

To the Graduate Council:

I am submitting herewith a thesis written by Tim Pobst entitled “Statistical Temporal Analyses on Legacy Data for the GRSM Stream Survey.” I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Environmental Engineering.

---

Dr. John Schwartz, Major Professor

We have read this thesis  
and recommend its acceptance:

---

Dr. Bruce Robinson

---

Dr. Qiang He

Accepted for the Council:

---

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

To the Graduate Council:

I am submitting herewith a thesis written by Tim Pobst entitled “Statistical Temporal Analyses on Legacy Data for the GRSM Stream Survey.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Environmental Engineering.

Dr. John Schwartz, Major Professor

We have read this thesis  
and recommend its acceptance:

Dr. Bruce Robinson

---

Dr. Qiang He

---

Accepted for the Council:

Carolyn R. Hodges

---

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Statistical Temporal Analyses on Legacy Data for the GRSM Stream Survey

A Thesis Presented for  
The Master of Science  
Degree

The University of Tennessee, Knoxville

Tim Pobst

May 2014

© by Tim Pobst, 2014  
All Rights Reserved.

# Acknowledgements

I would like to thank... Dr. Schwartz, Keil Neff, Matt, and Steve Moore.

I would like to thank all the faculty and staff in the civil and environmental engineering program. I was never lacking in offered help, from Larry and Ken to the Professors and the secretaries.

I would like to thank Dr.Schwartz for giving me the opportunity for a graduate degree. My responsibilities involved a lot of field work in the Great Smokey Mountains and I never bored of bragging about it.

I would like to thank Matt Kulp and Steve Moore of wildlife and fisheries in the GRSM for funding me and any work done in the park.

I would like to thank Mei Cai and Keil Neff who were post-docs for much of the time I was working on my degree. Sometimes it was like having multiple bosses saying different things at once but most of time it was multiple mentors that could help me if one was unavailable.

I would like to thank Chris Rollison and Keil Neff for teaching me everything I need to know to help on the GRSM projects.

I would like to thank Jordan Hayes and Matt Aplin who were the two primary undergraduate helpers regarding GRSM work.

# Abstract

Abstract text goes here...

# Contents

<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Description of study area . . . . .	1
1.2 Acid Deposition and the GRSM . . . . .	1
1.3 The Stream Survey . . . . .	3
1.3.1 Database . . . . .	3
1.3.2 Elevation Bands . . . . .	4
1.4 Literature review . . . . .	7
1.5 Data smoothing . . . . .	8
1.6 objectives . . . . .	10
<b>2 Trend Analysis</b>	<b>14</b>
2.1 Methods . . . . .	14
2.2 Results . . . . .	17
2.2.1 Step-wise Julian date coefficients . . . . .	17
2.2.2 Julian date coefficients from time variables only . . . . .	18
2.2.3 Elevation trends . . . . .	19
2.2.4 Results by Comparison . . . . .	20
2.3 Discussion . . . . .	21

<b>3</b>	<b>Means Comparison</b>	<b>23</b>
3.1	Methods . . . . .	23
3.1.1	Introduction . . . . .	23
3.2	Results . . . . .	25
3.3	Discussion . . . . .	27
<b>4</b>	<b>Power Analysis</b>	<b>29</b>
4.1	Methods . . . . .	29
4.1.1	Introduction . . . . .	29
4.1.2	Body . . . . .	30
4.1.3	Procedures . . . . .	30
4.2	Results . . . . .	34
4.2.1	Post hoc . . . . .	34
4.2.2	A priori . . . . .	35
4.3	Discussion . . . . .	38
4.3.1	Post hoc . . . . .	38
4.3.2	A priori . . . . .	39
	<b>Bibliography</b>	<b>41</b>
<b>A</b>		<b>45</b>
A.1	Site Data . . . . .	45
A.2	Site data . . . . .	45
<b>B</b>	<b>Descriptive Statistics</b>	<b>48</b>
<b>C</b>	<b>Variable selection</b>	<b>50</b>
<b>D</b>	<b>Julian Date Coefficients</b>	<b>51</b>
D.1	Step-wise Method . . . . .	51
D.2	Temporal Variables . . . . .	51



<b>E</b>	<b>ANOVA/Bonferoni</b>	<b>54</b>
E.1	pH . . . . .	55
E.2	ANC . . . . .	56
E.3	Nitrate . . . . .	57
E.4	Sulfate . . . . .	58
<b>F</b>	<b>Post Hoc Power Analysis</b>	<b>59</b>
F.1	Step-Wise Variables . . . . .	59
F.2	Temperol variables . . . . .	59
<b>G</b>	<b>A priori analysis</b>	<b>62</b>
G.1	Power graphs . . . . .	62
G.1.1	pH . . . . .	62
G.1.2	ANC and Nitrate . . . . .	62
G.1.3	Sulfate . . . . .	62
G.1.4	Time Variables . . . . .	62
<b>Vita</b>		<b>67</b>

# List of Tables

1.1	Historical elevation bands for the 90 site survey. *Approximate percentages based on planimetering contour map . . . . .	5
1.2	Historical elevation bands for the 43 site survey . . . . .	5
1.3	These elevation classes were created to add more weight to the higher elevations . . . . .	7
2.1	Equations created through step-wise variable selection . . . . .	16
2.2	Dependents regressed against elevation (m) only. . . . .	19
3.1	Bonferoni comparisons between multiple groups . . . . .	25
4.1	Hypothesis tests from Statistical Methods in Water Resources by theUSGS (Helsel and Hirsch, 1992). . . . .	29
4.2	A priori calculation in G*power when alpha, ES, and power are set to .05, .15, and .80 respectively. . . . .	32
4.3	samples/year to achieve a power .80 ( $N_b$ ) . . . . .	33
4.4	Sample sizes at a power of .80 . . . . .	35
4.5	Years to acheive a power of .80 . . . . .	36
4.6	Necessary sites scenario for water quality variables . . . . .	36
4.7	Necessary sites scenario for time variables . . . . .	37
A.1	GRSM Stream Survey site descriptions . . . . .	46
A.2	Site Data . . . . .	47

B.1	Descriptive statistics of Water Quality in the GRSM . . . . .	49
C.1	List of variables used for step-wise variable selection. X's for variables selected by the step-wise method, O's if variable was added after the step-wise process. . . . .	50
D.1	Time trend results for specific elevation classes using variables from step-wise regression. <b>Bold</b> results are insignificant. . . . .	52
D.2	Time trend results for specific elevation classes using julian date, $\cosine(\theta)$ , and $\sin(\theta)$ only. <b>Bold</b> results are insignificant. . . . .	53
F.1	Post hoc power analysis using G*power and a calculated ES, alpha is .05. <b>Bold</b> results are insignificant. . . . .	60
F.2	Post hoc power analysis using G*power a calculated ES, an alpha of .05 with the variables: $\sin(\theta)$ , $\cosine(\theta)$ , and julian date only. <b>Bold</b> results are insignificant. . . . .	61

# List of Figures

1.1	Modeled atmospheric deposition of N and S for the year 2000 and presented in <a href="#">Weathers et al. (2006)</a> . . . . .	6
1.2	pH plotted vs. Elevation. With and without outliers. . . . .	9
3.1	Sulfate emmisions of Kingstion and Bull run against those measured in Noland high elevation site ( <a href="#">UTK, 2012</a> ). . . . .	24
G.1	pH Power Graph . . . . .	63
G.2	ANC and Nitrate Power Graphs . . . . .	64
G.3	Sulfate Power Graph . . . . .	65
G.4	Time Variables Power Graph . . . . .	66

# Chapter 1

## Introduction

### 1.1 Description of study area

The Great Smoky Mountains National Park (GRSM) which is located in the southern Appalachians spanning eastern Tennessee and western North Carolina is the second largest national park in the eastern united states. It contains roughly 100 species of native trees, over 1,500 flowering plants, 200 species of birds, 66 types of mammals, 50 native fishes, 39 kinds of reptiles, and 43 species of amphibians. This unique nature of the park has earned it the title of International Biosphere Reserve by the United Nations (NPS, 2013). The GRSM is one the most visited parks in the US and its conservation is a high priority for the National Park Services (NPS) who look after it. Park conservation includes monitoring streams for the consequences of acid deposition. Acid deposition negatively affects the 3,000 km of streams present in the GRSM, impacting every living thing in the park which rely on the water quality in the park.

### 1.2 Acid Deposition and the GRSM

Acid deposition is characterized as wet deposition (rain and snow), dry deposition (gases and particles), and fog or cloud deposition (occult). These three weather

modes transport and deposit the pollution of the industrialized world. The top contributors of pollution to acid deposition are gas engines for transportation and industrial plants. Power plants expel sulfur oxides ( $\text{SO}_x$ ) and nitrogen oxides ( $\text{NO}_x$ ) through smoke stacks high into the atmosphere where they react and fall to the earth as acid deposition. The upper elevations of the GRSM receive some of the highest loading rates of acidifying nitrogen and sulfur species in North America (Johnson et al., 1992). Once the acid deposition constituents have entered in the environment they react with the surface waters, in the soil, and on man-made structures (Board et al., 1983). This begins to acidify the surface waters which can harm any life forms that interact with it, including the soils as well as streams. Along with adding acid to streams the added pollutants affect the alkalinity of water, or its ability to buffer a change in acidity, and results in a pH change. As the pH level lowers, the ability of surface waters to sustain aquatic life also decline, this is called acidification. And while many factors affect the ability for the survival of aquatic life, acidification can be directly related to a lot of species survival.

Acidification of bodies of water can be either chronic or episodic. Chronic acidification occurs when the body of water has constant low acid neutralizing capability (ANC); which creates a large area of nearly un-inhabitable water where aquatic life would struggle to survive. Episodic acidification describes a rapid increase of acidity due to large surges of pollutants usually from snow melts or heavy rains. While chronic acidification may inhibit habitation, episodic acidification can kill aquatic life by quickly dropping the pH of streams. A literature review in Neff et al. (2009) approximates a pH of 6 for negative biological effects and a pH of 5 for mortality for trout in the park. Stream pH levels between 5 to 6 can become toxic in the presence of aluminum through leaching and base cation exchange. This toxicity can be harmful to eggs and fry in very soft waters in the lower end of the range (Robinson et al., 2008).

## 1.3 The Stream Survey

The stream survey began as part of the park's Inventory and Monitoring program of the GRSM in 1993. It collects grab samples multiple times per year from multiple sites in order to monitor the health of the streams in the park. There are nearly 500 sites listed in the stream survey but the number of sites actually monitored has dwindled to the 43 sites used for this paper. Currently, samples are collected from 32 sites every two months and an additional 11 samples are collected twice per year. These samples cover streams from 6 GRSM stream systems. Every sample is measured for pH, ANC, conductivity, acid anions ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ , ammonia ( $\text{NH}_4^+$ ), the base cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ , and  $\text{Na}^+$ ), dissolved metals (Al, Cu, Fe, Mn, Si and Zn). A ManTech<sup>TM</sup> autotitrator was used for pH, ANC, and conductivity. A Dionex<sup>TM</sup> ion chromatograph (IC) was used for the analysis of  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ , and  $\text{NH}_4^+$ . A Thermo-Scientific<sup>TM</sup> Inductively Coupled Plasma - Atomic Emission Spectrometry (ICP-AES) was used for the study of  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , Al, Cu, Fe, Mn, Si and Zn.

### 1.3.1 Database

All data is collected, under contract, for the NPS. Sample data such as time, location, pH, and constituent concentrations are aggregated in spread sheets and formatted to NPS specifications. This data goes all the way back to the beginning of the survey in 1993. Along with specific sample measurements each sample is labeled by its site ID, which indicates location. Several important characteristics are known for each site such as stream name, geology, and elevation.

The database is dynamic and changes along with changes in the stream survey and lab methods. Currently the collection, analysis, and formatting falls under Dr. Schwartz of the University of Tennessee Civil and Environmental Engineering Department. He inherited it from Dr. Robinson of the same department, who inherited it from the Forestry Department. A difference in analysis and formatting

methods between the two departments is obvious in the data. There are many more outliers present in the data curated by the Forestry Department which makes smoothing and statistically analyzing that half of the data harder. In 2003 the survey changed from collecting grab samples from 90 sites to the current 43 sites (Odom, 2003). This makes the full database inconsistent and consistent data is important in statistical analysis. Another inconsistency is the baseflow/stormflow classification which ends in 2010, while data up to 2012 will be analyzed here. The baseflow/stormflow classification is useful when preparing the data for statistical analysis.

### 1.3.2 Elevation Bands

If the two most important location classifications for the survey data is site ID and stream system, then the third is elevation. Elevation was found to be a dominant driver for predicting water quality among the parks stream. Overall, results from the Biotics Effects report found that stream pH and ANC decreased at  $-.32$  units and  $-35.73$  eq L<sup>-1</sup> respectively, per 1,000-ft elevation gain (Meijun Cai, 2013). This difference in pH due to elevation occurs because of many factors including occult deposition which effects the higher elevations greater. When studying pH in the GRSM elevation bands are used to characterize elevation. Historically sites were divided into eleven elevation bands of arbitrary 500 ft intervals. The resulting distribution of sites are bottom heavy.

Table 1.1 represents the concerns of Kenith Odom presented as table 38 in Odom (2003). This dissertation remodeled the survey from 90 sites to 43 and this table was used to suggest more high elevation sites. The survey was reduced from 90 sites down to 43 but the elevational distribution was not fixed. Table 1.2 shows the percentage of sites per elevation bands for the 43 site survey just as Table 1.1 does for the 90 site survey. There is not much difference between the percentages.

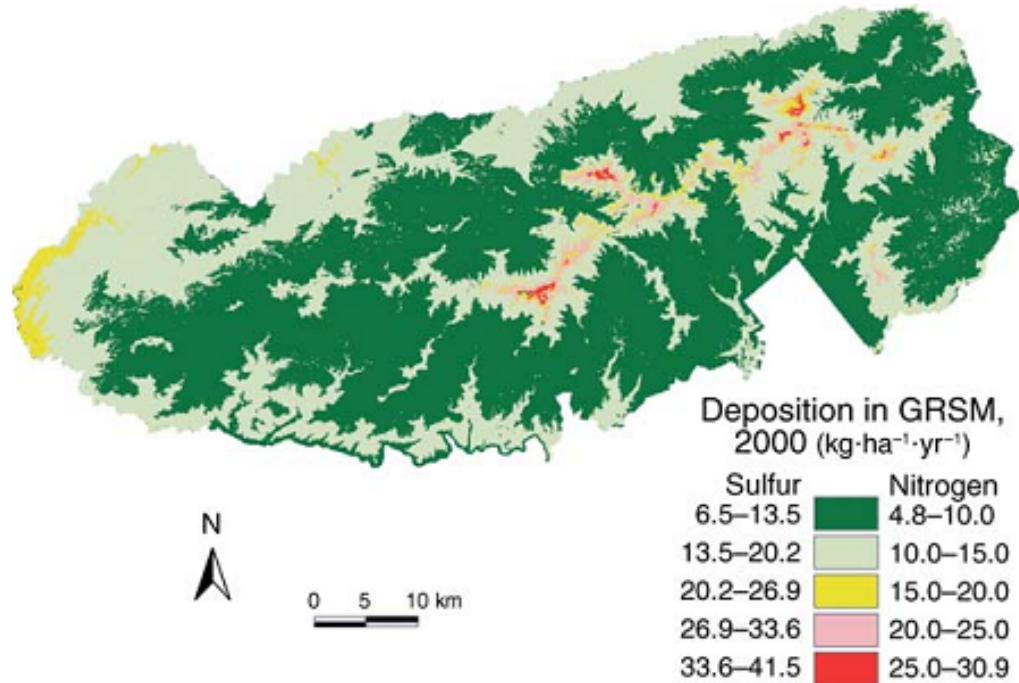


Elevation class	Range of elevation (ft) MSL	Number of sampling sites	Percent of NPS area*	Percent of sampling sites
1	<1000	0		
2	1000-1500	7		
3	1500-2000	13	43.3	65.0
4	2000-2500	16		
5	2500-3000	18		
6	3000-3500	13	27.4	20.5
7	3500-4000	4		
8	4000-4500	5	21.2	12.1
9	4500-5000	5		
10	5000-5500	1	8.1	2.4
11	>5500	1		

**Table 1.1:** Historical elevation bands for the 90 site survey. \*Approximate percentages based on planimetering contour map

Elevation class	Range of Elevation m(ft)	Number of sampling sites	Percent of sampling sites
1	<304.8 (<1000)	0	
2	304.8-457.2 (1000-1500)	4	
3	457.2-609.6 (1500-2000)	4	67.4
4	609.6-762 (2000-2500)	9	
5	762-914.4 (2500-3000)	12	
6	914.4-1066.8 (3000-3500)	6	16.3
7	1066.8-1219.2 (3500-4000)	1	
8	1219.2-1371.6 (4000-4500)	3	14.0
9	1371.6-1524 (4500-5000)	3	
10	>1524 (>5000)	1	2.3

**Table 1.2:** Historical elevation bands for the 43 site survey



**Figure 1.1:** Modeled atmospheric deposition of N and S for the year 2000 and presented in [Weathers et al. \(2006\)](#).

