University of Tennessee, Knoxville

TRACE: Tennessee Research and Creative Exchange

Masters Theses

Graduate School

12-2009

Characterizing Episodic Stream Acidification Using a Concentration-Duration-Frequency Methodology in Watersheds of the Great Smoky Mountains National Park

John Leland Mauney III
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by John Leland Mauney III entitled "Characterizing Episodic Stream Acidification Using a Concentration-Duration-Frequency Methodology in Watersheds of the Great Smoky Mountains National Park." 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.

John S. Schwartz, Major Professor

We have read this thesis and recommend its acceptance:

R. Bruce Robinson, Glenn A. Tootle

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

To the Graduate Council:

I am submitting herewith a thesis written by John Leland Mauney III entitled “Characterizing Episodic Stream Acidification Using a Concentration-Duration-Frequency Methodology in Watersheds of the Great Smoky Mountains National Park.” 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.

John S. Schwartz, Major Professor

We have read this thesis
and recommend its acceptance:

R. Bruce Robinson

Glenn A. Tootle

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

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

Characterizing Episodic Stream Acidification Using a Concentration-Duration-Frequency Methodology in Watersheds of the Great Smoky Mountains National Park

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

John Leland (Lee) Mauney III
November, 2009

Abstract

Episodic stream acidification occurs as storm events temporarily reduce acid neutralizing capacity and pH. Stream acidification is suspected to have damaging effects on the health of aquatic ecosystems and biota and is dependent on various watershed characteristics such as drainage area, elevation, slope, and surficial geology. Here, a stochastic modeling approach is applied to continuous pH data for multiple stream monitoring sites within the Great Smoky Mountains National Park in order to characterize episodic acidification responses during stormflows for different streams. The approach summarizes voluminous pH data recorded by water quality sondes at 15-minute intervals into concentration-duration-frequency relationships. Unique to this study is the ability to characterize the episodic acidification response to watershed attributes without using baseflow or single-point stormflow measurements. A slope metric of mean pH event duration, a measure of episodic acidification response was determined to correlate with basin area and elevation. In contrast, baseflow studies have shown elevation to be the main driver of chronic acidification. It appears that during stormflows transport and flushing of stored anions and cations govern the response of streams included in this study.

Acknowledgments

I would like to thank Dr. John Schwartz for serving as my major professor and advisor. His knowledge, expertise, and guidance played an essential role in this project and in my graduate education. I would also like to express gratitude to Dr. Glenn Tootle and Dr. Bruce Robinson. Their expert knowledge and experience proved invaluable in this study. I thank them for sitting on my committee and taking time to help with my thesis defense.

I am especially grateful to Mei Cai and Keil Neff for assistance in numerous aspects of my graduate education. Mei provided aid in excel and chemistry. Keil's GIS knowledge was indispensable for this project. In addition, Keil has been a mentor and friend, motivating me in both my studies and my life.

Special thanks to fisheries biologists Steve Moore and Matt Kulp of the Great Smoky Mountain National Park for providing this project, data, and such a beautiful study site to research.

This project could not have been possible without the assistance of fellow University of Tennessee students (Mary Ann Grell, Sam Matthews, Russell Freda, David Jacobs, Chris Rolison, Joe Parker, Will Cantrell, Tom Zimmerman, and Cody Moser).

To my family and friends, thank you for your unyielding encouragement, support, and love. You are a source of happiness and inspiration, and I appreciate all you have done for me.

Table of Contents

List of Figures

List of Tables

Note: This research is being submitted as a manuscript to a scientific journal. Additional research information that was not deemed publishable is located in the appendices for supplementary documentation of findings. Referenced figures and tables are shown in Appendix A.

1 INTRODUCTION

Episodic stream acidification occurs when rainfall and/or snowmelt runoff cause short-term decreases of pH and acid neutralizing capacity (ANC) in surface water, particularly in watersheds with base-poor soils [*Driscoll et al.*, 2001; *Wigington et al.*, 1996b]. Stream acidification is prevalent in the eastern United States where atmospheric acid deposition is relatively high [*Cooper et al.*, 2000; *Driscoll et al.*, 2003; *Sullivan et al.*, 2007]. Anthropogenic sources of sulfur dioxide and nitrogen oxides are emitted from fossil fuel combustion in coal-fired power plants and motorized vehicles [*Herlihy et al.*, 1991; *Wigington et al.*, 1996b]. These acids enter watersheds through wet deposition or when storm events flush ions accumulated by dry deposition into streams, resulting in decreased ANC and pH depressions in streamflow. Air quality has improved during the past two decades and consequent declines in acid deposition have been observed [*NADP*, 2007]. Improvements in air quality can be attributed to implementation of Title IV of the 1990 Clean Air Act Amendments, improved technologies including advanced scrubber systems for power plant exhaust and catalytic converters on motor vehicles, and increased use of low-sulfur coal [*Driscoll et al.*, 2001; *NADP*, 2007; *Stoddard et al.*, 1999]. Despite reductions in deposition, episodic acidification continues to impact stream and forest systems and threatens aquatic biota in northern Europe and eastern United States, specifically the Appalachian highlands [*Deviney et al.*, 2006; *Evans et al.*, 2006; *Wigington et al.*, 1996b].

Both chronic and episodic stream acidification have been studied extensively, and regional water quality surveys and models have been developed [*Baker et al.*, 1996; *Deviney et al.*, 2006; *Herlihy et al.*, 1993; *R. B. Robinson et al.*, 2008; *Tranter et al.*, 1994; *Wigington et al.*, 1996a]. Baseflow and stormflow are related spatially since the same geochemical and physiographic characteristics influence water quality in any particular stream location. Chronic stream acidification studies have shown long-term trends in response from physical watershed characteristics, but acute acidification response can be dependent on different complex biogeochemical processes. Watershed characteristics that influence acidification response are many, and include basin area, topography, and basin geology and soils [*Clow and Sueker*, 2000; *Herlihy et al.*, 1998; *O'Brien et al.*, 1993; *Sullivan et al.*, 2007]. Elevation and elevational gradients impact on acid-base chemistry have been documented in detail [*Johnson et al.*, 1981; *R. B. Robinson et al.*, 2008]. Studies of upland watersheds have shown that catchments at higher elevations with smaller drainage areas are more susceptible to acidification [*Baldigo and Lawrence*, 2001; *Deviney et al.*, 2006; *Hesthagen et al.*, 1999]. Bedrock composition and surficial geology can impact pH response through dictation of buffering capacity and the release of base cations during chemical weathering [*Clow and Sueker*, 2000; *Silsbee and Larson*, 1982]. Streams draining catchments with sili-clastic bedrock are particularly sensitive to acidification, and when exposed produce additional acid inputs [*MacAvoy and Bulger*, 1995]. Understanding how complex watershed-scale processes affect system responses of atmospheric deposition to stream acidification is imperative to evaluate how anthropogenic and natural acid inputs impact stream health.

Empirical relationships between stream chemistry and spatial statistics describing basin land cover composition have been shown [*Clow and Sueker*, 2000; *Sullivan et al.*, 2007], however a technique to quantify episodic response over time as a function of environmental factors, or watershed attributes, remains largely unexplored. The advent of water quality monitoring devices (sondes) has made frequent and voluminous amounts of data possible. Robinson and Roby [2006] successfully demonstrated that data from thousands of stream pH measurements can be summarized by developing concentration-duration-frequency (CDF) curves to provide time-connected durations and frequencies of episodic events. This CDF curve methodology, similar to precipitation intensity-duration-frequency curves, produces a pH event duration slope metric for a specific stream site that characterizes the response to acid episodes. Historically, descriptive statistics of chemical parameters from a limited number of samples have demonstrated important relationships in stream acidification, but have not adequately characterized time-series responses, including those of acid episodes. Concentration frequency histograms have been shown to illuminate less obvious trends in the data, but are limited in that they do not reflect the frequency and intensity of the time-connected durations. CDF methodology provides the foundation to compare intensities, durations, and frequencies of episodic low pH events in order to parameterize the episodic nature of stream acidification spatially, and can provide critical information to understand the impact of stream acidification on aquatic biota. It is valuable to predict the spatial extent of acidification on stream chemistry and forecast episodic extremes associated with acidification.

Understanding deleterious effects of stream acidification on aquatic ecosystems and fish populations is of utmost importance in watersheds impacted by acid deposition, including the Great Smoky Mountains National Park (GRSM) located in the Southern Appalachians. The GRSM is particularly susceptible to episodic storm events because of elevated rates of atmospheric acid deposition and geology with low buffering capacity. Episodic stream acidification is suspected to be a primary cause of native brook trout (*Salvelinus fontinalis*) extirpation in GRSM headwater streams [*Neff et al.*, 2009; *R. B. Robinson et al.*, 2008]. Although chronic acidification occurs in some streams within the GRSM [*R. B. Robinson et al.*, 2008], acute pH depressions and increases in inorganic monomeric aluminum associated with storm events are chiefly toxic to trout [*Wigington et al*., 1996b]. Studies concerning the impacts of these events on fish physiology are somewhat rare because episodic acidification events are complex and site specific [*Van Sickle et al.*, 1996]. Biological consequences of episodic acid events are dependent on the extent of the reduction in pH, duration of the pH event, and frequency associated with successive acid episodes [*Calta*, 2002]. Knowledge of spatial and temporal patterns of episodic stream acidification related to watershed attributes is required to fully understand ecological processes that lead to brook trout extirpation in some watersheds but not in others [*Neff et al.*, 2009].

Using the CDF methodology to characterize unique episodic acidification responses for each stream during stormflows, the objectives of this study are to: (1) relate the CDF metric, specifically the pH event duration slope, to watershed attributes, and (2) explore whether the pH event duration slope is representative of a watershed using multiple time periods with long-term data from a single stream site. Unique to this study is the ability to characterize the episodic acidification response to watershed attributes, not baseflow or single-point stormflow measurements. Severity and response from episodes can vary spatially and seasonally.

2 METHODS

2.1 Study Area

The GRSM (850 square miles) is located in the Blue Ridge physiographic region of the southern Appalachians in eastern Tennessee and western North Carolina (Figure 1). There are more than 3000 kilometers of headwater streams in the GRSM, which support a great number of fish species, amphibians, and benthic invertebrates that must be protected from impairment.

The region is physiographically characterized by rugged topography, heavily forested slopes, with steep mountain streams. Altitudes in the GRSM range from 300 m to 2,025 m. Watershed geology is predominately potassium feldspar sandstone, intermixed with siltstone, shale, and slate. Siliciclastic sulfidic slate (SSS), comprised of Anakeesta and Copperhill bedrock formations, are of particular interest. Anakeesta, a carbonaceous phyllite, is found is various watersheds within the park and is a potentially significant source of acidification for surface waters [*Huckabee et al.*, 1975]. Anakeesta and Copperhill formations are usually non-reactive until exposed to air and water, at which point are oxidized releasing acid and heavy metals. The GRSM geology provides little buffering capacity, with 96% of all monitored stream sites having ANC less than 200 μeq/L; 59% have ANC concentrations less than 50 μeq/L and 21% have a baseflow pH less than 6.0 [*R. B. Robinson et al.*, 2008]. Soils are thin and consist of rocky, sandy loams. The GRSM is not heavily influenced by agriculture or urbanization and approximately 80% of the GRSM is comprised of deciduous forests.

The climate of GRSM is perhumid mesothermal with seasonal temperature variation and precipitation distributed throughout the year [*Busing*, 2005]. The average annual rainfall varies significantly throughout the park with lower elevations generally receiving nearly 127 cm and some high elevation sites nearly 216 cm [*Busing*, 2005]. The pH of precipitation is about 4.5 in the GRSM region [*USEPA*, 1999]. Summer and early spring generally have the most abundant precipitation, with rainfall averages of 12.7 cm and 20.3 cm per month in lower and upper elevations, respectively. Autumn is the driest season, with rainfall averages around 7.6 cm and 12.7 cm per month in lower and upper elevations, respectively.

2.2 Study Design

Monitoring sites with continuous pH data available for close to one year were compiled for eighteen locations within the GRSM (Figure 1). The watersheds selected for study are representative watersheds of the GRSM, typified by steep gradients and thin sandy loams that provide poor buffering capacities. Sonde data collection period ranged from June 2003 to September 2009, including durations from nearly one year to four years (Table 1). Parkwide water quality monitoring has produced an extensive water chemistry database. Yellow Springs Instruments (YSI) sondes or Eureka Manta multi-parameter sondes were equipped at each site, recording stage, pH, conductivity, temperature, and at some sites turbidity at 15-minute or 30-minute intervals. The degree of stream acidification can be classified based on a number of parameters including acid neutralizing capacity, pH, and aluminum concentration. Streamwater pH is an important water quality parameter because it is a good indicator of stream health, including the stream's ability to sustain aquatic life.

A particularly valuable monitoring site within the GRSM dataset is the Noland Divide Watershed (NDW), the longest continuously monitored water quality site in the park and one of the more intensively studied high-elevation watersheds within the eastern United States

[Robinson et al., 2003]. Two first-order streams (SW and NE streams) drain the area headwaters and merge downstream as Noland Creek. NDW has been monitored for pH with sonde data since 1991, as part of the Inventory and Monitoring program of the GRSM, to track long-term depositional trends and understand their consequences. Such a unique and voluminous dataset allows for temporal analysis to determine the period length necessary to adequately characterize a watershed using the CFD methodology.

2.3 Development of CDF Curves

Poisson Arrival Approach

CDF curves are developed by the Poisson arrival approach, where downward episodic pH spikes during stormflows are stochastically modeled by observing the frequency and duration of events below a pH level of interest (pH_0). The pH (negative log of [H^+]) is plotted over time for a hypothetical stream (Figure 2). Figure 2 can be described as a chemograph, a plot of changing chemical concentration with respect to time. Here, pH is the changing chemical concentration and the time is in days. This stochastic type of problem is supported by a substantial theoretical development history [*Cramer and Leadbetter*, 1967; *Todorovic*, 1978]. For this application, pH_0 is the pH of interest and the plot shows crossings of the pH_0 level through episodic downward spikes. Time duration of the crossings (D_m) is the connected time the measured pH is less than pH_0. The arrival of episodic events and their durations are assumed to be random and similar to a Poisson process.

The duration of Poisson occurrences is the length of time between successive up and down crossings of the pH level of interest. In a Poisson arrival process, the number of arrivals is described by the Poisson distribution:

$$P(m) = \lambda^m e^{-\lambda} / m! \qquad \textbf{1}$$

Where P(*m*) is the probability of m occurrences in the time interval, *m* is the number of occurrences in the time interval, and λ is the mean number of occurrences in the time interval.

The length of time or duration of the event is described by the exponential distribution [*Anderson et al.*, 1993]. The exponential distribution has been proposed to describe the duration of exceedances for acid shocks in streams [*Bobba et al.*, 1990]. Mathematically, the exponential distribution function describing the durations is:

$$P(D \geq d) = e^{-d/\mu} \qquad \textbf{2}$$

Where $P(D \geq d)$ is the probability that the duration of a specific event (D) exceeds the duration of interest, *d*, and μ is the mean duration of all events. Equation (2) can also be written in the form

$$N_d = N_T \cdot e^{-d/\mu} \qquad \textbf{3}$$

or

$$\ln(N_d) = \ln(N_T) - d / \mu \qquad \textbf{4}$$

where N_d is the expected number of events exceeding a specified pH with a duration ≥ *d*, and N_T is the total number of events in a time period.

Construction of CDF Curves

An Excel® Visual Basic for Applications macro was created to form a list of durations for each event in which the pH exceeded an arbitrary cutoff value (Table 2). The cutoff values were assigned as every 0.5 change in pH. The frequency of events and their duration get shorter for extreme pH events, below the average pH. As pH increases above the average, the frequency of events again become lower, but with longer durations. Values from Table 2 were plotted along with fitted curves (Figure 3). The dependent axis is normalized by adjusting to events per year by multiplying each y-coordinate by a ratio of 365 over the total duration (days). For the CDF plots, "concentration" is represented by the pH cutoff, the x-axis is the "duration" of an event greater than or equal to the specified pH cutoff, and the y-axis represents "frequency" or more specifically the number of events per year for events greater than or equal to the specified pH cutoffs and duration. CDF plots can approximate the frequency of various time-connected durations of episodic events, e.g., the expected number of events per year that the pH will be less than 5.0 for at least one day.

Since the time resolution of sonde data is typically 15 minutes, extremely short-term events are inaccurately represented and are likely caused by other processes instead of storm events. Thus, durations less than or equal to one hour are truncated and not included. Most of the short duration events of one hour of less were concentrated near the average pH, therefore had little effect whether they were included or excluded from the regression.

Gaps in data from equipment malfunction were observed and had to be addressed. All gaps in record greater than one day are detailed in Appendix B. These breaks in data were ignored and the data stitched together. Stitching describes linking the last available pH data point before the break to the first point after the break, therefore assuming fifteen minute duration between the points. Robinson and Roby noted that this could potentially truncate pH depression events that are ongoing before or after the break, however other approaches, for example, assuming that the event is half over at the break, would automatically and arbitrarily add another pH event [2006].

Development of Characteristic Equations

The data in Table 2 were fitted to Equation (3) by least squares linear regression, using number of events as the dependent variable versus duration as the independent variable. The mean event duration for each pH level is the fitting parameter, μ, for the exponential distribution. The values of μ produced from the regression were then plotted against their respective pH (Figure 4), which gives the average duration of a pH event versus pH. For example, the characteristic equation for Greenbrier Ramsey site is

$$\mu = 1 \cdot 10^{-9} \cdot e^{4.028 \cdot pH} \qquad 5$$

The μ is the mean time-connected duration of all occurrences in which the pH is below a certain pH level. Characteristic equations like Equation 5 facilitate comparisons to various physical watershed attributes due to the strong exponential trend that μ has with pH. Here, the power of the characteristic equation (4.028) depicts the shape of the pH vs. time duration plot of a stream, more specifically the shape of the average duration of episodic pH drops during stormflow. For watersheds in general, higher power values describe sharper and likely shorter duration drops from baseflow pH levels. The change in shape can best be illustrated on a chemograph (Figure 2), where an increase in power is synonymous with a decrease in duration, $D_{m,}$ especially for low pH events. One over the power value is estimated as the slope of the characteristic equation and can be described as the “pH event duration slope metric.” The pH event duration slope metric, denoted S_μ, remains constant independent of total duration. Baseflow pH conditions shift the curve to the left or right, e.g., low pH during baseflow at higher elevations shifts the curve left, but whether baseflow pH has an affect on the pH event duration slope metric is not known.

The metric S_μ is the best metric to use for comparisons to physical watershed attributes because the only alternative option in Equation 5 is the intercept value (1E-9). The intercept value is dependent on the total duration of the data set, thereby making the it useless in comparisons between pH data of different total durations.

Further inquiry of this methodology may be addressed by Robinson and Roby’s [2006] manuscript detailing the development of this technique. In that manuscript, the authors explored use of other equation distributions including bi-Poisson and expanded exponential forms, which did not improve curve fitting.

2.4 GIS Analyses

Physical attributes in contributing watersheds were quantified using geographic information system (GIS) software. Data were collected from the National Park Service GIS database. ArcGIS 9.3 was utilized to evaluate physical watershed and sub-watershed characteristics for each monitoring site. ArcHydro® tools and Spatial Analyst® tools were used to delineate watersheds. Each sampling point location is, by definition, the lowest elevational point within its respective watershed. It should be noted that some watersheds are made up of smaller sub-watersheds, making their attributes cumulative. Zonal statistics were applied to each watershed to evaluate drainage basin area in km^2; SSS in km^2 and % coverage; elevation in meters for minimum, maximum, range, mean, and at monitoring site; and slope in meters for minimum, maximum, range, mean.

2.5 Statistical Analyses

Spatial Analysis for Characteristic Equations vs. Watershed Attributes

The pH event duration slope and watershed attributes for each monitoring site were compared using Pearson coefficients and regression, after being checked for normality. Least squares linear regression was employed to observe relationships between pH event duration slope and watershed attributes. Forward and backward stepwise regression and best subsets regression was used to develop bivariate models for estimating the pH event duration slope metric as a function of watershed attributes. The predictor with the highest partial F value is entered into the model first with additional predictors being individually entered if the addition is significant ($F > 4$) in the regression equation. Likewise, if retaining a predictor is not significant ($F < 4$) in the regression equation it will be removed. Strength and fit of the models were evaluated using R^2, adjusted R^2, predicted R^2, S Value, and Mallows' Cp. Multicollinearity was addressed using the variance inflation factor (VIF). A VIF value above 10 is often used to indicate multicollinearity concerns between variables [*Helsel and Hirsch*, 2002]. Autocorrelation within residuals were checked by the Durbin-Watson statistic. If adjacent observations are correlated, the regression model will underestimate the standard error of the coefficients causing the predictors to seem more significant than they are. Because different selection criteria are used in each model, it is possible that the two bivariate regression techniques will lead to different models. The simplest model that explains a comparable amount of the variability and adheres to the assumptions of regression was therefore chosen as the best model. Only significant variables were used in the regression models (*p* value of 0.05). Statistical analyses were performed in Minitab 15 and SPSS 17.0.

Temporal Analysis of Noland Divide Data

A temporal analysis of long-term NDW data was used to determine whether a one-year data period is adequate to characterize acidification response. Five characteristic equations are developed at the Noland Divide SW site, including four one-year characteristic equations and one characteristic equation summarizing the entire four year period. Chi-square goodness-of-fit test was performed on the dataset to test whether characteristic equations follow similar distributions. Testing was not preformed on S_μ because flawed distribution analysis result when all values are less than one. Instead, a chi-square goodness-of-fit test is conducted on the pH event duration for all five characteristic equations. Here, the null hypothesis is that data follow the same distribution.

3 RESULTS

3.1 CDF Curves and Watershed Attributes

Characteristic equations, i.e. Equation 5, developed for each of the eighteen study sites. Kolmogorov-Smirnov tests and the normal probability plot show the pH event duration slope (S_μ) to be normally distributed. In summary of the study sites, the mean pH event duration is 4.71, median is 4.55, and range is 3.94, respectively (Table 1). The minimum pH event duration of 2.81 characterizes the Greenbrier upper site response. 6.75, the maximum pH event duration, characterizes the Lost Bottom Creek site response. Figure 5 summarizes characteristic trends obtained for all study sites.

Among the study sites, watershed areas ranged from 0.086 km^2 to 118 km^2; SSS covers from none to 79% of watershed area per site, with a maximum area of 19.6 km^2; sonde monitoring elevations ranged from 414 to 1694 m; maximum elevation per watershed ranged from 1465 m to 2019 m; the maximum observed slope was near 72% (Table 3). Generally, slopes and elevations are highest in headwater areas and lowest in downstream reaches. However, since downstream watershed attributes include contributing headwater basin attributes, attribute values are cumulative. For example, the highest monitoring site elevation is at Noland Creek, which does not correspond to the highest maximum elevation watershed values, which occur at the Greenbrier monitoring sites.

3.2 Relationships between pH Event Duration Slope and Watershed Attributes

Correlation analyses showed significant watershed attribute variables that relate to the pH event duration slope metric (S_μ), including drainage area (km^2), SSS (km^2), maximum elevation (m), elevation range (m), maximum slope (%), and slope range (%) (Table 4). All significant correlations ($p < 0.05$) have positive Pearson tau coefficients, signifying that pH event duration slope increases as area, SSS, elevation, and slope increase. Consequently, pH event duration has an inverse relationship with area, SSS, elevation, and slope.

Single variable relationships do not serve as best predictor models, but the regression plots can illuminate obvious trends in the data. Drainage area demonstrated the strongest linear relationship by least squares linear regression, with R^2=57.8, R^2 adjusted=55.3, and R^2 predicted=45.4 (Figure 6), followed by maximum elevation (R^2=43.9, R^2 adjusted=40.4, R^2 predicted=0), elevation range (R^2=41.9, R^2 adjusted=38.3, R^2 predicted=16.76), maximum slope (R^2=30.5, R^2 adjusted=26.1, R^2 predicted=0.02), slope range (R^2=25.5, R^2 adjusted=20.8, R^2 predicted=0), and SSS in km (R^2=22.5, R^2 adjusted=17.7, R^2 predicted=0). To note, best subsets regression validates those five variables as best one variable predictor models (Appendix D).

Drainage area plotted against the pH event duration slope metric, S_μ, (Figure 5) reveals the possible existence of two separate linear relationships, one for areas of 0 to 40 km^2 and a second for areas including 60 to 120 km^2. A threshold may exist somewhere between 40 and 60 km^2, where a change in trend may result. This theory was further examined in Figure 7, by illustrating separate linear relationships. For the thirteen points grouped below the threshold line, including areas of 0 to 40 km^2, R^2=21.3, R^2 adjusted=14.2. For the five points grouped above the threshold line, including areas of 60 to 120 km^2, R^2=50.7 and R^2 adjusted=34.3.

