due mostly to the time of year and not as much because of other water quality factors.

Just like in PC1, the first RF gives a negative value to Sp Cond, therefore all the assumptions made in section 3.3.1.1 (PC1) apply to RF1. I called this factor *Seasons*.

3.3.2.2 Rotated Factor 2

For factor 2, Sp Cond, Rel, Vol and Precip were grouped. I called this factor *Capacity* because of the relation of the included parameters to the volume of water in the reservoir.

This factor explains 16% of the data variability and just like PC1. Sp Cond has a negative correlation when loading this factor. The direction of this factor could determine the water quality of the reservoir, in other words when Sp Cond is high and Rel, Vol and Precip are low, the factor will go in the left direction. Knowing which direction the factor will go can help understand the scatterplots shown in Section 3.3.3.

3.3.2.3 Rotated Factor 3

Factor 3, pH and DO were the variables that loaded most of the factor which are common characterizers of water chemistry. NH3-N and PO4-P were not included in any of the factors analyzed which suggests that the variation in the dataset can be predicted using other parameters in this model. I called this factor *Water Quality* because its similarity with PC2.

3.3.3 Scatterplots of Principal Components and Rotated Factors

The categorical data used to color the scatterplots are presented and explained in the following sections. In these plots the axis are based on either the PC or RF groupings. For example in Figure 3-3 the lower left plot is PC1 (x-axis) versus PC3 (y-axis) and in the upper left it is PC1 (x-axis) versus PC2 (y-axis). As expected, these PCs are independent of each other and show no visible correlation. The points are colored by the category variables, in the case of Figure 3-3 is depth. This can show groups and whether the categories are related to the various PCs and RFs. Discussions of the PCs or RFs and how they might be related to each categorical variable are found in each section.

3.3.3.1 Depth

The categorical depths used were; Surface, above thermocline, below thermocline and bottom. The scatterplot matrix of the PCs is shown in Figure 3-3. The scatterplot matrix of the RFs is shown in Figure 3-4.

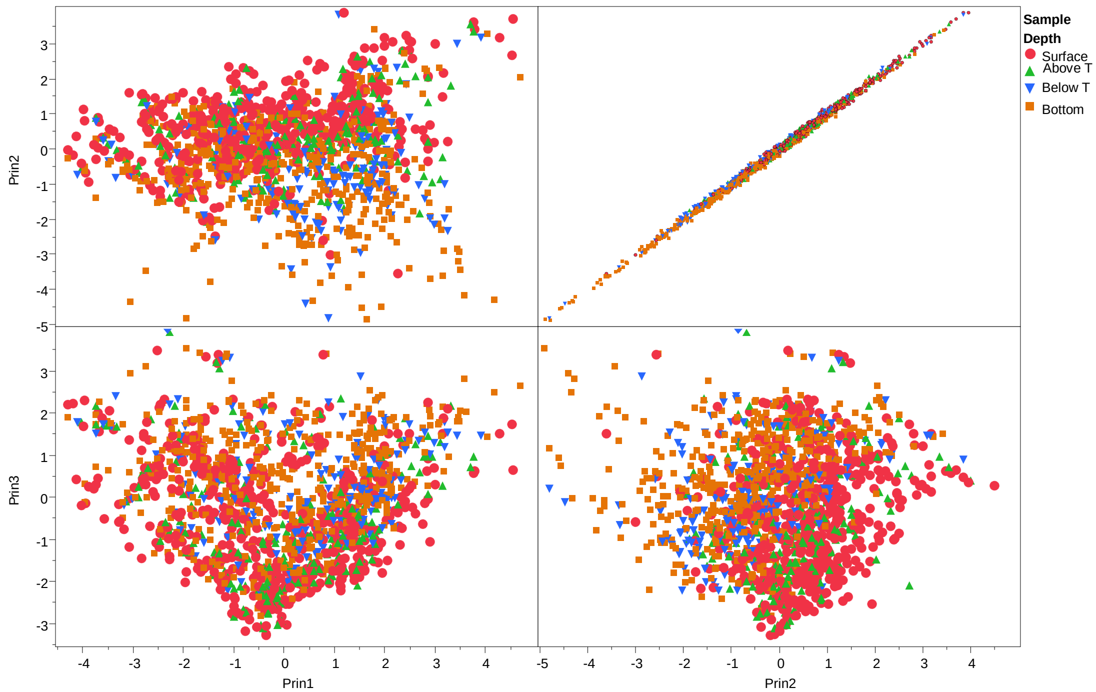

Figure 3-3. Scatterplot Matrix of PCs Colored by Depth

For the depth categories, RFs show segmentation much better than the PCs. Especially when plotting Factor 3 versus Factor 1 and 2 (See bottom pane of Figure 3-4). This may be due to the fact that DO and pH (the loading variables of Factor 3) are correlated to the depth where the sample was taken. Depth plays and important factor in explaining the changes in water quality of the reservoir. There are no great differences between Surface and Above Thermocline, the same as between Below Thermocline and Bottom. This suggests that the thermocline in general separates the quality of the water in the reservoir.

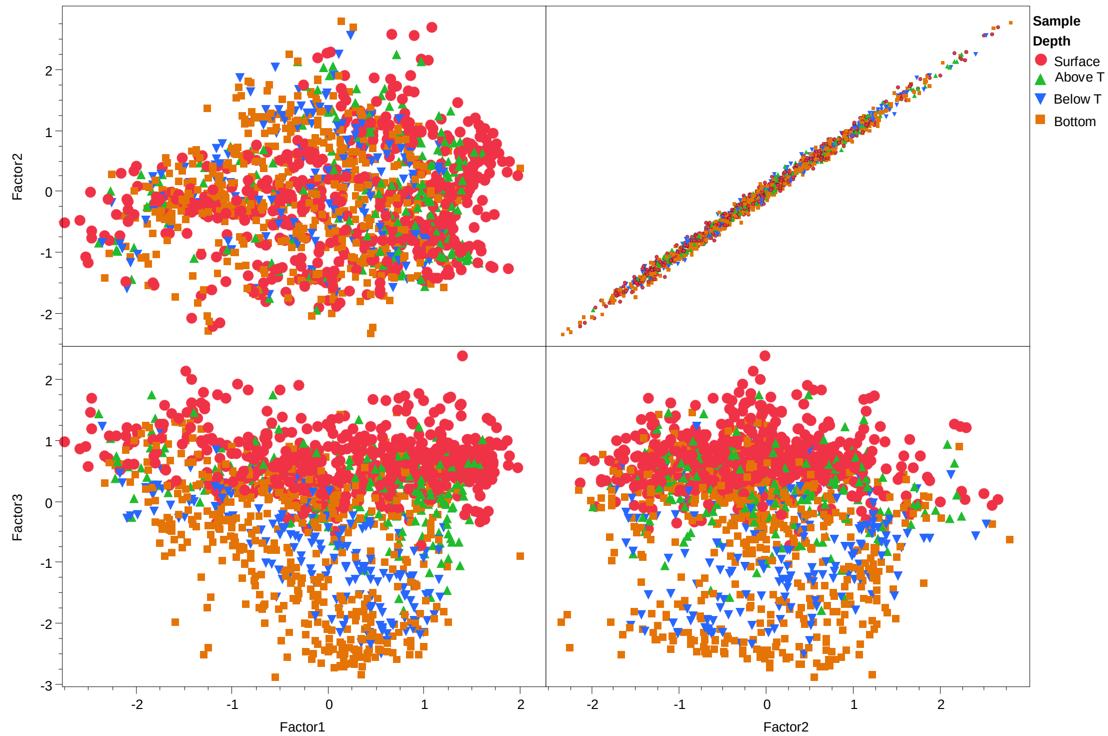

Figure 3-4. Scatterplot Matrix of Factors Colored by Depth

3.3.3.2 Seasonality (According to Water Temperature Distribution)

The season distribution categories were those obtained from the analysis shown in the season distribution section (Sections 3.1.1 and 3.1.2). The scatterplot matrix of the PCs is shown in Figure 3-5. The scatterplot matrix of the RFs is shown in Figure 3-6.

The categories used in these plots are; Cold (December, January, February, March, and April), Semi-Cold (May and November), Semi-Warm (June and October), Warm (July and October) and Transition (August). All the plots show segmentation even though the RF plots show it better than the PC plots. This clearly shows that variation in the reservoir is highly related to the time of year.

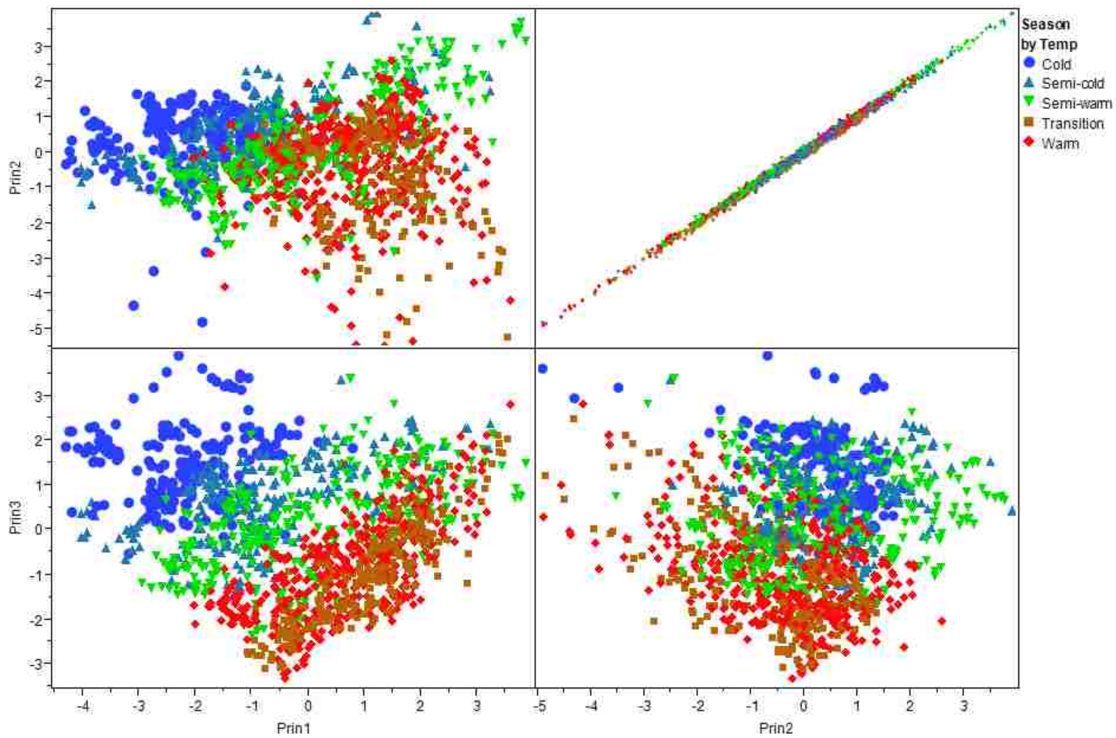

Figure 3-5. Scatterplot Matrix of PCs Colored by Proposed Season

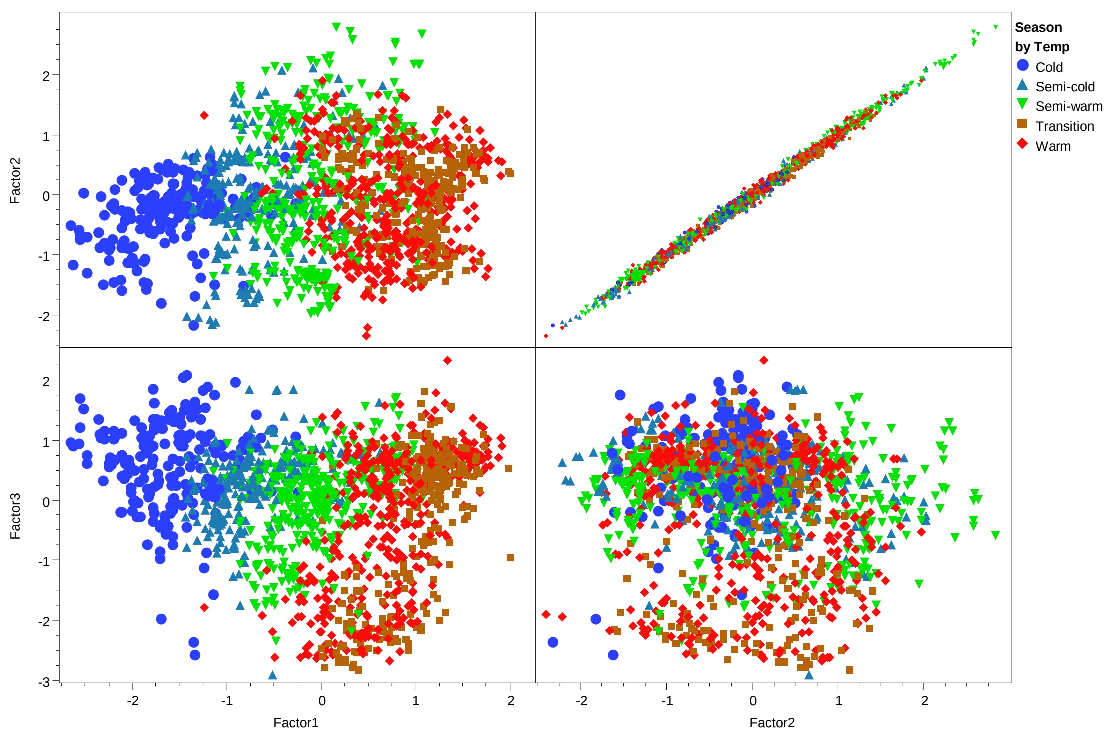

Figure 3-6. Scatterplot Matrix of Factors Colored by Proposed Season

3.3.3.3 Seasonality (Traditional Distribution)

Tradition notes that from March to May it was called Spring, from June to August is Summer, from September to November is Fall and from December to February is Winter. The scatterplot matrix of the PCs using these as categorical variables is shown in Figure 3-7. The scatterplot matrix of the RFs is shown in Figure 3-8.