For overall acidification of the GRSM, the high elevation bands could be the most important in the survey but they have the least amount of representation. As can be seen in [Figure 1.1](#), which is a model, the highest deposition of sulfur and nitrogen is at the highest elevations. Sites in these areas continue to receive low pH values in samples. The tenth elevation class according to [Table 1.2](#) has only one site in it. That whole elevation band is represented by only one spot in the GRSM and that is not very strong.

Without adding sites, the easiest way to fix this is to reorganize the elevation bands. For this paper the elevation bands were redesigned to try and strengthen

the higher elevations. A cluster analysis was explored for the task but it was not successful, there was too much variation to cluster by elevation only.

**Table 1.3:** These elevation classes were created to add more weight to the higher elevations

Elevation Classes	Meters (Feet)	n	Site #
1	304.8-609.6 (1000-2000)	5	13, 23, 24, 30, 479
2	609.6-762 (2000-2500)	9	4, 311, 268, 480, 310, 483, 147, 148, 484
3	762-914.4 (2500-3000)	13	114, 481, 482, 149, 66, 492, 137, 293, 270, 493, 485, 144, 224
4	914.4-1066.8 (3000-3500)	4	143, 142, 73, 71
5	1066.8-1371.6 (3500-4500)	4	74, 221, 251, 233
6	1371.6 < (4500 <)	2	253, 234

Table 1.3 contains all the sites of the 43 site survey not removed as outliers, so the 36 sites analyzed in this paper. Each of the following statistical analyses will use these elevation bands to classify elevation for the stream survey data.

## 1.4 Literature review

One of the objectives of this paper is to find time trends among the stream survey data. Trend analysis is a great way to characterize the park's water quality using data collected through the stream survey. It is used to state the condition of the parks water bodies while trying to predict where the water quality is headed in the future. A trend analysis on the stream survey data was conducted in 2002 and published in Robinson et al. (2008) and then again in 2009 for the Biotics Effects report (?). This statistical procedure is used to discover sudden and gradual trends over time. Of the ten elevation bands analyzed in Robinson et al. (2008) six had negative Julian date coefficients and the other four had no trend. Of the 67 sites studied in the biotic effects report most showed no trend, 22 showed an increase in pH and 2 showed a

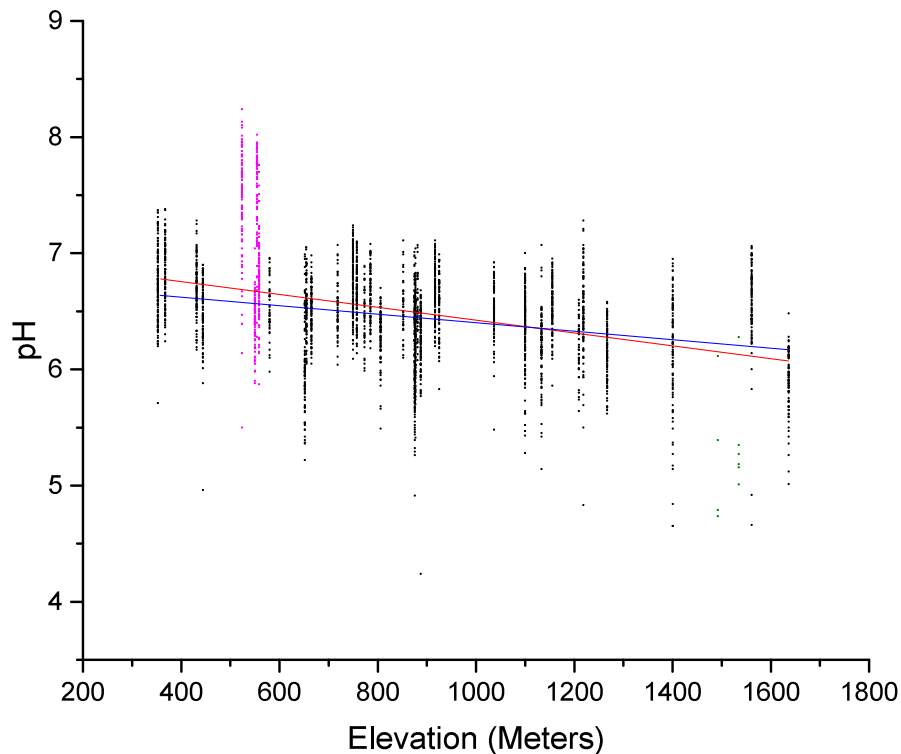
decrease(?). The trend analysis will use stream survey data from 1993 to 2012 using the statistical program SPSS for analysis.

Twenty years of data were available for this paper from the years 1993 to 2012. A single trend line containing all 20 years is unrealistic because it will not show why there is a difference in the previous trend analyses or if there was a change in trend after 2008. The opposite trends reported in Robinson et al. (2008) and ? suggest an inflection point in the trend line somewhere between 2002 and 2009. For this reason, and for easier comparison of results, a separate data set will be sectioned off from 1993 to 2002 to equal the years analyzed in Robinson et al. (2008). A third data set will be created after the year 2008 because this is the year that the Kingston and Bull run power plants installed scrubbers onto their smoke stack exhaust. The hypothesis being the  $\text{SO}_4^{2-}$  concentrations will be noticeably different, and this difference could indicate a need for further study. These three time sets will be analyzed separately.

## 1.5 Data smoothing

All of the statistical analysis was completed in statistical software. Initial data smoothing and influential data points were found using JMP 9.

Several plots were created in order to reduce the variance of the data before any statistical analysis was attempted. JMP was chosen for this task based on its ease of use in plotting data. A plot of pH vs. time visually represents the existence of a positive trend in the pH data from 1993 to 2012. Figure 1.2 shows the pH vs. elevation plot. It shows two trend lines, one which represents the trend of all of the data points and the other represents the trend after the influential points are removed. Both of the trends are negative as elevation increases but the trend line containing the influential points is steeper. pH was plotted against the month that the sample was collected to check for seasonality. Seasonality was expected and found for pH over one year. The variance caused by seasonality will be removed with sine and cosine functions.



**Figure 1.2:** pH plotted vs. Elevation. With and without outliers.

Much of the variance in [Figure 1.2](#) can be attributed to known influences in the stream survey data: Abrams creek watershed, sites that are affected by anakeesta geology, and stormflow ([Neff et al., 2012](#)). Comparatively Abrams is a low elevation, low slope area where the underlying geology is Cades Sandstone, which buffers against acid rain very well. This sandstone contributes to high ANC values which in turn keep the mean pH levels higher than the rest of the sites in the survey. Site numbers 237 and 252 are sites which are down hill of road cuts that have exposed the underlying anakeesta formation to runoff. The anakeesta formation contains sulfidic slate, which can have the same negative effect of acid deposition, and keeps the pH values of streams very low.

Stormflow is both influential and detrimental to GRSM water quality. Storms can bring high intensity rain fall which can very quickly reduce the pH and ANC of streams. In streams with low ANC and pH, episodic stream acidification can be harmful to aquatic life (Neff et al., 2009). Stormflow runoff can have a higher contribution to stream acidification than baseflow because it transports protons left in the upper layers of the soil by acid deposition. Stream acidification caused by stormflow can cause base cation exchange through the leaching of the surrounding soil. When the inherent base cation minerals run out, excess  $H^+$  and Al will be released into the water. Increasing the  $H^+$  concentration will lower the pH, and Al is toxic to biotics. Healthy streams can rebound to normal pH values; unhealthy streams can have permanently lowered ANC due to leaching. Measurements taken from stormflow can show uncharacteristically low pH values and high amounts of metals from leaching. In this way, stormflow is sometimes considered an influential group on the rest of the data, because the measurements are significantly different from the average. Dr. Cai characterized all of the available water quality data between 1993 and 2010 as storm flow or baseflow; this work is summarized in ?. Water quality data after 2010 had not been characterized. If all stormflow observations are to be kept in the data, the years 2011 and 2012 would need to be characterized. Quick analyses were run to see how influential stormflow was on the data as a whole, and it turned out that some were and many were not. Instead of throwing out all of the stormflow observations at once, single influential observations could be explained by stormflow and removed. They can be removed on a case by case basis during the regression method.

## 1.6 objectives

Objectives of this study were to:

- characterize time trends in stream pH and acidic anions among elevation ranges in order to assess whether conditions are improving or degrading, and

- characterize sampling variance based on available water quality data, within the context of time and elevation, to support development of the GRSMs Vital Signs Monitoring Program. In addition to supporting this Program, a cost analysis of selected water quality sampling schemes will be summarized. The format of this thesis will follow these two objectives.
- Has stream pH and acid anion concentrations changed among three time periods (1993-2002, 2003-2008, and 2009-2012), and among six elevation ranges (1000-2000ft, 2000-2500ft, 2500-3000ft, 3000-3500ft, 3500-4500ft, 4500+)?
  - ANOVA
  - Time trends
- What is the statistical power for water quality parameters based on frequency and elevational location based on historic data?
  - Post Hoc Analysis
  - A Priori Analysis

The thesis is organized into two separate chapters following the two above research questions. Each chapter will follow the technical format of introduction, methods, results, and discussion.

An aspect of this program is a high elevation site, Noland Divide Watershed. NDW is a small, 17.4 ha, watershed about 2 km east of Clingmans Dome, which is the second highest peak in the eastern US. NDW contains sites to measure wet and dry deposition as well as "grab" samples to study in the lab. This site is a continuation of the Integrated Forest Study (IFS) and was taken over by the park in 1991.

Information leading to a better understanding of water quality conditions per watershed characteristics aids GRSM resource staff in managing the Parks natural resources.

Other than chemistry the acidity of the park streams can be affected by geographical attributes such as geology and elevation. Stream acidification is more pronounced in higher elevation watersheds with underlying base-poor geology

The acidification patterns along an elevational gradient occur throughout the Appalachian region similar to the data collected in the GRSM streams.

This figure shows all pH data from 1993 to 2012 vs. Elevation (m). The red trend line includes abrams, sites 237 and 252. These outliers are removed for the blue trend line.

Figure 1 is a graph of all measured pH values for Stream Survey between the years of 1993 and 2012 graphed against the elevations they were collected. Two notable figures are here, one that contains all pH data from 1993 to 2012 and another that has removed the outliers Abrams stream system along with sites 252 and 237. The steeper sloped line represents the data set that still contains Abrams, sites 237, and 252. This graph clearly illustrates pH trending downwards as elevation increases. Increases in sulfate or nitrate in runoff and streams through acid deposition can cause an increase leaching of base cations, hydrogen ion, aluminum, and decreased ANC (Sullivan, 2004). Leached aluminum can end up in streams and is toxic to fish (Driscoll 2003). Rainfall and fog in the GRSM affect elevations above 4000 feet first higher elevations have steeper slopes which correlate to both thinner soils and base poor geology. All these factors contribute to lower pH in higher elevations.

Neff et al. (2013) observed this trend, but found other basin factors correlate to chronic and episodic acidification. Other basins factors include drainage area size, vegetation, and soil properties, which factors all are change within an elevational gradient.

In support of GRSM natural resource management, stream water quality has been monitored since 1993, and this long-term dataset was recently statistically analyzed and commonly referred to as the Biotic Effects Report (Cai et al. 2013).

Conductivity, chloride, and base cations were also found to significantly decrease with elevation gain. Sulfate showed no significant trend with elevation, however



nitrate was found to significantly increase with elevation gain. The GRSM 2011 Annual Water Quality Report compared pH trend lines representing the current 43 sites from 1993-2010 with 2011. The data showed lines of similar slopes with different intercepts, which was interpreted to mean increasing pH at all elevations in GRSM streams. Acid deposition increases with elevation in the GRSM and the higher elevation streams would experience increased sulfate, and prolonged acidification if soil desorption becomes a dominant geochemical watershed process which could occur if pH increased to 6.0 and sulfate dropped below 50 eq L<sup>-1</sup>. From a management perspective, the Biotic Effects Report contains limitations in the analyses to assess long-term changes because locations sampled have changed over time and most current sample locations are at lower elevations.

Table 1 shows the current historical elevation classes with the number of sites they include and the percent of sites in each band. Although there were limitations, the biotic effects report provided essential information on how to improve future water quality and biological monitoring programs.

The National Park Service is currently developing a Vital Sign Monitoring Program for a number of national parks around the country, which includes the GRSM (Annual Administrative Report For Inventories and Vital Signs Monitoring FY 2010). The goal of the GRSM Vital Sign Monitoring Program is to assess long-term changes in ecosystem health, including both terrestrial and aquatic environments. In addition, the program will integrate these environments and include existing data on basin factors associated with water quality.

Although it is widely recognized that elevational gradient is important, a question about what elevation ranges are changing versus those that are not needs to be quantified in order to design an effective sampling program. In addition, data must be collected at a set frequency to provide sufficient statistical power to make valid interpretations of change over time.

# Chapter 2

## Trend Analysis

### 2.1 Methods

Water quality is an ongoing concern for the park. The acidification of the streams can have significant negative effects on wildlife and vegetation. The stream survey collects water samples all over the park to monitor the health of the water.

The data used in these analyses are collected through the park wide stream survey. The stream survey includes six stream systems and five of them are collected every two months and analyzed in a lab for many water quality variables including pH, ANC,  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$  and some metals.

Two more factor divisions of the data include dividing the data by elevation classes and four dependent variables (pH, ANC,  $\text{NO}_3^-$ , and  $\text{SO}_4^{2-}$ ).

Dividing all the data into three different time sets, six elevation bands and studying four different dependents will create 72 different trend lines.

#### Instruments

The regression process includes preparing the data and identifying influential observations. The output of a step-wise regression analysis performed in SPSS can be configured to complete many different analyses in order to smooth the data. The

different tests applied for this paper include tests for normality, heteroscedasticity, cook's D, DFBETAS, and DFFITS. As observations were identified by cook's d, DFBETAS, and or DFFITS as influential, they were individually analyzed to determine what made them influential. Modification or removal of an influential observation had to be justified, or it would remain an outlier. An example of modification of the data included a pH value that read 16.47 was changed to 6.47. Another example is that some conductivity values were obvious copies of the ANC value for the same observation. These conductivity values were removed. Some influential observations were not as obvious and if they could not be labeled as storm flow or human error they would be kept. After sufficient attention was given to the influential observations the analysis was re-run and more influential observations could be found, and attention would need be given to these also. This process was completed for all four of the dependent variables, (pH , ANC,  $\text{NO}_3^-$ , and  $\text{SO}_4^{2-}$ ).

The step-wise variable selection process requires a list of variables to choose from. These variables are reported in [Table C.1](#). The variables chosen for this list were chosen from those chemistry values recorded in the full stream survey dataset. One benefit of choosing only variables directly from the stream survey dataset is a high ease of repeatability for the future. The step-wise process regulates entry into the equations by the probability of the F statistic. If this statistic were between .05 and .10 then the variable could stay. The variables selected were used to create the fixed models presented in [Table 2.1](#). If any of the time variables were chosen by the step-wise method then the others were added. This was done to ensure the Julian date coefficient was present along with  $\sin(\theta)$  and  $\cos(\theta)$  for seasonality. Many variables are present in the stream survey database, some are measurements but others were derived. Mathematically seasonality can be modeled with the  $\sin(\theta)$  and  $\cos(\theta)$  variables as shown in [Helsel and Hirsch \(1992\)](#). They represent each day of the year as a fraction of the year and place the lowest pH on January 1 and the highest on July 1. The variable BC (base cations) represent the sums of the  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ , and  $\text{Na}^+$  measurements. Correlations were run between each of the proposed

variables and both ANC and BC were found to be better described as  $\log_2(ANC)$  and  $\log_2(BC)$ .

