THESIS

PREDICTION OF SELENIUM IN SPRING CREEK AND FOSSIL CREEK, COLORADO

Submitted by

Adam L. Pierce

Department of Ecosystem Science and Sustainability

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2014

Master's Committee:

Advisor: John D. Stednick

Randall B. Boone
Christopher I. Thornton

ABSTRACT

PREDICTION OF SELENIUM IN SPRING CREEK AND FOSSIL CREEK, COLORADO

The role and importance of selenium as an environmental contaminant has gained widespread attention among research scientists, natural resource managers, and federal and state regulatory agencies during the last two decades. Selenium has been listed on Colorado's Clean Water Act Section 303(d) List of Impaired Waters for Spring Creek and Fossil Creek in the city of Fort Collins. Selenium is one of the most hazardous of the trace metals, following mercury, with a narrow range between dietary deficiency and toxicity. Identifying selenium sources and understanding the environmental processes controlling how selenium is introduced to streams is critical to managing and mitigating the effects of elevated concentrations.

A modeling approach was used to predict selenium concentrations with exploratory variables including 15 geospatial landscape parameters, precipitation, and streamflow for 5 sub-watersheds within Spring Creek and Fossil Creek watersheds. A correlation analysis was applied with surface water selenium concentrations and the better exploratory variables identified. Selected variables were used in a multiple linear regression model. Various combinations of different variables determined the best performing model, and included the area of shale, area of moderate to strongly alkaline soils, and the length of streams with an adjusted R^2 of 0.99, [Se µg/L = 24.038 + 9.516(ALK) – 0.782(STR) -1.039(SHL)]; where ALK = area (km^2) of moderate to strongly alkaline soils; STR = length (km) of streams; SHL = area (km^2) of shale.

Additional multiple linear regression models were developed in ArcGIS® using Ordinary Least Squares (OLS) Regression, and Geographically Weighted Regression (GWR) with area weighted geospatial variables. The best performing OLS model used only area (km^2) of wetlands, with an adjusted R^2 of 0.98, [Se μg/L = -6.584 + 170.509(wetlands)]. Similarly, the best performing GWR model included area of wetlands, with an adjusted R^2 of 0.98. The second best performing GWR model included area of shale, with an adjusted R^2 of 0.66.

Limitations of this model include a very small sample size of water quality sampling stations, which limits the statistical power of multiple regression models used. Additional techniques applied in basin delineations with landscape element coupling for identification of hydrologic and/or chemical response units can further develop the platform for future modeling efforts targeting unmonitored watersheds.

ACKNOWLEDGEMENTS

I gratefully acknowledge my advisor Dr. John Stednick for his guidance, expertise, and willing involvement in this project and throughout my graduate school career. I was very fortunate to study under someone with such dedication, excitement, and unique perspectives in the fields of water quality and watershed science. I extend my gratitude to Dr. Randall Boone for serving as a committee member and offering advice, constructive criticism, and for assisting in the development of my geospatial research concepts. I am also grateful for Dr. Chris Thornton's contributions as a committee member and for providing me with fundamental and technical support in the development of this research. His assistance and encouragement in professional growth and career endeavors have been most appreciated and respected.

Completion of this research would not have been possible without the support and love provided by my family and friends. My parents, Lee and Sandi, deserve special mention for their unwavering support and love, and for always believing in me. Their encouragement to pursue my dreams has allowed me to continue undertakings previously not thought possible. Finally, I would like to thank my sister Emily and brother Aaron for their continual support and endless encouragement.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION **1**

1.1 Sources and Effects of Selenium 1

1.2 Problem Definition 11

1.3 Statistical Models 14

1.4 GIS Models 15

1.5 Hypothesis and Objectives 19

CHAPTER 2 - METHODS **21**

2.1 Site Description 21

2.2 Water Quality and Streamflow 23

2.3 Geographic Information System 27

2.4 Statistical Model 42

2.5 GIS Statistical Model 45

CHAPTER 3 – RESULTS **48**

3.1 Overview 48

3.2 GIS Cartographic Display of Landscape Elements 48

3.3 Statistical Analysis of Se Concentrations 54

3.4 Geospatial Statistical Analysis of Landscape Parameters 60

3.5 Statistical Model Results 60

3.6 GIS Statistical Model Results 66

CHAPTER 4 – DISCUSSION **70**

4.1 Integration of Landscape Elements within a GIS Environment 70

4.2 Development of Traditional and GIS Statistical Models for Se Prediction 72

CHAPTER 5 – CONCLUSIONS 78

CHAPTER 6 - RECOMMENDATIONS 81

LITERATURE CITED 83

APPENDIX............ 90

APPENDIX A - Soil Data Viewer User Interface (pHwater)............ 91

APPENDIX B - Aggredation Report 93

APPENDIX C - USDA’s NRCS Soil Survey Classification 97

APPENDIX D - Land Use Land Cover Classification Key 99

APPENDIX E - OLS map output of regression residuals (over/under predictions) 103

APPENDIX F - Spatial Autocorrelation Report 104

APPENDIX G - Summary of OLS Results and OLS Diagnostics 105

APPENDIX H - Geographic Weighted Regression Results Output............ 106

LIST OF TABLES

Table 1 – Chemical forms of selenium in the environment (adopted from Fordyce, 2005).... 6

Table 2 – Colorado's Section 303(d) List of Impaired Waters on the Cache la Poudre River (CDPHE, 2010c). 12

Table 3 – Water quality sampling stations and locations on Spring Creek and Fossil Creek24

Table 4 – Streamflow gauge stations on Spring Creek and Fossil Creek........ 26

Table 5 – Landscape elements used in GIS model with data source and website. 28

Table 6 – Original and reclassified NHD flow line types for Spring Creek and Fossil Creek watersheds from USGS, 2010 35

Table 7 –SSURGO soil pH values found in the Spring Creek and Fossil Creek watersheds with modified class categories from USDA, 1993........ 37

Table 8 – Land Use Land Cover original and reclassified classes in the Spring Creek and Fossil Creek watersheds from MRLC, 2012 38

Table 9 –Landscape elements, derived parameters, spatial statistic types and geospatial variables 41

Table 10 – Summary statistics of selenium concentrations from Spring Creek........ 56

Table 11 – Summary statistics of selenium concentrations from Fossil Creek 58

Table 12 – Landscape parameter values for Spring Creek and Fossil Creek sub-watersheds. 61

Table 13 – Landscape parameter values are area-weighted for Spring Creek and Fossil Creek sub-watersheds. Total area values are not weighted........ 62

Table 14 – Correlation analysis output with Pearson coefficients indicated in bold and underlined for strong variable relationships with selenium 63

Table 15 – Continuation of the correlation analysis output........ 63

Table 16 – Multiple regression dependent and independent variables 65

LIST OF FIGURES

Figure 1 – Eh-pH diagram of Se in soils (from Dissanayake and Chandrajith, 1999) 8

Figure 2 – Location map of the Cache la Poudre River and study area of Spring Creek and Fossil Creek near Fort Collins, Colorado (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset , National Atlas) 22

Figure 3 – Map of Colorado Department of Public Health and Environment water quality sampling stations and City of Fort Collins streamflow gauging stations on Spring Creek and Fossil Creek (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset).. 25

Figure 4 – Map of Spring Creek and Fossil Creek delineated watershed boundaries (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset)..... 30

Figure 5 - Map of sub-basin watershed boundaries within Spring Creek and Fossil Creek watershed boundaries. Numbers denote sub-basin watershed labels. (Data source: USDA – NRCS Geospatial Data Gateway) 32

Figure 6 – Map of sub-watershed layers delineated upstream of each CDPHE water quality sampling station for Spring Creek and Fossil Creek with CDPHE water quality sampling stations (Data source: USDA – NRCS Geospatial Data Gateway) 33

Figure 7 – ModelBuilder layout showing workflow from left to right from top image to bottom image... 40

Figure 8 – Map of irrigated parcels within Spring Creek and Fossil Creek watershed boundaries (Data source: CWCB – DWR – CDSS Colorado Decision Support Systems)......... 49

Figure 9 – Map of Cretaceous age geology rock type shale within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS Geospatial Data Gateway)........... 50

Figure 10 – Map of land cover data classes within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS Geospatial Data Gateway) 51

Figure 11 – Map of soil pH values from the soil surface of 0 cm to 122 cm depth within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS – SSURGO Soil Survey Geographic Database)... 52

Figure 12 – Map of hydrography flowline types within Spring Creek and Fossil Creek watershed boundaries (Data source: USGS National Hydrography Dataset)............................ 53

Figure 13 – Plot of selenium concentrations over time from water quality samples taken from Spring Creek at Shields Street ... 55

Figure 14 - Plot of selenium concentrations over time from water quality samples taken from Spring Creek at College Avenue ... 55

Figure 15 – Plot of selenium concentrations over time from water quality samples taken from Spring Creek at Edora Park ... 55

Figure 16 – Plot of selenium concentrations over time from water quality samples taken from Fossil Creek at College Avenue ... 57

Figure 17 – Plot of selenium concentrations over time from water quality samples taken from Fossil Creek at Trilby Road ... 57

Figure 18 – Probability distribution functions. Cumulative frequency vs. selenium concentrations with 85^{th} percentile concentrations identified ... 59

Figure 19 – Histogram depicting the 85^{th} percentile of selenium concentrations with the CDPHE chronic selenium TVS identified ... 59

Figure 20 - Measured versus predicted 85th selenium concentration values (μg/L) for Excel multiple linear regression and Geographic Weighted Regression models. ... 68

Figure 21– Geographic Weighted Regression output with the wetland coefficient mapped for each sub-watershed ... 69

CHAPTER 1 – INTRODUCTION

1.1 Sources and Effects of Selenium

Selenium (Se) is a nonmetal related to sulfur (S) and tellurium (Te), and is toxic at elevated concentrations (Lemly, 1993; Presser, et al., 1994). Sources of Se include irrigation drainwater, fly-ash from coal-fired power plants, weathering of seleniferous shales and soils, and releases from metallic ore mining and smelting (Canton and Derveer, 1996). Other sources are selenized fertilizer and municipal wastewater discharge (Shamberger, 1983; Fordyce, 2005). Distributed naturally across the globe, areas underlain by Cretaceous marine or sedimentary rocks that are weathered and eroded can produce high selenium soils in many areas of the western United States, notably the San Joaquin Valley of California (Presser, et al., 1994), Montrose Arroyo Basin (Butler, 2001), Lower Gunnison River Basin (Butler and Leib, 2002), and Grand Valley of the Colorado River in Western Colorado (Leib, 2008). Shales are the principal sources of selenium-toxic soils of the Rocky Mountain foothills (Shamberger, 1983).

Little industrial application was made of selenium until the early 20th century when it began to be used as a red pigment, and to improve glass and ceramic manufacturing (Fordyce, 2005). Since the 1930s, selenium is widely used in insecticides, fertilizer, and anti-dandruff/antifungal pharmaceuticals, as well as the manufacture of electrodes, photocells, selenium cells, semiconductor fusion mixtures, toning baths in photography and X-ray xerography (Shamberger, 1983; Fordyce, 2005).

Bioconcentration, bioaccumulation and biomagnification of selenium in aquatic and terrestrial systems has been studied, but only in the last few decades. Adverse effects in aquatic birds due to selenium had not been reported until 1986, when high rates of embryonic deformity and death in wild aquatic birds was documented in the San Joaquin Valley of California at the Kesterson National Wildlife Refuge (Ohlendorf et al., 1986). The wildlife refuge is the terminus of subsurface agricultural drains with Se concentrations ranging between 100 and 350μg/L. This study showed deformities that were often multiple and included missing or abnormal eyes, beaks, wings, legs and feet due to Se biomagnifications, attributed to Se toxicosis. The selenium source was determined to be selenium carried by irrigation drainwater (Ohlendorf et al., 1986).

Because concerns were expressed by the United States Congress and environmental groups that adverse effects from irrigation drainwater might occur elsewhere in the nation, the United States Department of the Interior (USDOI) implemented the National Irrigation Water Quality Program (NIWQP) (Seiler, 1995). The purpose of this program is to complete comprehensive surveys of USDOI irrigation-drainage land and facilities by making evaluations of water quality, biologic, and geologic data, which are used to identify areas with contamination problems that warrant reconnaissance investigations. Reconnaissance investigations were designed to use existing information and the results of field-screening studies conducted by USDOI to determine whether irrigation drainage has caused, or has the potential to cause, harmful effects on human health, fish and wildlife, or beneficial use of water (Seiler, 1995). The middle Arkansas River basin, in southeastern Colorado and southwestern Kansas was selected for a reconnaissance investigation in 1988, where samples of water, bottom sediment, and biota were collected. Surface water selenium concentrations were found to range from 1μg/L to 52μg/L and the maximum concentration found in ground water was 29μg/L. The maximum concentration found in bottom sediment was 5.4µg/L. It was determined that drainage from irrigated land underlain by marine shale was the selenium source (Mueller et. al., 1991).

Selenium has been found to bioconcentrate and biomagnify in the planktonic food chain with resultant dietary toxicity to fish and aquatic birds (May et al. 2008). A study in the Solomon River Basin of Kansas found that all benthic invertebrate samples (n=20) except one and 97% of the fish sample set (n=195) exhibited selenium concentrations considered to be ranked as a high hazard. Dietary toxicity and reproductive impairment were found to occur in fish and aquatic birds in excess of 5µg/L and 4µg/L, respectively. In Belews Lake, NC, adverse effects in green sunfish (*Lepomis cyanellus*) were observed with selenium concentrations as low as 5 to 10µg/L, and dietary selenium toxicity has been shown in similar concentrations of 5 to 10µg/L dry weight (Sorensen et al., 1984; Goettl and Davies, 1978). Biomagnification of selenium in water can be concentrated from 100 to more than 30,000 times in the food organisms eaten by fish and wildlife, which exposes them to a highly concentrated dietary source of contamination further affecting offspring in eggs (Lemly, 1999).

Selenium is one of the most hazardous of the trace metals, following mercury (Luoma and Rainbow, 2008), yet nutritionally essential in small amounts. Of all the elements, it has one of the narrowest ranges between dietary deficiency (<40µg/day) and toxic levels (>400µg/day) (Fordyce, 2005). It has been stated that everything is toxic, it is just a matter of dose; however, toxicity can be defined by the dose with which causes adverse health effects or more specifically by the factors influencing uptake, critical organ, critical dose, critical effects, and biological half-life (Nordberg and Cherian, 2005). Bioaccumulation and biomagnification can occur in aquatic insects, fish, plants and terrestrial animals greatly increasing the threat of contaminate exposure, yet further research is needed to understand the complex nature of selenium due to the various elemental forms and physiological factors that affect toxicity. There is still a great deal of uncertainty about harmful doses of selenium to humans, but a maximum recommended dietary intake of 400μg/day has been proposed (WHO, 1996).

Many of the symptoms found in terrestrial animals with high selenium intake can be expected to occur in humans as well through drinking water or other dietary intake from foods. On the basis of selenium requirement studies, a range of 50-200μg/day has been recommended by the U.S. National Research Council (NRC) for adults, however plants and animals retain the element in great concentrations due to its bioaccumulative nature in which selenium can be bioconcentrated 200-6000 times (Fordyce, 2005). Therefore the most important exposure route to selenium for animals and humans is the food eaten, as concentrations are orders of magnitude greater than in water and air in most cases (WHO, 1996).

Acute oral doses of selenite and other selenium compounds cause symptoms such as nausea, diarrhea, abdominal pain, chills, tremor, and numbness in limbs, irregular menstrual bleeding, and marked hair loss (WHO, 1996). Irrigation and flooding can provide the mechanism in seleniferous soils for plant and crop uptake of precipitated selenium via wetting and drying of soils. Drinking water from wells can receive contaminated water through the percolation of surface water and movement of selenium bearing groundwater. One documented case of selenium toxicity in a water source shows a family exposed for about 3 months to well-water containing 9,000μg/L selenium. They suffered from loss of hair, weakened nails, and mental symptoms but recovered when they stopped using the water from the well (WHO, 1996).

The relationship between human illness and a drinking water supply containing high and low concentrations of selenium was examined in two representative samples of residents from a rural Colorado community. One sample group was exposed to high concentrations of selenium (50 to 126 μg/L), and the other group was exposed to lower concentrations of selenium (1 to 16 μg/L). The drinking water standard and Maximum Contaminant Level (MCL) for Se in Colorado is 50μg/L (CDPHE, 2010b). Urine of persons using the high selenium water supply contained significantly more selenium than urine of persons using the low selenium water, but there were no significant differences between the two groups in the incidence or prevalence of any disease entity studied (Hammer, 1981).

Although selenium is classified formally as a nonmetal in the periodic table of elements, it is also referred to as a metalloid as it exhibits properties that fall between metals and nonmetals (Chapman et al., 2009). Selenium is located in the oxygen group 6A and has an atomic number of 34. Selenium has been shown to occur in four oxidation states and many forms (Table 1). The most common form in flowing rivers is dissolved selenate, although selenite, ogano-selenide and/or elemental selenium can also become the primary form depending on biological transformation and site characteristics.

Forms and concentrations of selenium in a soil solution are governed by various physical-chemical factors expressed in terms of pH, dissociation constants, solubility products, and oxidation-reduction potentials (Geering et al., 1968). Sediment redox potential and pH were found to be the key factors in the biochemistry of selenium in

Table 1 – Chemical forms of selenium in the environment (adopted from Fordyce, 2005)

Oxidative state	**Chemical forms**
Se^{2-}	Selenide (Se^{2}, HSe^{-}, H_2Se_{aq})
Se^{0}	Elemental selenium (Se_0)
Se^{4+}	Selenite ($SeO^{2-}{}_3$, $HSeO^{-}{}_3$, H_2SeO_{3aq})
Se^{6+}	Selenate ($SeO^{2-}{}_4$, $HSeO^{2-}{}_4$, H_2SeO_{4aq})
Organic Se	Selenomethionine, Selenocycteine

relation to its solubility and its distribution among the various chemical species in a laboratory study conducted using Kesterson Reservoir sediments in California (Masscheleyn et al., 1990). Four different redox levels (-200, 0, 200, and 450 mV) and four suspension pH levels (6.5, 7.0, 8.5, and 9) were selected and maintained during the incubation period. In general, selenium solubility was found to be low with low redox levels and high with high redox potentials. The pH affected both the levels and chemical forms of dissolved selenium, but the selenium solubility was lowest in the incubations at neutral pH. Total soluble selenium concentrations substantially increased upon oxidation or increase in sediment redox potential. Selenide comprised 80-100% of the total soluble selenium under reduced conditions (-200 mV). Oxidation of selenide to selenite was rapid, and above 200 mV selenite slowly oxidized to selenate. Under highly oxidized conditions (450 mV), selenate became the major species in solution constituting 95% at higher pH levels (8.5, 9) to 75% at lower pH levels (7.5, 6.5) (Masscheleyn et al., 1990). The species distribution with its various oxidation states and pH can be seen in the derived Eh-pH diagram (Figure 1). The Eh-pH diagram shows that in acidic and neutral soils, inorganic Se occurs as very insoluble Se^{4+} complexes of oxides and oxyhydroxides. In neutral and alkaline soils, Se^{6+} is the main oxidation state, being soluble and hence more readily available for uptake by plants (Dissanayake and Chandrajith, 1999). The area enclosed by solid lines shows the stability field and the normal range of surface conditions where redox levels are between -0.4 and 1.0V and pH levels between 4 and 8.5s.u.

Increasing the residence time of water in ponds and when recycling occurs in canals and/or backwater areas, as well as plant productivity and contact time between sediment

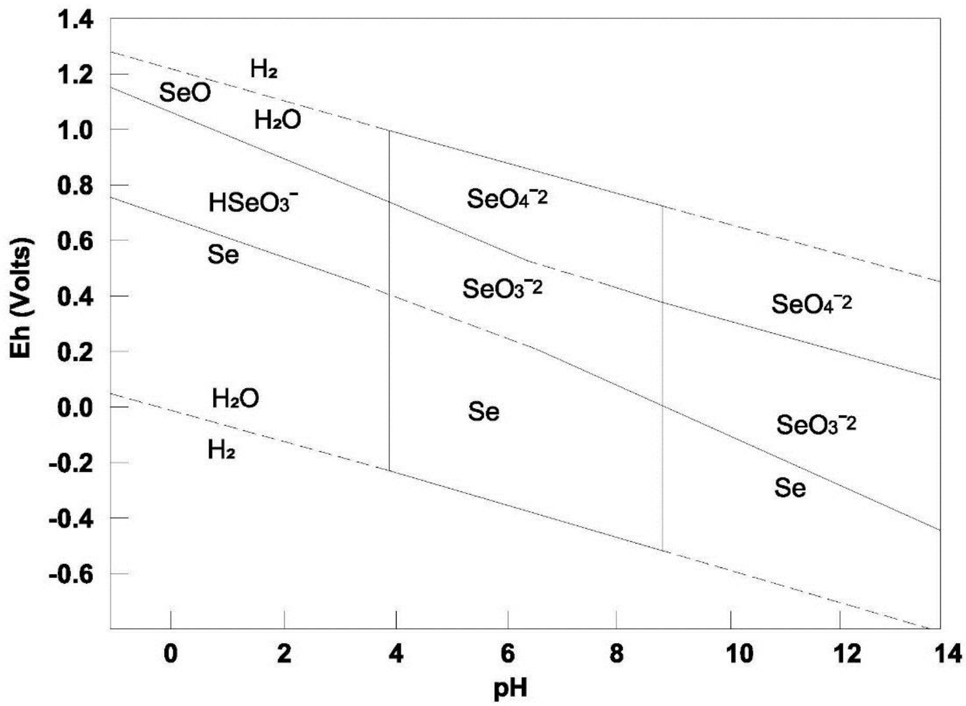


Figure 1 – Eh-pH diagram of Se in soils (from Dissanayake and Chandrajith, 1999)

and water all drive selenium transformations (Luoma and Rainbow, 2008). Plant uptake of selenate is slow and the release of the transformed products builds a pool of selenite and particulate organo-selenide in solution. Bioavailability grows with time because of the slow rate at which selenite and organo-selenide transform to selenate (Luoma and Rainbow, 2008). The proportion of selenium in the water column can be variable due to microbial activity, which promotes the reduction of the selenate form, as well as adsorption to sediment particulates followed by settling. Ninety percent of the total selenium in an aquatic system may be in the upper few centimeters of sediment (Lemly and Smith, 1987). Still, high flows and flood events in river systems disrupt the river bed and re-suspend selenium bearing sediments back into the water column where they can become bioavailable again through further reactions. With respect to persistent contaminants, aquatic systems are dynamic and selenium can be cycled back into the biota and remain at elevated levels for years after waterborne inputs of selenium are stopped (Lemly, 1997).

Across the western USA, 640,000 km^2 of land have the proper combination of climate, geology and soils to be Se-contaminated; 8,100 km^2 of these lands are irrigated (Seiler et al., 1999). Regions with an arid or semi-arid climate, where evaporation is greater than precipitation, and source rocks made of organic rich shales or marine sedimentary rocks are lands susceptible to selenium contamination. These arid and semi-arid regions of the country are also much more likely to redistribute selenium concentrations and loads into receiving waterways through irrigation for crop production and disturbances associated with land use changes. Irrigation leaches soluble forms of selenium and transports them through the soil (Ohlendorf et al., 1986). Urban development and mining operations can also greatly increase the mobility of selenium.

Grading or other soil mixing activities, especially in and around streams further exacerbates the problem by introducing seleniferous material to water, thereby dissolving and mobilizing selenium (Luoma and Rainbow, 2008).

The United States Environmental Protection Agency's (EPA) National Recommended Water Quality Criteria for chronic levels of selenium is 5μg/L (EPA, 1995), which is based upon the aquatic life table value criteria in EPA's 1987 Selenium Criteria Document (EPA, 1987). The EPA's criteria are not required law but recommendations for the guidance of states and tribes to adopt their own water quality standards for the protection of aquatic life and human health. Concentrations in undisturbed waters are as low as 0.07μg/L Se yet concentrations above 0.2μg/L Se can suggest contamination, whether enhanced by anthropogenic activities or natural processes (Luoma and Rainbow, 2008).

Water quality standards for selenium as selenate (Se^{6+}) in the State of Colorado is the Table Value Standard (TVS), acute = 18.4μg/L and chronic = 4.6μg/L (CDPHE, 2010a). Attainment of chronic metal standards in the state of Colorado, specifically dissolved metals in both streams and rivers, is based upon the 85th percentile of the ranked data (CHPHE, 2009). Chronic means the level not to be exceeded by the concentration as an average of all samples collected during a thirty day period to protect genera from chronic toxic effects, whereas acute means the level not to be exceeded for either a single sample or calculated as an average during a one-day period (CDPHE, 2012). Acute and chronic values adopted as stream standards for Se; however, are levels not to be exceeded more than once every three years on average (CDPHE, 2012).

TVS numeric values serve as interim guidance for the Water Quality Control Commission (WQCC) in establishing numeric standards for specific basins and individual stream segments. Standards are segment specific and may be different due to the bioaccumulation nature of selenium, and the adsorption to particulate matter.

1.2 Problem Definition

The role and importance of selenium as an environmental contaminant has gained widespread attention among research scientists, natural resource managers and federal and state regulatory agencies during the last two decades (Lemly, 1993; Chapman et al., 2009). Section 303(d) of the Federal Clean Water Act (CWA) requires states to report to the U.S. Environmental Protection Agency (USEPA) a list of water bodies that are water-quality impaired. Recently, selenium has been listed on Colorado's Section 303(d) List of Impaired Waters for several segments of the lower Cache la Poudre watershed around the city of Fort Collins (CDPHE, 2010c). Stream segments of interest are segments where the 85^{th} percentile of the ranked data for Se was above the (chronic) table value standard of 4.6μg/L (Table 2).

The Water Quality Control Division (WQCD) conducts water quality assessments across the state triannually, which are reported in the 305(b) Report fulfilling the obligation to the CWA. The 2010 Update to the 2008 305(b) Report was released as an integration report that includes the 2008-2009 water quality assessments, as well as provide the State's revised assessments that were conducted over the past five years. Roughly 94,455 miles of the state's 105,344 river miles were assessed and 10,673 miles of Colorado streams and rivers were found to be impaired and require a Total Maximum Daily

Table 2 – Colorado's Section 303(d) List of Impaired Waters on the Cache la Poudre River (CDPHE, 2010c).

Colorado's Section 303(d) List of Impaired Waters - April 2010		
Waterbody Identification WBID	Segment Description	Clean Water Act Section 303(d) Impairment
COSPCP11	Mainstem of the Cache la Poudre River from Shields Street in Ft. Collins to a point immediately above the confluence with Boxelder Creek	Selenium
COSPCP12	Cache la Poudre River, Box Elder Creek to South Platte River	Selenium
COSPCP13a	All tributaries to the Cache la Poudre River, including all wetlands, from the Monroe Gravity canal to the confluence with the South Platte River	Selenium

Load analysis (TMDL) (CDPHE, 2010d). Selenium was found to impair 7,478.41 miles of Colorado's rivers, whereas the second leading cause of impairment for Colorado rivers is *Escherichia coli*, affecting 1,666.02 river miles (CDPHE, 2010d). It is the CDPHE's objective to prioritize the impaired stream segments of the Cache la Poudre River for development of a TMDL. Waterbody identification COSPCP13a, which includes the Cache la Poudre River and all tributaries through the Fort Collins area, was assigned a low priority for TMDL development followed by restoration efforts (CDPHE, 2010c). Developing a TMDL involves the quantification of loads for the pollutant of interest, helping to identify point and non-point sources and a contaminant reduction strategy. This process takes into account all available data on the characterization and sources of selenium, including current hydrologic and water quality conditions. Fossil Creek, a tributary to the Cache la Poudre River in Fort Collins, Colorado, was first listed on the CWA Section 303(d) List for selenium in 2006 (EPAb, 2012), followed by other Cache la Poudre River segments in 2010 (Table 2). A TMDL analysis has not been completed for any of the Cache la Poudre River segments listed for selenium (EPAb, 2012).