Even though there is a clear segmentation between the PCs and the RFs, this segmentation is not as clear as the segmentation show based on the seasonal categories determined using temperature distributions (Figure 3-5 and Figure 3-6).

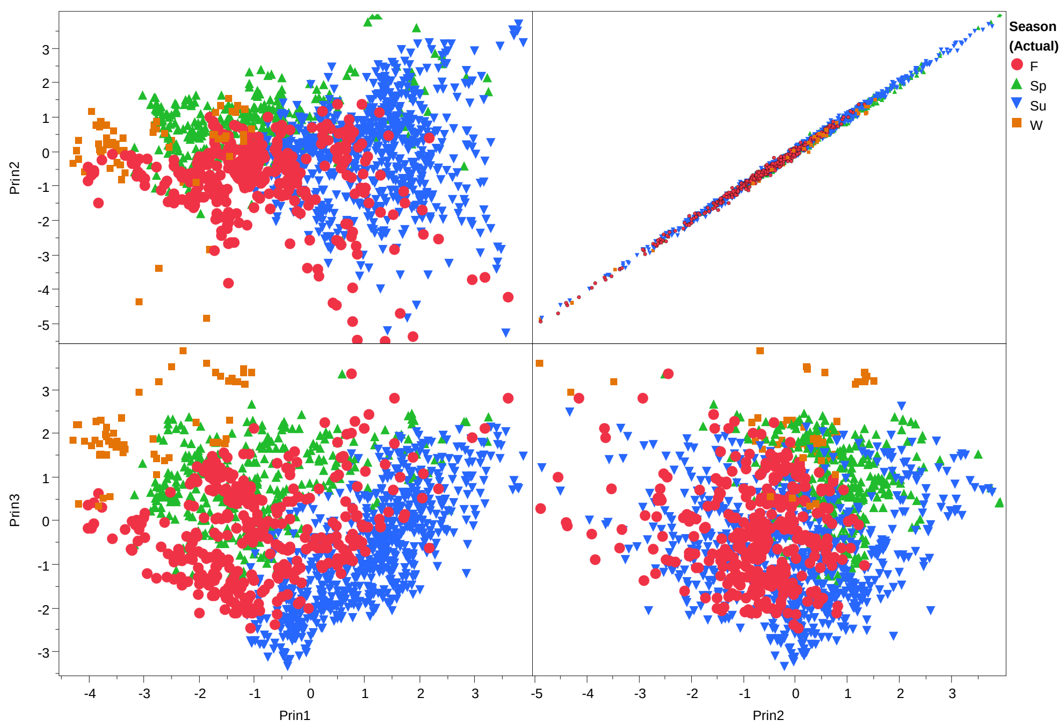

Figure 3-7. Scatterplot Matrix of PCs Colored by Actual Season

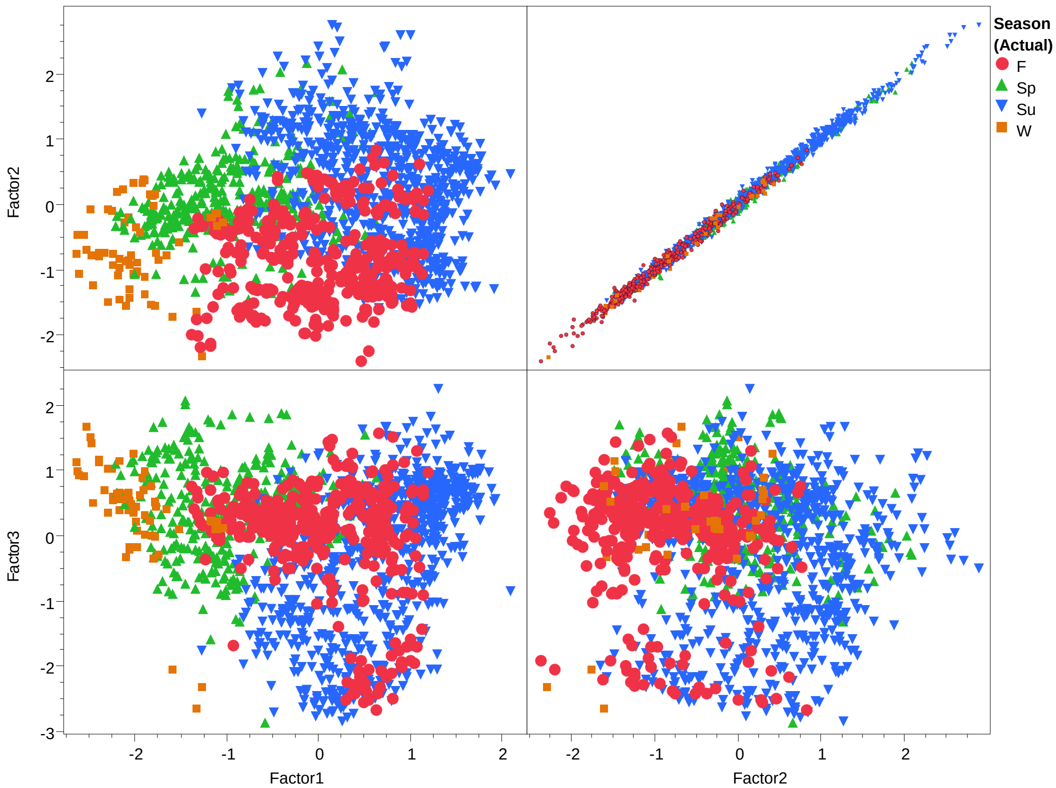

Figure 3-8. Scatterplot Matrix of Factors Colored by Actual Season

3.3.3.4 Location

The data were categorized into the four sampled sites; NearDam, MidLake, UpperEnd and Wallsburg. The scatterplot matrix of the PCs is shown in Figure 3-9. The scatterplot matrix of the RFs is shown in Figure 3-10.

Unexpectedly and unlike the segmentation shown by depth (See Figure 3-4), there are no strong signs of segmentations between these variables when colored by location where the sample was taken. This indicates that Deer Creek is well mixed and water quality is not significantly different based on location.

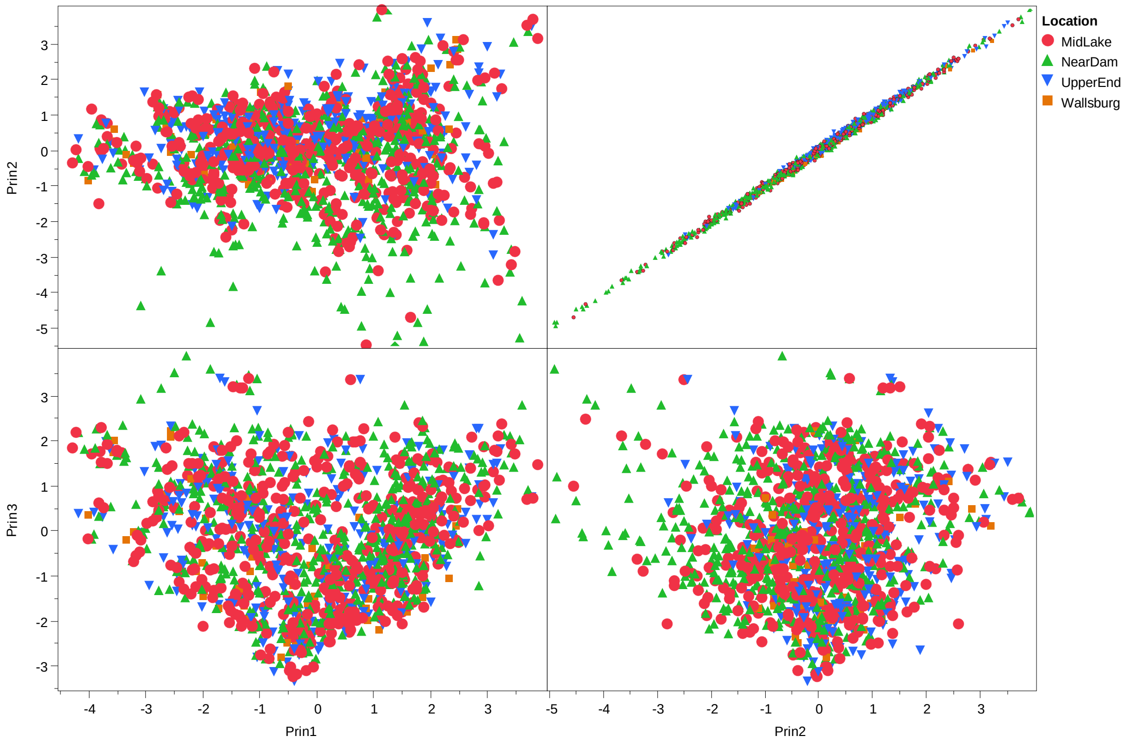

Figure 3-9. Scatterplot Matrix of PCs Colored by Location

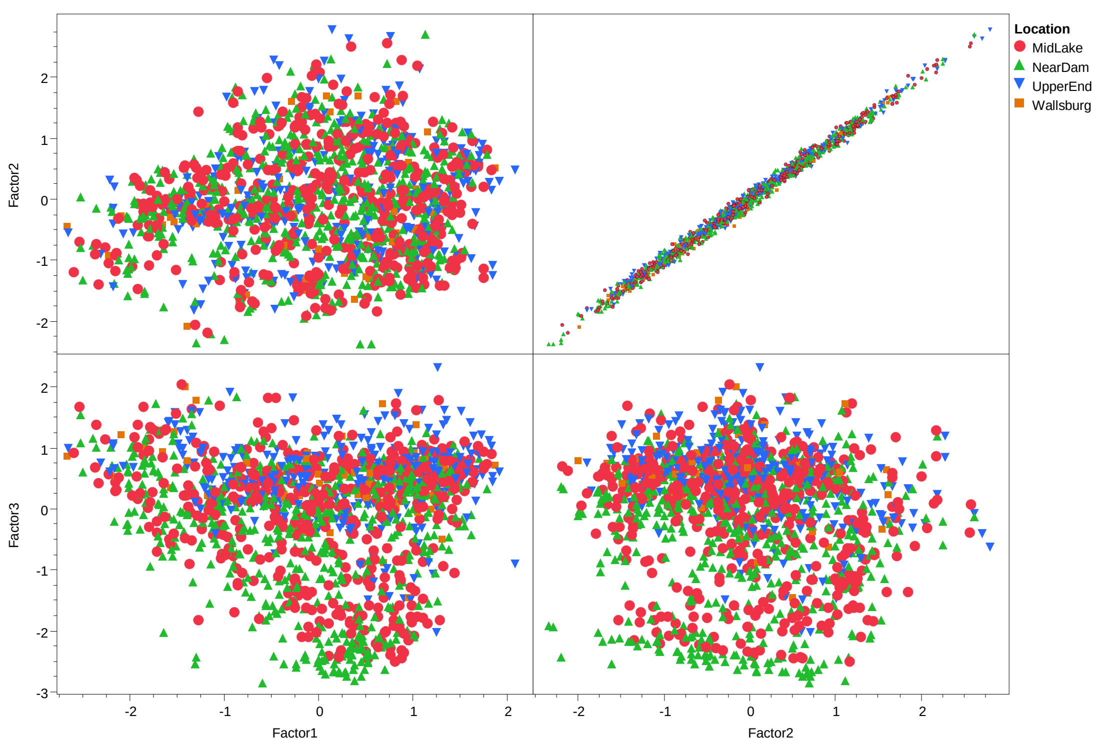

Figure 3-10. Scatterplot Matrix of Factors Colored by Location

3.3.3.5 Year (with Respect to the Construction of Jordanelle Dam)

I called from 1980 to 1986 *Before Jordanelle*, from 1987 to 1992 *Construction* and from 1993 to 2007 *After Jordanelle*. The scatterplot matrix of the PCs highlighting these categories is shown in Figure 3-11. The scatterplot matrix of the RFs is shown in Figure 3-12.