**Table 2.1:** Equations created through step-wise variable selection

Dependent (n)	Model	Adjusted $r^2$	Model p
pH (3116)	$.673 \times \log_2(\text{Sum Base Cations}) + (-.368 \times \text{NO}_3) + (.262 \times \text{Julian Day}) + (-.266 \times \text{SO}_4) + (-.050 \times \cos(\theta))$	0.630	<0.001
ANC (3116)	$(.415 \times \text{Sum Base Cations}) + (-.185 \times \text{SO}_4) + (.595 \times \text{Conductivity}) + (-.102 \times \text{NO}_3) + (.019 \times \text{Julian Date}) + (.005 \times \text{Cl}) + (.005 \times \sin(\theta))$	0.984	0.049
NO <sub>3</sub> (3116)	$(-.295 \times \text{SO}_4) + (-3.183 \times \text{ANC}) + (2.19 \times \text{Conductivity}) + (.923 \times \text{Sum Base Cations}) + (.120 \times \text{Julian Date}) + (.051 \times \text{Cl}) + (.047 \times \sin(\theta)) + (.031 \times \cos(\theta))$	0.498	0.017
SO <sub>4</sub> (3116)	$(-.166 \times \text{NO}_3) + (2.318 \times \text{Conductivity}) + (-3.229 \times \text{ANC}) + (1.033 \times \text{Sum Base Cations}) + (.042 \times \text{Julian Date})$	0.720	<0.001

The difficulty in modeling a time trend is the high amount of variation within the datasets. While trying to determine a time trend other variables are added besides those that explain a trend in time. All of the equations contain the time variables (julian date,  $\sin(\theta)$ , and  $\cos(\theta)$ ) along with the chosen chemical variables. Because of the difficulty of explaining what the Julian date coefficient really means along side the chemical variables a second set of equations was created for analysis. Theses equations use only the three time variables to describe each of the dependents.

Elevation was not a significant predictor for any of the dependent water quality variables chosen. The dependent variables were regressed using simple linear regression against elevation in meters in order to determine their trends by elevation. These trends encompass all elevations; no elevation bands were used.

## 2.2 Results

Julian date coefficients are reported in Robinson et al. (2008) for each of the eleven historic elevation classes and across each of the dependent variables (pH, ANC,  $\text{NO}_3^-$ , and  $\text{SO}_4^{2-}$ ). Julian date coefficients for this paper were reported in similar tables. 144 different Julian date coefficients were calculated and are presented in two tables. Table D.1 records the Julian date coefficients calculated using the equations in Table 2.1 and Table D.2 records the Julian date coefficients for equations containing only the three time variables. Each trend line is represented by its Julian date coefficient, the  $r^2$  value for the trend line, and its statistical significance.

2 of the 72 trend lines in Table D.1 are insignificant. In contrast 50 of the 72 trend lines in Table D.2 are insignificant. Setting the linear regression  $\alpha$  at .05 forces any trend with a p-value greater than .05 to be insignificant. Insignificance rejects the hypothesis that  $\beta(\text{the coefficient}) \neq 0$ . A p-value greater than .05 means that there is greater than a 5% chance that  $\beta = 0$  or in this case the Julian date coefficient = 0.

### 2.2.1 Step-wise Julian date coefficients

#### pH

The Julian date coefficients In Table D.1 for pH showed negative time trends in three statistically significant regression lines, all in the time range of 1993-2002. These lines were in elevation classes 2, 3, and 5. There is one degrading trend in the third time set (2009-2012) and in the fifth elevation class but it is insignificant. Most of the trend lines report that pH is increasing over time.

#### ANC

Trends for ANC fluctuate while evaluating across time sets and elevation classes . Eleven of the lines are positive, and seven are negative. Two of the three negative trends for ANC in set 2 have a smaller slope in set 3, and one of the degrading trends

in set 2 becomes positive in set 3. When comparing time set 2 to set 3, ANC trends are growing over time.

### **Nitrate**

$\text{NO}_3^-$  trends in set 2 are all positive. In set 3  $\text{NO}_3^-$  has a decreasing trend in elevation class 4. The  $\text{NO}_3^-$  trends for set 1 are half positive and half negative. But from the years 2003 to 2008 all of the  $\text{NO}_3^-$  trends are positive. In set 3, the trend in elevation class 4 has a negative trend.

### **Sulfate**

$\text{SO}_4^{2-}$  has mixed positive and negative trends for set 1 but all positive trends for set 2. Half of the  $\text{SO}_4^{2-}$  trends in set 3 are negative ( 1, 3, and 6).

## **2.2.2 Julian date coefficients from time variables only**

In [Table D.2](#) only 20 of the 72 regression lines are significant, those that have acceptable p-values less than .05.

### **pH**

The dependent variable pH in set 1 has zero significant lines, set 2 and 3 combined are slightly less than half insignificant trend lines. The insignificance of the trend lines leaves them untrustworthy, but the trend values themselves are quite similar to those calculated in [Table D.1](#).

### **ANC**

There are only two significant regression lines in for ANC in [Table D.2](#). Elevation class 5 in set 1 has a decreasing trend at -.148, there are no significant lines in set 2 and set 3 elevation class 5 has a positive trend at .891.

## Nitrate and Sulfate

$\text{NO}_3^-$  and  $\text{SO}_4^{2-}$  both had negative trends in set 1 class 1. These are the only significant decreasing trends exhibited for either  $\text{NO}_3^-$  or  $\text{SO}_4^{2-}$  in [Table D.2](#). Both have positive trends in set 2 at elevation classes 1,2,4 and 6, and neither variable have significant lines in set 3.

### 2.2.3 Elevation trends

**Table 2.2:** Dependents regressed against elevation (m) only.

set	Dependent	n	slope	$r^2$	per +1000m
1	pH	1357	.000	.173	-0.411
	ANC	1354	-.056	.199	-56.227
	$\text{NO}_3^-$	1161	.032	.372	32.211
	$\text{SO}_4^{2-}$	1343	.037	.108	37.371
	SBC	1358	.013	.005	13.065
2	pH	997	.000	.094	-0.391
	ANC	997	-.051	.157	-50.970
	$\text{NO}_3^-$	995	.031	.307	30.677
	$\text{SO}_4^{2-}$	1029	.036	.098	35.793
	SBC	1031	.016	.009	15.537
3	pH	757	.000	.061	-0.286
	ANC	757	-.036	.087	-35.689
	$\text{NO}_3^-$	757	.026	.195	25.924
	$\text{SO}_4^{2-}$	757	.030	.101	29.715
	SBC	757	.020	.014	19.905

The aim of [Table 2.2](#) is to calculate the change in water quality values for every 1000 meters of elevation. The base cations were added as a dependent for this analysis. All of the pH and ANC values decrease as elevation increases and all of the  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ , and base cations dependents increase as elevation increases. Every value in the right most column decreases for the water quality dependents as the table moves forward in time sets except for the base cations.

## 2.2.4 Results by Comparison

In comparing table 4 from Robinson et al. (2008) with Table D.1 from this study, it needs to be noted that the elevation classes are different and the data sets have slightly changed throughout the years. The largest difference is the reduction of 90 sites to 43. Abrams was not included in this analysis but was included in Robinson et al. (2008). This difference could explain the differences seen in the old elevation classes from Robinson et al. (2008) of 1,2, and 3 and elevation class 1 in this study. Two sites (237, 252) that are in the new elevation class 6 were left out of the statistical analysis as influential observations. These correspond to historic elevation class 9.

One interesting comparison between table 4 of Robinson et al. (2008) and set 1 of this study are the differences in pH coefficients. All of the pH trends presented in table 4 of Robinson et al. (2008) are negative which is what led to the statements that pH is dropping and can continue to dangerous levels in the future. However, only half the time trend trends for set 1 for pH found in this study were negative in Table D.1. All of the rest of the pH trends for Julian date for both trend analyses are positive when they are significant.

**pH and ANC** For a data set of 92 sites within the time frame of 1993 to 2009 ? reports a decrease for pH and ANC of -0.32 pH units and -35.73  $\mu\text{eq L}^{-1}$  per 1000-ft elevation gain or 302-m elevation gain respectively. These values are close to those found in this study for the years of 2009-2012, but the slopes in set 1 and 2 are much steeper. In set 3, pH is significantly lower with a trend of -.0286 pH units per 1000-m gain and ANC is a little bit lower with a trend of -35.689  $\mu\text{eq L}^{-1}$  per 1000-m gain (Table 2.2).

**Nitrate and Sulfate** The positive  $\text{SO}_4^{2-}$  trends seem to decrease by 2  $\mu\text{eq L}^{-1}$  between set 1 and set 2 in Table 2.2 and then by 6  $\mu\text{eq L}^{-1}$  between set 2 and 3. In contrast, an insignificant negative trend with elevation was found in ? for the years 1993 to 2009.  $\text{NO}_3^-$  follows a similar pattern as  $\text{SO}_4^{2-}$  in Table 2.2 which is also in



agreement with findings in [Weathers et al. \(2006\)](#). As the trends for  $\text{NO}_3^-$  and  $\text{SO}_4^{2-}$  decrease over the time sets the base cations increase by  $2 \mu\text{eq L}^{-1}$  between set 1 and set 2 and then by almost  $5 \mu\text{eq L}^{-1}$  between set 2 and set 3.

## 2.3 Discussion

It is interesting that the step-wise process did not choose elevation as an independent variable for any of the dependent variables. [Figure 1.2](#) clearly shows a decreasing trend for pH while increasing the elevation. Individual elevation classes might be too small to show a significant elevation trend. Increasing acidification with increased elevation was observed in ? will analyzing the entire 1993 to 2009 dataset available. This suggests that there is an elevation trend it is just not as important as other factors when studying acidification in the GRSM.

A trend in time is also clearly evident with a simple plot of pH vs. time but the mostly insignificant trends of [Table D.2](#) suggest otherwise. The three time variables alone are not enough to explain the dependent variables. [Robinson et al. \(2008\)](#) found that pH was decreasing over time when looking at stream survey data between 1993 to 2002, although this study found that most of the trends in that period are negative, the trends for 2009 to 2012 are all positive as well as the trends for 2003 to 2008. This is in agreement with values reported in ?. The differences between the results in [Robinson et al. \(2008\)](#) and those in [Table D.1](#) and [Table D.2](#) imply that water quality is worse in the past but is getting better. Both [Robinson et al. \(2008\)](#) and ? used more than double the sites of this study and [Robinson et al. \(2008\)](#) allowed Abrams to stay in the data. The differences in the data can account for differences in the results but it is safe to say that water quality in the park is getting healthier.

$\text{SO}_4^{2-}$  has more decreasing trends for the years 2009 to 2012 than in any other time set. This is not surprising based on the values shown in [Figure 3.1](#) in which  $\text{SO}_4^{2-}$  concentrations at the high elevation site Noland begin to drop along with emissions from Kingston and Bull run power plants. It is surprising that ? found an

insignificant but negative trend in  $\text{SO}_4^{2-}$  as elevation increases while this study shows only increasing elevation trends for all time sets. When looking at a graph of  $\text{SO}_4^{2-}$  vs. elevation there are many higher elevation outliers present, these outliers could make the difference in findings.

Water quality is increasing. pH and ANC are rising and the pollutants  $\text{NO}_3^-$  and  $\text{SO}_4^{2-}$  are decreasing. The concerns of lowering pH raised in Robinson et al. (2008) are now not as important as those for  $\text{SO}_4^{2-}$  desorption raised in ?. The lack of elevation trend in  $\text{SO}_4^{2-}$  was attributed to high elevation soil adsorption of depositional  $\text{SO}_4^{2-}$  and a statement was made that  $\text{SO}_4^{2-}$  remains absorbed to soil particles as long as soil water chemistry remains high in  $\text{SO}_4^{2-}$  concentration and low in pH (Cai et al., 2011). The slope for the elevation trend of  $\text{SO}_4^{2-}$  over the three sets is decreasing but most of the mean  $\text{SO}_4^{2-}$  concentrations listed in Table B.1 are increasing through time along with pH.

The advantage of using regression for trend analysis is its prediction abilities but regression is more difficult than the nonparametric methods of trend analysis. Tests for normality and heteroscedasticity along with variable transformations take care of forcing the usually nonparametric water quality data to be parametric. Nonparametric tests are more robust and do not require as much preparations to run and in the end are more reliable. Robinson et al. (2008) predicted negative trends and 9.4 years for the historic elevation class between 914 and 1067 meters to reach a pH of 6.00. This corresponds exactly to this study's elevation band 4 which received an increasing pH trend in all three time sets. The differences being the sites used and the equations formed through the step-wise process. The equations in Robinson et al. (2008) follow the theory behind acidification much more closely where as the equations created in this study used variables already available in the running stream survey dataset. Prediction is hard and unless it is absolutely necessary to use then the Mann-Kendal test for trends would be much easier, more reliable and more robust (Helsel and Hirsch, 1992).

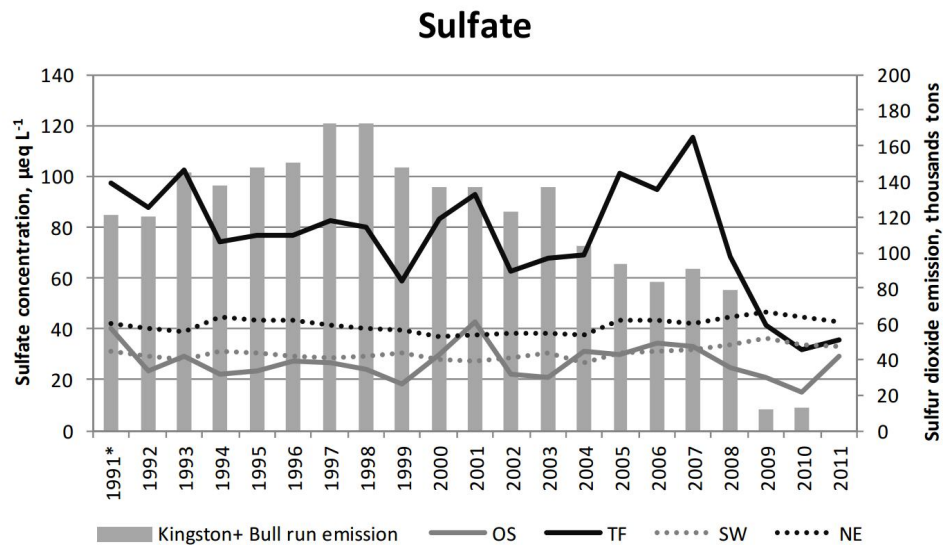
# Chapter 3

## Means Comparison

### 3.1 Methods

#### 3.1.1 Introduction

Bull run and Kingston power plants installed scrubbers in the year 2008 as mandated by the EPA. These scrubbers may have significantly reduced the amount of sulfur dioxide emitted by the smoke stacks. According to [Figure 3.1](#), which is a bar chart depicting the sum of sulfur dioxide emissions of Kingston and Bull run power plants, the sulfur dioxide concentration dropped from 80 thousand tons in 2008 to about 15 thousand tons in 2009. Interestingly the through fall  $\text{SO}_4$  concentrations dramatically decline from about  $115 \mu\text{eq L}^{-1}$  in 2007 to about  $30 \mu\text{eq L}^{-1}$  in 2010. The hypothesis is that the decrease in sulfur dioxide emissions could correlate to the decrease in  $\text{SO}_4$  concentrations measured in Noland Divide through fall. And assuming that the sulfur dioxide emissions from Kingston and Bull run power plants affect the whole GRSM park then there may be signals for this affect in the data. To examine this each time set will be tested against each other by way of means comparison methods.



**Figure 3.1:** Sulfate emissions of Kingston and Bull run against those measured in Noland high elevation site (UTK, 2012).

## Instruments

ANOVA is the standard means comparison method. But it cannot compare more than two groups at once, and a method is needed that can compare all three time sets. The Bonferroni multiple comparisons method is an option in the SPSS statistical program and can compare more than two groups at a time. The explanation of the Bonferroni method given by the SPSS manual is that this method uses *t* tests to perform pairwise comparisons between group means and that the observed significance level is adjusted for the fact that multiple comparisons are being made (IBM, 2013).