The WQCD prioritizes impaired stream segments for TMDL development to identify where the Division and public should focus their resources. Priorities are initially based on consideration of the severity of impairment to the classifications for the segments, and secondary factors include: endangered or declining native species, public interest, and administrative needs (CDPHE, 2009). Adequate data, as well as quality statistical analysis, is not always available and stream segments considered high priority may need further review. Waters designated "low" and "medium priority" may be targeted for further data collection and analysis as well, before they may be amenable to TMDL development.

Statistics are important in the analysis of ambient water quality data for the characterization of the waterbody in question and statistical models can further facilitate solutions to remediate the pollutant(s).

1.3 Statistical Models

Statistical models have been used extensively for the prediction, estimation and assessment of pollutant loads in surface water. A water quality model has been compared for accuracy against two different rating curve methods in five watersheds in Maryland for estimating nutrient loads, which are used to determine current loads for a waterbody ensuring that TMDLs are met (Dorianne and Moglen, 2008). The load-derived method and the concentration-derived method were applied forming a regression relationship and differences were examined between model fitness using Nash-Sutcliffe (NS) coefficients. Coefficients between 0.0 and 1.0 are considered acceptable, and coefficients <0 are considered unacceptable (Nash and Sutcliffe, 1970).

The load-derived method was characterized by developing a rating curve between the pollutant load and the corresponding discharge. The concentration-derived method was characterized by developing a rating curve between the measured pollutant concentration and the corresponding discharge. According to the NS coefficients, the concentration-derived method performed slightly better. Users of rating curves to develop pollutant loads; however, must recognize that the load-derived method is fundamentally flawed because it is based on a relationship between two dependent variables: load and discharge (Dorainne and Moglen, 2008). It was determined that the correct rating curve approach in estimating nutrient load, in the absence of continuous observations, should be based on the independent variables of pollutant concentration and discharge.

There is a need for selenium characterization using geospatial methods that determine sources of selenium contributing to surface waters and a water quality model designed to predict concentrations. While most water quality models constitute contaminant loads, concentration values should be investigated and just as important is the ability to predict the concentrations for unmonitored locations. Due to the lack of resources and funding, many water quality monitoring and assessment agencies are unable to predict or verify the water quality in these areas, making management decisions difficult. The integration and use of a GIS for the prediction of selenium in surface water can greatly benefit the understanding of selenium dynamics and regulatory agencies.

1.4 GIS Models

The importance of water quality deterioration due to contaminant inputs, and the need for effective targeting of mitigation approaches and river catchment management, has led to the development of a variety of water quality models (Rothwell, 2010). The use of a GIS as a tool for water quality and natural resource management has become of increasing interest, as the software development becomes more sophisticated and applicable to decision making without allocating unnecessary time and money in the field. GIS has been used to predict surface water quality problems associated with nonpoint-source agricultural pollution (Gilliland and Baxter-Potter, 1987, Tim and Jolly, 1994, Tong and Chen, 2002), heavy metals and mine waste pollution (Kern and Stednick, 1993, Xiao and Ji, 2007), nutrient contamination (Rothwell et. al., 2010), and groundwater contamination via hard-rock mining pollution (Rodda et al., 1999). Landscape characteristics are the basis and often considered the most important factors in spatial analysis used to determine the relationships linking landscape variables with water quality and the contaminant(s) in question; whether it is the source, causal mechanism, or the variability associated with diurnal, temporal, and spatial fluctuations.

A case study in the Tri-State Mining District (Missouri, Kansas, and Oklahoma), used multi-temporal Landsat® imagery to characterize land use and land cover and quantify the relationship between landscape metrics and surface water quality (Xiao and Ji, 2007). Landscape characteristics were significantly correlated to stream water quality in mine waste-located watersheds, and affected the water quality as much as in agricultural or urban watersheds. Regression analysis results showed landscape metrics could account for as much as 77% of the variance of water quality variables, suggesting that landscape metrics were useful in predicting water quality. The proportion of mine waste area accounted for <60% of the variance of heavy metal concentrations in stream water, suggesting that it is possible to predict water quality with only a single landscape metric; although the predicting power of single metric models would be limited. Catchment characteristics along the riparian zone have also been used to predict chemical parameters in a GIS (Smart et al., 2001). In particular, the parent material and geochemistry of the riparian zone, when combined with a simple hydrological flow path model, could be used to accurately predict stream water chemistry.

Landscape and/or catchment characteristics have also been used to develop Hydrological Response Units (HRUs) in a GIS (Kouwen et al., 1993, Ocampo et al., 2006, Vigiak et al., 2006). HRUs are individual units or subbasin groupings captured by homogenous elements in relation to hydrologic responses. Testing the practicability of defining hydrologic response units as combinations of soil, land use and topography for modeling infiltration at the hillslope and catchment scales has been attempted (Vigiak et al., 2006). Due to spatial heterogeneities of rainfall variability, canopy interception, soil sealing and infiltration, identifying hydrologic response units was difficult. With the addition of chemical dynamics in landscape elements when identifying HRUs, Chemical Hydrologic Response Units (CHRUs) may be developed. A GIS was coupled with hydrochemical software and synoptic water sampling and analysis to develop a new technology for the identification of water quality problems, one that identifies potential sources and addresses these sources before a substantive impact occurs (Kern and Stednick, 1993). The method developed, Chemical-Hydrologic Resource Information System, enabled the user to visualize metal concentrations in space, identify stream segments probably below the water quality standard, and those predicted to exceed the standard. CHRUs and GIS analysis has also been used to delineate spatial distribution in the Broel River Basin in Germany to compare model calculated chemical balances with measured output at gages (Bende et al., 1995).

The assistance of spatial analysis to identify sources and/or predict concentrations of selenium in a GIS has been attempted before. A regression model used a GIS to identify and assess potential sources of selenium in the Kendrick Reclamation Project Area of Wyoming (See et al., 1992). This area has a long history of selenium contamination showing effects on invertebrates, fish and terrestrial animals. The model was developed using median selenium discharges from each subbasin as a dependent variable regressed on measured physical and chemical characteristics of the hydrologic subbasins as the independent variables. Results of this study indicated two things; one that soils with typical total selenium concentrations can supply large selenium discharges to tributary streams if irrigation intensity as measured by area of irrigated land and length to irrigation canals, is large. The second is that the use of GIS-generated information for a regression model provided a method that generally defines areas that might constitute sources of selenium in streams. In terms of selenium concentration prediction, a geostatistical case study was developed through the use of ordinary and ordinary lognormal kriging, and through Bayesian estimation for predicting the selenium concentration in soils using data that are subject to measurement error (Orton et al., 2009).

Statistics within the framework of a GIS or geostatistics are very useful and important in the analysis of numerical data in that it provides a graphical representation of the spatial data distribution. Spatial statistics are a set of exploratory techniques used to describe and model spatial distributions, patterns, processes and relationships (ESRI, 2011). Spatial statistics use the following in their mathematics: area, length, proximity, orientation, and spatial relationships.

Traditional statistics is similarly the study of the collection, organization, analysis, and interpretation of data. Statistical models describe how one or more variables are related to one or more variables. Statistical tests can be applied to statistical models to help determine if a hypothesis is true or quantify evidence against a hypothesis being true.

Despite the State's efforts in water quality monitoring for Se in the surface waters of the Cache la Poudre River basin, no other data regarding potential sources, mobility or TMDL development could be found. A better site characterization is needed, as well as a better understanding of the spatial extent of contamination in the Cache la Poudre River basin. A predictive model based on a landscape and catchment investigation for this region will provide data of source inputs (natural and anthropogenic), and landscape dynamics (e.g. slope, soils, geology, and land use) that contribute to the accumulation and transport mechanisms of Se. Identifying point source inputs of selenium, such as wastewater treatment plants and mining operations, as well as non-point sources, such as selenium bearing rocks and soils can additionally be used in modeling techniques to observe the distribution of selenium concentrations in streams.

The purpose of this study is to develop a modeling approach to predict Se concentrations using specific landscape properties and a statistical evaluation of a landscape assessment necessary for the characterization of the lower Cache la Poudre River watersheds. The modeling approach could be applied in small to large watersheds, such as the 22.19 km^2 watershed of Spring Creek in Fort Collins, CO to the 4,895 km^2 watershed of the Cache la Poudre River in north-central Colorado (USGS, 2012a).

1.5 Hypothesis and Objectives

Although chemical transport modeling based on similar landscape properties and spatial analysis in a GIS for selenium and watershed management has been applied before, the need to understand dynamics within the Fort Collins regional environment and to be able to predict selenium concentrations in Spring Creek and Fossil Creek is important for management practices and potential TMDL development. The study hypothesis is that a statistical model using spatial landscape elements can predict selenium concentrations in these surface waters.

Study objectives:

1. Integrate a Geographic Information System (GIS) with the following landscape elements: elevation, land use land cover, soil pH, geology, irrigation, and hydrography data for application in selenium prediction and watershed assessment
2. Develop a traditional statistical model using geospatial landscape data, streamflow, and precipitation to predict selenium concentrations in Spring Creek and Fossil Creek
3. Develop a GIS spatially weighted statistical model using geospatial landscape data, streamflow, and precipitation to predict selenium concentrations in Spring Creek and Fossil Creek

CHAPTER 2 - METHODS

2.1 Site Description

Cache la Poudre River

The Cache la Poudre River Basin is the largest tributary drainage to the South Platte River Basin. The headwaters begin in Rocky Mountain National Park in north-central Colorado as a high-gradient stream descending quickly in an easterly direction along Highway 14 in the "Poudre Canyon." The river exits the mountainous region at the mouth of the canyon near the town of Bellvue in a more gentle landscape of rolling to relatively flat high plains. Flowing southeast through the cities of Fort Collins, Timnath and Windsor, the Cache la Poudre River ends at the confluence with the South Platte River east of the city of Greeley, Colorado (Figure 2).

Study Area

The study area is located near Fort Collins, Colorado. The city of Fort Collins has a population of 143,986 (U.S. Census, 2010) with a semi-arid climate. Elevation is roughly 1,500 m above sea level, and the average yearly precipitation is 355 to 383 mm (WRCC, 2012). Areas of interest specifically are Spring Creek and Fossil Creek from Horsetooth Reservoir in the foothills at the western edge of Fort Collins to their confluences with the Cache la Poudre River southeast of Fort Collins and south of Timnath (Figure 2).

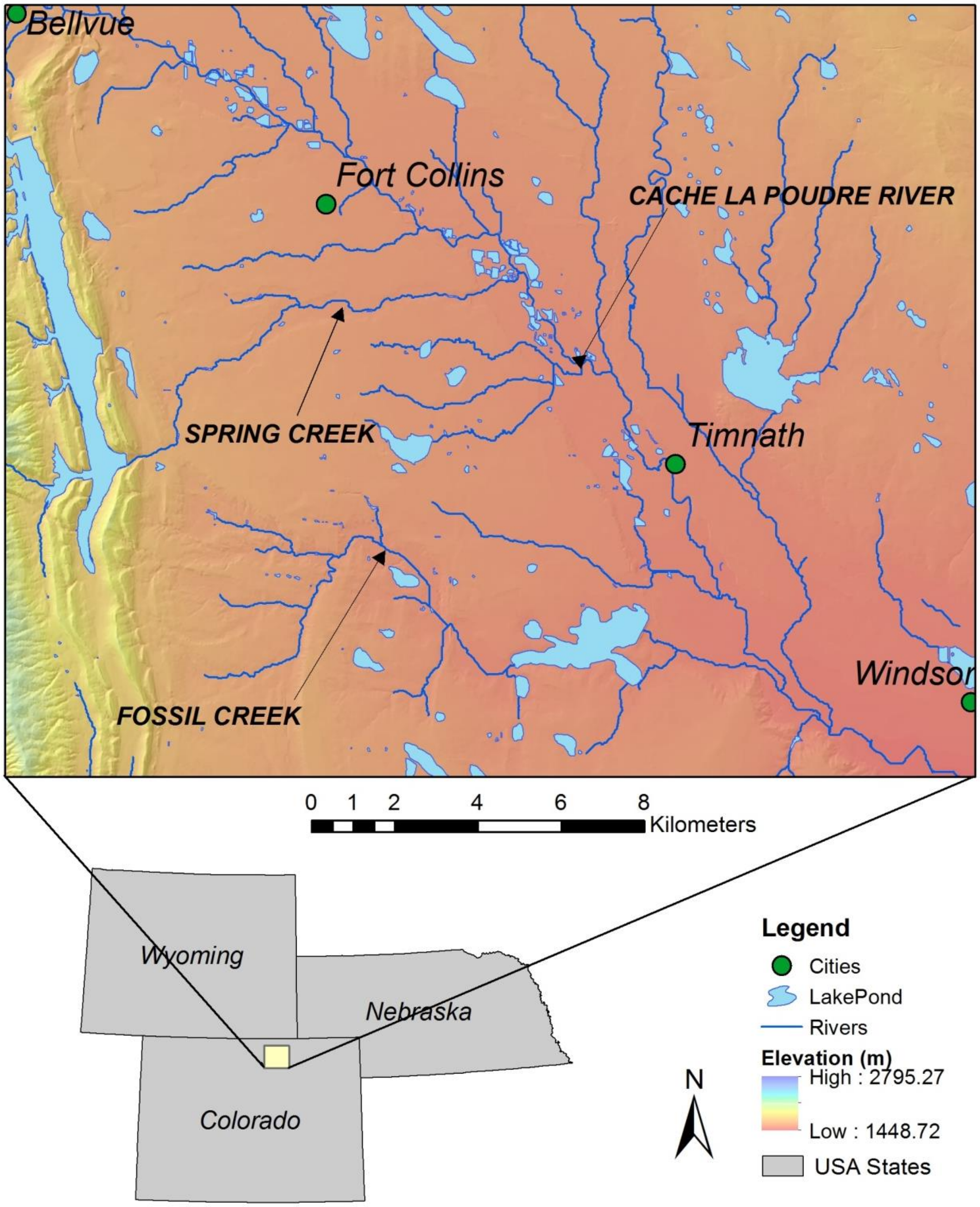


Figure 2 – Location map of the Cache la Poudre River and study area of Spring Creek and Fossil Creek near Fort Collins, Colorado (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset , National Atlas)

2.2 Water Quality and Streamflow

Water Quality Data

Water quality and streamflow data were organized into a database spreadsheet within Microsoft® Excel®. The water quality sampling efforts conducted by the WQCD, which were used to list sections of the lower Cache la Poudre River watershed on the CWA 303(d) Impaired Waters List, include the following: selenium (μg/L), hardness as $CaCO_3$ (mg/L), pH (s.u.), temperature (°C), dissolved oxygen (mg/L), location coordinate values for sampling stations, and the time and date of sampling measurements. This particular dataset was provided by the WQCD and included sampling dates between January 2003 and December 2008, and between March 2006 and December 2008 for Spring Creek and Fossil Creek, respectively (Hillegas, 2011). This represents all the available data. Within the study area, there are 3 sampling stations on Spring Creek and 2 sampling stations on Fossil Creek (Table 3; Figure 3).

Streamflow Data

Streamflow data were gathered by the City of Fort Collins Utilities Department and provided by the Department of Civil and Environmental Engineering at Colorado State University (Olson, Chris; personal communication, June 2012). The location coordinate values, river stage (ft), river discharge (cfs), and the date of measurements were provided. Within the study area, there are 5 stream gauging stations on Spring Creek and 3 on Fossil Creek (Table 4; Figure 3). Streamflow measurements are from January 2003 and July 2006, and between September 2006 and December 2008 for Spring Creek and Fossil Creek, respectively. Much of the streamflow data are from February or March through October

Table 3 – Water quality sampling stations and locations on Spring Creek and Fossil Creek

Colorado Department of Public Health and Environment Water Quality Sampling Stations			
Stream/River	**Location**	**Latitude**	**Longitude**
Spring Creek	Shields Street	40.56242	-105.0956
Spring Creek	College Avenue	40.562478	-105.077309
Spring Creek	Edora Park	40.56467	-105.044916
Fossil Creek	College Avenue	40.51341	-105.07644
Fossil Creek	Trilby Road	40.494716	-105.051287

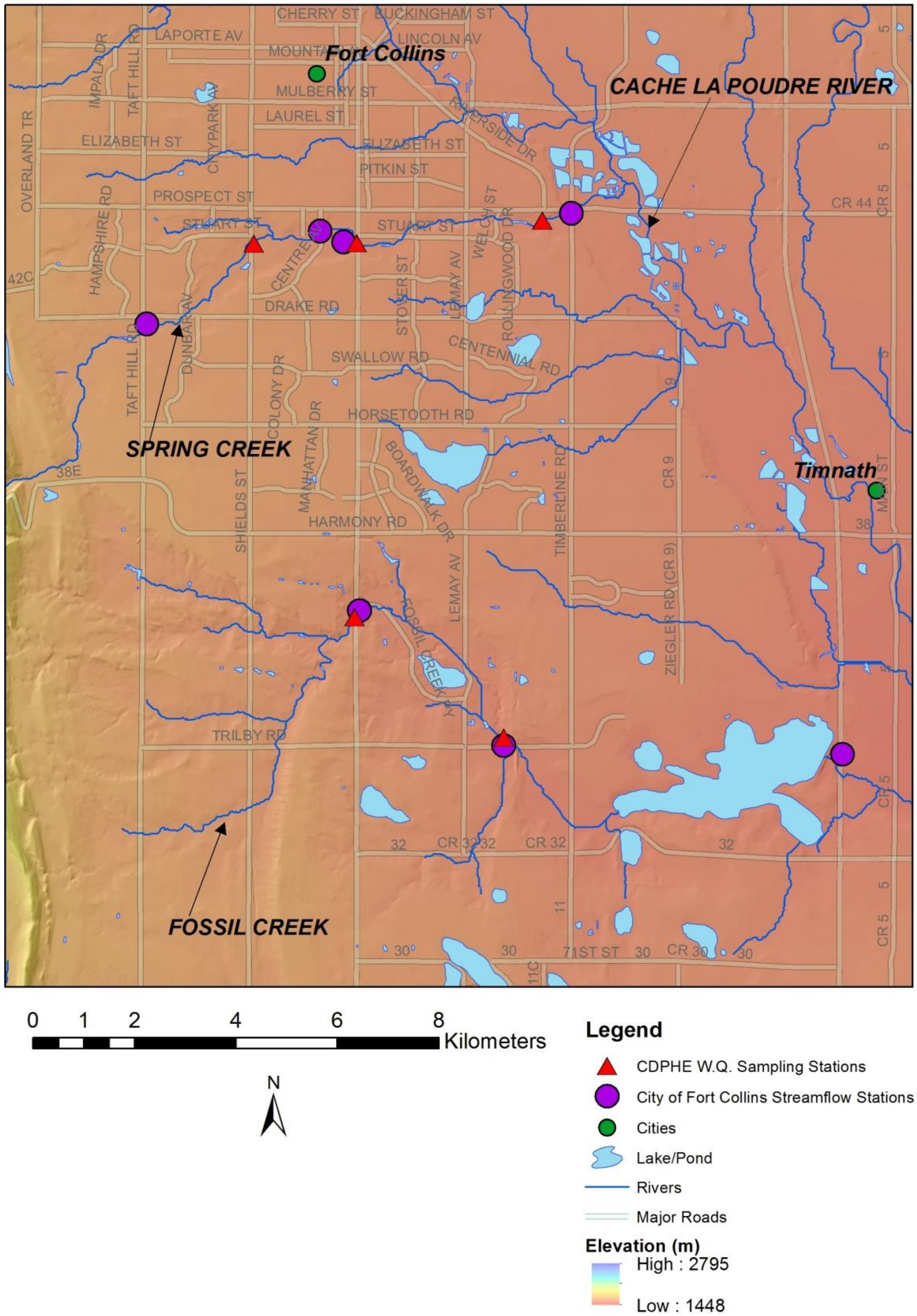


Figure 3 – Map of Colorado Department of Public Health and Environment water quality sampling stations and City of Fort Collins streamflow gauging stations on Spring Creek and Fossil Creek (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset)

Table 4 – Streamflow gauge stations on Spring Creek and Fossil Creek

City of Fort Collins Streamflow Gauge Stations				
Gauge ID	**Stream/River**	**Location**	**Latitude**	**Longitude**
6083	Spring Creek	Taft Hill Road	40.551632	-105.11448
6203	Spring Creek	Center Avenue	40.564097	-105.08378
6053	Spring Creek	Burlington Northern Rail Road (BNRR)	40.562655	-105.079607
6163	Spring Creek	Timberline Road	40.566550	-105.03930
6093	Fossil Creek	College Avenue	40.512928	-105.076858
6293	Fossil Creek	Fairway Estes Dam	40.49351	-104.991174
6283	Fossil Creek	Trilby Road	40.494716	-105.051287

or November. Streamflow data are incomplete. Missing streamflow values were estimated using the drainage-area ratio method (Emerson et. al., 2005).

Precipitation Data

Precipitation data were acquired from the National Oceanic and Atmospheric Administration's National Climatic Data Center (NCDC, 2013). Data from the NCDC includes daily precipitation measurements for Fort Collins. The dates selected for this data correspond to the water quality sampling efforts for selenium, which were between January 2003 and July 2006, and between September 2006 and December 2008 for Spring Creek and Fossil Creek, respectively.

2.3 Geographic Information System

Data Sources

A geospatial model was built using ArcGIS® Version 10.1 , ESRI®, within a Microsoft® Corporation Windows® 7 Home Premium based Hewlett-Packard® Company desktop computer. Landscape element datasets used in the basin assessment for the Se prediction model were downloaded from a variety of sources (Table 5).   Soils data were accessed from the United States Department of Agriculture (USDA) - National Resources Conservation Service's (NRCS) Soil Survey Geographic Database (SSURGO). Soils pH layer can then be retrieved, which is found in the soil chemical properties within the attribute folders of the soils database. Geology and Land Use Land Cover datasets were downloaded from the USDA – NRCS Geospatial Data Gateway. National Elevation Dataset (NED) 1/3 arc second (approximately 10m), also referred to as a Digital Elevation Model (DEM), was

Table 5 – Landscape elements used in GIS model with data source and website.

Geographic Information Datasets		
Landscape Elements	**Data Source**	**Website**
Soils	USDA-NRCS SSURGO (Soil Survey Geographic Database)	http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm
Geology	USDA-NRCS Geospatial Data Gateway	http://datagateway.nrcs.usda.gov/
Land Use Land Cover	USDA-NRCS Geospatial Data Gateway	http://datagateway.nrcs.usda.gov/
Elevation	USGS Seamless Data Warehouse	http://nationalmap.gov/viewer.html
Hydrography	USGS National Hydrography Dataset	http://nhd.usgs.gov/data.html
Irrigation	CWCB - DWR - CDSS (Colorado Decision Support Systems)	http://cdss.state.co.us/GIS/Pages/Division1SouthPlatte.aspx

accessed from the United States Geological Survey's (USGS) Seamless Data Warehouse. Hydrography data came from the USGS National Hydrography Dataset (NHD) database. The Colorado Water Conservation Board (CWCB) and Colorado Water Resource's water management system, Colorado's Decision Support System (CDDS), was used to attain the most recent irrigated land coverage. Irrigation geospatial data consists of polygons representing land areas that are consistently irrigated or watered.

Watershed Delineation

Geospatial landscape assessment involves the delineation of Spring Creek and Fossil Creeks' watersheds. Spring Creek is located centrally in Fort Collins, whereas Fossil Creek is located on the southern outskirts of Fort Collins and has additional natural spaces and agricultural land. Original source DEM raster layer was processed using the Fill, Flow Direction, and Flow Accumulation Tools within Spatial Analyst, an extension of ArcMap, creating several raster hydrology layers. Two Pour Point vector (point) shapefiles were created, one for each watershed, as well as the specification of the pour points themselves. The final step in creating the watersheds was to run the Watershed Tool with the Flow Direction and Pour Point layers, thereby delineating the watersheds (Figure 4). Raster based watersheds were then converted into vector based polygon features, allowing the imported vector landscape element layers to be analyzed with spatial statistics.

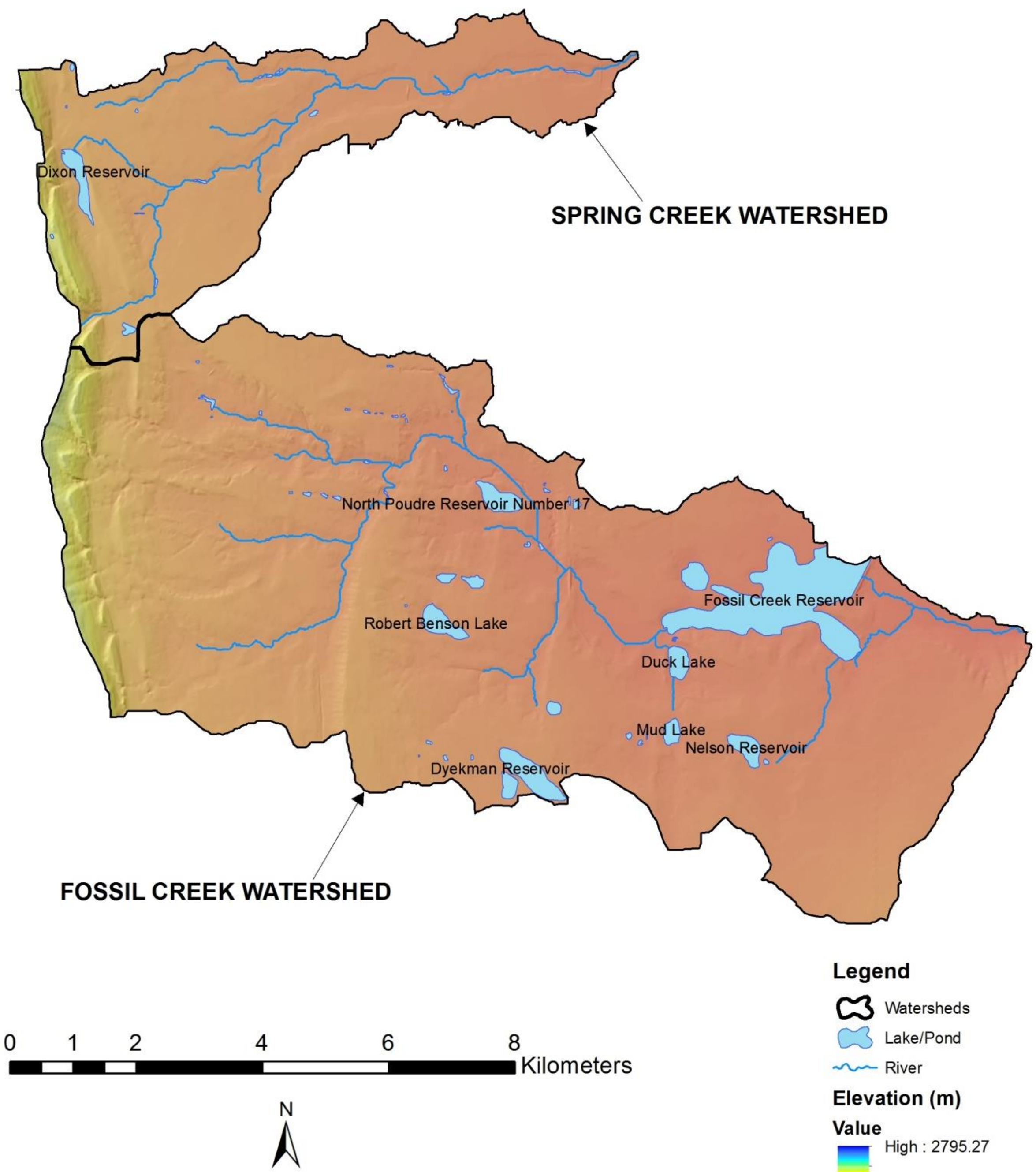


Figure 4 – Map of Spring Creek and Fossil Creek delineated watershed boundaries (Data sources: USDA – NRCS Geospatial Data Gateway, USGS National Hydrography Dataset)

Sub-watershed delineation

Delineation of sub-basin watersheds is needed for an approach and future development of a selenium prediction model using CHRUs. Note that the term sub-basin watershed refers to the smaller delineated watershed boundaries of CHRUs. Not only can these sub-basin watersheds be linked chemically and hydrologically for flow path modeling, but the statistical assessment of existing landscape elements identified spatially by the use of sub-basin watersheds is useful for characterizing the study area, as well as for determining the most influential landscape element(s) on surface water selenium concentrations. In order to delineate sub-basin watersheds within the Spring Creek and Fossil Creek drainage basins, a series of terrain preprocessing tools were run to generate stream layers. There are several stream tools in the Arc Map extension Arc Hydro Toolbox under the Terrain Preprocessing tool set, which aid in the development of watershed delineations, such as the Stream Definition, Stream Segmentation, and Catchment Grid Delineation tools. Catchment Polygon Processing tool, the last step in watershed delineation, was run creating a polygon feature class enabling the application of spatial statistics and data extraction techniques.