Stepwise linear regression includes drainage basin area and maximum elevation, resulting in R^2 of 74.5, R^2 adjusted of 71.1, R^2 predicted of 42.8, VIF of 1.16, and Durbin-Watson statistic of 1.26. The resulting prediction equation is $S_\mu = -0.156 + 0.000816 * \text{Area (km}^2) + 0.000184 *$ Maximum Elevation (m). Figure (7) illustrates predicted pH event duration slope (S_μ) from the regression model verses actual 1 / μ from field data. Figure 8 shows increasing variability as drainage basin area becomes larger in size. One may infer that as drainage area becomes larger, it becomes increasingly difficult to predict consistent acidification mechanisms due to the unique attributes and hydrological responses associated with sub-watersheds.

3.3 Temporal Variation of CDF Curves for a Single Watershed

Chi-square goodness-of-fit test shows that the null hypothesis, that all pH event duration metric values for the five CDF curves at Noland Divide SW site fit the same distribution, cannot be rejected ($\alpha = 0.1$, p-value = 0.94). The summary plot of μ vs. pH event at Noland Divide SW Site (Table 5, Figure 9) illustrates the similar trends of five curves at different time periods for the same watershed. For pH 5.0, mean duration ranges from 0.17 to 0.47 days, for pH 5.5 mean duration ranges from 0.77 to 3.5 days, and for pH 6.0 mean duration ranges from 3.5 to 158 days. For pH 6.5, only two mean durations exist. The four one-year curves have the same total duration near one year (360 days). Since the curve including all four years is the sum of the one-year curves, the total duration is near four years (1444 days). Here, the pH event duration slope is similar between the five curves, but the curve for near four years duration shifts the duration per event pH curve upward. This implies that as long as the total duration is near one year or greater, total duration has minimal influence on slope of μ vs. pH event.

4 DISCUSSION

CDF curves aided in summarizing large amounts of sonde data that can be fitted by an exponential distribution, and capture more information than other approaches using general descriptive statistics or time series analysis. Descriptive statistics alone do not adequately characterize the episodic nature of water quality parameters and require relatively few data points [*Kneale and Howard*, 1997]. Time series analysis supports development of predictive stochastic models, but does not allow for site characterization with field data. CDF curves allow for quantitative spatial and temporal comparisons of intensities, durations, and frequencies of episodic events. The importance of CDF relationships has been recognized by others, but previous studies have not evaluated spatial multi-site field data characterization [*Schwartz et al.*, 2008]. Robinson and Roby [2006] developed and applied the CDF curve methodology, similar to precipitation intensity-duration-frequency curves, to pH data for four sites in the Little Pigeon River watershed of the GRSM, but this technique is not restricted to pH data. Schwartz et. al. [2008] illustrated CDF curve development for stream turbidity data to assess biological stream impairment from siltation and to support development of sediment total maximum daily loads (TMDLs).

Watershed drainage area appears to be a dominant driver in stream response from episodic acidification of streams monitored for this study, explaining 65% of the variance. Herlihy et al. determines almost all acidic (ANC<0) streams in the Appalachian highlands are located in small upland watersheds, less than 20 km^2 in area [1993]. During stormflow, water is routed through upper soil layers that are acidic due to acid deposition or other natural processes, like the flushing of organic acids or base cation dilution [*Wellington and Driscoll*, 2004]. The stormwater has less time to react with base cations in the soil and generally is more acidic upon reaching streamflow [*Wigington et al.*, 1996a]. Also, smaller high-elevation watersheds can have positively charged soil surface conditions causing sulfate adsorption and base cation desorption [*Cai et al.*, 2009]. This depletion of base cations contribute to episodic ANC depressions [*Castro and Morgan*, 2000]. Deyton et al. [2009] showed that base cations contribute to reductions in episodic pH and ANC depressions for stormflow in larger watersheds in the GRSM. Therefore, increased drainage basin area may increase base cations available in stormflows.

Basin elevation has a strong influence on stream response from episodic acidification, explaining 44% of variance alone, and when coupled with basin area, explains 75% of variance. Previous storm event studies in the GRSM show significant stream pH drops at higher elevations [*R. B. Robinson et al.*, 2008]. The potential for a variable pH response between different sites in GRSM have been shown, where baseflow pH decreased by 0.72 pH units for every 1000 meter increase in elevation [*Roby*, 2005]. High elevation watersheds typically have low bedrock weathering rates that can diminish neutralizing capacity. In high elevation watersheds, acidification may be more a function of acidic anion inputs. Interestingly, basin area and elevation are sometime considered surrogates for stream size.

This study is not meant to diminish the complexity of the episodic acidification process, nor propose a simple equation that explains all the biogeochemical processes and environmental drivers that cause acidic episodes in streams. Inherent randomness of the driving variables and the randomness of the hydrologic system cause difficulty in explaining or predicting these processes [*Maidment*, 1993]. Some significant variables are intercorrelated and the strength of the observed correlations does not necessarily imply cause-effect. High elevation sites are generally more likely to have a smaller area, have steeper slopes, and be underlain by siliceous formations.

Period length and frequency of sonde data collection required are dependent on the episodic nature of pH events for a study. Seasonality and hydrologic patterns can have impacts on stream chemistry. Stream water pH is temporally variable in response to seasonal patterns in plant and microbial activities and dynamic hydrological conditions [*Sullivan et al.*, 2007]. Although the extent and magnitude of episodic acidification varies by site and with meteorological conditions, some generalities exist. For example, during the spring season, baseflow pH is likely to be depressed to its lowest levels of the year [*Herlihy et al.*, 1993]. Likewise, ANC values are usually at a maximum during summer base flow events [*Driscoll et al.*, 2001]. Episodic acidification is most common during seasons of high precipitation. Deyton et al. indicate large storms preceded by long, dry periods cause the largest pH depressions [2009]. Also, hydrologic flow paths typically shift from deeper soils and geologic strata of watersheds during baseflow to shallower, more acidic soils during storm events [*Wigington et al.*, 1990]. Despite temporal concerns, data collected in this study for approximately a one-year period was sufficient to characterize the episodic nature of GRSM pH events, as observed by the chi-square goodness-of-fit results. This outcome suggests that although hydraulic and climatic factors affect baseflow chemistry, episodic acidification of streams for this study induced by stormflow is dependent on watershed attributes and the complex biogeochemical processes in the particular watersheds.

As with any methodology, limitations are associated with use of CDF curves to characterize watershed response. First, the scale for this study is small, consisting of drainage basins no larger than 118 km^2. Larger scale watershed studies, such as regional or eco-regional studies, may find different associations and/or additional landscape variables that might affect stream pH. An advantage to having a larger data set with large sample size is ability to quantify relationships at multiple spatial scales. Second, continuous pH data for one year, preferably at fifteen minute durations, is required to produce characteristic CDF curves for any basin. Such data are not likely available for many watersheds of interest and producing such a data set is expensive and time intensive. Third, another data availability concern is that the assignment of landscape statistics depends on the availability of high quality gridded spatial data. The method of assigning proportions of landscape type within GIS grid squares is also approximate. Last, there is inherent variability in any data set, due to factors like equipment calibration, mechanical malfunctions, and operator errors; and these errors may have more of an effect on data sets with small durations. The limitations mentioned here are likely to be encountered under any attempt to model catchment scale water chemistry by linking landscape type, or watershed attributes, to water chemistry parameters. A limitation specific to CDF curve development is the requirement for enough pH variability to adequately regress a pH event duration metric. Newt Prong is an additional monitoring site where a characteristic equation was developed, but the metrics could not be applied in this study because of the small variability in pH data. Figure 10 illustrates that pH never goes below 6.0 or above 7.0. Two data points for the pH vs. μ regression causes difficulty in characterizing this site with any confidence.

In summary, the focus of this research was to apply the CDF curve methodology to summarize stream response and characterize stormflow based on physical watershed attributes. Significant stream response drivers include basin drainage area, elevation, surficial geology, and slope. Overall, this study validates that the CDF methodology applied to continuous pH data near one year in duration can adequately characterize a watershed's response to acidification.

REFERENCES

REFERENCES

Anderson, D. R., D. J. Sweeney, and T. A. Williams (1993), *Statistics for business and economics*, Minneapolis.

Baker, J. P., et al. (1996), Episodic acidification of small streams in the northeastern United States: Effects on fish populations, *Ecological Applications*, *6*(2), 422-437.

Baldigo, B. P., and G. B. Lawrence (2001), Effects of stream acidification and habitat on fish populations of a North American river, *Aquatic Sciences*, *63*(2), 196-222.

Bobba, A. G., D. C. L. Lam, W. G. Booty, D. S. Jefries, and R. L. Thomas (1990), Stochastic analysis of acid shocks generated by mixed hydrological processes, *Water Air Soil Pollut.*, *53*(3-4), 239-250.

Busing, R. T. (2005), Temperate Forest: Great Smoky Mountains, Tennessee, USA, 1978-1992. Data set. NPP (Net Primary Production), edited, Oak Ridge National Laboratory - Distributed Active Archive Center.

Cai, M., J. Schwartz, R. Robinson, S. Moore, and M. Kulp (2009), Long-Term Effects of Acidic Deposition on Water Quality in a High-Elevation Great Smoky Mountains National Park Watershed: Use of an Ion Input–Output Budget, *Water, Air, & Soil Pollution*.

Calta, M. (2002), Does brown trout (*Salmo trutta* L.) larval development retardation caused by short-term exposure to low pH and elevated aluminum concentration affect a second episode of toxicity?, *J. Appl. Ichthyol.*, *18*, 210-215.

Castro, M. S., and R. P. Morgan (2000), Input-output budgets of major ions for a forested watershed in western Maryland, *Water, Air and Soil Pollution*, *119*, 121-137.

Clow, D. W., and J. K. Sueker (2000), Relations between basin chacteristics and stream water chemistry in alpine/subapline basins in Rocky Mountain National Park, Colorado, *Water Resour. Res.*, *36*(1), 49-61.

Cooper, D. M., A. Jenkins, R. Skeffington, and B. Gannon (2000), Catchment-scale simulation of stream water chemistry by spatial mixing: theory and application, *Journal of Hydrology*, *233*(1-4), 121-137.

Cramer, H., and M. R. Leadbetter (1967), *Stationary and related stochastic processes*, New York.

Deviney, F. A., K. C. Rice, and G. M. Hornberger (2006), Time series and recurrence interval models to predict the vulnerability of streams to episodic acidification in Shenandoah National Park, Virginia, *Water Resour. Res.*, *42*(W09405 (1-14)), doi:10.1029/2005WR004740.

Deyton, E. B., J. S. Schwartz, R. B. Robinson, K. J. Neff, S. E. Moore, and M. A. Kulp (2009), Characterizing episodic stream acidity during stormflows in the Great Smoky Mountains National Park, *Water Air Soil Pollut.*, *196*(1-4), 3-18.

Driscoll, C. T., K. M. Driscoll, M. J. Mitchell, and D. J. Raynal (2003), Effects of acidic deposition on forest and aquatic ecosystems in New York State., *Environmental Pollution*, *123*, 327-336.

Driscoll, C. T., G. B. Lawrence, A. J. Bulger, T. J. Butler, C. S. Cronan, C. Eagar, K. F. Lambert, G. E. Likens, J. L. Stoddard, and K. C. Weathers (2001), Acidic deposition in the northeastern United States: sources and inputs, ecosystem effects, and management strategies, *BioScience*, *51*(3), 180-198.

Evans, C. D., D. M. Cooper, S. Juggins, A. Jenkins, and D. Norris (2006), A linked spatial and temporal model of the chemical and biological status of large, acid-senstive river network, *Science of the Total Environment*, *365*, 167-185.

Helsel, D. R., and R. M. Hirsch (2002), Statistical methods in water resources. USGS.

Herlihy, A. T., P. R. Kaufmann, and M. E. Mitch (1991), Stream chemistry in the Eastern United States: 2. Current sources of acidity in acidic and low-acid-neutralizing capacity streams, *Water Resour. Res.*, *27*(4), 629-642.

Herlihy, A. T., J. L. Stoddard, and C. B. Johnson (1998), The relationship between stream chemistry and watersheds land cover data in the mid-Atlantic region, U.S., *Water, Air and Soil Pollution*, *105*, 377-386.

Herlihy, A. T., P. R. Kaufmann, M. R. Church, P. J. Wigington, J. R. Webb, and M. J. Sale (1993), The effects of acidic deposition on streams in the Appalachian Mountains and Piedmont Region of the Mid-Atlantic United States, *Water Resour. Res.*, *29*(8), 2687-2703.

Hesthagen, T., I. H. Sevaldrud, and H. M. Berger (1999), Assessment of damage to fish population in Norwegian lakes due to acidification, *Ambio*, *28*(2), 112-117.

Huckabee, J. W., C. P. Goodyear, and R. D. Jones (1975), Acid rock in the Great Smokies: unanticipated impact on aquatic biota of road construction in regions of sulfide mineralization, *Trans. Am. Fish. Soc.*, *104*(4), 677-684.

Johnson, N. M., C. T. Driscoll, J. S. Eaton, G. E. Likens, and W. H. McDowell (1981), Acid-rain, dissolved aluminum and chemical-weathering at the Hubbard-Brook Experimental Forest, New Hampshire., *Geochimica Et Cosmochimica Acta*, *45*(9), 1421-1437.

Kneale, P. E., and A. Howard (1997), Statistical analysis of algal and water quality data, *Hydrobiologia*, *349*, 59-63.

MacAvoy, S. E., and A. J. Bulger (1995), Survival of brook trout (*Salvelinus fontinalis*) embroys and fry in streams of different acid sensitivity in Shenandoah National Park, USA, *Water, Air and Soil Pollution*, *85*, 445-450.

Maidment, D. R. (1993), *Handbook of Hydrology*, McGraw-Hill Professional, New York.
NADP (2007), National Atmospheric Deposition Program 2007 annual summary*Rep.*, Champaign, Illinois.

Minitab (2007), Minitab 15 Statistical Software, edited by StatGuide.

Neff, K. J., J. S. Schwartz, T. B. Henry, R. B. Robinson, S. E. Moore, and M. A. Kulp (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.

O'Brien, A. K., K. C. Rice, M. M. Kennedy, and O. P. Bricker (1993), Comparison of episodic acidification of mid-Atlantic Upland and Coastal Plain streams, *Water Resour. Res.*, *29*(3), 3029-3039.

Robinson, R. B., and J. C. Roby (2006), Concentration-duration-frequency curves for pH in a stream in the Great Smoky Mountains, *Journal of Environmental Engineering*, *132*(12), 1600-1605.

Robinson, R. B., T. W. Barnett, G. R. Harwell, S. E. Moore, M. Kulp, and J. S. Schwartz (2008), pH and acid anion time trends in different elevation ranges in the Great Smoky Mountains National Park, *Journal of Environmental Engineering-ASCE*, *134*(9), 800-808.

Roby, C. J. (2005), Water Quality Characterization, Stormwater Analyses, and Statistical Modeling of Stream Water Quality in the Little Pigeon River Watershed of the Great Smoky Mountains National Park, University of Tennessee, Knoxville.

Schwartz, J. S., M. Dahle, and R. B. Robinson (2008), Concentration-duration-frequency curves for stream turbidity: possibilities for assessing biological impairment, *Journal of the American Water Resources Association*, *44*(4), 879-886.

Silsbee, D. G., and G. L. Larson (1982), Water quality of streams in the Great Smoky Mountains National Park, *Hydrobiologia*, *89*, 97-115.

Stoddard, J. L., et al. (1999), Regional trends in aquatic recovery from acidification in North America and Europe, *Nature*, *401*(6753), 574-578.

Sullivan, T. J., J. R. Webb, K. U. Snyder, A. T. Herlihy, and B. J. Cosby (2007), Spatial distribution of acid-sensitive and acid-impacted streams in relation to watershed features in the Southern Appalachian Mountains, *Water Air Soil Pollut.*, *182*(1-4), 57-71.

Todorovic, P. (1978), Stochastic models of floods, *Water Resour. Res.*, *14*(2), 345-356. Tranter, M., T. D. Davies, P. J. Wigington, and K. N. Eshleman (1994), Episodic acidification of freshwater systems in Canada - physical and geochemical processes, *Water, Air and Soil Pollution*, *72*, 19-39.

USEPA (1999), Progress report on the EPA acid rain program*Rep.*, Washington, D.C. Van Sickle, J., J. P. Baker, H. A. Simonin, B. P. Baldigo, W. A. Kretser, and W. E. Sharpe (1996), Episodic acidification of small streams in the northeastern United States: Fish mortality in field bioassays, *Ecological Applications*, *6*(2), 408-421.

Wellington, B. I., and C. T. Driscoll (2004), The episodic acidification of a stream with elevated concentrations of dissolved organic carbon, *Hydrological Processes*, *18*, 2663-2680.

Wigington, P. J., T. D. Davies, M. Tranter, and K. N. Eshleman (1990), Episodic acidification of surface waters due to acidic deposition.*Rep.*, National Acid Precipitation Assessment Program, Washington D.C.

Wigington, P. J., D. R. DeWalle, P. S. Murdoch, W. A. Kretser, H. A. Simonin, J. Van Sickle, and J. P. Baker (1996a), Episodic acidification of small streams in the northeastern United States: ionic controls of episodes, *Ecological Applications*, *6*(2), 389-407.

Wigington, P. J., J. P. Baker, D. R. DeWalle, W. A. Kretser, P. S. Murdoch, H. A. Simonin, J.

Van Sickle, M. K. McDowell, D. V. Peck, and W. R. Barchet (1996b), Episodic acidification of small streams in the northeastern United States: Episodes Response Project, *Ecological Applications*, *6*(2), 374-388.

APPENDIX A: Referenced Figures and Tables

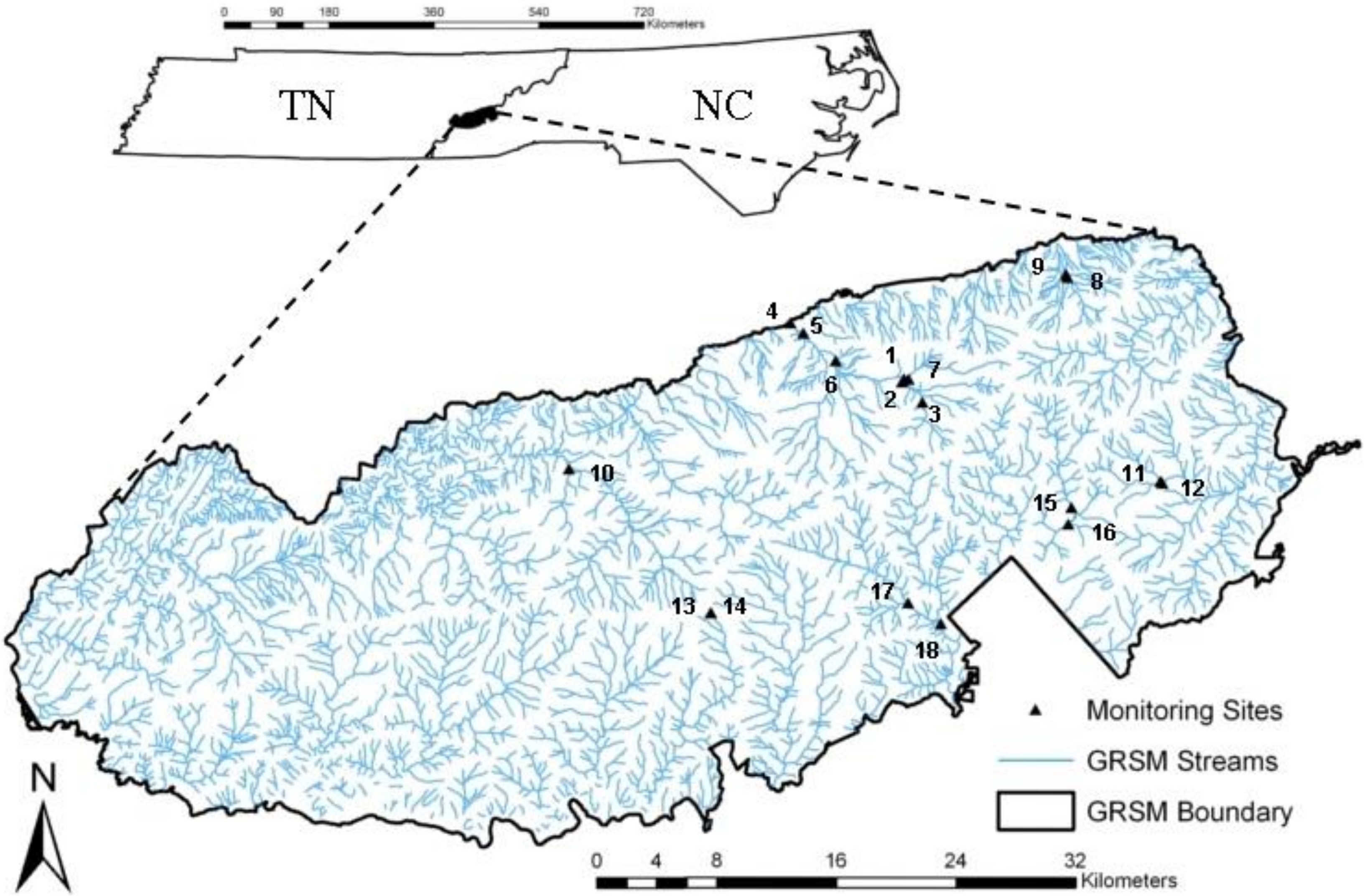


Figure 1. Monitoring Site Locations Within the GRSM, Located At The Tennessee And North Carolina Border. Map ID #'s Correspond To Table 1.

Table 1. Monitoring Site Details Including Duration and Determined Characteristic Values

Map ID # (Figure 1)	Monitoring Site Name	Monitored Time Range	Total Duration (days)	No. of Data Points	pH Event Duration	pH Event Duration Slope (Sμ)	R^2
1	Ramsey Prong	2/17/06 - 8/14/08	520	49811	4.055	0.247	0.937
2	Middle Prong	2/28/06 - 7/3/08	789	75615	4.003	0.250	0.993
3	Eagle Rocks	3/9/06 - 12/11/07	528	50591	5.478	0.183	0.923
4	Greenbrier-1 (Downstream)	8/25/03 - 9/15/04	324	30714	3.037	0.329	0.979
5	Greenbrier-2 (Middle)	9/3/03 - 8/28/04	340	32243	3.560	0.281	0.961
6	Greenbrier-3 (Upper)	8/20/03 - 9/15/04	275	25071	2.813	0.355	0.958
7	Greenbrier-4 (Ramsey)	6/19/03 - 5/25/04	327	31384	4.028	0.248	0.978
8	Rock Prong	4/2/08 - 9/5/09	504	48349	5.647	0.177	0.975
9	Cosby Creek	4/24/08 - 9/8/09	501	48112	5.476	0.183	0.953
10	Jakes Creek	5/29/08 - 9/7/09	453	43492	6.020	0.166	0.993
11	Lost Bottom Creek	4/23/08 - 9/4/09	499	47919	6.753	0.148	0.951
12	Palmer Creek	4/23/08 - 9/4/09	499	47935	6.006	0.167	0.993
13	Noland Creek - NE	11/13/03 - 12/10/06	810	77785	4.643	0.215	0.951
14	Noland Creek - SW	10/22/03 - 12/2/08	1443	138240	5.529	0.181	0.972
15	Straight Fork-Upper	10/27/04 - 3/10/07	637	61077	4.453	0.225	0.953
16	Straight Fork-Lower	10/27/04 - 3/2/07	649	62211	4.191	0.239	0.960
17	Oconaluftee River-Upper	5/25/05 - 7/13/09	1078	103485	5.198	0.192	0.977
18	Oconaluftee River-Lower	9/4/08 - 7/13/09	261	25092	3.957	0.253	0.923

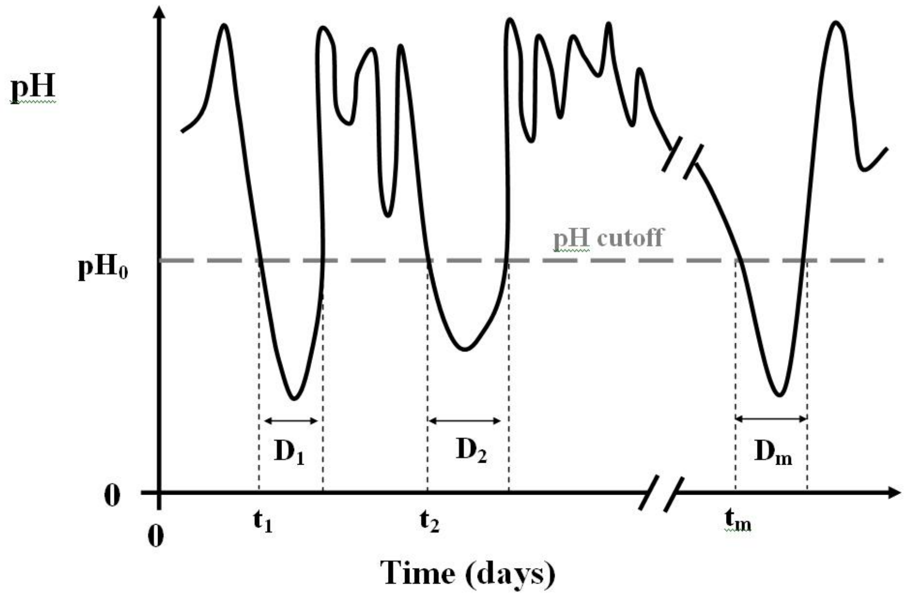


Figure 2. Example of Crossings of pH0 Criterion for a Hypothetical Plot of pH vs. Time

Table 2. Ranked Duration of Events for pH Cutoffs at the Greenbrier Ramsey Site.