The Bonferroni method will create two specific types of outputs. The first is a line graph showing the means of each group analyzed. And the second is a table of pairwise

listings of all the groups compared to each other. This table contains 95% confidence intervals and the significance associated with each comparison. These confidence intervals are produced by the difference in means between the two groups being compared. If the confidence interval includes zero then the groups are statistically the same or equal.

Using SPSS and the Bonferroni method three time sets (93-02, 03-08, 09-12) will be compared at six elevation class levels and across four water quality variables (pH, ANC, NO<sub>3</sub>, and SO<sub>4</sub>). Each group compared is the same stream survey data analyzed in chapter 2 and chapter 4 of this paper.

## 3.2 Results

**Table 3.1:** Bonferoni comparisons between multiple groups

Elevation Classes	pH			ANC			Nitrate			Sulfate		
	1-2	1-3	2-3	1-2	1-3	2-3	1-2	1-3	2-3	1-2	1-3	2-3
1	≠	≠	≠	=	=	=	≠	=	=	=	=	=
2	=	=	=	=	≠	=	≠	≠	=	≠	≠	=
3	≠	≠	≠	=	≠	=	=	≠	≠	=	=	=
4	=	≠	≠	=	=	=	=	=	=	=	=	=
5	≠	≠	≠	=	≠	≠	≠	=	≠	=	=	=
6	=	≠	≠	=	=	=	=	=	=	=	=	=

Table 3.1 reports the Bonferoni comparison means between the four water quality variables (pH, ANC, NO<sub>3</sub>, SO<sub>4</sub>) in one time set against the same water quality variable in another time set by elevation bands. In the table there are three columns per water quality variable. And each column represents the comparison of two groups of the same variable in different times. If the Bonferroni comparison found the groups to be equal then equality was represented by an equal sign. If the groups are not equal an not equal sign was used to show this. All groups that were found to be equal were

insignificant and all groups that were unequal are significant at the 0.05  $\alpha$  level. The bonferoni figures are presented in [section E.1](#). There are six figures for each of the water quality variables, one for each of the elevation classes. For the most part the figures representing pH seem to be continuing a similar rate of change through the years, but this is also misleading because the time sets do not contain equal numbers of years.

The set comparisons between pH are the first comparisons presented. pH contains more unequal sets than any other water quality variable. Mostly all the sets are different except between elevation class 2 which is all the same and class 4 and 6 which show sets 1 and 2 being equal. If a pronounced elevational trend existed for pH in the GRSM, this trend would be visible in the Bonferroni figures. Following the means of each time set through the different elevation classes the largest mean should be in elevation class 1 and the smallest in elevation class 6. Unfortunately elevation class 2 always contains the lowest means instead of Elevation class 6. And elevation class 3 behaves as if it should be between elevation class 5 and 6.

In contrast to the pH figures, the ANC figures do not all have similar rates of change. In the odd numbered classes ANC reached a peak in set 2 and dropped for set 3. All of the ANC figures have a decreasing trend from set 2 to 3 except for class 2 which is steadily increasing. The comparison found a lot of equality between the means of the ANC sets. Elevations 1, 4, and 6 are all equal. At elevations 2 and 3 set 1 and 3 are the only sets that are unequal. And at elevation 5 sets 1 and 2 are equal but set 3 is not. The set means presented in the ANC figures vary greatly in concentration. Classes 1 and 2 are more than double the means of the other classes. The ANC concentrations of elevation class 2 are the lowest which helps explain why the pH of elevation class 2 is also the lowest. It is important to note here that even though class 2's concentrations are the lowest, they are also the only concentrations that are increasing.

For  $\text{NO}_3$  elevation classes 4 and 6 are all the same and elevation 3 shows sets 1 and 2 being equal where 3 is not. Elevation class 1 is the opposite of expected showing all

being equal except for sets 1 and 2. In elevation class 2 sets 3 and 2 are equal and in elevation class 5 sets 3 and 1 are equal. The odd numbered figures for  $\text{NO}_3$  all have decreasing mean values from set 2 to 3. In elevation classes 2 and 4 the mean values for set 3 are higher than those in set 2 but the rate of increase is slowing. Elevation 6 is always increasing. The  $\text{NO}_3$  figures show mostly decreasing concentrations over time, except for class 6. The odd classes all have decreasing negative trends from set 2 to set 3 while classes 2 and 4 have decreasing positive trends between set 2 and 3.

$\text{SO}_4$  shows all three sets being equal across all elevation classes except for class 2. Class 2 shows only sets three and two being equal. These figures can sometimes be misleading when visually comparing groups and it is always best to have the confidence intervals on hand. For example when looking at the  $\text{SO}_4$  figures many of the means look different across the time sets, but according to the table, except for class 2, they are all equal across time sets. All of the set 3 means for  $\text{SO}_4$  are larger than their respective set 2 means except for those in class 2 which has negative trends throughout.

### 3.3 Discussion

Because this analysis was completed after the trend analysis and descriptive statistics for these data sets were available certain results were expected. Such as the increase in pH over time and the abnormal ANC concentrations. The apparent decrease in ANC overtime as indicated in the Bonferroni figures was unexpected both because this is in contrast to the Julian date coefficients for ANC and pH has an increasing trend overtime. Because the Bonferroni method calculates significant means for the water quality vector the difference between the figures and the trends suggests error in the trend analysis models.

The focus of this chapter was to investigate the decline in sulfur dioxide emissions from the Kingston and Bull run power plants and how it may impact the decline in  $\text{SO}_4$  concentrations in the through fall measurements of the Noland divide high elevation

site. If a correlation existed it would be apparent in the pH results, but especially in the  $\text{SO}_4$  results. A significant negative difference in means for set 3 compared to sets 1 and 2 would support the hypothesis. The inequalities present between the sets of pH data display a possible connection to the decline in sulfur dioxide pollution but there are many other factors that effect stream pH. The comparisons between the  $\text{SO}_4$  sets are unfortunately mostly equal. This suggests a bank of  $\text{SO}_4$  with a more steady output evidenced in the measured stream concentrations versus the quickly declining emissions measured as proposed input. The pollution is taken in by the forest and slowly released into the streams. More time may be needed to see affects of the scrubbers.



# Chapter 4

## Power Analysis

### 4.1 Methods

#### 4.1.1 Introduction

The statistics performed in [chapter 2](#) come with an inherent amount of error. This error in trend analysis is defined as type II or  $\beta$  and can be seen in [Table 4.1](#). The

**Table 4.1:** Hypothesis tests from Statistical Methods in Water Resources by theUSGS ([Helsel and Hirsch, 1992](#)).

	$H_0$ is true	$H_0$ is false
Fail to Reject $H_0$	Correct decision Prob(correct decision) $= 1 - \alpha$	Type II error Prob(Type II error) $= \beta$
Reject $H_0$	Type I error Prob (Type I error) $= \alpha$ <b>Significance level</b>	Correct decision Prob (correct decision) $= 1 - \beta$ <b>Power</b>

trend analysis tests the hypothesis that a trend exists in the data. The null hypothesis is that there is no trend or the coefficient is zero. In simpler terms this means that the hypothesis for the trend analysis is that there is a positive julian date coefficient for each dataset and the null hypothesis is that there is not one.  $\beta$  describes the failure

to reject a false null hypothesis or in the case of this paper a failure to detect a trend in the data when there really is one. The opposite of  $\beta$  is the probability that a trend will be detected when it exists and is called the power of the test. A trend line with a power of 1.00 means that there is a 100% chance that the calculated slope is not zero. The post hoc power of all 144 regression lines from [chapter 2](#) were calculated with the help of the statistical program G\*power. An a priori power analysis can be used to help plan for the future by choosing a desired power and planning around it. This analysis was also completed in G\*power.

### 4.1.2 Body

The objectives of the power analysis are two fold. It will calculate the power of each trend line and also give a more efficient sample size for desired powers. The statistical program G\*power requires four inputs total. Three of these inputs are passed from the output of the trend analysis: number of observations (N), adjusted  $r^2$  and number of predictors. The fourth input is ES or effect size which is calculated by G\*power before the analysis. G\*power is a free power analysis program written by four german psychology professors. It can compute both post hoc and a priori analysis for many different statistical tests. A post hoc analysis was performed for both julian date coefficient tables from [chapter 2](#), [Table F.1](#) and [Table F.2](#). In contrast to the post hoc analysis the a priori analysis only needs to be calculated for each of the four dependent variables. This output then has to be manipulated in excel in order to determine a number of sites per elevation bands.

### 4.1.3 Procedures

#### Post hoc

The most popular power analysis methods originate from Jacob Cohen who outlined his approach in "A Power Primer" ([Cohen, 1992a](#)). Cohen displayed ways to calculate the power for eight different tests the last of which is the F test for multiple and

multiple partial correlation. The different tests are represented by their differences in calculating ES. Effect size is the only input that needs to be calculated before the analysis can be completed. The equation for the ES of a regression model presented by Cohen is equal to the correlation coefficient divided by one minus the correlation coefficient. This equation can be described as the ratio of explained to unexplained variation for the regression model. For the post hoc analysis this equation will be used to calculate an specific ES for each model presented in [Table F.1](#) and [Table F.2](#) which correspond to the models in [chapter 2](#). This calculation is completed in G\*power after inputing the correlation coefficient. G\*power uses ES along with the  $\alpha$  used for the regression model, the number observations, and number of predictors in the model to output the power of the F test. This power will be between 0 and 1.00 and will be the power acquired by the models using past data and a calculated ES.

## A priori

The a priori analysis is more conditional than the straight forward calculations for the post hoc analysis. Instead of outputting a power value like the post hoc analysis, G\*power will output a value for the number of observations. The inputs for this analysis are  $\alpha$ , desired power, number of predictors, and ES. All of these inputs can be changed or manipulated based on the desired outcome. For this analysis the assumption was made that the same trend analysis as the one completed in [chapter 2](#) would be attempted in the future. Based on this assumption the same step-wise equations constructed in [Table 2.1](#) can be used to help chose the number of predictors and  $\alpha$ .

The most encompassing way to present an a priori analysis is through a power graph. The power graphs plot power on the y-axis and number of observations on the x-axis. Using this as a tool a planner can choose a desired power and get the corresponding number of observations.

Choosing an ES value and desired power will be a matter of convention. To make choosing the ES value easier Cohen has stated small, medium, and large ES values for each of the eight tests described in [Cohen \(1992a\)](#). Concerning the multiple and

multiple partial correlation test he decided on .02, .15, and .35 respectively. All of these ES values can be graphed in the power graphs by plotting different ES values as curves on the same plot. Even though all powers and all ES values can be plotted at once, some times it is useful to choose specific values. An ES value of .15 was settled upon after the power graphs for all three conventions per dependent variable were made. .02 was too small, requiring very high values for numbers of observations to reach a decent power. ES values of .35 can acquire small numbers for numbers of observations thus achieving a decent power level easier. But .15 is less than half of .35 so it minimizes the chances for insignificant results and the numbers of observations are reasonable to reach higher powers. If no argument can be made for any other desired power then Cohen suggests .80. This is chosen for its reasonable ratio of Type I error to Type II error which reflects their importance. If the power is .80 then  $\beta = .20$  and  $\alpha = .05$  and this makes the Type II error four times as likely as Type I error (Cohen, 1992b). These choices are presented in Table 4.2.

**Table 4.2:** A priori calculation in G\*power when alpha, ES, and power are set to .05, .15, and .80 respectively.

	Number of predictors	$N_a$
pH	6	98
ANC	8	109
Nitrate	8	109
Sulfate	7	103
Time	3	77

The a priori power analysis can be manipulated to calculate a number of sites per elevation band for the stream survey in the GRSM. First, samples per year per elevation band are counted for the 2012 year and will be represented by n. Next the results from Table 4.2 are divided by samples per year per elevation band to get the

number of years it will take at the 2012 sampling rate to reach a power of .80.

$$yrs. = \frac{N_a}{n} \quad (4.1)$$

But, in order to get to the number of sites per elevation band required to reach a power of .80, the years will have to be held constant. If the future trend analysis is to be completed using the equation with only time variables then 77 samples will need to be collected in one year to reach a power of .80 according to [Table 4.2](#). But if the future trend analysis is to be completed using the step-wise equations from [Table 2.1](#) then at least 109 samples will need to be collected in one year to satisfy the requirements for ANC and NO<sub>3</sub>. For the step-wise equations N will be rounded up to 110 and labeled  $N_b$ . These are presented in [Table 4.3](#), which has been calculated

**Table 4.3:** samples/year to achieve a power .80 ( $N_b$ )

Years	1	2	3	4
Water Quality Variables	110	55	37	28
Time Variables	77	39	26	19

out to four years. So that instead of completing the trend analysis after one year, one could wait four years and only need to collect 28 samples per year. Subtracting the number of samples collected in one year per elevation band in 2012 from the number of samples needed to be collected per year to reach a power of .80 will provide the number of samples needed per elevation band to receive a power of .80 ( $N_c$ ).

$$N_c = N_b - n \quad (4.2)$$

To get an estimation for the number of sites needed per elevation band to achieve a power of .80, the number of samples needed per elevation band to receive a power of

.80 ( $N_c$ ) were divided by six which is number of times each site is sampled per year.

$$\#Sites = \frac{N_c}{6} \quad (4.3)$$

## 4.2 Results

### 4.2.1 Post hoc

The results of the post hoc analysis on both trend analyses are reported in [Table F.1](#) and [Table F.2](#). The tables of results are broken into the four analyzed water quality variables (pH, ANC, NO<sub>3</sub>, SO<sub>4</sub>) and divided into the three time sets (93-02, 03-08, 09-12), and then further divided into the six elevation classes. Each regression line computed in [chapter 2](#) are represented by their number of observations, the adjusted  $r^2$ , the calculated ES, and finally their actual power. Of the 72 lines evaluated for power in [Table F.1](#) only eight of them were less than 1.00. Only two of the trend lines in [Table F.1](#) were insignificant. The trend line analyzed in Nitrate set 3 class 5 was insignificant but the adjusted  $r^2$  was negative and therefore the power could not be found. The trend line in pH set 3 class 5 was also insignificant and received the lowest actual power of .28. Of the 72 lines evaluated for power in [Table F.2](#) 52 of the them were insignificant. So already before the post hoc analysis is completed the trend analysis says the regression lines might not even have a trend, and power is the probability of finding a trend if there is one. By convention insignificant trends are ignored and their powers will be ignored here as well. 20 of the trend lines are significant and their powers range from .26 to 1.00. 11 of the 20 significant trend lines are above a power of .80 and 2 are .99 or greater.

## 4.2.2 A priori

### Power graphs

The traditional presentation for an a priori power analysis is the power graph. This graph plots power on the y-axis and number of observations on the x-axis and the curve will start at (0,0) and asymptotically approach a power of 1.00. Each plotted curve represents an ES of either .15 or .35, .02 is not represented because it was unreasonable. Four graphs were drawn for the five dependent variables (pH, ANC, NO<sub>3</sub>, SO<sub>4</sub>, and Time) The only real variation between the graphs are the number of predictors which are counted from [Table 2.1](#). The dependent water quality dependents ANC and NO<sub>3</sub> are represented by the same graph because the ANC and NO<sub>3</sub> equations contain the same number of predictors. All of the power graphs created by the a priori analysis have similar shapes. The curves are all the same, but their placement on the graphs are different. The power graphs for pH, ANC and NO<sub>3</sub>, SO<sub>4</sub>, and Time are plotted in the appendices [Figure G.1](#), [Figure G.2](#), [Figure G.3](#), and [Figure G.4](#) respectively. [Table 4.4](#) was created for easier comparison. This table

**Table 4.4:** Sample sizes at a power of .80

ES	0.15	0.35
pH	97	45
ANC and NO3	98	51
SO4	103	48
Time	76	35

shows the sample size values for both ES curves at a power of .80. Again all are similar except for the time graph, which has at least half as many predictors in its time trend equation as the others.

### A priori manipulation

In this section a scenario is presented in which the results of the a priori analysis are manipulated to achieve the number of sites required per elevation band to receive a

power of .80. This scenario was followed through with both methods of trend lines.

