IMPROVING REGRESSION ANALYSIS TO PREDICT AQUATIC LIFE POTENTIAL IN THE STREAMS OF AUSTIN, TEXAS

A Thesis

by

HAIYAN HUANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Patricia Smith
Committee Members,	Huilin Gao
	Roger Glick
	Fouad Jaber
Head of Department,	Ronald Kaiser

December 2016

Major Subject: Water Management and Hydrological Science

ABSTRACT

The aim for this study was to test if the City of Austin (COA) would gain more accurate estimations on Aquatic Life Potential (AQP) by regression equations with common watersheds characteristics, rather than using one general equation from Glick et al. (2010) for all twenty-four monitoring sites. Therefore, four objectives including grouping monitoring sites according to shared characteristics, creating regression models to predict AQP in each group, selecting the best regression model and comparing the best equation with the regression from Glick et al. (2010) were established.

The monitoring sites were divided into five groups based on five hydrological characteristics. The multivariate regression analysis and simple linear regression analysis were performed in each group to form five sets of regression equations. The R^2 (Coefficient of Determination), standard error, the distribution of the monitoring sites, the goodness-of fit statistics of NSE (Nash-Sutcliffe efficiency index) and RMSE (Root Mean Square Error) were used to compare for the best equation. The R^2, NSE and RMSE were also used to compare the best regression with the regression equation from Glick et al. (2010). The results indicated that the best predicting equations were from the Impervious Cover group with R^2 larger than or equal to 0.68 and RMSE less than or equal to 11.99. That is much better than the regression equation from Glick et al. (2010) which has a slightly higher R^2 of 0.70, but much larger RMSE of 15.84. Therefore, the hypothesis that grouping monitoring sites based on common charachteristics would result in better predictive models for AQL than one equation for all watersheds in the City of Austin was approved.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of Professor Patricia Smith and Fouad Jaber of the Department of Biological and Agricultural Engineering and Professor Huilin Gao of the Department of Civil Engineering. Dr. Roger Glick from City of Austin provided the data analyzed for Section III. Haiyan Huang completed all work for the thesis independently.

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Patricia Smith for encouraging me to work on this topic, supervising me until the end of my thesis, and promptly responding to my emails. I would also like to thank Dr. Huilin Gao and Dr. Roger Glick for assisting me as committee members. I am thankful to my family members and friends for their best wishes. A special thanks goes to all members in Writing Center for helping me with wording issues.

May 17th 2016

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

Humans have been using streamflow for centuries in a pursuit of a better life, from water supply for expanding cities to irrigation for crops. The same is true for instream aquatic species. Aquatic species are constantly seeking ideal environments in which to reproduce and thrive. Numerous studies have investigated the factors which might affect the instream quality of life for aquatic vegetation, macroinvertebrates, and vertebrates such as the Barton Spring Salamander (*Eurycea sosorum*) found in the natural springs in Austin, Texas (Bowles et al., 2006; McClintock, 2002; Paul and Meyer, 2001; Turner, 2009; U.S. Fish and Wildlife Service, 2005).

Evolution of the base soil in the streambed, alterations in drainage size, adjustments in streamflow and baseflow and even changes in land use practices may impact stream water quality, and potentially have an adverse influence on the ecosystem's health (Austin, 2014; Omernik, 1986; Richter, 2011b; Sung et al., 2011). The city of Austin (COA) uses the Aquatic Life Score (AQL), a measure of benthic macroinvertebrate communities, as one of the sub-indices in the Environmental Integrity Index (EII), a program to monitor and assess the physical, biological and chemical integrity of their creeks and streams. AQL is scored from 0 to 100, where higher AQL scores indicate a better environment for aquatic species survival. In addition to the AQL, the EII also includes measures of water quality, sediment quality, non-contact recreation, contact recreation, and habitat quality in its overall estimate of (McClintock, 2002). The difficulty in continuously measuring the factors making up the index, AQL in particular, has resulted in relatively infrequent assessments, approximately every 2 to three years (McClintock, 2002). Therefore, it is necessary to predict AQL, termed the Aquatic Life Potential (AQP), based on the factors, particularly the hydrologic factors, that effect benthic macroinvertebrate communities.

The significance of using AQP to predict AQL is easily seen. Having equations to predict AQP allows those making decisions regarding future developments in the COA to assess the effects on the EII through simulation of scenarios of change. Accurate equations to predict AQP equation allows more frequent assessments of AQL to be made while at the same time reduce heavy financial burden of frequent measurements. In 2010, the COA created a multiple regression equation to be use for all monitoring sites to determine AQP based on several hydrologic factors. The best set of regression equations were developed based on previous five watershed characters, like the Drainage Area (DA) Group, the Impervious Cover (IC) Group, the Baseflow Ratio Group, the Baseflow Group, and the Channel Geology Group, in twenty-four watersheds in and around Austin to verify the hypothesis for this research. The importance of all crucial factors will be determined from regression analysis as well as the backward elimination process. The five grouping criteria were made according to given numerical criteria ranges, which would be stated in Section III.

1.2. Hypothesis for this Research

The hypothesis for this research is that more accurate estimates of AQP can be made when watersheds are grouped based on similar characteristics (impervious cover %, drainage area, streamflow, baseflow, channel geology etc.), and a set of regression equations is developed for each of these groupings, rather than a single equation for all watersheds.

1.3. Objectives

The specific objectives done to test the research hypothesis are as follows:

1.3.1. Group monitoring sites according to shared characteristics

1.3.2. Create a regression model to predict AQP for each resulting group of watersheds.

1.3.3. Select the one "best" monitoring site grouping method based on regression analysis results and goodness-of-fit to observed AQL.

2. LITERATURE REVIEW

2.1. The Importance of Ecological Health in Streams

Ecological health has been defined as the "goal for the condition at a site that is cultivated for crops, managed for tree harvest, stocked for fish, urbanized, or otherwise intensively used" (Karr 1996). Long term monitoring of stream health is important because changing land use and other forms of human activity can have significant impacts on biological and physicochemical conditions of stream systems (Platt and Connell 2003). There is a need to ensure instream species are not endangered by instream conditions that favor invasive species, that streamflow is free from most nonpoint source pollution and that instream flows are sufficient for survival.

Sousa (1984) defined an ecological disturbance as "a discrete, punctuated killing, displacement, or damaging of one or more individuals (or colonies) that directly or indirectly creates an opportunity for new individuals (or colonies) to become established". Instream ecological disturbances can lead to changes in species diversity (Platt and Connell, 2003), and alter the length of instream food chains (Marks et al., 2000; Post, 2002). Eutrophication is one example of an ecological disturbance (Ansari, 2011). When the concentration of nutrients, phosphorus, nitrate and sulfate in particular, exceed the digestive ability in a water body, aquatic vegetation such as green algae multiply quickly consuming the available instream dissolved oxygen (DO) (Ansari, 2011). As a consequence, aquatic species either die off or decrease in diversity.

Instream flows are essential for all aquatic life, providing a basic habitat for aquatic species. Flow patterns determine both the quality and quantity of physical habitats for instream aquatic organisms (Fritz and Dodds, 2005). Instream species have evolved survival strategies that correspond to the natural flow regime, frequency, and the rate of change of flow conditions, making it necessary to understand the relationship between stream hydrology and aquatic species health (Poff and Ward, 1989). Unimpaired in-stream ecosystems are more likely to keep the adaptive capacity to face potential environmental changes, such as climate change (Baron et al., 2003). A healthy stream ecosystem is able to capture or transform excessive nutrients to other absorbable forms for other aquatic species. Changes in land cover, soil type and canopy cover can affect the water resource distribution of aquatic systems, (City of Austin, 2015a; Duncan, 2012; Richardson et al., 2007). Riparian zones, which are the transitional areas between aquatic systems and the streamside land environment, reflect the biodiversity conditions in surrounding watersheds. Well maintained vegetated riparian areas around urban streams add quality of life for provide supplementary recreational opportunities for the people living around them (City of Austin, 2015a). A sound aquatic ecosystem provides perfect habitat for existing species to reproduce and survive in a stable environment (Baron et al., 2003; Richter et al., 1996).

2.2 EII Scores in Austin and Its Applications in Austin

Water quality in streams and creeks, is not only important for human needs, but also for aquatic life including plants, mammals, and amphibians (Baron et al., 2003). In 1974, the Austin City Council adopted the Bicentennial Project with the goal of preserving and enhancing all creeks and waterways for municipal use function as well as future ecosystem health function (City of Austin, 2009a). Since 1974, several studies have been done to maintain or improve the aquatic health of Austin streams (City of Austin, 2009a; Michael et al., 2010), including the Lake Austin Watershed Ordinance (1978) and the Barton Creek Watershed Ordinance (1980). In 1986, the COA passed the Comprehensive Watersheds Ordinance (CWO) for NPS control measuring which specified requirements for water quality, locations of water quality buffer zones and limitations for impervious cover (IC) (Johns, 1991).

The Environmental Resource Management Division (EMR) of COA developed the Environmental Integrity Index (EII) in 1996. The goal was to use the EII as a tool to "assess the current water quality conditions of Austin's watersheds, and to have a baseline for long term evaluation of our water resource", based on specific water resource requirements of the Clean Water Act Section 303d list (McClintock, 2002). Ranging from 0 to 100, a higher EII score in a watershed indicates better biological, chemical and physical health. High scores also indicate that watersheds have been less impacted by their surrounding projects or programs (Herrington, 2003). The EII score is a weighted mean value of six annual sub-indices, water quality, sediment quality, non-contact recreation, contact recreation, habitat quality, and aquatic life. The EII score for each watershed in the City of Austin was initially calculated over a three-year period (1996-1999). Data requirements from the Texas Commission of Environmental Quality (TCEQ) Clean Water Program increased the frequency to once every two years (Herrington, 2011). The EII score is not only useful for estimating water quality in both urban and rural areas, but also validating estimation of ecological health of aquatic resources made by local agencies which merely use chemical analysis for water quality assessments (McClintock, 2002).

EII scores are widely used in combination with other index values for various purposes (Richter, 2011; Scoggins et al., 2013; Thompson, 2007). For example, the Index of Riparian Integrity (IRI), which is a macro-scale tool built using land use and aerial photography in Austin, uses the EII to measure the integrity of riparian zones along creeks and streams in Austin. EII scores were used as the dependent variable in combination with Drainage Acre Thresholds (DAT) of 64 acre and 640 acre in regression models used to estimate health in riparian zones of all creeks in COA's jurisdiction (Scoggins et al., 2013).

A third example is the Austin Lake Index (ALI). Following similar logic to and using modified methodology than the EII, the Nature Conservancy developed a new assessment for the physical, chemical and biological conditions in three watersheds which were not covered in EII (Richter, 2011). The ALI is applied on a larger scale with more sub-metrics and more frequent measurements than the EII. Habitat and aquatic life scores in ALI calculation are used by decision makers to identify aquatic systems suitable for potential habitat restoration in lakes instead of streams. (Richter, 2011). Although there was not much difference in ALI scores in lakes compared to the original assessment, they can still serve as a baseline for the City of Austin to detect other potential environmental changes in area lakes (Richter, 2011).

2.3. Aquatic Life Score (AQL)

The Aquatic Life Score (AQL) is one part of part of the EII score. The AQL score is composed of several factors affecting aquatic life. Those factors are the macroinvertebrate community structure, the diatom community structure, percent algae cover, chlorophyll-a index, and the presence or absence of fish making up 30%, 25%, 10%, 25%, and 10% of the total AQL score, respectively (McClintock, 2002). The COA uses several biologic metrics computed based on samples acquired for use in the EII to compute the AQL. Since the AQL score represents the overall aquatic health of a monitoring site, the COA tends to use this index more often than EII. A detailed description of the methodology to calculate AQL can be found in the EII Quality Assurance Project Plan (COA, 2008). The AQL score in the Barton Creek in 1994 was 74, it was the least impacted creek, and the Barton Creek was in the Good category (McClintock, 2002). The acceptable level in each factors varies, for example, it is considered good for aquatic healthy living when chlorophyll-a index is higher than 50 while the diatom community structure score needs to be higher than 65 (McClintock, 2002). AQL scores can be used to examine the effect of NPS pollution loads on stream health to assess the effectiveness of NPS pollution control regulations (City of Austin, 2015b).

2.4. Relating Hydrologic Characteristics to Aquatic Life

Several studies regarding the relationship between stream hydrology and ecological life score of species have focused on northern regions of United States (Zimmerman, 2010; Poff et al., 1997). However, there lacks more regional studies in

Austin, Texas to investigate sub daily stream hydrological characters with instream aquatic life scores for watersheds.

$$AQP = 63.417 + 3.914 \times \ln Q_{90} + 12.041 BF_1 - 18.227 T_{dry} \tag{2.1}$$

where

Q_{90} is defined in Table 1 (ft^3/s),

BF_1 is the baseflow calculated from measured streamflow using one pass through a baseflow filter (ft^3/s)

and T_{dry} is the fraction of time when the mean daily flow was less than 0.1 ft^3/s (decimal)

Glick et al. (2010) used multiple regression analysis to create a predictive equation for AQP to use for all watersheds in the COA, based on several hydrologic characteristics. The Q_{mean} (mean daily flow during 1958-2007), Q_{peak} (peak daily flow during 1958-2007), Q_{90} (the daily flow rate that is over 10 percent of the time during 1958-2007), $+_{mean}$ (the average of all positive differences between consecutive daily values) and $-_{mean}$ (the average of all negative differences between consecutive daily values) are the area-adjusted measurements and they were a log-normally distributed environmental statistics, so they were included in the analysis. The flow measurement score, BF_1, was used in the final equation. F_{LN} and T_{dry} had negative correlation with AQP value from regression analysis.

2.5 The Importance of Watershed Characteristics for Ecosystem Health

In order to create the most appropriate groupings of watersheds to predict AQP, several factors must be examined to determine which would provide the most accurate assessment of ecosystem health. The following is a discussion of five such factors that have shown to have such an influence.

Drainage Area and Slope

The importance of drainage area (DA) on ecosystem health is mainly related to the volume of baseflow and the frequency of peak flows. In addition, the drainage basin area in square miles (A) is related to the overland runoff (D), according to Fetter's book regarding hydrographs of regular storm events (2001).

Asquith (1998) studied peak-flow frequency predication in tributaries of the Colorado River, in which monitoring stations were divided into two groups based on their Contributing Drainage Area (CDA). Watersheds with CDA of less than 32 square miles, were termed small basins, while those with CDA larger than 32 square miles, were termed large basins (Asquith, 1998). They found regressions to be used to estimate the peak flow frequency under 50, 67 and 90% confidence level and the regression for 67% confidence level was presented in text (Asquith, 1998). In the drainage criteria manual, Dodson (1989) set up standards for hydrologic analysis in determining peak flow in Texas and he classified watersheds into three groups based on DA, watersheds smaller than 50 acres, watersheds between 50 to 640 acres, and the watersheds larger than 640 acres. For example, when a watershed is less than 50 acres in size, it is common to use the Rational Method to determine the peak discharge but when the size increase to be between 50 to 640 acres, the local runoff rate curve could be used to find the peak flow rate (Dodson, 1989).

Drainage area is an indicator of habitat size for aquatic species which has a direct relationship with AQP score (Richter, 2011). Changes in habitat size affect total food storage, and the availability of shelter. Therefore, drainage area is one factor necessary for understanding the benthic macroinvertebrates and fish populations (Richter, 2011a; USEPA, 2012).

In a study in Massachusetts Weiskel et al. (2010) found with no change in other factors, like flow length and drainage area, aquatic communities of invertebrates and other aquatic plants tend to have a more diverse and healthier life as the slope of the for streambed increases. Fend et al. (2005) found that the slope for average stream channels size and the stream elevation are two important factors for stream environment and the availability of water flow in Santa Clara Valley, California. Additionally, the study found benthic macroinvertebrate abundance was also related with the percent of impervious cover near impacted watersheds.

Baseflow

Baseflow is a major water supply to rivers, springs and streams during dry weather (Zhan, 2015). The definition for baseflow is “The part of stream flow that is neither from nor runoff results from seepage of water from the ground into a channel slowly over time.” (Xu et al, 2011). The prevention of the reduction of baseflow volume as a result of urbanization is one of the missions in the COA Watershed Protection Plan (City of Austin, 2015b). Baseflow has shown to be a major explanatory variable for streamflow patterns prediction in Austin (Porras and Scoggins, 2013). Tennant (1976) found that when baseflow decreased, the sustainability of fish species in streams declined. In its most recent watershed master plan, instream baseflow is considered an important factor in determining aquatic life in Austin streams (City of Austin, 2015b).

Abel Porras and Scoggins (2013) found that baseflow in the streams of Austin have a positive relationship with the frequency of wet days. With the probability of dry or wet days in each stream annually, people can be informed by the basic ecological health status of this stream. Therefore, using the baseflow data for gaining the expected aquatic score is an important parameter for predicting aquatic life score.

Streamflow

Streamflow is one of the major components in the continental hydrological cycle and the interactions between streamflow from lakes, rivers and oceans strongly impact the response of hydrologic systems to atmospheric recycling (Fetter, 2001; Zhan, 2015). The continuous streamflow is the primary condition for aquatic species survival because it not only maintains the morphology of a basin or channel, but also provides a stable living environment for all aquatic life needs as well as supports fluvial processes during both the wet and dry periods (Austin, 2014). Austin (2014) stated the importance of average streamflow for aquatic and terrestrial organisms is from its provision of adequate food and habitats for the entire ecosystem. Most aquatic species tend have a better ability to reproduce in streams with regular streamflow patterns (Allan and Castillo, 2007).

Impervious Cover

Impervious cover (IC) percentage can be defined as the percentage of the drainage area that prevents the infiltration of water into the ground, such as roads, parking areas, concrete, and buildings. IC is an important parameter of instream hydrology because it alters streamflow patterns, changes the timing for peak flow and greatly reduces the storage rate during precipitation events (Arnold Jr and Gibbons, 1996; Dunne and Leopold, 1978; Paul and Meyer, 2001).

Since 2000, Austin has been one of the fastest growing metropolitan cities with nearly 20% increase in population (U.S. Census Bureau, 2009). As the population of a city rapidly increases, the land use is likely to be changed as well. Various changes in land use, not only limited to the increase in IC during the urbanization process, can cause dramatic ecological disturbances in watersheds (Sousa, 1984). Furthermore, observed urbanization has caused changes in watershed hydrology, such as severe degradation of aquatic systems within the watershed, reduction of total drainage area, and loss of floodplains and wetlands (Arnold Jr and Gibbons, 1996; Hollis, 1975; Klein, 1979; Tennant, 1976). Tennant (1976) defined severe degradation of aquatic systems as impervious surfacing reaches 30 to 70 percent surrounding the watershed. Under this situation, adequate watershed restoration is needed (Tennant 1976). Glick et al. (2010) also found a negative relationship between water quality and impervious cover in streams in Texas. Sung et al. (2011) concluded that urbanization had lowered groundwater levels and reduced the soil moisture in Austin. As a result of rapid development of urban area in Austin in the late 1990s, increases in impervious cover near streams significantly impacted the hydrology of the streams.

Multiple studies have found that the IC percentage has obvious adverse effects on benthic macroinvertebrates and assemblages of fish in streams (Klein, 1979;

Schueler, 1994). Klein (1979) in research on twenty-seven small watersheds in the Piedmont province of Maryland looked at the relationship between stream quality and the extent of watershed urbanization. It was concluded that when more than 65 percent of watershed covered by impervious materials, the baseflow was reduced by 90 percent (Schuler, 1994).

Channel Geology

Texas is unique among states in United States due to its geographic location and various topographic diversity. There are in total ten ecoregions in Texas (Texas Parks and Wildlife Department, 2015), Austin and its surrounding cities are composed of two main ecological regions. The Prairie (Backland Prairie) group is a common ecoregion in Austin area (Omernik, 1986; Texas Parks and Wildlife Department, 2015). Usually the Backland Prairie land allows growth of crops and vegetation because the soil is rich with average rainfall between 28 to 40 inches every year (Texas Parks and Wildlife Department, 2015).

Goodness-Of-Fit Statistics

Since it is necessary to validate these models generated from this study by statistical test, the NSE and RMSE values were applied to testify how close the predicted AQP from the AQP predicting equations fits with the observed AQL provided by the COA. From previous studies, the NSE and RMSE are two of the most frequently used goodness-of-fit indicators in evaluating deviations between observed and predicted data of various hydrologic model's (Croke, 2009; Harmel and Smith, 2007; Moriasi et al., 2007). The following paragraphs described the NSE and RMSE in brief but more detailed discussion on the indicators can be found in Legates and McCabe (1999).

The NSE is a widely used tool to assess and quantify the goodness-of-fit of hydrological models goodness-of-fit on a scale ranging from negative infinity to 1.0 (Gupta and Kling, 2011; Harmel et al., 2010; Legates and McCabe, 1999; Nash and Sutcliffe, 1970). NSE is calculated as (Moriasi et al., 2007):

$$NSE=1-\left[\frac{\sum_{i=1}^{i=n}(O_i-P_i)^2}{\sum_{i=1}^{i=n}(O_i-\bar{O})^2}\right] \tag{2.2}$$

where:

n is the total number of observations;

O_i is the i^{th} observation of the evaluated variable, AQL in this study;

P_i is the i^{th} predicted value of the evaluated variable, AQP in this study; and

$\bar{O}$ is the mean of the observed data for the evaluated variable, mean AQL for this study.

One advantages of the NSE index is its flexibility. For example, NSE was used to evaluate a nonlinear regression model to predict the amount of sediment transported between wind-driven and rain-impacted flow events in Belgium (Erpul et al., 2003). In a case study conducted by Moriasi et al. (2007), NSE was used to compare monthly streamflow simulated using SWAT2005 with observed monthly streamflow. A negative NSE value indicates the residual variance in the observations is larger than the residual variance in the predications. In other words, the model is not able to accurately predict because the error in data is too large (Gupta and Kling, 2011). A positive NSE value indicates the residual variance in the observed data is smaller than the residual variance in predicted data. Therefore, a positive NSE value means the generated equation for prediction is able to provide accurate predictions (Beven and Young, 2013; Moriasi et al., 2007). The closer model efficiency is to 1, the better the predictive ability of the model (Nash and Sutcliffe 1970. However, one of the most obvious drawbacks in NSE is from squaring the residuals, making it very sensitive to outliers (Harmel et al., 2010).