There were 13 sub-basin watersheds and 21 sub-basin watersheds delineated in Spring Creek and Fossil Creek, respectively (Figure 5). Sub-watershed delineations also included the creation of separate watershed basin layers that captured all the land area upstream of each CDPHE water quality sampling stations (Figure 6). Note that the term sub-watershed refers to the medium sized delineated watershed boundaries upstream of the CDPHE stations. Through this process, three sub-watershed layers were created within the Spring Creek study area and two sub-watershed layers were created within in the

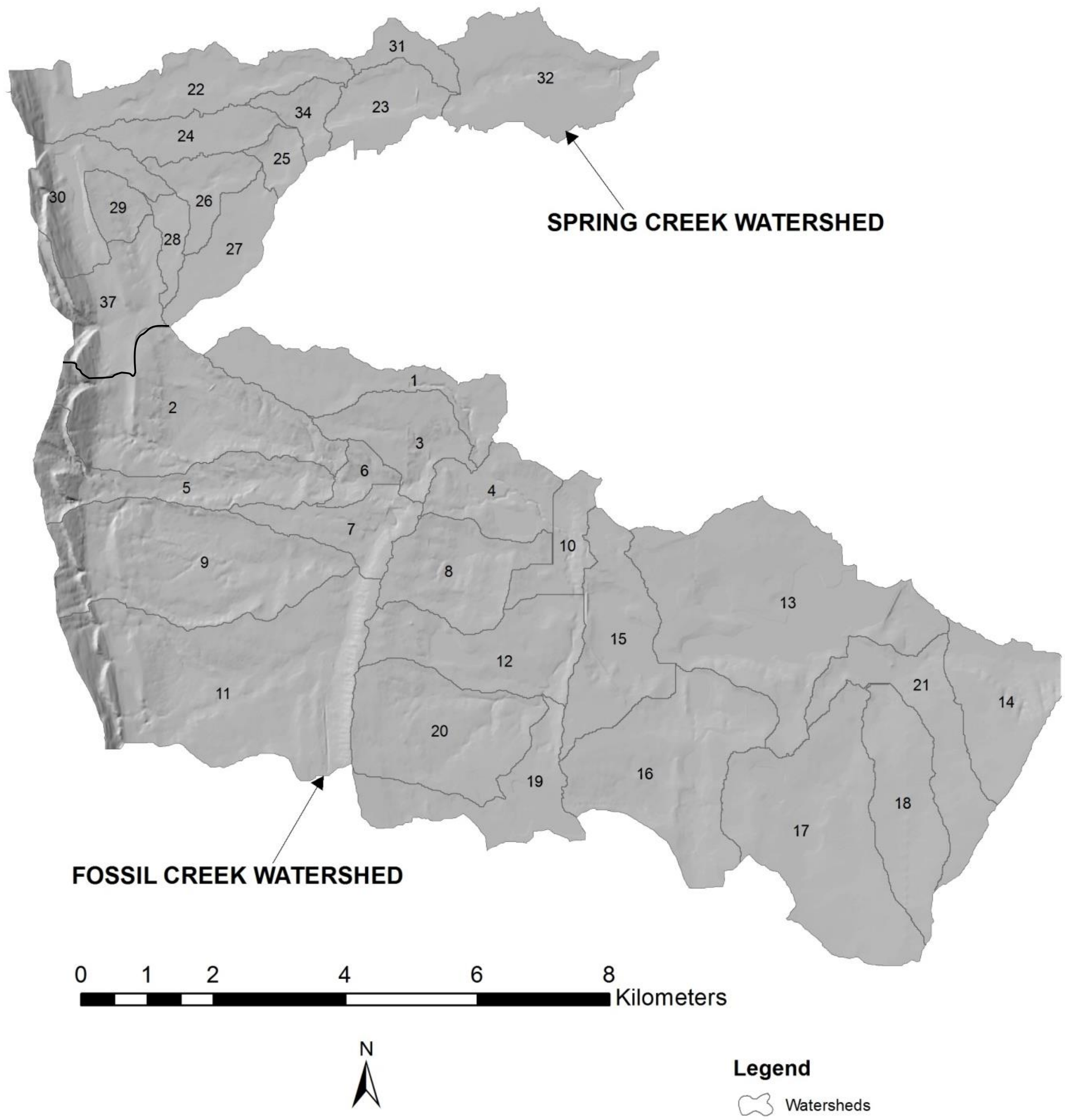


Figure 5 - Map of sub-basin watershed boundaries within Spring Creek and Fossil Creek watershed boundaries. Numbers denote sub-basin watershed labels. (Data source: USDA – NRCS Geospatial Data Gateway)

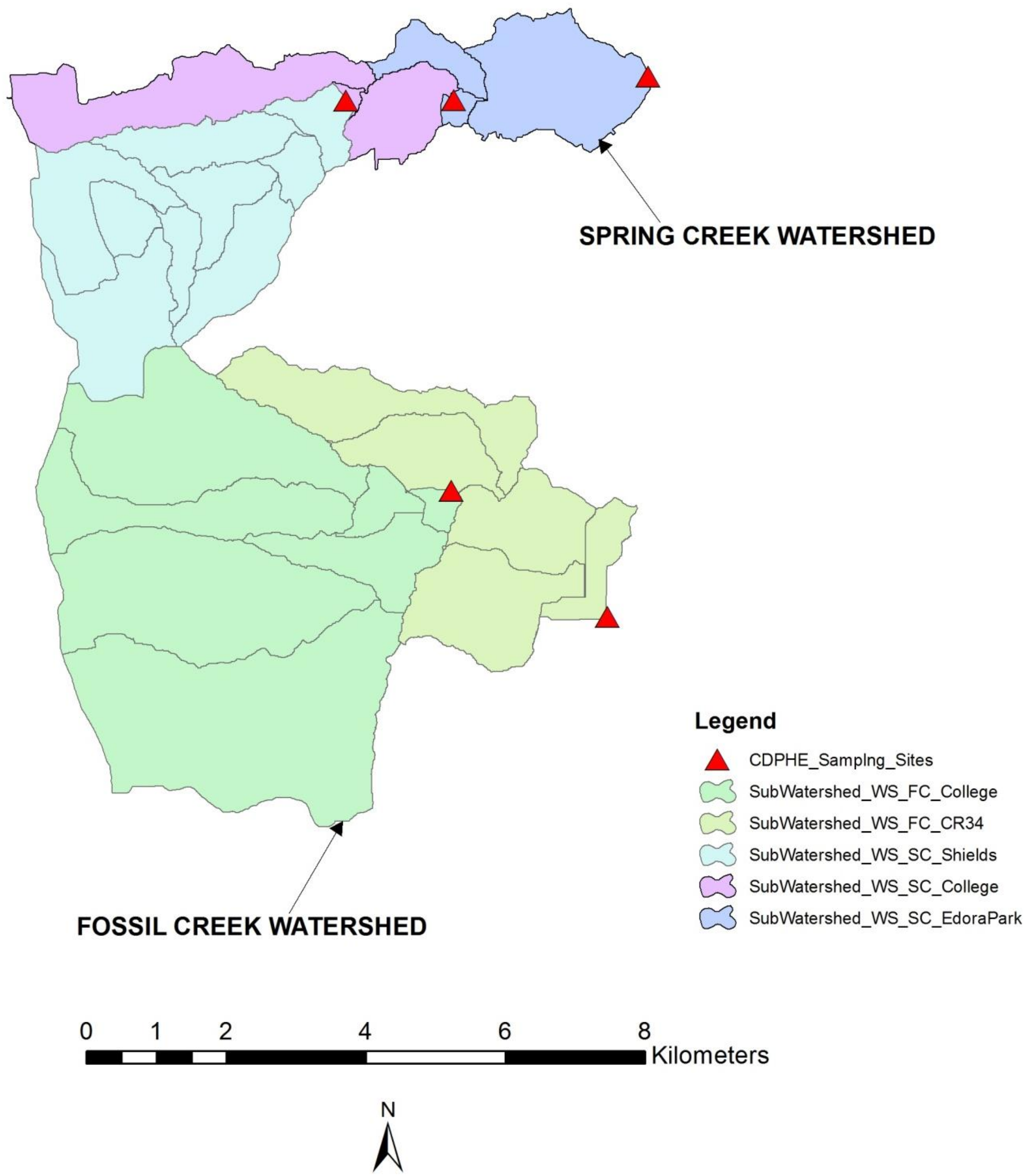


Figure 6 – Map of sub-watershed layers delineated upstream of each CDPHE water quality sampling station for Spring Creek and Fossil Creek with CDPHE water quality sampling stations (Data source: USDA – NRCS Geospatial Data Gateway)

Fossil Creek study area with respect to the number of CDPHE water quality sampling stations (Table 3). Although Figure 6 shows the sub-watershed layers overlayed on top of each other, each colored layer (e.g. sub-watershed layer) includes all areas upstream of its' respective water quality sampling station. It is important to delineate the spatial area upstream of each water quality sampling station for statistical purposes with respect to sampled selenium concentrations.

A. National Hydrography Dataset:

Originally, the NHD data set consisted of many types of flow lines. Because NHD data were clipped to the watersheds of Spring Creek and Fossil Creek in north-central Colorado, only the data that exists in this geographical location was actually used and needed. For example, there was a flow line type of Coastline, which is not applicable here. The NHD data within Spring Creek and Fossil Creeks watersheds consists of five flow line types with their associated class cell values and class name. Because these 5 flow line types are not all necessary for modeling or assessment purposes, they were reclassified into the two flow line types of Stream/river and Canal/ditch (Table 6.) The reclassified flow line types are also meant to replicate landscape elements used to assess potential selenium sources in a similar study in the Kendrick Reclamation Project Area in Wyoming (See et al., 1992).

B. Soil Survey Geographic Database:

The surface soil pH dataset was retrieved through the Soil Survey Geographic Database (SSURGO). Using the Soil Data Viewer 6.0 interface, an add-in to ArcMap, the soil

Table 6 – Original and reclassified NHD flow line types for Spring Creek and Fossil Creek watersheds from USGS, 2010

Flow Line Types and Reclassification		
	Cell Value	***Class Name***
Original Flow Line Types	334	Connector
	336	Canal/ditch
	428	Pipeline
	460	Stream/river
	558	Artificial Path
Reclassified Flow Line Types	460	Stream/river
	336	Canal/ditch

database can be accessed for many different soil properties and characteristics. A map form of the data can be generated, as well as "User Interface Descriptions" and an "Aggregation Report" (Appendix A and B). The soil pH dataset had 7 pH values with a range of 6.7 to 8.5 s.u. Since 7 soil pH values would make statistical evaluations of landscape element parameters complex, the dataset was reclassified using the category class format in which 3 soil pH category classes were developed (Table 7). Because sediment pH has been found to be a key factor in the biochemistry of selenium in relation to its solubility, the separation of the higher pH values into 2 alkaline classes (slightly alkaline and moderate to strong alkaline), creating 3 classes instead of 2, was done to better differentiate the effect soil pH may have on surface water selenium concentrations (Masscheleyn et al., 1990). The soil pH classes were based on USDA's NRCS soil survey classification (USDA, 1993) (Appendix C).

C. Land Use Land Cover Dataset:

There were 15 land use land cover types represented in the existing dataset in the Spring Creek and Fossil Creek watersheds. Because 15 classes would make statistical evaluations of landscape element parameters complex, the dataset was reclassified into 5 classes (Table 8). Land use land cover thematic classes are categorized by the National Land Cover Database (NLCD), which serves as the definitive Landsat – based, 30-meter resolution, land cover database for the nation; whose products are created by the Multi-Resolution Land Characteristics Consortium (USGS, 2012c; MRLC, 2012). The land use land cover classification key was attached with the original downloaded data

Table 7 –SSURGO soil pH values found in the Spring Creek and Fossil Creek watersheds with modified class categories from USDA, 1993

Soil pH Values and Category Classes	
pH Value (s.u.)	**Class**
6.7 - 7.1	*Neutral*
7.2 - 8.1	*Slightly Alkaline*
8.2 - 8.5	*Moderate to Strong Alkaline*

Table 8 – Land Use Land Cover original and reclassified classes in the Spring Creek and Fossil Creek watersheds from MRLC, 2012

Land Use Land Cover Original and Reclassified Classes		
	Cell Value	***Class Name***
Original Land Use Land Cover Classes	11	Open Water
	21	Developed, Open Space
	22	Developed, Low Intensity
	23	Developed, Medium Intensity
	24	Developed, High Intensity
	31	Barren Land (Rock, Sand, Clay)
	41	Deciduous Forest
	42	Evergreen Forest
	43	Mixed Forest
	52	Shrub/Scrub
	71	Grassland/Herbaceous
	81	Pasture/Hay
	82	Cultivated Crops
	90	Woody Wetlands
	95	Emergent Herbaceous Wetlands
Reclassified Land Use Land Cover Classes	11	Open Water
	21	Developed
	41	Vegetated
	82	Agriculture/Barren
	90	Wetlands

(Appendix D). Out of the 15 classes, there were only 5 completely distinct class types (e.g. developed, vegetated, water). Basically, any land use land cover class that exhibits similar characteristics was grouped together into one class. For example, the original land use land cover classes showed 4 different classes of "Developed," which were separated based on severity. Open space, low intensity, medium intensity, and high intensity "Developed" classes were reclassified into one singular "Developed" class. All vegetation type classes including shrub/scrub and grassland/herbaceous classes were grouped together. Similarly, all agriculture or non-vegetated classes including barren land, pasture/hay, and cultivated crops were grouped together into the agriculture/barren class. It is important to group the agricultural and barren type lands together as these have been shown in arid to semi-arid lands, such as the Fort Collins study area, to either be selenium contaminated or contribute to (Seiler et al., 1999).

ModelBuilder and Geospatial Variables

Geospatial statistics were used to characterize the landscape within the study area of Spring Creek and Fossil Creeks' watersheds. Using ArcMap's ModelBuilder, a series of analysis tools were completed in one step, which included several overlay and statistical processing tools (Figure 7). ModelBuilder is an application used to create, edit, and manage models, similarly known as workflows that string together sequences of geoprocessing tools, feeding the output of one tool into another tool as input (ESRI, 2012). The landscape parameter layers, which are the previously reclassified landscape element layers, were added to ModelBuilder along with spatial statistical processing tools resulting in geospatial variables (Table 9).

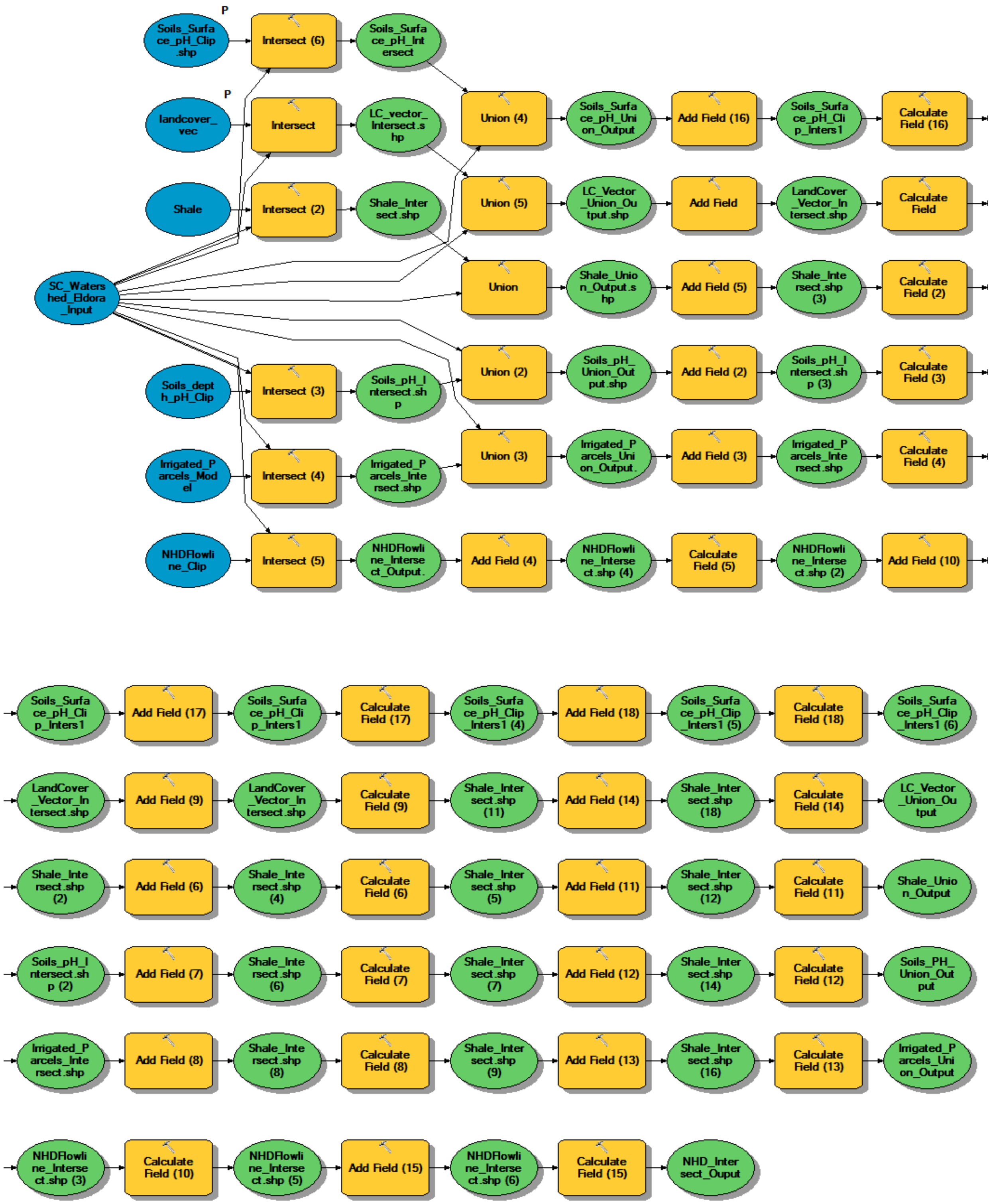


Figure 7 – ModelBuilder layout showing workflow from left to right from top image to bottom image

Table 9 –Landscape elements, derived parameters, spatial statistic types and geospatial variables

Geospatial Variable Analysis			
Landscape Element	**Landscape Parameter**	**Spatial Statistic Type**	**Geospatial Variable**
Geology	Shale	Calculate Area	Area of shale
Irrigation	Irrigated Land	Calculate Area	Area irrigated
Land Use Land Cover	Ag-Barren/Vegetated/ Developed/Wetland/ Open Water	Calculate Area	Area of Ag-Barren/ Vegetated/Developed/ Wetland/Open Water
Soils	pH Acidic/Neutral/Alkaline	Calculate Area	Area of Neutral/Slightly Alkaline/Moderate to Strongly Alkaline/Other
Hydrography	Streams	Line Statistics	Length of Streams/Canals

The following details the steps applied in the ModelBuilder environment:

1. Watershed layers were added as inputs and intersected with landscape sub-parameter layers, which were also added as inputs
2. The union tools were added as inputs, which were then strung between the resulting union output layers and the originally added input watershed layers
3. All of the successive processing steps involved the addition of attribute fields as inputs followed by spatial statistical types, such as area and length

2.4 Statistical Model

Selenium Prediction Statistical Model

Relationships between selenium concentrations, geospatial variables, streamflow, and precipitation were examined through application of a statistical package. Multiple linear regression analysis was performed using Microsoft® Excel software's add-in "Anaylsis Toolpak" as a prediction model for surface water selenium concentrations in Spring Creek and Fossil Creek. The general form for a multiple liner regression model is as follows (adapted from USGS, 1991):

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

where y is the response variable and $_i$ is the number of observations

β_0 is the intercept

β_1 is the slope coefficient for the first explanatory variable

β_2 is the slope coefficient for the second explanatory variable

β_k is the slope coefficient for the kth explanatory variable, and

x_{i-k} are the independent variables

ε is the remaining unexplained noise in the data (the error).

Dependent variables consisted of the 85^{th} percentile values of ranked selenium concentrations from 3 CDPHE water quality stations on Spring Creek and 2 CDPHE water quality stations on Fossil Creek resulting in a total of 5 dependent variable values. 17 exploratory variables consisted of 15 spatial variables determined geo-statistically, streamflow, and precipitation. Spatial variables were extracted from the GIS for each watershed drainage basin layer upstream of each water quality sampling station resulting in a total of 5 values for each of the 15 geospatial variables (Figure 6). The 5 streamflow stage values applied in the model were first organized by date and location to correspond with CDPHE selenium water quality sampling efforts. Secondly, streamflow values were developed from averaging the hourly measured stage values between 8:00am and 6:00pm for each day CDPHE water quality samples were taken, which were also sampled between the times of 8:00am and 6:00pm. Lastly, these values were then averaged at each location with respect to the 5 CDPHE water quality stations resulting in 5 independent streamflow values. Precipitation values were first organized by date to correspond in time (day/month/year) with selenium water quality samples, which were organized by the 5 water quality stations. The precipitation dataset was reduced by averaging the daily precipitation values from each of the 5 sets of values resulting in 5 individual values for inclusion into the statistical model.

A Pearson product-moment correlation analysis was applied with the dependent and exploratory variables for examination of relationships. Pearson correlation coefficient (Pearson *r*) allowed the determination of strength and direction of a relationship between two variables. This helped to enable a reduction of exploratory variables exhibiting the strongest relationships with selenium for inclusion into various multiple regression models before determining the best fitting multiple regression model. The multiple regression model was assessed using the adjusted coefficient of determination (R^2), which provides a measure of how well results are likely to be predicted by the statistical model (Draper and Smith, 1998). In addition, an *F-test* was computed for the coefficients associated with each exploratory variable to test the overall significance between the observed and predicted selenium values.

Statistical Model Validation

To evaluate the performance of the statistical model, the Nash–Sutcliffe efficiency (NSE) method was used. The NSE was applied to the observed and predicted values for model assessment. The NSE has been broadly used in water quality assessment and modeling studies (Motovilov, et al., 1999, Moriasi, et al., 2007, Prasad, et al., 2011). Nash–Sutcliffe efficiency is a normalized statistical measure that explains the model efficiency as a fraction of the observed value variance that is reproduced by the model (Nash and Sutcliffe, 1970). NSE specifies the fit of observed values compared to the predicted values in a 1:1 line. NSE is computed as shown in equation 1:

$$NSE = 1 - \left[\frac{\sum_i^n (o_i - p_i)^2}{\sum_i^n \left(o_i - o_{avg}\right)^2}\right]$$

where o_i is the observed value, p_i is the predicted value and o_{avg} is the mean of the observed data (Prasad, et al., 2011). NSE ranges between -∞ and 1.0. Values between 0.0 and 1.0 are considered acceptable levels of model performance, whereas values $\leq$ 0.0 indicate that the mean observed value is a better indicator than the predicted value, which indicates unacceptable performance (Moriasi et al., 2007).

2.5 GIS Statistical Model

GIS Selenium Prediction Statistical Model

Relationships between selenium concentrations, geospatial variables, streamflow, and precipitation were explored in a Geographic Information System environment. Within the GIS, Ordinary Least Squares (OLS) regression was a necessary step in building a properly specified spatial regression model for use in Geographic Weighted Regression (GWR). OLS was performed as a prediction model for surface water selenium concentrations in Spring Creek and Fossil Creek, but more importantly as a statistical tool for identifying exploratory variable(s) for inclusion in a GWR model. OLS is a global regression model, which computes a single equation calibrated using data from all features, which assumes the relationships between the dependent variable and exploratory variable(s) are static or fixed. GWR is a local regression model, which computes an equation for every feature in the dataset where each equation is calibrated using data from nearby features thereby allowing relationships to vary across the study area. The general form for OLS regression is identical to what was presented for multiple linear regression. With OLS regression, the actual magnitude of a response variable is modeled as a function of the magnitudes of one or more continuous explanatory variables; however, it was the probability of being in one of the two response groups that is modeled when the response is a binary categorical variable (USGS, 1991).

OLS regression was performed using the identical dependent variables applied in the multiple regression model, but an area weight was applied to the exploratory variables to account for the spatial aspects of nested sub-watersheds overlaying each other. Dependent variables consisted of 5 selenium concentration values determined from the 85th percentile of ranked selenium concentrations from the water quality stations. Exploratory variables consisted of 15 area weighted geospatial variables, streamflow, and precipitation, which were developed for each watershed drainage basin layer upstream of each CDPHE water quality station resulting in a total of 5 values for each variable. Similar to the Microsoft® Excel multiple regression modeling efforts, multiple OLS regression model runs were completed by inserting and deleting exploratory variables to determine the best fitting model.

OLS Statistical Model Assessment and Validation

A summary and diagnostics report was written as output after OLS processing is completed. The components of the OLS summary and diagnostics reports help to determine model performance and model significance among other model validity measures. There are 6 checks that must be assessed and met in determining a properly specified model before moving on to GWR (ESRI, 2013).

1. Model performance - The OLS model was assessed for performance using a coefficient of determination (R^2).
2. Statistically significant coefficients - Coefficients of explanatory variable(s) were assessed by the Probability and Robust Probability (p-value).