**Table 4.5:** Years to acheive a power of .80

Elevation Bands	Site #	Current n/yr	pH	ANC NO <sub>3</sub>	SO <sub>4</sub>	Time variables
1	13 ,23, 24, 30, 479	26	3.77	4.19	3.96	2.96
2	4, 311, 268, 480, 310, 483, 147, 148, 484	34	2.88	3.21	3.03	2.26
3	114, 481, 482, 149, 66, 492, 137, 293, 270, 493, 485, 144, 224	62	1.58	1.76	1.66	1.24
4	143, 142, 73, 71	24	4.08	4.54	4.29	3.21
5	74, 221, 251, 233	22	4.45	4.95	4.68	3.50
6	253, 234	12	8.17	9.08	8.58	6.42

Table 4.5 records the six elevation bands along with the site numbers that belong to them. In the column labeled, current n per year, the amount of samples collected per elevation band in the year 2012 are tabulated. Using Equation 4.1 the number of years needed for each variable to reach a power of .80 is calculated. Looking at the table there are 26 samples collected in elevation band one in one year. In order to compute a trend line that receives a power of .80 with pH as the dependent, samples would need to be collected for 3.77 years before the trend line can be computed. The largest number of years is elevation class six for a trend line in ANC or NO<sub>3</sub> which requires 9.08 years. This is because it only contains two sites.

**Table 4.6:** Necessary sites scenario for water quality variables

Elevation Bands	#Samples required				# sites required			
	1 yr	2 yrs	3 yrs	4 yrs	1 yr	2 yrs	3 yrs	4 yrs
1	84	29	11	2	14	5	2	0
2	76	21	3	-7	13	4	0	-1
3	48	-7	-25	-35	8	-1	-4	-6
4	86	31	13	4	14	5	2	1
5	88	33	15	6	15	6	2	1
6	98	43	25	16	16	7	4	3



**Table 4.7:** Necessary sites scenario for time variables

Elevation Bands	#Samples required				# sites required			
	1 yr	2 yrs	3 yrs	4 yrs	1 yr	2 yrs	3 yrs	4 yrs
1	51	13	0	-7	9	2	0	-1
2	43	5	-8	-15	7	1	-1	-2
3	15	-24	-36	-43	3	-4	-6	-7
4	53	15	2	-5	9	2	0	-1
5	55	17	4	-3	9	3	1	0
6	65	27	14	7	11	4	2	1

Tables [Table 4.6](#) and [Table 4.7](#) correspond to the two trend analysis completed in [chapter 2](#). Each table is broken down into two sides, the left side contains number of samples while the right side contains number of sites. Each side is grouped by elevation band and calculated out to four years. [Equation 4.2](#) is used to calculate the number of samples required and these numbers are divided by six to get the right side of the tables. Each elevation band is evaluated to determine the number of observations required to achieve a power of .80 with an ES of .15 for a trend analysis using the same equations as [Table 2.1](#). The different columns tell how many years will be needed before the trend analysis can be completed. In [Table 4.6](#) for elevation class 3, 48 more samples need to be collected if a trend line for the water quality dependents with a power of .80 is to be created after one year. But if a trend line can wait to be created after two years, then there is a surplus of seven samples per year. If four years can be waited there is a surplus of 35 samples which on the right side of the table translates into a surplus of 6 whole site locations per year. [Table 4.7](#) works the same way as [Table 4.6](#) but of course it uses different variables for the trend lines.

## 4.3 Discussion

### 4.3.1 Post hoc

#### Step-wise equations

By reviewing the results of the post hoc analysis after an a priori analysis has been completed using the same regression equations it is easier to see why the results were outstanding for the step-wise equations and awful for the time based equations. Knowing that an a priori analysis on the step-wise equations will give the number of observations for a power of .80 and an ES of .15 to be about 110 from [Table 4.2](#). It can easily be seen that as the number of observations in [Table F.1](#) deviate from 110 the power declines. And because the ES is a ratio of the adjusted  $r^2$  it declines with the as the  $r^2$  does. But for a calculated ES of .15 the adjusted  $r^2$  doesn't need to be very high. The analyzed trend line for pH in time set 3 elevation class 5 has an adjusted  $r^2$  of .158 and the ES is .19. The large conventional ES given by Cohen is .35 and only 3 of the trend lines analyzed here were below that, all in pH. Assuming that a power of .80 and an ES of .15 is ideal, then this post hoc analysis uses to many observations. One way to have less observations would be to use fewer years in the analysis. Another way would be to use less sites in the survey.

#### Time variable based equations

The differences in powers between the two post hoc analysis can not be the number of observations because the number of observations used in [Table F.1](#) are the same as those [Table F.2](#). The differences are between the adjusted  $r^2$ , which are very low for [Table F.2](#), and leads to the low ES values. Overlooking the fact that most of the regression models for the time variable analysis are insignificant, most of the powers calculated in [Table F.1](#) are not terrible. Of the 20 significant lines eleven have a power equal to or above .80.

### 4.3.2 A priori

The proper results for the a priori power analysis are [Table 4.2](#) and the [power graphs](#). The power graphs themselves show every possible power and the number of observations needed to achieve it. But they are based on the specific step-wise equations that were created using this specific dataset. Because the step-wise process uses the data to create the equations, every time new data is added the equations could change. The a priori analysis assumes that these same equations, with the same number of variables, will be used to detect trends in the future. But if data is added and the step-wise process is used again then the equations may change. A more static set of equations would ease this pressure.

These power graphs can still be used by managers and planners as an educated guess. After the number of observations for a desired power is determined from the graphs the observations can be placed into the survey in a logical manner in order for better efficiency. Each chosen power and ES value can represent a different scenario. One such scenario was carried out for a power of .80 and an ES of .15. Although any value in the power graphs can be chosen these values were chosen as the most efficient.

The results of this scenario can solve two concerns of the survey: the always present money problem, and the more scientific problem of a lack of high elevation sites. The greatest results presented in [Table 4.6](#), which is the table corresponding to the step-wise equations, are those results in the last column. This column states that if the trend analysis is completed after four years of collecting data then two sites can be removed from the survey. And assuming that cost of the survey is related to the number of sites, then removing sites will save money. But removing two sites is just the outcome of a redistribution suggested by the scenario. In fact one site should be removed from elevation class two and six from class three. One site each need to be added to classes five and six and three should be added to class six. There are too

many sites in the lower elevation classes of two and three and not enough sites in the higher elevation classes of four, five, and six.

# Bibliography

# Bibliography

- Board, E. S. et al. (1983). *Acid Deposition: Atmospheric Processes in Eastern North America*. National Academies Press. [2](#)
- Cai, M., Schwartz, J. S., Robinson, R. B., Moore, S. E., and Kulp, M. A. (2011). Long-term annual and seasonal patterns of acidic deposition and stream water quality in a great smoky mountains high-elevation watershed. *Water, Air, & Soil Pollution*, 219(1-4):547–562. [22](#)
- Cohen, J. (1992a). A power primer. *Psychological bulletin*, 112(1):155. [30](#), [31](#)
- Cohen, J. (1992b). Statistical power analysis. *Current directions in psychological science*, 1(3):98–101. [32](#)
- Helsel, D. R. and Hirsch, R. M. (1992). *Statistical methods in water resources*, volume 49. Elsevier. [viii](#), [15](#), [22](#), [29](#)
- IBM (2013). Ibm spss statistics base 22. [24](#)
- Johnson, D. W., Lindberg, S. E., et al. (1992). *Atmospheric deposition and forest nutrient cycling. A synthesis of the Integrated Forest Study*. Springer-Verlag. [2](#)
- Meijun Cai, J. S. S. (2013). Biological effects of stream water quality on aquatic macroinvertebrates and fish communities within great smoky mountains national park. [4](#)

- Neff, K. J., Schwartz, J. S., Henry, T. B., Robinson, R. B., Moore, S. E., and Kulp, M. A. (2009). Physiological stress in native southern brook trout during episodic stream acidification in the great smoky mountains national park. *Archives of environmental contamination and toxicology*, 57(2):366–376. 2, 10
- Neff, K. J., Schwartz, J. S., Moore, S. E., and Kulp, M. A. (2012). Influence of basin characteristics on baseflow and stormflow chemistry in the great smoky mountains national park, usa. *Hydrological Processes*. 9
- NPS (2013). Nature & science. <http://www.nps.gov/grsm/naturescience/index.htm>. Accessed: 2014-01-05. 1
- Odom, K. R. (2003). *Assessment and redesign of the synoptic water quality monitoring network in the great smoky mountains national park*. PhD thesis. 4
- Robinson, R. B., Barnett, T. W., Harwell, G. R., Moore, S. E., Kulp, M., and Schwartz, J. S. (2008). ph and acid anion time trends in different elevation ranges in the great smoky mountains national park. *Journal of Environmental Engineering*, 134(9):800–808. 2, 7, 8, 17, 20, 21, 22
- UTK (2012). 2011 water quality annual report. Technical report. x, 24
- Weathers, K. C., Simkin, S. M., Lovett, G. M., and Lindberg, S. E. (2006). Empirical modeling of atmospheric deposition in mountainous landscapes. *Ecological Applications*, 16(4):1590–1607. x, 6, 21

# Appendix



# Appendix A

## A.1 Site Data

## A.2 Site data

	Site ID	Site Description	Watershed
1	173	Mill Creek above Abrams Creek	Abrams
2	174	Abrams Creek below Cades Cove	Abrams
3	488	Mill Creek at Pumphouse on Forge Creek Road	Abrams
4	489	Abrams Creek 300 m below trailhead bridge	Abrams
5	142	Beech Creek above Lost Bottom Creek	Cataloochee
6	143	Lost Bottom Creek (Cataloochee Creek)	Cataloochee
7	144	Palmer Creek above Pretty Hollow Creek	Cataloochee
8	147	Lower Cataloochee Creek	Cataloochee
9	148	Lower Little Cataloochee Creek	Cataloochee
10	149	Middle Cataloochee Creek at bridge	Cataloochee
11	293	Rough Fork at Caldwell House	Cataloochee
12	493	Palmer Creek at Davidson Branch Trail	Cataloochee
13	4	Lower Rock Creek	Cosby
14	114	Cosby Creek at log bridge	Cosby
15	137	Upper Rock Creek (Cosby Creek)	Cosby
16	492	Camel Hump Creek off Low Gap Trail	Cosby
17	221	Hazel Creek above cascades	Hazel
18	224	Hazel Creek just below Proctor Creek Confluence	Hazel
19	310	Bone Valley Creek (Hazel Creek)	Hazel
20	311	Hazel Creek below Haw Gap Creek	Hazel
21	479	Hazel Creek at Campsite 86	Hazel
22	480	Haw Gap Creek at bridge near Campsite 84	Hazel
23	481	Little Fork above Sugar Fork Trail	Hazel
24	482	Sugar Fork above Little Fork	Hazel
25	483	Sugar Fork above Haw Gap Creek	Hazel
26	484	Hazel Creek at Cold Spring Gap Trail	Hazel
27	485	Walker Creek above Hazel Creek Trail	Hazel
28	13	Little River at boundary	Little
29	23	Lower Middle Prong Little River	Little
30	24	Lower West Prong Little River	Little
31	30	West Prong Little Pigeon at Headquarters	Little
32	66	West Prong Little Pigeon at Chimneys Picnic Area	Little
33	71	Road Prong above barrier cascade	Little
34	73	Walker Camp Prong above Road Prong	Little
35	74	Walker Camp Prong above Alum Cave Creek	Little
36	233	Walker Camp Prong above Alum Cave	Little
37	234	Upper Road Prong	Little
38	237	Walker Camp Prong at last bridge	Little
39	251	Beech Flats above US 441 loop	Oconaluftee
40	252	Beech Flats below roadcut	Oconaluftee
41	253	Beech Flats above roadcut	Oconaluftee
42	268	Oconaluftee River below Smokemont	Oconaluftee
43	270	Beech Flats at Kephart Footbridge	Oconaluftee

**Table A.1:** GRSM Stream Survey site descriptions

	Site ID	Elevation (ft)	Elevation (m)	slope	Latitude	Longitude	Historical Elevation Classes	New elevation classes
1	173	1715	522.73	35.68	35.59104	-83.85361	3	3
2	174	1715	522.73	10.27	35.59186	-83.85308	3	3
3	488	1790	545.59	4.04	35.58349	-83.83446	4	1
4	489	1710	521.21	32.78	35.59145	-83.85397	4	1
5	142	3300	1005.84	32.42	35.63565	-83.14537	5	2
6	143	3280	999.74	35.69	35.63625	-83.14481	6	2
7	144	2990	911.35	35.66	35.63900	-83.13078	5	2
8	147	2460	749.81	16.84	35.66688	-83.07277	4	3
9	148	2475	754.38	7.58	35.66913	-83.07283	4	3
10	149	2550	777.24	4.45	35.64627	-83.07554	5	3
11	293	2755	839.72	18.73	35.62442	-83.11391	5	4
12	493	2840	865.63	33.10	35.63462	-83.11943	6	6
13	4	2080	633.98	6.11	35.76133	-83.21044	3	1
14	114	2510	765.05	13.71	35.74863	-83.20066	5	2
15	137	2750	838.20	22.92	35.74616	-83.21630	5	2
16	492	2730	832.10	25.86	35.74457	-83.19876	5	6
17	221	4000	1219.20	30.02	35.54632	-83.58283	8	3
18	224	2999	914.00	17.92	35.53212	-83.62234	6	3
19	310	2240	682.75	19.63	35.49994	-83.68014	4	4
20	311	2155	656.84	26.20	35.49377	-83.68852	4	5
21	479	1740	530.35	39.70	35.47233	-83.71933	3	5
22	480	2201	671.00	10.07	35.49474	-83.68873	4	5
23	481	2540	774.19	30.90	35.50256	-83.70835	5	5
24	482	2540	774.19	38.66	35.50236	-83.70859	5	6
25	483	2320	707.14	34.29	35.49947	-83.69494	4	6
26	484	2475	754.38	9.11	35.50331	-83.65930	5	1
27	485	2860	871.73	5.17	35.52249	-83.63101	6	1
28	13	1100	335.28	44.21	35.66763	-83.71450	2	1
29	23	1150	350.52	5.96	35.65724	-83.70979	2	1
30	24	1150	350.52	31.60	35.65682	-83.71017	2	1
31	30	1430	435.86	2.17	35.68819	-83.53672	2	1
32	66	2680	816.86	17.92	35.63723	-83.49484	5	2
33	71	3400	1036.32	31.28	35.63440	-83.47032	6	2
34	73	3360	1024.13	28.98	35.63476	-83.46931	6	2
35	74	3820	1164.34	18.07	35.62912	-83.45102	7	2
36	233	4255	1296.92	21.86	35.61830	-83.42718	8	3
37	234	5000	1524.00	23.93	35.60975	-83.45043	10	3
38	237	4520	1377.70	30.21	35.62409	-83.41692	9	3
39	251	4010	1222.25	19.03	35.60226	-83.41533	8	3
40	252	4680	1426.46	33.32	35.60666	-83.43391	9	3
41	253	4760	1450.85	26.42	35.60682	-83.43510	9	3
42	268	2169	661.00	3.31	35.55293	-83.30937	4	4
43	270	2799	853.00	22.92	35.58641	-83.36400	5	4

**Table A.2:** Site Data

# Appendix B

## Descriptive Statistics

Table B.1: Descriptive statistics of Water Quality in the GRSM

Set	Class	pH			ANC meql			Nitrate meql			Sulfate meql						
		N	Minimum	Maximum	Mean	N	Minimum	Maximum	Mean	N	Minimum	Maximum	Mean				
1993-2002	1	327	4.96	7.90	6.57	327	-20.74	1534.47	149.76	275	0.00	49.94	12.04	325	12.32	85.01	36.09
	2	393	5.32	7.00	6.25	392	-7.43	182.95	40.75	377	1.37	73.76	26.62	390	0.00	159.51	51.68
	3	400	4.65	8.24	6.44	398	-19.97	1624.49	158.44	365	0.00	96.13	26.14	391	0.00	262.37	54.00
	4	121	6.18	7.11	6.50	120	24.45	178.00	75.84	105	2.16	28.29	11.90	119	12.34	77.74	25.16
	5	116	6.07	7.05	6.50	116	41.34	162.76	77.06	66	1.23	10.55	4.35	116	7.51	79.98	26.14
	6	110	5.77	7.06	6.41	110	15.64	165.02	68.01	81	1.56	60.46	21.13	110	14.71	61.16	28.35
2003-2008	1	255	5.22	7.95	6.65	255	-37.09	1314.56	173.48	252	0.50	62.75	16.56	261	10.00	93.23	38.85
	2	289	4.83	7.07	6.32	289	-1.88	145.95	42.20	296	0.62	67.12	29.20	298	11.64	152.55	48.19
	3	299	4.65	8.10	6.55	299	-26.45	1591.06	172.82	297	0.13	95.72	27.69	308	10.44	490.01	54.25
	4	119	5.95	7.06	6.58	119	23.36	128.28	69.90	121	1.87	55.67	17.51	123	13.88	61.31	29.04
	5	35	5.98	7.03	6.50	35	36.37	115.80	77.84	30	1.45	26.48	7.59	37	12.18	117.46	30.54
	6	97	5.79	7.05	6.44	97	6.73	130.63	55.68	98	1.09	72.79	24.88	101	10.02	65.53	34.31
2009-2012	1	191	5.42	8.02	6.77	191	-0.02	1377.93	164.72	191	0.22	62.14	16.31	190	14.61	113.83	39.63
	2	212	4.91	7.28	6.47	212	-11.74	174.52	44.45	212	4.43	72.17	30.08	212	13.45	125.36	47.41
	3	228	4.73	7.96	6.68	228	-18.28	1535.69	160.14	228	1.04	72.16	26.23	228	13.59	317.63	58.15
	4	97	6.20	7.08	6.68	97	25.70	107.58	64.13	97	0.54	34.67	18.72	97	19.89	46.66	29.33
	5	29	6.30	7.11	6.77	29	40.10	115.94	73.55	29	0.21	83.68	6.44	29	16.78	109.18	36.16
	6	76	4.24	7.09	6.52	76	-3.92	114.28	46.15	76	0.16	79.04	32.17	76	15.72	63.32	37.05

# Appendix C

## Variable selection

**Table C.1:** List of variables used for step-wise variable selection. X's for variables selected by the step-wise method, O's if variable was added after the step-wise process.