Since using more than one goodness-of-fit measure for calibration helps to reduce uncertainty and increase the confidence level (White and Chaubey, 2005), the RMSE was also used to measure goodness-of-fit. RMSE is calculated as (Harmel et al., 2010):

$$\text{RMSE} = \sqrt{N^{-1} \sum_{i=1}^{N} (O_i - P_i)^2} \tag{2.3}$$

where

N is the number of observations;

O_i is the observed value used in regression equation;

P_i is the predicted value from regression.

RMSE is a type of error index which indicates the absolute fit between observed and predicted data (Moriasi et al., 2007). While R^2 and NSE are relative assessments of goodness-of-fit, RMSE is an absolute measure of fit (Haan, 2002). The predicted values fit with observed values perfectly when the RMSE equals 0; therefore, lower values of RMSE indicate a better fit (Franz and Hogue, 2011; Moriasi et al., 2007). The RMSE of is a valuable error index for this study for two reasons. First, the RMSE value shows absolute error in the units of the variable being examined (Moriasi et al., 2007). Secondly, Chai and Draxler (2014) suggested that the RMSE best represents datasets with normally distributed errors (2014). In this study all errors from the regression where checked to ensure they were normally distributed.

3. METHODOLOGY

3.1. Study Area

The study area is located in the City of Austin, TX, defined by 22 United States Geological Service (USGS) stream monitoring sites and two stream monitoring sites served by the City of Austin. Twenty-one of the sites are located in Travis County (30° 16' 0" N, 97° 44' 34" W), the remaining three in surrounding counties (Fig. 1). Table 1 contains a list of the monitoring sites by watershed name, identity number and site name. In total, there were 14 different watersheds included in this study defined by the COA (City of Austin, 2015a), some containing more than one monitoring site. For example, the Barton Creek Watershed contains four monitoring sites.

The climate in Austin is subtropical, with temperatures in the summer exceeding 90° F more than 80% of the time and relatively mild winters (NOAA, 2007). Average rainfall in this region is approximately 33 inches per year (NOAA, 2007). Larger amounts of rainfall usually occur in the spring and summer (from April to September). When rainfall depths exceed five inches in a single rain event, flooding events can occur (NOAA, 2007).

Austin is located in the conjunction of two ecoregions; the Blackland Prairie, and Central Texas Plateau (Omernik, 1986; Omernik and Griffith, 2013). Austin overlies two major geologic regions. The Glen Rose Limestone or the Edwards Limestone, has strong resistance to erodibility by rainfall and runoff, and is mostly distributed in the western part of Austin. Buda Limestone, found in the eastern part of Austin, has less resistance to erosion than the Edwards Limestone (Jordan, 1977). Buda Limestone is harder and more resistant to erosion in its upper regions and less so in its lower regions. Purves and Tarrant soils are the dominant soil types in the eastern part of Austin around the Buda Limestone (Jordan, 1977). In western Austin, Walnut Clays occur around Edwards Limestone (Jordan, 1977).

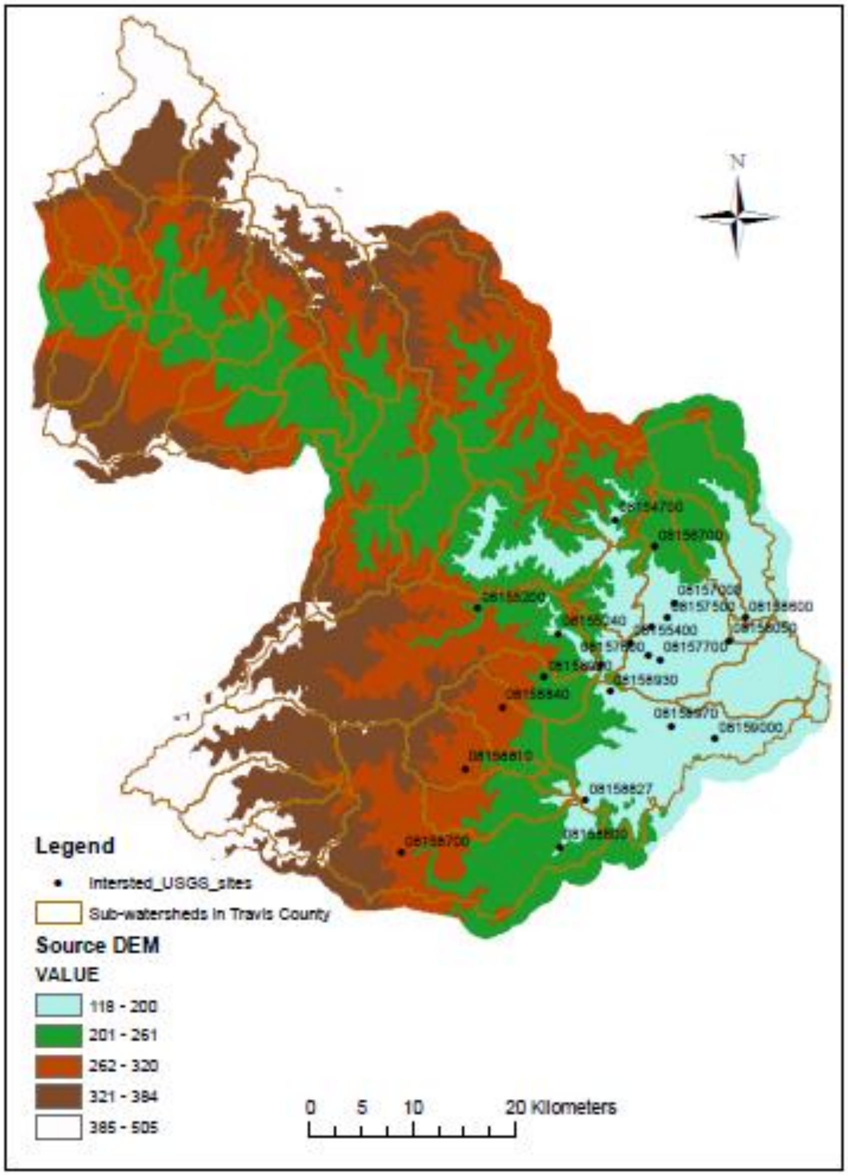

Figure 1. Study area in Austin, TX with USGS monitoring sites and sub-watersheds

Figure 1 shows that most of the USGS monitoring sites have relatively low elevations—below 200 meter ASL. According to the elevation data provided by the COA, in Travis County, the slopes for all 24 sites were less than 7.6 degrees or fairly flat.

Table 1. Twenty-four United States Geological Service (USGS) and COA monitoring sites used in this study.

Watershed Name	Station ID	Station Name
Bull Creek	08154700	Bull Creek at Loop 360
Barton Creek	08155200	Barton Creek at SH 71
Barton Creek	08155240	Barton Creek at Lost Creek Blvd.
Barton Creek	08155300	Barton Creek at Loop 360
Barton Creek	08155400	Barton Creek above Barton Springs
Shoal Creek	08156700	Shoal Creek at Northwest Park
Shoal Creek	08156800	Shoal Creek at 12th Street
Waller Creek	08157000	Waller Creek at 38th Street
Waller Creek	08157500	Waller Creek at 23rd Street
East Bouldin Creek	08157600	East Bouldin Creek at S. 1st Street
Blunn Creek	08157700	Blunn Creek near Little Stacy Park
Boggy Creek	08158050	Boggy Creek at US 183
Walnut Creek	08158600	Walnut Creek at Webberville Road
Onion Creek	08158700	Onion Creek near Driftwood, TX
Onion Creek	08158800	Onion Creek at Buda, TX
Bear Creek	08158810	Bear Creek below FM 1826
Onion Creek	08158827	Onion Creek at Twin Creeks Road
Slaughter Creek	08158840	Slaughter Creek at FM 1826
Williamson Creek	08158920	Williamson Creek at Oak Hill, TX
Williamson Creek	08158930	Williamson Creek at Manchaca Road
Williamson Creek	08158970	Williamson Creek at Jimmy Clay Road
Onion Creek	08159000	Onion Creek at US 183
Bear Creek	FBU	Bear Creek at FM 1826
Fort Branch	FTB	Fort Branch Creek at Webberville Road

Data

Table 2 contains a description of the data collected needed to either group the watersheds or to create the regression equations in this study. The physical characteristics for the monitoring sites which were assumed constant throughout the study period, drainage area (DA), longest flow path (L) to the outlet, and slope (S), are shown in Table 3. The impervious cover % (IC) shown in Table 3 was used to group watersheds. These IC values came from 2013, which was the most complete record. The data for these physical characteristics were obtained from the COA (Zhu, 2015).

Table 2. Data needed for study.

Streamflow Variables		
Variables	**Definition**	**Unit**
Q_B	Daily average baseflow	ft^3/s
Q_S	Daily average streamflow	ft^3/s
Q_{peak}	Peak daily streamflow	ft^3/day
$Q_B\%$	Baseflow ratio or percent of daily average streamflow that was contributed by baseflow	dimensionless
Q_{S90}	The 90th percentile streamflow or daily average streamflow rate exceeded 10% of the time during the period	ft^3/day
Q_{S+}	The mean of all positive differences between consecutive daily values, rise rate	ft^3/day
Q_{S-}	The mean of all negative differences between consecutive daily values, fall rate	ft^3/day
Monitoring Site Physical Characteristics		
Variables	**Definition**	**Unit**
DA	Drainage area	square miles
IC	Impervious Cover	%
S	Mean slope for selected site	%
L	The longest flow path for each watersheds	Feet

COA provided twenty years (1993-2013) of 15-min streamflow and 15-min baseflow for most of the monitoring sites (Zhu, 2015). Data from COA were then aggregated into daily average streamflow and baseflow using MATLAB. For sites or time periods that were missing data from the COA−for example site 08158810 in 1997 was missing the 15 minutes mean daily baseflow and streamflow data− they were obtained from the USGS (USGS, 2016). The SWAT digital filter was used to calculated the ratio between baseflow and streamflow for each site. Data for all of the streamflow and baseflow metrics used in the regression analysis are shown in Appendix A by monitoring site and time period.

Measured Aquatic Life Scores (AQL) and IC values used in the regression analysis are shown in Appendix B by monitoring site and date. Limited data from the COA on AQL and IC values required some assumptions be made to fill in the data record so there was a sufficient amount of data to create significant regression equations. The first assumption was that the AQL, which was typically reported for a two to three years period by COA, did not change within that period for a particular monitoring site. The second assumption was that the IC% was the same for two different monitoring sites located a short distance from one another within the same watershed. Lastly, if data on a physical characteristic, like the Impervious Cover, was missing for a monitoring site, it was assumed to have the same value as the nearest monitoring site within the same watershed or, if that was lacking, with a nearby watershed.

Table 3. Monitoring site physical characteristics.

Station ID	Drainage Area (DA) (sq. miles)	Longest Flow Path (L) (feet)	Slope (S) (%)	Impervious Cover (%)
08154700	22.3	27567.6	3.3	20.6
08155200	89.7	392.0	2.1	5.6
08155240	107.0	70010.2	3.9	6.4
08155300	116.1	4752.8	3.5	7.2
08155400	125.1	34671.1	4.1	7.8
08156700	6.5	719.5	6.3	54.3
08156800	12.3	5969.4	3.9	51.6
08157000	2.3	1188.19	3.4	48.7
08157500	4.1	103831.2	2.6	54.3
08157600	2.4	8777.8	3.4	54.4
08157700	1.2	66862.2	6.6	58.2
08158050	13.1	455.3	3.1	43.4
08158600	51.3	392.0	5.9	36.4
08158700	124.3	1338.3	2.1	6.6
08158800	166.3	9129.8	7.6	36.4
08158810	12.2	49391.8	6.6	6.7
08158827	181.2	546905.4	4.4	8.8
08158840	8.2	48563.5	4.7	6.5
08158920	6.3	4589.6	4.8	9.4
08158930	19.0	4505.8	2.5	20.7
08158970	27.6	5696.3	4.2	34.2
08159000	321.2	21238.0	6.4	29.8
FBU	5.5	N/A	N/A	19.4
FTB	2.6	N/A	N/A	44.0

3.2. Methodology for Objective 1

Five different grouping schemes were created using the 24 monitoring sites. Grouping categories were drainage area, daily average baseflow, baseflow ratio, impervious cover and channel geology. These characteristics of the monitoring sites were selected according to the relationship they each have with aquatic life as described in the literature review.

Drainage Area

Drainage areas were determined from the stream network delineation completed by the COA Watershed Protection Department in 2000. Drainage area categories used in previous studies were either too coarse (Asquith, 1998) or too fine (Dodson, 1989) to apply in this study. Therefore, monitoring sites were placed into one of three newly defined groups; very small−drainage area is less than 10 square miles; small− drainage area is between 10 and 100 square miles; and large− drainage area greater than 100 square miles.

Baseflow

Daily average baseflow was used to group the monitoring sites based on the permanence of baseflow over the 1993-2013. The monitoring sites were characterized as strictly permanent (observed baseflow over 85% of the period), strictly impermanent (no baseflow was observed over 85% of the period), and semi-permanent (watersheds not included in either of previous two groups) (Berhanu et al., 2015; City of Austin, 2000; Porras and Scoggins, 2013).

Baseflow Ratio

The baseflow ratio, the proportion of total streamflow that is contributed by baseflow, is an indicator of a stream's ability to support ecosystem services including fish habitat, wildlife, recreation, and other related environment resources. The ratio was calculated using the average daily baseflows and streamflows from 1993 to 2013. Monitoring sites were divided into one of three groups according to Tennant (1976): Poor − the average daily baseflow was less than or equal to 10% of average daily streamflow; Good − the average daily baseflow was between 10% and 30% of the average daily streamflow; and Optimal – average daily baseflow was greater than 30% of the average daily streamflow.

Impervious Cover

Glick et al. (2010) found a negative relationship between water quality and impervious cover in streams in Texas. Tennant (1976) defined a stream as "severely degraded" and in need of restoration if its surrounding watershed had imperious cover from 30 to 70% of the drainage area. Monitoring sites categorized as according to Schueler (1994) as: sensitive− IC is less than 10% of the drainage area; impacted−IC is greater than 10% but less than 25% of the drainage area; and non-supporting − IC is greater than 25% but less than 100% of the drainage area.

Channel Geology

The COA provided a final grouping describing the prominent channel geology of each monitoring site based on the prominent ecoregion and their professional judgement. Sites were divided into one of three subgroups: rock−streambeds primarily located in the western portion of the study area and are more resistant to erosion; prairie−streambeds primarily located in the eastern portion of the watershed and are less resistant to erosion; and transitional−monitoring sites located in the area between rock and prairie.

3.3. Methodology for Objective 2

Regression analysis was used to analyze the statistical relationships between one dependent variable, Aquatic Life (AQL) score, and several independent variables for the purpose of predicting Aquatic Life Potential (AQP). The independent variables, shown in Table 2 include: average daily (Q_B), average daily streamflow (Q_S), daily peak streamflow (Q_{peak}), 90th percentile average daily streamflow (Q_{S90}), daily streamflow mean rise rate (Q_{S+}) and daily streamflow mean fall rate(Q_{S-}), selected to reflect different components of the streamflow regime. Watershed characteristics were described by four additional variables: drainage area (DA), impervious cover (IC), longest flow path (L), and mean slope (S).

The type of regression used for a particular subgroup, multivariate or simple linear, was dependent on the amount of data available for the regression. Two conditions had to be met for multivariate regression to be used. The minimum number of observations required to do a multivariate regression is the number of independent variables plus one (p+1). In addition, to avoid overfitting the number of independent variables should not exceed 20 to 25% of the number of observations (Haan, 2002). The rule of thumb applied here was the number of observations must be 4 times the number of independent variables.

Multivariate Regression Analysis

In the cases where there was enough data, multivariate regression analysis using backward elimination was done using the following steps:

1. Microsoft Excel was used to create a multiple linear regression equation for the full model using all of the independent variables in Table 2 in the form (Haan 2002):

   $$Y' = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_m * X_m + \varepsilon \qquad (3.1)$$

   Where β_0 is the intercept, β_1, β_2 and β_m are the regression coefficients of the corresponding independent variables X_1, X_2 and X_m, Y' is the AQP value and ε is the residual between the observed value of the dependent variable and the predicted value of the dependent variable, in this case AQL and AQP, respectively.

2. Backward elimination was used to select the “best” regression model. Several factors were used for eliminating variables that did not contribute significantly to the model.

   a. First, a correlation matrix was created using all of the independent variables for the full regression model. When two variables are highly linearly correlated in a multiple linear regression model, they are both trying to explain the same thing in the model. If one of them is not removed than the variation in AQP that either would explain if used alone would be “split” causing one or both of them to be insignificant (Haan, 2002). Any two variables with a linear correlation value higher than 0.75 (Scoggins, 2000) were candidates for removal from the model.

b. Insignificant independent variables were selected for removal by testing the hypothesis that $\beta_m = 0$, or that a particular β is statistically not different from 0, adding nothing to the regression equation. In order to do this test, several assumptions about the residuals must be met. Specifically, tests are made to insure that the residuals are independent, identically and normally distributed (i.i.d).

c. The mean of the residuals was calculated to ensure that it is close to zero. All residual values were ranked from the smallest to the largest with a corresponding observation number and the empirical probability of each error was determined where $\mathrm{prob}(\varepsilon < \varepsilon_i) = \frac{\mathrm{rank}}{(n+1)}$. The mean and standard deviation of the residuals (standard error) were used with the empirical probability for each residual value in the NORM.INV function in Microsoft Excel to calculate the z-value for each residual. The z-values were plotted against the ranked residuals in a normal probability plot to check for normality in the distribution. A linear plot indicates that the residuals are normally distributed.

d. Once all the residuals were determined to be i.i.d., the p-values, the probability that the observed result is obtained by chance, of the regression coefficients were checked. If the coefficient had a p-value greater than α=0.05 than that independent variable is insignificant in the regression equation and is a candidate for removal. In the case of two highly correlated variables if both variables have p-values higher than 0.05, the variable with higher p-value was removed. Variables were removed one or two at a time.

3. After X_m is removed, the regression is recalculated and both the standard error (s_e), and the correlation coefficient (R^2) are used to determine the effect on the regression equation (Haan, 2002). If the standard error increased significantly or the R^2 value dropped significantly than the variable was put back into the equation.
4. Steps 2 and 3 were repeated until the best model was found (all or most independent variables were significant, with the highest R^2 and lowest s_e).

With ten dependent variables available for each equation there were some cases where the number of observations did not allow the use of all independent variables in the full model. When the number of observations in a group was less than or equal to the number of independent variables minus one ($n \leq m - 1$), nine in this case, the matrix of the correlations between the independent variables was used to select variable for deletion before completing the multiple regression steps. If there was still not enough data to do a multivariate regression than simple linear regression was used.

Simple Linear Regression

For subgroups which did not have enough observations to conduct a significant multivariate regression analysis, simple linear regressions analysis was done between AQL and each of five of the hydrologic variables from Table 2 (Q_B, Q_S, Q_{S-}, Q_{S+}, Q_{peak}).

Q_{90} was not used in this analysis because it had an almost perfect correlation with Q_{peak} and was redundant. The best fitting simple regression model from the six variables was used to represent that subgroup. In this way, each subgroup had a regression model associated with it that included at least one hydrologic variable.

The steps in doing simple linear regression were as follows:

1. A simple linear regression was calculated using Microsoft Excel with the observed AQL values as independent variable and one of five streamflow variables as the dependent variable.
2. This process is repeated until all five streamflow variables are used to create a simple linear regression model.
3. The best linear equation was selected based on the highest R^2 value, the lowest standard error and lowest p-value in the regression coefficient.

Regression analysis was repeated for all subgroups identified in objective 1, resulting in 5 sets of regression equations, one set for each type of group (DA, IC, baseflow, baseflow ratio and channel geology); each set containing one equation per subgroup (i.e. very small drainage area subgroup, small drainage area subgroup, large drainage area subgroup).

3.4. Methodology for Objective 3

Selection of Best Grouping Scheme

In order to select the best method for grouping the monitoring sites for predicting AQP all five sets of regression equations were compared using their R^2, standard error, NSE, RMSE values and the distribution of the monitoring sites among the subgroups. The selection process was as follows:

1. For each subgroup within the five major categories the best fitting regression equation (multivariate or simple linear) was used to calculate AQP. AQP values were calculated for every available AQL value in the subgroup, based on the streamflow metrics and watershed characteristics for the time-period and monitoring site associated with that AQL. For example, site 8154700 had an AQL value from 1993. It was grouped into the Small Watershed subgroup of DA group. The AQP value was generated based on the best fitting regression equation from the small watershed group, based on the watershed metrics from 1993 and the watershed characteristics for that site.
2. The NSE and RMSE were calculated between measured AQL and predicted AQP for each subgroup according to equations 2.2 and 2.3.
3. The R^2, standard error, distribution of the monitoring sites among the subgroups, NSE, and RMSE, were used to select one of the grouping schemes as the "best" for predicting AQP.

Comparison to Previous Work

In order to determine whether grouping watersheds based on similar characteristics will provide more reliable predictions of AQP than using one equation for all watersheds, a comparison between the goodness-of-fit measures for the equation developed by Glick et al. (2010) were compared to the goodness-of-fit measures from the "best" set of equations selected in this study.