3. Defensible variable relationships - Coefficients of explanatory variable(s) were assessed for the proper type of relationship expected
4. No multicollinearity - Redundancy among variables was assessed by the Variance Inflations Factor (VIF)
5. Normal distributed residuals - Model bias was assessed by the Jarque-Bera statistic.
6. Randomly distributed model residuals - Residual spatial autocorrelation was assessed by running the Spatial Autocorrelation (Moran's I) tool.

Overall model statistical significance was assessed by the Joint F-Statistic and Join Wald Statistic (ESRI, 2013).

A Geographic Weighted Regression model was then developed using the statistically significant exploratory variable(s) from the OLS results, which was were shown to provide a properly specified model for use in GWR. GWR is used to better refine results first obtained from the OLS regression model with the addition of a regression equation fit to every feature in the dataset. Furthermore, the regression equations were calibrated from nearby features and the associated equations allowing relationship to vary across the study area. The GWR model was assessed using the adjusted coefficient of determination (R^2) and compared with the OLS regression results.

CHAPTER 3 – RESULTS

3.1 Overview

A selenium assessment study was performed using surface water quality data taken from 3 sampling stations on Spring Creek and 2 sampling stations on Fossil Creek of the Cache la Poudre River basin in Fort Collins, Colorado between the years of 2003 and 2008. This is the most recent data. The 2012 Update to the 2010 305(b) Report does not indicate any new or revised water quality assessments for Spring Creek or Fossil Creek. Additionally, the 2014 Update to the 2012 305(b) Report was cancelled. A GIS was developed with the following landscape elements: elevation, land use land cover, soil pH, geology, irrigation, and hydrography to predict the 85th percentile selenium concentration. Traditional and GIS statistical models were developed to predict selenium concentrations in Spring Creek and Fossil Creek.

3.2 GIS Cartographic Display of Landscape Elements

Cartographic maps of the geospatial data were created for data extraction and application in statistical models, as well for spatial assessment and visualization purposes. The following maps of landscape elements and parameters include: irrigation (Figure 8), geology (Figure 9), land use land cover (Figure 10), soil pH (Figure 11), and hydrography (Figure 12).

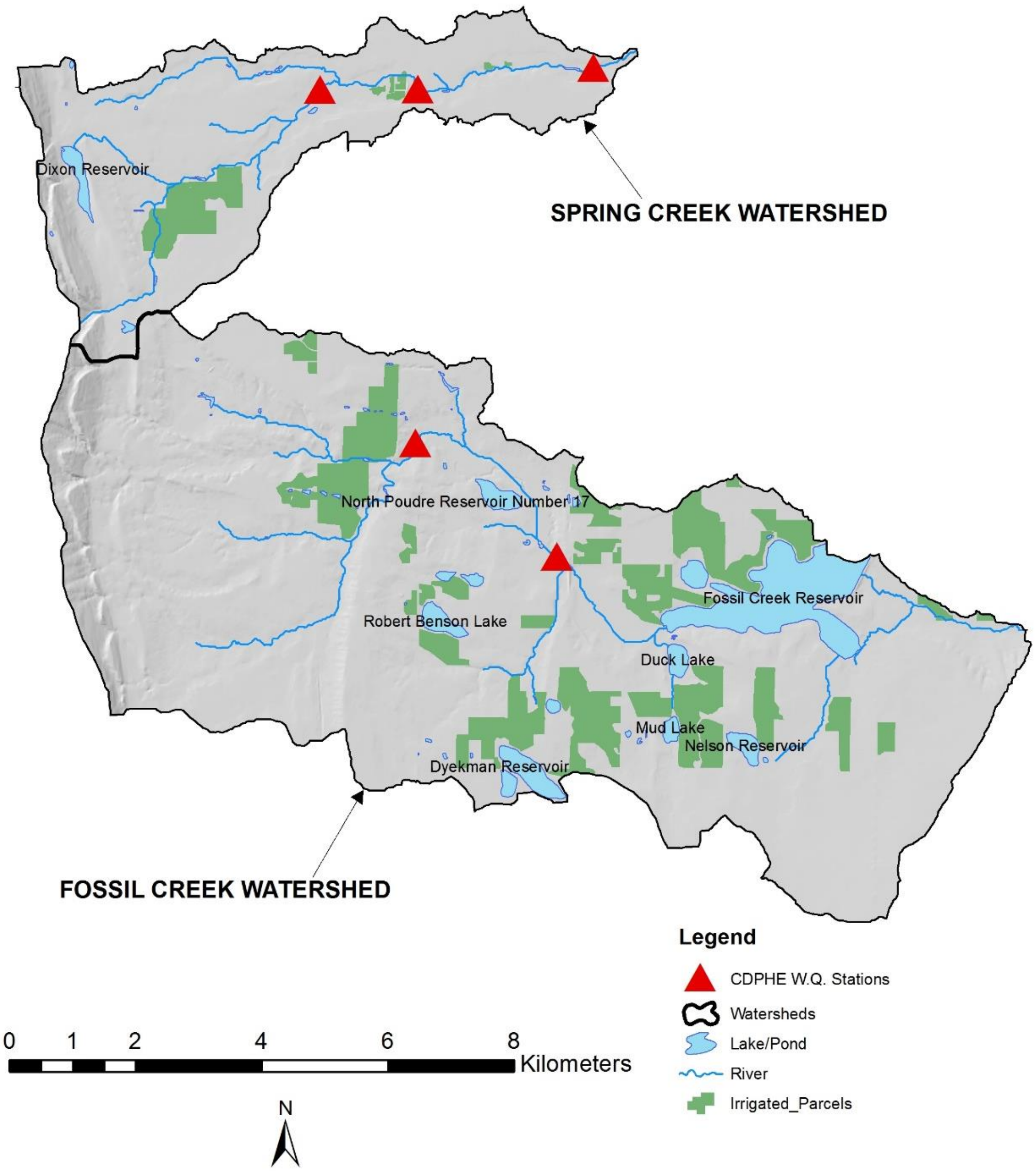


Figure 8 – Map of irrigated parcels within Spring Creek and Fossil Creek watershed boundaries (Data source: CWCB – DWR – CDSS Colorado Decision Support Systems)

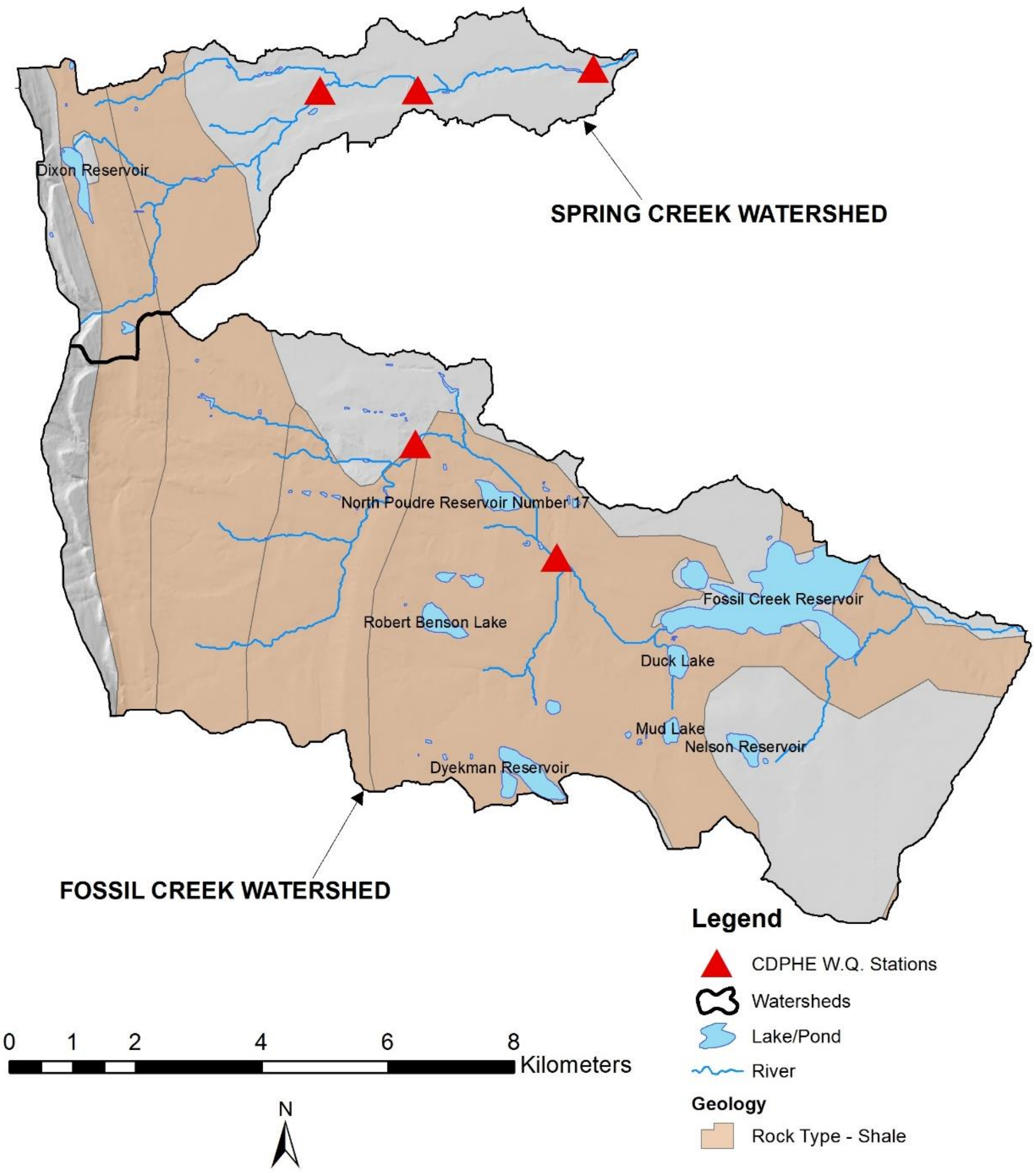


Figure 9 – Map of Cretaceous age geology rock type shale within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS Geospatial Data Gateway)

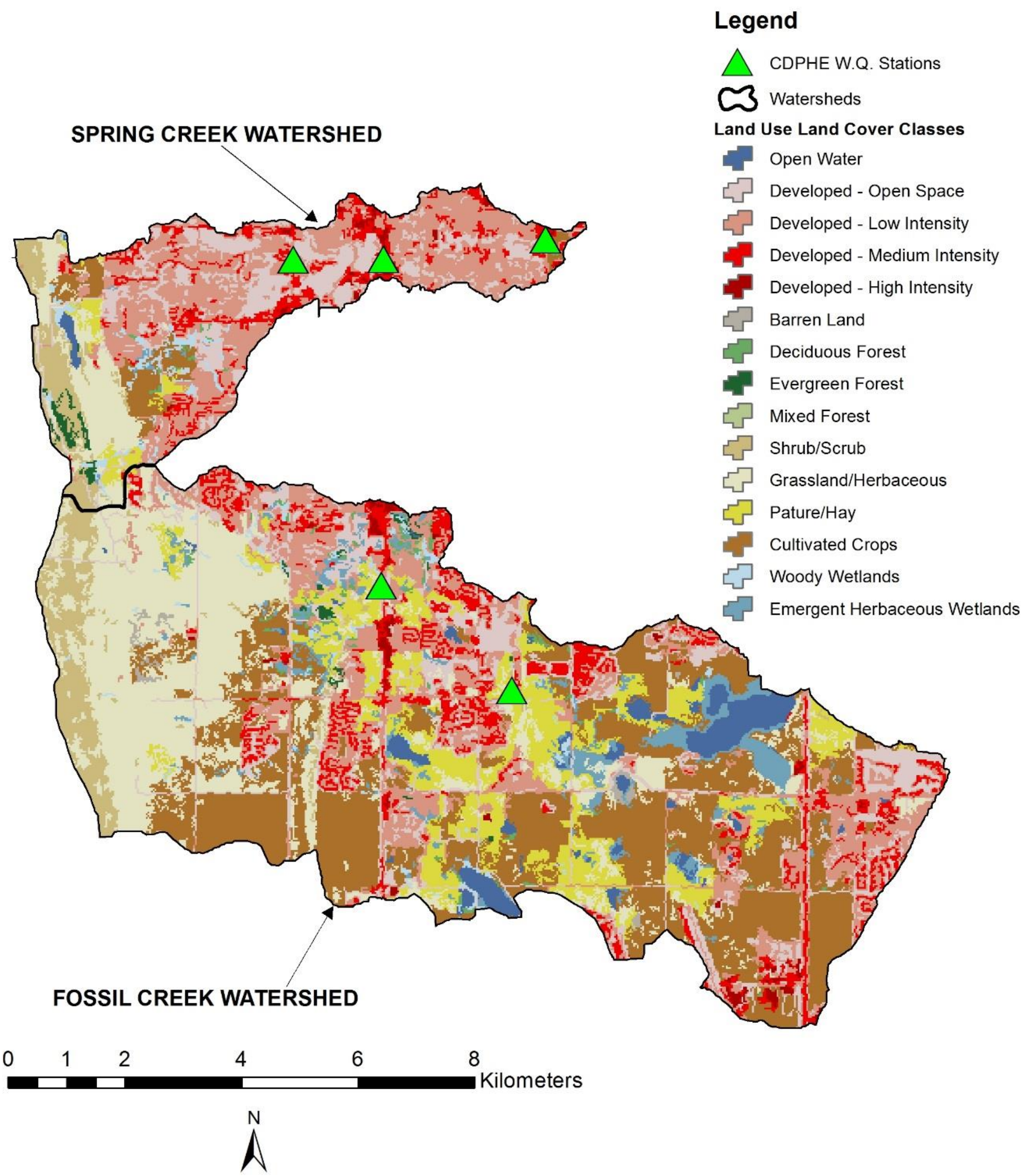


Figure 10 – Map of land cover data classes within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS Geospatial Data Gateway)

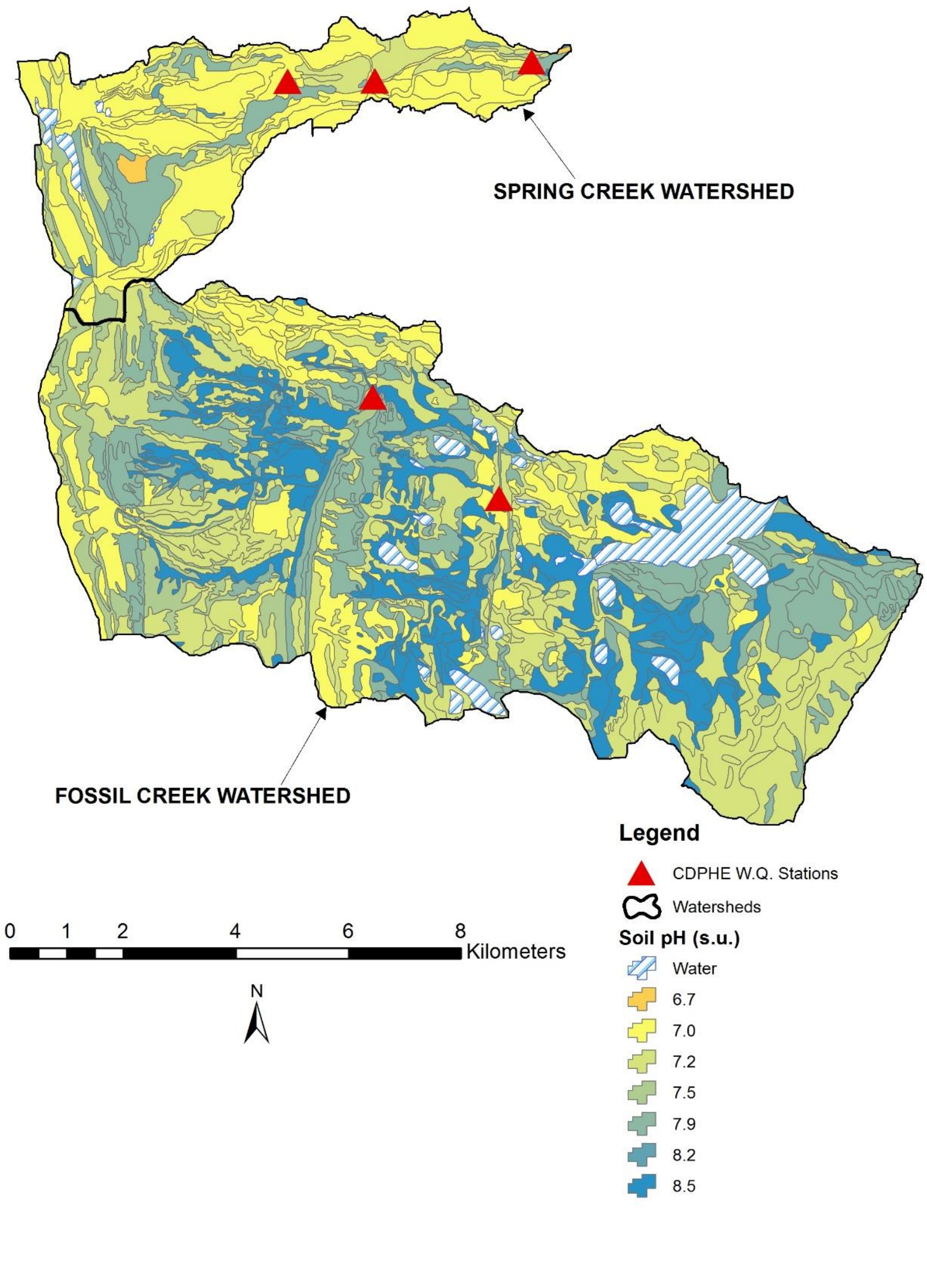


Figure 11 – Map of soil pH values from the soil surface of 0 cm to 122 cm depth within Spring Creek and Fossil Creek watershed boundaries (Data source: USDA – NRCS – SSURGO Soil Survey Geographic Database)

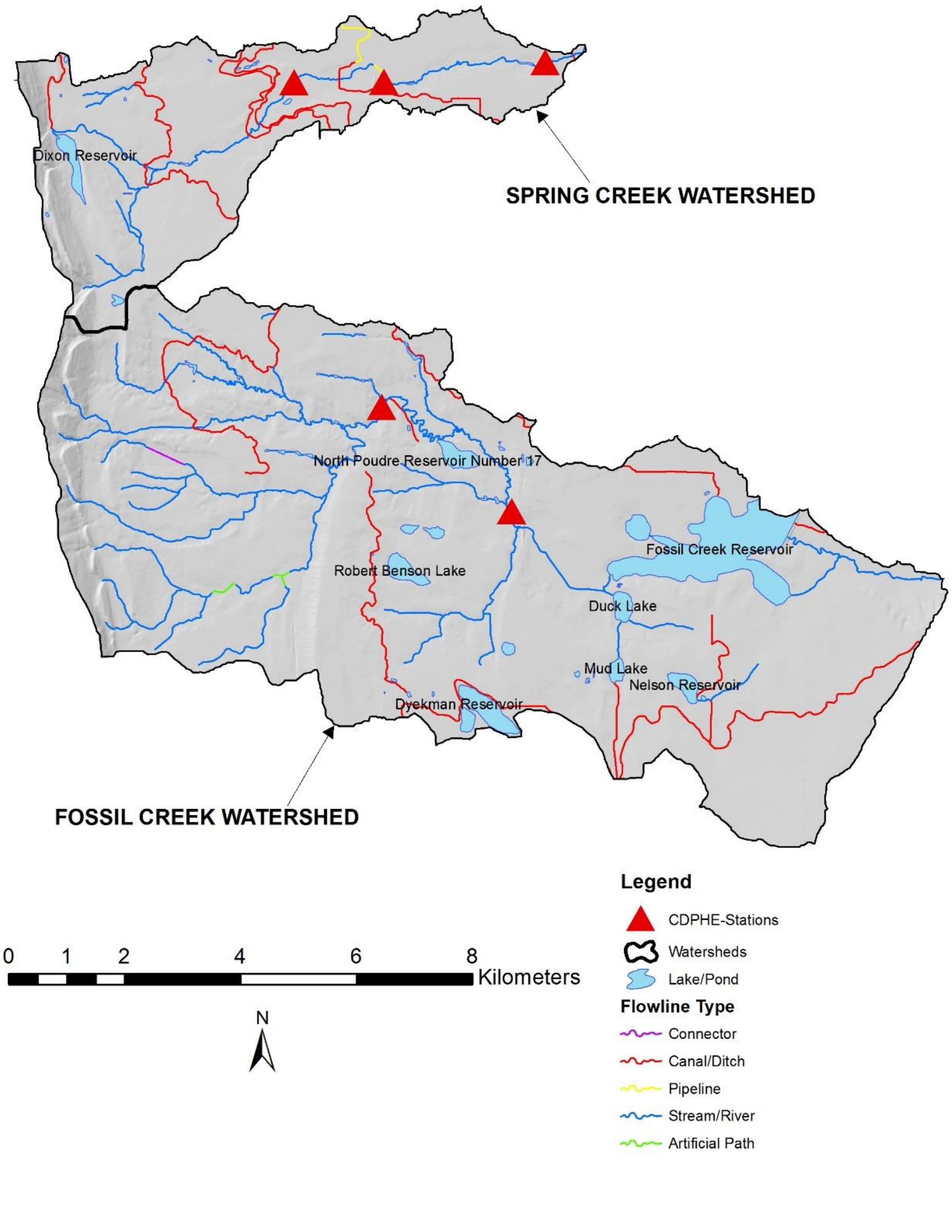


Figure 12 – Map of hydrography flowline types within Spring Creek and Fossil Creek watershed boundaries (Data source: USGS National Hydrography Dataset)

3.3 Statistical Analysis of Se Concentrations

Surface water selenium concentrations for Spring Creek were plotted over time (Figures 13 to 15) and summary statistics prepared (Table 10). Out of the 25 samples taken at the Shields Street station, 19 samples were higher than the Colorado Table Value Standard (TVS) of 4.6μg/L. The College Avenue and Edora Park sampling station each had 9 samples taken, all of which were below the Colorado TVS.

Surface water selenium concentrations for Fossil Creek were plotted over time (Figures 16 and 17) and summary statistics were prepared (Table 11). Out of the 10 samples taken at the College Avenue station, 8 samples were higher than the Colorado TVS and out of the 10 samples taken at the Trilby Road station, 5 samples were above the Colorado TVS.

Although many selenium concentration values were above the Colorado chronic Table Value Standard of 4.6μg/L, it is the 85th percentile of ranked data that determines the standard and whether or not the stream reach in question is listed as impaired (CDPHE, 2009). Selenium data were ranked to determine the 85th percentile at each sampling station (Figure 18). The 85th percentile values were also plotted in a bar graph with the Colorado chronic TVS for selenium indicated to reflect which stream segments are not meeting the current selenium water quality standard (Figure 19). The bar graph shows that the 85th percentile of ranked selenium data is 8.2μg/L for the Spring Creek at Shields Street sampling station. Spring Creek at College Avenue and Edora Park show that the 85th percentile of ranked data is below the chronic TVS at 2.7μg/L and 2.1μg/L, respectively. The 85th percentile for Fossil Creek at College Avenue and at Trilby Road is 13.7μg/L, higher than the chronic TVS.

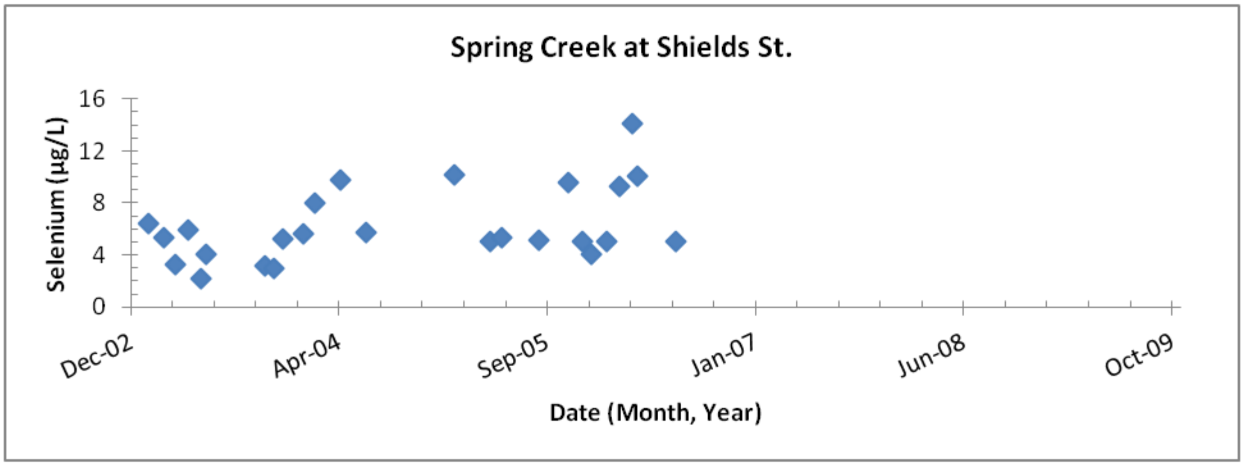


Figure 13 – Plot of selenium concentrations over time from water quality samples taken from Spring Creek at Shields Street

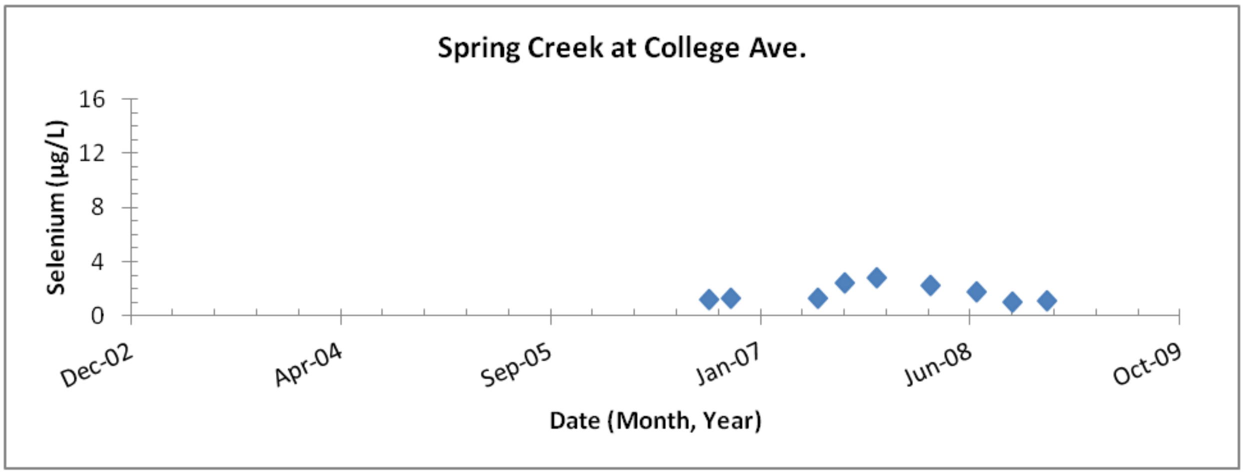


Figure 14 - Plot of selenium concentrations over time from water quality samples taken from Spring Creek at College Avenue