Available Variables	comments	Dependents for step-wise regression			
		pH	ANC	NO <sub>3</sub>	SO <sub>4</sub>
pH	Dependent				
ANC	Dependent			X	X
NO <sub>3</sub>	Dependent	X	X		X
SO <sub>4</sub>	Dependent	X	X	X	
Julian Date			X	X	X
Month					
Year					
Julian Date Days	Seasonality	X			
$\sin(\theta)$	Seasonality	O	X	X	O
$\cos(\theta)$	Seasonality	X	O	X	O
Sum Base Cations			X	X	X
Conductivity			X	X	X
Chloride			X	X	
Elevation (m)					
Slope					
$\log_2$ (ANC)					
$\log_2$ (Base Cations)		X			
Number of predictors		6	8	8	7

# Appendix D

## Julian Date Coefficients

### D.1 Step-wise Method

### D.2 Temporal Variables

**Table D.1:** Time trend results for specific elevation classes using variables from step-wise regression. **Bold** results are insignificant.

Time set	Elevation class	Elevation range m (ft)	Number of sites	Julian date coefficient, eq/L or pH units (model adjusted $r^2$ ) (p-value)			
				pH	ANC	Nitrate	Sulfate
1993-2002	1	304.8-609.6 (1000-2000)	5	0.069	0.007	0.034	-0.096
				0.712	0.985	0.503	0.569
				0.000	0.000	0.000	0.000
	2	609.6-762 (2000-2500)	9	-0.091	-0.036	-0.037	0.019
				0.388	0.603	0.699	0.766
				0.000	0.000	0.000	0.000
	3	762-914.4 (2500-3000)	13	-0.010	0.008	-0.013	0.024
				0.693	0.971	0.359	0.590
				0.000	0.000	0.000	0.000
	4	914.4-1066.8 (3500-3500)	4	0.019	0.015	0.058	0.061
				0.205	0.709	0.410	0.402
				0.000	0.000	0.000	0.000
	5	1066.8-1371.6 (3500-4500)	4	-0.157	-0.082	0.288	-0.133
				0.165	0.760	0.328	0.566
				0.010	0.000	0.000	0.000
	6	1371.6< (4500<)	2	0.218	0.067	-0.011	0.092
				0.505	0.802	0.871	0.716
				0.000	0.000	0.000	0.000
2003-2008	1	304.8-609.6 (1000-2000)	5	0.150	-0.004	0.038	0.039
				0.781	0.996	0.551	0.673
				0.000	0.000	0.000	0.000
	2	609.6-762 (2000-2500)	9	0.275	0.033	0.044	0.044
				0.348	0.779	0.816	0.893
				0.000	0.000	0.000	0.000
	3	762-914.4 (2500-3000)	13	0.156	0.005	0.072	0.034
				0.663	0.996	0.637	0.923
				0.000	0.000	0.000	0.000
	4	914.4-1066.8 (3500-3500)	4	0.249	-0.028	0.092	0.110
				0.400	0.779	0.405	0.343
				0.000	0.000	0.000	0.000
	5	1066.8-1371.6 (3500-4500)	4	0.137	-0.020	0.204	0.135
				0.300	0.739	0.562	0.884
				0.027	0.000	0.001	0.000
	6	1371.6< (4500<)	2	0.359	0.127	0.074	0.161
				0.317	0.812	0.832	0.844
				0.000	0.000	0.000	0.000
2009-2012	1	304.8-609.6 (1000-2000)	5	0.106	-0.002	0.026	-0.052
				0.894	0.989	0.376	0.536
				0.000	0.000	0.000	0.000
	2	609.6-762 (2000-2500)	9	0.218	0.069	0.121	0.039
				0.606	0.862	0.735	0.887
				0.000	0.000	0.000	0.000
	3	762-914.4 (2500-3000)	13	0.056	0.007	0.019	0.050
				0.766	0.997	0.598	0.915
				0.000	0.000	0.000	0.000
	4	914.4-1066.8 (3500-3500)	4	0.413	-0.006	-0.013	-0.068
				0.593	0.772	0.635	0.529
				0.000	0.000	0.000	0.000
	5	1066.8-1371.6 (3500-4500)	4	<b>-0.115</b>	0.901	<b>0.098</b>	0.015
				<b>0.158</b>	0.540	<b>-0.272</b>	0.658
				<b>0.130</b>	0.001	<b>0.975</b>	0.000
	6	1371.6< (4500<)	2	0.289	0.059	0.097	-0.059
				0.286	0.809	0.881	0.861
				0.000	0.000	0.000	0.000



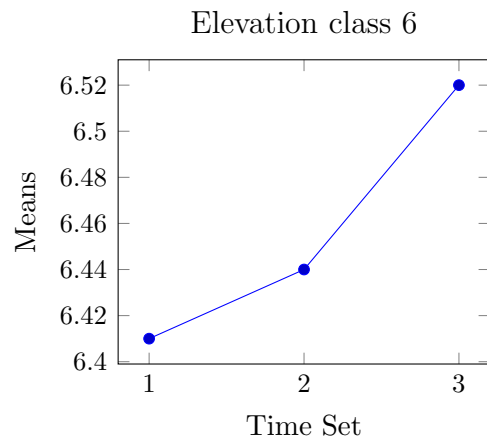
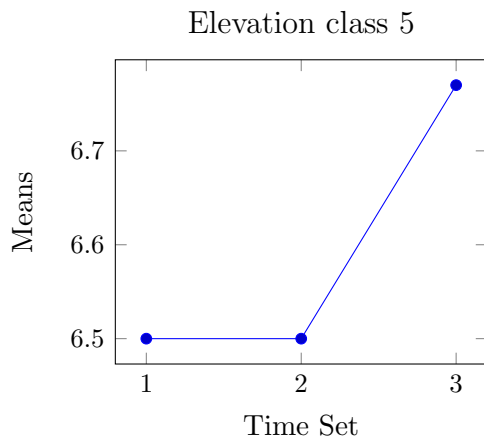
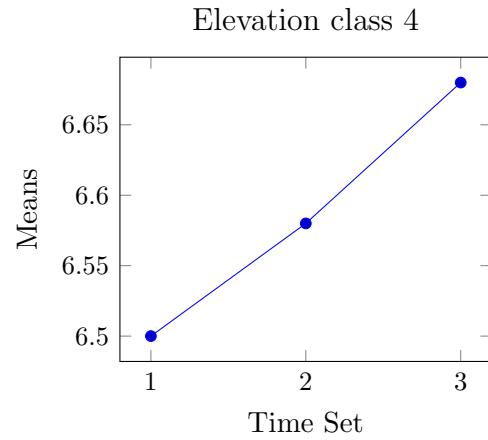
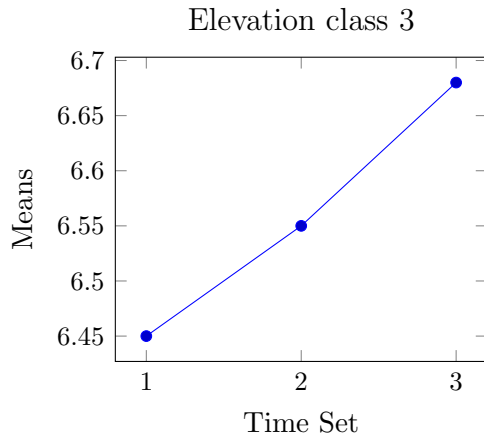
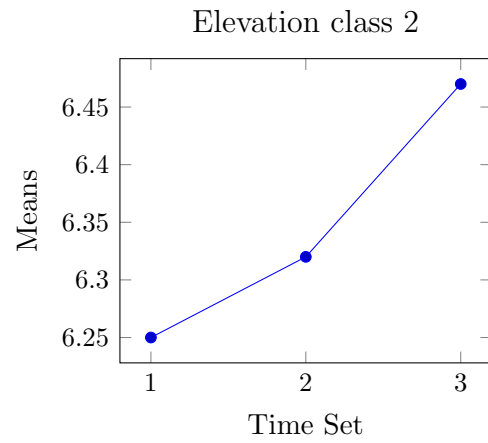
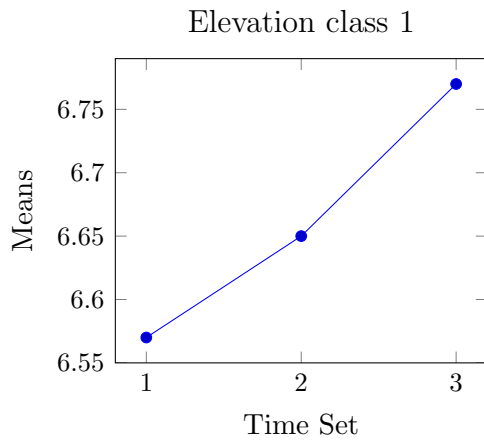
**Table D.2:** Time trend results for specific elevation classes using julian date, cosine( $\theta$ ), and sine( $\theta$ ) only. **Bold** results are insignificant.

Time set	Elevation class	Elevation range m (ft)	Number of sites	Julian date coefficient, eq/L or pH units (model adjusted r <sup>2</sup> ) (p-value)			
				pH	ANC	Nitrate	Sulfate
1993-2002	1	304.8-609.6 (1000-2000)	5	<b>0.054</b>	<b>0.089</b>	-0.138	-0.190
				<b>0.047</b>	<b>0.024</b>	0.016	0.045
				<b>0.321</b>	<b>0.106</b>	0.022	0.001
	2	609.6-762 (2000-2500)	9	<b>-0.090</b>	<b>-0.060</b>	<b>-0.060</b>	<b>-0.075</b>
				<b>0.128</b>	<b>0.189</b>	<b>0.017</b>	<b>0.009</b>
				<b>0.060</b>	<b>0.195</b>	<b>0.248</b>	<b>0.142</b>
	3	762-914.4 (2500-3000)	13	<b>-0.012</b>	<b>-0.030</b>	<b>-0.048</b>	<b>-0.047</b>
				<b>0.013</b>	<b>0.000</b>	<b>-0.004</b>	<b>-0.004</b>
				<b>0.817</b>	<b>0.550</b>	<b>0.365</b>	<b>0.355</b>
	4	914.4-1066.8 (3500-3500)	4	<b>-0.047</b>	<b>-0.151</b>	<b>-0.009</b>	<b>0.095</b>
				<b>0.059</b>	<b>0.294</b>	<b>-0.027</b>	<b>-0.016</b>
				<b>.597</b>	<b>0.055</b>	<b>0.926</b>	<b>0.313</b>
	5	1066.8-1371.6 (3500-4500)	4	<b>-0.151</b>	-0.148	0.330	<b>0.092</b>
				<b>0.051</b>	0.381	0.120	<b>-0.010</b>
				<b>.100</b>	0.047	0.006	<b>0.331</b>
	6	1371.6< (4500<)	2	<b>.156</b>	<b>-0.016</b>	<b>-0.208</b>	<b>-0.036</b>
				<b>.096</b>	<b>0.075</b>	<b>0.092</b>	<b>-0.009</b>
				<b>.092</b>	<b>0.863</b>	<b>0.058</b>	<b>0.707</b>
2003-2008	1	304.8-609.6 (1000-2000)	5	.139	<b>0.009</b>	0.155	0.192
				0.040	<b>0.001</b>	0.061	0.043
				0.025	<b>0.888</b>	0.012	0.002
	2	609.6-762 (2000-2500)	9	0.145	<b>-0.090</b>	0.178	0.138
				0.061	<b>0.081</b>	0.043	0.014
				0.012	<b>0.114</b>	0.002	0.017
	3	762-914.4 (2500-3000)	13	<b>0.103</b>	<b>-0.006</b>	<b>0.047</b>	<b>0.099</b>
				<b>0.020</b>	<b>-0.003</b>	<b>-0.003</b>	<b>0.006</b>
				<b>0.075</b>	<b>0.925</b>	<b>0.418</b>	<b>0.085</b>
	4	914.4-1066.8 (3500-3500)	4	0.235	<b>-0.029</b>	0.193	0.192
				0.148	<b>0.180</b>	0.086	0.023
				0.007	<b>0.728</b>	0.030	0.035
	5	1066.8-1371.6 (3500-4500)	4	<b>0.135</b>	<b>-0.112</b>	<b>-0.176</b>	<b>0.067</b>
				<b>-0.069</b>	<b>0.337</b>	<b>-0.082</b>	<b>-0.024</b>
				<b>0.466</b>	<b>0.443</b>	<b>0.401</b>	<b>0.701</b>
	6	1371.6< (4500<)	2	0.204	<b>-0.108</b>	0.236	0.307
				0.081	<b>0.094</b>	0.046	0.074
				0.041	<b>0.274</b>	0.020	0.002
2009-2012	1	304.8-609.6 (1000-2000)	5	<b>0.111</b>	<b>0.026</b>	<b>-0.036</b>	<b>-0.092</b>
				<b>0.028</b>	<b>0.000</b>	<b>0.018</b>	<b>0.005</b>
				<b>0.122</b>	<b>0.718</b>	<b>0.619</b>	<b>0.207</b>
	2	609.6-762 (2000-2500)	9	0.141	<b>0.017</b>	<b>0.020</b>	<b>-0.062</b>
				0.052	<b>0.056</b>	<b>0.011</b>	<b>-0.010</b>
				0.037	<b>0.800</b>	<b>0.767</b>	<b>0.376</b>
	3	762-914.4 (2500-3000)	13	<b>-0.034</b>	<b>-0.027</b>	<b>-0.036</b>	<b>0.078</b>
				<b>-0.009</b>	<b>-0.002</b>	<b>-0.004</b>	<b>-0.007</b>
				<b>0.611</b>	<b>0.684</b>	<b>0.592</b>	<b>0.246</b>
	4	914.4-1066.8 (3500-3500)	4	0.405	<b>0.032</b>	<b>-0.067</b>	<b>-0.129</b>
				0.200	<b>0.161</b>	<b>-0.016</b>	<b>-0.011</b>
				0.000	<b>0.733</b>	<b>0.518</b>	<b>0.215</b>
	5	1066.8-1371.6 (3500-4500)	4	<b>-0.031</b>	0.891	<b>0.052</b>	<b>-0.414</b>
				<b>0.218</b>	0.466	<b>-0.039</b>	<b>-0.076</b>
				<b>0.934</b>	0.007	<b>0.904</b>	<b>0.347</b>
	6	1371.6< (4500<)	2	0.264	<b>0.083</b>	<b>-0.021</b>	<b>-0.214</b>
				0.039	<b>0.058</b>	<b>-0.016</b>	<b>0.007</b>
				0.023	<b>0.462</b>	<b>0.859</b>	<b>0.068</b>