The regression analysis done by Glick et al. (2010) resulted in the equation described in the literature review (Equation 2.1) as:

$$AQP = 63.417 + 3.914\ LN\left(Q_{90}\right) + 12.041BF_1 - 18.227T_{dry}$$

where:

Q_{90} is defined in Table 2 (ft^3/s)

BF_1 is the daily average baseflow after one pass of a baseflow filter, Q_B, in Table 2 (ft^3/s)

and T_{dry} is the fraction of time when the mean daily flow was less than 0.1 ft^3/s

AQP values,Q_{90}, BF_1, and T_{dry} data were obtained from Glick et al.(2010) during 1997-2007 period because this period has the most abundant data for all metrics. After gaining the summations for all the differences between AQL and AQP as well as the differences between AQL and the mean of AQL, the NSE and RMSE between AQL and AQP were calculated using eqs. 3.2 and 3.3. The resulting metrics were compared to those calculated for the “best” grouping regression equations; comparing one set of NSE and RMSE from the Glick et al. (2010) equation to three sets of NSE and RMSE from this study.

4. RESULTS AND DISCUSSION

4.1. Objective 1: Monitoring Site Grouping Results

Drainage Area Groups

The 24 monitoring sites were divided almost equally between the three drainage area subgroups (Table 4). The nine very small, eight small and seven large monitoring sites were 39, 248 and 1140 square miles, respectively.

Table 4. COA monitoring sites grouped according to drainage area (DA).

Subgroups	Site Number	Watershed Name	Drainage Area (square miles)
Very Small Watershed (DA < 10)	08157700	Blunn Creek	1.2
	08157000	Waller Creek	2.3
	08157600	East Bouldin	2.4
	FTB	Fort Branch	2.6
	08157500	Waller Creek	4.1
	FBU	Bear Creek	5.5
	08158920	Williamson Creek	6.3
	08156700	Shoal Creek	6.5
	08158840	Slaughter Creek	8.2
Small Watershed (10 < DA< 100)	08158810	Bear Creek	12.2
	08156800	Shoal Creek	12.3
	08158050	Boggy Creek	13.1
	08158930	Williamson Creek	19.0
	08154700	Bull Creek	22.3
	08158970	Williamson Creek	27.6
	08158600	Walnut Creek	51.3
	08155200	Barton Creek	89.7
Large Watershed (DA > 100)	08155240	Barton Creek	107.0
	08155300	Barton Creek	116.1
	08158700	Onion Creek	124.3
	08155400	Barton Creek	125.1
	08158800	Onion Creek	166.3
	08158827	Onion Creek	181.2
	08159000	Onion Creek	321.2

The drainage area for a monitoring site is defined by the land surface area that drains through it based on elevation. The closer the monitoring sites are to one another, the smaller the drainage area. From Fig. 1, it can be seen that the larger drainage area sites are found in the Onion Creek Watershed in the southern part of Travis County and the Barton Creek Watershed on the western side of Austin, relatively less developed areas that have fewer monitoring sites. The very small watersheds, not surprisingly were mostly found in the center of the City of Austin, an artifact of the larger number of monitoring sites in a smaller area. As shown in Table 4, several watersheds had monitoring sites in more than one group including the Williamson, Bear, Barton, and Shoal Creek Watersheds.

Baseflow Groups

Although three subgroups were defined for this group, the monitoring sites fell into one of two subgroups, strictly permanent baseflow, or semi-permanent baseflow, with 12 monitoring sites in each (Table 5). For the most part, monitoring sites within the same watershed fell into the same baseflow subgroup including three of four monitoring sites in the Barton and Onion Creek Watersheds and all of the monitoring sites in the Waller and Williamson Creek Watersheds, confirming that the baseflow was not largely different from site to site within individual watersheds.

There does not appear to be a direct connection between the drainage area and the baseflow subgroup. The large drainage area monitoring sites were divided fairly evenly between the strictly permanent and semi-permanent baseflow subgroups.

Table 5. COA Monitoring sites grouped according to % of time site had observed baseflow (Q_B %).

Subgroups	Site Number	Watershed Name	% of Time with Observed Q_B
Strictly Permanent (Q_B > 85% of Period)	08157600	East Bouldin	89.9
	FTB	Fort Branch	92.0
	08154700	Bull Creek	92.3
	08158600	Walnut Creek	94.7
	08156700	Shoal Creek	95.4
	FBU	Bear Creek	96.0
	08158827	Onion Creek	96.2
	08155400	Barton Creek	98.0
	08155240	Barton Creek	99.8
	08155200	Barton Creek	99.9
	08157000	Waller Creek	100.0
	08157500	Waller Creek	100.0
Strictly Impermanent (No Q_B > 85% of Period)	No Sites		
Semi-Permanent Baseflow (15% < Q_B < 85% of Period)	08158920	Williamson Creek	42.9
	08158800	Onion Creek	43.6
	08158840	Slaughter Creek	47.5
	08156800	Shoal Creek	48.2
	08158930	Williamson Creek	49.7
	08157700	Blunn Creek	66.4
	08158700	Onion Creek	71.0
	08155300	Barton Creek	72.0
	08158810	Bear Creek	72.5
	08158970	Williamson Creek	75.8
	08159000	Onion Creek	77.0
	08158050	Boggy Creek	81.9

The study period (1993-2013) was relatively wet because none of the watersheds had baseflow less than 42% of the time during this period. It is very likely that if a different time period were used, monitoring sites would shift into different subgroups, including the strictly impermanent baseflow group, particularly if conditions were very dry.

Baseflow Ratio Groups

The three subgroups formed by considering different values of the baseflow ratio were termed poor, good and optimal based on their ability to support ecosystems. Two monitoring stations were in the poor ratio group, nine in the good ratio group and 13 in the optimal ratio group (Table 6).

Table 6. COA Monitoring sites grouped according to baseflow ratio (Q_B/Q_S).

Subgroups	Site Number	Watershed Name	Q_B/Q_S
Poor $Q_B/Q_S < 0.10$	08158930	Williamson Creek	0.1
	08156800	Shoal Creek	0.1
Good $0.10 < Q_B/Q_S \leq 0.30$	08156700	Shoal Creek	0.1
	08158970	Williamson Creek	0.1
	08158050	Boggy Creek	0.1
	08157700	Blunn Creek	0.1
	08157000	Waller Creek	0.2
	08157600	East Bouldin Creek	0.2
	08158600	Walnut Creek	0.3
	08157500	Waller Creek	0.3
	08158920	Williamson Creek	0.3
Optimal $Q_B/Q_S > 0.30$	08159000	Onion Creek	0.3
	08155240	Barton Creek	0.4
	08158840	Slaughter Creek	0.4
	08155300	Barton Creek	0.5
	08154700	Bull Creek	0.5
	08158800	Onion Creek	0.5
	08158810	Bear Creek	0.5
	08155200	Barton Creek	0.6
	08158700	Onion Creek	0.6
	FBU	Bear Creek	0.6
	FTB	Fort Branch	0.8
	08155400	Barton Creek	0.8
	08158827	Onion Creek	0.9

All seven monitoring sites in the large drainage area group (greater than 100 square miles) had optimal baseflow ratios. As indicated earlier, these larger drainage areas are in relatively undeveloped areas. All nine sites in the very small drainage area subgroup fell either into the good or optimal baseflow ratio subgroup. The small drainage area monitoring sites (between 10 and 100 square miles) had mixed results falling into each of the 3 baseflow ratio subgroups.

As expected, the watersheds that fell into the strictly permanent baseflow category all had favorable baseflow ratios. Additionally, the two watersheds that fell into the poor baseflow ratio subgroup had baseflow less than 50% of the time. What is more surprising is that the remaining three sites that also had baseflow less than 50% of the time all fell into or very close to the optimal subgroup (ratio of 0.3 or better).

Impervious Cover Groups

Table 7 contains the results of the 24 monitoring sites grouped by their impervious cover percentage. The final grouping contained eight sites in the sensitive watershed subgroup, three sites in the impacted watershed subgroup, and 13 sites in the non-supporting watershed subgroup.

As expected, the very small drainage area monitoring sites clustered around the most densely populated areas of Austin have the highest IC. 66% of the very small and 50% of the small drainage area groups fell into the non-supporting IC subgroup.

IC did not appear to be highly correlated with Q_B as there was no obvious pattern related to the IC and the permanence of a site's baseflow. However, a majority of the sites with good baseflow ratios (8 of the 9) fell into the non-supporting IC subgroup

(IC > 25%). This indicates a stronger relationship of IC to streamflow volume than to baseflow in these watersheds. As IC increases, runoff from a monitoring site increases and the baseflow ratio decreases.

Table 7. COA Monitoring sites grouped according to impervious cover % (IC%).

Subgroups	Site Number	Watershed Name	IC (%)
Sensitive Watershed (IC % < 10%)	08155200	Barton Creek	5.6
	08155240	Barton Creek	6.4
	08158840	Slaughter Creek	6.5
	08158700	Onion Creek	6.6
	08158810	Bear Creek	6.7
	08155300	Barton Creek	7.2
	08155400	Barton Creek	7.8
	08158827	Onion Creek	8.8
	08158920	Williamson Creek	9.4
Impacted Watershed (10%< IC% < 25%)	FBU	Bear Creek	19.4
	08154700	Bull Creek	20.6
	08158930	Williamson Creek	20.7
Non-supporting Watershed (IC% > 25%)	08159000	Onion Creek	29.8
	08158970	Williamson Creek	34.2
	08158600	Walnut Creek	36.4
	08158800	Onion Creek	36.4
	08158050	Boggy Creek	43.4
	FTB	Fort Branch Creek	44.0
	08157000	Waller Creek	48.7
	08156800	Shoal Creek	51.6
	08156700	Shoal Creek	54.3
	08157500	Waller Creek	54.3
	08157600	East Bouldin Creek	54.4
	08157700	Blunn Creek	58.2

Channel Geology

Monitoring sites divided according to their channel geology provided by the COA are shown in Table 8. Eleven sites were categorized as rock, four as prairie and ten sites as transitional. Most monitoring sites in the urban areas of Austin were grouped as transitional or prairie (Figure 1). The rock group contained the 11 westernmost monitoring sites. The prairie group contained the three easternmost monitoring sites. Transitional monitoring sites were those that occurred between these two ecoregions.

Table 8. COA Monitoring sites grouped according to channel geology.

	Rock	Transitional	Prairie
Site Number	08154700	08156700	08158050
	08155200	08156800	08158600
	08155240	08157000	FTB
	08155300	08157500	
	08155400	08157600	
	08158700	08157700	
	08158800	08158827	
	08158810	08158930	
	08158840	08158970	
	08158920	08159000	
	FBU		

Discussion of the Groups

The distributions of the 24 monitoring sites among subgroups varied widely between the groups. The drainage area group had the most uniform distribution of monitoring sites among subgroups (Table 3). All remaining groups has one subgroup that had significantly fewer monitoring sites than the other two ranging from zero in the baseflow grouping to three in the IC and channel geology groups. Therefore, it was not reasonable to expect that monitoring sites would be evenly distributed among subgroups regardless of which grouping scheme was used.

There was a very strong negative correlation between drainage area and IC. Of the seven monitoring sites in the large drainage area group, five of them had IC less than 25%. As stated earlier, these larger watersheds are in relatively undeveloped areas rural or suburban areas with less development and fewer monitoring sites, so this finding is not unexpected.

There was also a very strong negative correlation between IC and the baseflow ratio (Q_B/Q_S). Of the 13 monitoring sites in the optimal baseflow ratio group (ranging from 33 to 85%), 10 of them had IC values less than 25%, and seven of them had drainage areas greater than 90 square miles. As the IC increased the baseflow ratio decreased. Eight of the nine monitoring sites with baseflow ratios between 10 to 30% had IC > 25%. At the same time the groups showed no real correlation between IC and Q_B, the baseflow volume. As stated above, this can be explained by the effect that IC has on runoff. As IC increases the amount of rainfall converted to runoff increases. Higher runoff leads to higher overall streamflow, but the baseflow contribution remains relatively unchanged. Therefore, the baseflow ratio decreases and the stream becomes less favorable to aquatic life. This finding is supported by study by Richter et al (1996) which found that the IC has negative relationship with macroinvertebrate diversity. It is reasonable to assume that macroinvertebrate diversity would also be affected by the size of drainage area in this study area. However, the drainage areas in this study are strongly tied to developed area. As areas become more developed and more monitoring sites are added, the drainage areas for each monitoring site will decrease. It is likely that if more monitoring sites were added to these larger watersheds prior to development, the tie between drainage area size and IC would be much more tenuous.

From the standpoint of overall stream health, the results are mixed. Half of the monitoring sites are non-supporting based on their IC values. However, all but two of the monitoring sites have good or optimal baseflow/streamflow ratios. These baseflow ratios are strongly influenced by the rainfall during the study period. During dryer periods, baseflow ratios would most likely change.

4.2. Objective 2: Regression Analysis Results

Drainage Area Group

The number of observations of AQL in each subgroup guided the type of regression (multiple or simple linear) that would be possible (significant). In the case of the DA groups, the small and large drainage area subgroups, had thirty-three and twenty-five observations, respectively, allowing for multiple regression. The very small drainage area subgroup had only 8 observations of AQL, not enough to create a significant multiple regression equation, and so the best fitting simple linear regression was found.

Small Drainage Area Subgroup

Using backward elimination, nine of the ten independent variables were eliminated (Q_{90}, Q_{S+}, Q_B, Q_{S-}, S, Q_{peak}, DA, L, and IC) in this order from the regression equation predicting AQP for the small drainage area group. The final equation, a function of IC, can be found in Table 9 along with the R^2 and standard error of the regression. Table 10 shows the summary statistics describing the observed AQL and predicted AQP for all of the drainage area groups. The normal probability plot of the residuals from the final equation, shown in Figure 2 indicates that the residuals are normally distributed, and therefore, hypothesis tests on the regression equation are valid. Figure 3 shows the fit between the observed AQL and the predicted AQP. The R^2 of 0.16 and large standard error indicate a very poor fit, particularly in the extremes.

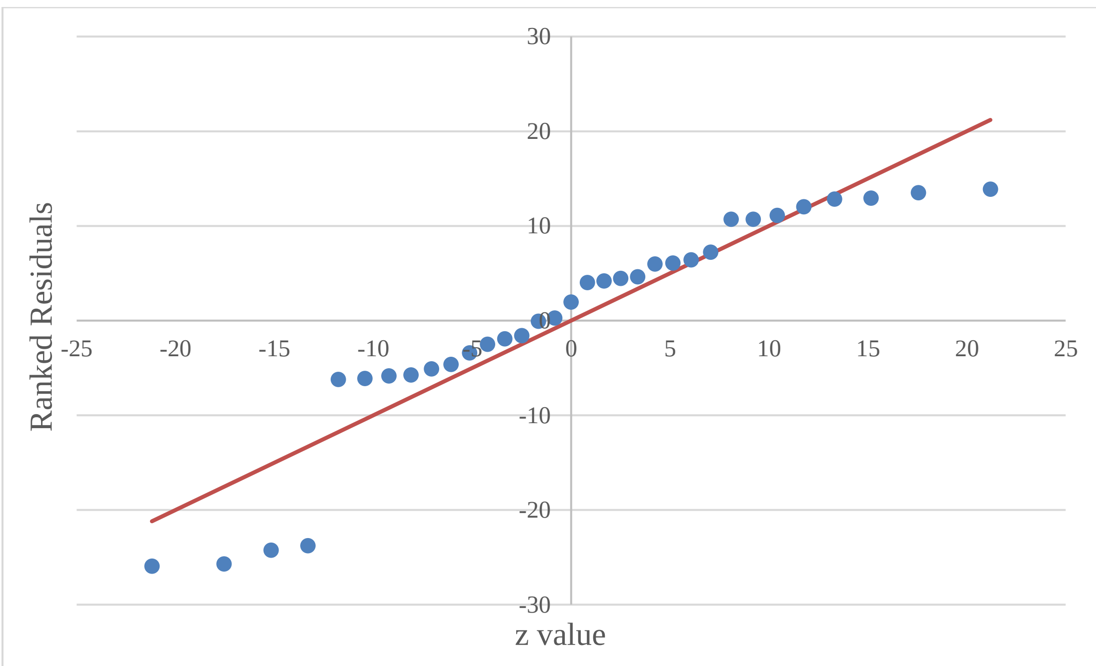

Figure 2. Normal probabililty plot for the residuals of the regression equation for AQP for the small drainage area subgroup.

Table 9. Best fitting regression equations for each subcategory in the drainage area group.

DA	*Regression Equation (AQP=)*	*n*	R^2	S_e
Small Watershed	$71.2551+0.2006\times Q_S$	33	0.16	11.57
Large Watershed	$93.2182+0.1060\times DA-0.0244\times Q_S+0.1134\times Q_B-6.5951\times S+0.0001\times L-0.0754Q_{S-}$	25	0.94	0.77
Very Small Watershed	$67.8300+1.2682\times Q_S$	8	0.53	3.08

Table 10. Summary statistics describing AQL and AQP for the drainage area groups.

	Small				Large				Very Small			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
AQL	77.97	12.43	51	91	88.76	2.72	86	92	75.25	4.17	73	82
AQP	77.97	4.96	71	88	88.76	2.64	85	93	75.25	3.04	71	79

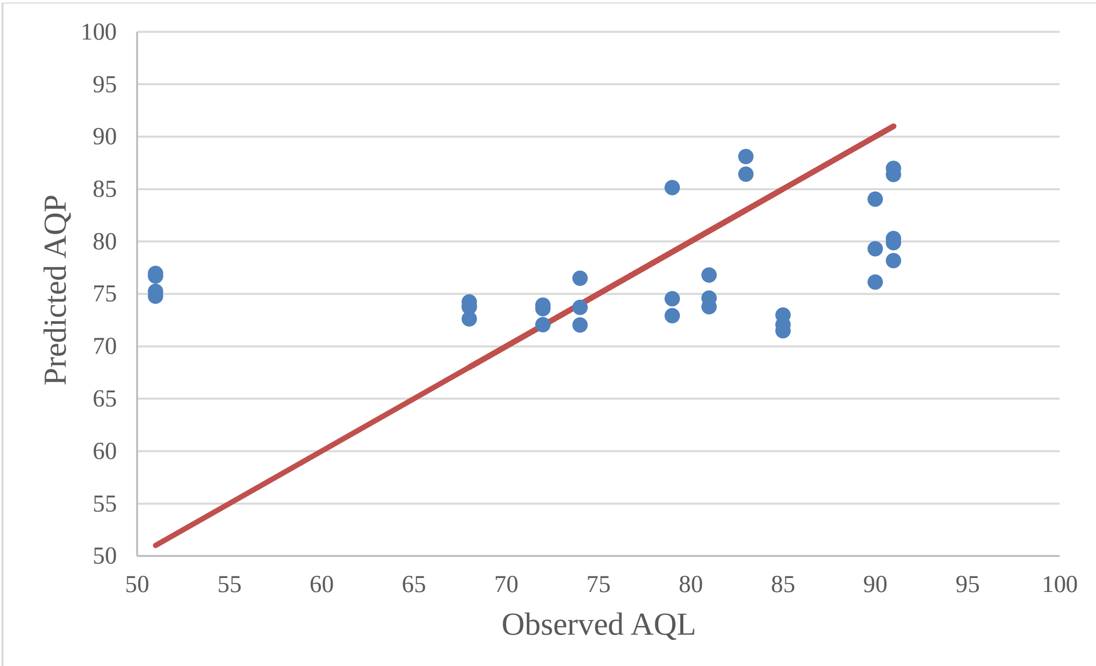

Figure 3. Observed AQL vs. predicted AQP for the small drainage area subgroup.

Large Drainage Area Subgroup

Four of the independent variables (Q_{S90}, Q_{S+}, IC, and Q_{peak}) were removed from the full model for predicting AQP in the large drainage area subgroup using the backward elimination process. The remaining variables, DA, Q_S, Q_B, S, L, and Q_{S-}, formed the equation found in Table 9, resulting in an R^2 of 0.94. Figure 4 shows the normal probability plot of the residuals between measured AQL and predicted AQP for large drainage area sites. The fit is linear for the most part, but shows some deviation in the extremes indicating a deviation from the normal distribution. Figure 5 shows the goodness-of-fit between the observed AQL and predicted AQP. The fit in the large drainage area subgroup was much better than the small drainage area subgroup, which much less variation throughout the range of the data.

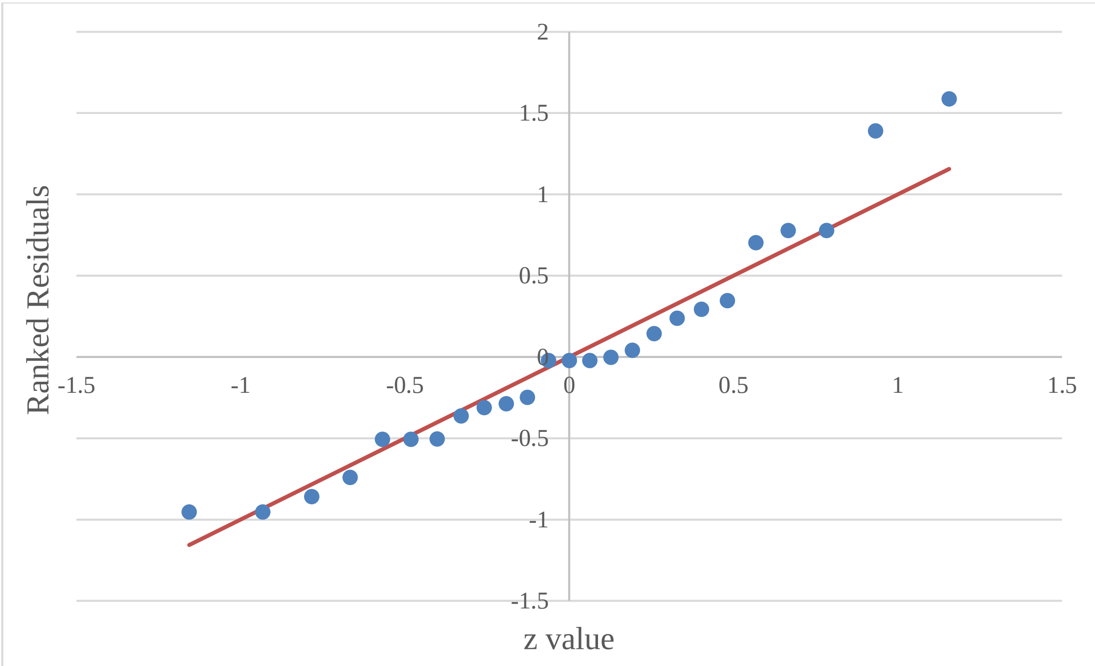

Figure 4. Normal probabililty plot for the residuals of the regression equation for AQP for the large drainage area subgroup.