		Duration of Events (days)				
Event No.	Event No. Normalized	pH ≤ 4.5	pH ≤ 5.0	pH ≤ 5.5	pH ≤ 6.0	pH ≤ 6.5
1	1.12	0.33	1.54	6.55	98.99	326.98
2	2.23	0.26	1.30	5.60	65.39	–
3	3.35	0.23	1.20	5.28	27.72	–
4	4.47	0.23	1.18	3.02	11.28	–
5	5.58	0.20	0.94	2.86	10.78	–
6	6.70	0.17	0.79	2.78	10.38	–
7	7.81	0.17	0.79	2.77	6.42	–
8	8.93	0.16	0.73	2.68	6.40	–
9	10.05	0.15	0.58	2.34	4.61	–
10	11.16	0.11	0.56	1.99	2.76	–
11	12.28	0.05	0.54	1.91	1.64	–
12	13.40	0.05	0.52	1.78	1.27	–
13	14.51	–	0.47	1.61	0.90	–
14	15.63	–	0.46	1.42	0.87	–
15	16.74	–	0.40	1.39	0.40	–
16	17.86	–	0.24	1.17	0.30	–
17	18.98	–	0.20	1.05	0.27	–
18	20.09	–	–	0.98	0.25	–
19	21.21	–	–	0.81	0.16	–
20	22.33	–	–	0.78	0.11	–
21	23.44	–	–	0.74	0.08	–
22	24.56	–	–	0.64	0.07	–
23	25.67	–	–	0.59	0.06	–
24	26.79	–	–	0.56	–	–
25	27.91	–	–	0.47	–	–
↓	↓	–	–	↓	–	–
36	40.19	–	–	0.06	–	–
Total Number of Events		12	17	36	23	1

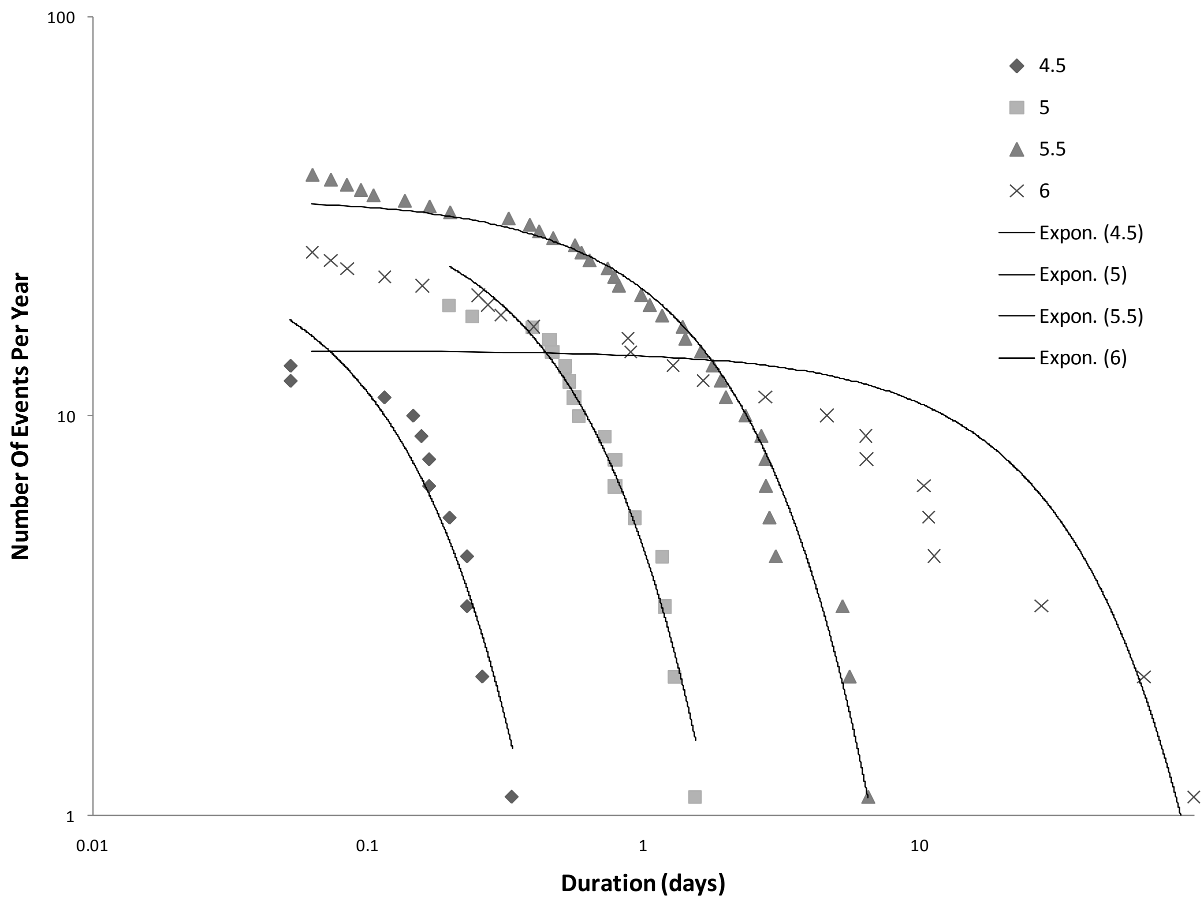


Figure 3. For the Greenbrier Ramsey Site, Data Points Represent the Number of Events Per Time With pH Greater Than the Specified pH Cutoff vs. Duration (days), and the Corresponding Exponential-Fitted Curves.

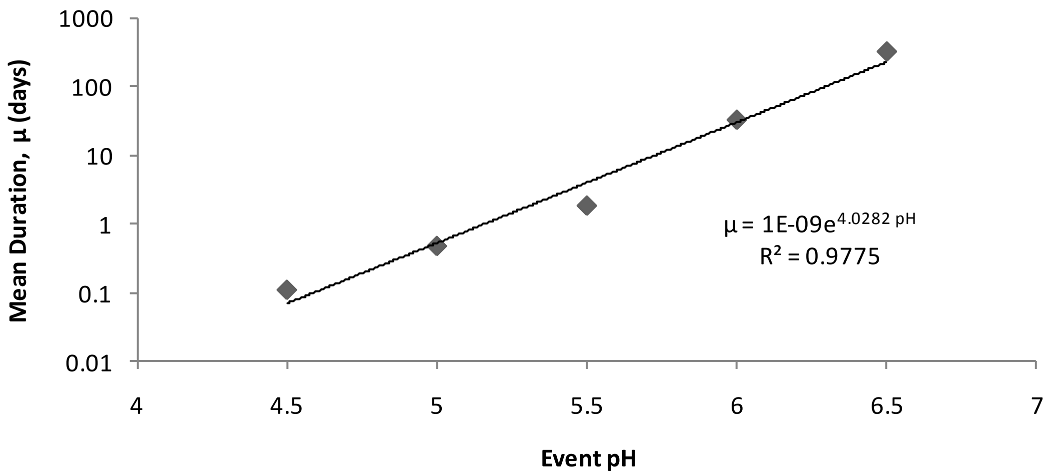


Figure 4. Best-Fit Exponential Regressions for pH vs. μ at Greenbrier Ramsey Site.

Table 3. Summary Table For Watershed Attributes At Monitoring Site Locations

Monitoring Site Name	State	Longitude	Latitude	Area (km^2)	*SSS (km2)	*SSS (% Coverage)	Maximum Elevation (m)	Elevation Range (m)	Average Elevation (m)	Sonde Site Elevation (m)	Maximum Slope (%)	Slope Range (%)	Average Slope (%)
Ramsey Prong	TN	289243.0	3953451.5	10.29	0.00	0.0%	2019	1177	1419	847	69.1	68.8	25.1
Middle Prong	TN	288790.8	3953281.0	38.77	1.17	3.0%	2019	1225	1407	795	71.8	71.6	28.5
Eagle Rocks	TN	290091.3	3951877.5	10.47	1.17	11.2%	1808	843	1445	969	71.8	71.6	30.5
Greenbrier-1 (Downstream)	TN	281543.1	3957626.8	117.71	10.00	8.5%	2019	1605	1146	414	71.8	71.7	28.4
Greenbrier-2 (Middle)	TN	282375.4	3956895.5	111.10	10.00	9.0%	2019	1592	1174	427	71.8	71.7	28.8
Greenbrier-3 (Upper)	TN	284432.5	3954923.9	99.80	10.00	10.0%	2019	1532	1233	487	71.8	71.7	29.3
Greenbrier-4 (Ramsey)	TN	288940.2	3953502.1	10.47	0.00	0.0%	2019	1217	1410	804	69.1	68.8	25.1
Rock Prong	TN	300192.9	3959734.6	3.63	0.00	0.0%	1793	1167	1249	626	61.5	60.2	30.0
Cosby Creek	TN	300118.9	3959943.0	17.56	0.00	0.0%	1793	1188	1111	605	61.7	61.5	28.2
Jakes Creek	TN	266268.3	3948636.6	12.02	4.01	33.3%	1465	807	1096	659	60.3	60.2	22.5
Lost Bottom Creek	NC	305730.4	3945795.9	8.45	0.00	0.0%	1875	869	1423	1008	60.3	60.3	26.4
Palmer Creek	NC	305830.5	3945667.8	20.00	0.00	0.0%	1875	888	1380	987	60.3	60.3	25.8
Noland Creek - NE	NC	275243.2	3938508.6	0.088	0.00	0.0%	1904	212	1798	1694	46.7	41.3	19.7
Noland Creek - SW	NC	275232.7	3938492.6	0.086	0.00	0.0%	1918	227	1814	1692	52.8	48.6	19.8
Straight Fork-Upper	NC	299689.9	3944344.5	27.23	0.00	0.0%	1900	952	1425	948	61.2	61.0	28.4
Straight Fork-Lower	NC	299416.3	3943229.2	39.75	0.00	0.0%	1900	981	1378	919	62.3	62.3	28.2
Oconaluftee River-Upper	NC	288440.7	3938511.2	66.21	13.84	20.9%	1799	1133	1170	729	69.9	69.9	31.0
Oconaluftee River-Lower	NC	290587.2	3937002.9	105.62	19.58	18.5%	1895	1229	1205	666	69.9	69.9	30.1

All elevation values represent elevation above sea level

*SSS represents sililiclastic slate, comprised of anakeesta and copperhill surficial geology features

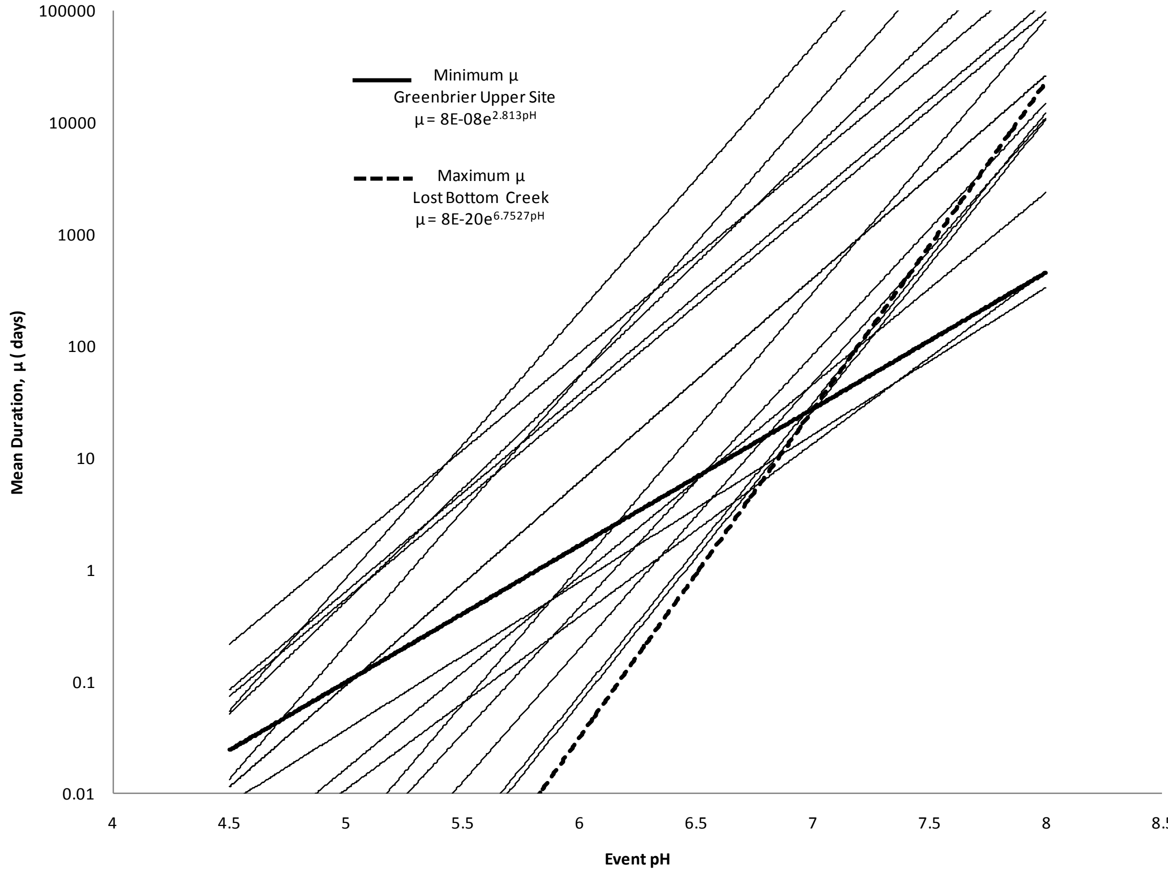


Figure 5. Summary Chart of pH Event Duration Slopes at All Monitoring Sites

Table 4. Pearson Correlation and p values for pH Event Duration Slopes (S_μ) vs. Watershed Attributes

	Area (km^2)	SSS (km^2)	SSS (% Coverage)	Maximum Elevation (m)	Elevation Range (m)	Average Elevation (m)	Sonde Site Elevation (m)	Maximum Slope (%)	Slope Range (%)	Average Slope (%)
Pearson coefficient	0.761	0.475	0.026	0.662	0.648	-0.231	-0.451	0.552	0.505	0.269
p-Value	< 0.001*	0.047*	0.917	0.003*	0.004*	0.356	0.06	0.018*	0.033*	0.28

* denotes significant *p* value (alpha = 0.05)

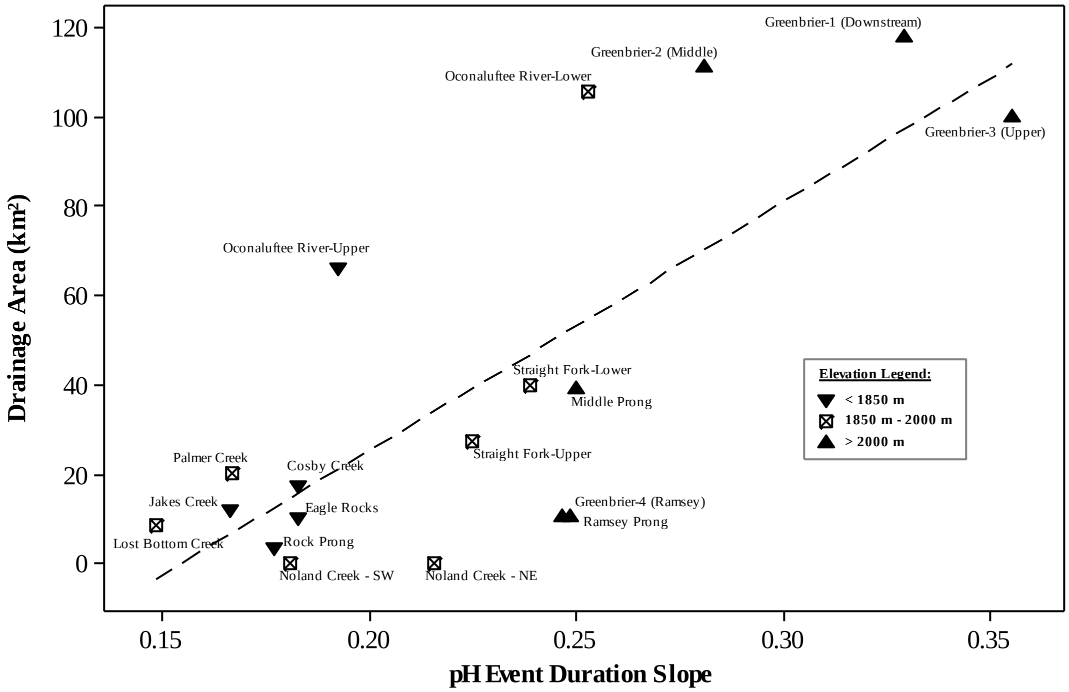


Figure 6. Linear Regression For Best One Variable Predictor Model. Area (km²) vs. pH Event Duration Slope with resulting regression equation pH Event Duration Slope (S_μ) = 0.184 + 0.00104*Area. Points Are Shown To Be In One Of Three Maximum Elevation Classes.

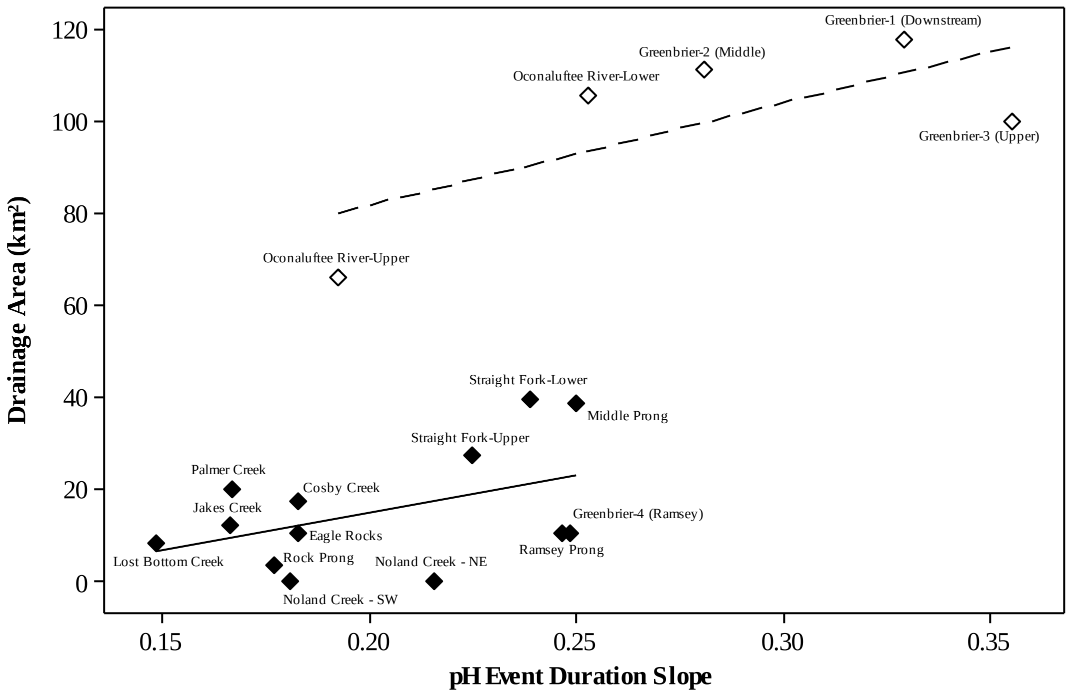


Figure 7. Two Linear Regression Fits For Best One Variable Predictor Model. Area (km^2) vs. pH Event Duration Slope (S_μ).

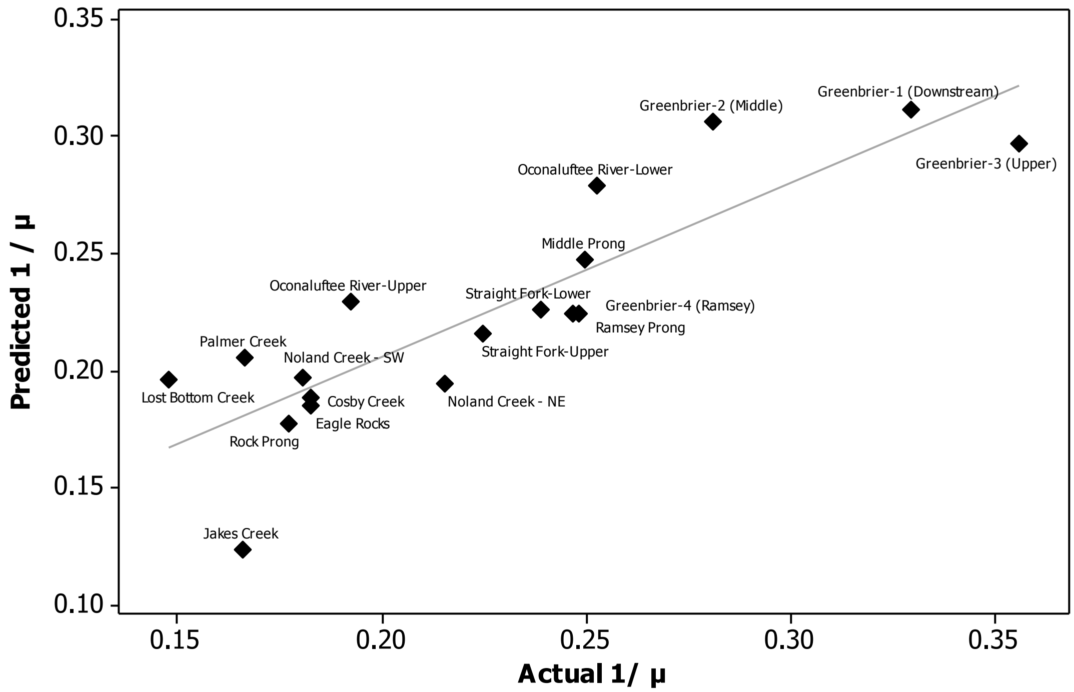


Figure 8. Predicted S_μ vs. Actual S_μ Using Stepwise Regression Equation S_μ = -0.156+0.000816*Area+0.000184*Maximum Elevation.