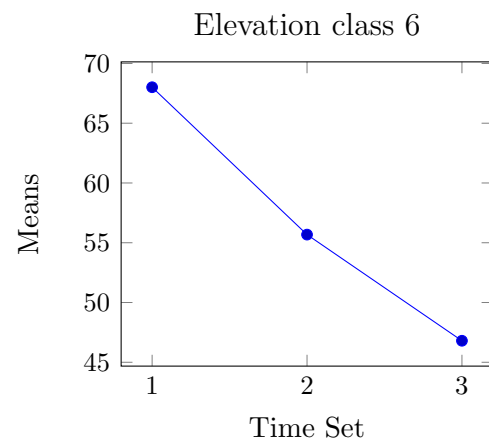
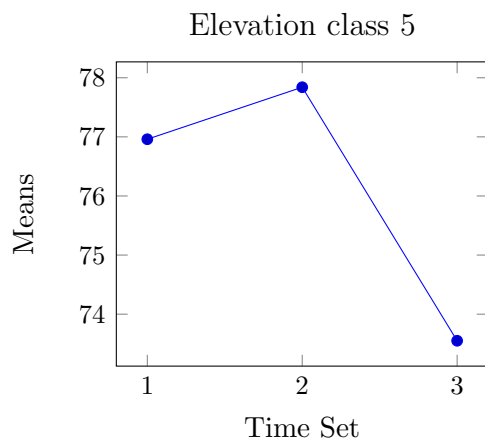
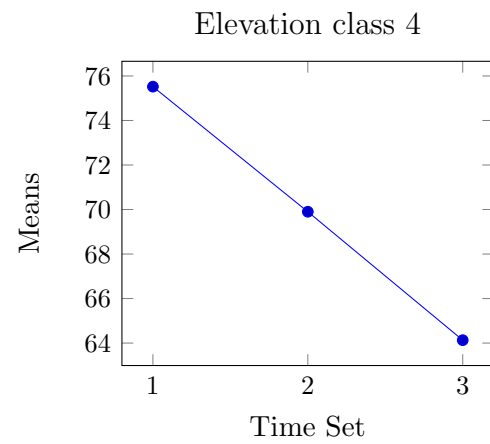
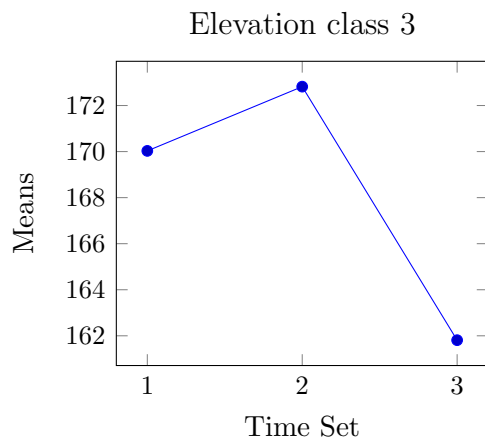
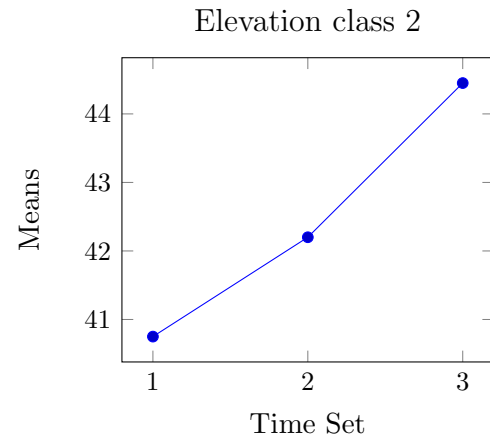
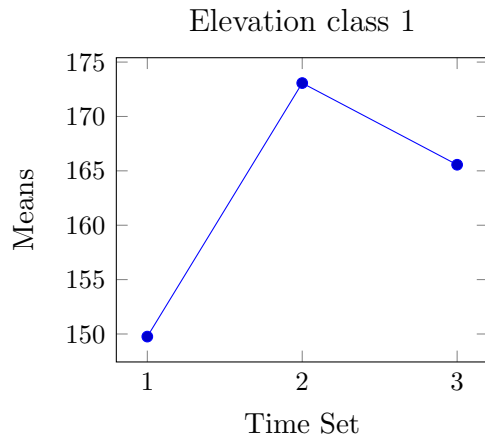
# Appendix E

## ANOVA/Bonferoni

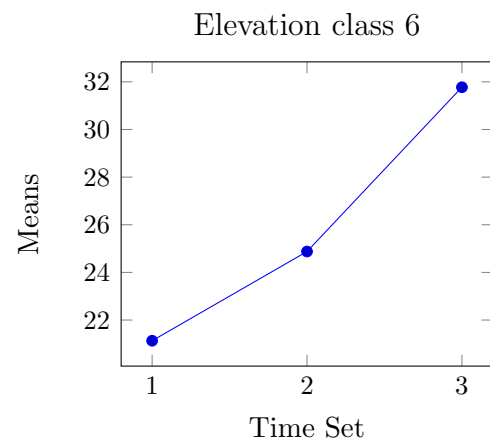
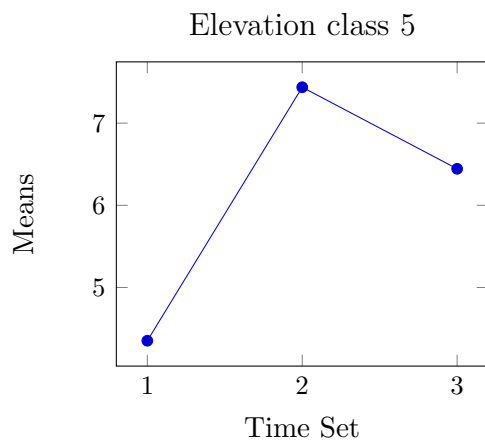
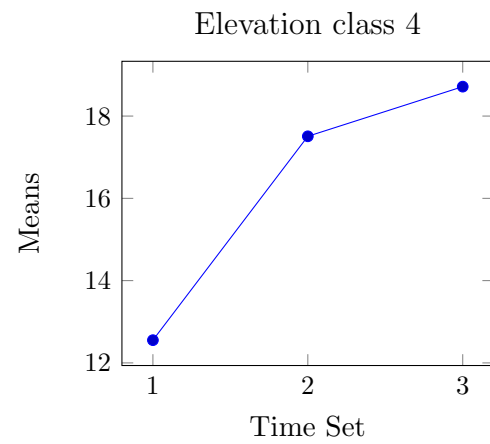
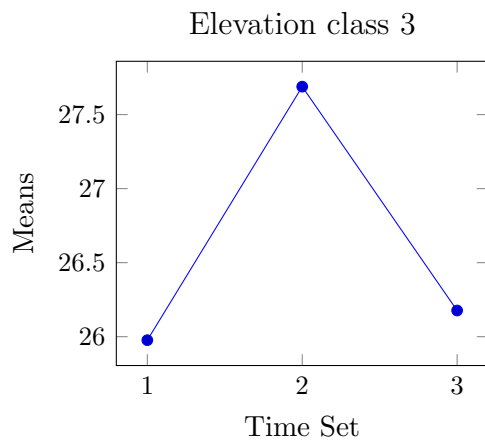
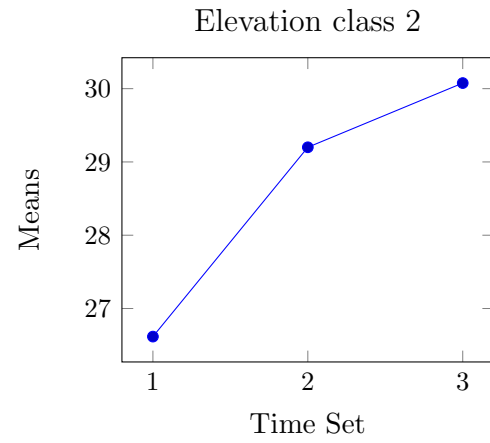
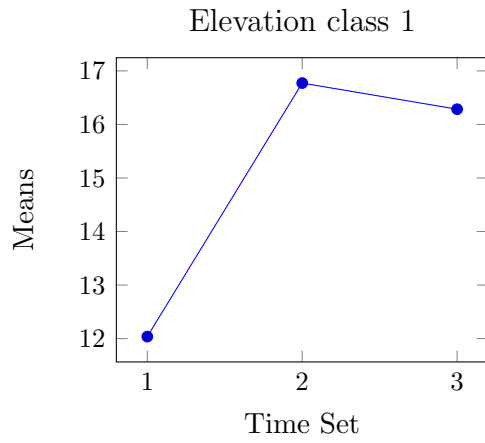
## E.1 pH



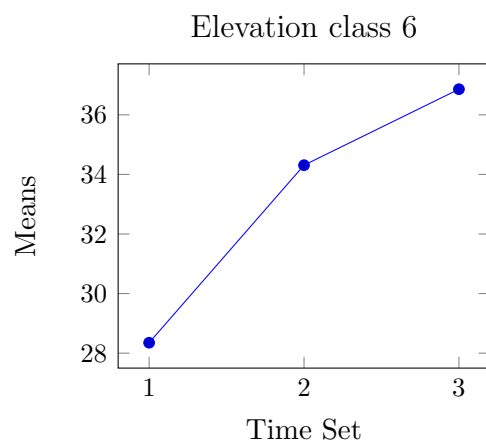
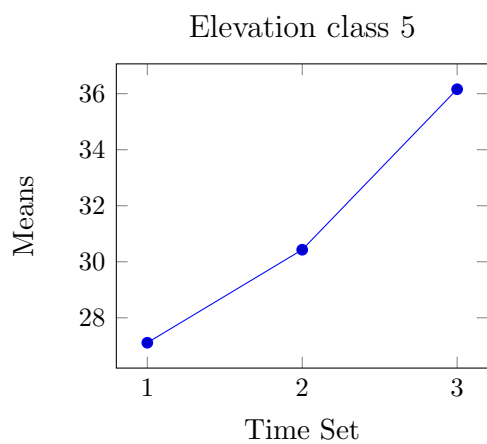
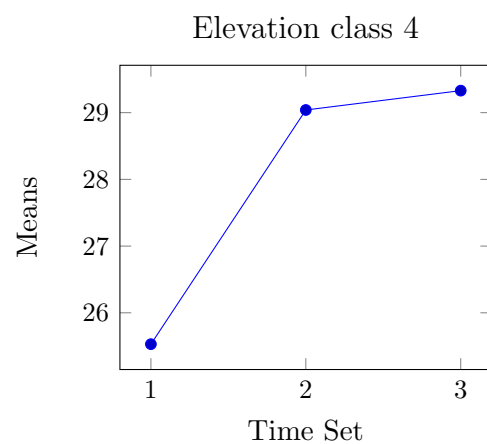
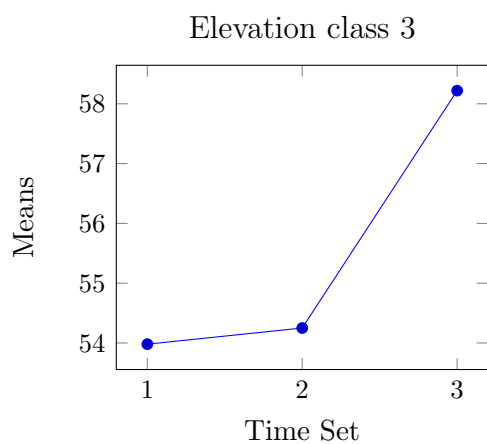
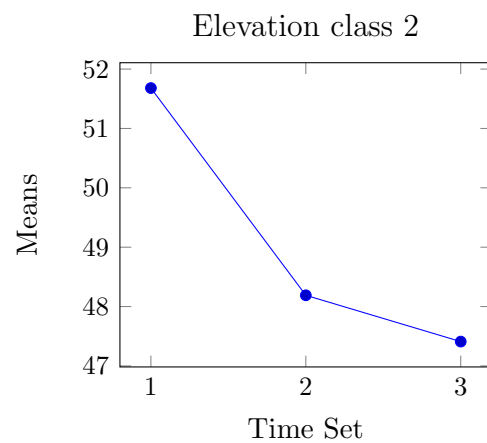
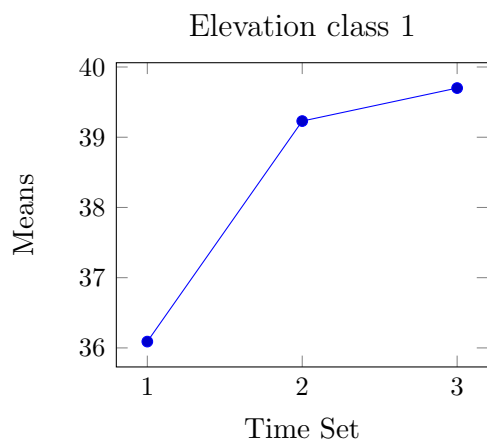
## E.2 ANC



## E.3 Nitrate



## E.4 Sulfate



# Appendix F

## Post Hoc Power Analysis

### F.1 Step-Wise Variables

### F.2 Temperol variables

**Table F.1:** Post hoc power analysis using G\*power and a calculated ES, alpha is .05. **Bold** results are insignificant.

Set	Class	pH			ANCmeqL			NitratemeqL			SulfatemeqL						
		N	Adjusted r <sup>2</sup>	Effect Size	Actual Power	N	Adjusted r <sup>2</sup>	Effect Size	Actual Power	N	Adjusted r <sup>2</sup>	Effect Size	Actual Power				
1993- 2002	1	327	0.712	2.47	1.00	327	0.985	65.67	1.00	275	0.503	1.01	1.00	325	0.569	1.32	1.00
	2	393	0.388	0.63	1.00	392	0.603	1.52	1.00	377	0.699	2.32	1.00	390	0.766	3.27	1.00
	3	400	0.693	2.26	1.00	398	0.971	33.48	1.00	365	0.359	0.56	1.00	391	0.590	1.44	1.00
	4	121	0.205	0.26	0.99	120	0.709	2.44	1.00	105	0.410	0.69	1.00	119	0.402	0.67	1.00
	5	116	0.165	0.20	0.96	116	0.760	3.17	1.00	66	0.328	0.49	0.98	116	0.566	1.30	1.00
	6	110	0.505	1.02	1.00	110	0.802	4.05	1.00	81	0.871	6.75	1.00	110	0.716	2.52	1.00
2003- 2008	1	255	0.781	3.57	1.00	255	0.996	249.00	1.00	252	0.551	1.23	1.00	261	0.673	2.06	1.00
	2	289	0.348	0.53	1.00	289	0.779	3.52	1.00	296	0.816	4.43	1.00	298	0.893	8.35	1.00
	3	299	0.663	1.97	1.00	299	0.996	249.00	1.00	297	0.637	1.75	1.00	308	0.923	11.99	1.00
	4	119	0.400	0.67	1.00	119	0.779	3.52	1.00	121	0.405	0.68	1.00	123	0.343	0.52	1.00
	5	35	0.300	0.43	0.74	35	0.739	2.83	1.00	30	0.562	1.28	0.98	37	0.884	7.62	1.00
	6	97	0.317	0.46	1.00	97	0.812	4.32	1.00	98	0.832	4.95	1.00	101	0.844	5.41	1.00
2009- 2012	1	191	0.894	8.43	1.00	191	0.989	89.91	1.00	191	0.376	0.60	1.00	190	0.536	1.16	1.00
	2	212	0.606	1.54	1.00	212	0.862	6.25	1.00	212	0.735	2.77	1.00	212	0.887	7.85	1.00
	3	228	0.766	3.27	1.00	228	0.997	332.33	1.00	228	0.598	1.49	1.00	228	0.915	10.76	1.00
	4	97	0.593	1.46	1.00	97	0.772	3.39	1.00	97	0.635	1.74	1.00	97	0.529	1.12	1.00
	5	29	<b>0.158</b>	0.19	0.28	29	0.540	1.17	0.96	29	<b>-0.272</b>	NA	NA	29	0.658	1.92	1.00
	6	76	0.286	0.40	0.99	76	0.809	4.24	1.00	76	0.881	7.40	1.00	76	0.861	6.19	1.00



**Table F.2:** Post hoc power analysis using G\*power a calculated ES, an alpha of .05 with the variables:  $\text{sine}(\theta)$ ,  $\text{cosine}(\theta)$ , and julian date only. **Bold** results are insignificant.