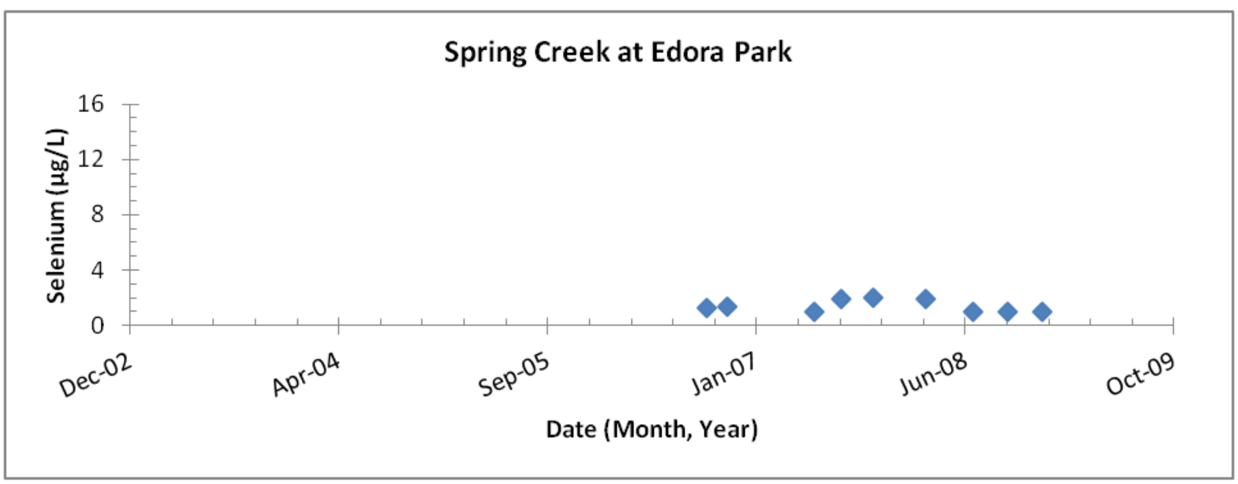


Figure 15 – Plot of selenium concentrations over time from water quality samples taken from Spring Creek at Edora Park

Table 10 – Summary statistics of selenium concentrations from Spring Creek

Statistic	***Spring Creek Shields Street***	***Spring Creek College Avenue***	***Spring Creek Edora Park***
Number of Samples	25	9	9
Mean (µg/L)	6.18	1.70	1.39
Median (µg/L)	5.30	1.32	1.25
Maximum (µg/L)	14.10	2.86	1.99
Minimum (µg/L)	2.10	1.00	1.00
Standard Deviation (µg/L)	2.84	0.67	0.44
Skewness	1.06	0.72	0.60

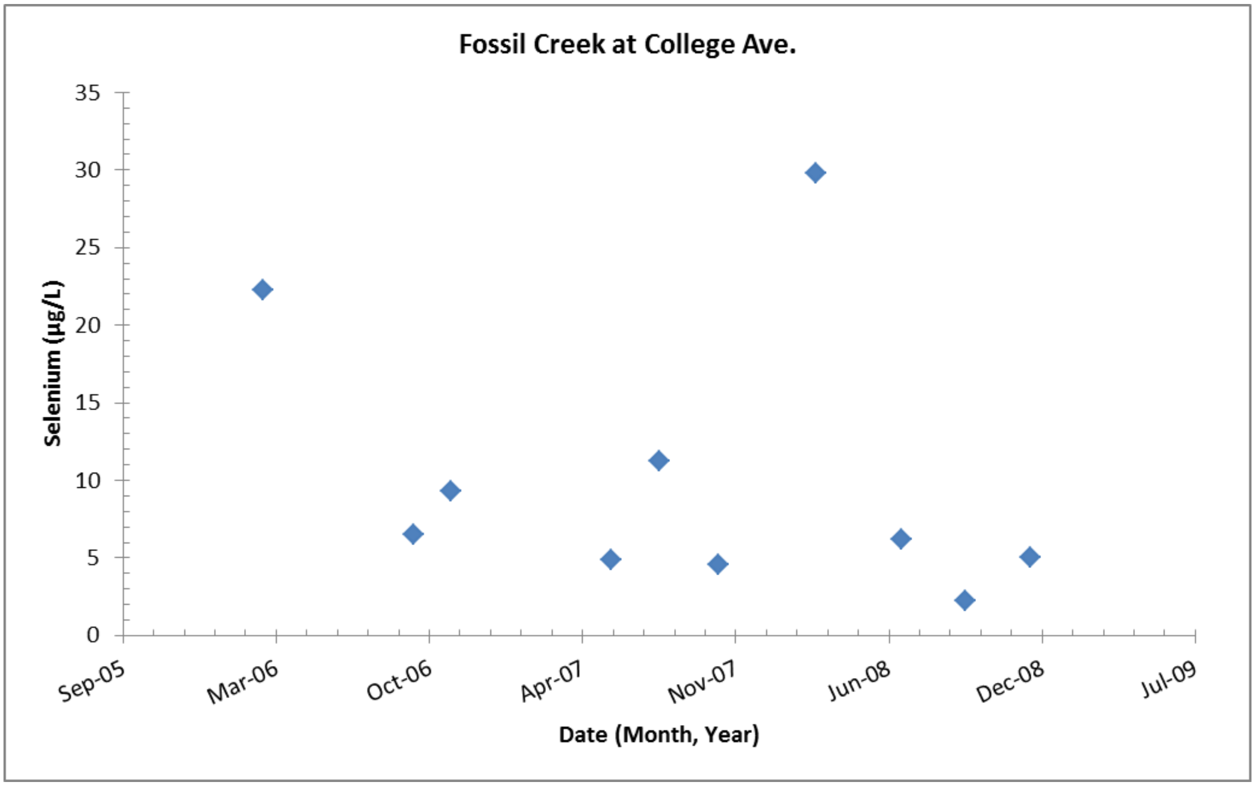


Figure 16 – Plot of selenium concentrations over time from water quality samples taken from Fossil Creek at College Avenue

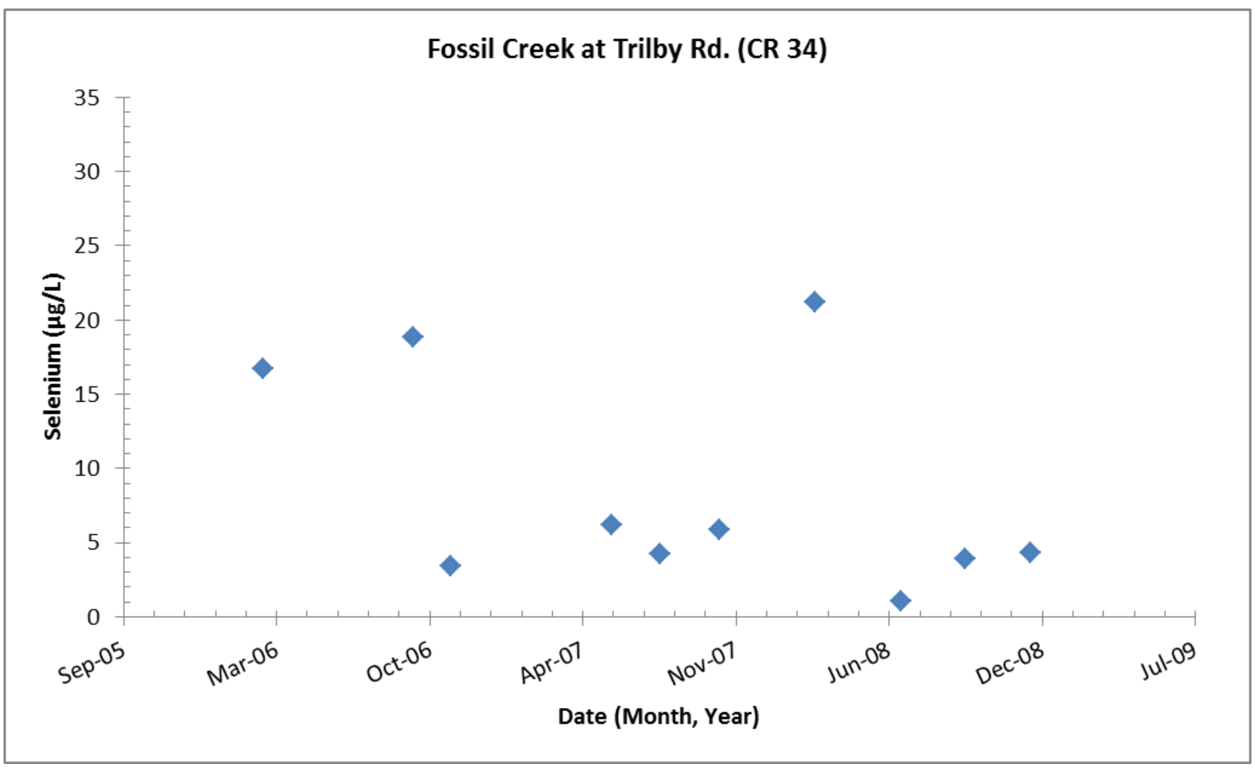


Figure 17 – Plot of selenium concentrations over time from water quality samples taken from Fossil Creek at Trilby Road

Table 11 – Summary statistics of selenium concentrations from Fossil Creek

Statistic	***Fossil Creek College Avenue***	***Fossil Creek Trilby Road***
Number of Samples	10.00	10.00
Mean (µg/L)	10.19	8.58
Median (µg/L)	6.34	5.08
Maximum (µg/L)	29.76	21.17
Minimum (µg/L)	2.20	1.03
Standard Deviation (µg/L)	8.89	7.34
Skewness	1.62	0.97

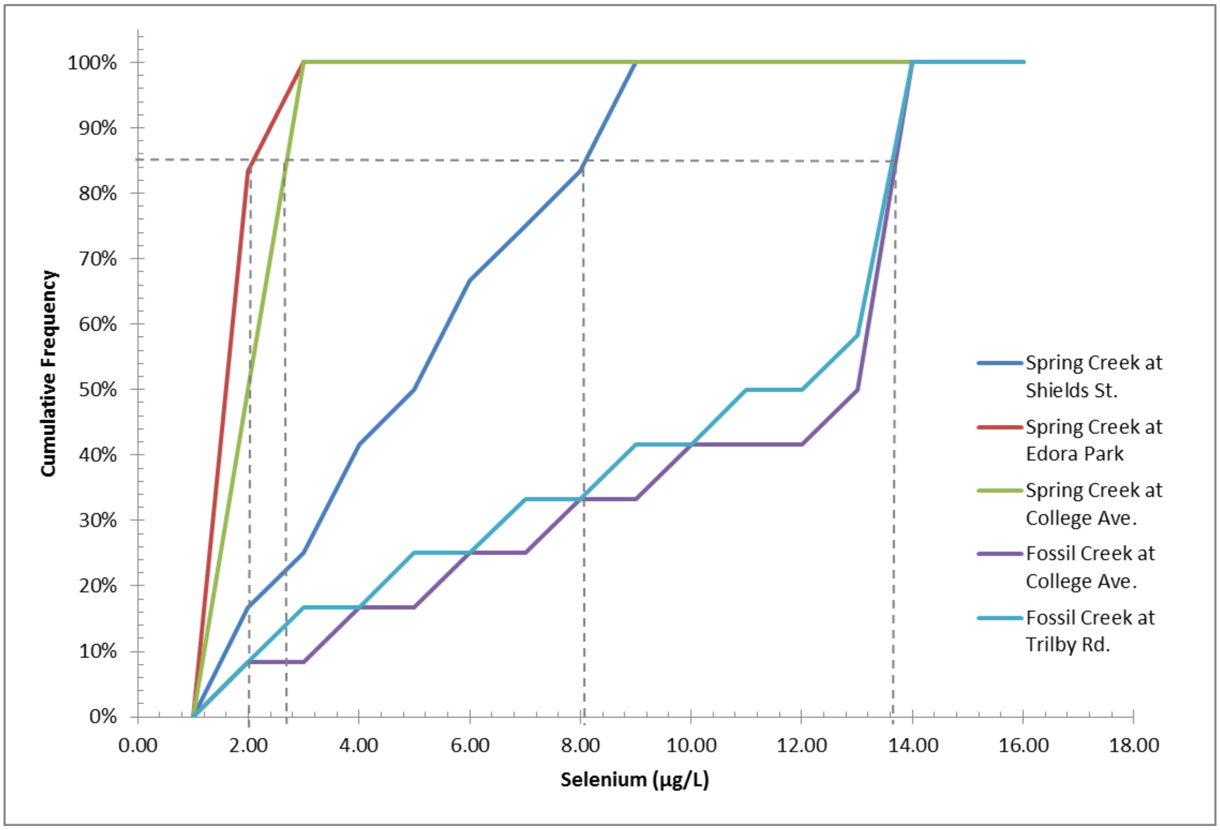


Figure 18 – Probability distribution functions. Cumulative frequency vs. selenium concentrations with 85th percentile concentrations identified

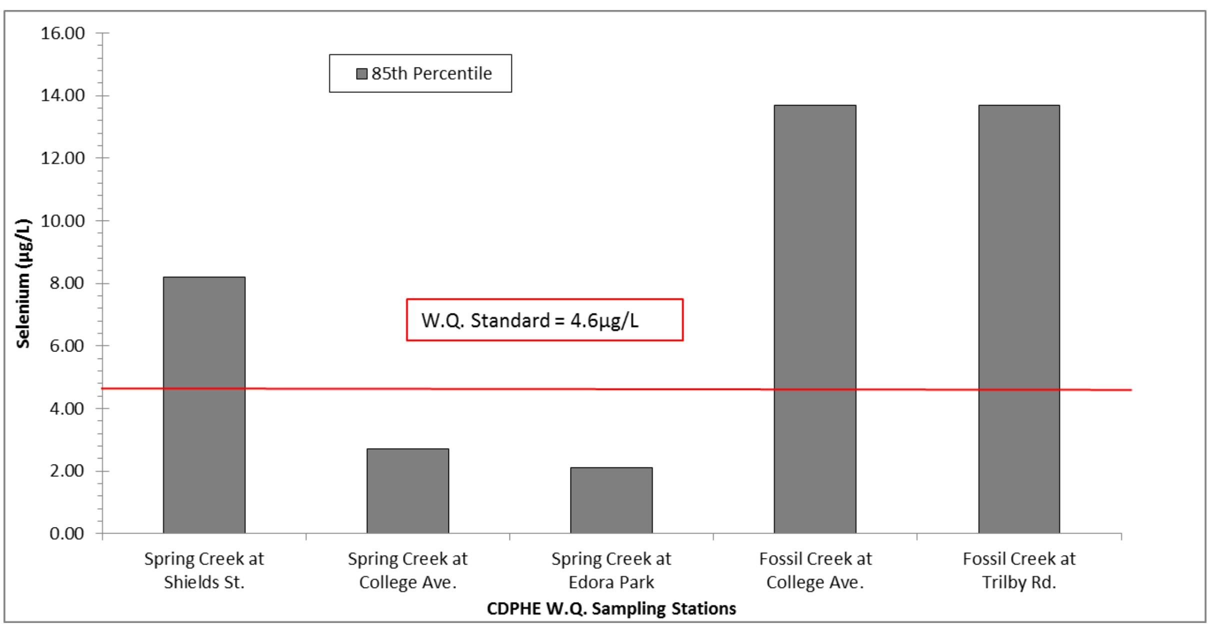


Figure 19 – Histogram depicting the 85th percentile of selenium concentrations with the CDPHE chronic selenium TVS identified

3.4 Geospatial Statistical Analysis of Landscape Parameters

Geospatial statistical analysis was performed using ModelBuilder in a GIS environment. As an approach to a selenium prediction model using CHRUs to predict selenium upstream of any water quality monitoring station, statistical processing in ModelBuilder workflows were run with landscape characteristics on sub-basin watershed layers. A total of 13 sub-basin watersheds were delineated in Spring Creek and a total of 21 sub-basin watersheds were delineated in Fossil Creek (Figure 5). Results of the statistical processing tools and Model Builder workflows were totaled and are listed by landscape parameters for Spring Creek and Fossil Creek sub-watersheds reflected by the 5 CDPHE water quality sampling station locations within Spring Creek and Fossil Creek (Table 12). The landscape parameter statistical values were area-weighted for the same 5 sub-watersheds of Spring Creek and Fossil Creek for inclusion in the GIS models (Table 13).

3.5 Statistical Model Results

Relationships between selenium concentrations, geospatial variables, streamflow, and precipitation were examined by a correlation analysis (Table 14 and Table 15). The Pearson correlation coefficient (Pearson *r*) allowed the determination of strength and direction of a relationship between two variables. Variables with a Pearson correlation coefficient between -0.8 and -1.0, and between 0.8 and 1.0 were considered to have a strong relationship with selenium. A total of 8 exploratory variables exhibited a Pearson *r* between -0.8 and -1.0, and between 0.8 and 1.0 which included: shale (0.847), agriculture (0.869), vegetated (0.883), pH_mod_alk (0.888), stream (0.831), water (-0.865), pH_neutral (-0.917), and pH_other (-0.808).

Table 12 – Landscape parameter values for Spring Creek and Fossil Creek sub-watersheds.

Landscape Parameters		**Sub - Watersheds**				
		SC Shields	**SC College**	**SC Edora**	**FC College**	**FC Trilby**
Area (sq km)		12.02	17.34	21.88	27.36	38.16
Irrigated Parcels (sq km)		1.16	1.39	1.30	1.37	2.38
Shale (sq km)		7.64	10.05	9.21	23.55	30.17
Land Cover (sq km)	*Ag/Barren*	1.41	1.93	1.77	6.46	7.94
	Developed	6.24	11.38	14.97	3.44	11.62
	Open Water	0.14	0.16	0.14	0.03	0.09
	Vegetated	3.64	4.78	4.38	16.65	17.22
	Wetland	0.58	0.67	0.62	0.78	1.28
Soil pH, 0 to 122cm deep (Standard Units grouped) (sq km)	*Slightly Alkaline*	11.40	17.34	21.13	22.33	31.04
	Moderate to Strongly Alkaline	0.07	0.11	0.10	5.01	6.81
	Neutral	0.21	0.23	0.21	0.00	0.00
	Other	0.34	0.48	0.44	0.01	0.31
NHD Flowline (sum of km)		20.70	37.43	41.35	52.58	72.18
NHD Flowline (km)	*Canal/Ditch*	9.77	22.11	24.30	9.59	16.24
	Stream/River	10.93	15.32	17.05	42.99	55.94

Table 13 – Landscape parameter values are area-weighted for Spring Creek and Fossil Creek sub-watersheds. Total area values are not weighted.

Landscape Parameters		Sub-Basin Watershed				
		SC Shields	**SC College**	**SC Edora**	**FC College**	**FC Trilby**
Area (sq km)		12.02	17.34	21.88	27.36	38.16
Irrigated Parcels (sq km)		0.13	0.10	0.08	0.07	0.12
Shale (sq km)		1.11	1.11	0.88	5.76	4.56
Land Cover (sq km)	*Ag/Barren*	0.2	0.21	0.17	1.68	1.30
	Developed	0.72	1.20	1.58	0.77	1.09
	Open Water	0.03	0.02	0.02	0.00	0.01
	Vegetated	0.69	0.64	0.50	3.88	2.82
	Wetland	0.08	0.06	0.05	0.12	0.12
Soil pH, 0 to 120cm deep (Standard Units grouped) (sq km)	*Slightly Alkaline*	1.62	2.03	2.25	5.50	4.53
	Moderate to Strongly Alkaline	0.01	0.02	0.01	0.96	0.80
	Neutral	0.02	0.02	0.01	0.00	0.00
	Other	0.06	0.06	0.05	0.00	0.02
NHD Flowline (sum of km)		2.56	3.97	4.14	11.25	9.31
NHD Flowline (km)	*Canal/Ditch*	0.92	2.38	2.40	1.78	1.73
	Stream/River	1.63	1.59	1.74	9.47	7.58

Table 14 – Correlation analysis output with Pearson coefficients indicated in bold and underlined for strong variable relationships with selenium

	Se_85th	*Precip_mm*	*Staff Gauge_m*	*Area*	*Irigation*	*Shale*	*Agriculture*	*Developed*	*Water*
Se_85th	1.000								
Precip_mm	0.012	1.000							
Staff Gauge_m	0.626	-0.433	1.000						
Area	0.642	-0.632	0.953	1.000					
Irigation	0.541	-0.410	0.959	0.883	1.000				
<u>Shale</u>	**<u>0.847</u>**	-0.469	0.884	0.938	0.827	1.000			
<u>Agriculture</u>	**<u>0.869</u>**	-0.455	0.856	0.923	0.788	0.998	1.000		
Developed	-0.626	-0.398	0.213	0.133	0.269	-0.197	-0.251	1.000	
<u>Water</u>	**<u>-0.865</u>**	0.296	-0.470	-0.630	-0.315	-0.792	-0.831	0.628	1.000
<u>Vegetated</u>	**<u>0.883</u>**	-0.458	0.780	0.880	0.694	0.979	0.990	-0.348	-0.898
Wetland	0.671	-0.395	0.970	0.918	0.986	0.905	0.876	0.116	-0.464
pH_slight_alk	0.482	-0.738	0.927	0.980	0.878	0.864	0.840	0.302	-0.502
<u>pH_mod_alk</u>	**<u>0.888</u>**	-0.404	0.865	0.916	0.795	0.996	0.998	-0.263	-0.825
<u>pH_neutr</u>	**<u>-0.917</u>**	0.377	-0.764	-0.854	-0.654	-0.963	-0.978	0.399	0.923
<u>pH_other</u>	**<u>-0.808</u>**	0.064	-0.208	-0.366	-0.044	-0.586	-0.639	0.810	0.952
Flowline	0.591	-0.707	0.926	0.992	0.878	0.927	0.909	0.161	-0.604
Canal	-0.766	-0.544	-0.016	-0.002	0.056	-0.312	-0.356	0.926	0.619
<u>Stream</u>	**<u>0.831</u>**	-0.493	0.895	0.954	0.824	0.998	0.996	-0.163	-0.793

Table 15 – Continuation of the correlation analysis output

	Vegetated	*Wetland*	*pH_slight_alk*	*pH_mod_alk*	*pH_neutr*	*pH_other*	*Flowline*	*Canal*	*Stream*
Se_85th									
Precip_mm									
Staff Gauge_m									
Area									
Irigation									
Shale									
Agriculture									
Developed									
Water									
Vegetated	1.000								
Wetland	0.799	1.000							
pH_slight_alk	0.788	0.885	1.000						
pH_mod_alk	0.984	0.883	0.826	1.000					
pH_neutr	-0.993	-0.768	-0.746	-0.978	1.000				
pH_other	-0.732	-0.210	-0.214	-0.639	0.779	1.000			
Flowline	0.868	0.906	0.989	0.895	-0.828	-0.331	1.000		
Canal	-0.415	-0.087	0.193	-0.388	0.487	0.769	0.063	1.000	
Stream	0.977	0.901	0.884	0.993	-0.963	-0.582	0.940	-0.282	1.000

A multiple linear regression analysis was performed using Microsoft® Excel as a prediction model for the 85th percentile surface water selenium concentrations in Spring Creek and Fossil Creek (Table 16). Different combinations of the strongly correlated variables were applied in a regression analysis before determining the best fitting multiple regression model.

A combination of 3 variables including pH_mod_alk (area of moderate to strongly alkaline soils) Pearson r = 0.888, stream (length of streams) Pearson r = 0.831, and shale (area of shale) Pearson r = 0.847, provided the best fit model with an adjusted R^2 value of 0.99, and a p-value of 0.01. ANOVA output for Significance F is 0.012. Regression coefficients for Intercept, pH_mod_alk, stream, and shale are 24.038, 9.516, -0.782, -1.039, respectively. P-values for the regression coefficients of Intercept, pH_mod_alk, stream, and shale are 0.01, 0.01, 0.03, and 0.06, respectively. A summary of the regression coefficients in a fitted line is as follows:

$$[Se\ \mu g/L = 24.038 + 9.516(ALK) - 0.782(STR) - 1.039(SHL)]$$

ALK = area (km^2) of moderate to strongly alkaline soils

STR = length (km) of streams

SHL = area (km^2) of shale

Statistical Model Validation Results

The Nash–Sutcliffe efficiency (NSE) method was used to evaluate the performance of the statistical model. NSE was applied to the observed and predicted values for model assessment (Table 20). Calculated NSE coefficient of 0.90 for SC Shields, 0.99 for SC College, 0.99 for SC Edora, 0.99 for FC College, 0.99 for FC Trilby, and 0.99 for the NSE total

Table 16 – Multiple regression dependent and independent variables

Variable Type	Regression Model Variables		Sub - Watersheds SC Shields	SC College	SC Edora	FC College	FC Trilby
y	*Selenium (µg/L)*		8.20	2.70	2.10	13.70	13.70
x	*Shale (sq km)*		7.64	10.05	9.21	23.55	30.17
x	*Land Cover (sq km)*	*Ag/Barren*	1.41	1.93	1.77	6.46	7.94
x		*Vegetated*	3.64	4.78	4.38	16.65	17.22
x		*Open Water*	0.14	0.16	0.14	0.03	0.09
x	*Soil pH, 0 to 122cm deep (Standard Units grouped) (sq km)*	*Mod. to Str. Alkaline*	0.07	0.11	0.10	5.01	6.81
x		*Neutral*	0.21	0.23	0.21	0.00	0.00
x		*Other*	0.34	0.48	0.44	0.01	0.31
x	*NHD Flowline (km)*	*Stream/River*	10.93	15.32	17.05	42.99	55.94

indicate an acceptable level of model performance for each individual sub-watershed predictions as well as for the overall model.

3.6 GIS Statistical Model Results

Relationships between selenium concentrations, geospatial variables, streamflow, and precipitation were explored in a Geographic Information System environment. Ordinary Least Squares regression was performed as a spatial prediction model for surface water selenium concentrations in Spring Creek and Fossil Creek, which is a necessary step in building a properly specified spatial regression model for use in Geographic Weighted Regression. Unlike the multiple regression model in Microsoft® Excel, the OLS regression model was developed using area-weighted variables (Table 13).

Similar to the Excel multiple regression model, OLS regression requires manual selection of variables. Wetlands (area of wetlands) provided the best fit model, with an adjusted R^2 of 0.98 and p-value < 0.05. Regression coefficients of Intercept (-6.584) and Wetlands (170.509) show p-values < 0.05. A summary of the regression coefficients in a fitted line is as follows:

$$[Se\ \mu g/L\ = -6.584 + 170.509(wetlands) - 0.782]$$

An output map of over/under predictions was produced after OLS regression processing was completed, which is used to help determine if key explanatory variables are missing due to clustering (Appendix E). To ensure regression residuals are spatially random, the Spatial Autocorrelation (Moran's I) tool produces an output for visual analysis (Appendix F). Summary results and a diagnostics output are also written after OLS processing to address the 6 checks in determining a properly specified model (Appendix G).

A Geographic Weighted Regression model was developed using the statistically significant exploratory variable "wetlands" from the OLS results. GWR was used to better refine results first obtained from the OLS regression model with the addition of a regression equation fit to every feature in the dataset (Appendix H). The GWR model with wetlands applied as the independent variable resulted in an adjusted R^2 value of 0.98. Shale (area of shale), provided the second best performing GWR model, with an adjusted R^2 of 0.66. The measured and predicted 85^{th} selenium concentration values were plotted against each other from both the best performing Excel multiple linear regression model and the best performing GWR model (Figure 20). Because GWR develops an equation for each feature, an example of regression coefficients for Intercept and Wetlands for the SC Edora sub-watershed is -6.585 and 170.535 with a local R^2 of 0.98. Similar to the OLS results output, GWR produced an over/under predictions map. More importantly, this output map has the ability to be reconfigured to map the wetland coefficients for each feature in the study area thereby identifying the most appropriate location(s) to address Se contamination through additional site investigation and management practices (Figure 21).