According to these scatterplot matrices, there are no significant differences in Deer Creek's water quality between these categories, at least no signs of a dramatic change because of the construction of the dam upstream Deer Creek. This analysis does not rule out a long term trend, only that statistically there is no great difference in these three populations.

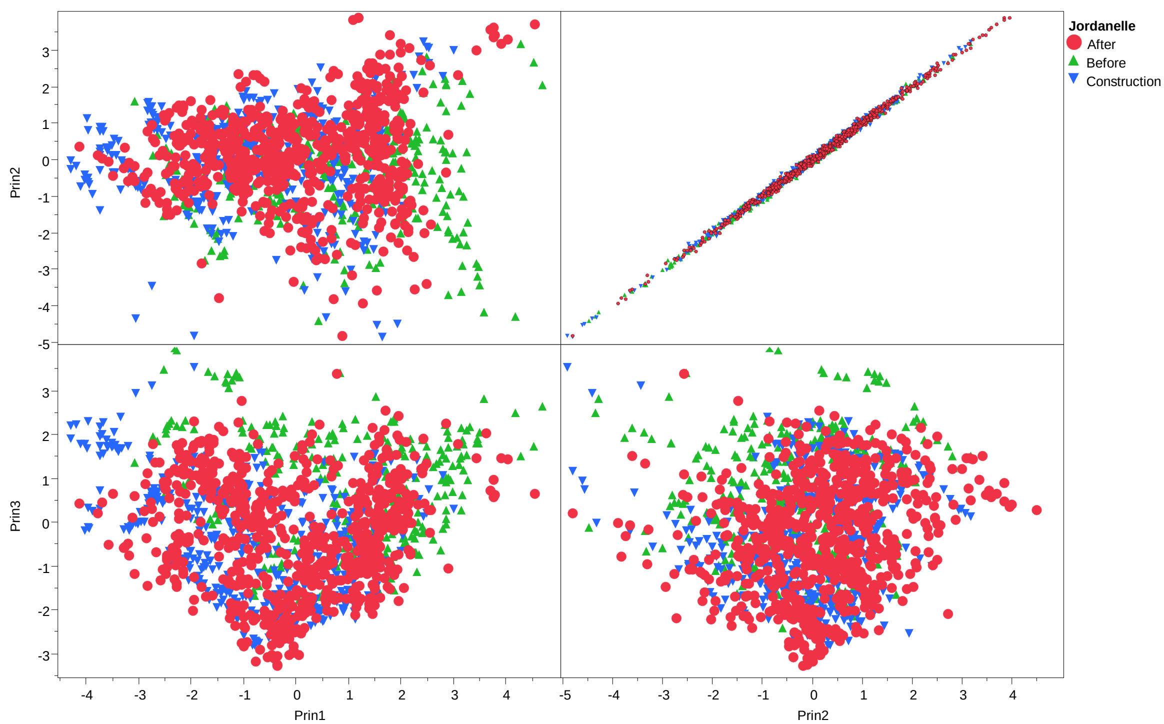

Figure 3-11. Scatterplot Matrix of PCs Colored by Construction of Jordanelle

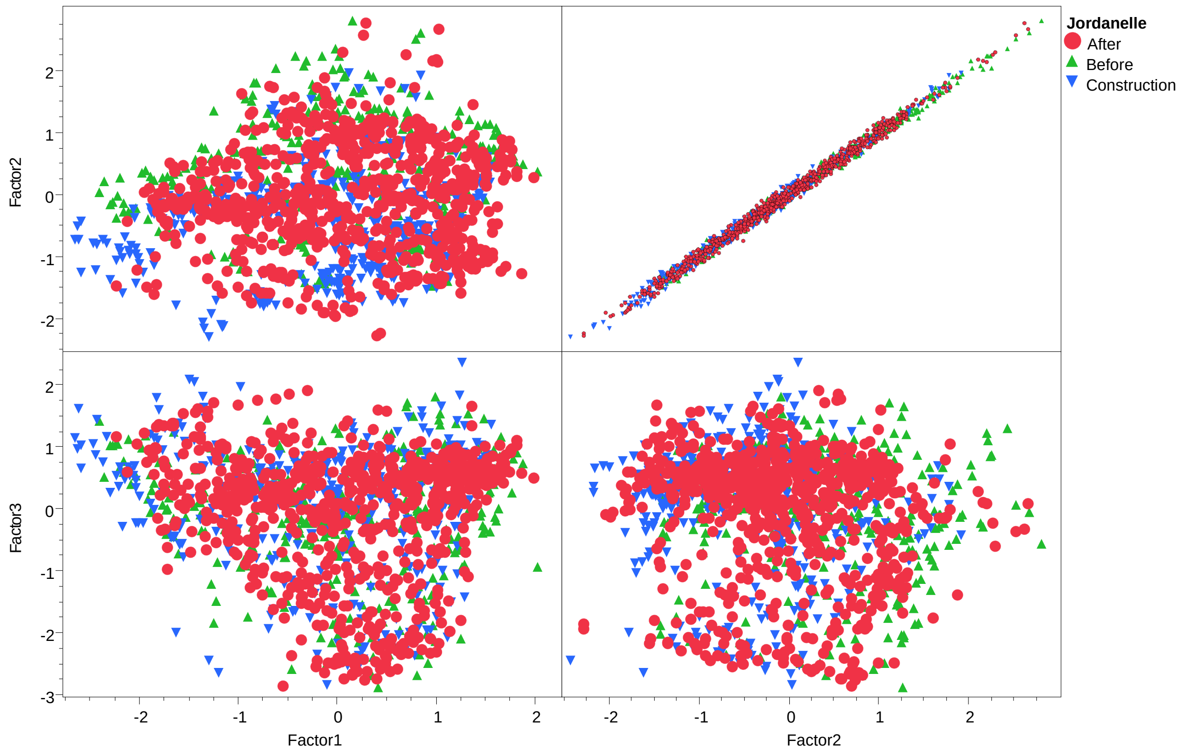

Figure 3-12. Scatterplot Matrix of Factors Colored by Construction of Jordanelle

3.3.3.6 Decade

The values were arranged by each five years; this is 80s, 85s, 90s, 95s, 00s and 05s. The scatterplot matrix of the PCs is shown in Figure 3-13. The scatterplot matrix of the RFs is shown in Figure 3-14.

Even though some of the markers tend to group together the scatterplot matrices do not show strong patterns. The continuous improvement of the water quality of Deer Creek, as seen in the box plots in Appendix D, is not apparent in these plots.

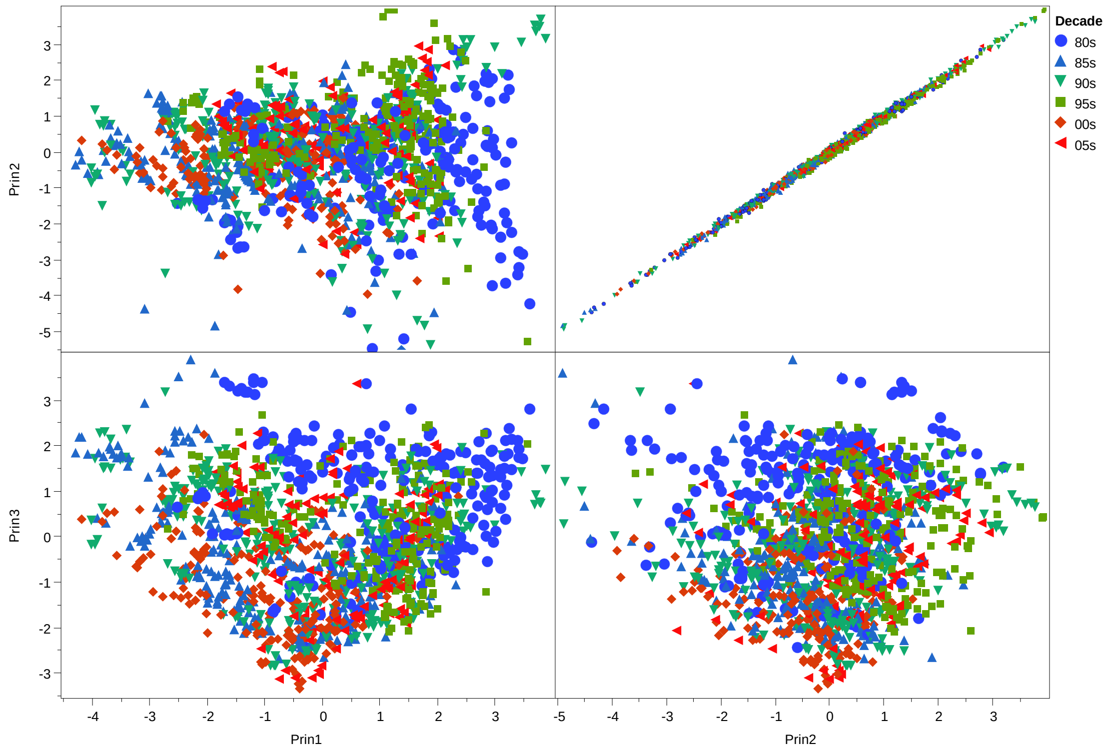

Figure 3-13. Scatterplot Matrix of PCs Colored by Decade

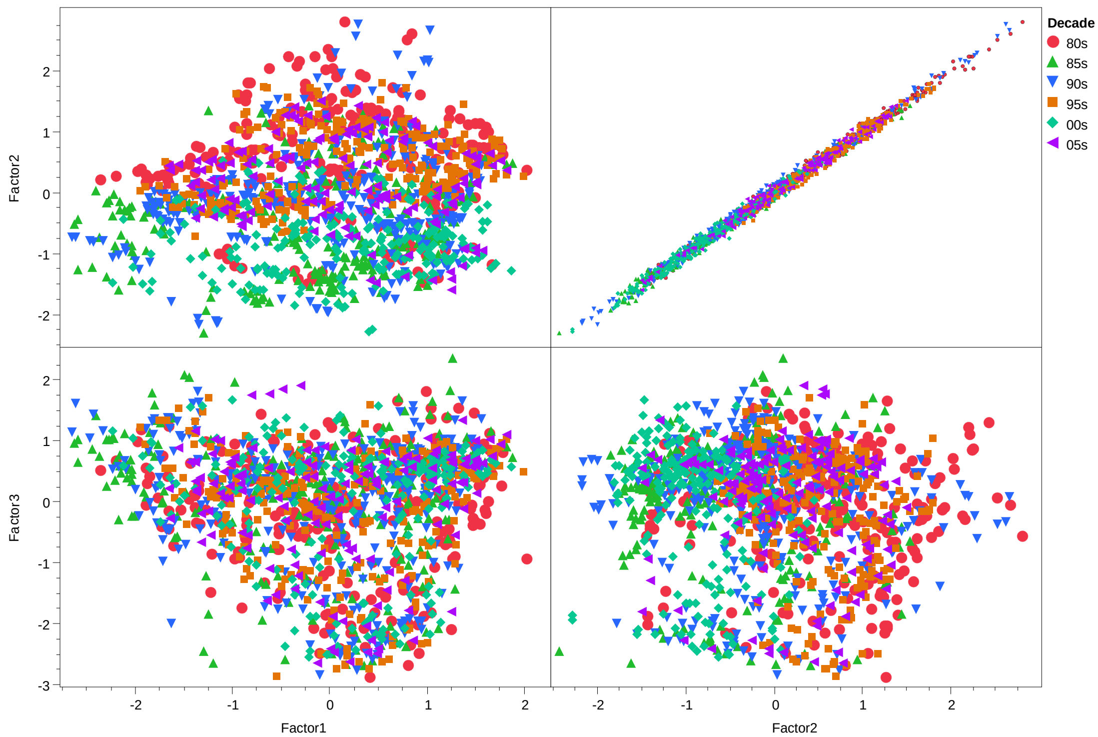

Figure 3-14. Scatterplot Matrix of Factors Colored by Decade

4 CONCLUSIONS

The tested season distribution proved to be effective for showing general variations in the reservoir. The first three PCs and RFs can explain (or predict) a cumulative 59% and 49% respectively of the variability in the reservoir, which is reasonable for a study of 28 years. The FA proved to be effective when grouping different variables by combining variables with similar characteristics. Depth and season showed visible trends when plotted as RFs. Location did not show any trends.