Set	Class	pH			ANCmeqL			NitratemeqL			SulfatemeqL						
		N	Adjusted r <sup>2</sup>	Effect Size	Actual Power	N	Adjusted r <sup>2</sup>	Effect Size	Actual Power	N	Adjusted r <sup>2</sup>	Effect Size	Actual Power				
1993- 2002	1	327	<b>0.047</b>	0.049	0.93	327	<b>0.024</b>	0.02	0.65	275	0.016	0.02	0.39	325	0.045	0.05	0.92
	2	393	<b>0.128</b>	0.15	1.00	392	<b>0.189</b>	0.23	1.00	377	<b>0.017</b>	0.02	0.55	390	<b>0.009</b>	0.01	0.32
	3	400	<b>0.013</b>	0.01	0.46	398	<b>0.000</b>	0.00	0.06	365	<b>-0.004</b>	NA	NA	391	<b>-0.004</b>	NA	NA
	4	121	<b>0.059</b>	0.06	0.61	120	<b>0.294</b>	0.42	1.00	105	<b>-0.027</b>	NA	NA	119	<b>-0.016</b>	NA	NA
	5	116	<b>0.051</b>	0.05	0.52	116	0.381	0.62	1.00	66	0.120	0.14	0.68	116	<b>-0.010</b>	NA	NA
	6	110	<b>0.096</b>	0.11	0.81	110	<b>0.075</b>	0.08	0.69	81	<b>0.092</b>	0.10	0.64	110	<b>-0.009</b>	NA	NA
2003- 2008	1	255	0.040	0.04	0.78	255	<b>0.001</b>	0.00	0.07	252	0.061	0.06	0.94	261	0.043	0.04	0.82
	2	289	0.061	0.06	0.96	289	<b>0.081</b>	0.09	0.99	296	0.043	0.04	0.87	298	0.014	0.01	0.37
	3	299	<b>0.020</b>	0.02	0.52	299	<b>-0.003</b>	NA	NA	297	<b>-0.003</b>	NA	NA	308	<b>0.006</b>	0.01	0.18
	4	119	0.148	0.17	0.97	119	<b>0.180</b>	0.22	0.99	121	0.086	0.09	0.80	123	0.023	0.02	0.26
	5	35	<b>-0.069</b>	NA	NA	35	<b>0.337</b>	0.51	0.93	30	<b>-0.082</b>	NA	NA	37	<b>-0.024</b>	NA	NA
	6	97	0.081	0.09	0.67	97	<b>0.094</b>	0.10	0.74	98	0.046	0.05	0.40	101	0.074	0.08	0.64
2009- 2012	1	191	<b>0.028</b>	0.03	0.47	191	<b>0.000</b>	0.00	0.05	191	<b>0.018</b>	0.02	0.31	190	<b>0.005</b>	0.01	0.11
	2	212	0.052	0.05	0.82	212	<b>0.056</b>	0.06	0.85	212	<b>0.011</b>	0.01	0.22	212	<b>-0.010</b>	NA	NA
	3	228	<b>-0.009</b>	NA	NA	228	<b>-0.002</b>	NA	NA	228	<b>-0.004</b>	NA	NA	228	<b>-0.007</b>	NA	NA
	4	97	0.200	0.25	0.99	97	<b>0.161</b>	0.19	0.96	97	<b>-0.016</b>	NA	NA	97	<b>-0.011</b>	NA	NA
	5	29	<b>0.218</b>	0.28	0.58	29	0.466	0.87	0.98	29	<b>-0.039</b>	NA	NA	29	<b>-0.076</b>	NA	NA
	6	76	0.039	0.04	0.27	76	<b>0.058</b>	0.06	0.39	76	<b>-0.016</b>	NA	NA	76	<b>0.007</b>	0.01	0.08

# Appendix G

## A priori analysis

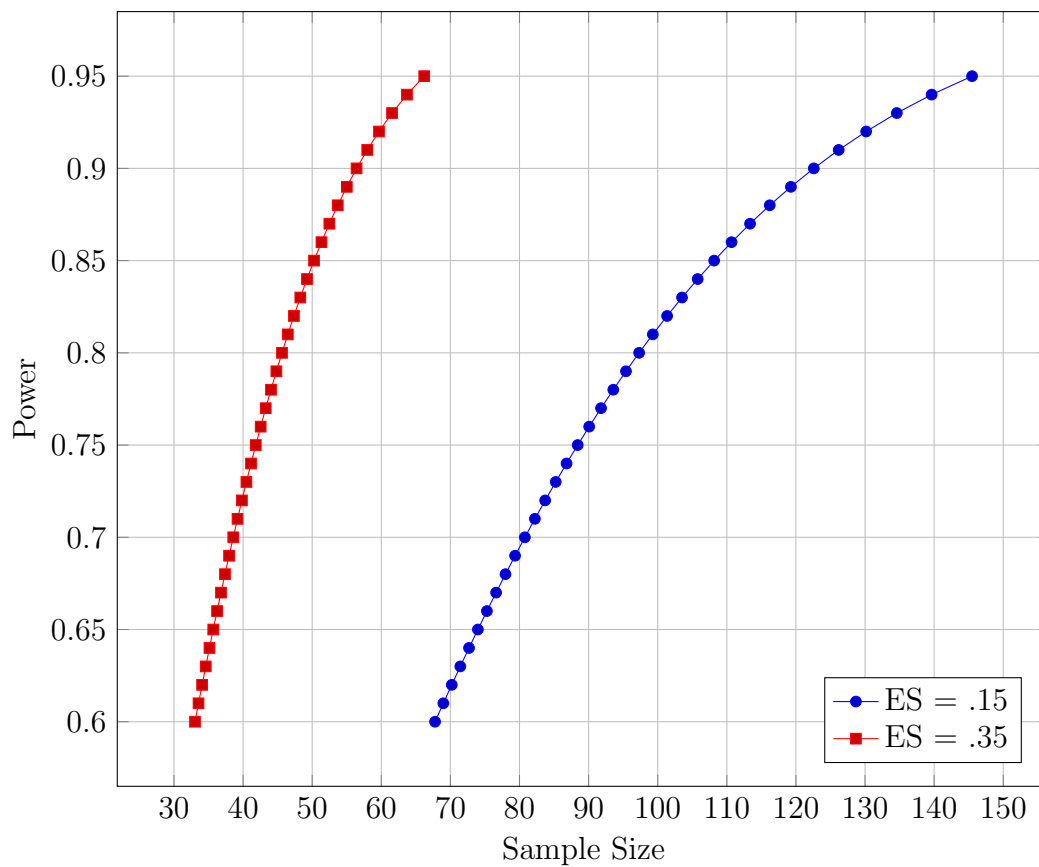
### G.1 Power graphs

#### G.1.1 pH

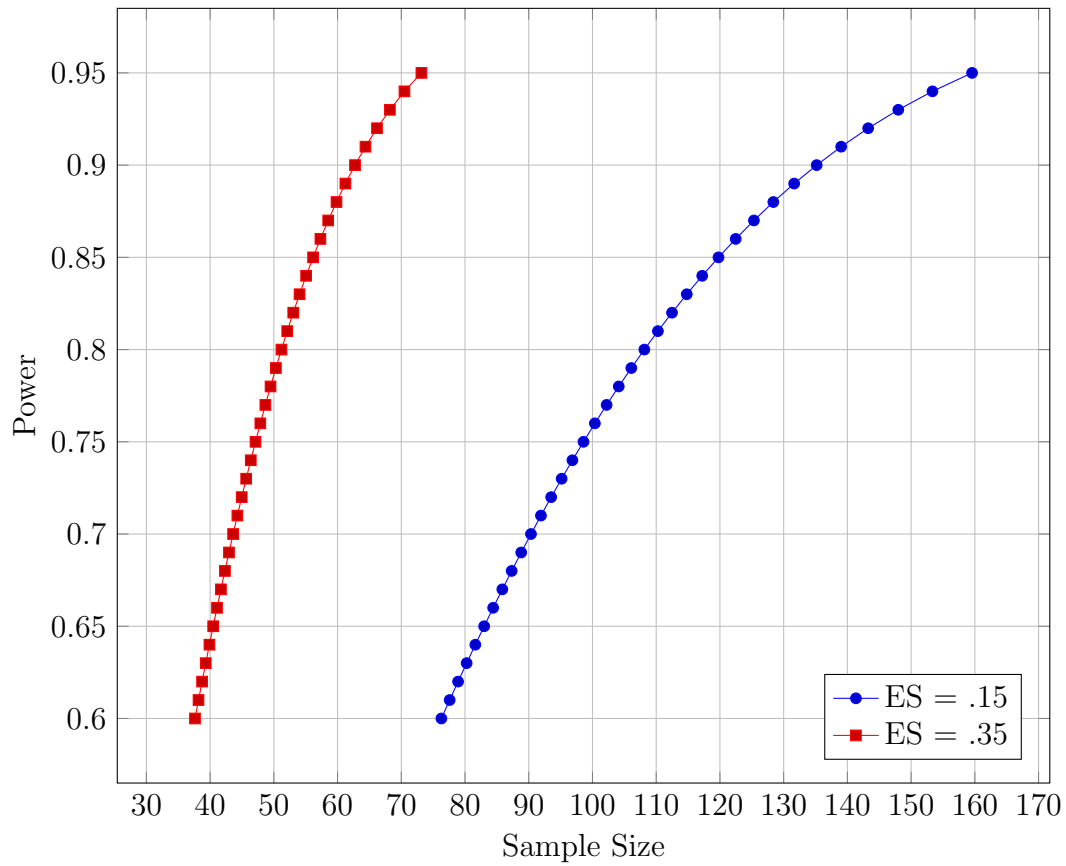
#### G.1.2 ANC and Nitrate

#### G.1.3 Sulfate

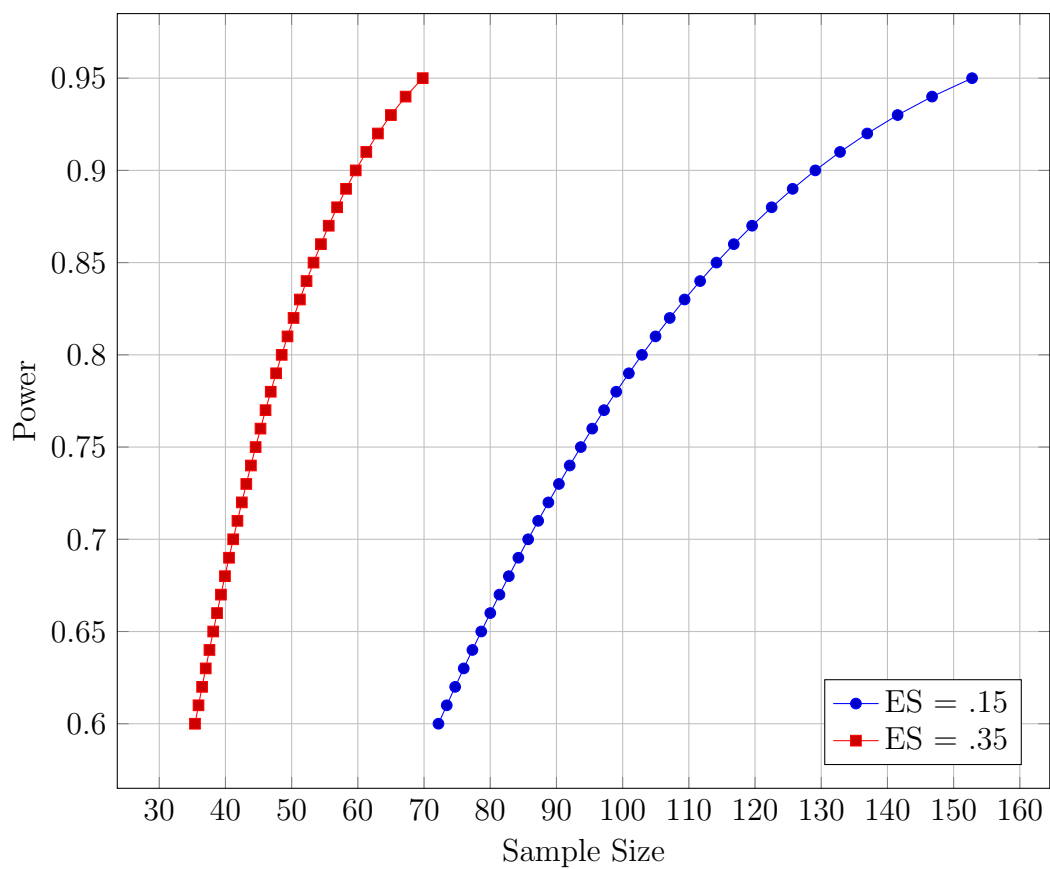
#### G.1.4 Time Variables



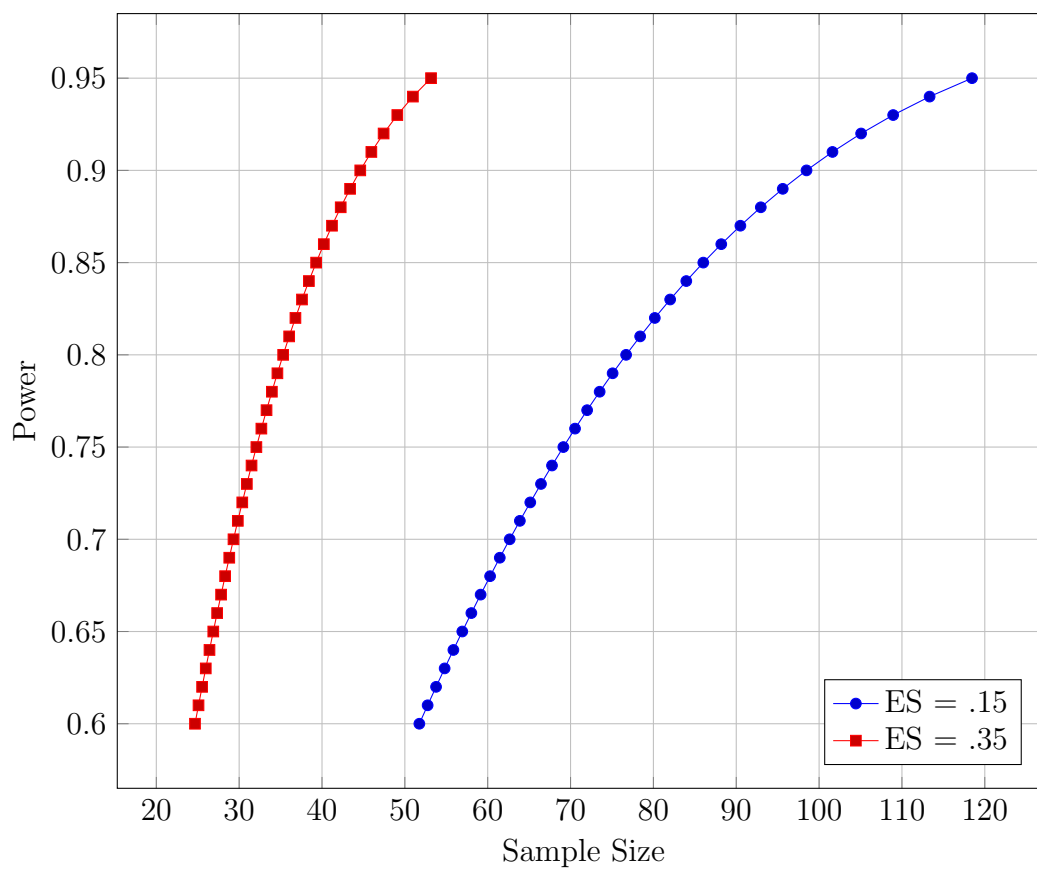
**Figure G.1:** pH Power Graph. The power is shown as a function of pH



**Figure G.2:** ANC and Nitrate Power Graphs. The power graphs for ANC and Nitrate are the same because they both have the same number of predictors.



**Figure G.3:** Sulfate Power Graph



**Figure G.4:** Time Variables Power Graph

# Vita

Tim Pobst was born in Nashville, TN on June 1st 1985 to George and Peggy Pobst. He graduated from Centennial High School near Franklin, TN and was accepted to the University of Tennessee immediately after. He was undecided for three years before deciding to try for a civil engineering degree and he finished it in spring of 2011. He stayed at the University of Tennessee to get a masters degree in environmental engineering under Dr. Schwartz.