Table 5. Summary Table of NDW Curves Used For Temporal Analysis

Time Range	Duration (days)	No. of Data Points	pH Event Duration	pH Event Duration Slope (S_μ)	R^2
10/22/03 - 11/1/04	359.9	34560	4.87	0.205	0.910
11/1/04 - 2/24/06	360.7	34560	7.15	0.140	0.977
2/24/06 - 9/13/07	361.9	34560	4.45	0.225	0.999
9/13/07 - 12/2/08	360.5	34560	5.01	0.200	0.937
10/22/03 - 12/2/08	1442.9	138240	5.53	0.181	0.972

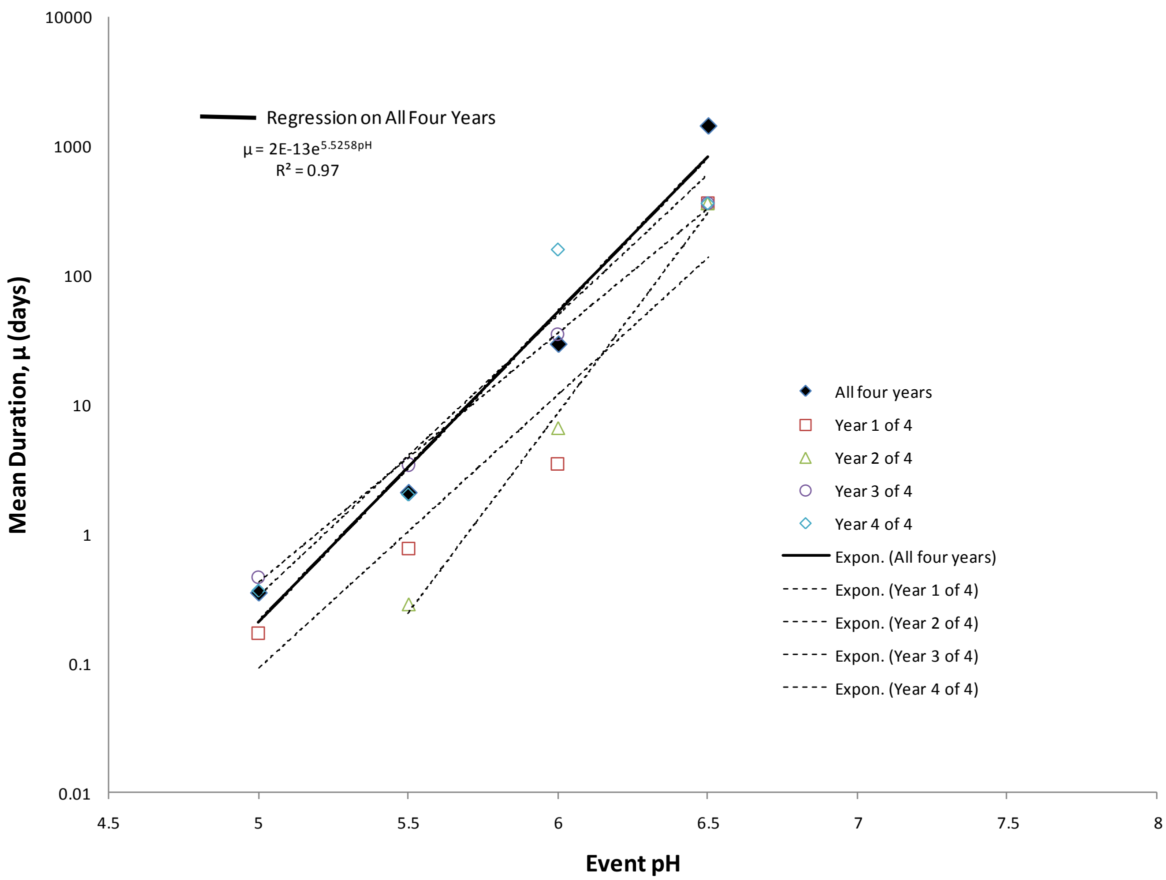


Figure 9. μ vs. Event pH For Temporal Analysis at NDW

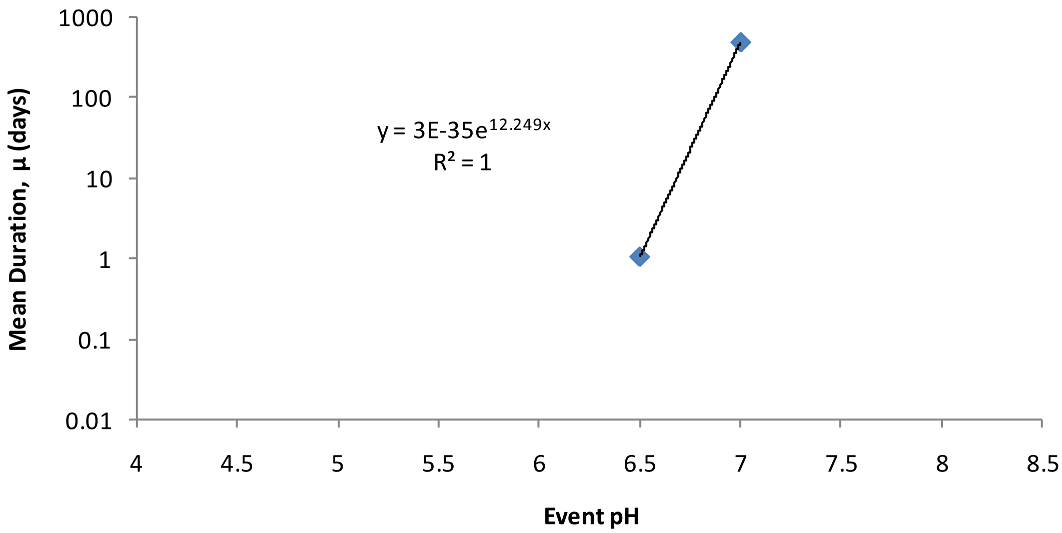


Figure 10. Best-Fit Exponential Regressions for pH vs. µ at Newt Prong Site (Only Two Data Points Available).

APPENDIX B: CDF Curves and Missing Data

pH vs μ Curves Used For Regressions

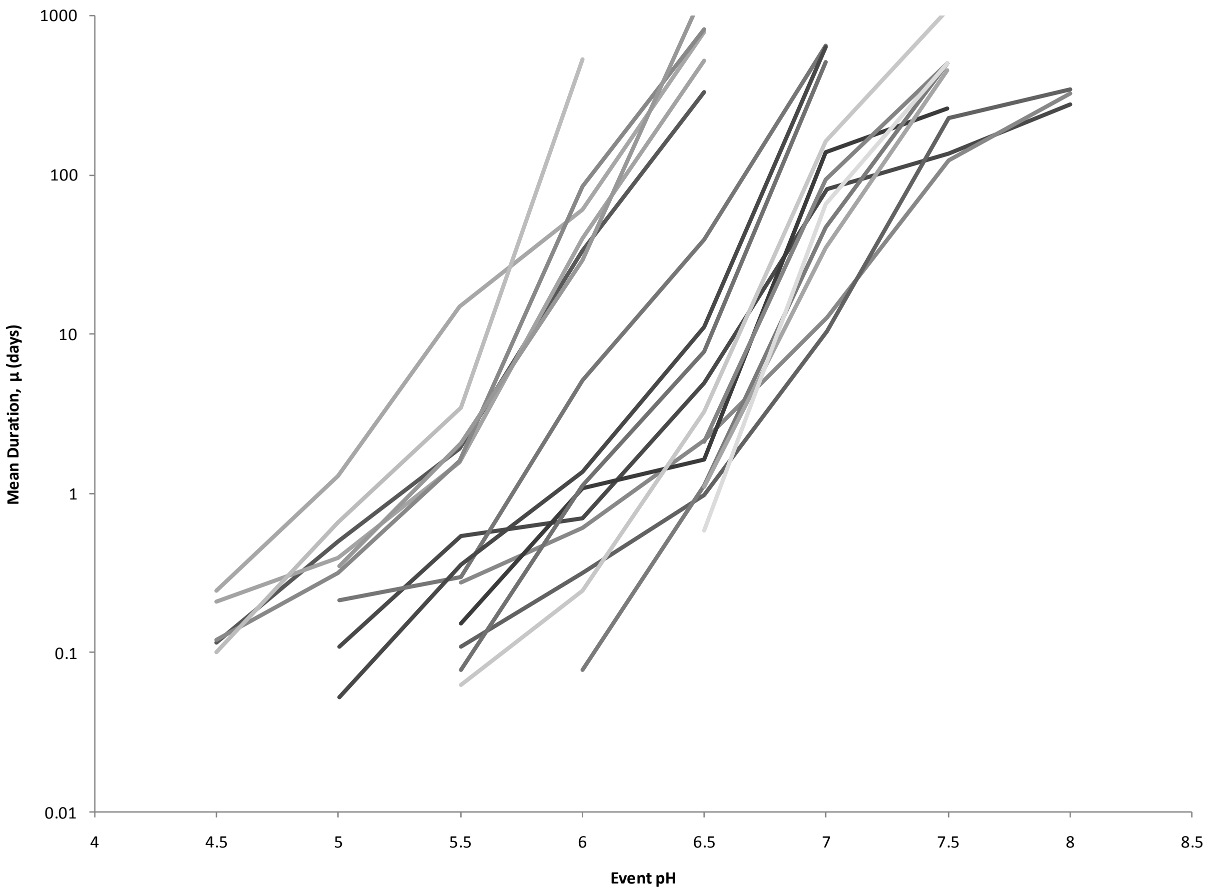


Figure 11. μ vs pH for All Sites With Points Simply Connected (No Regression Fits).

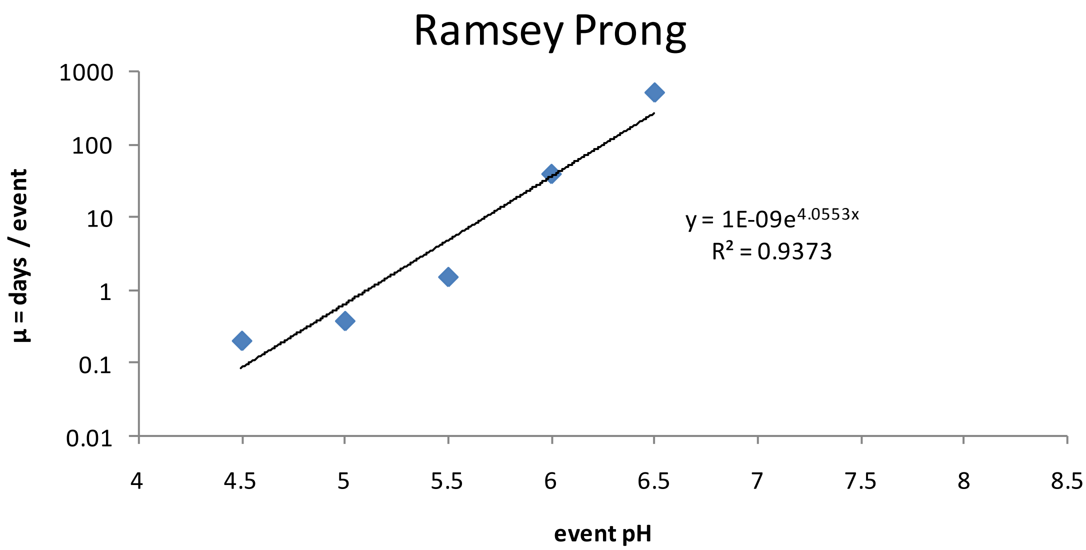


Figure 12. μ vs Event pH for Ramsey Prong Site.

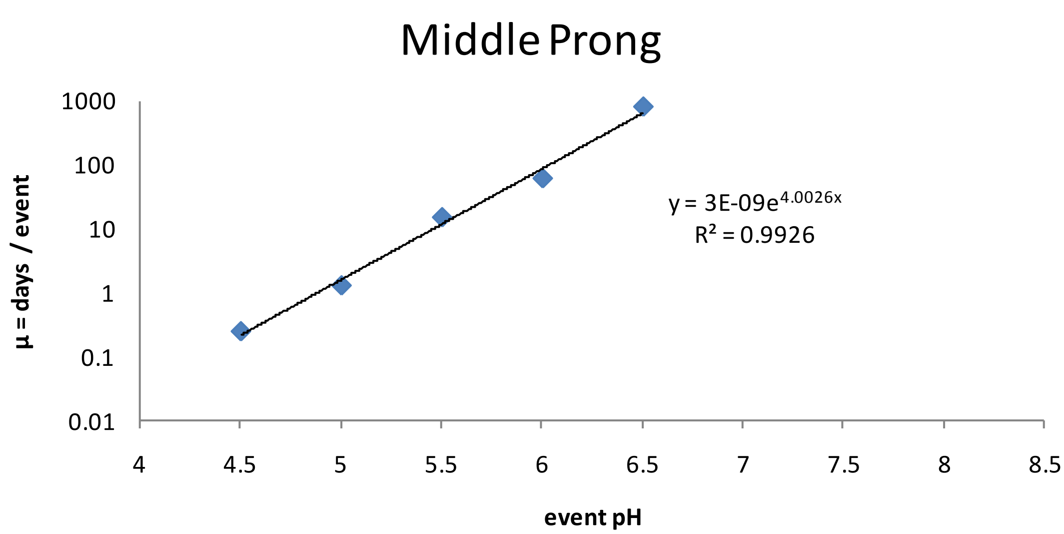


Figure 13. μ vs Event pH for Middle Prong Site.