The FA indicated that seasonality was the best predictor of variability of the reservoir, but only explained a portion of the variation. The rest of the variation could be explained (or predicted) based on just a few water quality parameters.

The water quality of the reservoir has improved since the dataset began in the 1980s. The improvement of the reservoir was probably due to a combination of best management practices rather than a specific event. PCA and FA do not show a difference in relationship between the PC or factors due to the construction of Jordanelle reservoir. However, trends in the data, as long as they do not affect the correlation between variables, are not represented in scatter plots or the PC or FA loading factors. Differences in these plots would indicate variations in the relationships between the PCs or factors.

Nonparametric approaches should be used when dealing with water quality data, traditional parametric approaches can result in misleading conclusions about the data.

Some of the results are driven by the inputs and methods performed. This suggests that the researcher must be very careful when picking the amount of variables to be included in a study like PCA and FA.

REFERENCES

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Eckhoff, D., Boyd, A., Gillette, C., & King, R. (2002). Deer Creek Reservoir Drainage, TMDL Study. Retrieved from

Fan, X., Cui, B., Zhang, K., Zhang, Z., & Shao, H. (2012). Water Quality Management Based on Division of Dry and Wet Seasons in Pearl River Delta, China. *CLEAN–Soil, Air, Water*.

Helsel, D., & Hirsch, R. (2002). Statistical methods in water resources: US Geological Survey Techniques of Water-Resources Investigations, book 4, chap. *A3, accessed June, 28*, 2007.

Helsel, D. R. (2012). *Statistics for Censored Environmental Data Using Minitab and R* (Vol. 77): Wiley.

Higgins, J. J. (2003). Introduction to modern nonparametric statistics.

Jolliffe, I. T. (2002). *Principal component analysis* (Vol. 2): Wiley Online Library.

Miller, J. B. (2008). East Canyon Reservoir CE-QUAL-W2 Model, 2008 Water Quality Assessment Utah DEQ Phosphorus TMDL. Hooper: JM Water Quality, LLC.

Palma, P., Alvarenga, P., Palma, V. L., Fernandes, R. M., Soares, A. M. V. M., & Barbosa, I. R. (2010). Assessment of anthropogenic sources of water pollution using multivariate statistical techniques: a case study of the Alqueva's reservoir, Portugal. *Environmental monitoring and assessment*, *165*(1), 539-552.

Praus, P. (2006). Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA*, *31*(4), 417-422.

Rahman, A., Al Bakri, D., Ford, P., & Church, T. (2005). Limnological characteristics, eutrophication and cyanobacterial blooms in an inland reservoir, Australia. *Lakes & Reservoirs: Research & Management*, *10*(4), 211-220.

Ramsey, F. L., & Schafer, D. W. (2002). *The statistical sleuth: a course in methods of data analysis* (Vol. 163): Duxbury/Thomson Learning Australia; Pacific Grove, CA.

Rangel-Peraza, J. G., De Anda, J., González-Farias, F., & Erickson, D. (2009). Statistical assessment of water quality seasonality in large tropical reservoirs. *Lakes & Reservoirs: Research & Management, 14*(4), 315-323.

SAS Institute Inc. (2007). JMP User's Guide Vol. 2011.

StatSoft. (2004). *Electronic statistics textbook*: StatSoft Inc.

UDEQ. (2000). *Deer Creek Reservoir*. State of Utah Retrieved from http://www.waterquality.utah.gov/watersheds/lakes/DEERCREK.pdf

Vega, M., Pardo, R., Barrado, E., & Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water research, 32*(12), 3581-3592.

Wong, P. M., Choi, S., & Niu, Y. (2002). A comparison of PCA/ICA for data preprocessing in a geoscience application. *IEEE Transactions on Geoscience and Remote Sensing, 40*(5), 410-420.

Wunderlin, D. A., Pesce, S. F., & Hued, A. C. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquia River Basin (Córdoba-Argentina). *Water Research, 35*(12), 2881-2894.

Yang, Y. H., Zhou, F., Guo, H. C., Sheng, H., Liu, H., Dao, X., & He, C. J. (2010). Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environmental monitoring and assessment*, 1-10.

APPENDIX A. COMPARISONS OF WATER TEMPERATURE BY MONTH

Table A-1. Comparison for All Pairs Using Tukey-Kramer HSD

Month	- Month	Difference	p-Value	Month	- Month	Difference	p-Value
Aug	Feb	15.62425	<.0001	Nov	Jan	6.0641	<.0001
Aug	Jan	15.39464	<.0001	Oct	Apr	5.95809	<.0001
Sep	Feb	14.44844	<.0001	Aug	Oct	5.42191	<.0001
Jul	Feb	14.40413	<.0001	May	Mar	5.3914	<.0001
Sep	Jan	14.21884	<.0001	May	Dec	5.25689	<.0001
Jul	Jan	14.17452	<.0001	Jun	Nov	4.67954	<.0001
Aug	Mar	13.52951	<.0001	Aug	Jun	4.65101	<.0001
Aug	Dec	13.395	<.0001	Sep	Oct	4.24611	<.0001
Sep	Mar	12.35371	<.0001	Apr	Feb	4.24424	<.0001
Jul	Mar	12.30939	<.0001	Jul	Oct	4.20179	<.0001
Sep	Dec	12.2192	<.0001	Nov	Mar	4.19897	<.0001
Jul	Dec	12.17488	<.0001	Nov	Dec	4.06446	<.0001
Aug	Apr	11.38	<.0001	Apr	Jan	4.01464	<.0001
Jun	Feb	10.97324	<.0001	Oct	Nov	3.90863	<.0001
Jun	Jan	10.74364	<.0001	Jun	May	3.48711	<.0001
Sep	Apr	10.2042	<.0001	Sep	Jun	3.4752	<.0001
Oct	Feb	10.20233	<.0001	Jul	Jun	3.43089	<.0001
Jul	Apr	10.15988	<.0001	May	Apr	3.24189	<.0001
Oct	Jan	9.97273	<.0001	Oct	May	2.7162	<.0001
Aug	Nov	9.33054	<.0001	Dec	Feb	2.22925	0.1621
Jun	Mar	8.8785	<.0001	Apr	Mar	2.14951	0.0087
Jun	Dec	8.74399	<.0001	Mar	Feb	2.09474	0.1407
Sep	Nov	8.15474	<.0001	Nov	Apr	2.04946	0.0002
Aug	May	8.13811	<.0001	Apr	Dec	2.015	0.0772
Jul	Nov	8.11042	<.0001	Dec	Jan	1.99964	0.6739
Oct	Mar	8.1076	<.0001	Mar	Jan	1.86513	0.7008
Oct	Dec	7.97309	<.0001	Aug	Jul	1.22012	0.0002
May	Feb	7.48614	<.0001	May	Nov	1.19243	0.141
May	Jan	7.25653	<.0001	Aug	Sep	1.1758	0.0016
Sep	May	6.96231	<.0001	Jun	Oct	0.77091	0.3213
Jul	May	6.91799	<.0001	Jan	Feb	0.22961	1
Jun	Apr	6.729	<.0001	Dec	Mar	0.13451	.
Nov	Feb	6.29371	<.0001	Sep	Jul	0.04432	.

Table A-2. Comparison for Each Pair Using Student's t

Month	- Month	Difference	*p*-Value	Month	- Month	Difference	*p*-Value
Aug	Feb	15.62425	<.0001	Nov	Jan	6.0641	<.0001
Aug	Jan	15.39464	<.0001	Oct	Apr	5.95809	<.0001
Sep	Feb	14.44844	<.0001	Aug	Oct	5.42191	<.0001
Jul	Feb	14.40413	<.0001	May	Mar	5.3914	<.0001
Sep	Jan	14.21884	<.0001	May	Dec	5.25689	<.0001
Jul	Jan	14.17452	<.0001	Jun	Nov	4.67954	<.0001
Aug	Mar	13.52951	<.0001	Aug	Jun	4.65101	<.0001
Aug	Dec	13.395	<.0001	Sep	Oct	4.24611	<.0001
Sep	Mar	12.35371	<.0001	Apr	Feb	4.24424	<.0001
Jul	Mar	12.30939	<.0001	Jul	Oct	4.20179	<.0001
Sep	Dec	12.2192	<.0001	Nov	Mar	4.19897	<.0001
Jul	Dec	12.17488	<.0001	Nov	Dec	4.06446	<.0001
Aug	Apr	11.38	<.0001	Apr	Jan	4.01464	<.0001
Jun	Feb	10.97324	<.0001	Oct	Nov	3.90863	<.0001
Jun	Jan	10.74364	<.0001	Jun	May	3.48711	<.0001
Sep	Apr	10.2042	<.0001	Sep	Jun	3.4752	<.0001
Oct	Feb	10.20233	<.0001	Jul	Jun	3.43089	<.0001
Jul	Apr	10.15988	<.0001	May	Apr	3.24189	<.0001
Oct	Jan	9.97273	<.0001	Oct	May	2.7162	<.0001
Aug	Nov	9.33054	<.0001	Dec	Feb	2.22925	0.0045
Jun	Mar	8.8785	<.0001	Apr	Mar	2.14951	0.0002
Jun	Dec	8.74399	<.0001	Mar	Feb	2.09474	0.0037
Sep	Nov	8.15474	<.0001	Nov	Apr	2.04946	<.0001
Aug	May	8.13811	<.0001	Apr	Dec	2.015	0.0018
Jul	Nov	8.11042	<.0001	Dec	Jan	1.99964	0.0427
Oct	Mar	8.1076	<.0001	Mar	Jan	1.86513	0.0468
Oct	Dec	7.97309	<.0001	Aug	Jul	1.22012	<.0001
May	Feb	7.48614	<.0001	May	Nov	1.19243	0.0038
May	Jan	7.25653	<.0001	Aug	Sep	1.1758	<.0001
Sep	May	6.96231	<.0001	Jun	Oct	0.77091	0.0115
Jul	May	6.91799	<.0001	Jan	Feb	0.22961	0.8066
Jun	Apr	6.729	<.0001	Dec	Mar	0.13451	0.8637
Nov	Feb	6.29371	<.0001	Sep	Jul	0.04432	0.874

APPENDIX B. DISTRIBUTIONS, QUANTILES AND MOMENTS

B.1 Release

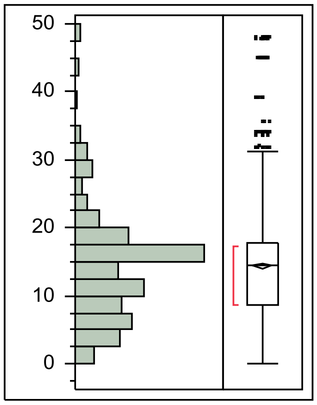

Figure B-1. Distribution of Release

Table B-3. Quartiles of Release

Percentage	Location	Value
100.0%	maximum	48.04
99.5%		48.04
97.5%		34.13
90.0%		24.07
75.0%	quartile	17.9
50.0%	median	14.46
25.0%	quartile	8.77
10.0%		4.426
2.5%		2.173
0.5%		0
0.0%	minimum	0

Table B-4. Moments of Release

Moments	Value
Mean	14.41
Std Dev	8.20
Std Err Mean	0.18
Upper 95% Mean	14.77
Lower 95% Mean	14.05
N	1985

B.2 Volume

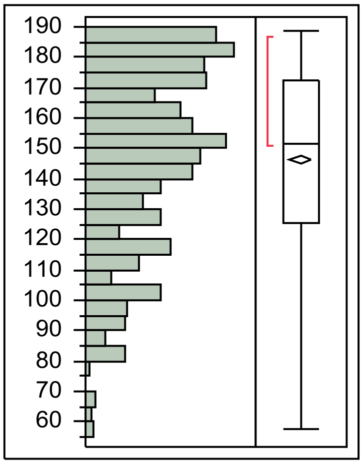

Figure B-2. Distribution of Volume

Table B-5. Quartiles of Volume

Percentage	Location	Value
100.00%	maximum	189
99.50%		189
97.50%		186.7
90.00%		183.9
75.00%	quartile	172.5
50.00%	median	151.2
25.00%	quartile	125.7
10.00%		100.9
2.50%		81.02
0.50%		63.678
0.00%	minimum	57.7

Table B-6. Moments of Volume

Moments	Value
Mean	146.25
Std Dev	30.78
Std Err Mean	0.69
Upper 95% Mean	147.60
Lower 95% Mean	144.89
N	1985.00