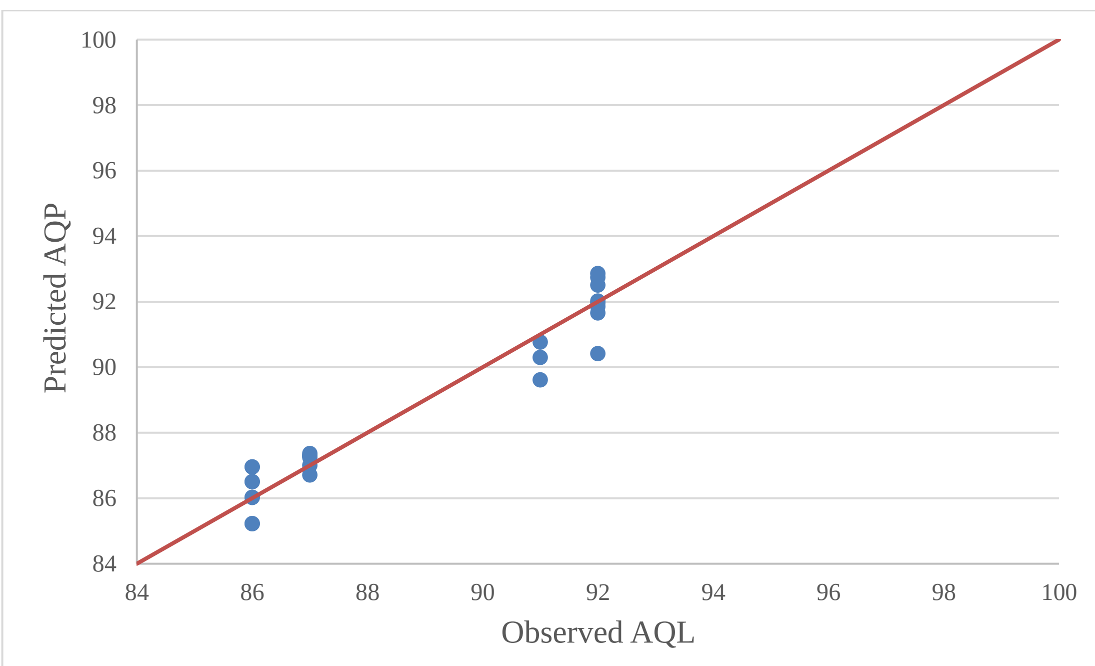

Figure 5. Observed AQL vs. predicted AQP for the large drainage area subgroup

Very Small Drainage Area Subgroup

There were not enough measured values of AQL in the very small DA subgroup to do a significant multiple regression. Therefore, simple linear regression analysis was done using each of the flow related independent variables (Q_S, Q_B, Q_{S+}, Q_{S-}, and Q_{peak}) with the observed AQL. The results of those regressions can be found in Table 11. Three of the five simple linear regression equations were not significant with p-values for the slope of the regression line much greater than $\alpha = 0.05$. The best fitting simple linear regression, came from the regression of Q_S with AQL, which had the highest R^2 value of 0.53 and lowest standard error of 3.07 and the most significant p-value (0.04) for the slope of the regression. Q_B also had a significant relationship with AQL with an R^2 value of 0.50, a standard error of 3.16 and a p-value on the slope of the regression line of 0.05. However, since Q_S had a slightly better relationship with AQL, that equation was selected to represent the very small drainage area group.

The normal probability plot (Figure 6) shows that the residuals from the very small drainage area plot follow a linear pattern. The plot of the measured AQL versus the predicted AQP (Figure 7), along with the R^2 show a fit that is very much influenced by one point in the higher end of the AQL range.

Table 11. Simple linear regression results for AQL versus each flow related independent variable for the very small drainage area subgroup.

	Q_S	Q_B	Q_{peak}	Q_{S+}	Q_{S-}
AQP Equation	$67.8300+1.2682\times Q_S$	$74.6315+1.3602\times Q_B$	$74.7161+0.0011\times Q_{peak}$	$77.1655+0.0046\times Q_{s+}$	$76.0801+0.0221\times Q_{S-}$
R^2	0.53	0.50	0.15	0.10	0.18
p-value	0.04	0.05	0.35	0.44	0.30
S_e	3.07	3.16	4.11	4.22	4.04

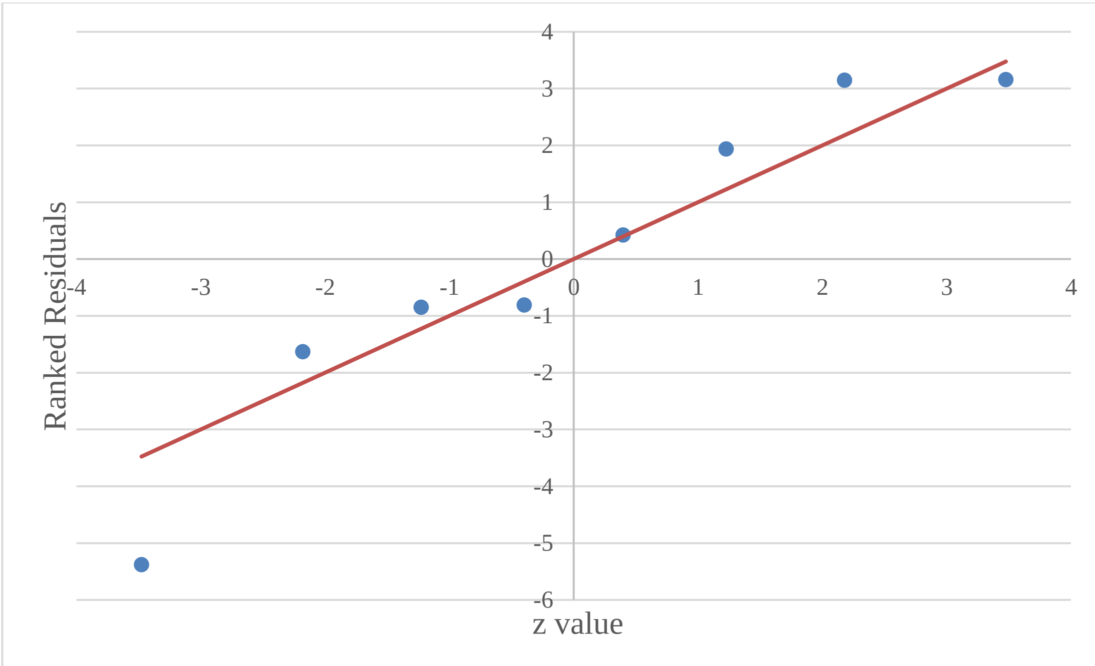

Figure 6. Normal probability plot for the residuals of the regression equation for AQP for the very small drainage area subgroup.

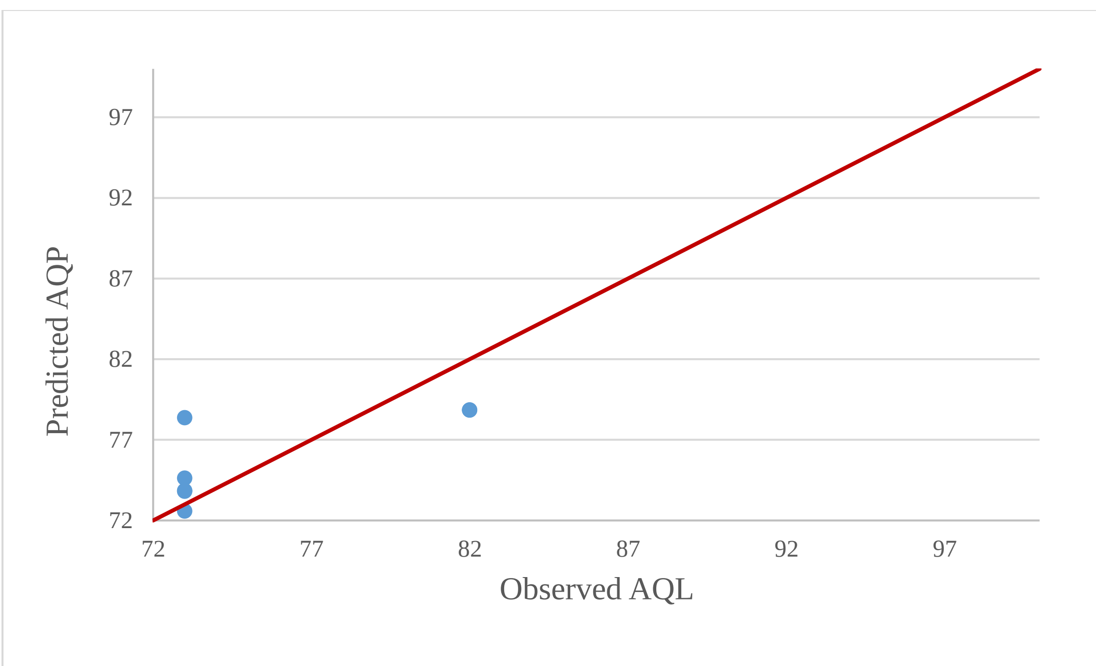

Figure 7. Observed AQL vs. predicted AQP for the very small drainage area subgroup

Discussion of the Drainage Area Regressions

The regression equation for the large DA subgroup had the best fit with the AQL data, with highest R^2 (0.94), and lowest standard error (0.77) out of the three subgroups. This subgroup also had the largest number of significant dependent variables (six of ten variables). Of those variables, half of them are related to the watershed characteristics, DA, S and L, not the flow characteristics of the monitoring site. The independent variable DA was only included in the large watershed subgroup regression equation. In fact, the large DA subgroup was the only one that contained any physical characteristics of the watershed as part of the final predictive equation. This may indicate that physical attributes of the watershed such as drainage area, slope and maximum flow length become more important as drainage area increases, while smaller watersheds are more dependent on the instream flow characteristics. In all three subgroups Q_S was included in the final predictive equations indicating a very strong relationship between drainage area, streamflow and aquatic life potential.

Both the large and very small DA subgroups had very limited data resulting in a small number of observed AQL values spread over the time period. AQL values were reported by COA for a period of 2 to 3 years and so were considered the same on an annual basis during that period resulting in the vertical lines in the data in figures 3, 5 and 7. Table 10 shows the summary statistics for AQL and AQP for all three subgroups. The nature of multiple regression insures that the means will be the same for AQL and AQP. The mean AQL and AQP decrease when moving from larger to smaller DAs. This is as expected since the very small and small DAs are centered around the more developed watersheds with higher risk to aquatic life. The small DA group showed the most variation in AQL values, with a much higher standard deviation and a very wide range between the lowest and highest AQL values. The drainage areas have the most significant variation in size, from very close to the very small DA group to very close to the large DA group. The difference in the standard deviation, minimum and maximum values between AQP and AQL are significant for the small DA group. Both the large and very small DA subgroup equations do a much better job of capturing the standard deviation, and extremes of the AQL.

Baseflow Groups

Although three subgroups were defined for the baseflow group, streams for the most part flow continuously in this area and there were no sites in the strictly impermanent subgroup. The strictly permanent subgroup and the semi-permanent subgroup had forty-seven and eighteen AQL observations, enough for multivariable regression equations to be developed.

Strictly Permanent Baseflow Subgroup

Eight of ten independent variables were removed (Q_{S90}, Q_{S+}, Q_B, IC, Q_{S-}, Q_{peak}, S, and DA) in this order in the strictly permanent baseflow group. The remaining variables, L, and Q_S, formed the equation found in Table 12, resulting in an R^2 of 0.40. Table 13 shows the summary statistics describing the observed AQL and predicted AQP for the two baseflow subgroups. Figure 8 shows the normal probability plot of the residuals between measured AQL and predicted AQP for sites in the strictly permanent baseflow subgroup, indicating a fairly good fit throughout most of the distribution. Figure 9 demonstrates the goodness-of-fit between the observed AQL and predicted AQP, and shows some significant deviations from the 1:1 line for lower values of the residuals.

Table 12. Best fitting regression equations for each subcategory in the baseflow group.

Baseflow	*Regression Equation*	*n*	R^2	*Se*
Strictly Permanent	$66.4889+0.1814\times Q_S+0.0002\times L$	47	0.40	9.61
Semi-Permanent	$106.2573 - 41.6681\times IC - 0.0005\times L$	18	0.94	3.24

Table 13. Summary statistics describing AQL and AQP for the baseflow groups.

	Strictly Permanent				Semi-Permanent			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min	Max
AQL	79.96	12.16	51	92	80.17	12.35	58	92
AQP	79.96	7.72	69	99	80.17	11.97	57	92

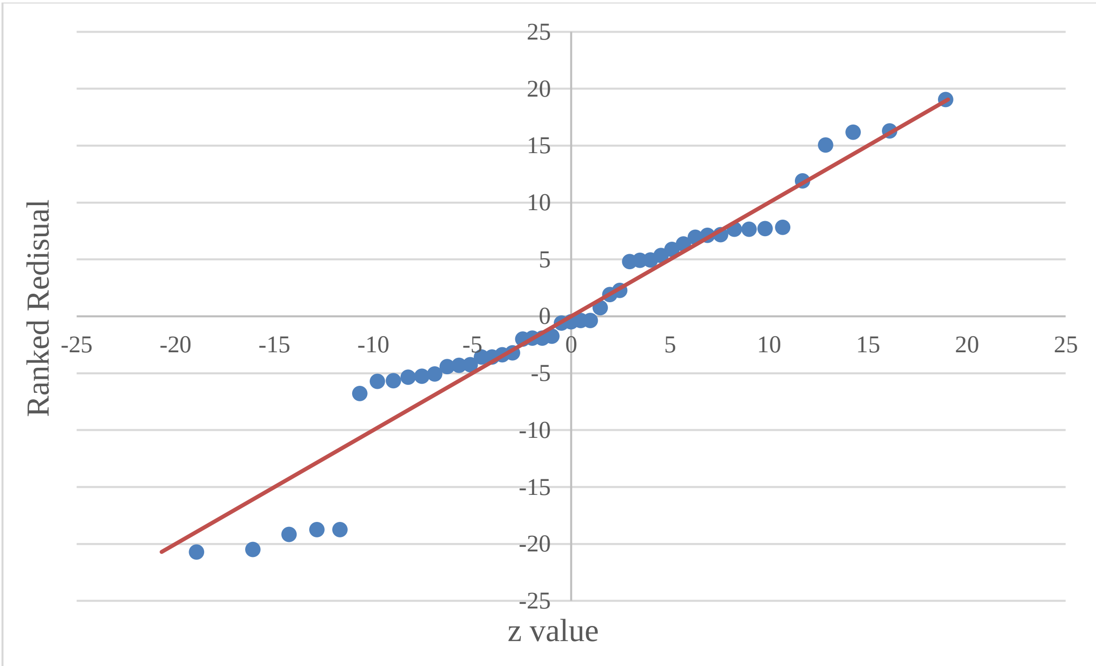

Figure 8. Normal probability plot for the residuals of the regression equation for AQP for strictly permanent baseflow subgroup.

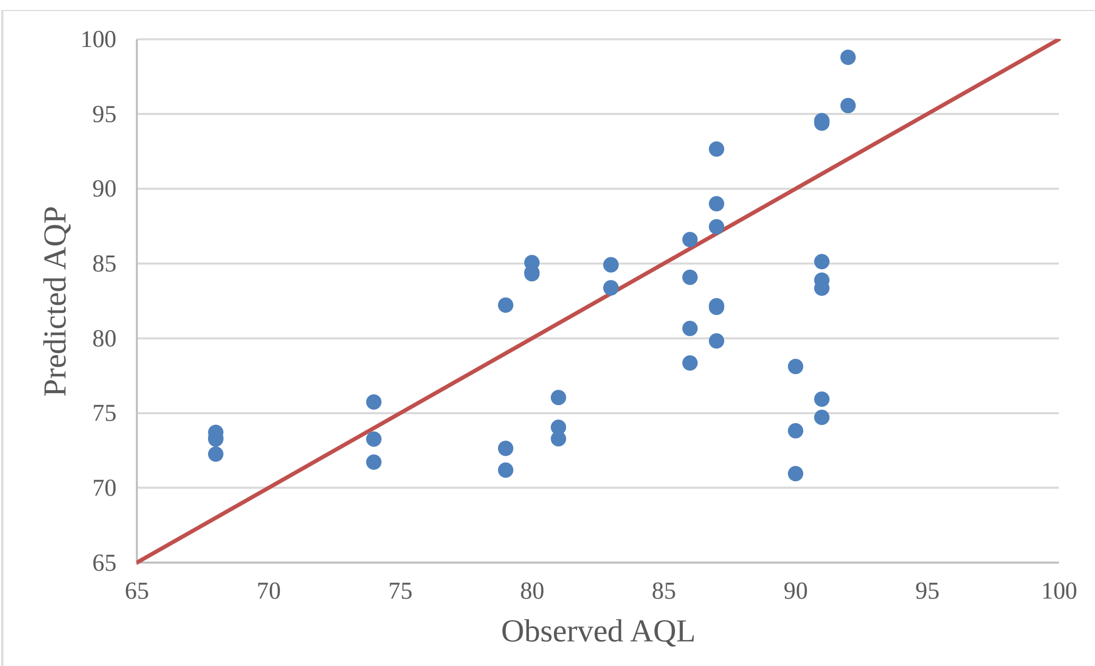

Figure 9. Observed AQL vs. predicted AQP for the strictly permanent baseflow subgroup

Semi-Permanent Baseflow Subgroup

Eight of ten independent variables (Q_{S90}, Q_B, Q_{S+}, S, DA, Q_S, Q_{peak}, and Q_{S-}) were removed in this order from the full model for predicting AQP in the semi-permanent baseflow subgroup using the backward elimination process. The remaining variables, L and IC, formed the equation found in Table 12, resulting in an R^2 of 0.94. Figure 10 shows the normal probability plot of the residuals between measured AQL and predicted AQP for sites in the semi-permanent baseflow subgroup. The normal probability plot shows that although the residuals of this subgroup have an increasing trend they do not fit a normal distribution. Figure 11 shows the goodness-of-fit between the observed AQL and predicted AQP. The fit in the semi-permanent baseflow subgroup was good but was based on limited data.

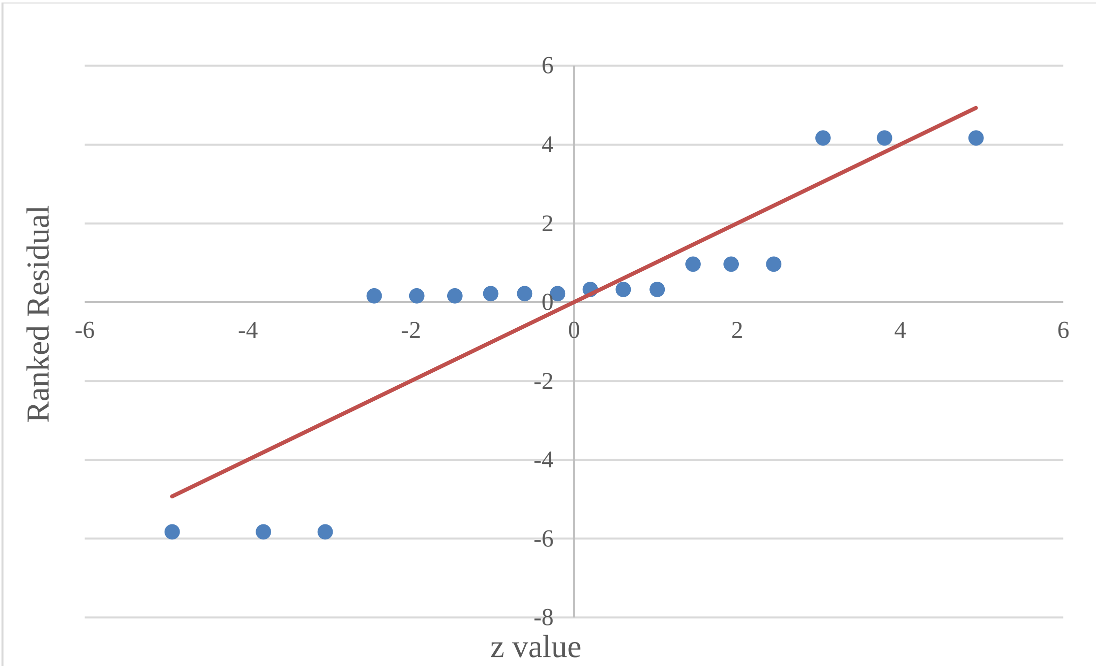

Figure 10. Normal probability plot for the residuals of the regression equation for AQP for the semi-permanent baseflow subgroup.

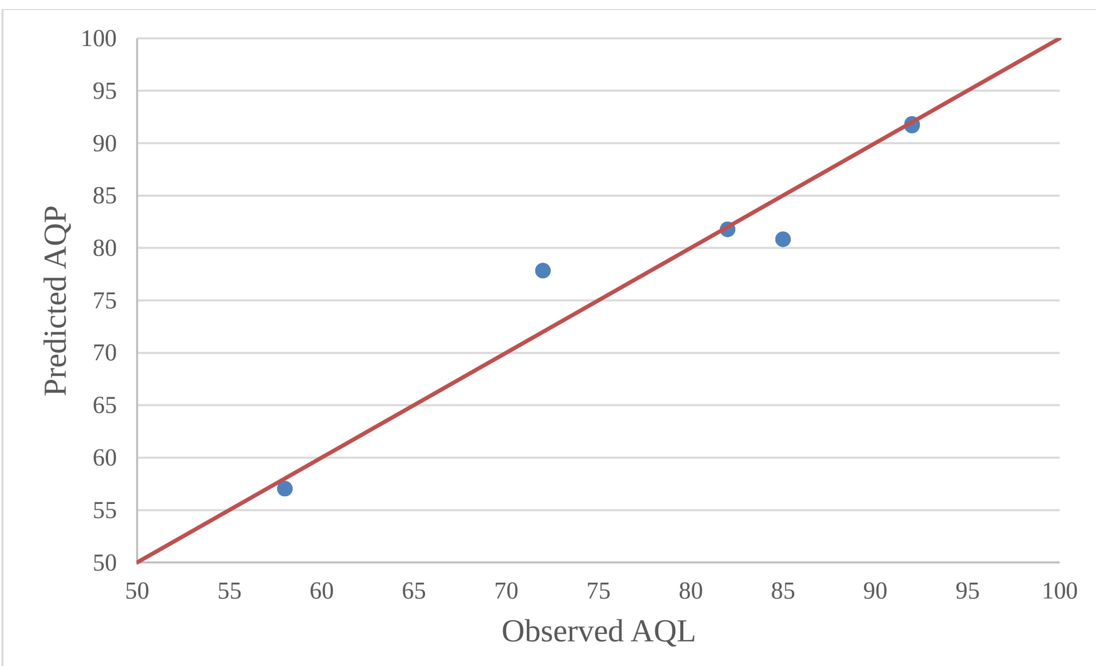

Figure 11. Observed AQL vs. predicted AQP for the semi-permanent baseflow subgroup.

