DISSERTATION

# RELATIONSHIPS AMONG ENVIRONMENTAL FACTORS AND STREAM WATER ION YIELDS OF WATERSHEDS IN THE UNITED STATES

Submitted by:

Khoa Van Phung

Department of Geosciences

In partial fulfillment of the requirements
For the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Fall 2006

# COLORADO STATE UNIVERSITY

October 24, 2006

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY KHOA VAN PHUNG ENTITLED: "RELATIONSHIPS AMONG ENVIRONMENTAL FACTORS AND STREAM WATER ION YIELDS OF WATERSHEDS IN THE UNITED STATES" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

## Committee on Graduate Work

DENIS J. DEAN

JIM C. LOFTIS

ROBERT C. WARD

Adviser FREEMAN M. SMITH

Co-Adviser JOHN D. STEDNICK

Department Head/Director SALLY J. SUTTON

## Abstract of Dissertation

### RELATIONSHIPS AMONG ENVIRONMENTAL FACTORS AND STREAM WATER ION YIELDS OF WATERSHEDS IN THE UNITED STATES

Understanding relationships among environmental factors and water quality constituents is very important for ion yield prediction, water quality management, and pollution prevention strategies. In this study, 13 stream water constituents – total dissolved solids, dissolved oxygen, pH, total nitrogen, orthophosphate, phosphorus, calcium, magnesium, sodium, potassium, sulfate, chloride, and bicarbonate from 47 watersheds in the United States were analyzed in relation to the environmental factors, including bedrock types (limestone, sandstone, and crystalline), land cover types, average annual precipitation, population density, watershed slope, and watershed area. Two regression analyses were applied, multiple linear regression using the "stepwise" method of model selection and principal component regression (PCR) followed by a prediction and validation process. Geographic information system (GIS) techniques were used to delineate watersheds, handle data, and examine the spatial distribution of the predicted ion yields.

The results of this study indicated that there was a strong exponential relationship among environmental factors and selected stream water ion yields. The most important factor affecting stream water ion yields was average annual precipitation, which either appeared uniquely in the multiple regression models or appeared in company with other factors, accounting for 12 to 94 percent of the variance of most annual ion yields. Bedrock type was the second importance for stream water ion yields. Some average annual ion yields of limestone watersheds are 2 to 5 times larger than those of sandstone or crystalline watersheds, including total dissolved solids, bicarbonate, magnesium, and calcium. In addition, ion yields in limestone and sandstone watersheds are more closely related to the environmental factors than those in crystalline watersheds. Population density was the third most important factor, which accounted for 26 to 61 percent (in company with other factors) of the variance of ion yields of total dissolved solids (sandstone), sodium (limestone), magnesium (sandstone), potassium (limestone, sandstone), chloride (limestone), and bicarbonate (sandstone).

Land cover was less important. The effects of land cover were significant in company with other factors and accounted for 15 to 45 percent of the variances in the predictions of ion yields in limestone and sandstone watersheds, including orthophosphate, phosphorus, magnesium, and sulfate (limestone watersheds), total dissolved solids, sodium, and bicarbonate (sandstone watersheds). Watershed slope was an ineffective predictor, significant only in limestone watersheds and accounted (in company with other factors) for 12 to 42 percent of the variance of ion yields of sodium, potassium, magnesium, and dissolved oxygen. Watershed area was not significant in any ion yield regression analysis.

In this study, multiple linear regression (MLR) provided simpler models but larger R square, smaller standard error of estimate, and could detect more significant equations than principal component regression (PCR).

The results of this study can be helpful to understand effects and importance of environmental factors on water quality, to predict average annual stream water ion yields, especially for watersheds where water chemistry data are not available, and to evaluate the explanation of the predicted results contributed by each environmental factor. The methods of this study may also be useful for those who want to develop their own models to predict water ion yields.

KHOA VAN PHUNG
DEPARTMENT OF GEOSCIENCES
COLORADO STATE UNIVERSITY
FORT COLLINS, CO 80523
FALL 2006

# ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Dr. Freeman M. Smith and my co-advisor, Dr. John D. Stednick for their special sentiments, enthusiasm, academic advice, editing, and corrections for my dissertation.

My sincerest gratitude and appreciation is extended to the academic committee, Dr. Denis J. Dean, Dr. Freeman M. Smith, Dr. Jim C. Loftis, Dr. John D. Stednick, and Dr. Robert C. Ward for giving hints, recommendations, and spending a lot of time to serve during the course of this research.

I would like to send many special thanks to the Ministry of Education and Training of Vietnam and the Embassy of Vietnam in the United States, who provided financial and administrative support for my study in Colorado State University.

I am especially grateful to Dr. James R. zumBrunnen and Dr. Philip L. Chapman for their statistical consultations and comments through out my research.