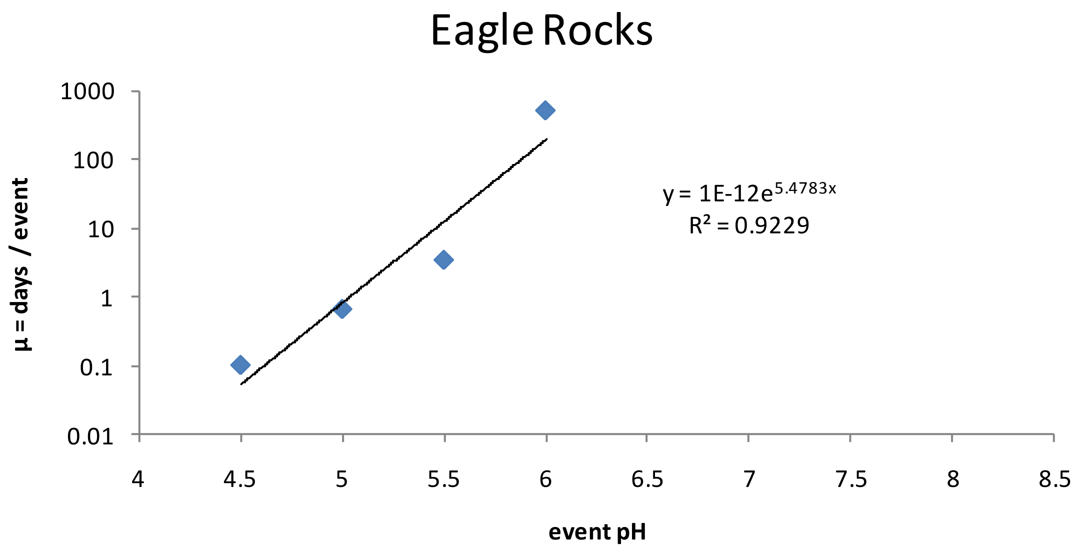


Figure 14. μ vs Event pH for Eagle Rocks Site

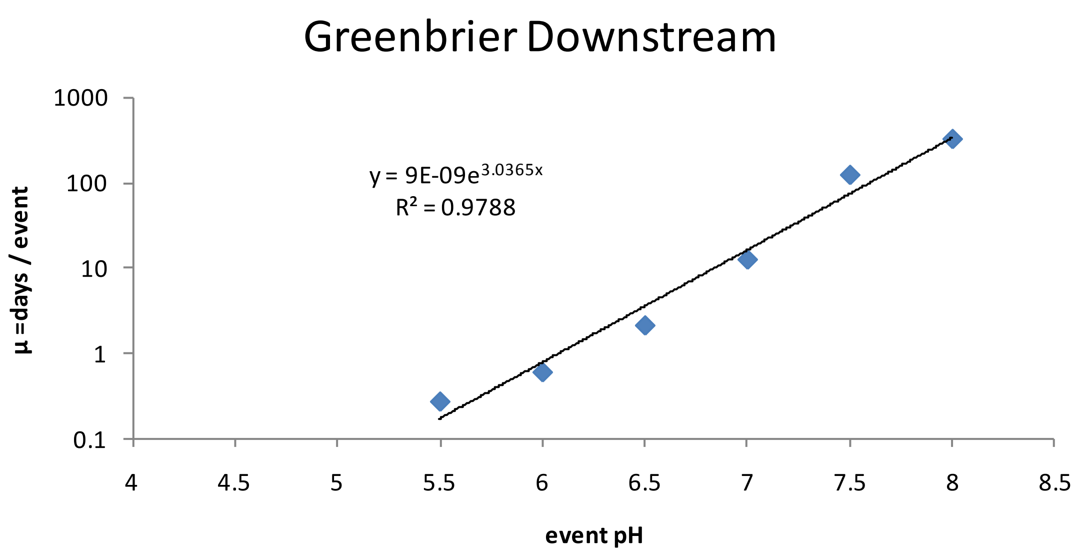


Figure 15. μ vs Event pH for Greenbrier 1 – Downstream Site

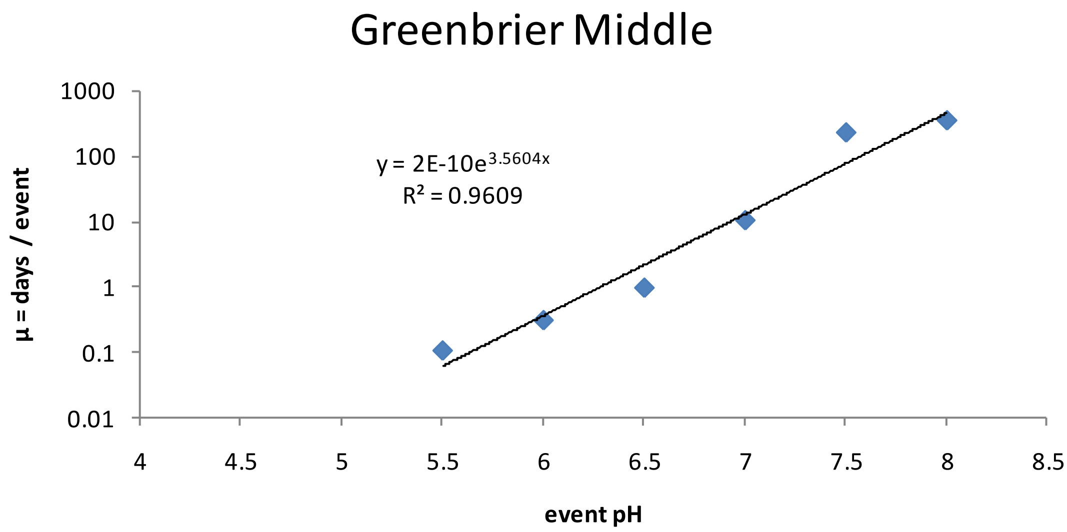


Figure 16. μ vs Event pH for Greenbrier 2 - Middle Site

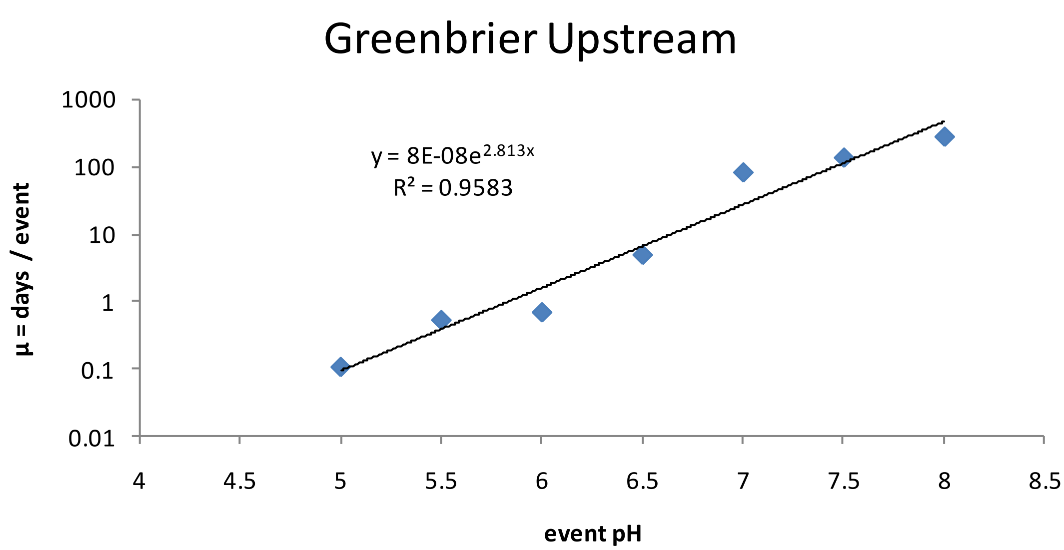


Figure 17. μ vs Event pH for Greenbrier 3 - Upstream Site

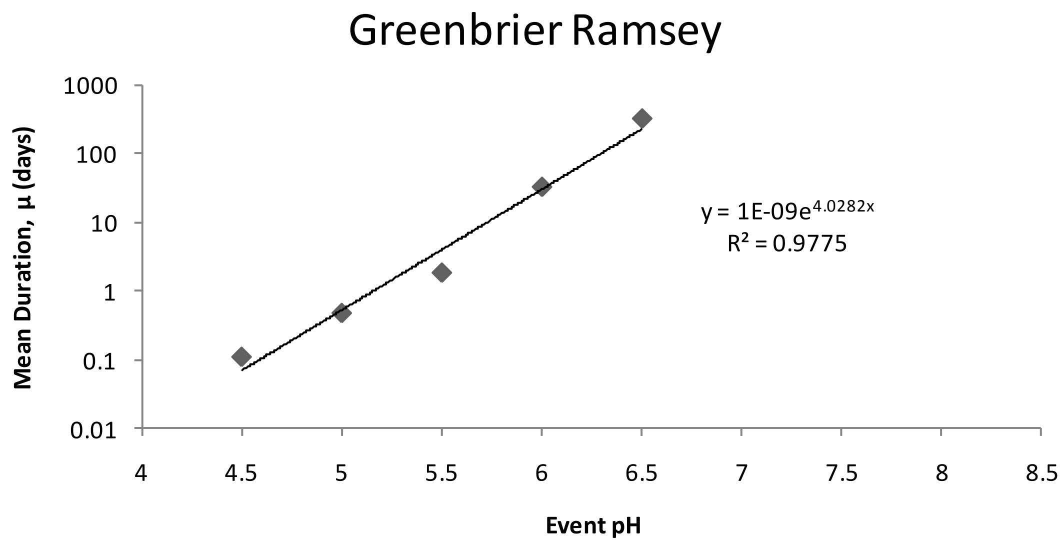


Figure 18. μ vs Event pH for Greenbrier 4 - Ramsey Site

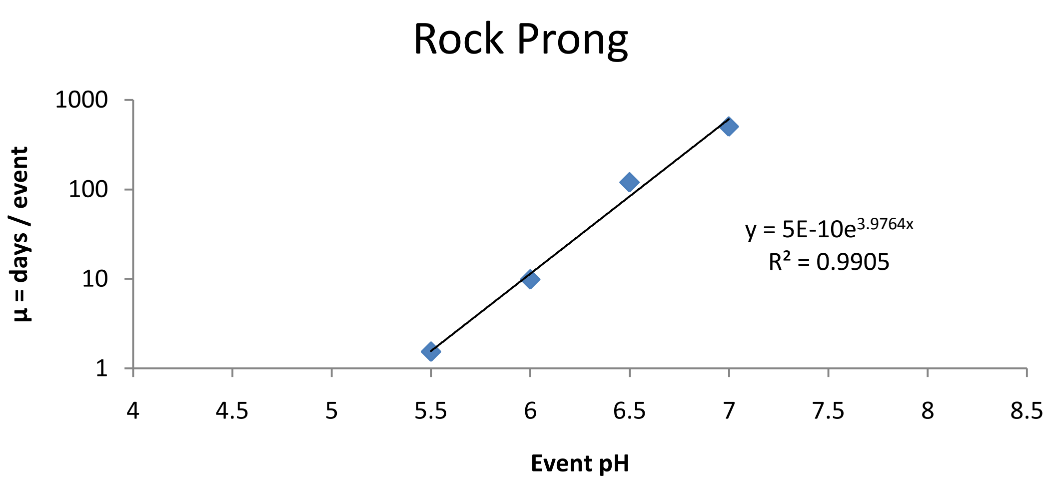


Figure 19. μ vs Event pH for Rock Prong Site

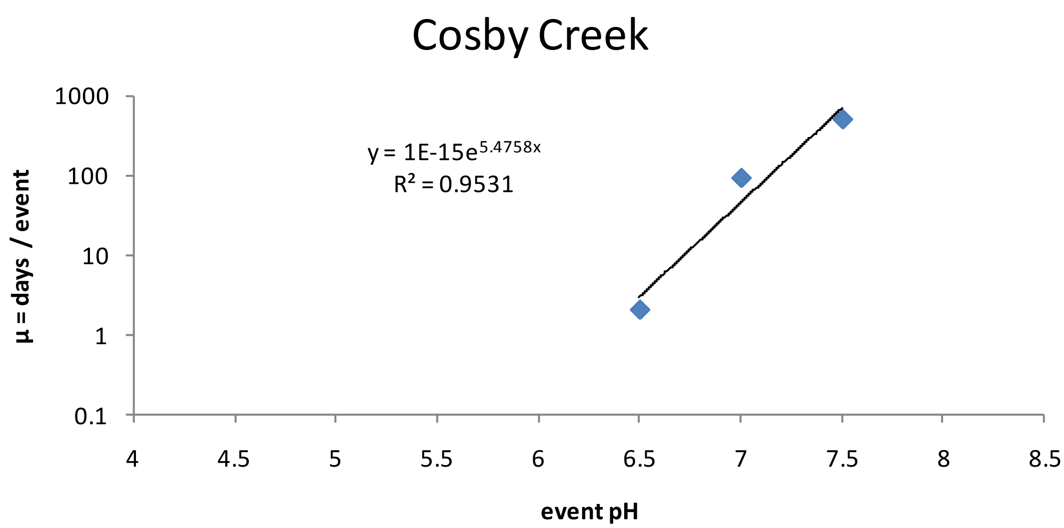


Figure 20. μ vs Event pH for Cosby Creek Site

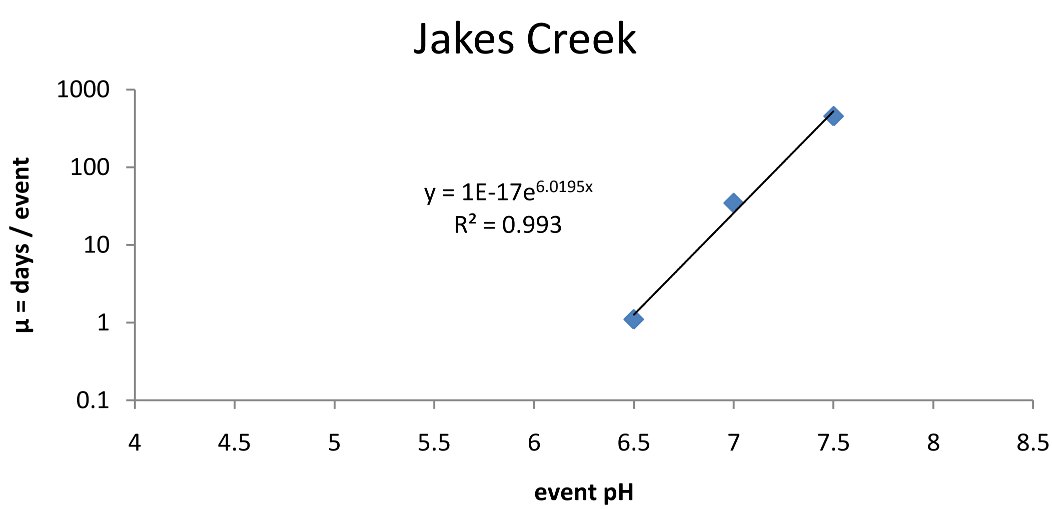


Figure 21. μ vs Event pH for Jakes Creek Site

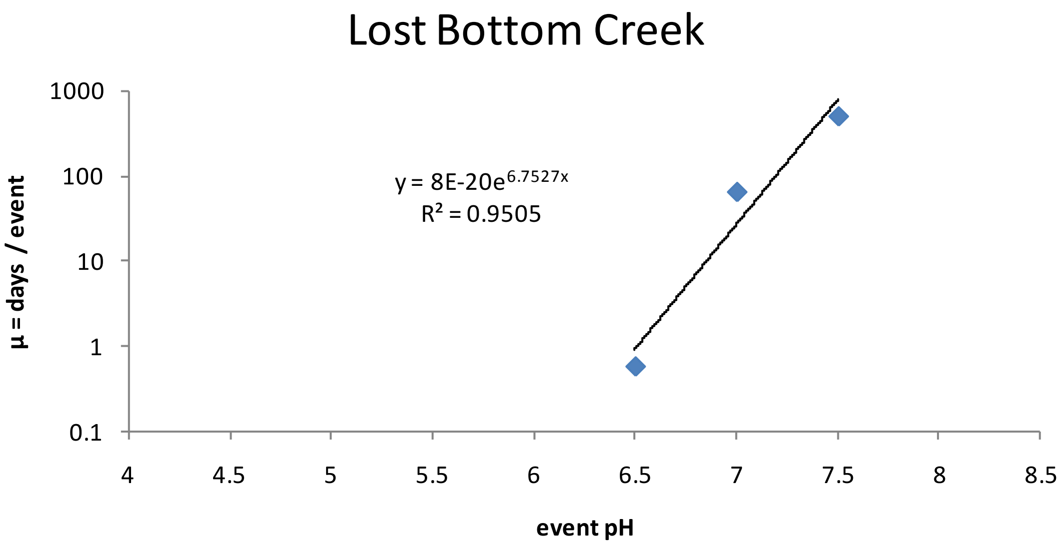


Figure 22. μ vs Event pH for Lost Bottom Creek Site

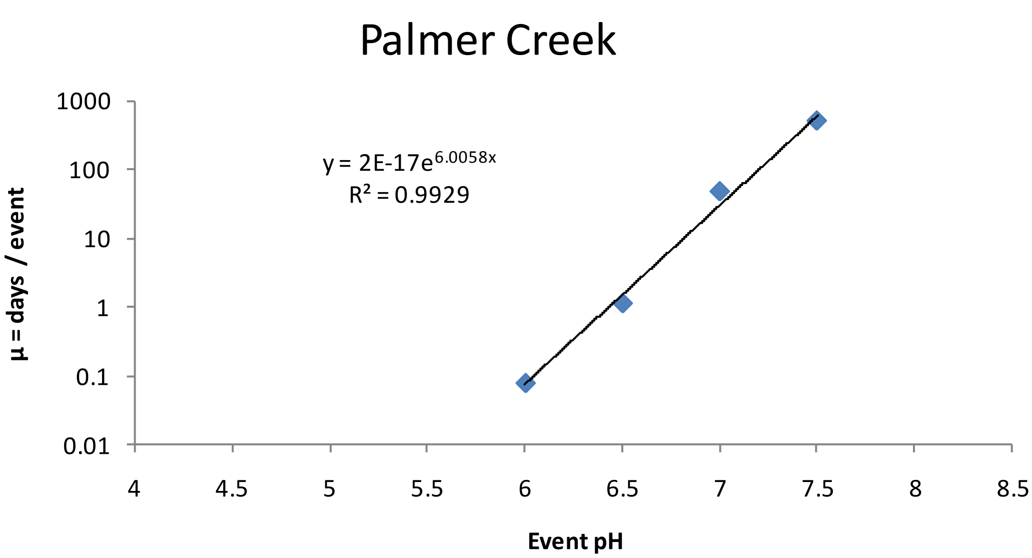


Figure 23. μ vs Event pH for Palmer Creek Site

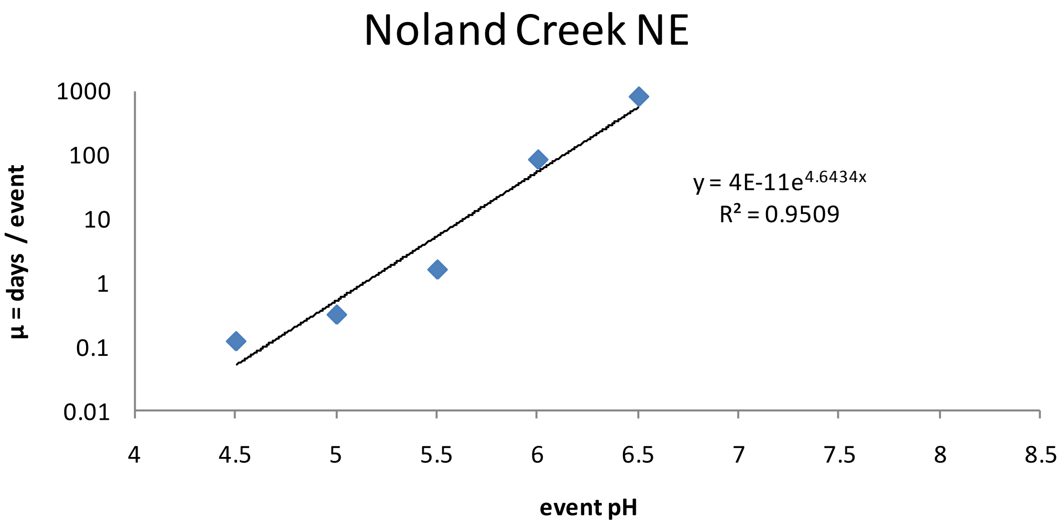


Figure 24. μ vs Event pH for Noland Creek - Northeast Site

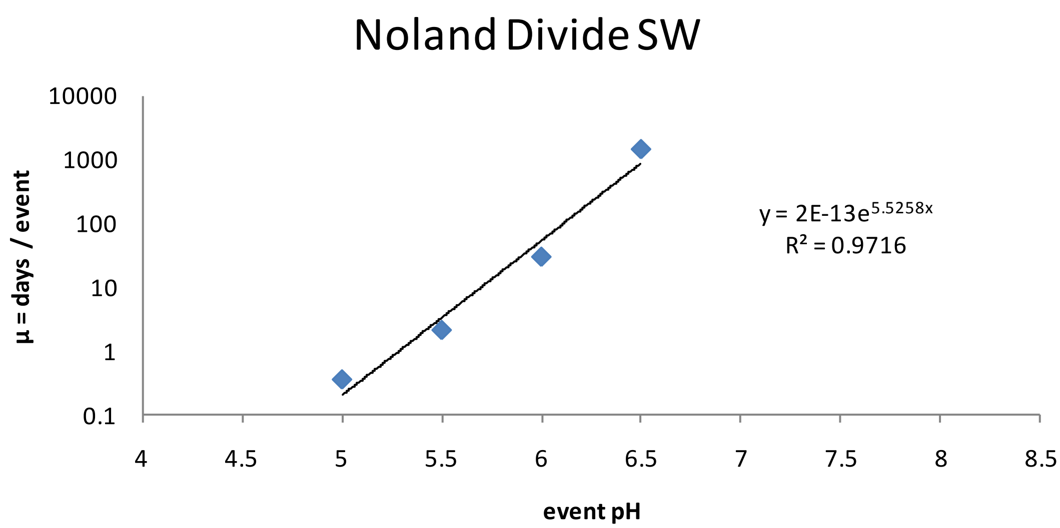


Figure 25. μ vs Event pH for Noland Divide - Southwest Site

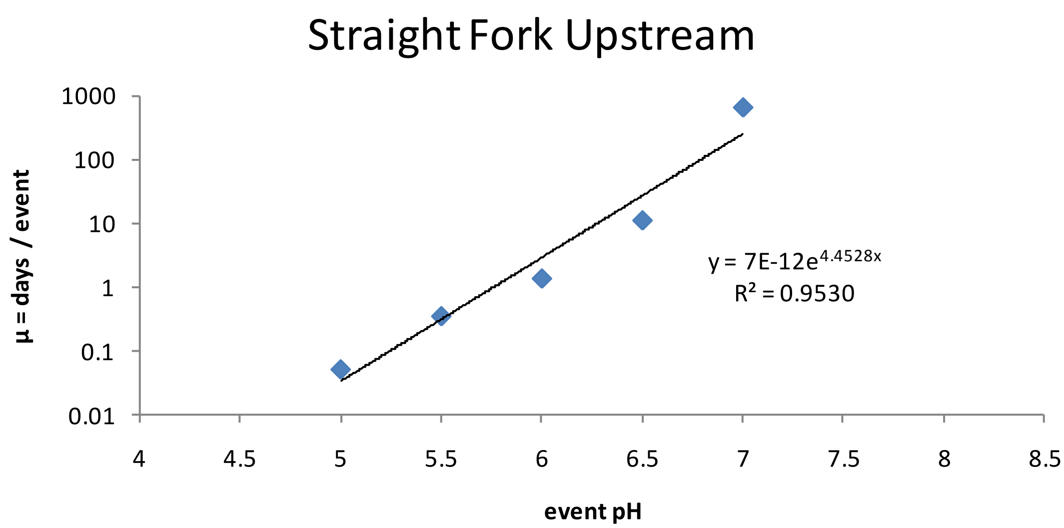


Figure 26. μ vs Event pH for Straight Fork - Upstream Site

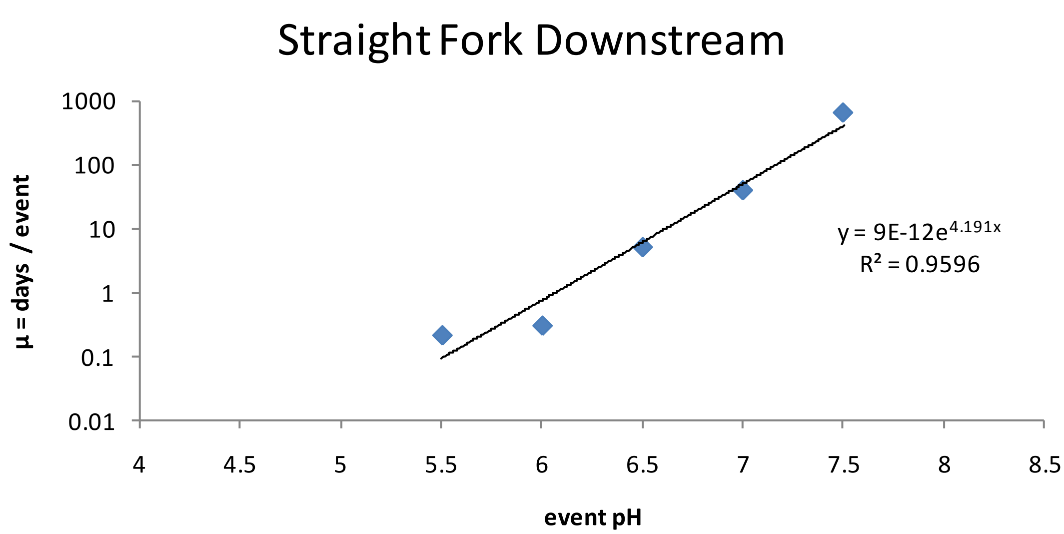


Figure 27. μ vs Event pH for Straight Fork - Downstream Site

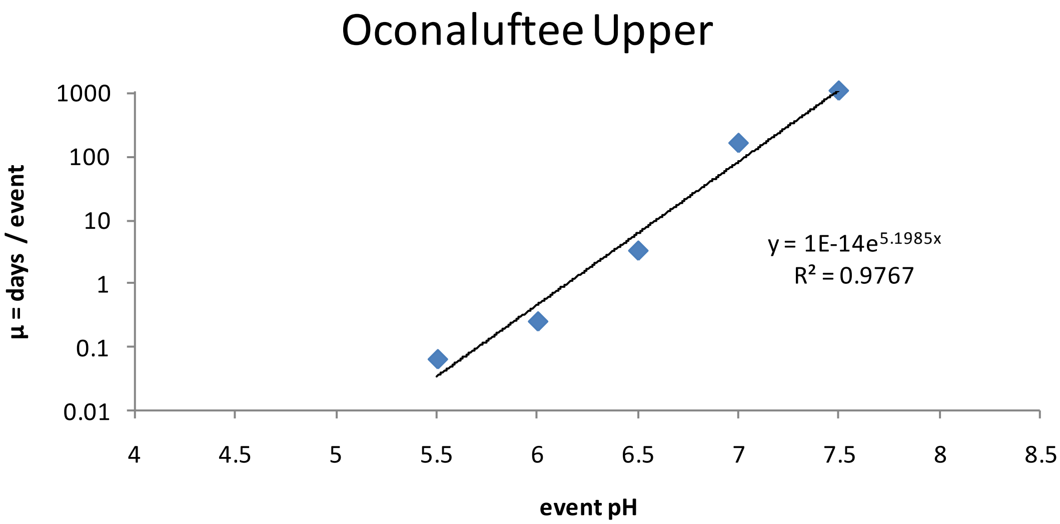


Figure 28. μ vs Event pH for Oconaluftee - Upstream Site

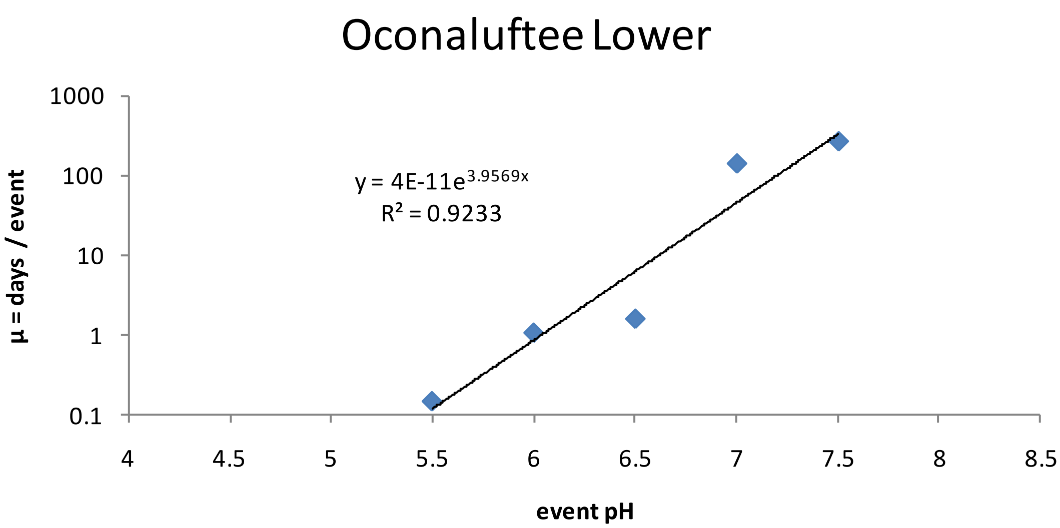


Figure 29. μ vs Event pH for Oconaluftee - Downstream Site

pH vs. μ Curves Used for Temporal Analysis at Noland Divide Watershed

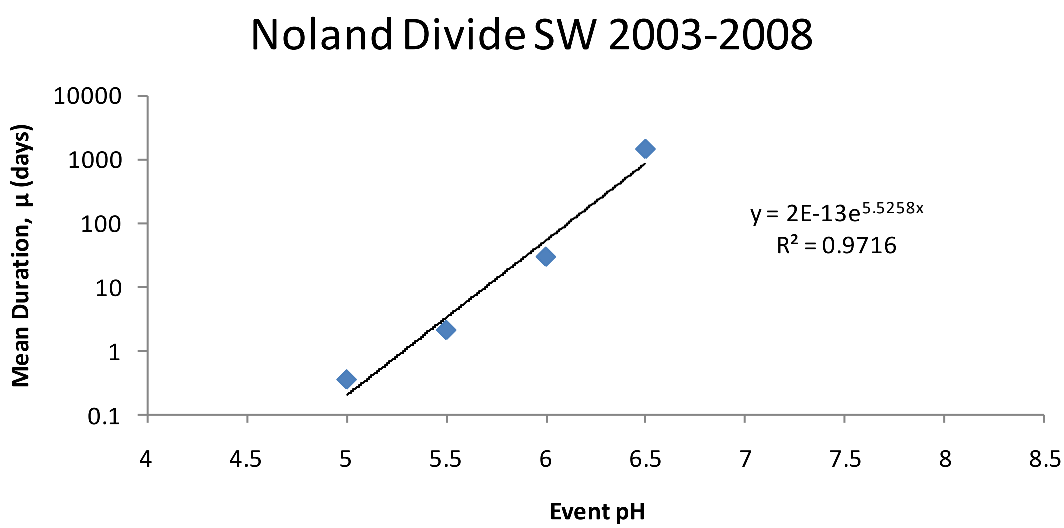


Figure 30. μ vs Event pH for NDW Site (All Four Years Of Data)

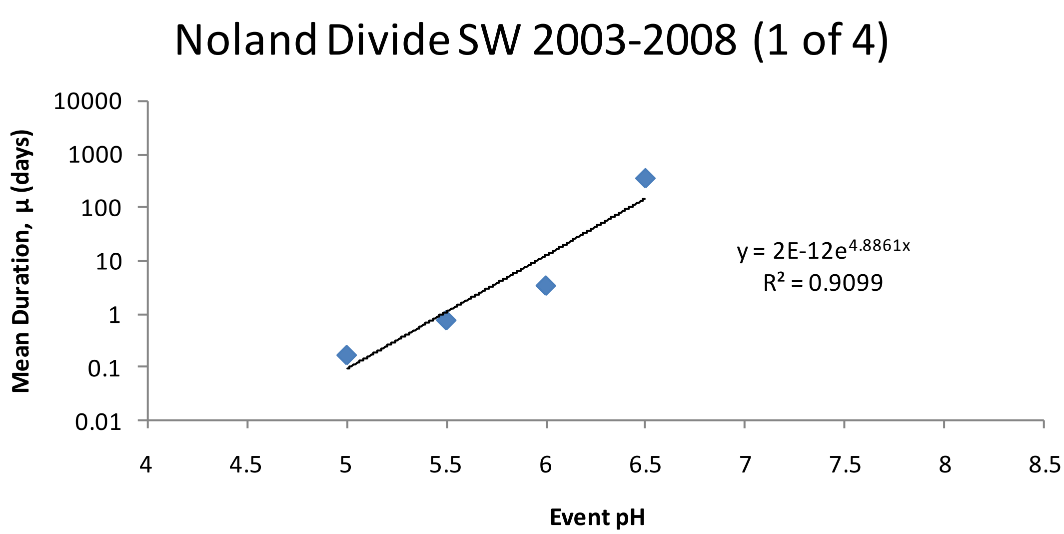


Figure 31. μ vs Event pH for NDW (Year One of Four)

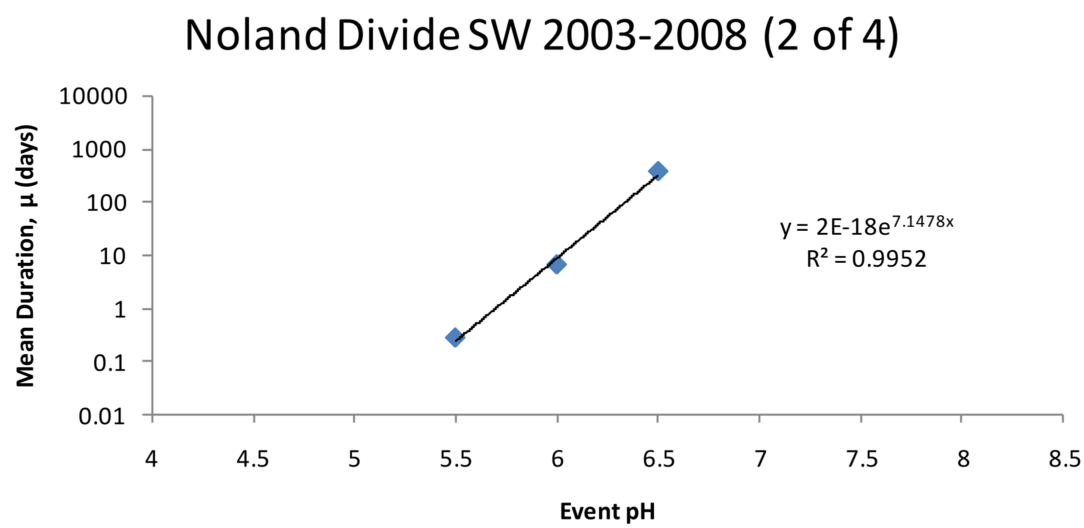


Figure 32. µ vs Event pH for NDW (Year Two of Four)

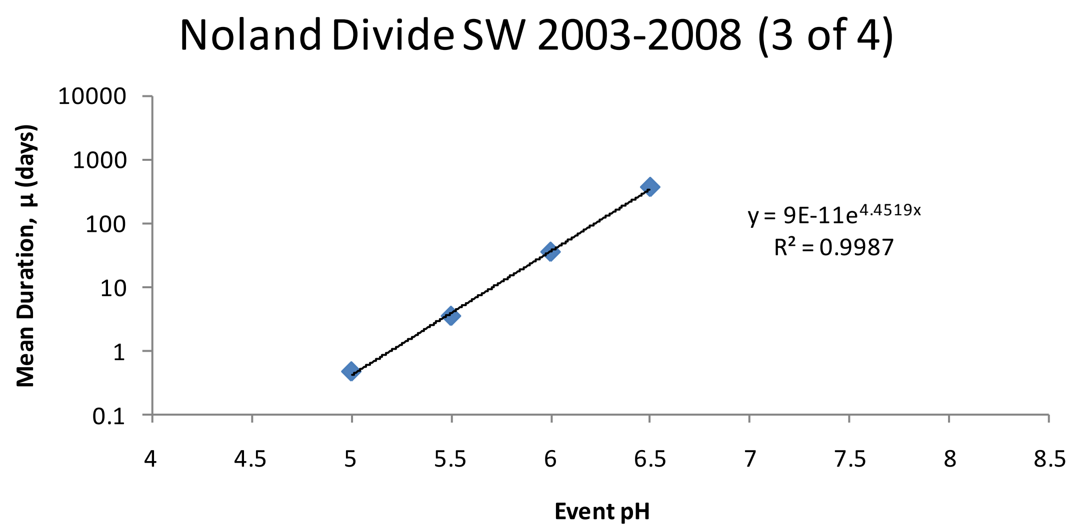


Figure 33. µ vs Event pH for NDW (Year Three of Four)