B.3 Inflow

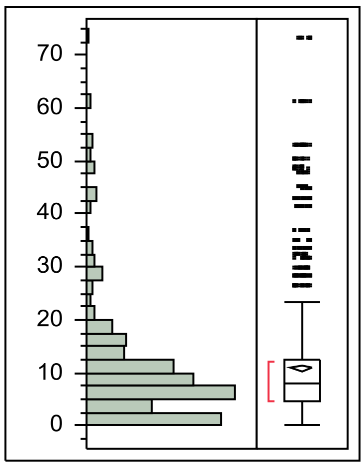

Figure B-3. Distribution of Inflow

Table B-7. Quartiles of Inflow

Percentage	Location	Value
100.00%	maximum	72.95
99.50%		61.12
97.50%		48.54
90.00%		26.24
75.00%	quartile	12.61
50.00%	median	7.81
25.00%	quartile	4.52
10.00%		0
2.50%		0
0.50%		0
0.00%	minimum	0

Table B-8. Moments of Inflow

Moments	Value
Mean	10.89
Std Dev	11.76
Std Err Mean	0.26
Upper 95% Mean	11.41
Lower 95% Mean	10.37
N	1985.00

B.4 PH

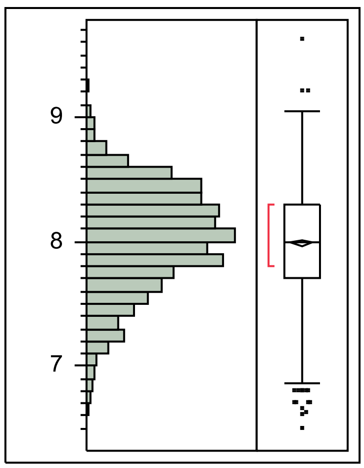

Figure B-4. Distribution of pH

Table B-9. Quartiles of pH

Percentage	Location	Value
100.00%	maximum	9.615
99.50%		8.9
97.50%		8.7
90.00%		8.5
75.00%	quartile	8.3
50.00%	median	8
25.00%	quartile	7.713
10.00%		7.4
2.50%		7.1
0.50%		6.739
0.00%	minimum	6.5

Table B-10. Moments of pH

Moments	Value
Mean	7.99
Std Dev	0.42
Std Err Mean	0.01
Upper 95% Mean	8.01
Lower 95% Mean	7.97
N	1876.00

B.5 Specific Conductivity

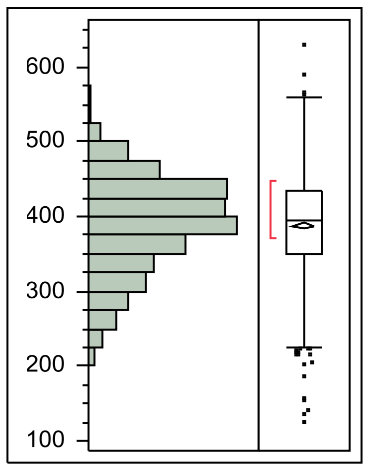

Figure B-5. Distribution of Specific Conductivity

Table B-11. Quartiles of Specific Conductivity

Percentage	Location	Value
100.00%	maximum	628
99.50%		530.7
97.50%		492
90.00%		463
75.00%	quartile	434
50.00%	median	396
25.00%	quartile	350
10.00%		297.5
2.50%		249.2
0.50%		213.2
0.00%	minimum	123

Table B-12. Moments of Specific Conductivity

Moments	Value
Mean	388.08
Std Dev	63.93
Std Err Mean	1.47
Upper 95% Mean	390.96
Lower 95% Mean	385.19
N	1886.00

B.6 Dissolved Oxygen

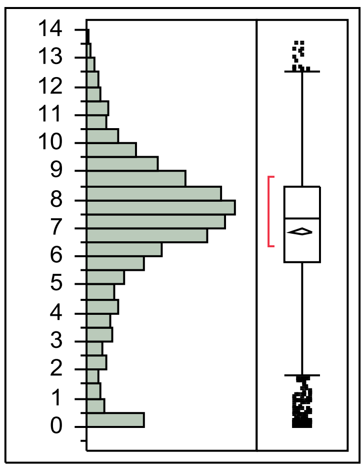

Figure B-6. Distribution of Dissolved Oxygen

Table B-13. Quartiles of Dissolved Oxygen

Percentage	Location	Value
100.00%	maximum	13.5
99.50%		12.7
97.50%		11.6
90.00%		9.8
75.00%	quartile	8.5
50.00%	median	7.34
25.00%	quartile	5.8
10.00%		2.652
2.50%		0.1
0.50%		0
0.00%	minimum	0

Table B-14. Moments of Dissolved Oxygen

Moments	Value
Mean	6.88
Std Dev	2.73
Std Err Mean	0.06
Upper 95% Mean	7.01
Lower 95% Mean	6.76
N	1823.00

B.7 Water Temperature

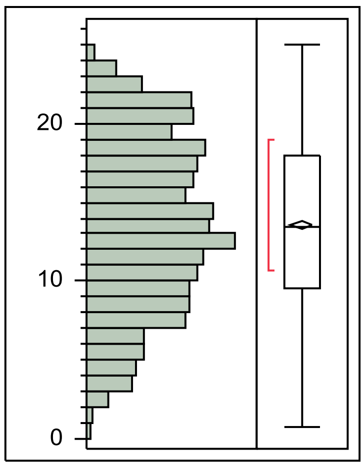

Figure B-7. Distribution of Water Temperature

Table B-15. Quartiles of Water Temperature

Percentage	Location	Value
100.00%	maximum	24.98
99.50%		23.88
97.50%		22.73
90.00%		20.97
75.00%	quartile	18
50.00%	median	13.5
25.00%	quartile	9.52
10.00%		6.291
2.50%		3.4
0.50%		1.941
0.00%	minimum	0.7

Table B-16. Moments of Water Temperature

Moments	Value
Mean	13.61
Std Dev	5.39
Std Err Mean	0.12
Upper 95% Mean	13.85
Lower 95% Mean	13.36
N	1880.00

B.8 Chlorophyll-a

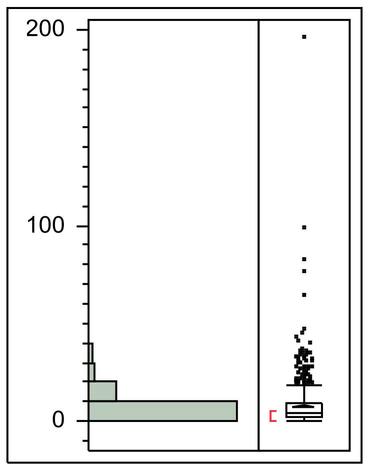

Figure B-8. Distribution of Chlorophyll-a

Table B-17. Quartiles of Chlorophyll-a

Percentage	Location	Value
100.00%	maximum	196
99.50%		75.97
97.50%		32.72
90.00%		16.14
75.00%	quartile	8.85
50.00%	median	4.7
25.00%	quartile	2.4
10.00%		0.9
2.50%		0.2
0.50%		0
0.00%	minimum	0

Table B-18. Moments of Chlorophyll-a

Moments	Value
Mean	7.53
Std Dev	11.14
Std Err Mean	0.39
Upper 95% Mean	8.31
Lower 95% Mean	6.76
N	797.00

B.9 Phosphate-Phosphorus

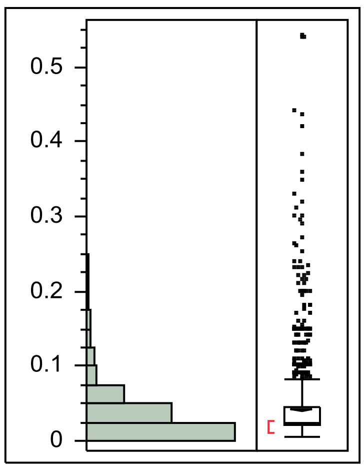

Figure B-9. Distribution of Phosphate-Phosphorus

Table B-19. Quartiles of Phosphate-Phosphorus

Percentage	Location	Value
100.00%	maximum	0.543
99.50%		0.348
97.50%		0.18
90.00%		0.08
75.00%	quartile	0.045
50.00%	median	0.025
25.00%	quartile	0.02
10.00%		0.012
2.50%		0.01
0.50%		0.005
0.00%	minimum	0.005

Table B-20. Moments of Phosphate-Phosphorus

Moments	Value
Mean	0.04
Std Dev	0.05
Std Err Mean	0.00
Upper 95% Mean	0.04
Lower 95% Mean	0.04
N	1814.00

B.10 Nitrite-Nitrate-Nitrogen

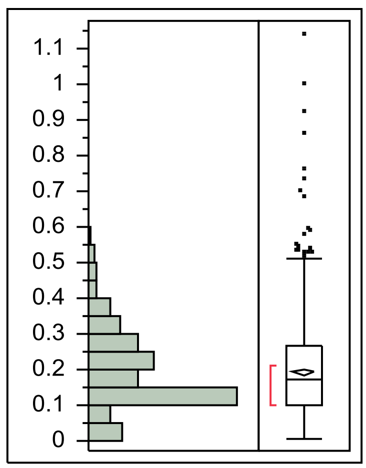

Figure B-10. Distribution of Nitrite-Nitrate-Nitrogen

Table B-21. Quartiles of Nitrite-Nitrate-Nitrogen

Percentage	Location	Value
100.00%	maximum	1.14
99.50%		0.727
97.50%		0.495
90.00%		0.354
75.00%	quartile	0.265
50.00%	median	0.17
25.00%	quartile	0.1
10.00%		0.07
2.50%		0.02
0.50%		0.01
0.00%	minimum	0.005

Table B-22. Moments of Nitrite-Nitrate-Nitrogen

Moments	Value
Mean	0.19
Std Dev	0.13
Std Err Mean	0.00
Upper 95% Mean	0.20
Lower 95% Mean	0.19
N	1222.00

B.11 Ammonia-Nitrogen

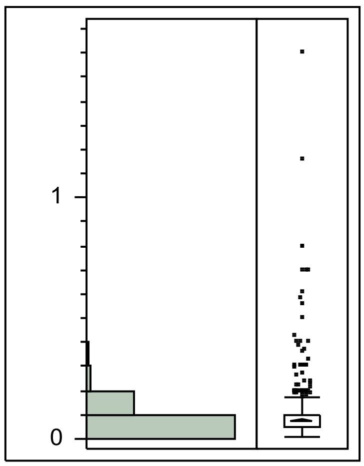

Figure B-11. Distribution of Ammonia-Nitrogen

Table B-23. Quartiles of Ammonia-Nitrogen

Percentage	Location	Value
100.00%	maximum	1.6
99.50%		0.574
97.50%		0.2
90.00%		0.1
75.00%	quartile	0.1
50.00%	median	0.05
25.00%	quartile	0.05
10.00%		0.05
2.50%		0.05
0.50%		0.005
0.00%	minimum	0.005

Table B-24. Moments of Ammonia-Nitrogen

Moments	Value
Mean	0.08
Std Dev	0.07
Std Err Mean	0.00
Upper 95% Mean	0.08
Lower 95% Mean	0.07
N	1656.00

B.12 Secchi Depth

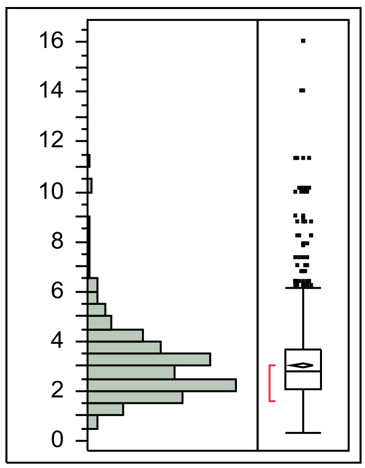

Figure B-12. Distribution of Secchi Depth

Table B-25. Quartiles of Secchi Depth

Percentage	Location	Value
100.00%	maximum	16
99.50%		10.69
97.50%		6.8
90.00%		4.6
75.00%	quartile	3.7
50.00%	median	2.75
25.00%	quartile	2.1
10.00%		1.5
2.50%		1.2
0.50%		0.751
0.00%	minimum	0.3

Table B-26. Moments of Secchi Depth

Moments	Value
Mean	3.04
Std Dev	1.54
Std Err Mean	0.04
Upper 95% Mean	3.12
Lower 95% Mean	2.97
N	1700.00