I greatly thank the faculty and staff of the Department of Environmental Management, the Faculty of Environmental and Forest Resource Management, the Forestry University of Vietnam, and the Ministry of Agriculture and Rural Development for their administrative support for my study in Colorado State University and for their encouragement to my dissertation. My many thanks go to members of the Vietnamese Student Association in Colorado (VISACO) for their friendship during the time of this research.

Finally and ultimately, my deepest love and thanks are dedicated to my mother and my father, Tran Thi Lan and Phung Van Ba, to my lovely wife, Kieu Thi Duong, and to my cheerful son, Phung Kim Bang for their love, unwavering faith, and cheers. I never would have accomplished this research without their encouragement and mental energy supports.

*This research is dedicated to my mother and father, Tran Thi Lan and Phung Van Ba, to my lovely wife, Kieu Thi Duong, and to my cheerful son, Phung Kim Bang, who taught me the true values of happiness, love, joy, and the strength to overcome the obstacles of life.*

## Table of Contents

ABSTRACT .......... iii

ACKNOWLEDGEMENTS .......... vi

List of Tables .......... xi

List of Figures .......... xv

I. INTRODUCTION .......... 1

II. LITERATURE REVIEW .......... 4

HYPOTHESIS, OBJECTIVES .......... 16

III. METHODOLOGY .......... 17

1. Bedrock types, station number and station geo-coordinates .......... 17
2. Sources and periods of selected data .......... 20
3. Watershed/Basin delineation .......... 22
4. Data calculations for the stream water constituents .......... 23
5. GIS data preparation and calculations for the environmental factors .......... 24
6. Building regression models for estimating constituent yield .......... 29
7. Multicollinearity diagnostics .......... 31
8. Determining the most effective factors on constituent yields .......... 34
9. Examining the effects of bedrock types on ion yields .......... 35

10. Stream water ion yield prediction, model validation, and spatial distribution of the predicted ion yields ........ 37

IV. RESULTS AND DISCUSSION ........ 38

1. Descriptions of the project watersheds ........ 38

*1. 1. Watershed delineation and corresponding areas* ........ 38

*1. 2. Basin characteristics (environmental factors)* ........ 41

*1. 3. Stream water ion yields (Kg/ha/yr)* ........ 46

*1. 4. Kolomogorov-Smirnov Tests* ........ 50

2. Effects of bedrock type on stream ion yields ........ 51

3. Effects of the other environmental factors on stream water ion yields ........ 62

*3.1. Effects of land cover* ........ 62

*3.2. Effects of Population* ........ 67

*3.3. Effects of Precipitation* ........ 70

*3.4. Effects of Slope* ........ 73

*3.5. Effects of Area* ........ 75

4. Effects of principal components of the environmental variables on stream water ion yields ........ 75

*4.1. Limestone watersheds* ........ 75

*4.2. Sandstone watersheds* ........ 81

*4.3. Crystalline watersheds* .................................................. 87

5. Summary of the principal component regression analysis ........................ 92

6. Stream water ion yield prediction, model validation, and spatial distribution of the predicted ion yields. .................................................................. 94

V. CONCLUSIONS.................................................. 106

VI. RECOMMENDATIONS .............................................. 110

LITERATURE CITED .............................................. 112

APPENDICES.................................................... 117

# List of Tables

Table 1. Data sources for the study........................................................................ 20

Table 2. General information of the selected USGS gaged stream stations in the United States........................................................................ 21

Table 3. Land cover classification and corresponding area of each class for station number 06478500 - JAMES R NEAR SCOTLAND, SD (sandstone)........................ 26

Table 4. Recombinations of land cover classes into 4 categories for station number 06478500 - JAMES R NEAR SCOTLAND, SD (sandstone).................................... 27

Table 5. Average annual precipitation for basin station number 06478500 (sandstone).. 28

Table 6. Descriptive statistics of the delineated watershed areas and the area differences in Table 7a (in square mile, SQMI).................................................... 39

Table 7a. Differences between the delineated and USGS areas of the project watersheds........................................................................ 40

Table 7b. Nonparametric test for the differences between delineated watershed areas and USGS watershed areas........................................................................ 41

Table 8. Descriptive statistics of the 4 land cover categories of the project watersheds (in square mile, SQMI)........................................................................ 42

Table 9. Percentage of the 4 land cover categories of the project watersheds (%)......... 42

Table 10. Average population of each watershed (people/SQMI, 1990)...................... 43

Table 11. Average annual precipitation of each watershed (in/year, 1961-1990). 44

Table 12. Average slope of each watershed (degree)............................................ 45

Table 13a. Descriptive statistics of average annual stream water ion yields of each bedrock type (Kg/ha/year)........................................................................ 47

Table 13b. Average annual ion yields (Kg/ha/yr) in watersheds of each bedrock type from Peters's research in 1984 and the current research (complied for the corresponding watersheds only)........................................................................ 48