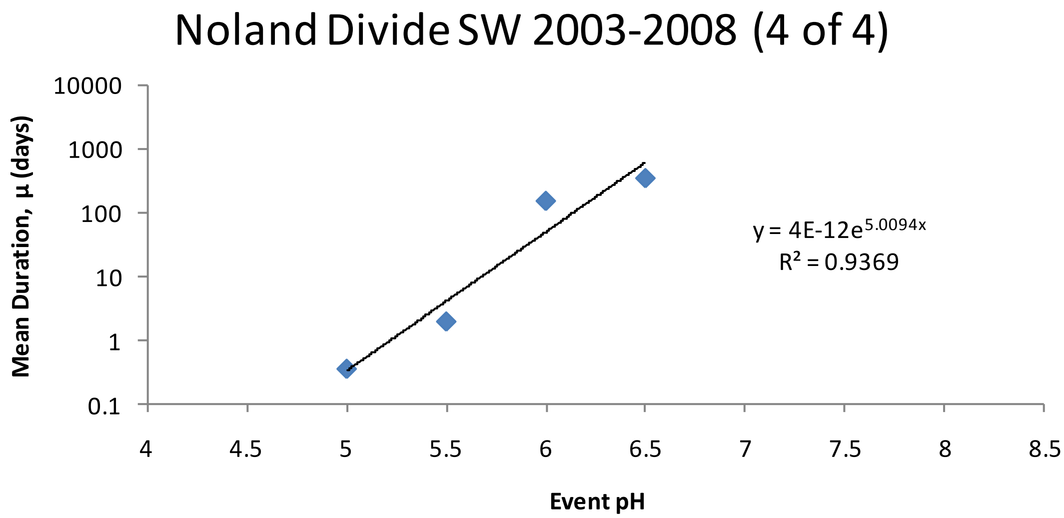


Figure 34. µ vs Event pH for NDW (Year Four of Four)

Missing Data Periods

Table 6. Missing Data Periods Greater Than One Day (Excluding NDW Sites)

	Ramsey Prong	Middle Prong	Eagle Rocks	Greenbrier-1 (Downstream)	Greenbrier-2 (Middle)	Greenbrier-3 (Upper)	Greenbrier-4 (Ramsey)	Rock Prong	Cosby Creek	Jakes Creek	Lost Bottom Creek	Palmer Creek	Straight Fork-Upper	Straight Fork-Lower	Oconaluftee River-Upper	Oconaluftee River-Lower
Date (From)	4/25/06	4/21/06	5/5/06	1/21/04	9/29/03	9/29/03	7/30/03	4/13/08	No Missing Data (>1 Day)	12/2/08	No Missing Data (>1 Day)	No Missing Data (>1 Day)	7/25/05	5/25/05	6/28/06	3/23/09
Date (To)	5/15/06	4/24/06	6/21/06	2/16/04	10/6/03	10/19/03	8/13/03	4/30/08		12/4/08			3/1/06	6/24/05	7/6/06	5/13/09
Duration (days)	20.3	3.5	47.3	26.0	7.2	20.1	13.6	17.5		1.8			219.0	30.2	7.8	50.8
Date (From)	7/28/06	5/5/06	10/29/06	5/29/04	12/6/03	10/27/03				5/8/09			5/13/06	9/25/05	9/13/06	
Date (To)	11/3/06	5/15/06	1/4/07	6/9/04	12/12/03	11/4/03				5/9/09			5/18/06	3/1/06	10/11/06	
Duration (days)	97.9	10.3	67.0	10.7	5.9	8.1				1.4			5.0	157.1	27.9	
Date (From)	12/21/06	9/22/06				11/18/03				6/29/09			12/3/06	4/28/06	7/5/07	
Date (To)	1/4/07	10/5/06				12/9/03				7/7/09			12/6/06	5/11/06	7/18/07	
Duration (days)	13.8	12.7				20.7				7.9			3.1	13.1	12.7	
Date (From)	6/17/07	9/7/07				1/21/04								1/7/07	8/5/07	
Date (To)	2/14/08	10/14/07				3/8/04								1/11/07	11/14/07	
Duration (days)	242.0	36.9				47.1								4.6	100.9	
Date (From)	3/18/08					6/3/04								1/16/07	1/21/08	
Date (To)	3/21/08					6/23/04								1/18/07	2/25/08	
Duration (days)	3.4					19.8								1.7	35.5	
Date (From)	7/20/08														7/31/08	
Date (To)	7/23/08														12/10/08	
Duration (days)	3.1														131.8	
Date (From)															12/29/08	
Date (To)															12/31/08	
Duration (days)															2.0	
Date (From)															2/19/09	
Date (To)															5/13/09	
Duration (days)															83.1	
Total Missing Days	380.5	63.4	114.3	36.7	13.2	115.8	13.6	17.5	0.0	11.0	0.0	0.0	227.0	206.6	401.6	50.8
Time Range	2/17/06 - 8/14/08	2/28/06 - 7/3/08	3/9/06 - 12/11/07	8/25/03 - 9/15/04	9/3/03 - 8/28/04	8/20/03 - 9/15/04	6/19/03 - 5/25/04	4/2/08 - 9/5/09	4/24/08 - 9/8/09	5/29/08 - 9/7/09	4/23/08 - 9/4/09	4/23/08 - 9/4/09	10/27/04 - 3/10/07	10/27/04 - 3/2/07	5/25/05 - 7/13/09	9/4/08 - 7/13/09
Duration (days)	520.0	788.7	527.8	323.6	339.8	275.1	327.0	503.7	501.3	453.0	499.2	499.4	636.6	649.1	1078.3	261.4
No. of Data Points	49811	75615	50591	30714	32243	25071	31384	48349	48112	43492	47919	47935	61077	62211	103485	25092

Table 7. Missing Data Periods Greater Than One Day (NDW SW Site Only)

	1st Year	2nd Year	3rd Year	4th Year
Date (From)	2/3/04	11/24/04	3/30/06	9/13/07
Date (To)	2/5/04	12/1/04	5/9/06	9/20/07
Duration (days)	2.0	7.0	39.9	6.8
Date (From)	2/7/04	12/14/04	6/19/06	10/5/07
Date (To)	2/12/04	12/16/04	7/3/06	10/24/07
Duration (days)	4.7	2.3	14.3	18.9
Date (From)	9/26/04	12/18/04	8/30/06	12/14/07
Date (To)	9/30/04	12/21/04	9/15/06	12/21/07
Duration (days)	3.6	2.9	16.0	7.1
Date (From)		5/4/05	9/21/06	12/24/07
Date (To)		7/27/05	9/29/06	12/26/07
Duration (days)		84.1	7.7	2.2
Date (From)		8/6/05	12/2/06	1/19/08
Date (To)		8/9/05	1/2/07	1/24/08
Duration (days)		2.7	31.1	5.2
Date (From)		8/25/05	1/14/07	2/27/08
Date (To)		8/31/05	1/18/07	3/6/08
Duration (days)		5.8	4.3	8.1
Date (From)			2/3/07	3/24/08
Date (To)			2/8/07	4/3/08
Duration (days)			4.7	10.1
Date (From)			5/14/07	4/12/08
Date (To)			8/1/07	4/18/08
Duration (days)			78.8	6.5
Date (From)				9/5/08
Date (To)				9/25/08
Duration (days)				19.8
Total Missing Days	10.2	104.9	196.7	84.7
Time Range	10/22/03 - 11/1/04	11/1/04 - 2/24/06	2/24/06 - 9/13/07	9/13/07 - 12/2/08
Duration (days)	359.9	360.7	361.9	360.5
No. of Data Points	34560	34560.0	34560	34560

Table 8. Missing Data Periods Greater Than One Day (NDW NE Site Only)

	NDW NE
Date (From)	12/31/03
Date (To)	2/12/04
Duration (days)	42.5
Date (From)	5/31/04
Date (To)	6/9/04
Duration (days)	9.3
Date (From)	9/1/04
Date (To)	10/21/04
Duration (days)	49.7
Date (From)	11/24/04
Date (To)	12/1/04
Duration (days)	7.3
Date (From)	12/14/04
Date (To)	12/16/04
Duration (days)	2.3
Date (From)	12/18/04
Date (To)	12/21/04
Duration (days)	2.9
Date (From)	7/13/05
Date (To)	10/20/05
Duration (days)	98.7
Date (From)	3/30/06
Date (To)	4/11/06
Duration (days)	12.1
Date (From)	7/15/06
Date (To)	7/20/06
Duration (days)	4.8
Date (From)	7/27/06
Date (To)	10/5/06
Duration (days)	69.5
Total Missing Days	229.5
Time Range	11/13/03 - 12/10/06
Duration (days)	809.8
No. of Data Points	77785

APPENDIX C: GIS Figures

Discussion of watershed variables:

Admittedly, the listed predictor variables are only a few of the possible predictor variables that could be integrated into this study. The variables included were known to have strong relationships to baseflow pH in the GRSM, prompting their use in this study of stormflow pH. Previous studies have incorporated limestone bedrock as a predictor of stream chemistry due to associated ANC increases [Jackson, 2006]. Limestone is only prevalent within certain regions of the GRSM and does not include monitoring sites used in this study. Basin soil type is difficult to classify per watershed, other than a majority soil type. Furthermore, soil types have been shown to be strongly correlated with elevation. Similarly, overstory vegetation, i.e., tree cover, is correlated with elevational zones, and although diverse in species, forests are the dominant land cover in the study area so comparison with non-forested basins is not possible. Although they are not physical watershed parameters, baseflow ANC and baseflow pH levels could provide further understanding of watershed response from episodic acidification.

Flow rate and precipitation were also not assessed for significance. Stream flow rates are not known for individual monitoring sites and mean daily flow does not adequately represent stormflow discharge. Similarly, site specific precipitation data is not available at such a small spatial scale and would likely be relatively static over the small study area. To compare these site specific data is needed.

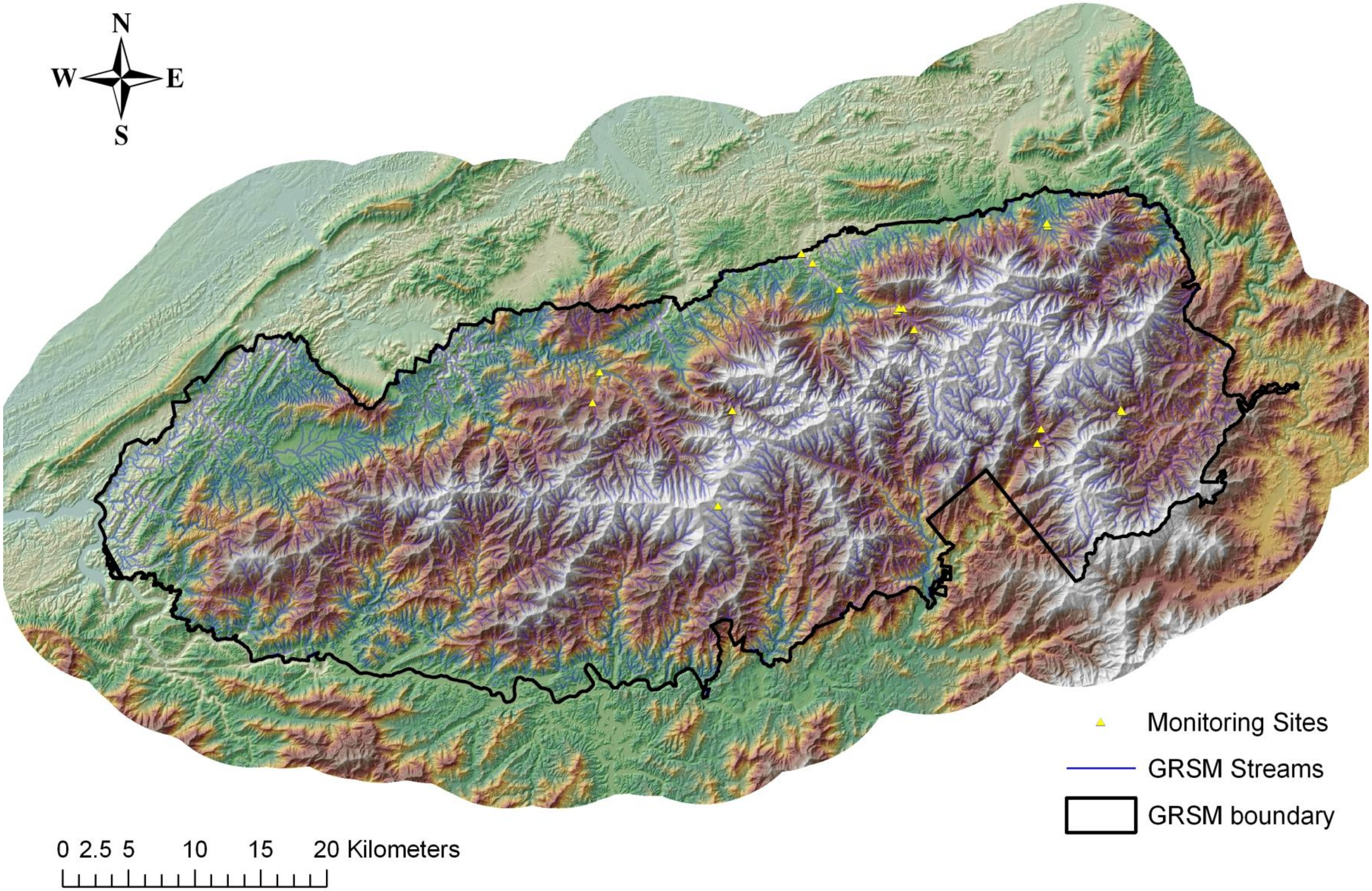


Figure 35. Map Depicting Topographical Features In And Surrounding GRSM

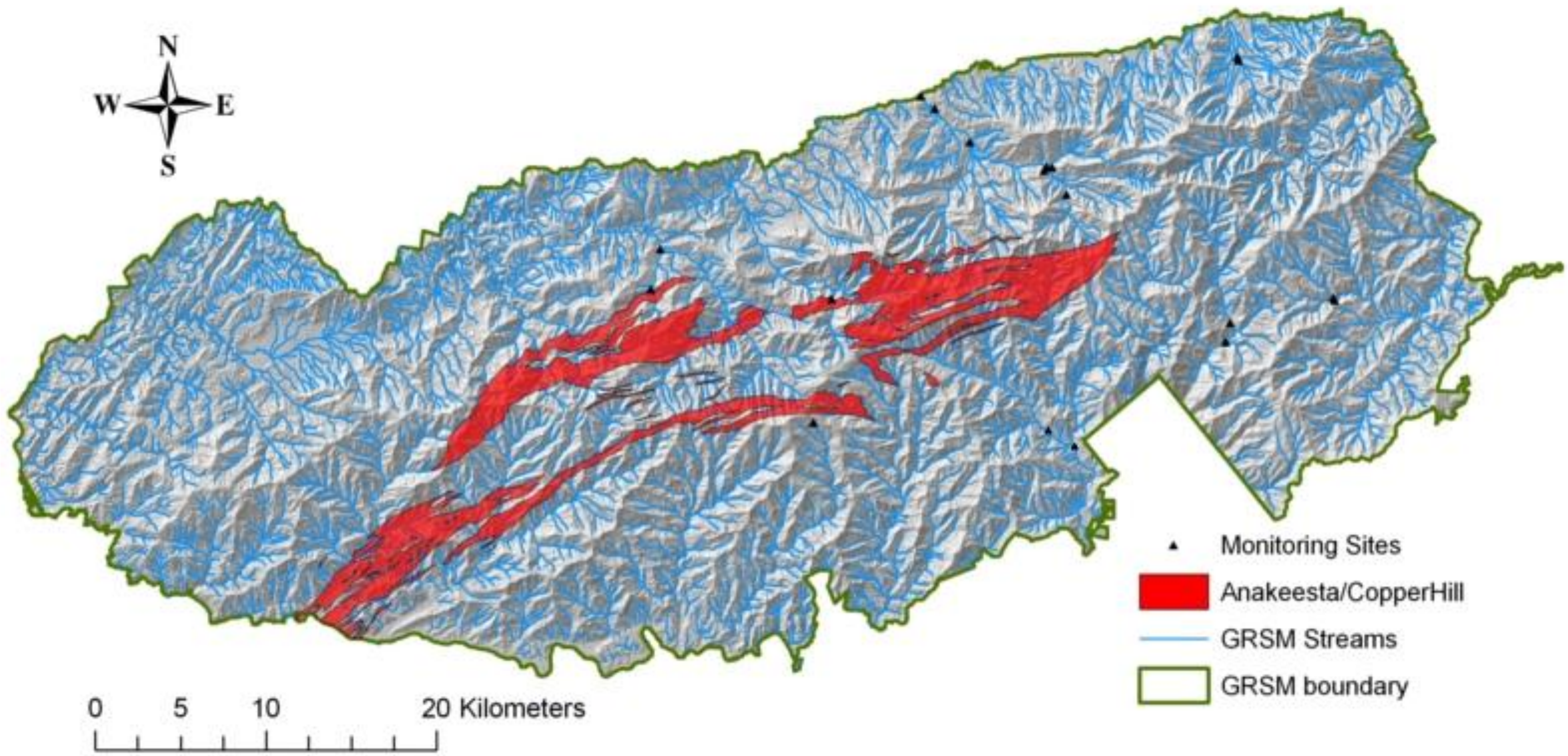


Figure 36. Anekeesta And Copperhill Formation In GRSM.

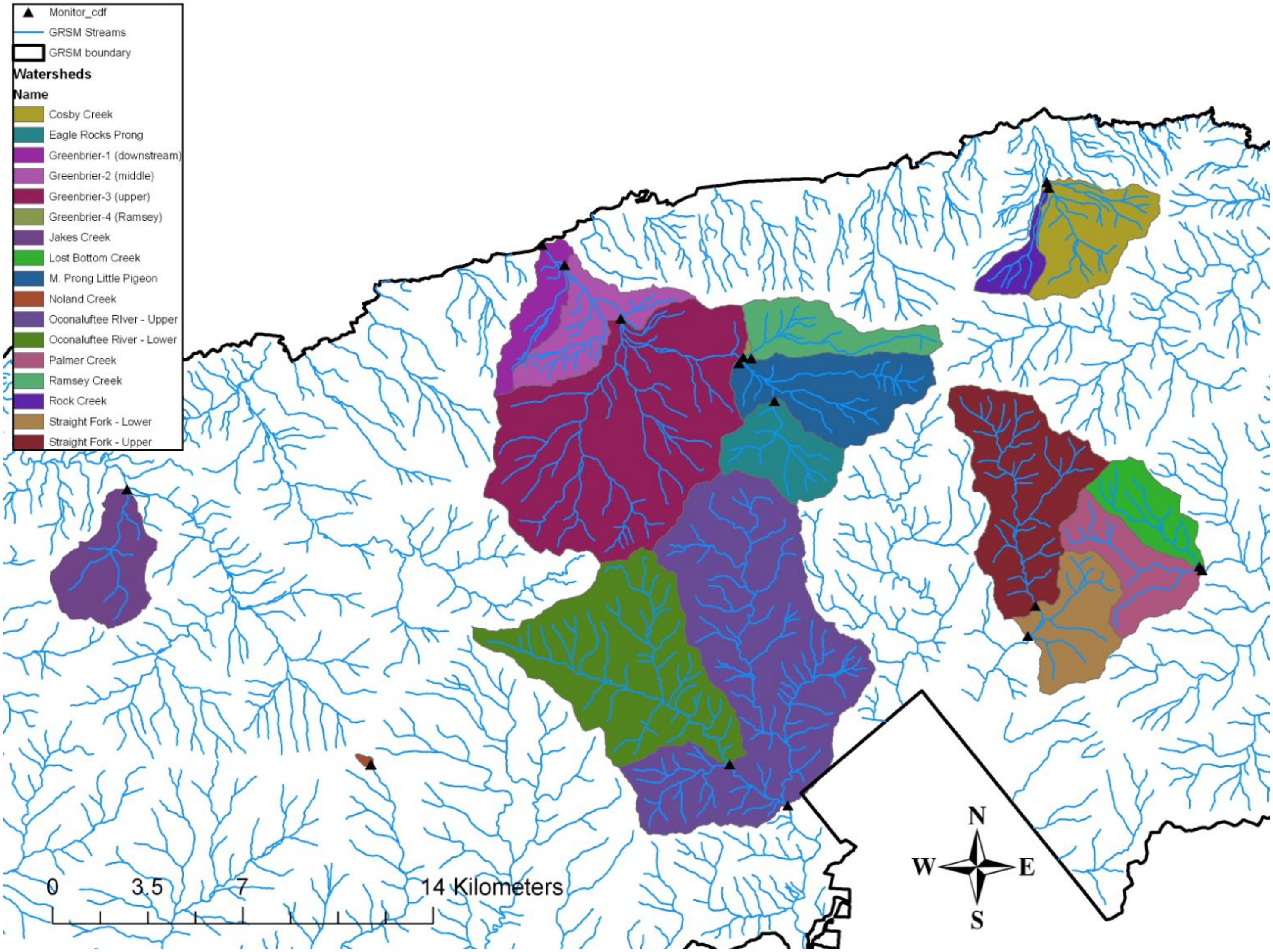


Figure 37. Drainage Area For Monitoring Sites Within GRSM. Several Watershed Include The Attributes Of Smaller Sub-Watersheds

APPENDIX D: Statistical Analyses

All statistical analyses were computed using Minitab or SPSS Statistical software. With spatial land statistics that summarize basin attributes and characteristic equations (e.g. Equation 5) developed to characterize pH episodes, an analysis of statistical relationships is feasible.

To note, Jakes Creek site seems to be an outlier in this study. Per discussions with Neff [2009], the data at Jakes Creek may be inaccurate due to poor site conditions and technical malfunctions with sonde equipment during the monitoring period. Although eighteen monitoring sites were included for this study, future work may exclude this site from analyses.

Test For Normality

Kolmogorov-Smirnov test with observations of the normal probability plot were used to assess the normality of the characteristic equations.

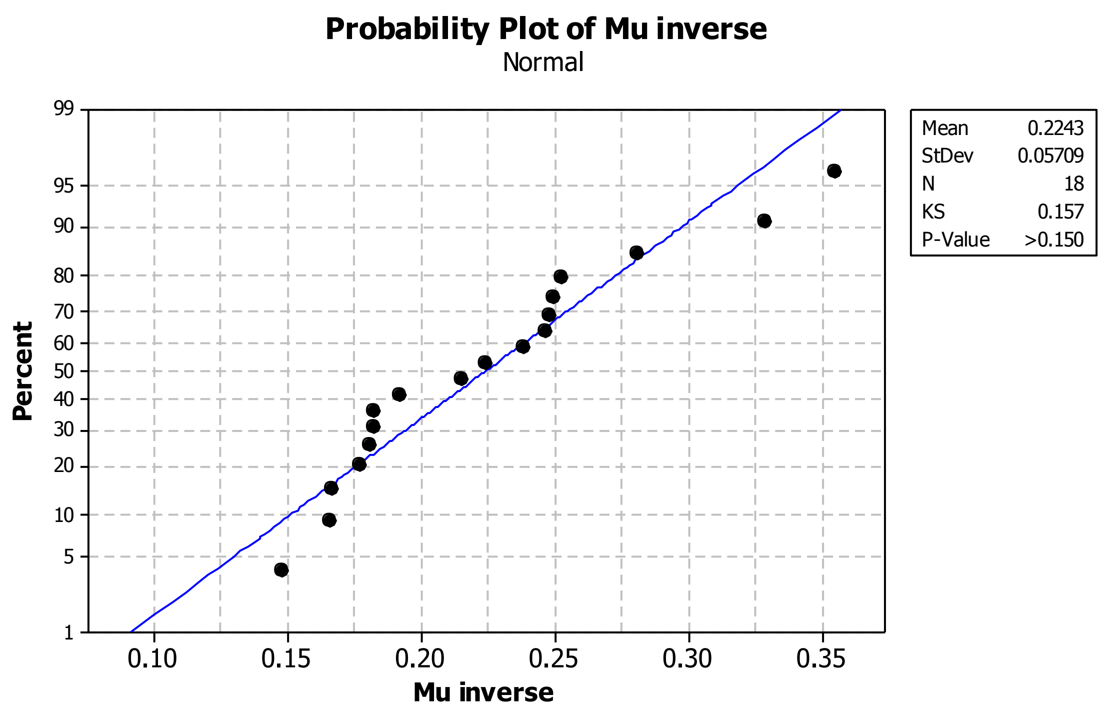


Figure 38. Kolmogorov-Smirnov Test For Normality.