Discussion of Baseflow Regression Equations

The longest flow path, L, was included in the final predictive equations for both the strictly permanent and semi-permanent baseflow subgroups. The semi-permanent group had a significantly higher R^2 (0.94) and the lower standard error (3.24) than the strictly permanent group. Table 13 shows that both the strictly and semi-permanent groups were similar in the mean, standard deviation and range of the AQL values, an indication that categorizing monitoring sites by baseflow does not result in significantly different groupings. Like the three drainage area groups, the strictly permanent baseflow group included Qs as a significant independent variable. The semi-permanent group equation contained only characteristics that describe the watershed, IC and L, and none that describe the instream flow conditions. However, as indicated in the grouping results, there appears to be a strong relationship between monitoring sites with high IC values and those with high Q_S values. It was somewhat surprising to find that although baseflow was used to create the subgroups in this group, Q_B was not a significant independent variable in either of the resulting equations.

Baseflow Ratio Groups

The optimal and good baseflow ratio subgroups had forty-eight and fifteen AQL observations, respectively, allowing for multiple regression analysis. The poor baseflow ratio subgroup had only six observations of AQL, not enough to create a significant multiple regression equation, and so the best fitting simple linear regression equation was found.

Poor Baseflow Ratio Group

There were not enough measured values of AQL in the poor baseflow ratio subgroup to do a significant multiple regression. Therefore, simple linear regression analysis was done using each of the flow related independent variables (Q_S, Q_B, Q_{S+}, Q_{S-} and Q_{peak}) with the observed AQL. The results of those regressions can be found in Table 14. All the simple linear regression equations were insignificant at the $\alpha = 0.05$ level. However, the equation containing Q_B is significant at the $\alpha = 0.10$ level, with an R^2 of 0.58 and a standard error or 3.19. Therefore, this equation was used to represent the poor baseflow ratio subgroup. Figure 12 shows the normal probability plot for the residuals of the AQP equation and indicates a somewhat normal distribution. Figure 13 shows the predicted AQL versus the observed AQP. The limited data in this group makes judging the goodness-of-fit difficult.

Table 14. Simple linear regression results for AQL versus each flow related independent variable for the poor baseflow ratio subgroup.

	Q_S	Q_B	Q_{peak}	Q_{S+}	Q_{S-}
AQP Equation	$64.5239+0.0796\times Q_S$	$52.8124+2.0023\times Q_B$	$58.8866 - 0.0002\times Q_{peak}$	$56.3431+0.0230\times Q_{S+}$	$58.3342 - 0.0024\times Q_{S-}$
R^2	0.00	0.58	0.01	0.11	0.01
p-value	0.94	0.08	0.82	0.52	0.84
S_e	15.04	3.19	4.86	4.62	4.87

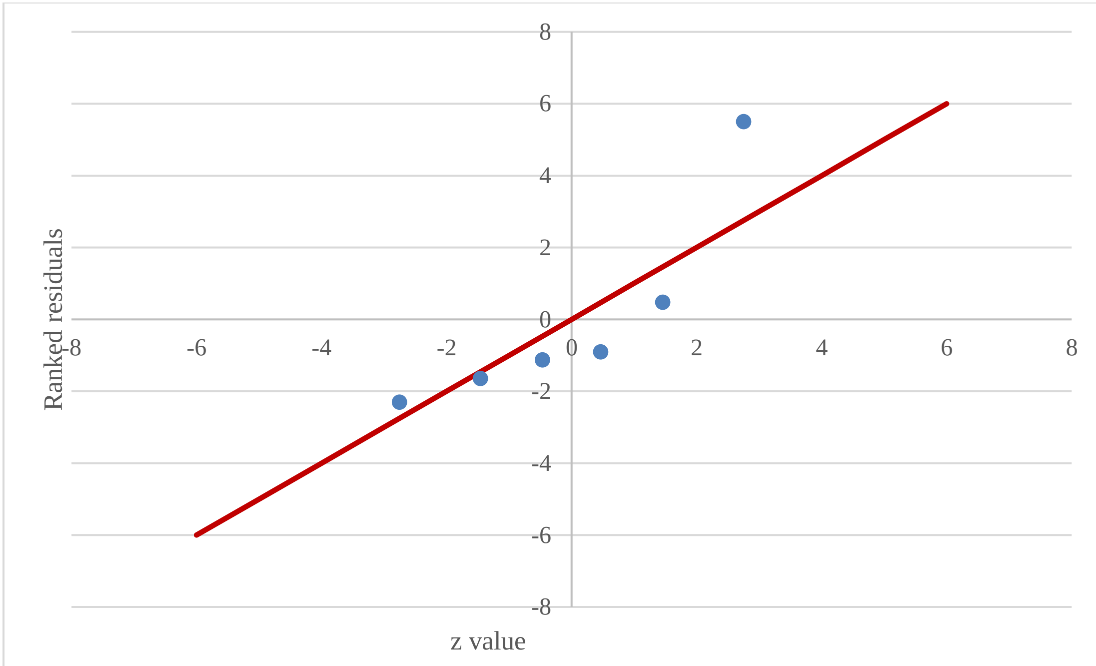

Figure 12. Normal probability plot for the residuals of the regression equation for AQP for the poor baseflow ratio subgroup.

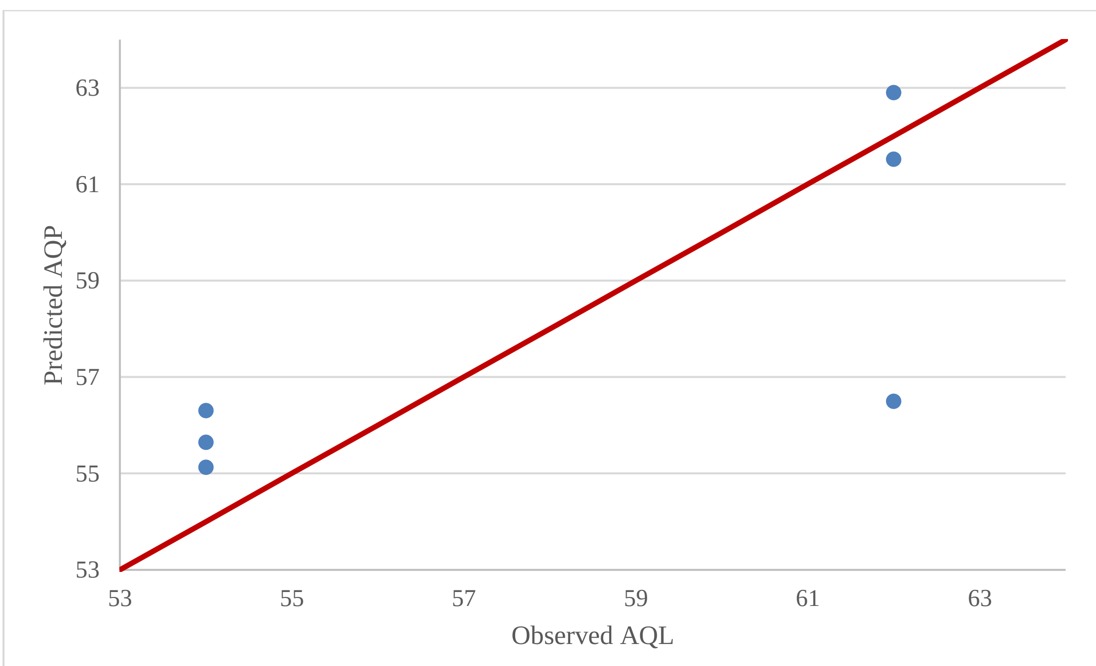

Figure 13. Observed AQL vs. predicted AQP for the poor baseflow ratio subgroup.

Good Baseflow Ratio Group

Eight of ten independent variables were removed from the full model for predicting AQP in the good baseflow ratio subgroup using the backward elimination process. The remaining variables, IC, and Q_B, formed the equation found in Table 15, resulting in an R^2 of 0.55. Figure 14 shows the normal probability plot of the residuals for the good baseflow ratio subgroup and shows deviation from the normal distribution throughout. Therefore, hypothesis tests on the regression equation are approximations. Figure 15 shows the goodness-of-fit between the observed AQL and predicted AQP, were there were significant deviations for larger values of AQL.

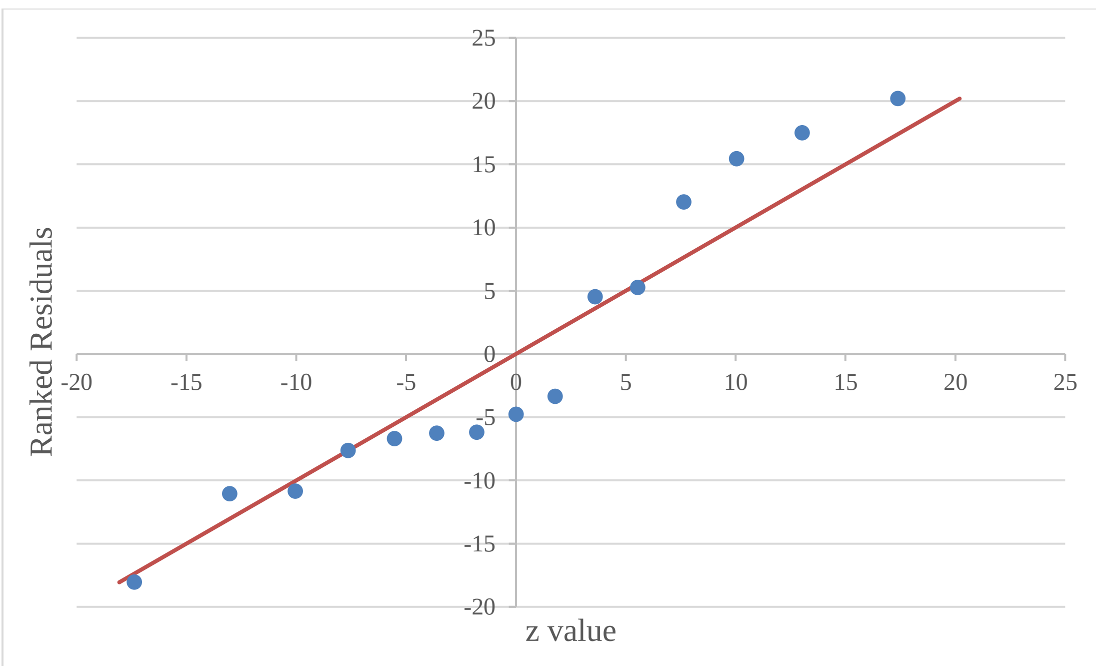

Figure 14. Normal probability plot for the residuals of the regression equation for AQP for the good baseflow ratio subgroup.

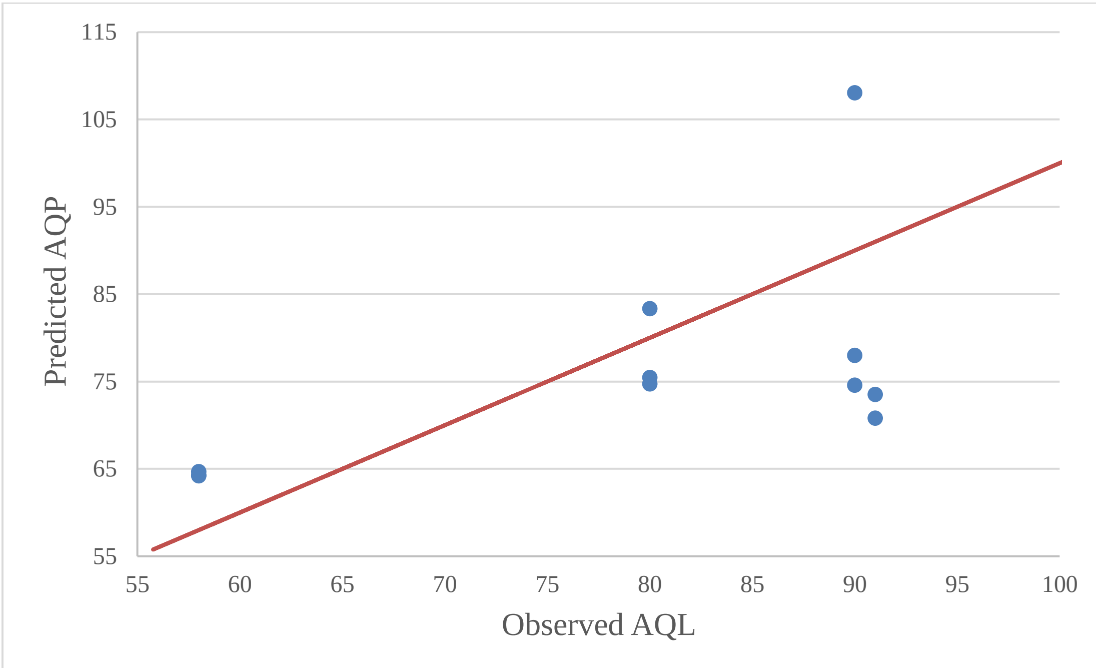

Figure 15. Observed AQL vs. predicted AQP for the good baseflow ratio subgroup.

Optimal Baseflow Ratio Group

Seven of ten independent variables (Q_{S90}, Q_{S+}, Q_B, Q_{S-}, Q_{peak}, S, and L) were removed from the full model for predicting AQP in the optimal baseflow ratio subgroup using the backward elimination process. The remaining variables, DA, IC, and Q_S, formed the equation found in Table 15, resulting in an R^2 of 0.61. The normal probability plot (Figure 16) shows the residuals for the optimal baseflow ratio subgroup fit a normal distribution throughout most of the range. Therefore, hypothesis tests on the regression equation are valid. Figure 17 shows the goodness-of-fit between the observed AQL and predicted AQP, showing an acceptable fit throughout the range of AQL.

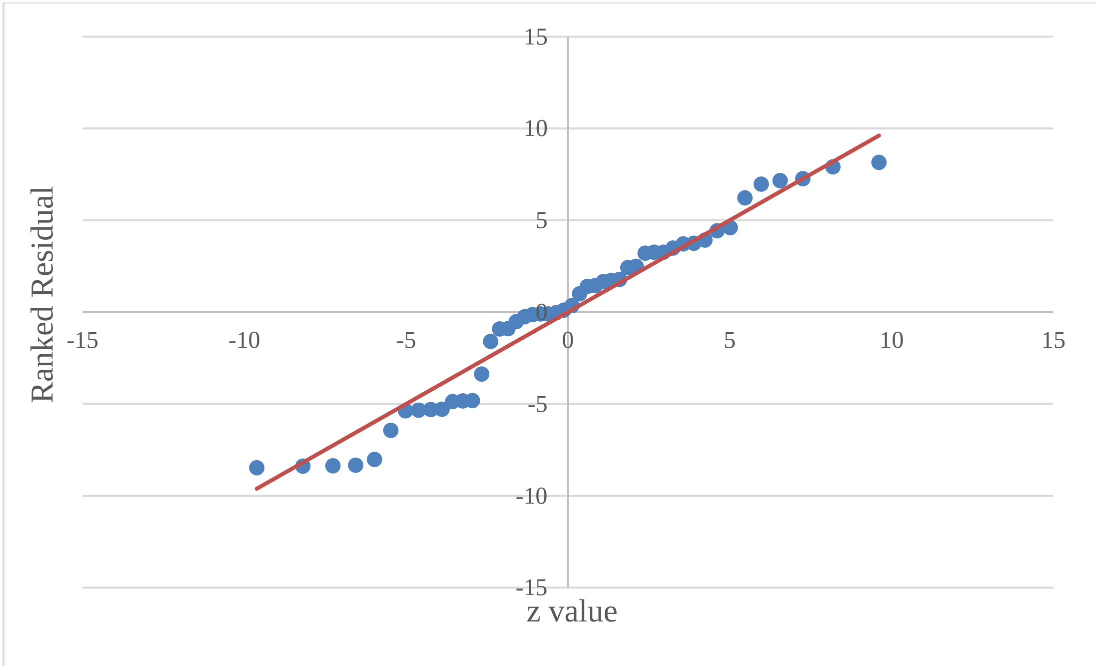

Figure 16. Normal probability plot for the residuals of the regression equation for AQP for the optimal baseflow ratio subgroup.

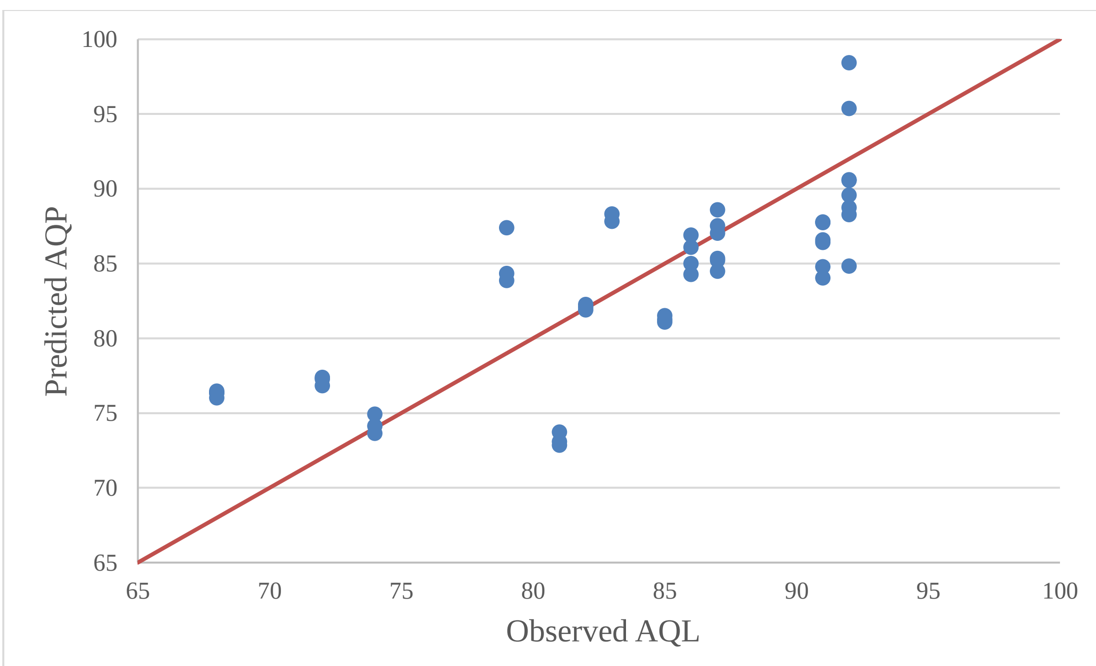

Figure 17. Observed AQL vs. predicted AQP for the optimal baseflow ratio subgroup.

Table 15. Best fitting regression equations for each subcategory in the baseflow ratio group.

Baseflow Ratio	*Regression Equation*	*n*	R^2	*Se*
Optimal	$81.9679+0.0363\times DA-61.2476\times IC+0.0578\times Q_S$	48	0.61	4.91
Good	$17.5792+120.6383\times IC+0.9508\times Q_B$	15	0.55	12.66
Poor*	$52.8123+2.0023\times Q_B$	6	0.57	3.19

* Significant at the $\alpha = 0.10$ level.

Table 16. Summary statistics describing AQL and AQP for the baseflow ratio groups.

	Optimal				Good				Poor			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
AQL	83.43	7.63	68	92	71.33	17.39	51	91	58	4.00	54	62
AQP	83.43	5.97	73	98	71.33	12.83	58	108	58	3.03	55	63

Discussion of Baseflow Ratio Regression Equations

Table 16 contains the summary statistics for AQL and AQP for all three baseflow ratio groups. The average value of AQL clearly increases when moving from poor to good to optimum baseflow groups, with values of 58, 71, and 83, respectively. This is a clear indication that the ratio of baseflow to streamflow has a significant effect on aquatic life. The three equations shown in Table 15 all contain either Q_S or Q_B, the variables that make up the baseflow ratio groups. The two equations that were developed using multiple regression also included IC as a significant variable, reinforcing the influence that impervious cover has on the relative proportions of baseflow and streamflow in an urban stream. All three equations have similar R^2 values ranging from 0.55 to 0.61, however the standard error for the good baseflow ratio group significantly larger than the standard error of the other two subgroups. As a result of its very high standard deviation, the good baseflow ratio group also has a maximum predicted value of AQP of 108, which exceeds the maximum of the AQL index itself, 100.

Impervious Cover Groups

The sensitive, impacted and non-supporting watersheds of IC subgroups had thirty-two, ten, and twenty-one AQL observations, respectively, allowing multiple regression analysis for all three groups.

Sensitive Watershed Group

Using backward elimination, seven of the ten independent variables were eliminated (Q_{S90}, Q_B, Q_{S+}, Q_{S-}, S, Q_{peak} and DA) in this order. The final equation to predict AQP in the sensitive watersheds, a function of IC, L, and Q_S, can be found in

Table 17 along with the R^2 and standard error of the regression. Table 18 contains the summary statistics for AQL and AQP for all three IC subgroups. The normal probability plot of the residuals from the final equation, shown in Figure 18 indicates that the residuals are normally distributed with some deviation in the positive tail; therefore, hypothesis tests on the regression equation are valid. Figure 19 shows the fit between the observed AQL and the predicted AQP. The R^2 of 0.54 indicates a moderate fit.