Table 13c. Nonparametric test for the differences between stream water ion yields of Peters's research (1984) and the current research........................................................ 48

Table 14. Kolomogorov-Smirnov Test for the stream ion yields combined from the three bedrocks (limestone, sandstone, and crystalline)........................................................ 50

Table 15. Kolomogorov-Smirnov Test for the environmental factors combined from the three bedrocks (limestone, sandstone, and crystalline)................................................ 51

Table 16. Nonparametric Test Statistics for the ion yield differences between two bedrock types of the three bedrock studied........................................................ 53

Table 17. Summary of the multiple linear regression results of the relationship between Log of ion yield and Log of the environmental factors in the three bedrocks................. 55

Table 18. Data for the test of effects of bedrocks on significant coefficient slope of *Precipitation* on pH of the three bedrock types........................................................ 56

Table 19. Model summary and the analysis of coefficients of the relationship between pH and precipitation in the combined data of the three bedrock types (except for Sandstone, all other data are in log-transformed)................................................ 57

Table 20. Data for the test of effects of bedrocks on significant coefficient slopes of precipitation on DO, pH, and on total nitrogen in sandstone and crystalline bedrock types........................................................................................ 58

Table 21. Model summary and the analysis of coefficients of the relationship between DO and precipitation in the combined data of sandstone and crystalline bedrock types.. 59

Table 22. Model summary and the analysis of coefficients of the relationship between pH and precipitation in the combined data of sandstone and crystalline bedrock types... 60

Table 23. Model summary and the analysis of coefficients of the relationship between total nitrogen and precipitation in the combined data of sandstone and crystalline bedrock types........................................................................................ 61

Table 24. The presence of significant slope coefficients of Water in the regression equations in limestone and sandstone bedrocks ($p < 0.05$)........................................ 62

Table 25. The presence of significant slope coefficients of Developed in the regression 64

equations in limestone bedrock (p < 0.05)........................................................

Table 26. The presence of significant slope coefficients of Forest in the regression equations in limestone bedrock (p < 0.05)........................................................ 65

Table 27. The presence of significant slope coefficients of Agriculture in the regression equations in limestone bedrock (p < 0.05)........................................................ 67

Table 28. The presence of significant slope coefficients of Population in the regression equations in limestone and sandstone bedrocks (p < 0.05)...................................... 68

Table 29. The presence of significant slope coefficients of Precipitation in the regression equations in limestone, sandstone, and crystalline bedrocks (p < 0.05)........ 71

Table 30. The presence of significant slope coefficients of Slope in the regression equations in limestone and crystalline bedrocks (p < 0.05)........................................ 74

Table 31. Extraction communities of the principal components of the log transforms of the 8 environmental variables in limestone watersheds............................................ 76

Table 32. Total variance explained for principal components in limestone watersheds. 76

Table 33. The Rotated Component Matrix of the log transforms of the 8 variables in limestone watersheds.................................................................................... 78

Table 34. Component Score Coefficient Matrix of the log transforms of the 8 variables in limestone watersheds.................................................................................... 78

Table 35. Regression Factor Scores for the first 2 principal components of the log transforms of the 8 variables in limestone watersheds............................................ 79

Table 36. Significant principal component regressions in limestone watersheds (p < 0.05).................................................................................................................. 80

Table 37. Extraction communities of the principal components of the log transforms of the 8 environmental variables in sandstone watersheds............................................ 81

Table 38. Total variance explained for principal components in sandstone watersheds.. 82

Table 39. The Rotated Component Matrix of the log transforms of the 8 variables in sandstone watersheds.................................................................................... 84

Table 40. Component Score Coefficient Matrix of the log transforms of the 8 variables in sandstone watersheds........ 84

Table 41. Regression Factor Scores for the first 2 principal components of the log transforms of the 8 variables in sandstone watersheds........ 85

Table 42. Principal component regressions in sandstone watersheds ($p < 0.05$) ......... 86

Table 43. Extraction communities of the principal components of the log transforms of the 8 environmental variables in crystalline watersheds........ 88

Table 44. Total variance explained for principal components in crystalline watersheds... 88

Table 45. The Rotated Component Matrix of the log transforms of the 8 variables in crystalline watersheds........ 89

Table 46. Component Score Coefficient Matrix of the log transforms of the 8 variables in crystalline watersheds........ 90

Table 47. Regression Factor Scores for the first 3 principal components of the log transforms of the 8 variables in crystalline watersheds........ 90

Table 48. Significant principal component regressions in crystalline watersheds ($p < 0.05$)........ 91

Table 49. Comparison of Multiple Regression and Principal Component Regression using Correlation and Standard Error of the Estimate for each bedrock type (more details are presented in Appendix 4, 5, 6)........ 93