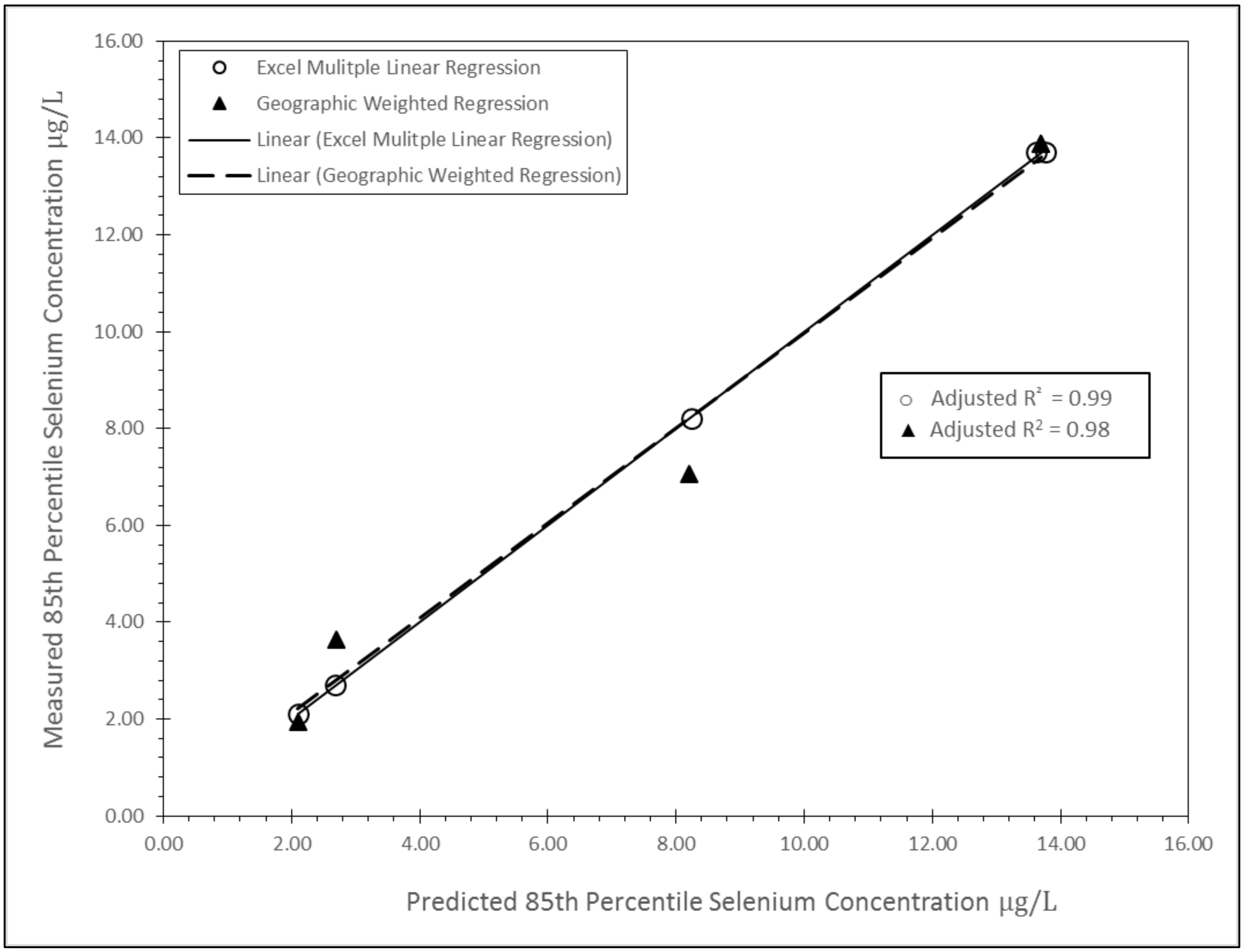


Figure 20 - Measured versus predicted 85th selenium concentration values (µg/L) for Excel multiple linear regression and Geographic Weighted Regression models.

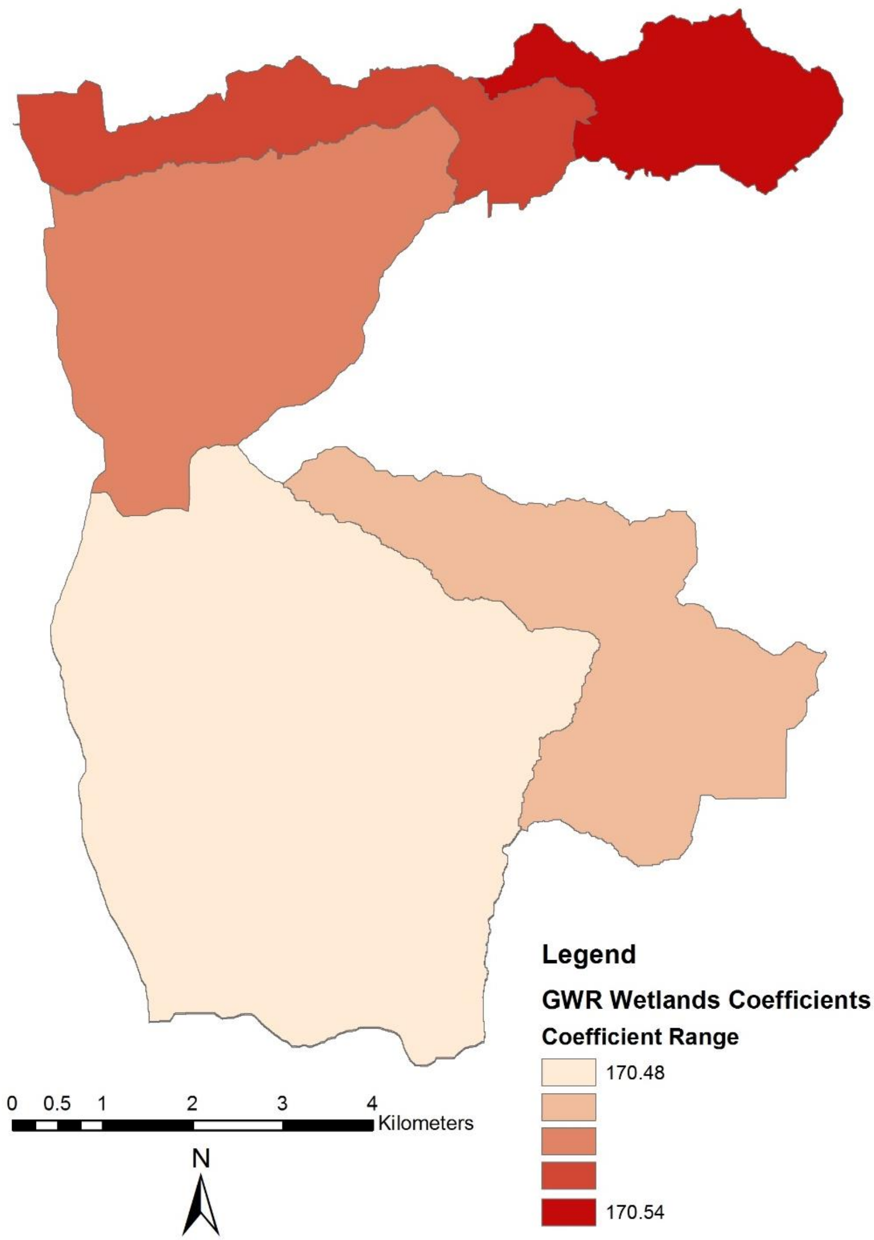


Figure 21– Geographic Weighted Regression output with the wetland coefficient mapped for each sub-watershed

CHAPTER 4 – DISCUSSION

4.1 Integration of Landscape Elements within a GIS Environment

Landscape Element Development and GIS Integration

Landscape elements were integrated into a Geographic Information System for model development for the prediction of surface water selenium concentrations along Spring Creek and Fossil Creek in Fort Collins, Colorado. The GIS was used to aid organizing, viewing, manipulating, and editing of the geospatial data. Past studies, including those of See et al., 1992; Tong and Chen, 2002; Xiao and Ji, 2007; Rothwell et al., 2010; Leib et al., 2012; and Phung and Stednick, 2012, have completed water quality investigations and predictions applying GIS generated landscape characteristics before. For this study, a literature review with respect to selenium and selenium response provided the initial background information required for watershed investigations and consequently the identification of landscape elements with linkages to selenium for study area characterization and integration within a GIS environment. Landscape elements included the following: elevation, land use land cover, soil pH, geology, irrigation, and hydrography data.

Landscape element classes from land use land cover and soil pH datasets were reclassified to better categorize properties that have been demonstrated in previous studies to show relationships with selenium (Ohlendorf et al., 1986, Masscheleyn et al., 1990, Seiler, 1995). Using ModelBuilder, a geoprocessing workflow tool within ArcGIS, spatial data was created, edited, modeled, and extracted specific to the study area and statistically processed for spatial physical characteristics (ESRI, 2012). Physical characteristics included the area and length of landscape elements, which provided statistical values and thus the geospatial variables.

Geospatial variables were first extracted out of the study area from smaller sub-basin watersheds as a future modeling approach using CHRUs (Figure 5). These smaller watershed boundaries were delineated through hydrologic response (e.g. slope and general topography) and could potentially identify locations of selenium contamination with the integrated geospatial variables. Practicability of defining hydrologic response units as combinations of spatial landscape elements for modeling purposes has been tested before (Vigiak et al., 2006). Hydrologic units could also potentially be used in a prediction model for unmonitored areas. Studies have been established where hydrologic response units were developed with the addition of chemical dynamics for development of CHRUs (Kern and Stednick, 1993; Bende et al., 1995). However; no studies were found linking selenium with CHRUs. Geospatial variables were successfully extracted in the much smaller scale sub-basin watersheds as a practical start for this type of spatial modeling. Methods used can be replicated and reproduced for other areas outside of this study area with respect to basin delineations and geospatial variable extraction. Chemical attributes or selenium concentrations however, were not acquired in these smaller sub-basin watersheds for this study.

Due to the location of the CDPHE water quality sampling locations, sub-watersheds were delineated upstream of 3 sites along Spring Creek and 2 sites along Fossil Creek. Geospatial variables were again successfully extracted from these 5 sub-watersheds for incorporation in selenium prediction model development. The process and methods used to integrate landscape elements within a GIS environment is not new, however; this study's focus on selenium and its geospatial relationships to landscape elements within the Fort Collins regional environment has not been done to date.

4.2 Development of Traditional and GIS Statistical Models for Se Prediction

Traditional and GIS statistical models were developed using geospatial landscape data to predict selenium concentrations in Spring Creek and Fossil Creek. Dependent variables consisted of the 85th percentile values of ranked selenium concentration values from 3 CDPHE water quality stations on Spring Creek and 2 CDPHE water quality stations on Fossil Creek resulting in a total of 5 dependent variable values. 17 exploratory variables consisted of 15 spatial variables determined geo-statistically, streamflow, and precipitation. GIS integrated geospatial variables were extracted for each watershed drainage basin layer upstream of each water quality sampling station resulting in a total of 5 values for each of the 15 geospatial variables.

Geospatial Variables

Geospatial variables were established by area and length of landscape elements bounded within watersheds. See et al., (1992) similarly used area and length of landscape elements for variables in a regression model; however, their study focused on median selenium discharges rather than selenium concentrations. Geospatial variables were used for study area characterization.

Correlation Analysis

A Pearson product-moment correlation analysis provided an assessment of the linkages between geospatial variables developed from landscape characteristics, precipitation, streamflow, and surface water quality selenium concentrations. To reduce the number of geospatial variables for inclusion in multiple regression models, regression coefficients from simple linear regression analyses were used as criterion values rather than Pearson correlation coefficients (Leib et al., 2012). A p-value of 0.01 was the first criterion used to select variables for multiple regression models. For this study's correlation analysis, variables with a Pearson correlation coefficient between -0.8 and -1.0, and 0.8 and 1.0 were considered to have a strong relationship with selenium, enabling a reduction of exploratory variables exhibiting the strongest relationships with selenium for inclusion into multiple regression models before determining the best fitting multiple regression model. A total of 8 exploratory variables exhibited this criteria, which included: shale (0.847), agriculture (0.869), vegetated (0.883), pH_mod_alk (0.888), stream (0.831), water (-0.865), pH_neutral (-0.917), and pH_other (-0.808). Choosing a lower cut-off value for Pearson *r* values would have created more variables for inclusion in regression models.

Other studies showed that the 8 exploratory variables chosen for this study's regression model development were related to elevated selenium concentrations. Shales have been shown to be the principal sources of selenium-toxic soils of the Rocky Mountain foothills of the United States (Shamberger, 1983). Median selenium discharge is highly correlated with the area of Cody Shale and total sub-basin area (See et al., 1992). Agriculture has been shown to be directly related to selenium, as well as when the regional environment exhibits areas underlain by Cretaceous marine or sedimentary rocks that are weathered, and when applied fertilizer is manufactured with selenium (Shamberger, 1983; Presser, et al., 1994; Fordyce, 2005). Coupled with the proper landscape characteristics, selenium contamination has also been determined to be caused by irrigation drainwater (Ohlendorf et al., 1986; Seiler, 1995). Selenate, the most important species of selenium in relation to this study and surface water quality, has been shown to be the most prevalent at higher pH levels (8.5 and 9 s.u.) when under highly oxidized conditions providing the use of moderately alkaline soils to be reasonable for model use (Masscheleyn et al., 1990). Soil pH geospatial variable applied in the correlation analysis for moderately alkaline soils was classified between the range of 8.2 and 8.5 pH standard units, which was also determined based on the range of soil pH s.u. found to occupy the study area, 6.7-8.5 s.u. (Table 7).

Traditional Statistical Model

Dependent variables applied in the multiple regression model consisted of 5 selenium concentration values determined from the 85th percentile of ranked selenium concentrations from 3 CDPHE water quality stations on Spring Creek and 2 CDPHE water quality stations on Fossil Creek (Table 16). Exploratory variables consisted of 3 variables including pH_mod_alk (area of moderate to strongly alkaline soils, stream (length of streams), and shale (area of shale), which provided the best fit model, adjusted R^2 = 0.99. Previous studies using regression models have found the variables of length of streams and area of shale to have influence on selenium concentrations (See et al., 1992; Leib et al., 2012). See et al., (1992) determined from a traditional stepwise regression analysis, that area of irrigated land and length of irrigation canals within each sub-basin were the most important factors explaining the variability of median selenium discharges ($R^2 = 0.97$). Leib et al., (2012) determined that sub-basin area and the area of irrigated area over Mancos shale were the most important landscape variables in determining the best fit stepwise regression model ($R^2 = 0.82$). Pearson *r* value for basin area for this study was 0.64, which suggests that selenium concentrations increase as basin area increases. The use of soil pH however, measured as the area of alkaline soils as used in this study, has not been found in any previous studies looking at selenium response and/or prediction.

An adjusted R^2 value of 0.99 from this study's regression model suggests that the overall model permanence is excellent and that the geospatial variables are good predictors of selenium. An *F-test* p-value of 0.01, indicates statistical significance between the observed and predicted selenium values. The NSE method was used to evaluate the performance of the statistical model. Calculated NSE coefficients of 0.90 for SC Shields, 0.99 for SC College, 0.99 for SC Edora, 0.99 for FC College, 0.99 for FC Trilby, and 0.99 for the NSE total indicate an acceptable level of model performance for each individual sub-watershed predictions as well as for the overall model. Prasad et al., (2011) calculated NSE coefficients of 0.82-0.86, 0.65-0.82, and 0.70-0.82 when predicting dissolved oxygen in three dimensions in Chesapeake Bay for a surface layer, middle layer, and bottom layer, respectively. This study's NSE scores further demonstrate that the model presented here has the ability to predict observed selenium concentrations and that the model could be used in management or regulatory settings for selenium prediction and might provide identification of selenium sources in streams.

GIS Statistical Model

Unique to this study, specifically relating to selenium response and prediction, is the use of ArcGIS's modeling relationships capabilities. Ordinary Least Squares regression was performed as a spatial prediction model for surface water selenium concentrations, as well as a necessary step in building a properly specified spatial regression model for use in Geographic Weighted Regression. The OLS regression model was developed using the area-weighted geospatial variables as an attempt to accommodate the spatial aspects and response of the variables in relation to the size of the watersheds they are occupying. Because the sub-watersheds themselves are "nested" in some cases or on top of each other, area-weighted variables were meant to eliminate the possible altercations associated with multiple area values being summed together from one sub-watershed to another. For example, the area of shale in relation to the area of the most downstream sub-watershed of SC Edora spatially includes the area of shale from both the upstream watersheds of SC College and SC Shields. By applying the area-weight, the area of shale in relation to the area of a sub-watershed is proportional to each sub-watershed modeled.

Wetlands, single exploratory variable, provided the best fit model after selection of variables. Xiao and Ji, 2007 found that the use of a single landscape metric, proportion of mine waste area, could predict surface water quality ($R^2 = 0.60$). The use of wetlands, measured as area of wetlands, as applied in this study has not been found in any previous studies pertaining to selenium response and/or prediction. This model's adjusted R^2 value of 0.98, suggests that the overall model permanence is excellent and that the geospatial variables are good predictors of selenium. The purpose of GWR is to better refine results from OLS regression. GWR results compared to OLS results, regarding adjusted R^2 values, does not seem to suggest an improvement. GWR results output map does however provide additional benefits not capable in other modeling methods. The independent variables' regression coefficients for each feature in the study area were mapped providing a visual display of areas that might provide the most effective management efforts with respect to selenium contamination (Figure 21). In this case, the mapped GWR wetland coefficients show Spring Creek to be the most important watershed for cost effective selenium management measures. The sub-watersheds of Spring Creek moving upstream show wetland coefficients to be more important than Fossil Creek sub-watersheds as well, possibly a spurious correlation due the nesting of watersheds and wetland area from upstream to downstream.

Previous studies with respect to selenium management and selenium modeling have not shown soil pH, specifically the area of moderate to strongly alkaline soils, or the area of wetlands to have been considered as selenium response variables. These landscape characteristics are shown here to be highly correlated with selenium. It is advisable to further research the use of landscape characteristics, with regards to area and length in a GIS for selenium prediction and management. Additionally, constraints existed on the number of variables that could be included in the regression models due to the sample size or number of selenium observations determined from using 5 sub-basins for which surface water selenium concentrations could be predicted.

CHAPTER 5 – CONCLUSIONS

In this study, 17 exploratory variables developed from landscape characteristics and environmental factors were analyzed in relation to the 85th percentile of ranked surface water selenium concentrations from 3 sub-watersheds in Spring Creek and 2 sub-watersheds in Fossil Creek in Fort Collins, Colorado. Landscape characteristics included sub-watershed area, shale, irrigated parcels; (land use land cover) ag/barren, developed, open water, vegetated, wetland; (soil pH) slightly alkaline, moderate to strongly alkaline, neutral, other; (NHD flowline) sum of flowline, canal/ditch, and stream/river. Environmental factors included precipitation and streamflow.

Integration of landscape elements within a GIS was relatively simple in terms of data collection and the uploading into an ArcGIS environment. Processes involved with geostatistical analyses worked well in that variables as measured by area and length were easily integrated within the traditional regression model and GIS regression models. Rather than using average or median values for relationship investigation and prediction similar to other studies, the 85th percentile values of ranked surface water selenium concentrations were used for which water quality standards for selenium are based.

A correlation analysis provided a clear depiction of strong relationships between exploratory variables and selenium. Results from this study indicate that the geospatial variables pH_mod_alk (measured as area of moderate to strongly alkaline soils), streams (measured as length of streams), and shale (measured as area of shale) provide the best fit traditional regression model using Microsoft® Excel, adjusted $R^2 = 0.99$. This model suggests these variables to be the most important and reliable predictors of selenium in

Spring Creek and Fossil Creek watersheds. Summary of the regression coefficients in a fitted line is as follows:

$$[Se\ \mu g/L\ = 24.038 + 9.516(ALK) - 0.782(STR) - 1.039(SHL)]$$

ALK = area (km^2) of moderate to strongly alkaline soils

STR = length (km) of streams

SHL = area (km^2) of shale

Results from this study also indicate that the geospatial variable wetlands, measured as area of wetlands, provided the best fit regression model using both Ordinary Least Squares Regression, adjusted R^2 = 0.98, and Geographically Weighted Regression, adjusted R^2 = 0.98, when area weighted. Summary of the OLS regression coefficients in a fitted line is as follows:

$$[Se\ \mu g/L\ = -6.584 + 170.509(wetlands) - 0.782]$$

GWR develops an equation for each feature; therefore, an example of regression coefficients for Intercept and Wetlands for the SC Edora sub-watershed is -6.585 and 170.535.

Because both GIS regression model methods resulted in an adjusted R^2 = 0.98, neither model can be said to be better than the other when comparing this particular statistical measure. A GIS based approach can however provide useful information not possible using only traditional regression model packages. The GIS approach provides spatial aspects to landscape elements led by visualization of patterns and through GWR analysis, identification of contaminated areas depicted through regression coefficients. Results from regression methods used in GWR, explicitly the regression coefficients, can be used to help prioritize areas, Spring Creek watershed in this case, where management practices and TMDL development might be most cost effective.

Constraints existed in both the traditional multiple linear regression model as well as in the GIS spatial Geographic Weighted Regression model. Limitations of both of these models include a very limited number of water quality sampling locations and subsequently a limited number of selenium concentrations. Due to these limitations, care should be applied when using any of these models in other areas of Colorado.

CHAPTER 6 - RECOMMENDATIONS

Additional modeling efforts and field investigations in the lower Cache la Poudre River system near Fort Collins including Spring Creek and Fossil Creek would increase the ability to predict selenium concentrations and gain a better understanding of sources and transport mechanisms with respect to selenium contamination.

Recommendations for future studies include:

1. A second Se prediction model developed and applied with the same methods present herein with the addition of the Box Elder Creek watershed near Fort Collins. The watershed size at the confluence with the Cache la Poudre River is a logical choice as an additional basin for model validation.
2. A larger number of selenium observation locations to increase the statistical strength when predicting selenium concentrations. This would involve an increase in selenium sampling locations, as well as the delineation of additional watersheds and subsequently additional geospatial variable values.
3. Future modeling efforts can be built out of the process outlined in this study for development of CHRUs. Selenium sampling from much smaller basin areas can be used for model development designed to predict selenium concentrations in upper watershed basins which are typically unmonitored.

Application to Regulatory Environment

Analysis of landscape characteristics from this study can provide regulatory agencies with watershed characterization data for selenium in Spring Creek and Fossil Creek watersheds'. Correlation and regression analyses results point to the landscape elements of shale, streams, wetlands, and soil pH as important selenium response variables. Watershed assessment studies regarding these suspected selenium sources and transport mechanisms could eventually lead to a depiction of the hydrologic and ecologic structure and function, which could allow the quantification of the role of selenium reduction strategies in water quality management. This information will help in TMDL development, and overall protection of the stream and local community health. Isolation of high risk areas for selenium will reduce cost and time for agencies as well. The spatial processes utilized and resulting data could also be used in future construction of a geospatial model calibrated with existing observed water quality data that predicts Se concentrations in chemical hydrologic response units. This model could further be designed for application in unmonitored watersheds, from local areas to extreme locations difficult to reach in the field saving money and time for regulatory agencies.

LITERATURE CITED

Bende, Ulrike; Flugel, Wolfgang-Albert; Kern J., Tim. 1995. Using GIS to delineate Chemical Hydrological Response Units (CHRUs) for hydrochemical modeling in a mesoscale catchment in Germany. Effects of Scale on Interpretation and Management of Sediment and Water Quality (Proceedings of a Boulder Symposium, July 1995). IAHS Publ. no. 226.

Butler L., David. 2001. Effects of Piping Irrigation Laterals on Selenium and Salt Loads, Montrose Arroyo Basin, Western Colorado. U.S. Department of the Interior. U.S. Geological Survey. Water-Resources Investigations Report 01-4204. Denver, Colorado.

Butler L., David and Leib J., Kenneth. 2002. Characterization of Selenium in the Lower Gunnison River Basin, Colorado, 1988-2000. U.S. Department of the Interior. U.S. Geological Survey. Water-Resources Investigations Report 02-4151. Denver, Colorado.

Canton P., Steven and Van Deveer D., William. 1997. Selenium Toxicity to Aquatic Life: An Argument for Sediment-Based Water Quality Criteria. Environmental Toxicology and Chemistry, Vol. 16, No. 6, pp. 1255-1259.

CDPHE. Colorado Department of Public Health and Environment, 2009. Water Quality Control Division. Section 303(d) Listing Methodology, 2010 Listing Cycle.

CDPHE. Colorado Department of Public Health and Environment, 2010a. Water Quality Control Commission. Regulation #38 Stream Classifications and Water Quality Standards for South Platte River Basin, Laramie River Basin, Republican River Basin, and Smoky Hill River Basin. 5 CCR 1002-38.

CDPHE. Colorado Department of Public Health and Environment, 2010b. Water Quality Control Commission. Colorado Primary Drinking Water Regulations. 5 CCR 1003-1. Amended 08/09/2010. Effective 11/30/2010.

CDPHE. Colorado Department of Public Health and Environment, 2010c. Colorado's Section 303(d) List of Impaired Water and Monitoring and Evaluation List. 5 CCR 1002-93. Regulation #93.

CDPHE. Colorado Department of Public Health and Environment, 2010d. Water Quality Control Division. Integrated Water Quality Monitoring and Assessment Report. State of Colorado. 2010 Update to the 2008 305(b) Report.

CDPHE. Colorado Department of Public Health and Environment, 2012. Water Quality Control Commission. Regulation #31. The Basic Standards and Methodologies for Surface Water. 5 CCR 1002-31. Amended: June 13, 2011. Effective: January 1.

Chapman, P.M.; Adams, W.J.; Brooks, M.L.; Delos C.G.; Luoma, S.N.; maher, W.A.; Ohlendorf, H.M.; Presser, T.S.; Shaw, D.P. 2009. Ecological Assessment of Selenium in the Aquatic Environment. Summary of a SETAC Pellston Workshop. Pensacola FL (USA): Society of Environmental Toxicology and Chemistry (SETAC).

CWCB. Colorado Water Conservation Board. DWR. 2012. Colorado Division of Water Resources. CDSS. Colorado's Decision Support Systems. Date accessed: January 24, 2012. http://cdss.state.co.us/GIS/Pages/Division1SouthPlatte.aspx

Dissanayake, C.B.; Chandrajith, Rohana. 1999. Medical geochemistry of tropical environments. Earth-Science Reviews. Vol. 47, pg. 219 – 258.

Dorianne, E. Shivers; Moglen, E. Glenn. 2008. Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads. Journal of Environmental Engineering © ASCE. 134:8 (610).

Draper N., Richard; Smith, Harry. 1998. Applied Regression Analysis. In Wiley Series in Probability and Statistics. New York: Wiley.

Emerson, G. Douglas; Vecchia, V. Aldo; Dahl, L. Ann. 2005. Evaluation of Drainage-Area Method Used to Estimate Streamflow for the Red River of the North Basin, North Dakota and Minnesota. U.S. Department of the Interior. U.S. Geological Survey. Scientific Investigations Report 2005-5017. Reston, Virginia.

EPA. Environmental Protection Agency, 1987. Ambient Water Quality Criteria for Selenium – 1987. Office of Water Regulations and Standards Criteria and Standards Division. EPA – 440/5-87-006. September.