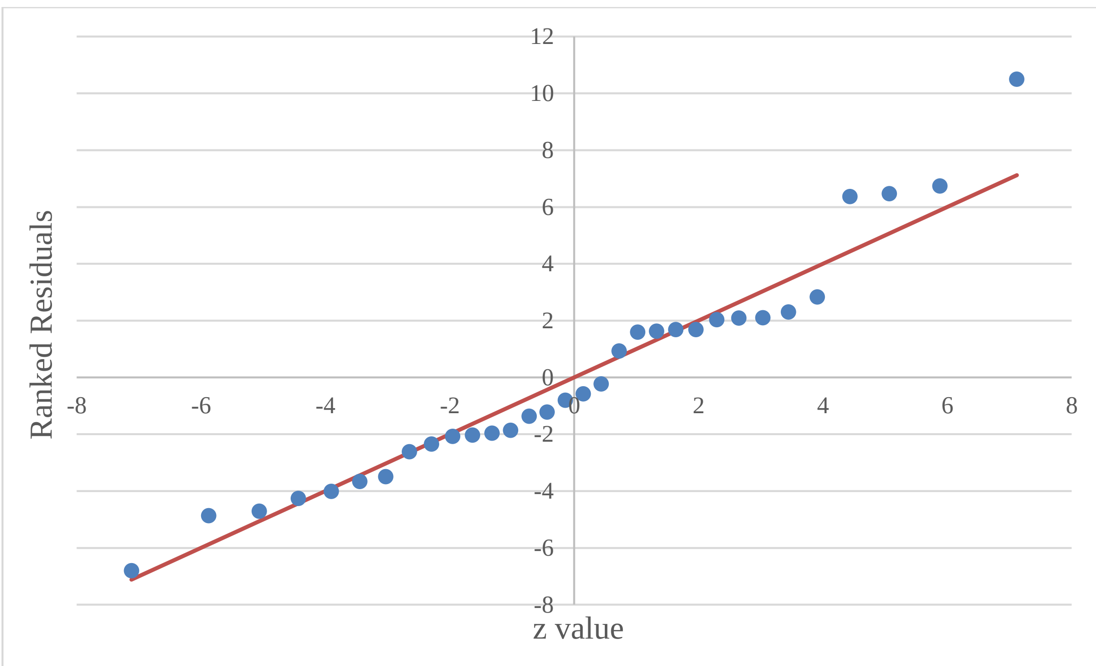

Figure 18. Normal probability plot for the residuals of the regression equation for AQP for the sensitive watershed subgroup.

Table 17. Best fitting regression equations for each subcategory in the impervious cover group.

IC Group	*Regression Equation (AQP=)*	*n*	R^2	*Se*
Sensitive Watershed	$80.8313 - 96.4598 \times IC + 7.85 \times 10^{-5} \times L + 0.0918 \times Q_S$	32	0.54	4.06
Impacted Watershed	$41.9829 + 218.8894 \times IC$	10	0.97	1.11
Non-supporting Watershed	$63.3630 - 0.0002 \times L + 1.3547 \times Q_{S-}$	21	0.68	13.26

Table 18. Summary statistics describing AQL and AQP for the impervious cover groups.

	Sensitive				Impacted				Non-supporting			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
AQL	84.88	5.35	72	92	73.7	5.68	68	81	70.52	22.39	36	93
AQP	84.88	4.14	76	94	73.7	5.58	68	80	70.52	18.52	43	102

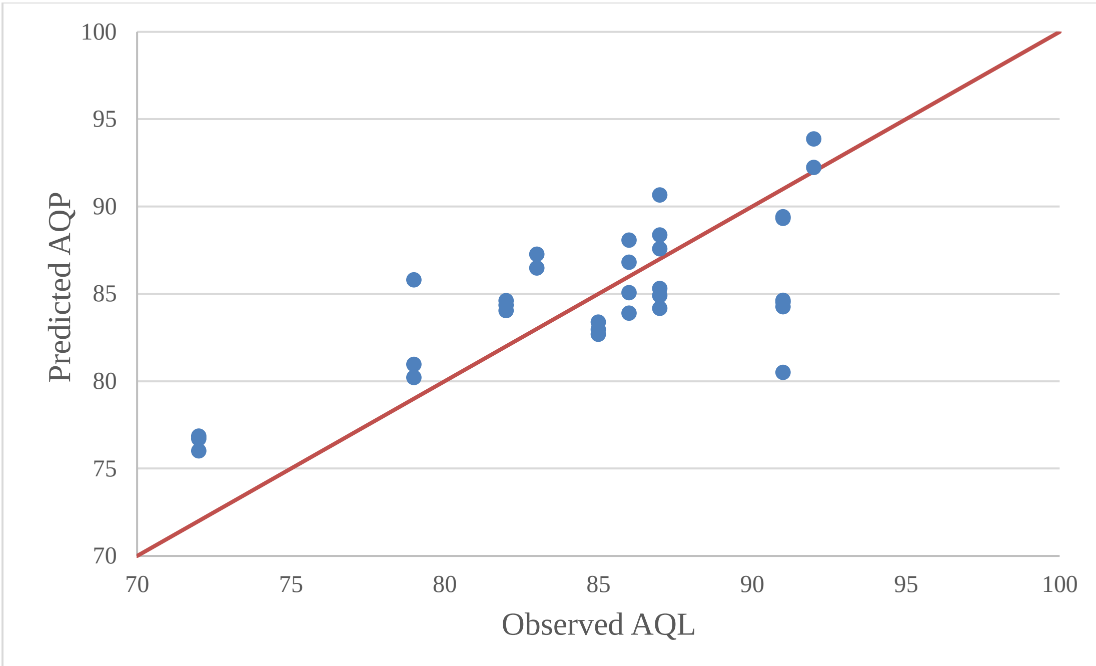

Figure 19. Observed AQL vs. predicted AQP for the sensitive watershed subgroup.

Impacted Watershed Group

Nine of the independent variables (L, Q_{S90}, S, DA, Q_{peak}, Q_{S-}, Q_B, Q_{S+}, and Q_S) were removed from the full model for predicting AQP in the impacted watershed subgroup using the backward elimination process. The remaining variable, IC, formed the equation found in Table 17 resulting in an R^2 of 0.97. Figure 20 shows the normal probability plot of the residuals between measured AQL and predicted AQP for impacted watershed and shows that the residuals do not fit a normal distribution, and so hypothesis on the regression are approximations. Figure 21 shows the goodness-of-fit between the observed AQL and predicted AQP. The plot and R2 show a very good fit, however, although there are 10 observations of AQL, the equation containing only IC resulted in only 3 AQL/AQP pairs.

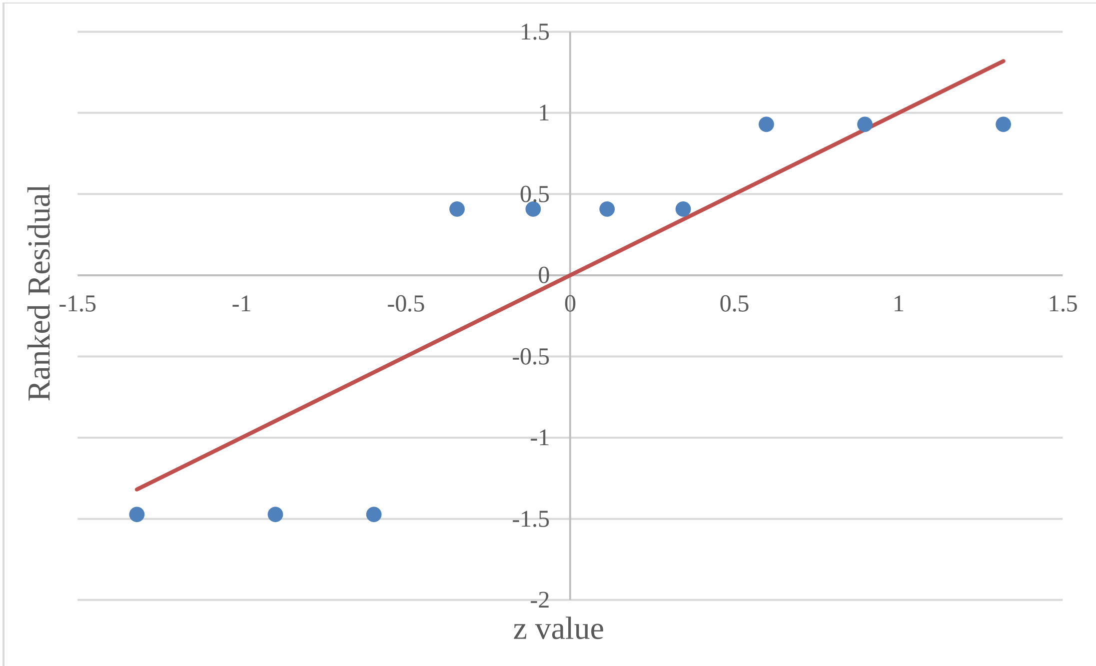

Figure 20. Normal probability plot for the residuals of the regression equation for AQP for the impacted watershed subgroup.

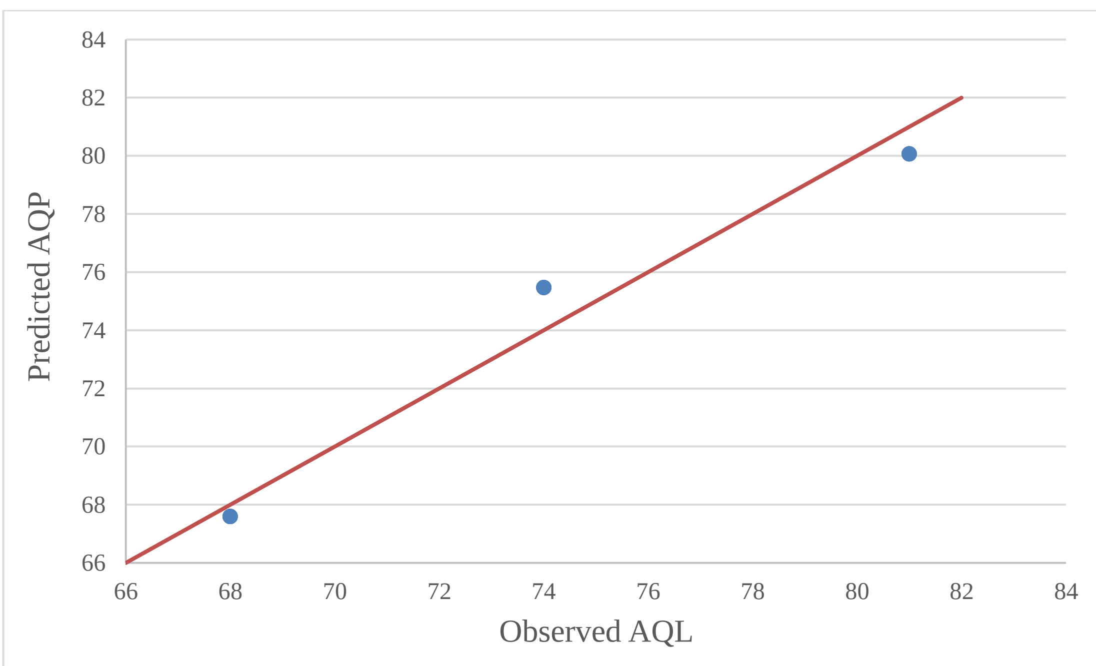

Figure 21. Observed AQL vs. predicted AQP for the impacted watershed subgroup.

Non-supporting Watershed Group

Eight of the independent variables (Q_{S90}, Q_S, Q_{S+}, DA, Q_{peak}, Q_{S-}, Q_B, and IC) were removed from the full model for predicting AQP in the non-supporting watershed subgroup using the backward elimination process. The remaining variables, Q_S and L formed the equation found in Table 17 resulting in an R^2 of 0.68. Figure 22 shows the normal probability plot of the residuals between measured AQL and predicted AQP for non-supporting watershed and shows that the residuals fit a normal distribution. Therefore, hypothesis tests on the regression equation are valid. Figure 23 shows the goodness-of-fit between the observed AQL and predicted AQP. The fit is acceptable but shows significant variation in larger values of AQL.

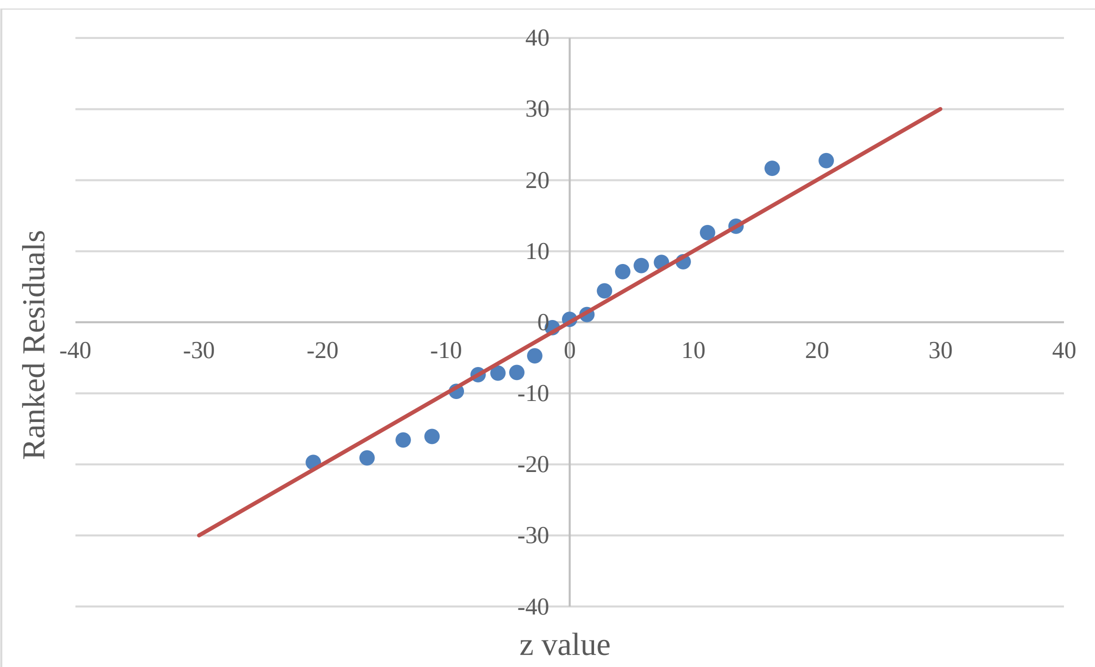

Figure 22. Normal probability plot for the residuals of the regression equation for AQP for the non-supporting watershed subgroup.

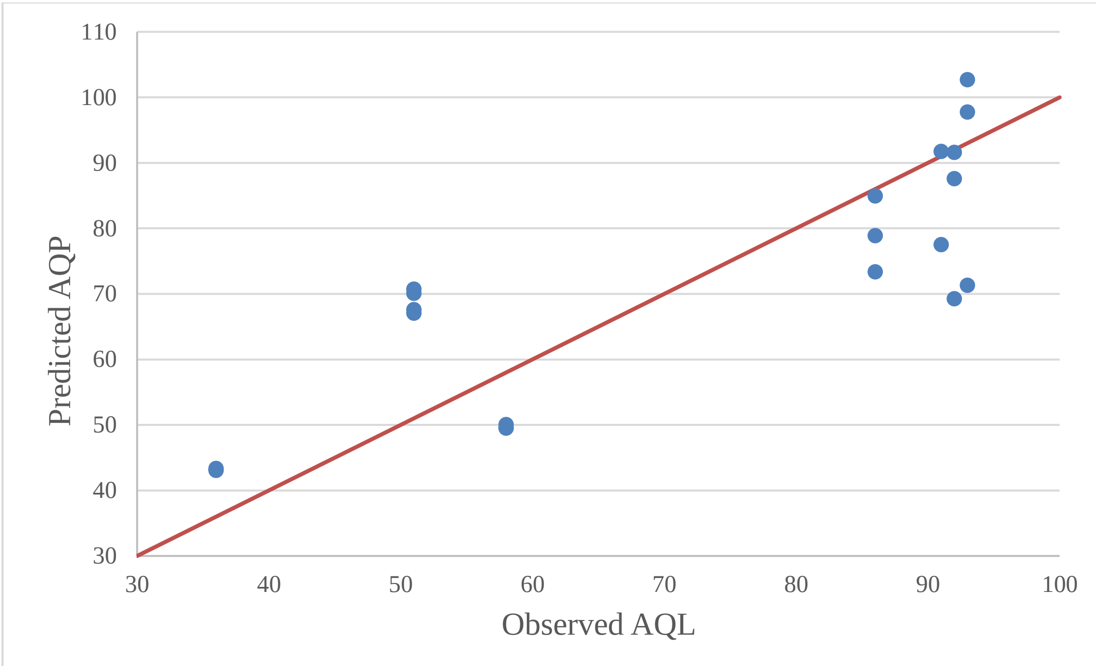

Figure 23. Observed AQL vs. predicted AQP for the non-supporting watershed subgroup.

Discussion of Impervious Cover Regression Equations

Like the baseflow ratio group, there is a decrease in the average AQL score when moving from sensitive to impacted to non-supporting watersheds, from 85 to 74 to 71, respectively (Table 18). However, the distinction between impacted and non-supporting is not clear based on the data from these watersheds. The non-supporting watersheds have the lowest values of AQL, but the maximum from the subgroup is larger than the maximums in both the sensitive and impacted groups. In fact, eleven of the AQL values from the non-supporting group are greater than 86. The remaining 10 range from 36 to 58. All of the AQL values from the sensitive watershed group are greater than 72. The standard deviations of AQL and AQP in the non-supporting group is quite high. This seems to indicate that grouping based on IC results does not clearly differentiate the effect of IC on AQL. Two of the regression equations, sensitive and impacted, did contain IC as an independent variable, but IC was not significant in the non-supporting watershed group. The longest flow path, L, was a factor in both the sensitive and non-supporting watershed groups, again showing the influence of watershed physical characteristics on AQL.

Channel Geology Groups

In the channel geology groups, the rock subgroup, had 46 observations, allowing for multiple regression. The transitional and prairie subgroups each had only 9 observations of AQL, not enough to create a significant multiple regression equation, and so the best fitting simple linear regression was found.

Rock Subgroup

Using backward elimination, eight of the ten independent variables were eliminated (Q_{S90}, Q_{S-}, Q_B, Q_{S+}, Q_{peak}, L, S, and Q_S) in this order. The final equation to predict AQP in rock subgroup, which including DA, and IC, as independent variables, can be found in Table 19 along with the R^2 and standard error of the regression. Summary statistics for AQL and AQP for all three channel geology subgroups can be found in Table 20. The normal probability plot of the residuals from the final equation, shown in Figure 24 indicates that although the residuals have a linear trend they are not normally distributed, and therefore, hypothesis tests on the regression equation are approximations. Figure 25 shows the fit between the observed AQL and the predicted AQL. The R^2 of 0.57 indicates an acceptable fit.

Table 19. Best fitting regression equations for each subcategory in the channel geology group.

Channel Geology Group	*Regression Equation (AQP)=*	n	R^2	S_e
Rock	$78.3933+0.1065\times DA - 38.7441\times IC$	46	0.57	5.24
Transitional	$67.4091+0.0016\times Q_{peak}$	9	0.61	9.92
Prairie	$22.3741+0.3246\times Q_{S+}$	9	0.87	8.03

Table 20. Summary statistics describing AQL and AQP for the channel geology groups.

	Rock				Transitional				Prairie			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
AQL	83.21	7.83	68	99	76.66	14.93	58	92	72.89	20.77	51	91
AQP	83.21	5.92	74	90	76.66	11.70	67	98	72.89	19.36	48	98

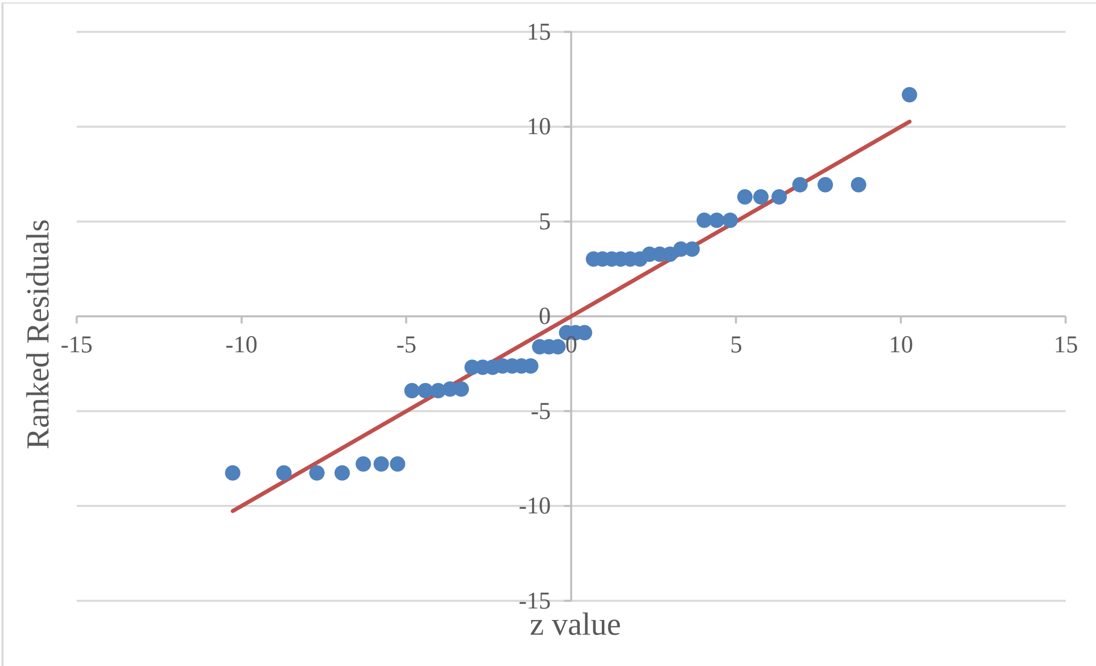

Figure 24. Normal probability plot for the residuals of the regression equation for AQP for the rock subgroup.