Table 50. Selected equations for predictions and validations of stream water ion yields in sandstone watersheds........ 94

Table 51. Predicted Ion Yields Using Multiple Regression, Principal Component Regression, and GIS techniques in sandstone watersheds........ 95

Table 52. Average Percent Error (APE) for Predicted Ion Yields Using Multiple Regression, Principal Component Regression, and GIS techniques in sandstone watersheds........ 96

Table 53. Standard deviation for predictions of multiple regression and PCR using GIS layers........ 99

# List of Figures

---

Figure 1. Methodology used for the study........................................................... 18

Figure 2. Locations of the selected USGS gaged stream stations for drainage watersheds underlain by single bedrock types in the United States.......................... 19

Figure 3. A sandstone watershed delineated from National Elevation Dataset............... 38

Figure 4. Comparison of average annual sulfate yields (kg/ha/yr) between Peters 1984 and the current research in limestone watersheds.............................................. 49

Figure 5. Comparison of average annual potassium yields (kg/ha/yr) between Peters 1984 and the current research in crystalline watersheds........................................ 49

Figure 6. Bivariate plot of annual yield of sodium (kg/ha/yr) and Water land cover (SQMI) in watersheds underlain by sandstone bedrock in the United States............... 63

Figure 7. Relationship between annual yield of phosphorus (kg/ha/yr) and Developed land cover (SQMI) in limestone watersheds in the United States............................ 65

Figure 8. Effects of population density (people/SQMI) on annual yields of bicarbonate and potassium (kg/ha/yr) in sandstone watersheds in the United States.................... 69

Figure 9. Effects of population density (people/SQMI) on annual yields of chloride (kg/ha/yr) in limestone watersheds in the United States........................................ 70

Figure 10. Effects of average annual precipitation (in/yr) on annual median of pH in the drainage watersheds underlain by limestone and sandstone bedrock types in the United States.................................................................................................. 72

Figure 11. Effects of average annual precipitation (in/yr) on annual median of pH in the drainage watersheds underlain by crystalline bedrock types in the United States.......... 73

Figure 12. Annual average yield of dissolved oxygen in relation to slope degree of drainage watersheds underlain by limestone bedrock in the United States................. 74

Figure 13. Scree plot of the initial eigenvalue of each component in sandstone watersheds............................................................................................... 83

Figure 14. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 06485500 - Big Sioux R At Akron, IA............ 100

Figure 15. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 06810000 - Nishnabotna River above Hamburg, IA........................................................................................................ 102

Figure 16. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 08053500 - Denton Ck nr Justin, TX........... 104

# I. INTRODUCTION

"It is important to understand the difficulties in identifying causes and, in particular, sources of pollution in impaired waters. For many waters, states and other jurisdictions classify the causes and sources of impairment as unknown. U. S. Environmental Protection Agency (EPA) and states are working to develop methodologies for both determining the causes and sources of impairment and describing the level of confidence in the classification." (U. S. Environmental Protection Agency, 2000). The causes and/or the sources of pollution in impaired waters are not easy to be identified partly due to the fact that factors affecting water quality vary greatly across landscapes (Vose, 2006, as cited in USDA, 2006). In other words, the processes controlling solute concentrations in surface water are affected by many environmental factors, including bedrock formation, climate, topography, biota and soils, atmospheric deposition, hydrology, land use, time, and anthropogenic activities (Peters, 1984; Hem et al., 1990; Brown, 2000). It would be expected that properties and magnitude of the impacts of environmental factors on water quality are complicated and may change profoundly through time and space. While affecting water quality, environmental factors simultaneously affect each other. Water quality, therefore, is typically conceptualized as a reflection of the environmental factors occurring within a watershed (Stednick, 1991; Brown 2000).

For a long time, watershed scientists have attempted to find the "causes" of of stream water ion yield variation. For example:

*Why do water quality constituents differ considerably from this watershed to other* *watersheds? Or what factors account for the fact that some watersheds have a higher/lower rate of constituent concentrations than others through the time? And*

*What can water quality managers do in terms of Best Management Practices to reduce non-point source water pollution in their watersheds?*

To answer those questions, it is important to define and assess the effects of environmental factors on water quality. This work is necessary to set up inexpensive, effective but simple models for getting master plans for water pollution (especially for nonpoint source pollution) prevention strategies. Based on these research findings, water resource managers and environmental agencies can have a suitable time allocation, technologies (e.g. Best Management Practices), and enforcements to protect water resources properly. For example, a well known development strategy named "Low Impact Site Design" has been developed and implemented to minimize effects of environmental factors on water quality from stormwater based on a "set of "micromanagement", nonstructural approaches for a more attractive, multifunctional landscape when compared to the conventional strategies that just rely on a "pipe-and-pond" approach." (Massachusetts Low Impact Development Toolkit, 2006). In addition, to deal with nonpoint source pollution, the "Nonpoint Source Management Programs", created by the Water Quality Act of 1987 (PL 100-4) through section 319 did require State governments to submit State Assessment Reports identifying significant nonpoint sources of pollution to the state's navigable waters (U.S. G.P.O., 1988).