EPA. Environmental Protection Agency, 1995a. National Recommended Water Quality Criteria, 1995. http://water.epa.gov/scitech/swguidance/standards/current/index.cfm

EPA. Environmental Protection Agency, 2012b. Watershed Assessment, Tracking & Environmental Results. Colorado 303(d) Listed Waters for Reporting Year 2006. http://iaspub.epa.gov/tmdl_waters10/attains_impaired_waters.impaired_waters_list?p_state=CO&p_cycle=2006

ESRI. Environmental Systems Research Institute. Introduciton to Spatial Pattern Analysis. Training Seminar. 2011 http://training.esri.com/gateway/index.cfm?fa=catalog.webCourseDetail&courseID=2031

ESRI. Environmental Systems Research Institute. ArcGIS Resources. 2012. http://resources.arcgis.com/en/help/main/10.1/index.html#//002w00000001000000

ESRI. Environmental Systems Research Institute. ArcGIS Resources. ArcGIS Help 10.1. 2013. Ordinary Least Squares (OLS) (Spatial Statistics) http://resources.arcgis.com/en/help/main/10.1/index.html#/Ordinary_Least_Squares_OLS/005p00000022000000/

Fordyce, Fiona. 2005. Essentials of Medical Geology. Impacts of the Natural Environment on Public Health. Chapter 15, pg. 373. Selenium Deficiency and Toxicity in the Environment. Elsevier Academic Press.

Geering, Harold R.; Cary, Earle E.; Jones, L.H.P.; Allaway, W.H. 1968. Solubility and Redox Criteria for the Possible Forms of Selenium in Soils. Soil Science America Proceedings. Vol. 32.

Hammer, Mark J. 1981. An Assessment of Current Standards for Selenium in Drinking Water. Ground Water. Vol. 19, No. 4.

Hillegas, Robert. 2011. Personal Communication. Environmental Data Unit. Water Quality Control Division. Colorado Department of Public Health and Environment. January 19.

Kern J. Tim; Stednick D., John. 1993. Identification of heavy metal concentrations in surface waters through coupling of GIS and hydrochemical models. Hydro GIS 93: Application of Geographic Information Systems in Hydrology and Water Resources (Proceedings of the Vienna Conference, April 1993). IAHS Publ. no. 211.

Kleinbaum G., David; Kupper L., Lawrence; Muller E. Keith; Nizam, Azhar. 1998. Applied Regression Analysis and Other Multivariable Methods. Duxbury Press. Brooks/Cole Publishing Company. An International Thomson Publishing Company.

Kouwen, N.; Soulis, E.D.; Pietroniro, A.; Donald, J.; Harrington, R.A. 1993. Grouped Response Units for Distributed Hydrologic Modeling. Journal of Water Resources Planning and Management. Vol. 119, No. 3, May/June.

Leib J., Kenneth. 2008. Concentrations and Loads of Selenium in Selected Tributaries to the Colorado River in the Grand Valley, Western Colorado, 2004-2006. U.S. Department of the Interior. U.S. Geological Survey. Scientific Investigations Report 2008-5036. Reston, Virginia.

Leib J., Kennteth; Linard I., Joshua; Williams A., Cory. 2012. Statistical Relations of Salt and Selenium Loads to Geospatial Characteristics of Corresponding Subbasins of the Colorado and Gunnison Rivers in Colorado. U.S. Department of the Interior. U.S. Geological Survey. Scientific Investigations Report 2012 – 5003. Reston, Virginia.

Lemly A., Dennis. 1993. Guidelines for Evaluating Selenium Data from Aquatic Monitoring and Assessment Studies. Environmental Monitoring and Assessment. 28: 83-100.

Lemly, A. D. 1997. Ecosystem Recovery Following Selenium Contamination in a Freshwater Reservoir. Ecotoxicology. Environ. Saf. 36, 275-281.

Lemly A., Dennis. 1999. Selenium Transport and Bioaccumulation in Aquatic Ecosystems: A Proposal for Water Quality Criteria Based on Hydrological Units. Ecotoxicology and Environmental Safety 42, 150-156.

Lemly A., Dennis; Smith G., J. 1987. Aquatic Cycling of Selenium: Implications for Fish and Wildlife, Fish and Wildlife Leaflet 12. U.S. Fish and Wildlife Service, Washington D.C.

Luoma N., Samuel; Rainbow S., Philip. 2008. Metal Contamination in Aquatic Environments: Science and Lateral Management. Cambridge University Press. Ch. 13, pg. 327.

Massscheleyn, Patrick H.; Delaune, Ronald D.; Patrick Jr., William H. 1990. Transformations of Selenium As Affected by Sediment Oxidation – Reduction Potential and pH. Environmental Science Technology. Vol. 24, No.1.

May W., Thomas; Fairchild F., James; Petty D., Jim; Walther J., Walther; Lucero, Jeff; Delvaux, Mike; Manring, Jill; Armbruster, Mike. 2008. An Evaluation of Selenium Concentrations in Water, Sediment, Invertebrates, and fish from the Solomon River Basin. Environ Monit Assess, 137: 213-232.

Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Binger, R.L.; Harmel, R.D.; Veith, T.L. Model. 2007. Evaluation for Systematic Quantification of Accuracy in Watershed Simulations. Transaction of the ASABE. American Society of Agricultural and Biological Engineers. ISSN 0001 – 2351. Vol. 50(3): 885 – 900.

MRLC. Multi-Resolution Land Characteristics Consortium. National Land Cover Database (NLCD). Date Accessed: January 12, 2012. http://www.mrlc.gov/index.asp

Mueller, K., David; DeWeese, R., Lawrence; Garner, Jack, A.; Spruill, B., Timothy. 1991. Reconnaissance investigation of water quality, bottom sediment, and biota associated with irrigation drainage in the middle Arkansas River basin, Colorado and Kansas, 1988-1999. U.S. Geological Survey. Water-Resources Investigations Report 91-4060. Denver, Colorado.

Nash, J.E.; Sutcliffe, J.V. 1970. River Flow Forecasting Through Conceptual models. Part 1 – A Discussion of Principles. Journal of Hydrology, 10:282 – 290.

NCDC. National Climatic Data Center. National Oceanic and Atmospheric Administration. Date Accessed: 4-2-2013.
http://www.ncdc.noaa.gov/IPS/coop/coop.html?_finish=0.8805662320693092

Nordberg, Monica; Cherian M., George. Essentials of Medical Geology. 2005. Impacts of the Natural Environment on Public Health. Chapter 18. Biological Responses of Elements. Elsevier Academic Press.

Ocampo, J., Carlos; Sivapalan, Murugesu; Oldham, E., Carolyn. 2006. Field exploration of coupled hydrological and biochemical catchment responses and a unifying perceptual model. Advances in Water Resources, Vol. 29, pp. 161 – 180.

Ohlendorf M., Harry; Hoffman J., David; Saiki K., Michael; Aldrich W., Thomas. 1986. Embryonic Mortality and Abnormalities of Aquatic Birds: Apparent Impacts of Selenium from Irrigation Drainwater. The Science of the Total Environment, 52: 49-63.

Olson, Chris. 2012. Personal Communication. Civil Engineering Department. Colorado State University. June.

Orton T.G.; Rawlins B.G.; Lark R.M. 2009. Using Measurements close to a detection limit in a geostatistical case study to predict selenium in topsoil. Geoderma. Vol. 152, 269 – 282.

Phung Van, Khoa; Stednick D., John. 2012. Environmental Factors Influencing Streamwater Chemistry in Selected Watersheds of the United States of America. Water – Research & Development. ISSN: 2249-2003. Vol. 2. No. 1 & 2. 1-11. January-August.

Prasad Krishna M., Bala; Long, Wen; Zhang, Xinsheng; Wood J., Robert; Murtugudde, Raghu. 2011. Predicting dissolved oxygen in the Chesapeake Bay: applications and implications. Aquatic Sciences, 73:437 – 451.

Presser S., Theresa; Sylvester A., Marc; and Low H., Walton. 1994. Bioaccumulation of Selenium from Natural Geologic Sources in Western States and Its Potential Consequences. Environmental Management. Vol. 18, No. 13, pp. 423-436.

Rodda J.E., Harvey; Demuth, Siegfried; Shankar, Ude. 1999. The application of a GIS-based decision support system to predict nitrate leaching to groundwater in southern Germany. Journal of Hydrological Sciences, 44(2). April.

Rothwell, J.J.; Dise, N.B.; Taylor, K.G.; Allott, T.E.H.; Scholefield, P.; Davies, H.; Neal, C. 2010. Predicting river water quality across North West England using catchment characteristics. Journal of Hydrology. 395, 153 – 162.

See B., Randolph; Naftz L., David; Qualls L., Charles. 1992. GIS-Assisted Regression Analysis to Identify Sources of Selenium in Streams. Water Resources Bulletin. American Water Resources Association. Vol. 28, No. 2. April.

Seiler L., Ralph. 1995. Prediction of Areas Where Irrigation Drainage May Induce Selenium Contamination of Water. Journal of Environmental Quality. 24:973-979.

Seiler L., Ralph; Skorupa P., Joseph; and Peltz A., Lorri. 1999. Areas susceptible to irrigation-induced selenium contamination of water and biota in the western United States. U.S. Geological Survey Circular 1180.

Shamberger, J. Raymond. 1983. Biochemistry of Selenium. Volume 2. Plenum Press, New York.

Shivers E., Dorianne; Moglen E., Glenn. 2008. Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads. Journal of Environmental Engineering. ASCE. Vol. 134. No. 8. August 1.

Smart R.P.; Soulsby C.; Cresser M.S.; Wade A.J.; Townend J.; Billett M.F.; Langan S. 2001. Riparian zone influence on stream water chemistry at different spatial scales: a GIS-based modeling approach, an example for the Dee, NE Scotland. The Science of the Total Environment. Vol. 280, 173 – 193.

Sorenson M.B., Elsie; Cumbie M., Peter; Bauer L., Thomas; Bell S., James; Harlan W. Charles. 1984. Histopathological, Hematological, Condition-Factor, and Organ Weight Changes Associated with Selenium Accumulation in Fish from Belews Lake, North Carolina. Archives of Environmental Contamination and Toxicology. Vol. 13, 153-162.

Tong T.Y., Susanna; Chen, Wenli. 2002. Modeling the relationship between land use and surface water quality. Journal of Environmental Management. 66, 377-393.

United States Census 2010. http://2010.census.gov

USDA. United States Department of Agriculture. 1993. Soil Survey Manual. Soil Survey Division Staff. Soil Conservation Service. Handbook 18. CH 3, Selected Chemical Properties.

USDA. United States Department of Agriculture. NRCS. Natural Resources Conservation Service. SSURGO. Soil Survey Geographic Database. Web Soil Survey. Date accessed: November 18, 2010. http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm

USDA. United States Department of Agriculture. NRCS. Natural Resources Conservation Service. Geospatial Data Gateway. Date accessed: January 24, 2012. http://datagateway.nrcs.usda.gov/

USGS. United States Geological Survey. Helsel R., Dennis; Hirsch M. Robert. 1991. Statistical Methods in Water Resources. Chapter A3. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation. Chapter 11. Multiple Linear Regression. Pg. 296. June.

USGS. United States Geological Survey. NHD. National Hydrography Dataset Data. Date accessed: January 22, 2010. http://nhd.usgs.gov/data.html

USGS. Unites States Geological Survey, 2012a. StreamStats. U.S. Department of the Interior. Date accessed: January 9, 2012. http://streamstats.usgs.gov/

USGS. United States Geological Survey, 2012b. Seamless Data Warehouse. The National Map Viewer. Date accessed: January 24, 2012. http://nationalmap.gov/viewer.html

USGS. United Stated Geological Survey, 2012c. The National Landcover Database. Fact Sheet 2012 – 3020. February 2012.

Vigiak, Olga; van Dijck, J. E., Simone; van Loon, Emiel, E.; Stroosnijder, Leo. 2006. Matching Hydrologic Response to Measured Effective Hydraulic Conductivity. Hydrologic Processes, Vol. 20, pp. 487 – 504.

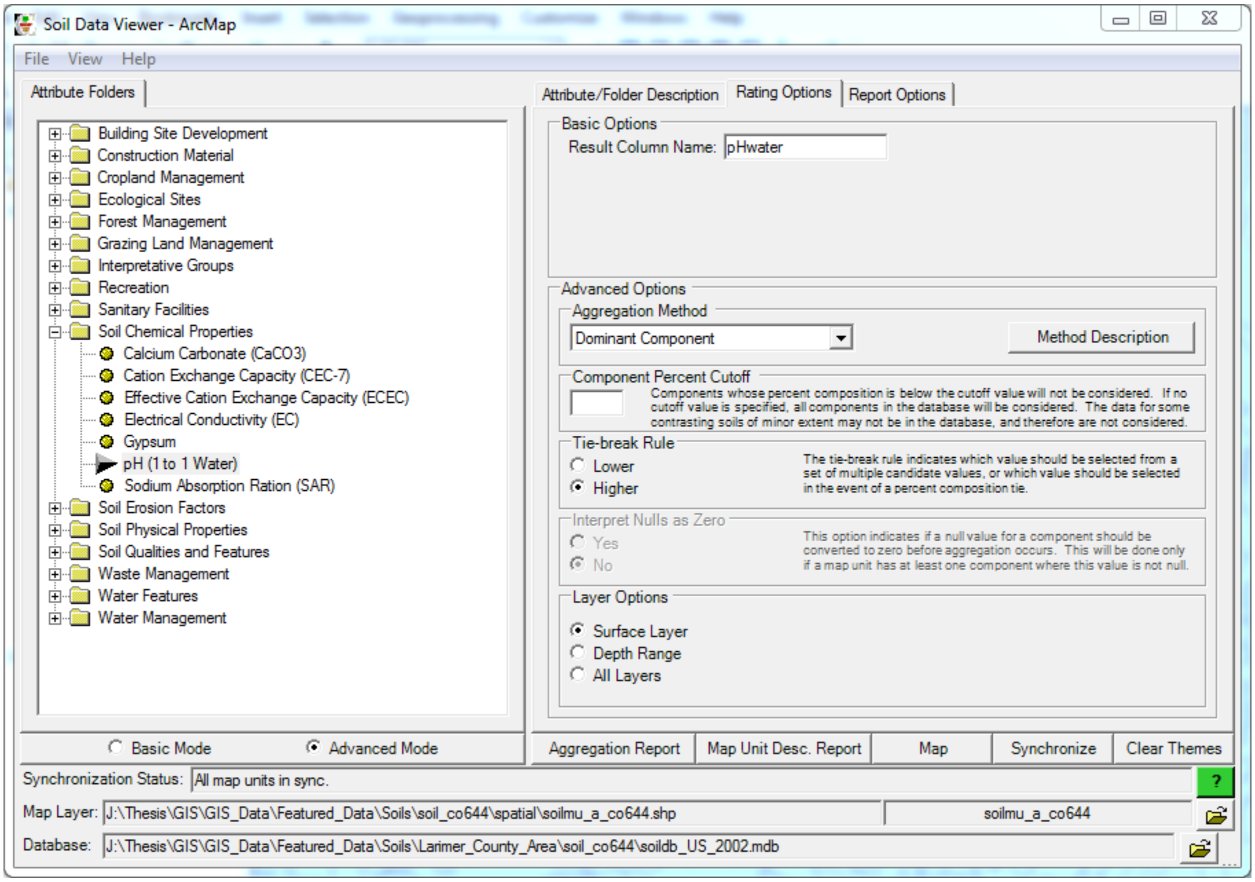
WRCC. Western Regional Climate Center, 2012. www.wrcc.dri.edu

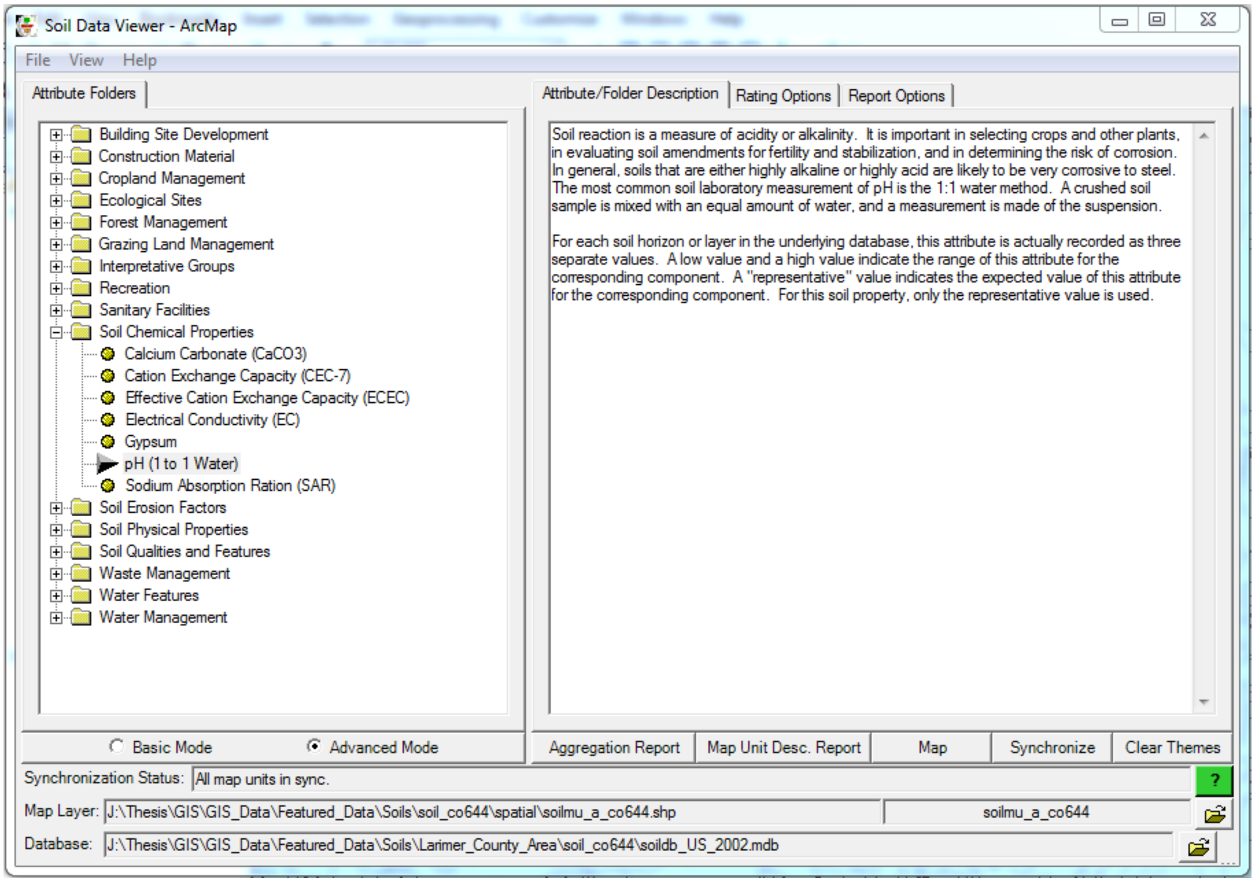
WHO. World Health Organization. 1996. Selenium in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality. Guidelines for drinking-water quality, 2nd ed. Vol. 2. World Health Organization, Geneva.

Xiao, Huaguao; Ji, Wei. 2007. Relating landscape characteristics to non-point source pollution in mine waste-located watersheds using geospatial techniques. Journal of Environmental Management. Vol. 82, 111-119.

APPENDIX

APPENDIX A - Soil Data Viewer User Interface (pHwater)




Soil Data Viewer - ArcMap
File View Help
Attribute Folders
Building Site Development
Construction Material
Cropland Management
Ecological Sites
Forest Management
Grazing Land Management
Interpretative Groups
Recreation
Sanitary Facilities
Soil Chemical Properties
Calcium Carbonate (CaCO3)
Cation Exchange Capacity (CEC-7)
Effective Cation Exchange Capacity (ECEC)
Electrical Conductivity (EC)
Gypsum
pH (1 to 1 Water)
Sodium Absorption Ration (SAR)
Soil Erosion Factors
Soil Physical Properties
Soil Qualities and Features
Waste Management
Water Features
Water Management
Attribute/Folder Description
Rating Options
Report Options
Soil reaction is a measure of acidity or alkalinity. It is important in selecting crops and other plants, in evaluating soil amendments for fertility and stabilization, and in determining the risk of corrosion. In general, soils that are either highly alkaline or highly acid are likely to be very corrosive to steel. The most common soil laboratory measurement of pH is the 1:1 water method. A crushed soil sample is mixed with an equal amount of water, and a measurement is made of the suspension.
For each soil horizon or layer in the underlying database, this attribute is actually recorded as three separate values. A low value and a high value indicate the range of this attribute for the corresponding component. A "representative" value indicates the expected value of this attribute for the corresponding component. For this soil property, only the representative value is used.
Basic Mode
Advanced Mode
Aggregation Report
Map Unit Desc. Report
Map
Synchronize
Clear Themes
Synchronization Status:
All map units in sync.
Map Layer:
J:\Thesis\GIS\GIS_Data\Featured_Data\Soils\soil_co644\spatial\soilmu_a_co644.shp
soilmu_a_co644
Database:
J:\Thesis\GIS\GIS_Data\Featured_Data\Soils\Larimer_County_Area\soil_co644\soildb_US_2002.mdb

APPENDIX B - Aggredation Report

pH (1 to 1 Water)

Layer Option: Surface Layer
Aggregation Method: Dominant Component
Tie-break Rule: Higher
Interpret Nulls as Zero: No

Larimer County Area, Colorado
Survey Area Version and Date: 3 - 12/12/2005

Map symbol	Map unit name	Rating
1	Altvan loam, 0 to 3 percent slopes	7.0
2	Altvan loam, 3 to 9 percent slopes	7.0
3	Altvan-Satanta loams, 0 to 3 percent slopes	7.0
4	Altvan-Satanta loams, 3 to 9 percent slopes	7.0
5	Aquepts, loamy	
6	Aquepts, ponded	
7	Ascalon sandy loam, 0 to 3 percent slopes	7.2
8	Ascalon sandy loam, 3 to 5 percent slopes	7.2
9	Bainville-Epping silt loams, 5 to 20 percent slopes	7.2
10	Bainville-Keith complex, 2 to 9 percent slopes	7.2
11	Baller-Carnero complex, 9 to 35 percent slopes	7.0
12	Baller-Rock outcrop complex, 15 to 45 percent slopes	7.0
13	Blackwell clay loam, 0 to 5 percent slopes	6.5
14	Boyle gravelly sandy loam, 3 to 9 percent slopes	6.7
15	Boyle gravelly sandy loam, 9 to 30 percent slopes	6.7
16	Boyle-Ratake gravelly sandy loams, 1 to 9 percent slopes	6.7
17	Boyle-Ratake gravelly sandy loams, 9 to 25 percent slopes	6.7
18	Breece coarse sandy loam, 0 to 3 percent slopes	7.0
19	Breece coarse sandy loam, 3 to 9 percent slopes	7.0
20	Breece coarse sandy loam, 9 to 30 percent slopes	7.0
21	Carnero loam, 3 to 9 percent slopes	7.5
22	Caruso clay loam, 0 to 1 percent slope	7.9
23	Clergern fine sandy loam, 2 to 10 percent slopes	7.2
24	Connerton-Barnum complex, 0 to 3 percent slopes	8.2
25	Connerton-Barnum complex, 3 to 9 percent slopes	8.2
26	Cushman fine sandy loam, 0 to 3 percent slopes	7.2
27	Cushman fine sandy loam, 3 to 9 percent slopes	7.2
28	Driggs loam, 0 to 3 percent slopes	7.0
29	Driggs loam, 3 to 25 percent slopes	7.0
30	Elbeth-Moen loams, 5 to 30 percent slopes	6.2
31	Farnuf loam, 2 to 10 percent slopes	7.0
32	Farnuf-Boyle-Rock outcrop complex, 10 to 25 percent slopes	7.0
33	Fluvaquents, nearly level	7.2
34	Fort Collins loam, 0 to 1 percent slopes	7.2
35	Fort Collins loam, 1 to 3 percent slopes	7.2
36	Fort Collins loam, 3 to 5 percent slopes	7.2
37	Fort Collins loam, 5 to 9 percent slopes	7.2
38	Foxcreek loam, 0 to 3 percent slopes	7.2
39	Gapo clay loam, 0 to 5 percent slopes	7.9
40	Garrett loam, 0 to 1 percent slopes	7.2
41	Garrett loam, 1 to 3 percent slopes	7.2
42	Gravel pits	
43	Haploborolls-Rock outcrop complex, steep	6.7
44	Haplustolls, hilly	7.5
45	Haplustolls-Rock outcrop complex, steep	7.5
46	Harlan fine sandy loam, 1 to 3 percent slopes	7.2

pH (1 to 1 Water)

Layer Option: Surface Layer
Aggregation Method: Dominant Component
Tie-break Rule: Higher
Interpret Nulls as Zero: No

Larimer County Area, Colorado
Survey Area Version and Date: 3 - 12/12/2005

Map symbol	Map unit name	Rating
47	Harlan fine sandy loam, 3 to 9 percent slopes	7.2
48	Heldt clay loam, 0 to 3 percent slopes	8.5
49	Heldt clay loam, 3 to 6 percent slopes	8.5
50	Keith silty clay loam, 0 to 3 percent slopes	6.7
51	Kildor clay loam, 0 to 6 percent slopes	7.0
52	Kildor-Shale outcrop complex, 5 to 30 percent slopes	7.0
53	Kim loam, 1 to 3 percent slopes	7.9
54	Kim loam, 3 to 5 percent slopes	7.9
55	Kim loam, 5 to 9 percent slopes	7.9
56	Kim-Thedalund loams, 3 to 15 percent slopes	7.9
57	Kirtley loam, 3 to 9 percent slopes	7.2
58	Kirtley-Purner complex, 5 to 20 percent slopes	7.2
59	Laporte-Rock outcrop complex, 3 to 30 percent slopes	7.9
60	Larim gravelly sandy loam, 5 to 40 percent slopes	7.0
61	Larimer fine sandy loam, 1 to 3 percent slopes	7.5
62	Larimer-Stoneham complex, 3 to 10 percent slopes	7.5
63	Longmont clay, 0 to 3 percent slopes	8.5
64	Loveland clay loam, 0 to 1 percent slopes	8.5
65	Midway clay loam, 5 to 25 percent slopes	7.5
66	Minnequa silt loam, 3 to 9 percent slopes	7.9
67	Minnequa-Laporte complex, 3 to 15 percent slopes	7.9
68	Miracle sandy loam, 5 to 25 percent slopes	7.2
69	Naz sandy loam, 1 to 3 percent slopes	6.5
70	Naz sandy loam, 3 to 25 percent slopes	6.5
71	Nelson fine sandy loam, 3 to 9 percent slopes	8.2
72	Newfork sandy loam, 0 to 3 percent slope	7.0
73	Nunn clay loam, 0 to 1 percent slope	7.0
74	Nunn clay loam, 1 to 3 percent slopes	7.0
75	Nunn clay loam, 3 to 5 percent slopes	7.0
76	Nunn clay loam, wet, 1 to 3 percent slopes	7.2
77	Otero sandy loam, 0 to 3 percent slopes	7.9
78	Otero sandy loam, 3 to 5 percent slopes	7.9
79	Otero sandy loam, 5 to 9 percent slopes	7.9
80	Otero-Nelson sandy loams, 3 to 25 percent slopes	7.9
81	Paoli fine sandy loam, 0 to 1 percent slopes	7.2
82	Pendergrass-Rock outcrop complex, 15 to 25 percent slopes	7.2
83	Pinata-Rock outcrop complex, 15 to 45 percent slopes	7.0
84	Poudre fine sandy loam, 0 to 1 percent slope	7.9
85	Purner fine sandy loam, 1 to 9 percent slopes	8.2
86	Purner-Rock outcrop complex, 10 to 50 percent slopes	8.2
87	Ratake-Rock outcrop complex, 25 to 55 percent slopes	7.0
88	Redfeather sandy loam, 5 to 50 percent slopes	5.8
89	Renohill clay loam, 0 to 3 percent slopes	7.2
90	Renohill clay loam, 3 to 9 percent slopes	7.2
91	Renohill-Midway clay loams, 3 to 15 percent slopes	7.2
92	Riverwash	7.9

pH (1 to 1 Water)