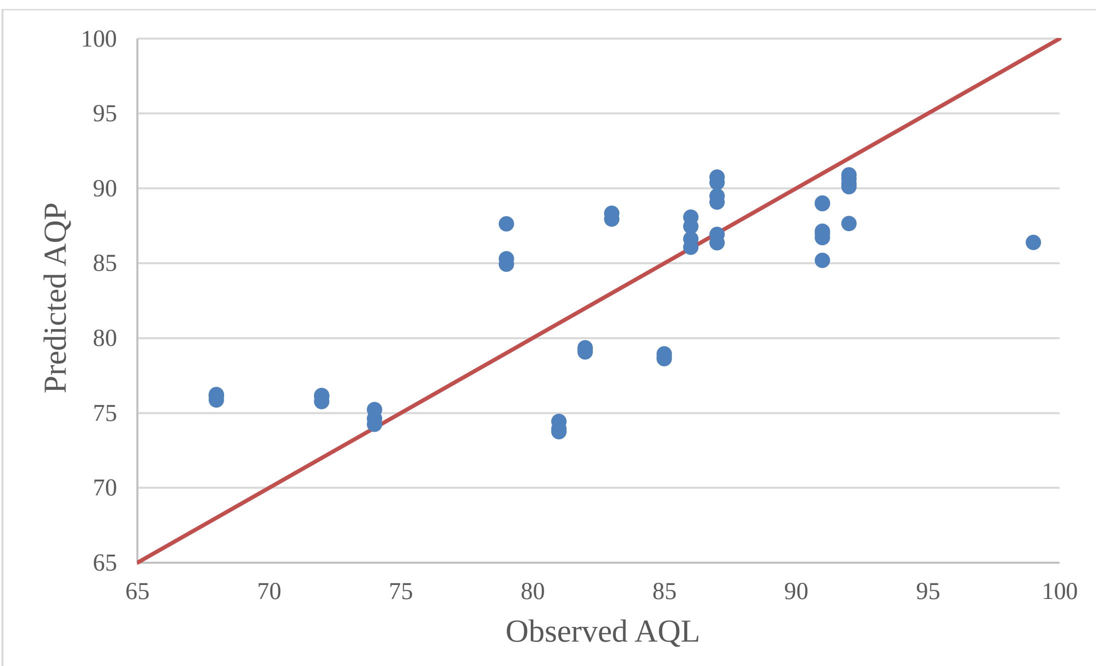

Figure 25. Observed AQL vs. predicted AQP for the rock subgroup.

Transitional Subgroup

There were not enough measured values of AQL in the transitional channel geology subgroup to do significant multiple regression. Therefore, simple linear regression analysis was done using each of the flow related independent variables (Q_S, Q_B, Q_{S+}, Q_{S-}, and Q_{peak}) with the observed AQL. The results of those regressions can be found in Table 21. All of the five simple linear regression equations were significant with p-values for the slope of the regression line less than $\alpha = 0.05$. The best fitting simple linear regression, came from the regression of Q_{peak} with AQL which had the highest R^2 value, 0.61, lowest standard error, 9.91, and the most significant p-value, 0.01, for the slope of the regression. The normal probability plot (Figure 26) shows that the residuals from the transitional channel geology subgroups plot do generally follow a linear pattern, but deviates from normal in both extremes. The R^2 indicates a moderate fit between the measured AQL versus the predicted AQP (Figure 27) with most of the points underestimated.

Table 21. Simple linear regression results for AQL versus each flow related independent variable for the transitional channel geology subgroup.

	Q_S	Q_B	Q_{peak}	Q_{S+}	Q_{S-}
AQP Equation	$70.2778 + 0.1496 \times Q_S$	$69.9649+0.2690\times Q_B$	$67.4091+0.0016\times Q_{peak}$	$70.0999+0.0241\times Q_{S+}$	$69.6004+1.1100\times Q_{S-}$
R^2	0.47	0.50	0.61	0.47	0.50
p-value	0.04	0.03	0.01	0.04	0.03
S_e	11.65	11.24	9.91	11.61	11.25

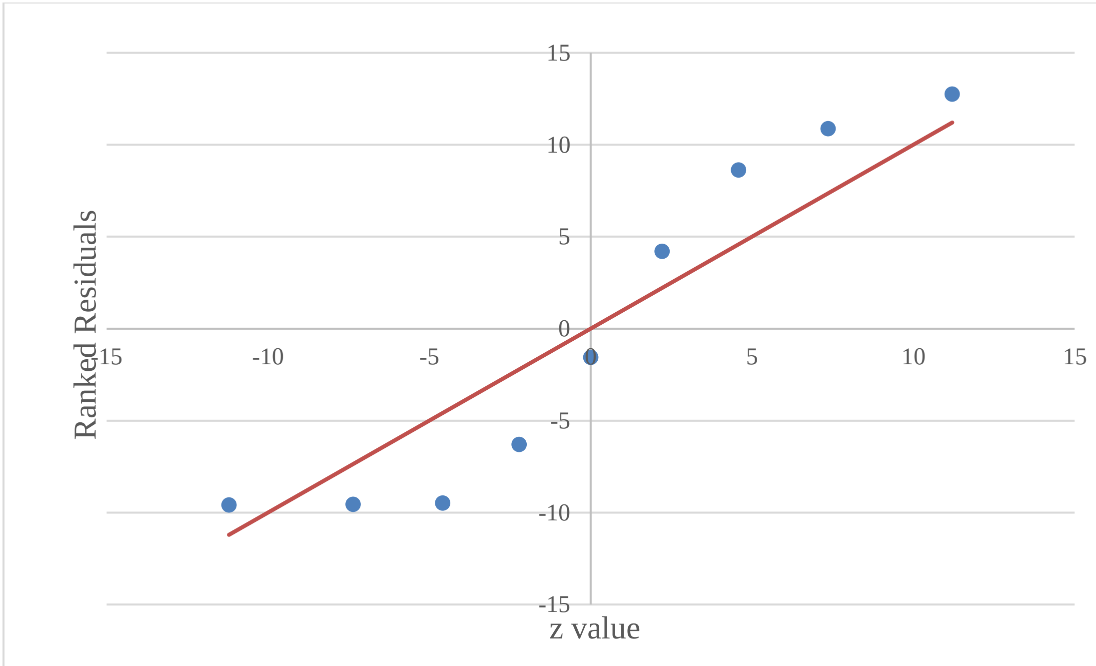

Figure 26. Normal probability plot for the residuals of the regression equation for AQP for the transitional channel geology subgroup.

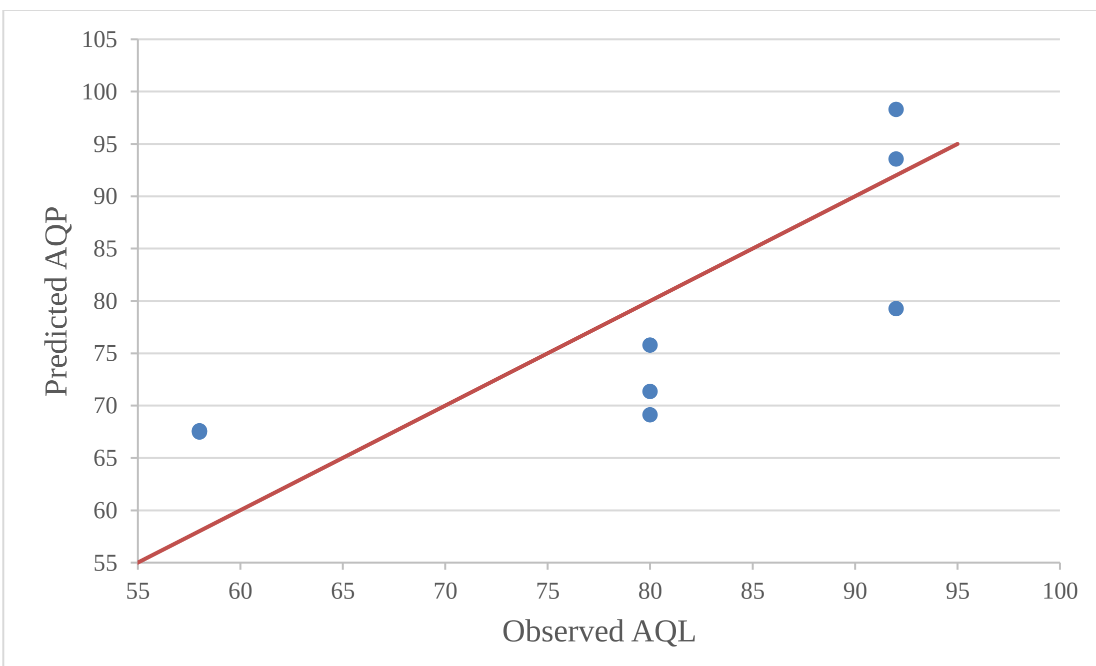

Figure 27. Observed AQL vs. predicted AQP for the transitional channel geology subgroup.

Prairie Subgroup

There were not enough measured values of AQL in the prairie subgroup to do significant multiple regression. Therefore, simple linear regression analysis was done using each of the flow related independent variables (Q_S, Q_B, Q_{S+}, Q_{S-}, and Q_{peak}) with the observed AQL. The results of those regressions can be found in Table 22. One of the five simple linear regression equations, Q_P, was insignificant with a p-value for the slope of the regression line of 0.68. The best fitting simple linear regression, came from the regression of Q_{S+}, with the highest R^2, of 0.87, the lowest standard error, 8.03, and the lowest p-value, 0.0002, for the slope of the regression. The normal probability plot (Figure 28) shows that the residuals from the prairie subgroup fit the normal distribution fairly well. The R^2 indicates a good fit between the measured AQL versus the predicted AQP, but figure 27 shows that the data is grouped in the extremes of AQL.

Table 22. Simple linear regression results for AQL versus each flow related independent variable for the prairie channel geology subgroup.

	Q_S	Q_B	Q_{peak}	Q_{S+}	Q_{S-}
AQP Equation	38.8151+0.9915×Q	$31.2962+2.5819\times Q_B$	$63.1182+0.0014\times Q_{peak}$	$22.3741+0.3245\times Q_{S+}$	$48.3840+2.6564\times Q_{S-}$
R^2	0.50	0.63	0.03	0.87	0.62
p-value	0.03	0.01	0.68	0.0002	0.01
S_e	15.67	13.54	21.91	8.03	13.65

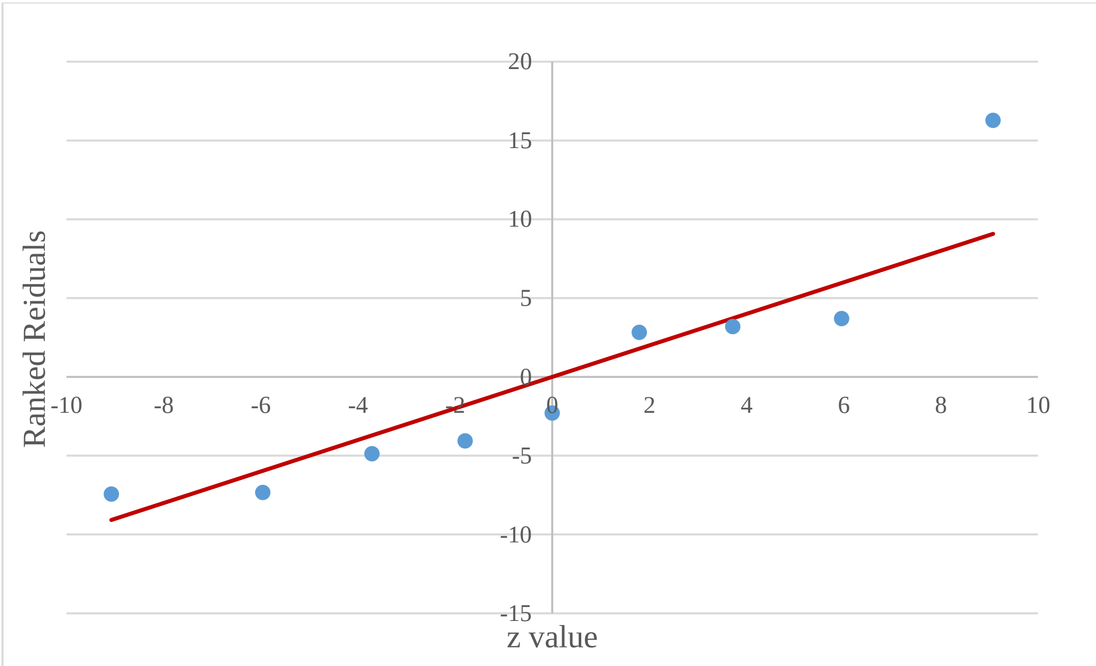

Figure 28. Normal probability plot for the residuals of the regression equation for AQP for the prairie channel geology subgroup.

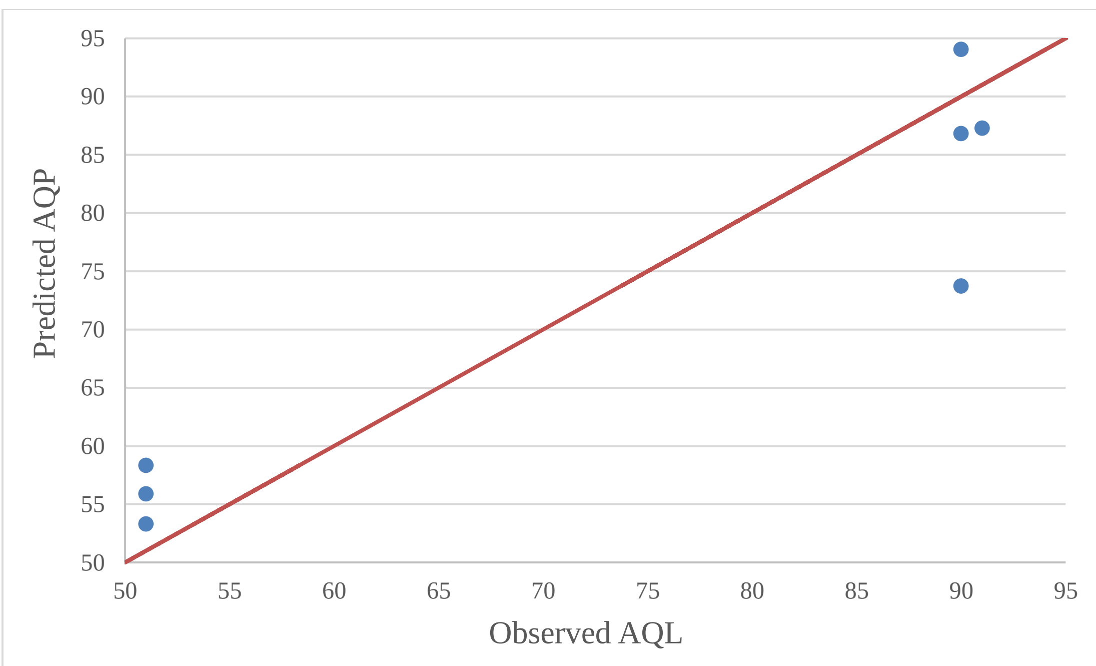

Figure 29. Observed AQL vs. predicted AQP for the prairie channel geology subgroup.

Discussion of Channel Geology Regression Equations

Table 20 shows that the average AQL increases when moving from prairie to transitional to rock, from 73 to 77 to 83. The standard deviations in the prairie and transitional groups are quite large, and their minimum and maximum values are similar, so there may not be a significant difference between them in terms of AQL. The prairie subgroup had the best fitting regression equation of the three with an R^2 of 0.87, but the standard error of the rock subgroup was the smallest. All three equations contained different independent variables. Rock was dependent only on the physical watershed characteristics of drainage area and impervious cover. Transitional was the only group of all the regression equations that had Q_{peak} as an independent variable and prairie was the only group to have Q_{S+} as an independent variable.

Discussion of All Regression Results

The regression analysis resulted in 14 equations for predicting AQP, 3 each in the drainage area, baseflow ratio, impervious cover and channel geology groups, and 2 in the baseflow group. Of those 14 equations, 6 of them contained Q_S as an independent variable, 6 contained IC, 5 contained L, 3 contained DA, 3 contained Q_B, 2 contained Q_S-, and Q_{S+}, Q_{peak} and S were all found in only one equation. Because Q_{90} was so closely correlated with Q_{peak} it was not used as one of the possible independent variables in the linear regressions, and it did not appear in any of the multiple regression equations. Of note is the number of equations that contain one or more watershed physical characteristics for predicting AQL because Glick et al. (2010) did not use these factors in their analysis. Nine of the 14 equations had one or more watershed physical characteristics and three of those contained only watershed physical characteristics as dependent variables (rock, impacted IC and semi-permanent baseflow). This may have been higher if the linear regressions had not been restricted to flow related characteristics only. IC was found in all of the groups except the DA group. L was found in 3 of the groupings, DA, baseflow and IC and DA was found in three of the groups, drainage area, baseflow ratio and channel geology.

This does not diminish the importance of the flow related characteristics. Eleven of the 14 equations had at least one flow related dependent variable. Q_S was found in all of the groups but channel geology. It is a little surprising that Q_B did not have more influence but it is important to remember that these streams were for the most part flowing continuously during the study period. The effect of Q_B may be more prominent when streams are drier.

4.3. Objective 3 Results: Select the "Best" Monitoring Site Grouping Method Based on Regression Analysis and Goodness-of-Fit to Observed AQL and Compare to Previous Work

Selection of Best Grouping Scheme

The Nash Sutcliffe Coefficient of Efficiency (NSE) and Root Mean Standard Error (RMSE) were calculated for each of the 14 regression equations in the five groups. NSE and RMSE values and the number of sites in each subgroup are shown in Table 23. These values along with the R^2 and standard errors of each of the equations were used to select one of the grouping schemes as the best representation for calculating AQP.

Table 23. NSE and RMSE values for each of the 5 groups and 14 regression equations predicting AQL.

	Subgroup	No. of Sites	NSE[a]	RMSE[b]
Drainage Area	Very small	9	0.53	2.66
	Small	8	0.16	11.22
	Large	7	0.94	0.65
Baseflow	Strictly permanent	12	0.40	9.30
	Semi-permanent	12	0.94	2.96
Baseflow Ratio	Poor	2	0.58	2.60
	Good	9	0.55	11.33
	Optimal	13	0.61	4.70
Impervious Cover	Sensitive watershed	9	0.53	3.79
	Non-supporting watershed	12	0.68	11.99
	Impacted watershed	3	0.97	0.99
Channel Geology	Rock	11	0.57	5.06
	Transitional	10	0.61	8.74
	Prairie	3	0.87	7.08

[a] NSE = Nash-Sutcliffe Efficiency
[b] RMSE = Root Mean Standard Error

The NSE values as a whole ranged from 0.16 (small drainage area group) to 0.97 (impacted IC group). All 14 NSE values were positive meaning that all of equations produced predictions for AQP that are better than just using the mean value of the AQL values. Moriasi et al. (2007) defined ranges for general performance of hydrologic and water quality models based on the NSE score, where an NSE between 0.75 and 1.00 is considered very good, between 0.65 and 0.75 is considered good, between 0.5 and 0.65 is considered satisfactory and less than 0.50 is considered unsatisfactory. Based on these categories all but two of the equations were at least satisfactory. Four of the equations can be considered very good, one was good, and the remaining seven were satisfactory.

The RMSE values ranged from 0.65 (large drainage area group) to 11.99 (non-supporting IC group). Most of the groups showed a wide range in the NSE and RMSE values within the group. The baseflow group had the most consistency between subgroups in terms of the NSE, but the RMSE varied widely.

Selection of the best grouping was based on distribution of the watersheds among the subgroups, a distinction between groups based on the measured AQL values, satisfactory or better NSE values and small values of RMSE. Groups that had 2 or more multiple regression equations were given priority over those that had several simple linear regressions that only included flow related independent variables.

Based on these criteria the Impervious Cover Group was selected as the best fitting group of equations to predicted AQL. There is a distinct pattern in the measured AQL values when moving from the lowest IC values (sensitive watersheds) to the highest IC values (non-supporting watersheds). All three equations were developed using multiple regression unlike the channel geology group where two of the equations were based on simple linear regression. The NSE values ranged from 0.53 to 0.97, higher than any other group and all in the satisfactory to very good range. The baseflow ratio group, for instance, had consistent NSE values among the subgroups, ranging from 0.55 to 0.61, but two of the three NSE values in the IC group were higher than 0.61 and the third was consistent with the baseflow ratio group (0.53). The drainage area group also contained one equation with an NSE that was very good (large drainage area) but also had one in the unsatisfactory category (0.16). While the non-supporting watersheds had the highest RMSE out of all of the equations developed, the RMSE for the other two subgroups were low.

4.4. Comparison to Previous Work

The hypothesis for this work was that grouping watersheds based on common characteristics would result in better predictive models for AQL than one equation for all watersheds in the City of Austin. Additionally, using independent variables that describe both the hydrology and the physical watershed will be more accurate than one just based on hydrology. To test this hypothesis the equation developed by Glick et al. (2010) that used only hydrologic characteristics of all the streams in the COA to predict AQP. The NSE and RMSE were calculated using the data used to develop that equation in 2010. Table 24 shows the results of this equation along with the results from the IC group, the best fitting group equations from this study.

Table 24. Comparison between Glick et al., (2010) AQP prediction equation and AQP with the best fitting group from this study.