There have been numerous studies examining the relationship among specific aspects of water quality and environmental characteristics. However, "relatively few have included the "common" ions (e.g., calcium and magnesium) and/or the entire range of environmental characteristics occurring within the United States" (Brown, 2000).

This study will examine the relationship between environmental factors (including geology (bedrock type), climatology (precipitation), anthropogenic (population), and land cover type) and 13 constituent ion yields, including dissolved oxygen, orthophosphate, phosphorus, pH, total nitrogen, sodium, potassium, magnesium, calcium, chloride, sulfate, bicarbonate, and total dissolved solids. This study also comprehensively takes advantage of the current techniques of multiple regression, principal component regression, and GIS-based models to investigate alternatives for prediction and delineation of annual ion loads on a spatially distributed basis.

# II. LITERATURE REVIEW

Annual dissolved ion yields in streamwater are mainly derived from few sources, including bedrock and soil types of the drainage basin, atmospheric deposition, and substances derived from human activities (Hem, 1970, as cited in Peters, 1984). It is difficult to determine the influence of a specific source on a stream system due to the fact that the relationships between dissolved ion yields (distributed by specific sources) and the yields actually transported in streams are complex and variable (Peters, 1984).

In the United States, research projects aimed at assessing and defining the effects of environmental factors on water quality spatially and temporally have been conducted since the early 1970's (Brown, 2000). The most common examples of those projects are the National Water Summary Report like *Hydrologic Events and Stream Water Quality* (U.S. Geological Survey, 1993) and the National Water Quality Inventory (U.S. Environmental Protection Agency, 2000). The National Water Summary is a series of publications by the U.S. Geological Survey focused on increasing public understanding of the nature, quality, use, and trends of the nation's water resources. Each volume of the National Water Summary series includes general papers on the specific topic and state summaries. Having been prepared since 1975 under Section 305(b) of the Federal Clean Water Act, the *National Water Quality Inventory Report to Congress* is a biennial report to Congress and the public about the quality characteristics of our nation's water bodies. This becomes the primary vehicle for identifying and informing about water quality problems in rivers and streams, sources of pollutants and stressors impacting water quality in the United States (U.S. Environmental Protection Agency, 2000).

Of the most common sources of environmental effects, the bedrock type underlying a drainage basin is considered a major source of influences on the chemical composition of the stream water in that basin (Douglas, 1968, as cited in Peters 1984). Stream draining granite basins were reported to have dissolved solids concentration about 2 to 3 times greater than that in streams draining quartzite basins (Miller, 1961, as cited in Peters 1984). Plutonic, granite, or sandstone bedrock basins were reported to contain lower concentrations of dissolved solids than those of basins containing sedimentary, limestone or dolomite bedrocks (e.g., Johnson and Reynolds, 1977, as cited in Peters 1984). According to Hack (1960, as cited in Peters 1984), the lowest concentrations of solutes were found in streams draining quartzite basins and the highest were found in streams underlain by carbonate basins. In general, streams of noncalcareous and nonevaporitic basins drained the lowest dissolved solids concentration, while the highest dissolved solids concentrations were found in the sedimentary basins (Peters, 1984). Streams draining limestone bedrock typically produced ion yields that were approximately twice those of streams draining sandstone or crystalline bedrock (Brown, 2000).

The atmosphere is also a major source of ions in stream water (Peters, 1984). In basins underlain by bedrocks containing granites or quartzites, the ions found in stream water can be attributed mostly to atmospheric precipitation (Fisher, 1968, as cited in Peters 1984). In the Southwestern United States, high concentrations of sulfate in precipitation were reported due to the oxidation of $H_2S$ produced by decaying organic matter (Junge and Werby, 1958, as cited in Peters 1984) while in the Northeastern United States, the most important source of sulfates found in streams was attributed to the localized industrial areas (Fisher et al., 1968; Likens et al., 1977, as cited in Peters 1984).

The in-stream environmental factors also affect ion yield. For example, as water temperature increases, mineral solubility usually increases although the relationship is complicated by other effects. At higher temperatures, water in equilibrium contains less $CO_2$ than at lower temperatures. Cold water, therefore, may have more carbonic acid available for weathering than warm water (Peters, 1984). According to Sweeting (1966, as cited in Peters 1984), pH can decrease when stream temperature decreases, but carbonate mineral dissolution can increase. Streamflow also affects chemical concentrations. For example, most major stream solute concentrations were flow dependent, indicating a decreased concentration with increasing streamflow (Peters, 2004).