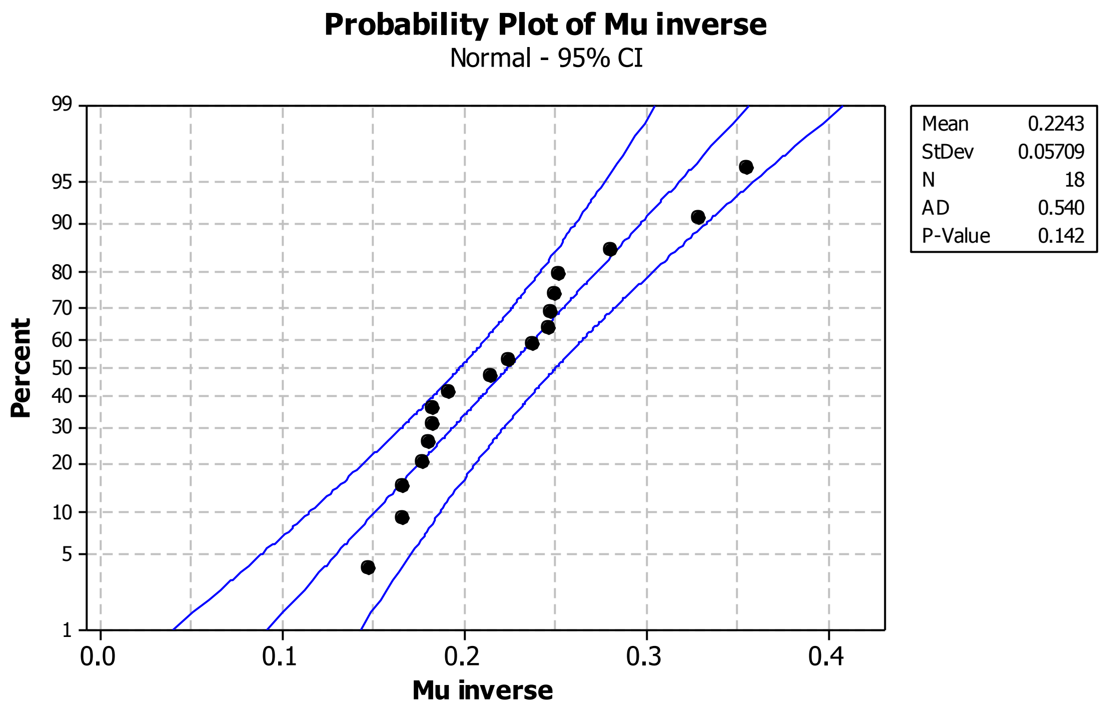


Figure 39. Normal Probability Plot Showing 95% Confidence Intervals

Correlation Analysis

Correlation analysis was used to determine significant linear relationships between basin characteristics and the characteristic curves. Pearson's R coefficient, Kendall's tau coefficient, and Spearman's Rho coefficient were assessed.

Table 9. Correlation Analysis Table for Pearson, Kendall's tau, and Spearman's rho Coefficients.

	Area (km^2)	SSS (km2)	SSS (% Coverage)	Maximum Elevation (m)	Elevation Range (m)	Average Elevation (m)	Sonde Site Elevation (m)	Maximum Slope (%)	Slope Range (%)	Average Slope (%)
Pearson coefficient	0.761	0.475	0.026	0.662	0.648	-0.231	-0.451	0.552	0.505	0.269
p-Value	< 0.001	0.047	0.917	0.003	0.004	0.356	0.06	0.018	0.033	0.28
Kendall's tau coefficient	0.516	0.365	0.218	0.615	0.647	-0.19	-0.438	0.588	0.599	0.242
p-Value	0.003	0.053	0.242	0.001	< 0.001	0.272	0.011	0.001	0.001	0.161
Spearman's rho coefficient	0.684	0.489	0.245	0.793	0.804	-0.205	-0.517	0.726	0.774	0.319
p-Value	0.002	0.04	0.328	< 0.001	< 0.001	0.414	0.028	0.001	< 0.001	0.197

Linear Regression

Least squares linear regression was employed to model relationships between pH event duration slope as the independent variable and watershed attributes as dependent variables.

Linear Regression – S_μ vs. Area

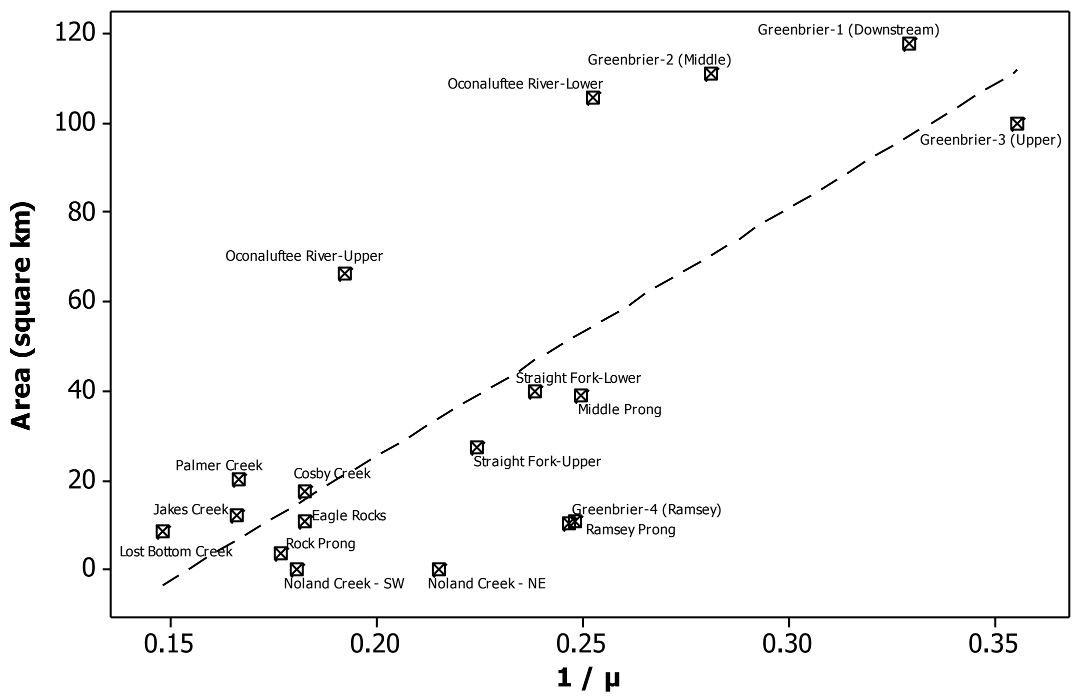


Figure 40. Least Squares Linear Regression For Area (km2) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = 0.184 + 0.00104 Area (km2)

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.18392	0.01247	14.75	0.000	
Area (km2)	0.0010401	0.0002220	4.69	0.000	1.000

S = 0.0382055   R-Sq = 57.8%   R-Sq(adj) = 55.2%

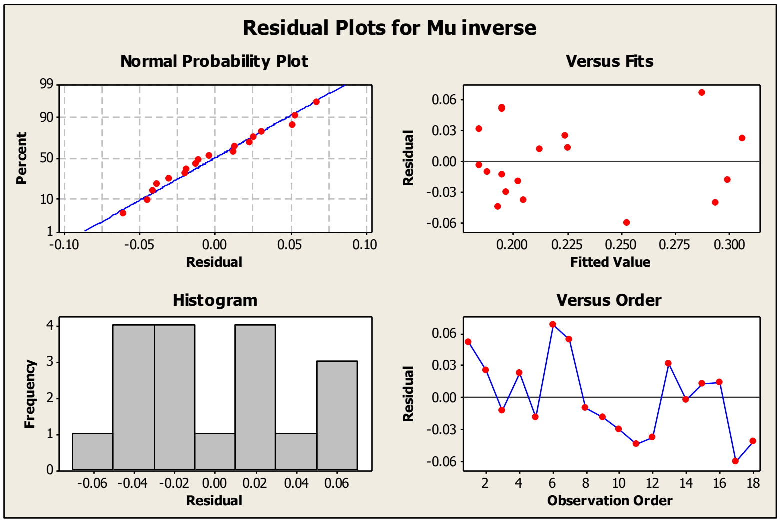
PRESS = 0.0302744   R-Sq(pred) = 45.35%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.032047	0.032047	21.95	0.000
Residual Error	16	0.023355	0.001460		

Total 17 0.055401

Durbin-Watson statistic = 1.25821



Linear Regression – S_{μ} vs. Maximum Elevation

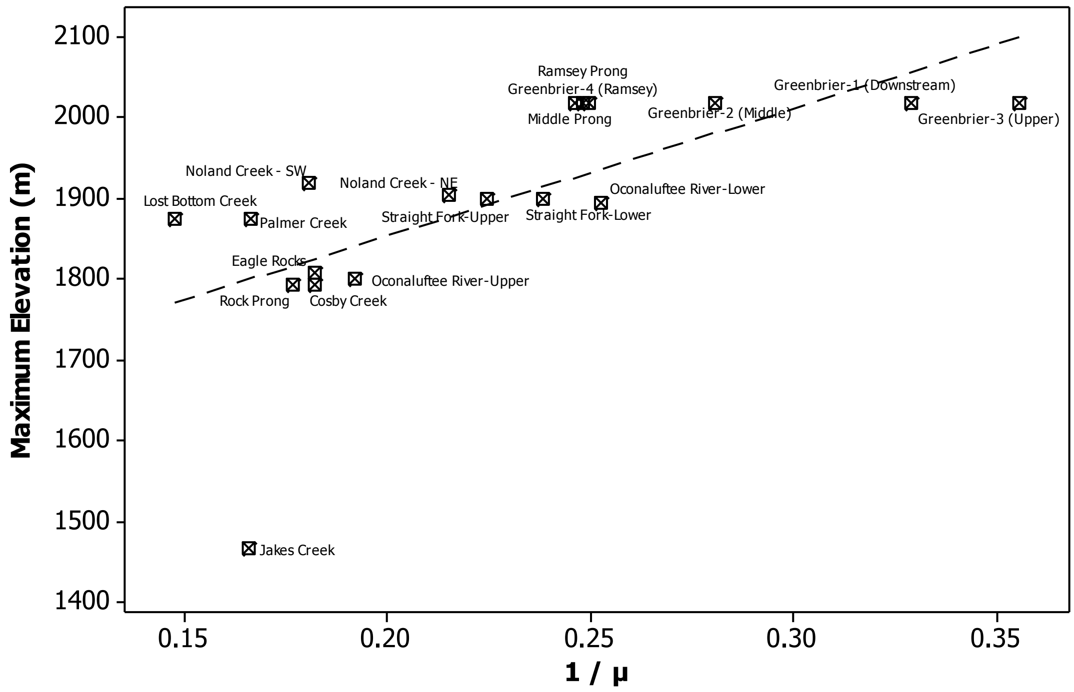


Figure 41. Least Squares Linear Regression For Maximum Elevation (m) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = - 0.301 + 0.000278 MAX_Elev

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.3006	0.1488	-2.02	0.060	
MAX_Elev	0.00027758	0.00007849	3.54	0.003	1.000

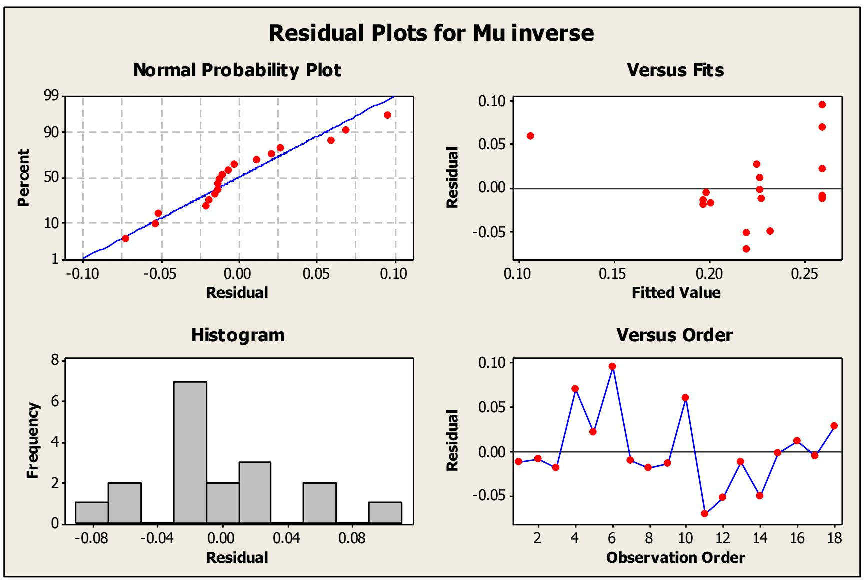
S = 0.0440848   R-Sq = 43.9%   R-Sq(adj) = 40.4%

PRESS = 0.0591687   R-Sq(pred) = 0.00%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.024306	0.024306	12.51	0.003
Residual Error	16	0.031096	0.001943		
Total	17	0.055401			

Durbin-Watson statistic = 1.85968



An outlier seems to exist in the regression plot for maximum elevation vs. pH event duration slope. Removing Jakes Creek here does improve the regression fit for maximum elevation, but removing outliers in this study was not deemed appropriate and all eighteen monitoring sites were included for statistical analyses.

Linear Regression – S_μ vs. Elevation Range

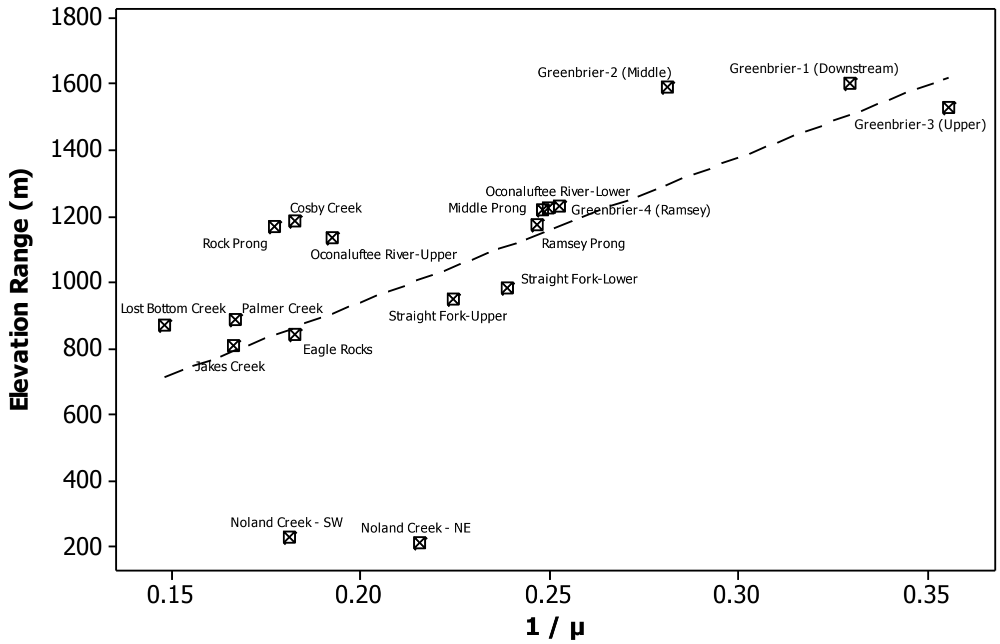


Figure 42. Least Squares Linear Regression For Elevation Range (m) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = 0.124 + 0.000096 RANGE_Elev

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.12410	0.03131	3.96	0.001	
RANGE_Elev	0.00009574	0.00002816	3.40	0.004	1.000

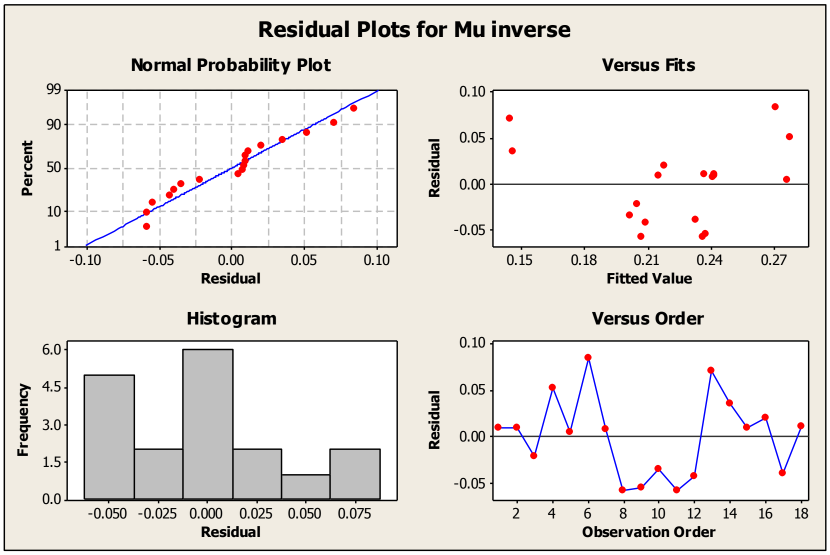
S = 0.0448338   R-Sq = 41.9%   R-Sq(adj) = 38.3%

PRESS = 0.0461162   R-Sq(pred) = 16.76%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.023240	0.023240	11.56	0.004
Residual Error	16	0.032161	0.002010		
Total	17	0.055401			

Durbin-Watson statistic = 1.49226


Residual Plots for Mu inverse
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Residual
Versus Order
Residual
Observation Order

Linear Regression – S_μ vs. Maximum Slope

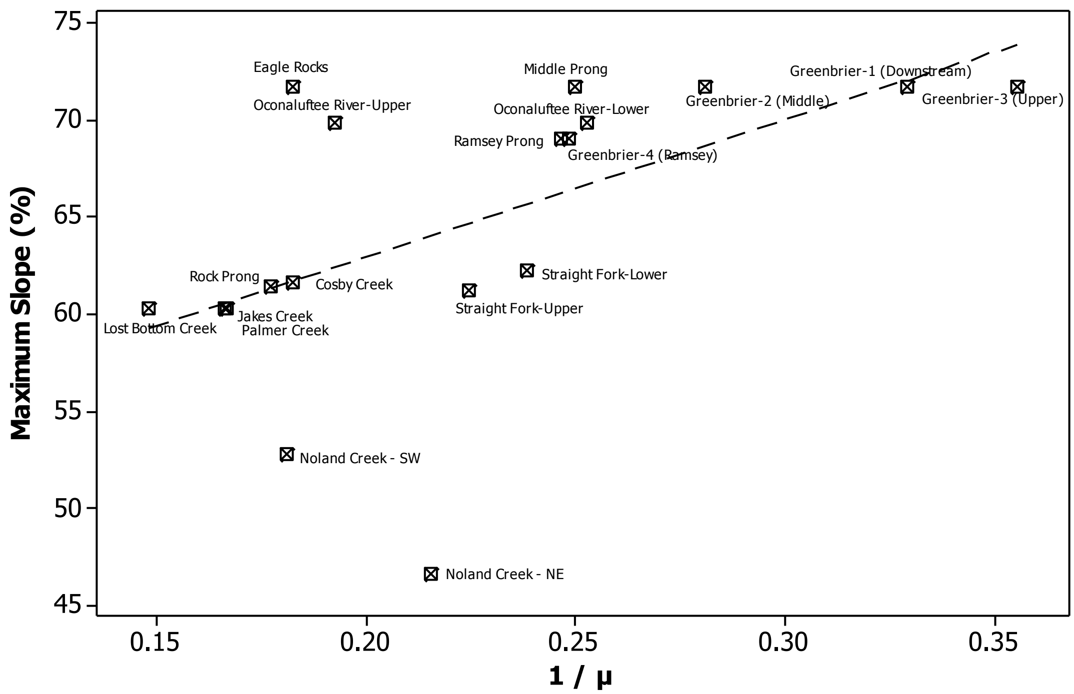


Figure 43. Least Squares Linear Regression For Maximum Slope (%) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = - 0.055 + 0.00432 MAX_Slope

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.0547	0.1060	-0.52	0.613	
MAX_Slope	0.004316	0.001630	2.65	0.018	1.000

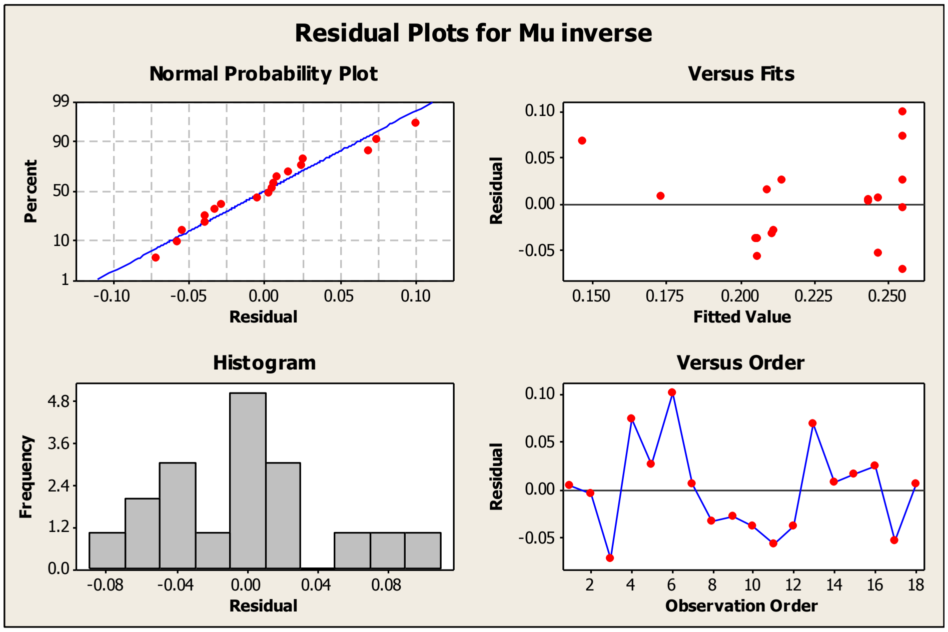
S = 0.0490663   R-Sq = 30.5%   R-Sq(adj) = 26.1%

PRESS = 0.0553917   R-Sq(pred) = 0.02%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.016881	0.016881	7.01	0.018
Residual Error	16	0.038520	0.002407		
Total	17	0.055401			

Durbin-Watson statistic = 1.83824


Residual Plots for Mu inverse
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Residual
Versus Order
Residual
Observation Order

Linear Regression – S_{μ} vs. Slope Range

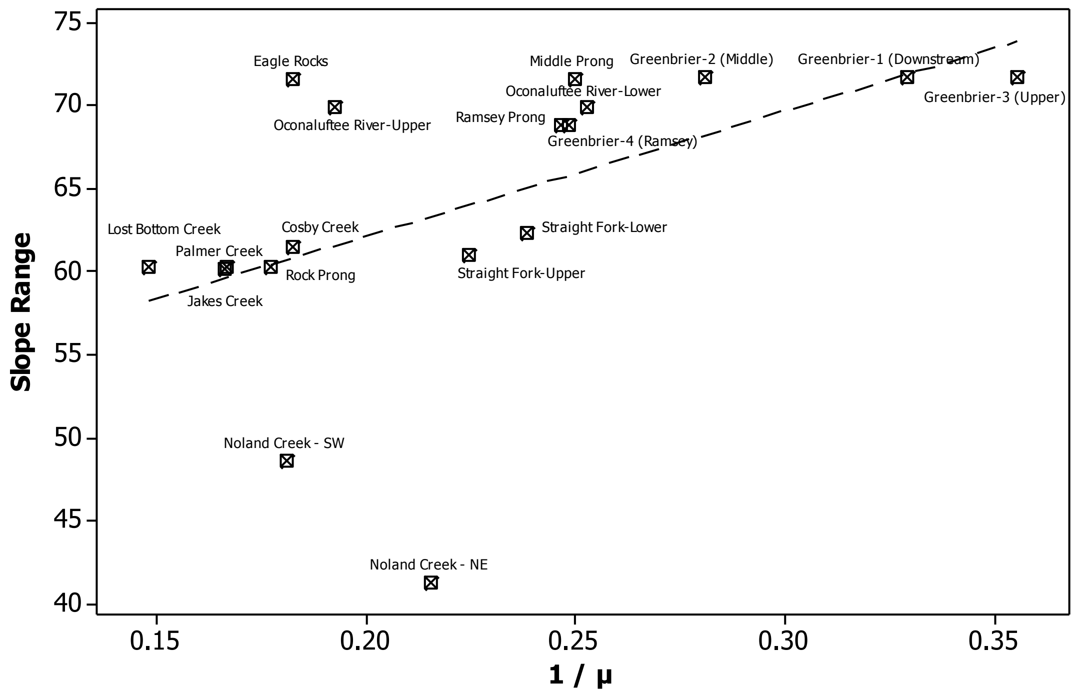


Figure 44. Least Squares Linear Regression For Slope Range (%) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = 0.0081 + 0.00338 RANGE_Slope