IC	*Regression Equation (AQP=)*	*n*	R^2	*Se*	*NSE*[a]	*RMSE*[b]
Sensitive Watershed	$80.8313-96.4598\times IC+7.85\times 10^{-5}\times L+0.0918\times Q_S$	32	0.54	4.06	0.53	3.79
Non-supporting Watershed	$63.3630-0.0002\times L+1.3547\times Q_{S-}$	21	0.68	13.26	0.68	11.99
Impacted Watershed	$41.9829+218.8894\times IC$	10	0.97	1.11	0.97	0.99
Glick et al. (2010)	$63.417+3.914\times \ln(Q_{90})+12.041\times BF1-18.227\times T_{dry}$	24	0.70	N/A	0.212	15.84

[a] NSE = Nash-Sutcliffe Efficiency
[b] RMSE = Root Mean Standard Error

The IC group equations and the Glick et al. (2010) equation have no independent variables in common. The R^2 value for the Glick et al. (2010) equation falls into the good category according to the categories established by Moriasi et al. (2007) and falls right in the middle of the collective R^2 values for the IC group. However, the NSE, although positive, is much smaller than the NSE for all of the IC group equations. Additionally, the RMSE is higher than the RMSE for all of the IC group equations. Therefore, in this limited application, it appears that grouping watersheds based on their percentage of impervious cover and using both hydrologic and physically based watershed characteristics does result in better predictions of AQP than using a single equation using only hydrologic variables that describe all watersheds.

5. CONCLUSION AND FURTHER IMPROVEMENTS

5.1. Conclusions

The conclusion from grouping watersheds was that the IC has strong negative relationship with the drainage area and the baseflow ratio (Q_B/Q_S). Since larger watersheds are located in less developed areas with fewer monitoring sites, so the IC value decreases in large drainage area than in small drainage area. The IC value also had negative relationship with baseflow ratio. When the IC rises, the baseflow ratio drops. This can be explained by runoff, the high IC leads more streamflow and more runoff but the baseflow remain the same. Therefore, the baseflow ratio drops. Another conclusion is the distributions of all monitoring sites among subgroups varies between the groups. Not every groups has the same uniform distribution like the drainage area group.

The conclusion gained from developing regression equations was that the physical characteristics of a watershed have the same importance as the flow characteristics of a watershed. This could be found from dependent variables in regression equations. Nine of the 14 equations had one or more watershed physical characteristics and three of those contained only watershed physical characteristics as dependent variables in regression equations. At the same time, flow characteristics were still crucial in regressions forming. There were more than half of equations had at least one flow related dependent variables in regression equations.

The conclusion gained from selecting best group of equations was that the IC group was selected to be the best fitting regression group based on the number of regression equations formed in the IC group, accuracy of estimations in IC group and the easiness to acquire IC data.

First of all, based on the number of groups formed for multivariate regression analysis, only the IC and DA categories were able in forming three subgroups within each category with more than one group for multivariate regression analysis. The baseflow category did had two subgroups for multivariate regression analysis but it merely formed two instead of three subgroups. The baseflow ratio group and the channel geology group were not enough had subgroups formed in these categories for multivariate regression analysis so these groups cannot be the representative monitoring sites.

Secondly, all groups in IC group can predict AQP more precisely and accurately because the R^2 in each groups is generally higher. Since the R^2 did illustrate the fitness between AQP with AQL, and therefore, the sensitive watershed subgroup, impacted watershed subgroup and the non-supporting subgroup, Table 17 shown that 54%, 68% and 97% estimated AQP values from this study fit in with observed AQL score. In Glick and his colleagues' article(2010), they also used R^2 as an indicator to illustrates the fitness of a model. I believe IC value is an ideal indicator of existing stream water quality and the instream species living score since it was reported that in California, impervious cover development around watershed might decreases in the baseflow rate in streams and may adversely affect the stream ecology (Kitchell, 2003). In addition, the NSE and RMSE in the impacted watershed IC group closer enough to 1 and 0 than the NSE and RMSE in the rock channel geology subgroup, which are the most favorable values. A positive NSE for an equation indicates the residual variance in AQP is less than the residual variance in AQL as expected (Beven and Young, 2013; Moriasi et al., 2007). Therefore, in this study, the closer NSE is to 1 of the IC groups, the better goodness of fit is between the AQP and AQL, so the more predictive of the equations are. The NSE value of 0.68 and 0.97 in the non-supporting and the impacted IC group is close enough to 1 and the NSE values illustrate the best goodness of fit in IC group.

Last but not the least, the easiness for acquiring the IC values. Since most of the IC values comes from the EII estimation with the COA, it should not be too hard to get access to them for grouping purpose. Thus, using the IC groups to representing all 24 monitoring sites was a good way to acquire the estimated AQP values and it might help to save measuring time and human labor for COA.

The conclusion gained after comparing to Glick et al (2010) with the best fitting group of regression equation was that they were not sharing any independent variables. The regression equation from the impervious cover group resulted better predictions of AQP than the equation from Glick et al (2010). As stated above, the optimal values for NSE and RMSE are 1 and 0. Though the R^2 value for the Glick et al. (2010) equation was in the good category according to Moriasi et al. (2007), the NSE value for Glick et al. (2010) equation was much smaller than the NSE for all of the IC group equations; and the RMSE value was higher than the RMSE for all of the IC group equations. It means the predictions of AQP from Glick et al (2010) was less accurate than predictions of AQP from the impervious cover group.

5.2. Limitations in This Study and Further Improvements

The limitations in this study majorly comes from two parts: the small sample size and the poor groups shifting ability. From this study, further work should focus on using all data from Glick and his colleagues' study in 2010 to measure the differences between NSE and RMSE values between Glick (2010) and this study so it could be more obvious on differences. The second improvement from this study could be use multiple non-linear regression to analysis the relationship between variables since there were more than one subgroups did not shown the normal distribution in the residual plots.

First of all, it is true that with a larger sample size in known AQL, it can give more reliable predictions on AQP. In this study, none of subgroups had AQL observations larger than 50. If the total number of observations on AQL could enlarged, it would help to increase the number of monitoring sites distribution in the channel geology group and the baseflow group. Therefore, these groups can be compared with IC group for NSE and RMSE value to better decide which group can represents all the monitoring sites. The further improvement on NSE and RMSE values might focus on using all data applied in Glick and his colleagues' regression to measure the NSE and RMSE values. So that these values might be able to illustrate the difference between regressions from this study and Glick and his colleagues' regression.

Secondly, the poor interchange ability in grouping method does exists. When a certain monitoring site changes its grouping in the future, as far as to the concern in this study, a new set of regression equations need to be built for AQP prediction purpose. For example, when the IC of a monitoring site increases, it would belongs to another subgroup. Only by having enough AQL observations data, as well as other hydrologic data, promises the generation of the multivariable regression equations or the simple linear equations. Maybe collapsing subgroups and just form groups would be better for subgroups does not have enough AQL observations. I do believe collapse subgroups for baseflow group would make the regressions more reliable to be used because it will have more AQL observations in one group for multivariable regression analysis. To reclassify all five groups in this study into either the Environmental classifications or the Hydrologic classifications (Olden et al., 2012) might be a better way to improve this study. From these classifications, it includes both the physical characters of a stream system and the hydrologic metrics like the streamflow and baseflow rate (Olden et al., 2012).

REFERENCES

Allan, J. D., and Castillo, M. M. (2007). *Strm eco: structure and function of running waters.* Dordrecht: Springer (2nd ed. Vol.3).

Ansari, A. A. (2011). *Eutrophication: causes, consequences and control.* Dordrecht ; New York : Springer. (1st ed. Vol.1) MI:ASABE

Arnold Jr., C. L., and Gibbons., C. J. (1996). Impervious surface coverage: The emergence of a key environmental indicator. *Journal of the American Planning Association,* 62(2), 243-258.

Asquith, W. H. (1998). Peak-flow frequency for tributaries of the Colorado River downstream of Austin, Texas. 98–4015. U.S. Geological Survery and Lower Colorado River Authority.

Austin, S. H. (2014). *Methods for estimating drought streamflow probabilities for Virginia streams*. Scientific investigations report: 2014-5145. Reston, Virginia: U.S. Department of the Interior, U.S. Geological Survey.

Berhanu, B., Seleshi, Y., Demisse, S. S., and Melesse, A. M.. (2015). Flow regime classification and hydrological characterization: A case study of ethiopian rivers. *Water (20734441)* 7(6),3149-3165.

Beven, K., and Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. *Water Resources Research* 49(8):5092-5098.

Chai, T., and Draxler, R. R.. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7(3):1247-1250.

City of Austin. (2000). The Code of City of Austin,Texas. Austin Texas goverment. Available at: https://www.municode.com/library/tx/austin/codes/land_development_code?nodeId=TIT25LADE_CH25-8EN_SUBCHAPTER_AWAQU_ART1GEPR_DIV1DEDEREAR_S25-8-1DE. Access date: March 23rd 2016

City of Austin. (2001). Watershed Protection Master Plan: Phase I Watersheds Report. COA-WPD-2001-02. Watershed Protection Report Series Available at: http://www.austintexas.gov/department/watershed-protection-master-plan Access date:15th June 2016

City Of Austin. (2009). WPDRD Master Plan Water Quality Problem Score Method Revision. SR-08-09 Watershed Protection and Development Review Department Environmental Resource Management Division: W. P. D. Review.

City of Austin. (2015a). Austin Watersheds List. City of Austin Official website. Available at: http://www.austintexas.gov/page/austin-watersheds-list. May 25th

City of Austin. (2015b). Watershed Protection Master Plan 2015 Update. Watershed Protetion: City of Austin. Available at: http://www.austintexas.gov/watershed_protection/publications/document.cfm?id=240265.

Croke, B. F. W. (2009). Representing uncertainty in objective functions: Extension to include the influence of serial correlation. *The Fenner School of Environment and Society* 13(17), :3372-3378.

Dunne, T.,& Leopold, L. B..(1978). *Water in environmental planning.* San Francisco: W. H. Freeman.

Erpul, G., Norton, L. D., &Gabriels, D..(2003). Sediment transport from interrill areas under wind-driven rain. *Journal of Hydrology* 276;184-197.

Fetter, C. W. (2001). *Applied hydrogeology (4th ed.).* Upper Saddle River, N.J.: Prentice Hall.

Franz, K. J., & Hogue,T. S..(2011). Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community. *Hydrology & Earth System Sciences* 15(11):3367-3382.

Fritz, K. M., & Dodds,W. K.. (2005). Harshness: characterisation of intermittent stream habitat over space and time. *Marine and Freshwater Research* 56(1);13-23.

Glick, R. H., Gosselink L., Bai B., & Herrington C..(2010). Impacts of stream hydrologic characteristics on ambient water quality and aquatic health in the Austin, Texas area. W. P. Department, ed. City of Austin.

Gupta, H. V., & Kling, H.. (2011). On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. *Water Resources Research* 47(10).

Haan, C. T. (2002). *Statistical methods in hydrology*. 2nd ed. Ames: Iowa State Press, 2002.

Harmel, R. D., & Smith, P. K.. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology*.

Harmel, R. D., Smith, P. K., & Migliaccio, K. W.. (2010). Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. *Transactions of the ASABE* 53(1), 55-63.

Herrington, C. (2003). Change in environmental integrity index values in the Austin, Texas, area (1996-2002). Watershed Protection and Development Review Department (WPDRD) SR-03-06. City of Austin Watershed Protection.

Herrington, C. (2011). Action plan items related to EII site scores - fiscal year 2010. SR-11-09. City of Austin Watershed Protection.

Hollis, G. E. (1975). The effects of urbanizstion on floods of different recurrence interval. *Water Resources Research* 11(3):431-435.

Johns, D. A. (1991). Water quality issue for Barton Creek and Barton Springs. Austin Geological Society: B. S. T. Force.

Jordan, M. A. (1977). Guidebook to the Geology of Travis County. The Student Geology Society. Available at: https://www.lib.utexas.edu/geo/ggtc/ch3.html. Aug 21 2015

Kitchell, C. A. (2003). Impacts of impervious cover on aquatic systems. Watershed Protection Research Ellicott City, MD 21043: C. f. W. Protection.

Klein, R. D. (1979). Urbainzation and stream quality impairment. *Journal of the American Water Resources Association* 15(4), 948-952.

Legates, D. R., & McCabe G. J.. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35(1), 233-241.

Marks, J. C., Power, M. E., and Parker, M. S.. (2000). Flood disturbance, algal productivity, and interannual variation in food chain length.20. Munksgaard International Publishers, Ltd.

McClintock, N. L. (2002). Environmental integrity index methodology. W. P. Department, ed. Water Quality Report Series: City of Austin

Moriasi, D. N., Arnold,J. G., and Van Liew, M. W.. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50(3), 885-900.

Nash, J. E., and Sutcliffe, J. V.. 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology* 10(3), 282-290.

NOAA. (2007). Austin climate summary. National Oceanic and Atmonospheric Adminstration. Available at: http://www.srh.noaa.gov/images/ewx/climate/ausclisum.pdf. Janauary 21st

Olden, J. D., Kennard, M. J., & Pusey, B. J.. (2012). A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* 5(4):503-518.

Omernik, J. M. (1986). *Ecoregions of the conterminous United States.* Corvallis, Or.: Corvallis Environmental Research Laboratory, U.S. Environmental Protection Agency.

Omernik, J. M., & Griffith, G. E.. 2013. Ecoregions of Texas (EPA). The encyclopedia of earth. Available at: http://www.eoearth.org/view/article/152207/. April 28th 2016

Paul, M. J., & Meyer, J. L..(2001). Streams in the urban landscape. *Annual Reviews Inc*. 21(17),456-468.

Platt, W. J., & Connell, J. H.. (2003). Natural disturbances and directional replacement of species. *Ecological Society of America*.17(08), 698-715

Poff, N. L., Allan, J. D., Bain, M. B., Karr,J. R., Prestegaard, K. L., Richter B. D., Sparks R. E., and Stromberg J. C.. (1997). The natural flow regime: A paradigm for river conservation and restoration. *BioScience* 47(11):769-784.

Porras, A., & Scoggins, A. M.. (2013). The flow permanence index: A statistical assessment of flow regime in Austin streams. City of Austin Watershed Protection Department

Post, D. M. (2002). The long and short of food-chain length. *Trends in Ecology & Evolution* 17(6), 269-271.

Richter, A. (2011). Creation of a multi-metric index for describing the environmental integrity of Austin-area lakes. City of Austin SR-11-19. Watershed Protection Department.

Richter, A. (2012). Modeling benthic algae growth in Lake Creek, a high nutrient stream in Austin, TX. 07. City of Austin Watershed Protection.

Richter, B. D. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* 10(4),1163-1174.

Schueler, T. (1994). The importance of imperviousness. *Watershed Protection Techniques* 1(3), 100-111.

Scoggins, M. (2000). Effects of hydrology on bioassessment in Austin, Texas.Avaiable at http://www.austintexas.gov/watershed_protection/publications/document.cfm?id=186264 City of Austin

Scoggins, M., Richter A., Meade, J., & Murray, K.. (2013). An index of riparian integrity for the Austin Area SR-13-09. Avaiable at http://www.austintexas.gov/watershed_protection/publications/document.cfm?id=206342 City of Austin.

Sousa, W. P. (1984). The role of disturbance in natural communities.5(28), 557-568. Annual Reviews Inc.

Tennant, D. L. (1976). Instram flow regimens for fish, wildlife, recreation and related environmental resources. *Fisheries* 1(4):6-10.

Texas Parks and Wildlife Department. (2015). Texas Ecoregions Available at: https://tpwd.texas.gov/education/hunter-education/online-course/wildlife-conservation/texas-ecoregions. April 6th

Thompson, J. (2007). Using benthic macroinvertebrates and GIS to assess and manage watershed health of the Colorado River basin. Center for Research in Water Resource: The Univerisity of Texas at Austin. Available at: https://www.crwr.utexas.edu/gis/gishydro05/Introduction/TermProjects/thompson.htm. Feburary 20th

U.S. Census Bureau. (2009). Population finder: Austin City, Texas. U.S. Deparment of Commerce. Available at: http://factfinder.census.gov. March 22 2016

Vitvar, T., Burns, D. A., Lawrence, G. B., McDonnell,J. J., & Wolock, D. M.. (2002). Estimation of baseflow residence times in watersheds from the runoff hydrograph recession: method and application in the Neversink watershed, Catskill Mountains, New York. *Hydrological processes* 16(9),1871-1877.

White, K. L., & Chaubey I.. 2005. Sensitivity analysis, calibration and validations for a multisite and multivariable SWAT model. *Journal of the American Water Resources Association* 41(5):1077-1089.

Xu, L., Liu, J., Jin, C., Wang, A., Guan, D., Wu, J., & Yuan, F. 2011. Baseflow separation methods in hydrological process research: A review. *Yingyong Shengtai Xuebao*, 22(11), 3073-3080.

Zhan, H. 2015. Lecture on Advanced Hydrogeology.

Zhu, T. 2015. EII sites baseflow and streamflow data.

Zimmerman, J. K. H., Lutz, K. A., Magilligan, F. J., Letcher, B. H., & Nislow K. H.. 2010. Determining the effects of dams on subdaily variation in river flows at a whole-basin scale [electronic resource]. *River research and applications* 26(10):1246-1260.

APPENDIX A

Site Number	Year	Streamflow	Baseflow	IC	AQP	AQL
08154700	1993	14.811	0.000	0.117	76.11	68.00
08154700	1994	12.369	0.000	0.117	79.90	68.00
08154700	1995	12.849	0.000	0.117	79.30	68.00
08154700	1996	6.834	0.000	0.117	80.28	68.00
08154700	1999	3.906	0.000	0.153	78.18	74.00
08154700	2000	12.309	0.000	0.153	84.03	74.00
08154700	2001	26.088	0.000	0.153	71.48	74.00
*08158600	1993	19.897	13.189	0.263	72.92	51.00
*08158600	1994	27.245	9.809	0.263	72.08	51.00
*08158600	1995	28.400	13.398	0.263	73.76	51.00
*08158600	1996	17.620	6.792	0.263	72.98	51.00
*08158600	1997	45.010	22.930	0.283	74.54	91.00
*08158600	1998	43.093	20.068	0.283	74.60	91.00
08155200	1997	84.043	60.055	0.029	76.81	83.00
08155200	1998	75.592	57.321	0.029	72.04	83.00
08158810	1999	1.131	1.026	0.022	86.39	85.00
08158810	2000	4.094	2.921	0.022	86.98	85.00
08158810	2001	8.605	7.140	0.022	73.72	85.00
08155200	1999	8.308	7.306	0.030	76.49	79.00
08155200	2000	16.386	13.872	0.030	72.63	79.00
08155200	2001	69.179	54.862	0.030	72.06	79.00
08158810	2002	11.655	7.529	0.094	73.74	72.00
08158810	2003	4.023	3.468	0.094	73.83	72.00
08158810	2004	13.330	7.751	0.094	86.42	72.00
08154700	2002	16.697	0.000	0.174	74.23	81.00
08154700	2003	12.503	0.000	0.174	73.59	81.00
08154700	2004	27.701	0.000	0.174	73.93	81.00
*08158600	2002	40.092	18.086	0.296	85.13	90.00
*08158600	2003	24.221	14.007	0.296	88.11	90.00
*08158600	2004	63.701	26.703	0.296	74.79	90.00
08155200	2002	75.435	47.345	0.052	75.25	91.00
08155200	2003	34.539	31.102	0.052	76.72	91.00
08155200	2004	78.404	52.252	0.052	76.72	91.00

(Appendix A continues)

Site Number	Year	Streamflow	Baseflow	IC	AQP	AQL
08155240	1993	33.597	0.000	0.027	86.02	86.00
08155240	1994	14.682	0.000	0.027	86.51	86.00
08155240	1995	47.474	0.000	0.027	85.22	86.00
08155240	1996	1.923	0.000	0.027	86.95	86.00
08155240	2002	90.417	64.147	0.055	90.76	91.00
08155240	2003	39.317	36.154	0.055	89.61	91.00
08155240	2004	91.391	65.516	0.055	90.30	91.00
08158700	2002	102.438	30.413	0.067	91.65	92.00
08158700	2003	42.838	12.114	0.067	92.50	92.00
08158700	2004	110.341	36.343	0.067	91.86	92.00
08155400	2002	84.268	61.291	0.038	87.00	87.00
08155400	2003	55.118	50.652	0.038	86.71	87.00
08155400	2004	92.763	67.260	0.038	87.25	87.00
*08159000	2002	131.415	83.856	0.097	92.74	92.00
*08159000	2003	47.962	36.339	0.097	90.41	92.00
*08159000	2004	184.198	98.229	0.097	92.86	92.00
08155240	1997	114.679	85.810	0.031	92.02	92.00
08155240	1998	96.890	78.329	0.031	91.96	92.00
08155240	1999	10.136	8.936	0.032	87.31	87.00
08155240	2000	22.535	17.975	0.032	87.29	87.00
08155240	2001	80.897	65.325	0.032	87.36	87.00
08158840	2002	6.261	4.555	0.009	81.78	82.00
08158840	2003	2.838	2.458	0.009	81.78	82.00
08158840	2004	9.158	5.034	0.009	81.78	82.00
*08157500	2002	2.550	0.658	0.500	73.56	80.00
*08157500	2003	3.744	0.862	0.500	75.12	80.00
*08157500	2004	5.363	1.394	0.500	77.24	80.00
*08157700	2002	4.747	1.228	0.385	76.43	58.00
*08157700	2003	4.715	1.815	0.385	76.39	58.00
*08157700	2004	8.317	1.557	0.385	81.10	58.00
*08158840	1997	8.683	5.554	N/A	81.58	82.00
*08158840	1998	8.689	6.009	N/A	81.59	82.00

(Appendix A Continues)

Site Number	Year	Streamflow	Baseflow	IC	AQP	AQL
*08156800	1999	2.883	1.157	N/A	N/A	54.00
*08156800	2000	8.423	1.415	N/A	N/A	54.00
*08156800	2001	12.48	1.743	N/A	N/A	54.00
*08156800	2002	5.4	1.84	N/A	N/A	62.00
*08156800	2003	18.9	4.35	N/A	N/A	62.00
*08156800	2004	7.21	5.04	N/A	N/A	62.00

*Site Number means the site was used in simple regression analysis because short on IC or observed AQL data or not had enough sites in the subgroup.
In this study, FTB and FBU were not used in either regression analysis because no IC or observed AQL data provided.