Forest land can play an important role and is among one of the most effective factors on water quality. Although the effects of forest land on runoff, stream flow, and erosion are both site- and species-specific (Calder, 2002), hundreds of research projects confirming the role of forest on improving water quality have been documented. For example, references and abstracts of more than 286 scientific articles particularly relevant to forest water quality were compiled in "A Forest Water Quality Literature Review" (Austin, 2006). Most of those articles were concerned with the effects of forest management and silviculture activities on water quality. In general, forest buffer strips are widely recommended for trapping nonpoint source pollutants and protecting surface water quality (Endreny, 2002). The presence of a forest or grass vegetative buffer predictably filtered between 40 and 95 percent of the runoff load where runoff carries sediment and/or particulate phosphorus (Haycock et al., 1997; Reed-Anderson et al., 2000, as cited in Endreny, 2002). “Even a Mile of Forest Makes a Difference in Water Quality”, was stated by researchers after conducting research for the USDA Forest Service Southern Research Station on the effects of forest on water quality. The research findings “showed a definite "cleaning" affect on the stream from passing through a mile of National Forest, with evidence of significant reductions in concentrations of chemicals such as nitrates, ammonium, and phosphorous” (Vose, J. and Clinton, B., 2006, as cited in USDA).

Human activities can profoundly affect stream water composition by contributing constituents directly through both point and non-point sources; indirectly by disturbing the soils and underlying rocks, and by releasing gases and smoke into the air (Peters, 1984). These elements then go back to the stream water system via atmospheric deposition and precipitation. The effects of human activities on the composition and loads of dissolved constituents in streams are the subject of many studies. Among those activities, population density and land use data have been shown to affect stream water quality in many cases (Peters, 1984). Population density was reported to be correlated well with all constituents transported in the river except phosphorus and silica (Ceasar et al., 1976, as cited in Peters 1984). Road salting activities have been considered a principal source of sodium and chloride in streams (Peters and Turk, 1981; Mattson and Godfrey, 1994). The level of potassium was typically increased by runoff from fertilized land (Berner and Berner, 1987, as cited in Brown, 2000). Total dissolved solids were reported to be affected by a number of human activities such as irrigation return flow (Engberg, 1983, as cited in Brown, 2000) and acid mine drainage (Herlihy et al., 1990, as cited in Brown, 2000). In their research (Day et al., 2002), the authors indicated that the degradation of water quality was more likely related to the increases in local populations and residential by-products rather than the increases in the associated infrastructure such as roads. The authors also suggested using the insights gained from their study to extrapolate the probable future impacts of growth on water quality for other watersheds that do not have the detailed data available.

The average yearly runoff of land covered by corn/soybean is about 4 times, and of commercial land in Central Indiana is about 24 times larger than that in forest (Frankenberger, undated). Therefore, the land use can affect water quality considerably. Land use was one of the most influential factors affecting nitrogen and phosphorous species in the Susquehanna River (Lystrom et al., 1978, as cited in Peters 1984). Disturbed land surface can cause considerably higher concentrations and loads of chemical constituents than undisturbed basins (Musser and Whetstone, 1964). According to Choi and Deal (2005), total nitrogen, total suspended particle, and total phosphorus in the year 2000 in five counties (Clinton, Jersey, Madison, Monroe, and St. Clair) in Illinois and five counties (St. Louis City, St. Louis, St. Charles, Jefferson, and Franklin) in Missouri were about 0.2 to 4.0 % in forest land, while those in agriculture land ranged from 55 to approximately 70%. Concentrations of sulphate increased with the loss of forest canopy by logging, and concentrations of nitrate increased due to increased nitrification of the forest floor caused by landscape excavation for drainage, reforestation, and liming (Havel, et al., 1999). Pollution from urban and agricultural land transported by precipitation and runoff was reported as the leading source of stream water impairment (U.S. Environmental Protection Agency, 2000). The increases of aquatic nutrient concentration have been related to the use of phosphorus detergents, urbanization, disposal of untreated human and animal wastes, and the leaching due to applications of nitrogen and phosphorus fertilizers (Peters and Donohue, 2001).

Today, one of the most influential factors affecting hydrology and water quality is attributed to human activity. Anthropogenic activities such as urbanization, transportation, farming, deforestation, and afforestation affect both water quality and quantity by altering not only water balance but also processes controlling water quality (Peters, 2000). In addition to human influences such as land use, soil type also affects water quality. Effects of soil on water quality have been documented in many studies. For example, total N and total P in tributaries of clay soils were lower than those in tributaries of siliceous sandy soils (Donohue, et al., 2001).