Layer Option: Surface Layer
Aggregation Method: Dominant Component
Tie-break Rule: Higher
Interpret Nulls as Zero: No

Larimer County Area, Colorado
Survey Area Version and Date: 3 - 12/12/2005

Map symbol	Map unit name	Rating
93	Rock outcrop	
94	Satanta loam, 0 to 1 percent slopes	7.0
95	Satanta loam, 1 to 3 percent slopes	7.0
96	Satanta loam, 3 to 5 percent slopes	7.0
97	Satanta loam, gullied, 3 to 9 percent slopes	7.0
98	Satanta Variant clay loam, 0 to 3 percent slopes	7.9
99	Schofield-Redfeather-Rock outcrop complex, 5 to 25 percent slopes	6.3
100	Stoneham loam, 0 to 1 percent slopes	7.2
101	Stoneham loam, 1 to 3 percent slopes	7.2
102	Stoneham loam, 3 to 5 percent slopes	7.2
103	Stoneham loam, 5 to 9 percent slopes	7.2
104	Sunshine stony sandy loam, 5 to 15 percent slopes	6.7
105	Table Mountain loam, 0 to 1 percent slopes	6.7
106	Tassel sandy loam, 3 to 25 percent slopes	7.9
107	Thedalund loam, 0 to 3 percent slopes	7.9
108	Thedalund loam, 3 to 9 percent slopes	7.9
109	Thiel gravelly sandy loam, 5 to 25 percent slopes	7.0
110	Tine gravelly sandy loam, 0 to 3 percent slopes	7.0
111	Tine cobbly sandy loam, 15 to 40 percent slopes	7.0
112	Trag-Moen complex, 5 to 30 percent slopes	6.7
113	Ulm clay loam, 0 to 3 percent slopes	7.2
114	Ulm clay loam, 3 to 5 percent slopes	7.2
115	Weld silt loam, 0 to 3 percent slopes	7.2
116	Wetmore-Boyle-Moen complex, 5 to 40 percent slopes	6.2
117	Wetmore-Boyle-Rock outcrop complex, 5 to 60 percent slopes	6.2
118	Wiley silt loam, 1 to 3 percent slopes	7.9
119	Wiley silt loam, 3 to 5 percent slopes	7.9
120	Ascalon loam, 0 to 6 percent slopes	7.6
121	Boyle-Boyle, thin solum, gravelly loams, 3 to 6 percent slopes	7.0
122	Boyle-Rock outcrop-Cathedral complex, 5 to 45 percent slopes	7.0
123	Cathedral-Boyle complex, 10 to 30 percent slopes	7.0
124	Evanston loam, 0 to 6 percent slopes	7.2
125	Evanston-Ipson association, 3 to 20 percent slopes	7.2
126	Ipson-Evanston complex, 6 to 30 percent slopes	7.2
127	Ipson-Trimad complex, 15 to 45 percent slopes	7.2
128	Merden, cool-Kovich complex, 0 to 3 percent slopes	7.9
129	Poposhia-Trimad complex, 3 to 15 percent slopes	7.9
130	Redthayne-Tyzak-Rock outcrop complex, 15 to 45 percent slopes	7.6
131	Tieside, north slopes-Rock outcrop complex, 10 to 45 percent slopes	7.9
132	Trimad-Blazon, thin solum-Rock outcrop complex, 20 to 45 percent slopes	8.2
133	Tyzak-Tyzak, thin solum-Rock outcrop complex, 30 to 50 percent slopes	7.9
134	Weed loam, 0 to 6 percent slopes	7.6
135	Dam	
136	Water	
137	Playas	

pH (1 to 1 Water)

Rating Options

Attribute Name: pH (1 to 1 Water)

Soil reaction is a measure of acidity or alkalinity. It is important in selecting crops and other plants, in evaluating soil amendments for fertility and stabilization, and in determining the risk of corrosion. In general, soils that are either highly alkaline or highly acid are likely to be very corrosive to steel. The most common soil laboratory measurement of pH is the 1:1 water method. A crushed soil sample is mixed with an equal amount of water, and a measurement is made of the suspension.

For each soil horizon or layer in the underlying database, this attribute is actually recorded as three separate values. A low value and a high value indicate the range of this attribute for the corresponding component. A "representative" value indicates the expected value of this attribute for the corresponding component. For this soil property, only the representative value is used.

Layer Option: Surface Layer

Aggregation Method: Dominant Component

Aggregation is the process by which a set of component attribute values is reduced to a single value to represent the map unit as a whole.

A map unit is typically composed of one or more "components". A component is either some type of soil or some nonsoil entity, e.g., rock outcrop. The components in the map unit name represent the major soils within a map unit delineation. Minor components make up the balance of the map unit. Great differences in soil properties can occur between map unit components and within short distances. Minor components may be very different from the major components. Such differences could significantly affect use and management of the map unit. Minor components may or may not be documented in the database. The results of aggregation do not reflect the presence or absence of limitations of the components which are not listed in the database. An on-site investigation is required to identify the location of individual map unit components.

For each of a map unit's components, a corresponding percent composition is recorded. A percent composition of 60 indicates that the corresponding component typically makes up approximately 60% of the map unit. Percent composition is a critical factor in some, but not all, aggregation methods.

For the attribute being aggregated, the first step of the aggregation process is to derive one attribute value for each of a map unit's components. From this set of component attributes, the next step of the aggregation process derives a single value that represents the map unit as a whole. Once a single value for each map unit is derived, a thematic map for soil map units can be generated. Aggregation must be done because, on any soil map, map units are delineated but components are not.The aggregation method "Dominant Component" returns the attribute value associated with the component with the highest percent composition in the map unit. If more than one component shares the highest percent composition, the corresponding "tie-break" rule determines which value should be returned. The "tie-break" rule indicates whether the lower or higher attribute value should be returned in the case of a percent composition tie.

The result returned by this aggregation method may or may not represent the dominant condition throughout the map unit.

Tie-break Rule: Higher

The tie-break rule indicates which value should be selected from a set of multiple candidate values, or which value should be selected in the event of a percent composition tie.

Interpret Nulls as Zero: No

This option indicates that a null value for a component should be converted to zero before aggregation occurs. This will be done only if a map unit has at least one component where this value is not null.

Soil Quality Information Sheet

Soil Quality Indicators: pH

USDA Natural Resources Conservation Service January 1998

What is pH?

Soil pH is a measure of the acidity or alkalinity in the soil. It is also called soil reaction.

The most common classes of soil pH are:

Extremely acid	3.5 – 4.4
Very strongly acid	4.5 – 5.0
Strongly acid	5.1 – 5.5
Moderately acid	5.6 – 6.0
Slightly acid	6.1 – 6.5
Neutral	6.6 – 7.3
Slightly alkaline	7.4 – 7.8
Moderately alkaline	7.9 – 8.4
Strongly alkaline	8.5 – 9.0

What is the significance of pH?

Availability of Nutrients
Soil pH influences the solubility of nutrients. It also affects the activity of micro-organisms responsible for breaking down organic matter and most chemical transformations in the soil. Soil pH thus affects the availability of several plant nutrients.

A pH range of 6 to 7 is generally most favorable for plant growth because most plant nutrients are readily available in this range. However, some plants have soil pH requirements above or below this range.

Soils that have a pH below 5.5 generally have a low availability of calcium, magnesium, and phosphorus. At these low pH's, the solubility of aluminum, iron, and boron is high; and low for molybdenum.

At pH 7.8 or more, calcium and magnesium are abundant. Molybdenum is also available if it is present in the soil minerals. High pH soils may have an inadequate availability of iron, manganese, copper, zinc, and especially of phosphorus and boron.

Micro-organisms
Soil pH affects many micro-organisms. The type and population densities change with pH. A pH of 6.6 to 7.3 is favorable for microbial activities that contribute to the availability of nitrogen, sulfur, and phosphorus in soils.

Pesticide Interaction
Most pesticides are labeled for specific soil conditions. If soils have a pH outside the allowed range, the pesticides may become ineffective, changed to an undesirable form, or may not degrade as expected, which results in problems for the next crop period.

Mobility of heavy metals
Many heavy metals become more water soluble under acid conditions and can move downward with water through the soil, and in some cases move to aquifers, surface streams, or lakes.

Corrosivity
Soil pH is one of several properties used as a general indicator of soil corrosivity. Generally, soils that are either highly alkaline or highly acid are likely to be corrosive to steel. Soils that have pH of 5.5 or lower are likely to be highly corrosive to concrete.

What controls soil pH?

The acidity or alkalinity in soils have several different sources. In natural systems, the pH is affected by the mineralogy, climate, and weathering. Management of soils often alters the natural pH because of acid-forming nitrogen fertilizers, or removal of bases (potassium, calcium, and magnesium). Soils that have sulfur-forming minerals can produce very acid soil conditions when they are exposed to air. These conditions often occur in tidal flats or near recent mining activity where the soil is drained.

The pH of a soil should always be tested before making management decisions that depend on the soil pH.

How is pH measured?

A variety of kits and devices are available to determine the pH in the field. The methods include:

- dyes
- paper strips
- glass electrodes.

Soil pH can change during the year. It depends on temperature and moisture conditions, and can vary to as much as a whole pH unit during the growing season. Since pH is a measure of the hydrogen ion activity [H^+], many different chemical reactions can affect it. Temperature changes the chemical activity, so most measurements of pH include a temperature correction to a standard temperature of 25 degrees C (77°F). The soil pH generally is recorded as a range in values for the soil depth selected.

(Prepared by the National Soil Survey Center in cooperation with the Soil Quality Institute, NRCS, USDA, and the National Soil Tilth Laboratory, Agricultural Research Service, USDA).

How is soil pH modified?

A soil pH below about 5.6 is considered low for most crops. Generally, the ideal pH range is between 6.0 and 7.0. Liming is a common method to increase the pH. It involves adding finely ground limestone to the soil. The reaction rate for limestone increases when soil temperatures are warm and soil moisture is high. If the limestone is more finely ground, the reaction is faster.

The amount of limestone to apply depends on the amount of organic matter and clay as well as the pH. Fertility testing laboratories that have local experience make this determination.

A soil pH that is more than about 8.0 is considered high for most crops. Soils that have a pH in this range are often also calcareous.

Calcareous soils have a high content of calcium carbonate. The pH of these soils does not change until most of the calcium carbonate is removed. Acids that are added to the soil dissolve the carbonates and lower the soil pH. Treatments with acid generally are uneconomical for soils that have a content of calcium carbonate of more than about 5%. Because phosphorus, iron, copper, and zinc are less available to plants in calcareous soils, nutrient deficiencies are often apparent. Applications of these nutrients are commonly more efficient than trying to lower the pH.

When the soil pH is above 8.6, sodium often is present. These soils generally do not have gypsum or calcium carbonates, at least not in the affected soil horizons. Addition of gypsum followed by leaching using irrigation is a common reclamation practice. However, salts flushed into drainage water may contaminate downstream waters and soils.

The application of anhydrous ammonia as a nitrogen fertilizer contributes to lowering the soil pH. In some parts of the country, applications of ammonia lower the surface soil pH from ranges of 6.6 to 7.3 to below 5.6. This reduction can be easily overlooked in areas of no-till cropping unless the pH is measured in the upper 2 inches.

Chemical amendments that contain sulfur generally form an acid, which lowers the soil pH.

Visit our Web site:
http://soils.usda.gov

APPENDIX D - Land Use Land Cover Classification Key

NLCD_classes

URL for NLCD 2006 - http://www.mrlc.gov/index.asp

National Land Cover Data Classification System Key

NOTE - All classes may NOT be represented in a specific State dataset. The class number represents the digital value of the class in the dataset.

RGB values:

Red - The value is arbitrarily assigned by the display software package, unless defined by user. Standard user defined ramp for NLCD project is start color light gray, end color red.

Green - The value is arbitrarily assigned by the display software package, unless defined by user. Standard user defined ramp for NLCD project is start color light gray, end color red.

Blue - The value is arbitrarily assigned by the display software package, unless defined by user. Standard user defined ramp for NLCD project is start color light gray, end color red.

Class value - Class Name

Water
 11. Open Water
 12. Perennial Ice/Snow

Developed
 21. Developed, Open Space
 22. Developed, Low Intensity
 23. Developed, Medium Intensity
 24. Developed, High Intensity

Barren
 31. Barren Land (Rock/Sand/Clay)
 32. Unconsolidated Shore*

Vegetated; Natural Forested Upland
 41. Deciduous Forest
 42. Evergreen Forest

Vegetated; Natural Shrubland
 51. Dwarf Scrub
 52. Shrub/Scrub

Herbaceous Upland Natural/Seminatural Vegetation
 71. Grassland/Herbaceous
 72. Sedge/Herbaceous
 73. Lichens
 74. Moss

Herbaceous Planted/Cultivated
 81. Pasture/Hay
 82. Cultivated Crops

Wetlands
 90. Woody Wetlands
 91. Palustrine Forested Wetland*
 92. Palustrine Scrub/Shrub Wetland*
 93. Estuarine Forested Wetland*

94. Estuarine Scrub/Shrub Wetland*
95. Emergent Herbaceous Wetlands
96. Palustrine Emergent Wetland (Persistent)*
97. Estuarine Emergent Wetland*
98. Palustrine Aquatic Bed*
99. Estuarine Aquatic Bed*

*Coastal NLCD class only

Class Definitions of the National Land Cover Dataset:

11. Open Water—All areas of open water, generally with less than 25 percent cover of vegetation or soil.

12. Perennial Ice/Snow—All areas characterized by a perennial cover of ice and/or snow, generally greater than 25 percent of total cover.

21. Developed, Open Space—Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses.
Impervious surfaces account for less than 20 percent of total cover.
These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.

22. Developed, Low Intensity—Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for
20-49 percent of total cover. These areas most commonly include single-family housing units.

23. Developed, Medium Intensity—Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for
50-79 percent of the total cover. These areas most commonly include single-family housing units.

24. Developed, High Intensity—Includes highly developed areas where people reside or work in high numbers. Examples include apartment
complexes, row houses, and commercial/industrial. Impervious surfaces account for 80 to 100 percent of the total cover.

31. Barren Land (Rock/Sand/Clay)—Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes,strip mines, gravel pits, and other accumulations of earthen material.
Generally, vegetation accounts for less than 15 percent of total cover.

32. Unconsolidated Shore*—Unconsolidated material such as silt, sand, or gravel that is subject to inundation and redistribution due to the action of water.
Characterized by substrates lacking vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable.
Erosion and deposition by waves and currents produce a number of landforms representing this class.

41. Deciduous Forest—Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover.
More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.

42. Evergreen Forest—Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover.
More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.

43. Mixed Forest–Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover.
Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.

51. Dwarf Scrub–Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20 percent of
total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.

52. Shrub/Scrub–Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20 percent of total vegetation.
This class includes true shrubs, young trees in an early successional stage, or trees stunted from environmental conditions.

71. Grassland/Herbaceous–Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80 percent of total vegetation.
These areas are not subject to intensive management such as tilling, but can be utilized for grazing.

72. Sedge/Herbaceous–Alaska only areas dominated by sedges and forbs, generally greater than 80 percent of total vegetation. This type can
occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.

73. Lichens–Alaska only areas dominated by fruticose or foliose lichens generally greater than 80 percent of total vegetation.

74. Moss–Alaska only areas dominated by mosses, generally greater than 80 percent of total vegetation.

81. Pasture/Hay–Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops,
typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.

82. Cultivated Crops–Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards.
Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled.

90. Woody Wetlands–Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or
substrate is periodically saturated with or covered with water.

91. Palustrine Forested Wetland*–Includes all tidal and non-tidal wetlands dominated by woody vegetation greater than or equal to
5 meters in height and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent.

92. Palustrine Scrub/Shrub Wetland*–Includes all tidal and non-tidal wetlands dominated by woody vegetation less than 5 meters in
height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent.
The species present could be true shrubs, young trees and shrubs or trees that are small or stunted due to environmental conditions.

93. Estuarine Forested Wetland*–Includes all tidal wetlands dominated by woody vegetation greater than or equal to 5 meters in height,
and all such wetlands that occur in tidal areas in which salinity due to

ocean-derived salts is equal to or greater than 0.5 percent.Total vegetation coverage is greater than 20 percent.

94. Estuarine Scrub/Shrub Wetland*-Includes all tidal wetlands dominated by woody vegetation less than 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent. Total vegetation coverage is greater than 20 percent.

95. Emergent Herbaceous Wetlands—Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

96. Palustrine Emergent Wetland (Persistent)*-Includes all tidal and non-tidal wetlands dominated by persistent emergent vascular plants, emergent mosses or lichens, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Plants generally remain standing until the next growing season.

97. Estuarine Emergent Wetland*-Includes all tidal wetlands dominated by erect, rooted, herbaceous hydrophytes (excluding mosses and lichens) and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and that are present for most of the growing season in most years. Perennial plants usually dominate these wetlands.

98. Palustrine Aquatic Bed*-The Palustrine Aquatic Bed class includes tidal and nontidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is below 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, detached floating mats, and rooted vascular plant assemblages.

99. Estuarine Aquatic Bed*-Includes tidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, kelp beds, and rooted vascular plant assemblages.

*Coastal NLCD class only

FAQ

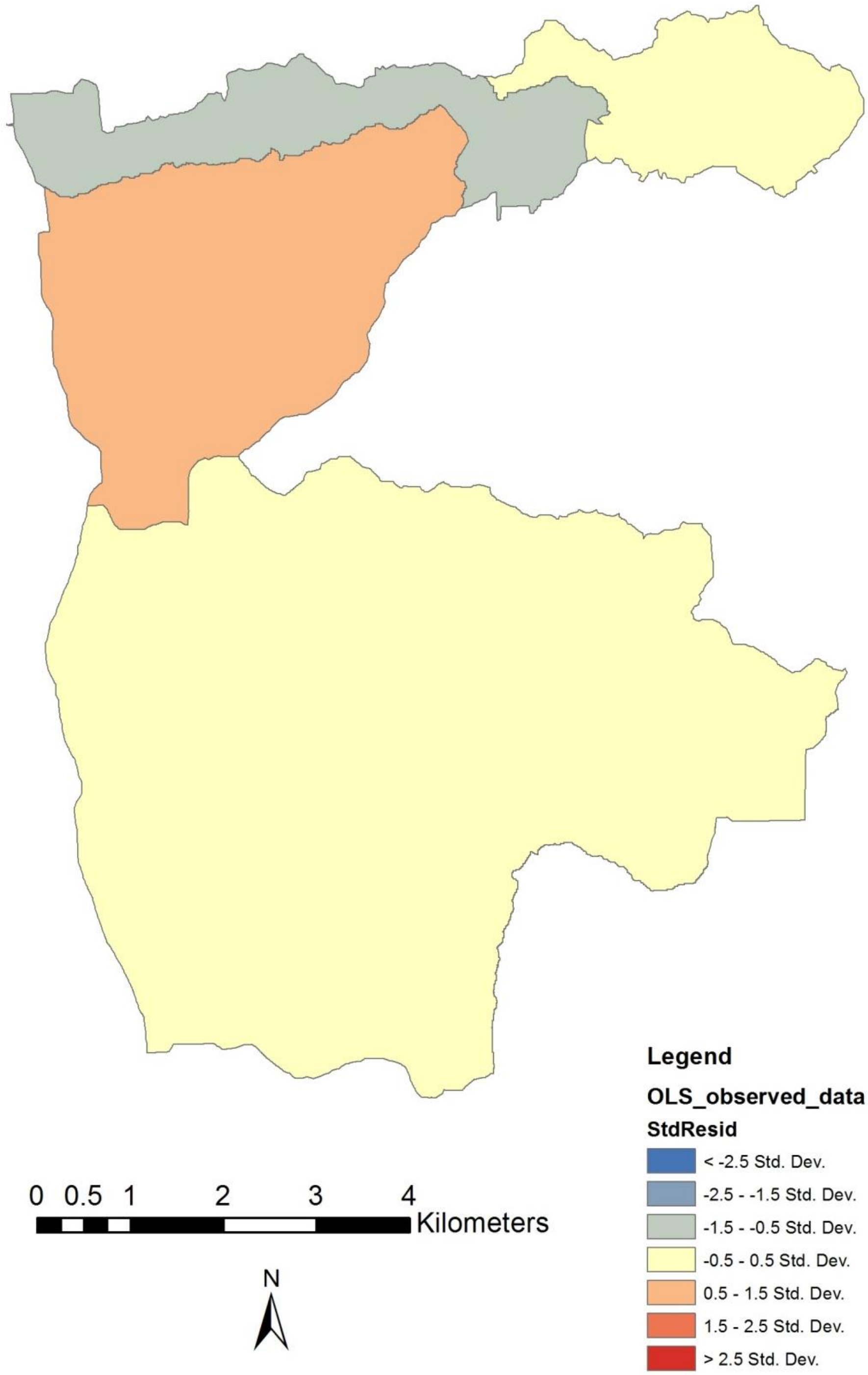
1.Symptom:

Showing details at a larger scale shows most of the details White. This usually happens when you are trying to show an entire state.

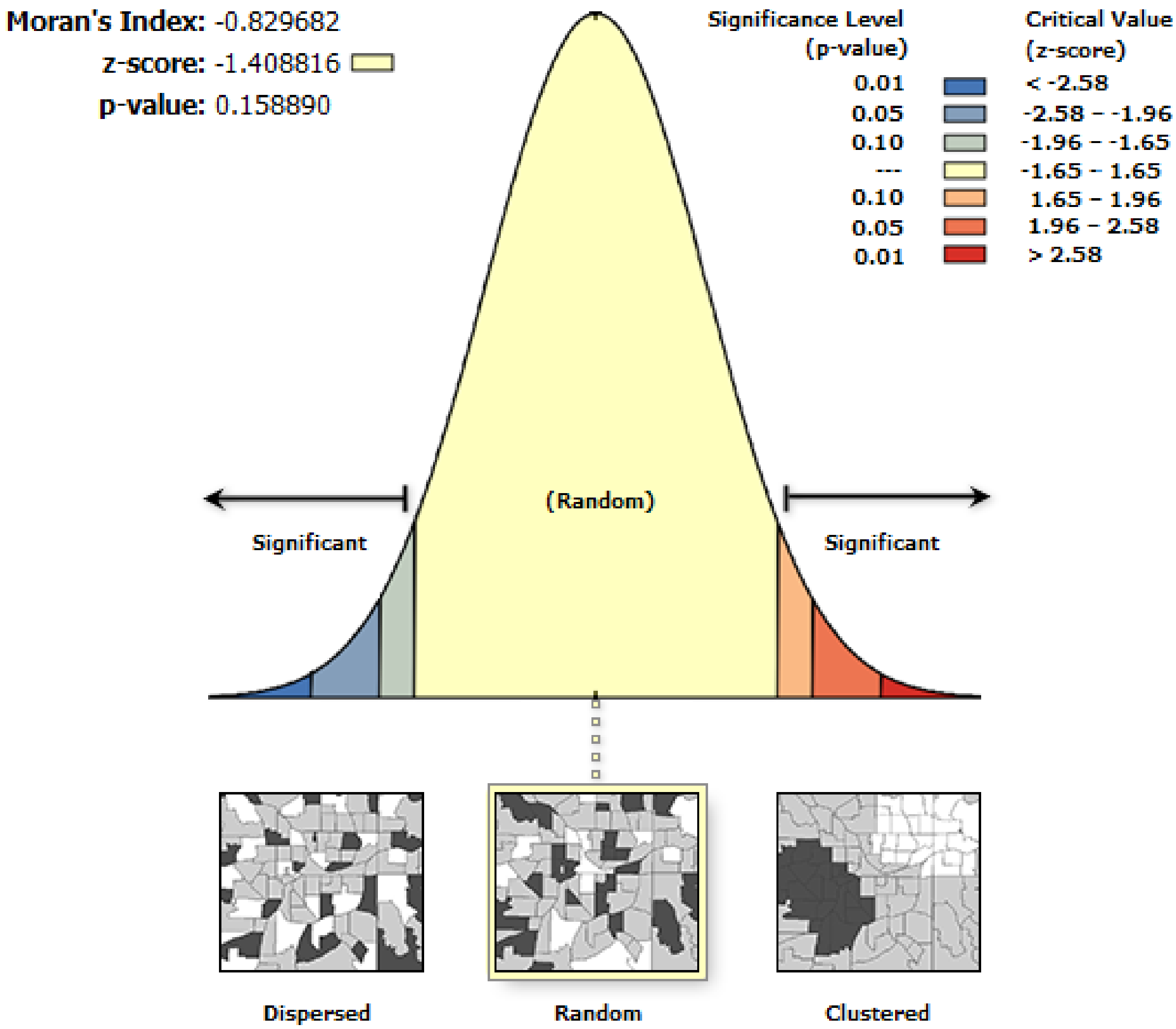
Solution:

If displaying data using ESRI (ArcMap) do not select the option create pyramids. In the Display menu in the Layer Properties, change the Resample option to "Majority (for discrete data).

APPENDIX E - OLS map output of regression residuals (over/under predictions)



APPENDIX F - Spatial Autocorrelation Report



APPENDIX G - Summary of OLS Results and OLS Diagnostics

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	-6.583797	1.208339	-5.448636	0.000027*	0.793889	-8.293093	0.000000*
WETLANDS	170.509264	13.295314	12.824764	0.000000*	6.373739	26.751841	0.000000*

OLS Diagnostics

Input Features:	ObservedData	Dependent Variable:	SELENIUM
Number of Observations:	5	Akaike's Information Criterion (AICc) [d]:	40.286874
Multiple R-Squared [d]:	0.982087	Adjusted R-Squared [d]:	0.976116
Joint F-Statistic [e]:	164.474568	Prob(>F), (1,3) degrees of freedom:	0.001023*
Joint Wald Statistic [e]:	715.660999	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	0.707427	Prob(>chi-squared), (1) degrees of freedom:	0.400299
Jarque-Bera Statistic [g]:	0.217338	Prob(>chi-squared), (2) degrees of freedom:	0.897027

APPENDIX H - Geographic Weighted Regression Results Output

```
Bandwidth           : 61063.55257188902
ResidualSquares     : 2.2891539092756861
EffectiveNumber     : 2.002507508216528
Sigma               : 0.87389337016953372
AICc                : 40.333395395056009
R2                  : 0.98210032234758926
R2Adjusted          : 0.97611379818101141
Succeeded at Thu May 29 09:30:15 2014 (Elapsed Time: 1.00 seconds)
```