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.00810	0.09319	0.09	0.932	
RANGE_Slope	0.003381	0.001445	2.34	0.033	1.000

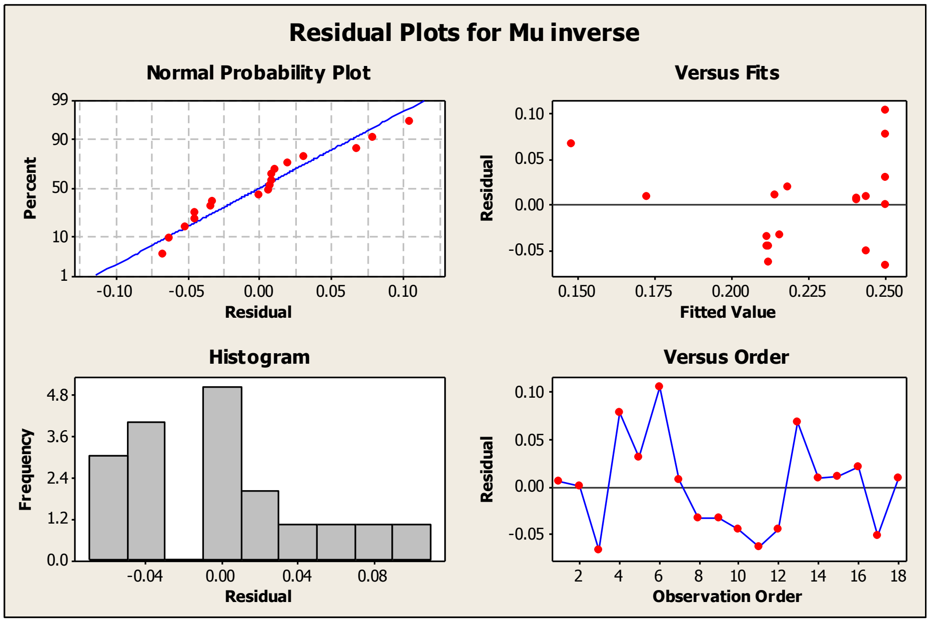
S = 0.0507931   R-Sq = 25.5%   R-Sq(adj) = 20.8%

PRESS = 0.0609225   R-Sq(pred) = 0.00%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.014122	0.014122	5.47	0.033
Residual Error	16	0.041279	0.002580		
Total	17	0.055401			

Durbin-Watson statistic = 1.72632


Residual Plots for Mu inverse
Normal Probability Plot
Percent
99
90
50
10
1
-0.10
-0.05
0.00
0.05
0.10
Residual
Versus Fits
Residual
0.10
0.05
0.00
-0.05
0.150
0.175
0.200
0.225
0.250
Fitted Value
Histogram
Frequency
4.8
3.6
2.4
1.2
0.0
-0.04
0.00
0.04
0.08
Residual
Versus Order
Residual
0.10
0.05
0.00
-0.05
2
4
6
8
10
12
14
16
18
Observation Order

Linear Regression – S_μ vs. SSS (km2)

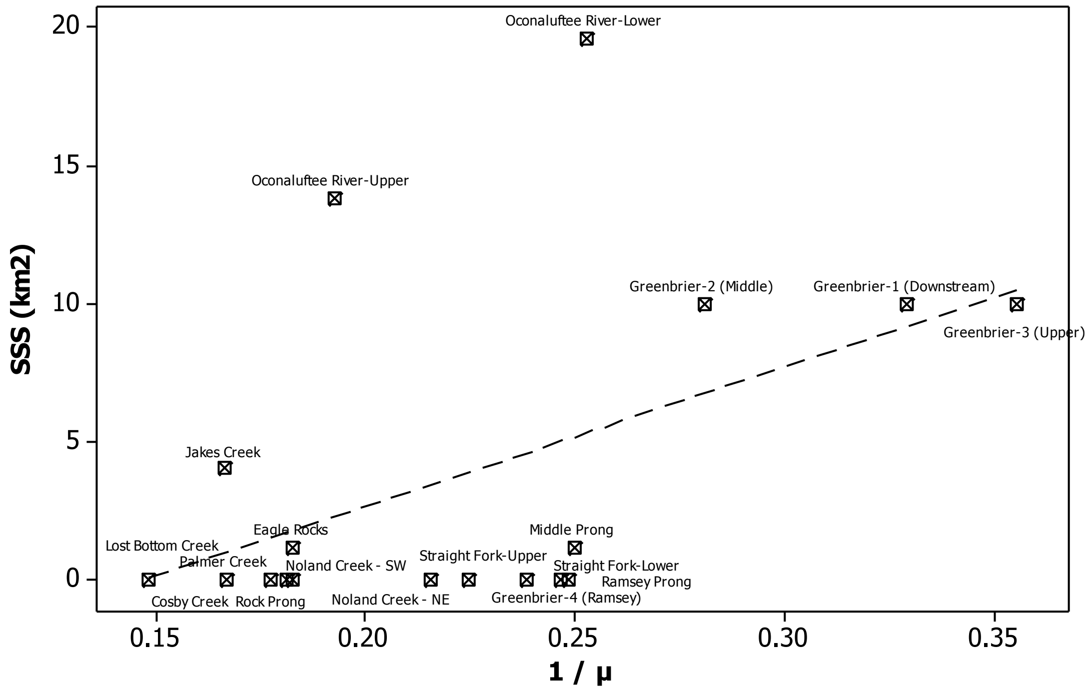


Figure 45. Least Squares Linear Regression For SSS (km^2) vs. pH Event Duration Slope.

The regression equation is
Mu inverse = 0.207 + 0.00447 SSS (km2)

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.20698	0.01462	14.16	0.000	
SSS (km2)	0.004475	0.002074	2.16	0.047	1.000

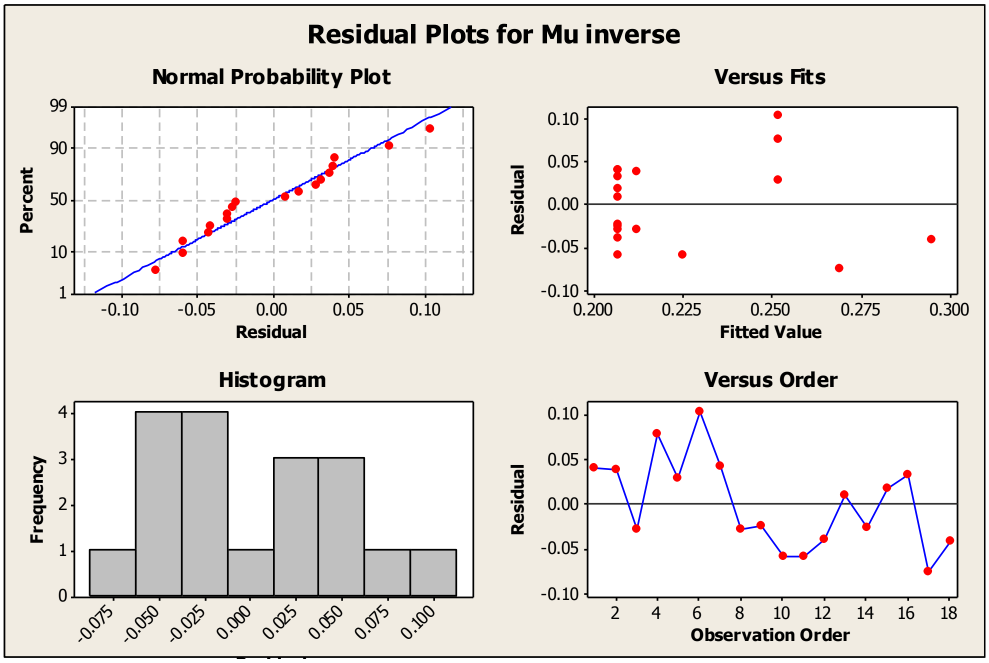
S = 0.0517914   R-Sq = 22.5%   R-Sq(adj) = 17.7%

PRESS = 0.0584610   R-Sq(pred) = 0.00%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.012484	0.012484	4.65	0.047
Residual Error	16	0.042918	0.002682		
Total	17	0.055401			

Durbin-Watson statistic = 1.23572


Residual Plots for Mu inverse
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Versus Order
Residual
Observation Order

Bivariate Regression Analyses

Only significant variables were used in the model (*p* value of 0.05). Strength and fit of the models were evaluated using R^2, adjusted R^2, predicted R^2, S Value, and Mallows' Cp. R^2, the coefficient of determination, is the percent of the total variance explained, where each additional variable used in the equation results in a higher R^2. Adjusted R^2 is an attempt to correct this weakness by adjusting the numerator and denominator by their respective degrees of freedom. Unlike R^2, adjusted R^2 can decline in value if the contribution to the explained deviation by the additional variable is less than the impact on the degrees of freedom. Predicted R^2 indicates how well the model predicts responses for new observations by using observations not included in model estimation. Predicted R^2 can prevent overfitting and is sometimes more useful for comparing models. S represents the standard distance data values fall from the fitted values, where better model response prediction produces lower S values [Minitab 15, 2009]. Small Mallows' Cp close to the number of predictors in the model plus the constant indicates that the model is relatively precise with small variance in estimating the regression coefficients and predicting future responses [*Minitab*, 2007].

Multicollinearity was addressed using the variance inflation factor (VIF). A VIF value above 10 is often used to indicate multicollinearity concerns between variables [*Helsel and Hirsch*, 2002]. Autocorrelation within residuals were checked by the Durbin-Watson statistic. If adjacent observations are correlated, the regression model will underestimate the standard error of the coefficients causing the predictors to seem more significant than they are. The range of values for the Durbin-Watson statistic is between 0 and 4 with a value of 2 generally indicating that no autocorrelation is apparent. Typically a value less than 2 indicates positive correlation and a value less than 1 describes substantial autocorrelation.

Stepwise Linear Regression

Forward and backward stepwise regression was used to develop bivariate models for estimating pH event duration slope metric as a function of observable watershed attributes. Stepwise regression removes and adds variables to the regression model to identify a useful subset of the predictor variables. The predictor with the highest partial F value is entered into the model first with additional predictors being individually entered if the addition is significant (F > 4) in the regression equation. Likewise, if retaining a predictor is not significant (F < 4) in the regression equation it will be removed. The F value is a measurement of the difference between individual distributions and the confidence of the difference, increasing as *p* values decrease.

Regression Output

F-to-Enter: 4 F-to-Remove: 4
Response is Mu inverse on 10 predictors, with N = 18

Step	1	2
Constant	0.1839	-0.1562
Area (km2)	0.00104	0.00082
T-Value	4.69	4.25
P-Value	0.000	0.001
MAX_Elev		0.00018
T-Value		3.13
P-Value		0.007
S	0.0382	0.0307
R-Sq	57.84	74.53
R-Sq(adj)	55.21	71.13
Mallows Cp	15.6	5.9
PRESS	0.030274	0.031694
R-Sq(pred)	45.35	42.79

Least squares linear regression on this model:

The regression equation is
Mu inverse = - 0.156 + 0.000816 Area (km2) + 0.000184 MAX_Elev

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.1562	0.1090	-1.43	0.172	
Area (km2)	0.0008159	0.0001920	4.25	0.001	1.161
MAX_Elev	0.00018445	0.00005885	3.13	0.007	1.161

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.041290	0.020645	21.94	0.000
Residual Error	15	0.014112	0.000941		
Total	17	0.055401			

Durbin-Watson statistic = 2.14717

Best Subsets Regression

A second bivariate regression technique, best-subsets regression, was used to develop alternative models for predicting pH event duration slope metric. The best subsets regression procedure is an efficient way to identify the best models with as few predictor variables as possible. Best subsets regression provides information on the fit of several different models, allowing model selection based on four distinct statistics, compared with stepwise regression which produces a single model based on a single statistic [*Minitab*, 2007]. Because different selection criteria are used in each model, it is possible that the two regression techniques will lead to different models.

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	Area (km^2)	SSS (km2)	Maximum Elevation (m)	Elevation Range (m)	Maximum Slope (%)	Slope Range (%)
1	57.8	55.2	11.9	0.038206	X					
1	43.9	40.4	20.5	0.044085			X			
1	41.9	38.3	21.7	0.044834				X		
1	30.5	26.1	28.8	0.049066					X	
1	25.5	20.8	31.8	0.050793						X
2	74.5	71.1	3.7	0.030672	X		X			
2	69	64.9	7.1	0.033826	X	X				
2	63.1	58.2	10.7	0.036894			X	X		
2	62.3	57.3	11.2	0.037315		X	X			
2	59.9	54.6	12.7	0.038478	X			X		
3	76.7	71.7	4.4	0.030395	X	X	X			
3	75.3	70	5.2	0.031273	X		X	X		
3	74.7	69.2	5.6	0.031663	X		X		X	
3	74.6	69.1	5.6	0.031718	X		X			X
3	69.7	63.2	8.7	0.034653	X	X			X	
4	80.1	74	4.2	0.029125	X	X			X	X
4	76.8	69.7	6.2	0.03141	X	X	X	X		
4	76.8	69.7	6.3	0.031428	X	X	X		X	
4	76.7	69.5	6.3	0.03152	X	X	X			X
4	76	68.7	6.7	0.031954	X		X		X	X
5	81.4	73.6	5.4	0.029304	X	X		X	X	X
5	80.9	73	5.7	0.02967	X	X	X		X	X
5	79.1	70.5	6.8	0.03103	X		X	X	X	X
5	76.9	67.3	8.2	0.032631	X	X	X	X		X
5	76.9	67.2	8.2	0.032684	X	X	X	X	X	
6	82.1	72.4	7	0.030006	X	X	X	X	X	X

Figure 46. Best Subsets Regression Output For All Variables

Least squares regression with 3 variables: Area (km2), SSS (km2), Max Elev (m)

The regression equation is
Mu inverse = - 0.087 + 0.00124 Area (km2) - 0.00307 SSS (km2)
+ 0.000145 MAX_Elev

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.0867	0.1243	-0.70	0.497	
Area (km2)	0.0012434	0.0004238	2.93	0.011	5.759
SSS (km2)	-0.003068	0.002718	-1.13	0.278	4.984
MAX_Elev	0.00014523	0.00006788	2.14	0.051	1.573

S = 0.0303955   R-Sq = 76.7%   R-Sq(adj) = 71.7%

PRESS = 0.0299208   R-Sq(pred) = 45.99%

Durbin-Watson statistic = 2.24392

Least squares regression on all 6 Variables:

The regression equation is
Mu inverse = - 0.135 + 0.00120 Area (km2) - 0.00429 SSS (km2)
+ 0.000057 MAX_Elev + 0.000043 RANGE_Elev + 0.0213 MAX_Slope
- 0.0188 RANGE_Slope

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.1352	0.1423	-0.95	0.362	
Area (km2)	0.0012044	0.0005493	2.19	0.051	9.927
SSS (km2)	-0.004290	0.003168	-1.35	0.203	6.951
MAX_Elev	0.00005667	0.00008495	0.67	0.518	2.529
RANGE_Elev	0.00004317	0.00005044	0.86	0.410	7.163
MAX_Slope	0.02134	0.01195	1.79	0.102	143.654
RANGE_Slope	-0.01881	0.01045	-1.80	0.099	149.904

S = 0.0300062   R-Sq = 82.1%   R-Sq(adj) = 72.4%

PRESS = 0.0308282   R-Sq(pred) = 44.35%

Durbin-Watson statistic = 2.48809

Discussion:

Best subsets regression produces the same two variable predictor model as stepwise linear regression, which has the lowest Cp of all models. However, the best subsets method also provides other possible models. For example, a model using three dependent variables, including area, SSS, and maximum elevation has R^2=76.6, R^2 adjusted=71.7, and R^2 predicted=46.0. The additional predictor variable SSS has no effect on R^2 adjusted, increases R^2 predicted, but the model has a higher VIF of 5.8 for area, with a Durbin-Watson statistic of 2.24. A regression with all six dependent variables yields a model with R^2=82.1, R^2 adjusted=72.4, and R^2 predicted=44.0, but has extremely high VIF values, near 140, signifying an extremely autocorrelated model.

Goodness-of-fit (Chi-Square)

Goodness-of-fit tests are used to test whether data follow a specific distribution. The chi-square test is a classic goodness-of-fit test which involves breaking data into groups and comparing those groups to the expected groups from the known distribution [EPA QA, 2006]. A four year data period is broken into five separate CDF curves, including four one-year duration curves and one curve developed for all four years combined.

μ vs. pH event at NDW demonstrates the influence low pH events have on characteristic curves. Year 2 has the worst fit among the one-year duration slopes, but if this curve had a μ value at pH 5.0 the curve would likely fit the slopes of the others. For demonstration purposes the lowest μ value at pH 5.0 of year 1, 3, and 4, was entered for the μ value at pH 5.0 for year 2. This change moves the μ value to 4.817, and fits very well as compared to the other yearly power values. This changes chi-square goodness-of-fit test results, showing the data to fit the same distribution with a *p*-value of 0.02.

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
1	5.526	0.2	5.4042	0.002745
2	4.886	0.2	5.4042	0.049689
3	7.148	0.2	5.4042	0.562681
4	4.452	0.2	5.4042	0.167774
5	5.009	0.2	5.4042	0.028900

N	DF	Chi-Sq	P-Value
27.021	4	0.811789	0.937

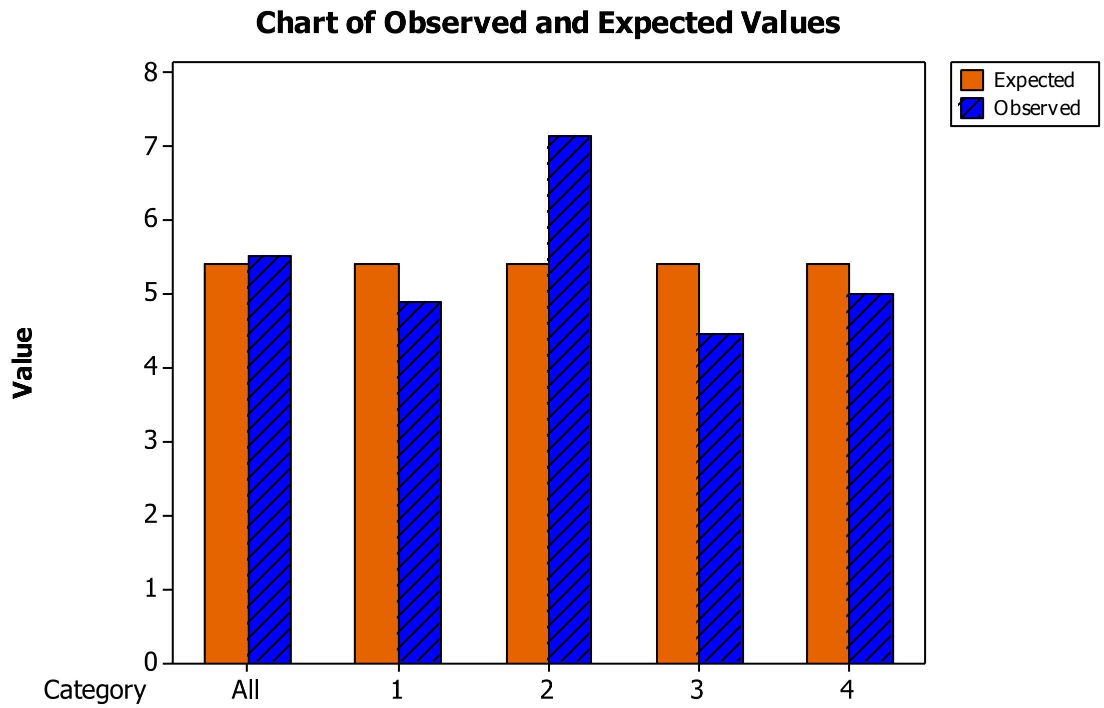


Figure 47. Chart of Observed and Expected Values for Temporal Analysis

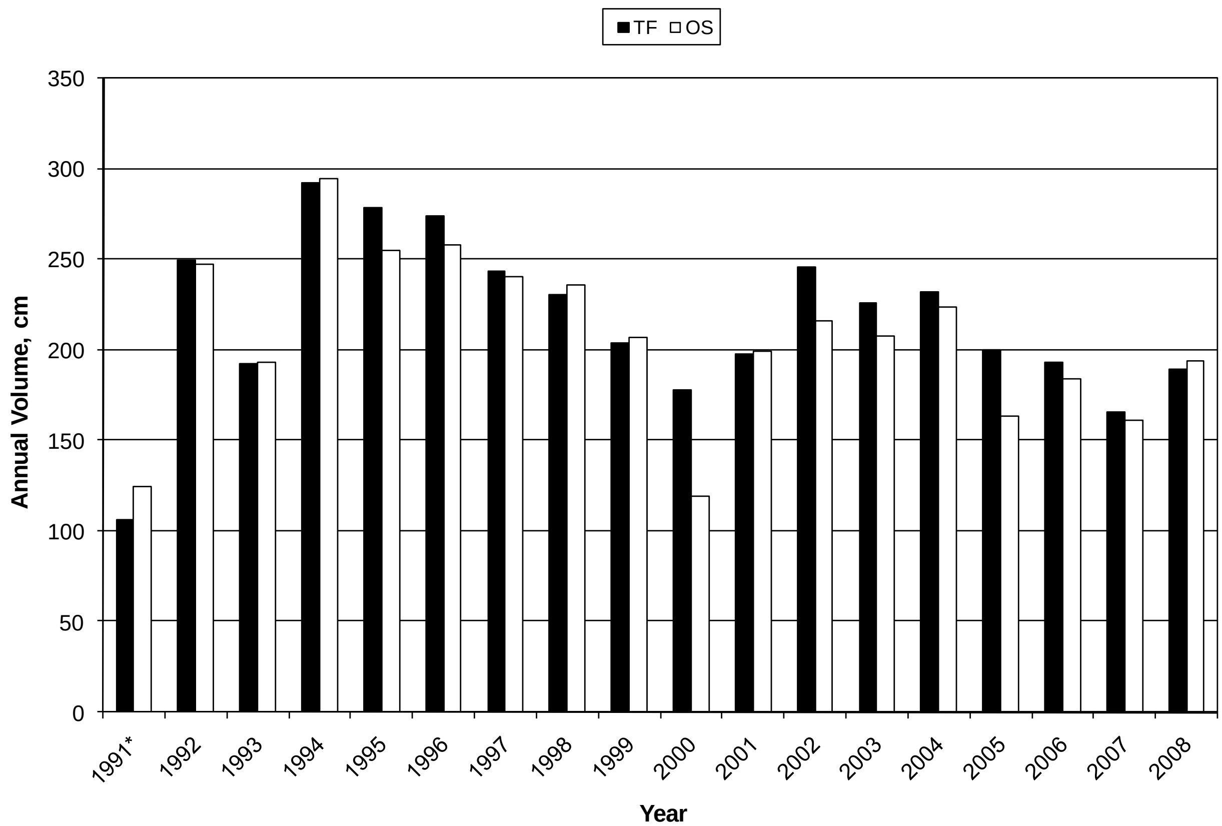


Figure 48. Annual Precipitation Volume Data Summarized For NDW From 1991 Through 2008.

VITA

Lee Mauney was born in Fulton, MS on May 12, 1985. His family later moved to Jackson, TN where he graduated high school in 2003. Following high school, he attended the University of Tennessee at Knoxville. While at UT for undergraduate studies, Lee served as Marshall of the UT Chapter of Chi Epsilon Civil Engineering Honor Society and as Outreach Coordinator for the American Society of Civil Engineers. In 2007, Lee traveled to Dominican Republic with UT's Engineers Without Borders for a water supply project to aid a small rural village. Lee benefited from a broad collection of work experience during his undergraduate time, including aiding his father with his general contracting company, working as an undergraduate research assistant, and looking after kids at a daycare. Lee spent his spring semester of 2007 in Atlanta, GA as a consulting co-op for Kimley Horn and Associates. Following graduation in 2008 with a Bachelor of Science degree in Civil Engineering, Lee decided to continue his studies. He was awarded a graduate research assistantship to research and monitor a water quality site in the Great Smoky Mountains National Park. Lee served as Vice President of Engineers Without Borders as a graduate student, and traveled to Guatemala for a water quality project in May of 2009. In December 2009, Mr. Mauney will receive a Master of Science degree in Environmental Engineering, with a focus in Water Resources Engineering. During his spare time, Lee loves to travel, run, hike, bike, play sports, read, and dabble in new things.