Topographic condition is also an important factor that indirectly affects water quality and ion yield via effect of it on runoff and soil erosion. As described in the "rational" erosion-estimating equation (Smith and Whitt, 1948, as cited in Renard et al., 1996) and in the empirical Universal Soil Loss Equation (USLE) (Renard et al., 1996), not only slope, but also the length of slope is taken into account. Size of watersheds can affect water quality and ion yields as well. The effect of watershed size is indirectly reflected via travel time; in turn, this affects the decay and decomposition processes of the contributing substances draining to the stream water system. The effects of watershed size on water quality can also be reflected indirectly via sediment yield and transport. However, conflicting results have been reported in previous studies about effects of subwatershed sizes on sediment yield (Arabi et al., 2004). For example, Bingner et al. (1997) and Jha et al. (2004) (as cited in Arabi et al., 2004) indicated that sediment yield at the outlet is very sensitive to the number and size of subwatersheds. However, FitzHugh and Mackay (2000) (as cited in Arabi et al., 2004) reported that observed sediment yield at the outlet was almost unaffected by the number and size of subwatersheds. Although slope and watershed size have not been extensively studied, they warrant attention because many ion concentrations are related to sediment and runoff transport.

One of the most comprehensive, national scale studies describing relationships among concentration or transport of a dissolved constituent in streams and various basin characteristics throughout the United States was conducted by Peters (1984). In this study, the relationships among seven major dissolved ions in streams – sodium, potassium, magnesium, calcium, chloride, sulfate, bicarbonate, and dissolved solids and four environmental factors including bedrock type, annual precipitation, population density, and stream temperature were examined by using linear and multiple linear regression analysis. Water ion yields were modeled as a linear function of the 4 environmental characteristics using ordinary least squares (OLS) regression. The general form of the linear regression equation was:

$$\mathbf{y_i = a + b_1x_1 + b_2x_2 + ... + b_nx_n}$$

where:

$y_i$ = the annual yield of constituent i,

$x_1$, $x_2$,..., $x_n$ = the bedrock type ($x_1$) and other basin characteristics.

In his study, Peters treated a parallel slope coefficient for a certain basin characteristic in different bedrocks when taking the class variable (dummy variable) into the multiple linear regression model to compare the effects of bedrock on ion yield. That assumption, however, may not be true. The author also combined the data from three bedrock types to do the regression analysis. Obviously, this approach is not as appropriate as doing regression analysis for each bedrock type separately.

Peters (1984) noted that, of the 4 factors analyzed, annual precipitation and the type of bedrock underlying a basin are the most important in determining the yields of dissolved major ions in streams. Population density was found to be effective in predicting only the yields of sodium, potassium, chloride, and sulfate in company with other variables. Stream water temperature had a negative effect on the yields of potassium, calcium, sulfate, bicarbonate, and dissolved solids. Peters suggested that the prediction of a constituent yield may be improved by determining the amount contributed from atmospheric deposition, and that stream water ion yield prediction may also be improved by taking more specific contributions from human activities, such as land use instead of population density in the multiple linear regression analysis. In addition, the prediction of ion yield can also be improved by developing multiple linear regression models for basins in each bedrock type separately (Peters, 1984).

### ***Summary of the previous methods used to predict water quality concentrations and ion yields:***

The various techniques used to identify the effects of environmental factors on water quality in large watersheds can be classified as one of two general approaches (Haith, 1976, as cited in Brown, 2000). The first approach is based on a particular set of environmental characteristics used to calculate the specified ion yields. In this approach, the study area is first partitioned into discrete areas of relatively homogenous conditions of interest (e.g., land use/cover). Water samples are then collected to define ion concentrations for each area of concern. Total ion yields are calculated by multiplying the mean concentration by the runoff volume. This approach is inexpensive and offers a relatively rapid means of prioritizing watersheds for water quality remediation plans (Adamus and Bergman, 1995; Brown, 2000). Nevertheless, there are some limitations involved in this approach. For example, this approach is considered unable to predict ion yields for combinations of the environmental characteristics which are not included in the sampled population (Brown, 2000).

The second approach is to model water quality loading based on a range of environmental characteristics. In this approach, the locations that integrate combinations of the environmental conditions are considered for the site selection (not relatively homogenous areas of interest) (Brown, 2000). Water samples collected at each site are then used to define ion concentrations. The total ion yield is calculated as average concentration multiplied by runoff volume. A function of the environmental characteristics occurring within a watershed is then built to model the prediction of ion yields in that watershed. This approach can bring back an advantage overwhelming the first approach, that is, we would be able to predict ion yields for combinations of environmental characteristics which are not included in the sampled population (Brown, 2000).

Traditionally, the ordinary least squares (OLS) regression was used to analyze the relationship between dependent and independent variables. However, in many situations, data collected do not satisfy OLS regression assumptions. To deal with those cases, alternative regression methods to OSL have been developed and they can provide greater predictive efficiency and comprehension of structure when data do not satisfy OLS regression assumptions (Efron and Tibshirani, 1991; Brown, 2000). Multiple linear regression analyses have been used in many studies to predict water concentration yield (Lystrom et al., 1978; Peters, 1984). Brown (2000) used a regression tree model to analyze the relationship between environmental factors and ion yields. The author noted that analogous to the OLS approach, regression tree models revealed the linear nature, non-constant variance, and the importance of variables observed in relationships between ion yields and environmental factors.