B.13 Air Temperature

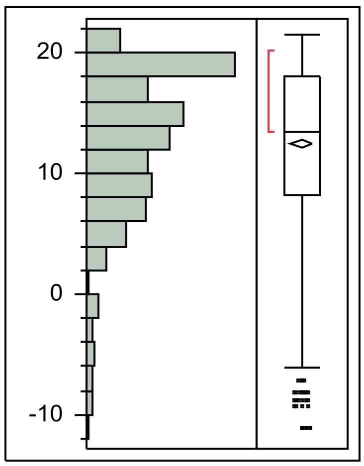

Figure B-13. Distribution of Air Temperature

Table B-27. Quartiles of Air Temperature

Percentage	Location	Value
100.00%	maximum	21.56
99.50%		21.28
97.50%		20.22
90.00%		19.22
75.00%	quartile	18.11
50.00%	median	13.5
25.00%	quartile	8.222
10.00%		4.5
2.50%		-4.722
0.50%		-8.833
0.00%	minimum	-11.11

Table B-28. Moments of Air Temperature

Moments	Value
Mean	12.45
Std Dev	6.41
Std Err Mean	0.14
Upper 95% Mean	12.73
Lower 95% Mean	12.16
N	1985.00

B.14 Precipitation

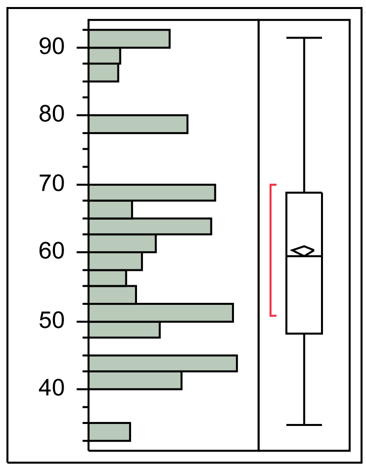

Figure B-14. Distribution of Precipitation

Table B-29. Quartiles of Precipitation

Percentage	Location	Value
100.00%	maximum	91.31
99.50%		91.31
97.50%		91.31
90.00%		85.27
75.00%	quartile	68.86
50.00%	median	59.59
25.00%	quartile	48.26
10.00%		41.28
2.50%		34.7
0.50%		34.7
0.00%	minimum	34.7

Table B-30. Moments of Precipitation

Moments	Value
Mean	60.25
Std Dev	15.42
Std Err Mean	0.35
Upper 95% Mean	60.93
Lower 95% Mean	59.57
N	1985.00

B.15 Depth

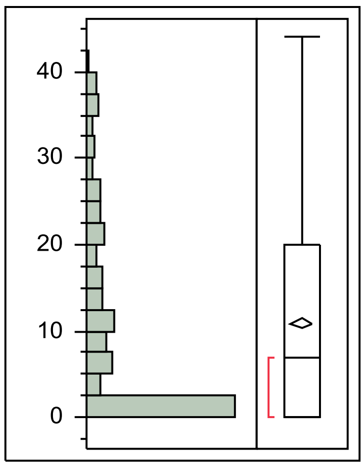

Figure B-15. Distribution of Depth

Table B-31. Quartiles of Depth

Percentage	Location	Value
100.00%	maximum	44
99.50%		40
97.50%		39
90.00%		30
75.00%	quartile	20
50.00%	median	7
25.00%	quartile	0.1
10.00%		0
2.50%		0
0.50%		0
0.00%	minimum	0

Table B-32. Moments of Depth

Moments	Value
Mean	10.89
Std Dev	11.97
Std Err Mean	0.27
Upper 95% Mean	11.42
Lower 95% Mean	10.36
N	1949.00

APPENDIX C. QUARTILES OF VARIABLES BY MONTH

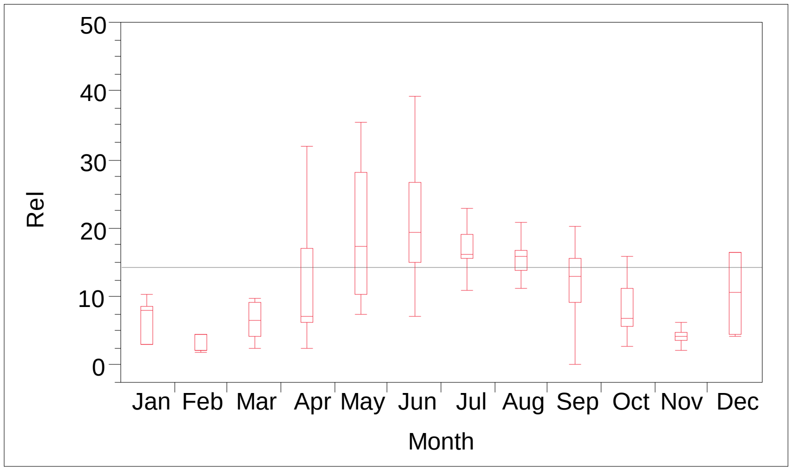

Figure C-1. Quartiles of Release by Month

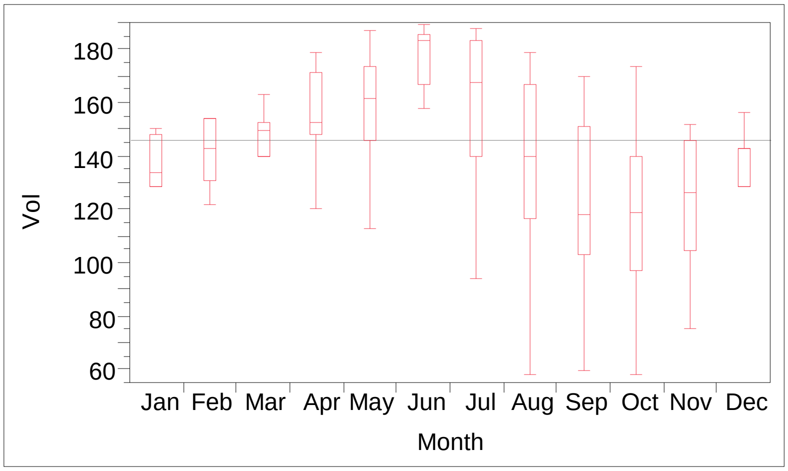

Figure C-2. Quartiles of Volume by Month

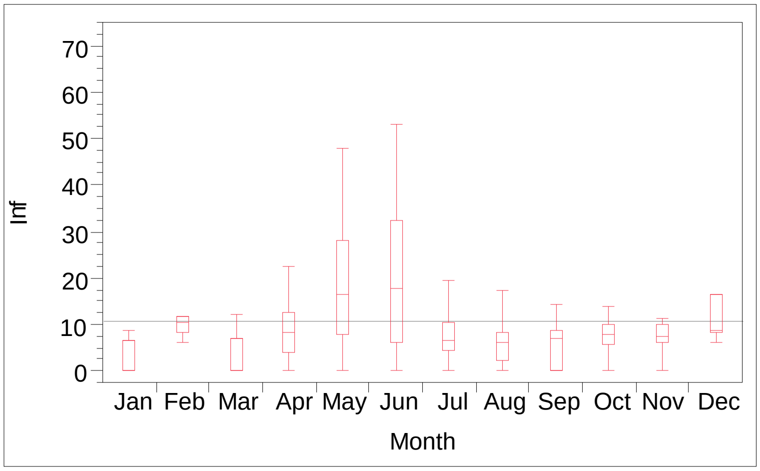

Figure C-3. Quartiles of Inflow by Month

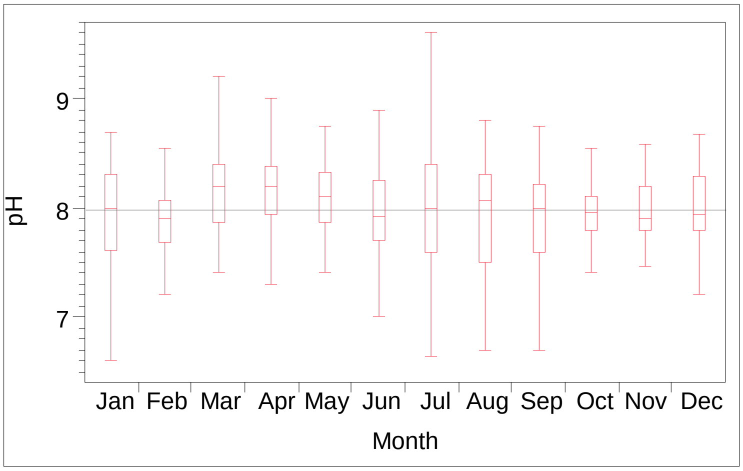

Figure C-4. Quartiles of PH by Month

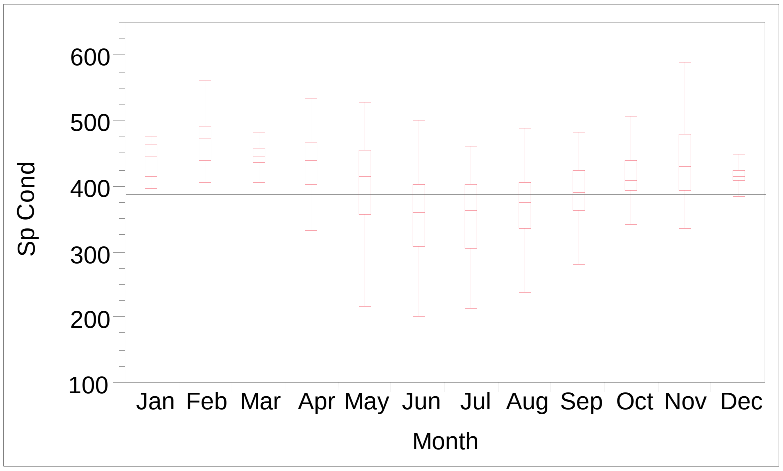

Figure C-5. Quartiles of Specific Conductance by Month

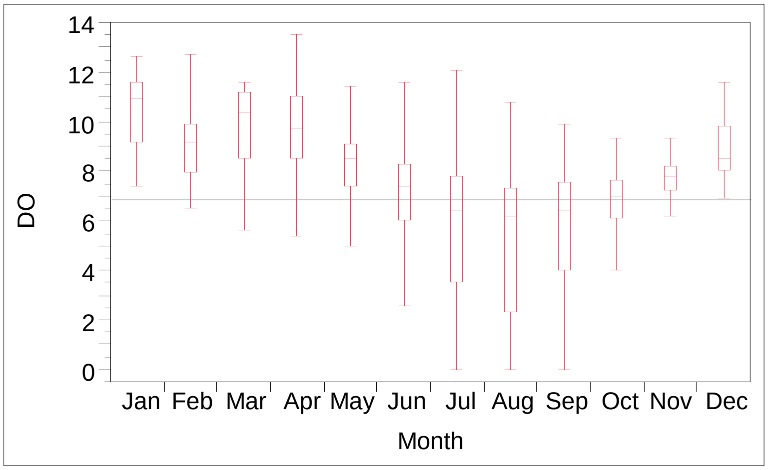

Figure C-6. Quartiles of Dissolved Oxygen by Month

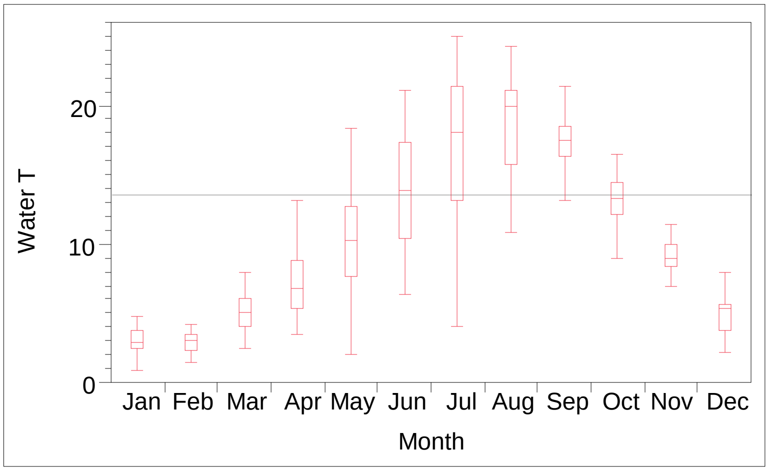

Figure C-7. Quartiles of Water Temperature by Month

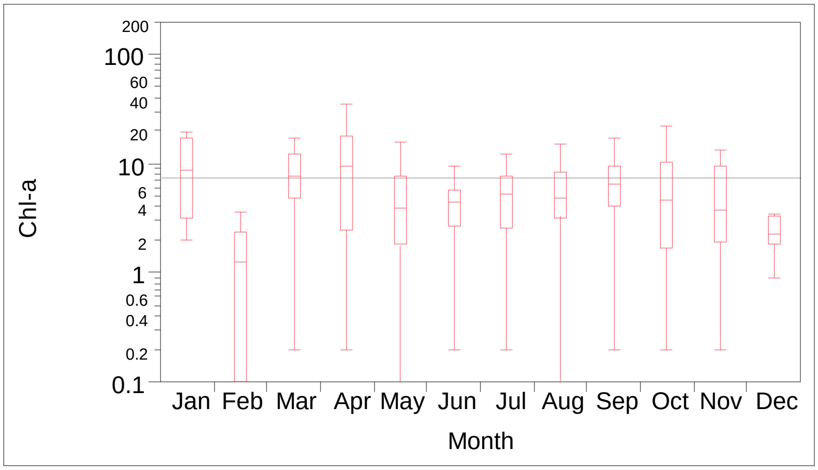

Figure C-8. Quartiles of Chlorophyll-a by Month

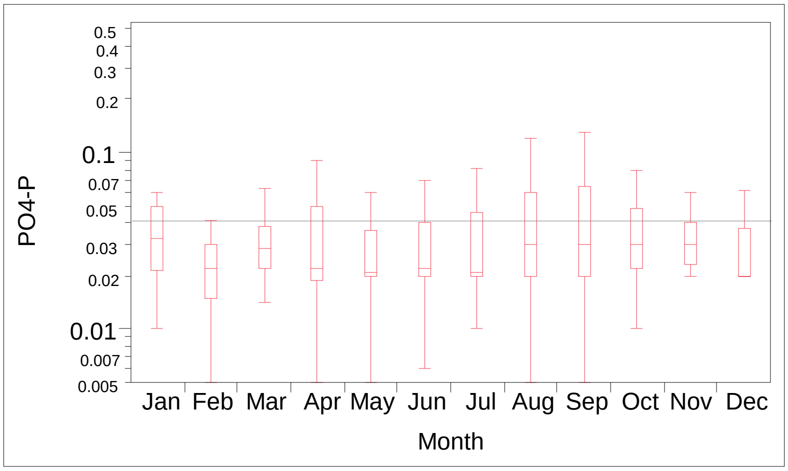

Figure C-9. Quartiles of Phosphate-Phosphorus by Month

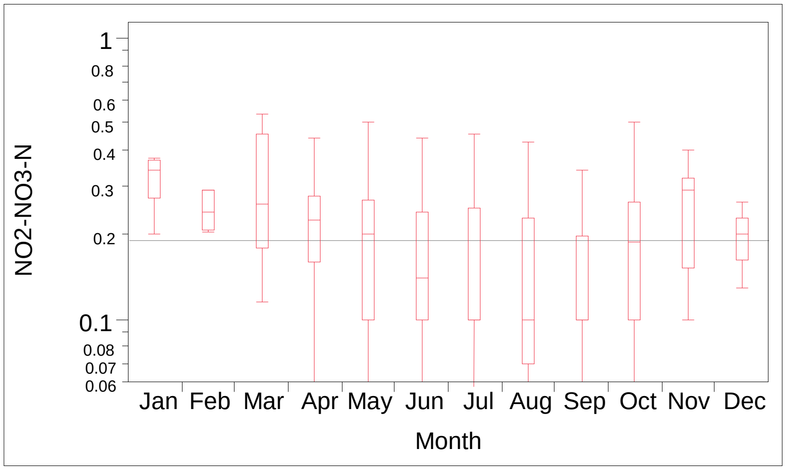

Figure C-10. Quartiles of Nitrite-Nitrate-Nitrogen by Month

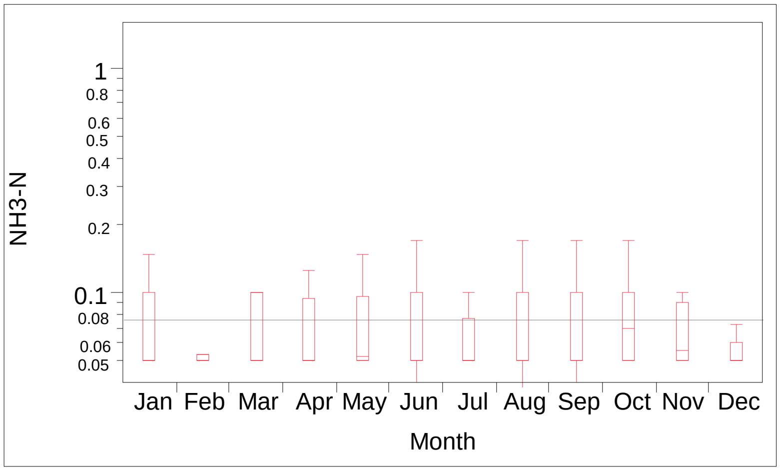

Figure C-11. Quartiles of Ammonia-Nitrogen by Month

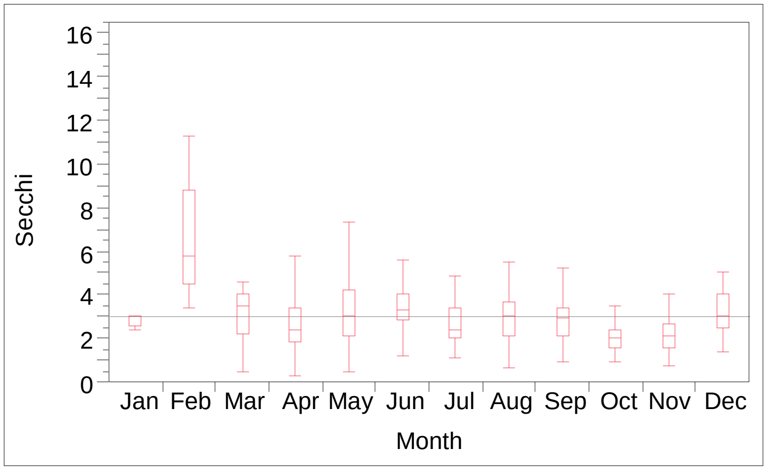

Figure C-12. Quartiles of Secchi-Depth by Month

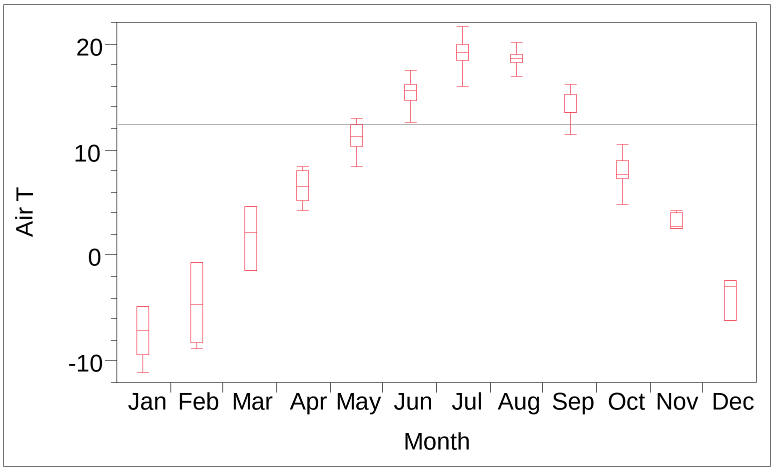

Figure C-13. Quartiles of Air Temperature by Month

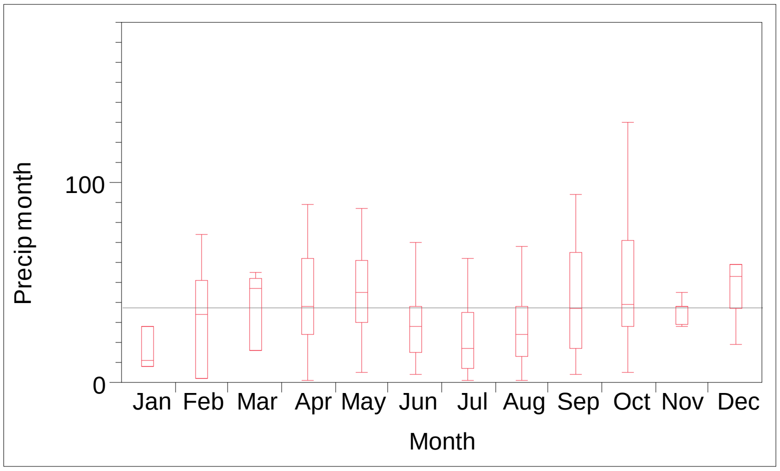

Figure C-14. Quartiles of Precipitation by Month

APPENDIX D. QUARTILES OF VARIABLES BY YEAR

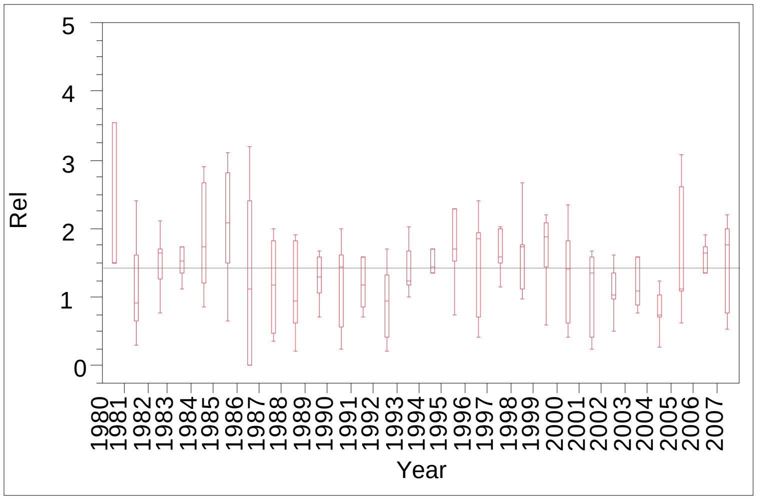

Figure D-15. Quartiles of Release by Year

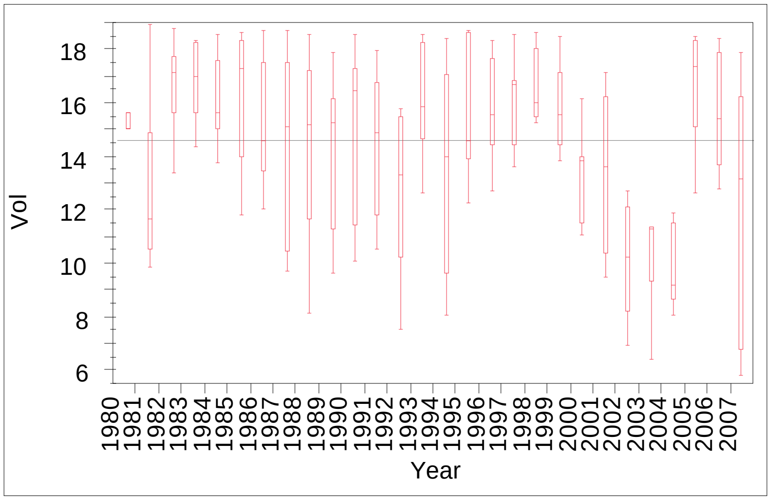

Figure D-16. Quartiles of Volume by Year

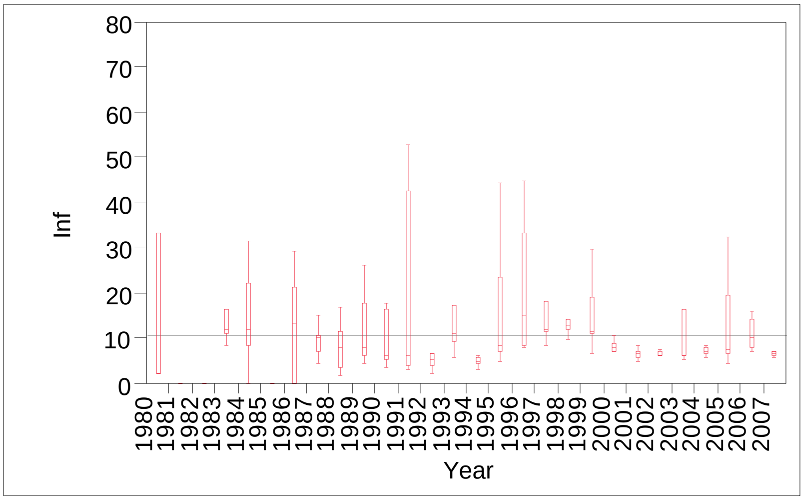

Figure D-17. Quartiles of Inflow by Year

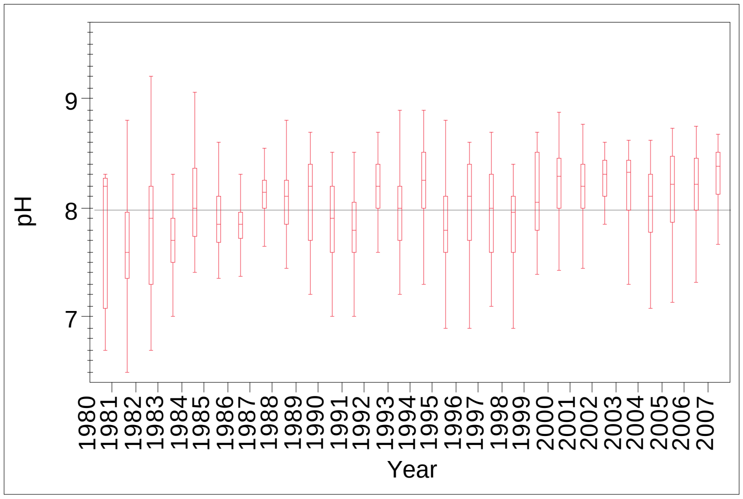

Figure D-18. Quartiles of PH by Year

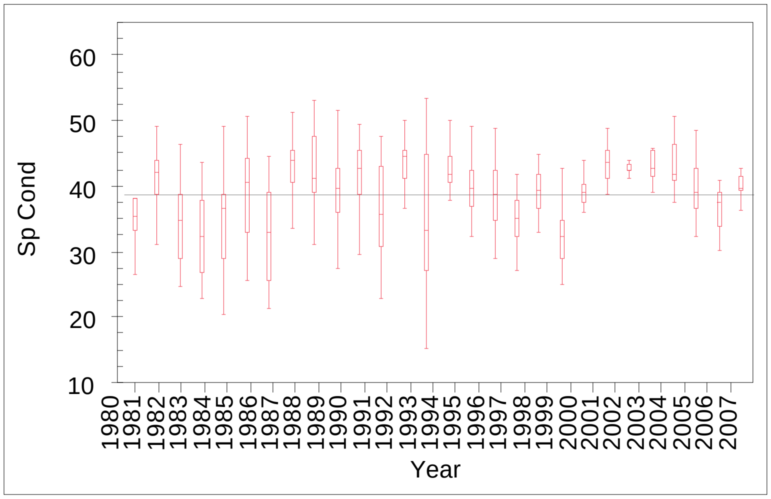

Figure D-19. Quartiles of Specific Conductivity by Year

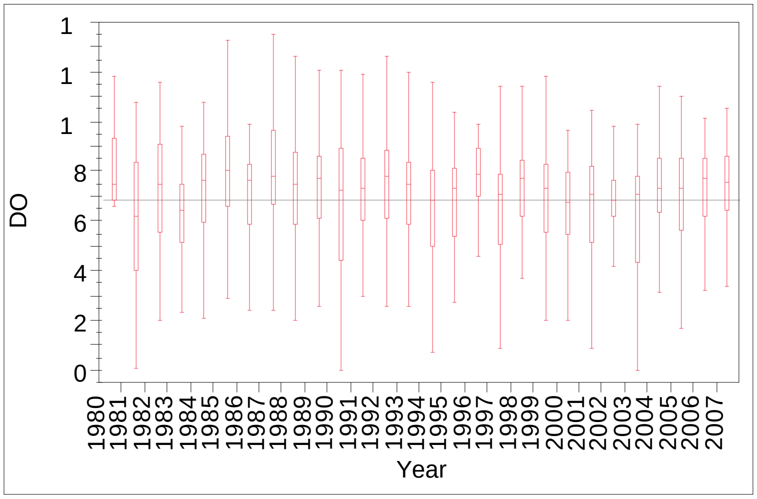

Figure D-20. Quartiles of Dissolved Oxygen by Year

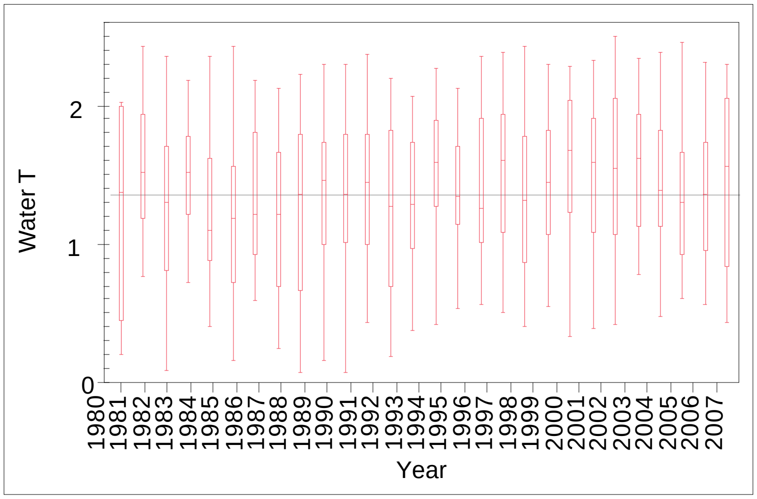

Figure D-21. Quartiles of Water Temperature by Year

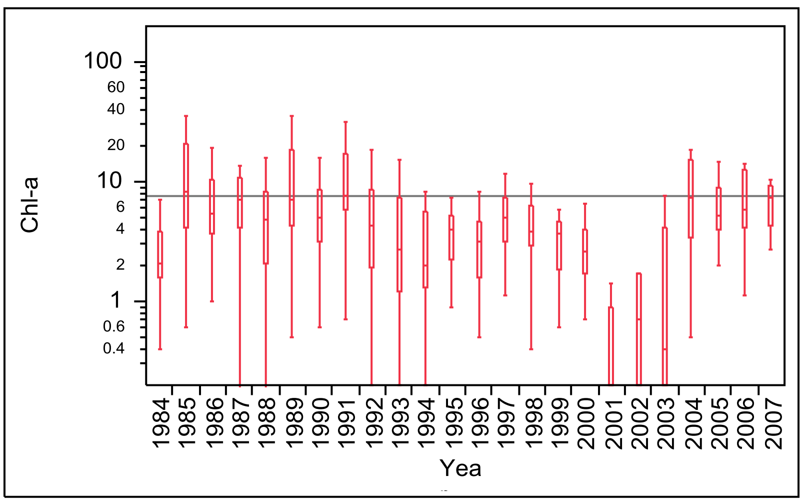

Figure D-22. Quartiles of Chlorophyll-a by Year

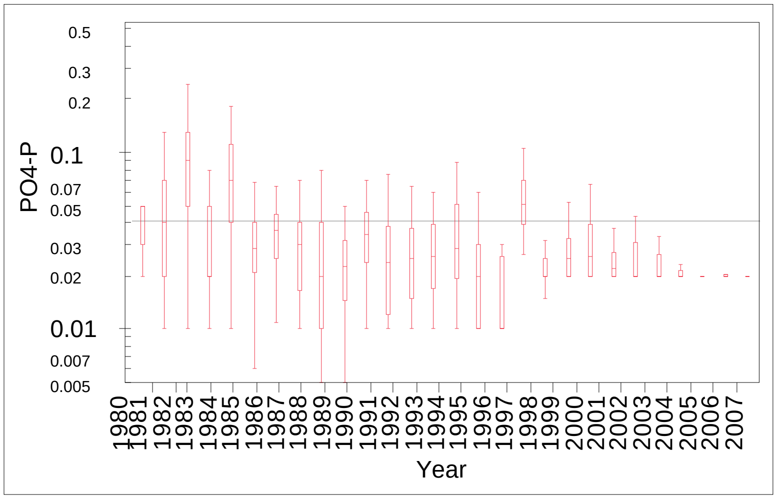

Figure D-23. Quartiles of Phosphate-Phosphorus by Year

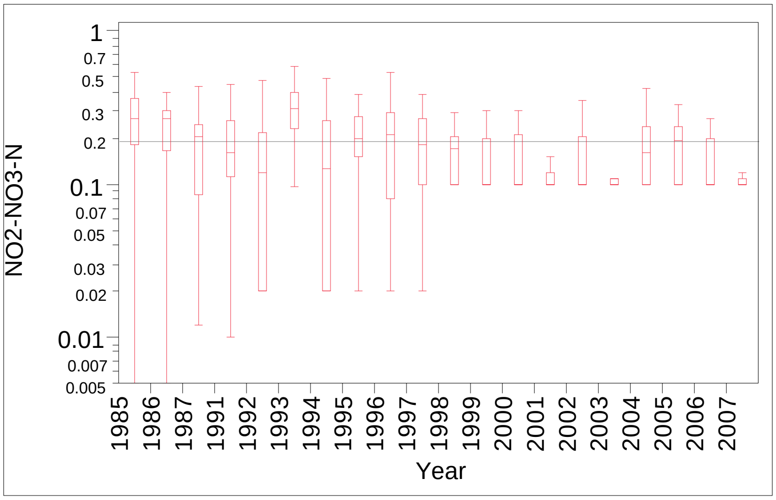

Figure D-24. Quartiles of Nitrite-Nitrate-Nitrogen by Year

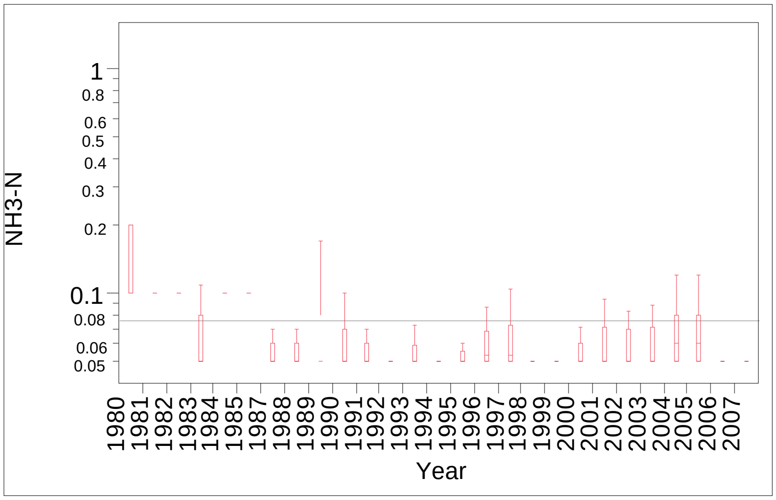

Figure D-25. Quartiles of Ammonia-Nitrogen by Year

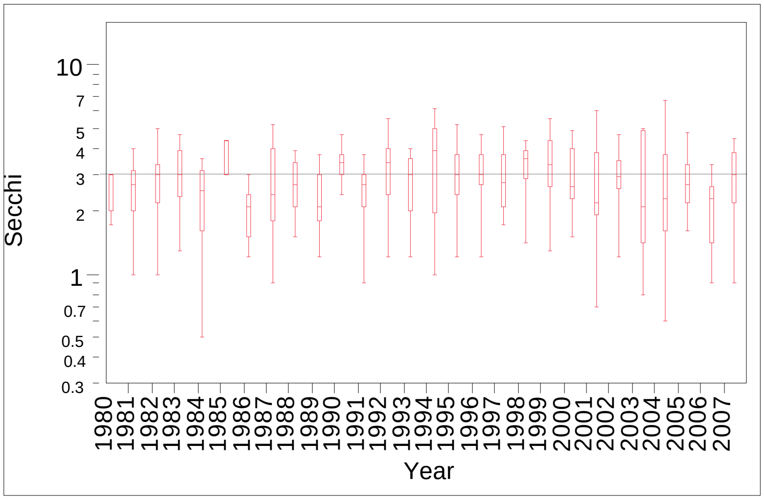

Figure D-26. Quartiles of Secchi Depth by Year

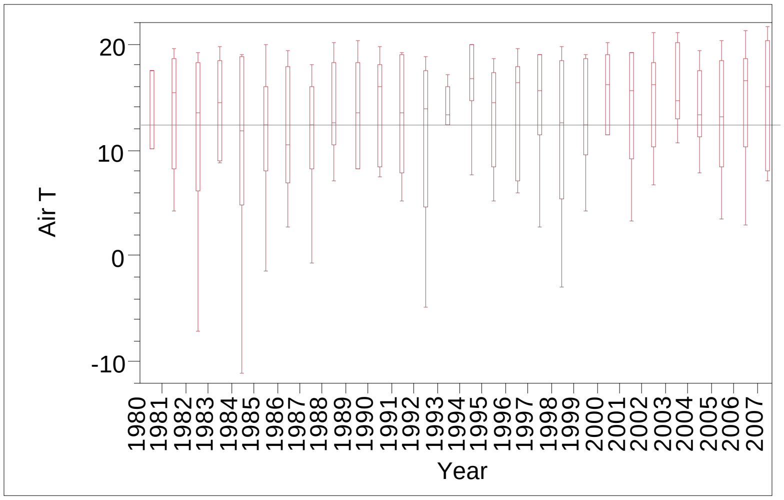

Figure D-27. Quartiles of Air Temperature by Year