The uses of linear and multiple linear regression analysis are among aggregation approaches. These approaches can calculate predicted discrete single values of the water concentration of interest. However, using these approaches, we do not know the spatial variation of the project concentrations and ion yields.

Recent advances in geographic information systems (GIS) and remote sensing technologies have enabled us to build more realistic relationships among environmental factors and water concentration. These techniques include disaggregating approaches that can bring in both temporal and spatial results, rather than just predicted discrete single values. This is important, especially when we want to predict water ion yields for large watersheds.

Some of the most comprehensive models developed for predicting spatial distribution of water concentrations are useful due to applications of GIS techniques (Saunders, 1996; Dartiguenave, 1997; Wong et al., 1997; Naranjo, 1997; Quenzer, 1998;

Melancon, 1999; Osborne, 2000; Khalil, 2002). Although these GIS models were developed from different data and places, they were done in a similar way, that is, the loading of a given project constituent is a function of runoff and estimated or expected mean concentration (EMC) of that constituent. The expected mean concentrations for each predefined land use/land cover and corresponding runoff coefficient are derived from previous studies. Runoff coefficients can be also calculated using the Curve Number (CN) method (e.g. Khalil, 2002). Then the model uses a grid laid over the landscape, accounting for the constituent loadings and runoff derived from each cell. By tracing the flow of water from cell-to-cell basis, the movement of a predefined constituent over the landscape and into the river system is simulated (e.g. Naranjo, 1997).

In general, as reported in those studies (e.g., in Saunders, 1996), most of them have either some or all of the following limitations. *First*, the transport of constituents is pre-assumed to be conservative with no material loss or decay/decomposition. In fact, that assumption might be contrary with reality. *Second*, only average annual assessments are considered, runoff flow and constituent/pollutant loads are pre-assumed to be steady state parameters from year to year and within any year. *Third,* constituent concentrations on the land surface are considered to be directly related to general land use categories and do not vary with more specific types of land uses (e.g., from event to event). *Fourth,* some constituent loads are estimated due to data limitation. *Fifth,* most of those studies used Expected Mean Concentration or Estimated Mean Concentration (EMC) derived from each pre-defined land use type to estimate total loadings for that type of land use. Since the available research findings of EMC are still very limited, the application of these models is somewhat limited. In addition, using runoff and EMC to estimate total loadings of ion yields is not easy to explain the contribution of environmental factors on the estimated results. Nevertheless, in spite of these assumptions and limitations, those GIS models still be useful to make the results easier to analyze in a visual format and can be used to represent both the observed and predicted concentration/ion yield data spatially in many cases.

# HYPOTHESIS

Water quality can be predicted from watershed characteristics including land cover, population, precipitation, topology, and geology in the United States.

# OBJECTIVES

(1) To develop and compare regression equations for annual ion yields measured at different watersheds located throughout the United States using water quality and environmental data available via the internet.

(2) To determine the importance (weight) of environmental factors on water quality constituent yields.

(3) To generate an alternative Principal Component Regression technique for a subset of watersheds to estimate annual water quality ion yields based on the environmental factors and compare the results from this technique with the results from the traditional multiple linear regression analysis approaches.

(4) To predict water quality and use the GIS techniques to display different areas affecting water quality in a subset of watersheds.

# III. METHODDOLOGY

The overall methodology used in this study (Figure 1) is described in detail in the following sections:

## 1. Bedrock types, station number and station geo-coordinates

Following an earlier study (Peters, 1984), 56 gaged stream stations in the conterminous United States and Hawaii having drainage basins underlain by limestone, sandstone, or crystalline bedrocks were selected for this study. However, due to the lack of available data, only 47 study areas were actually resampled (Table 2, Appendix 1). For each selected basin, the following data sets were complied:

(1) Annual yields of major dissolved ions, bicarbonate ($HCO_3$), calcium (Ca), chloride (Cl), dissolved oxygen (DO), total dissolved solids (TDS), magnesium (Mg), phosphate (P), potassium (K), orthophosphate ($PO_4$), sodium (Na), sulfate ($SO_4$), total nitrogen, and pH.

(2) Average annual precipitation; (3) Mean population density; (4) Average slope; (5) Total basin area.

(6) National Land Cover Dataset.

Watersheds were selected throughout the United States to provide a wide range in stream water constituents and environmental factors. The three rock types, limestone, sandstone, and crystalline, were chosen due to the reasons as the following (Peters, 1984): limestone is relatively abundant (constituting 20 percent by volume of the world's

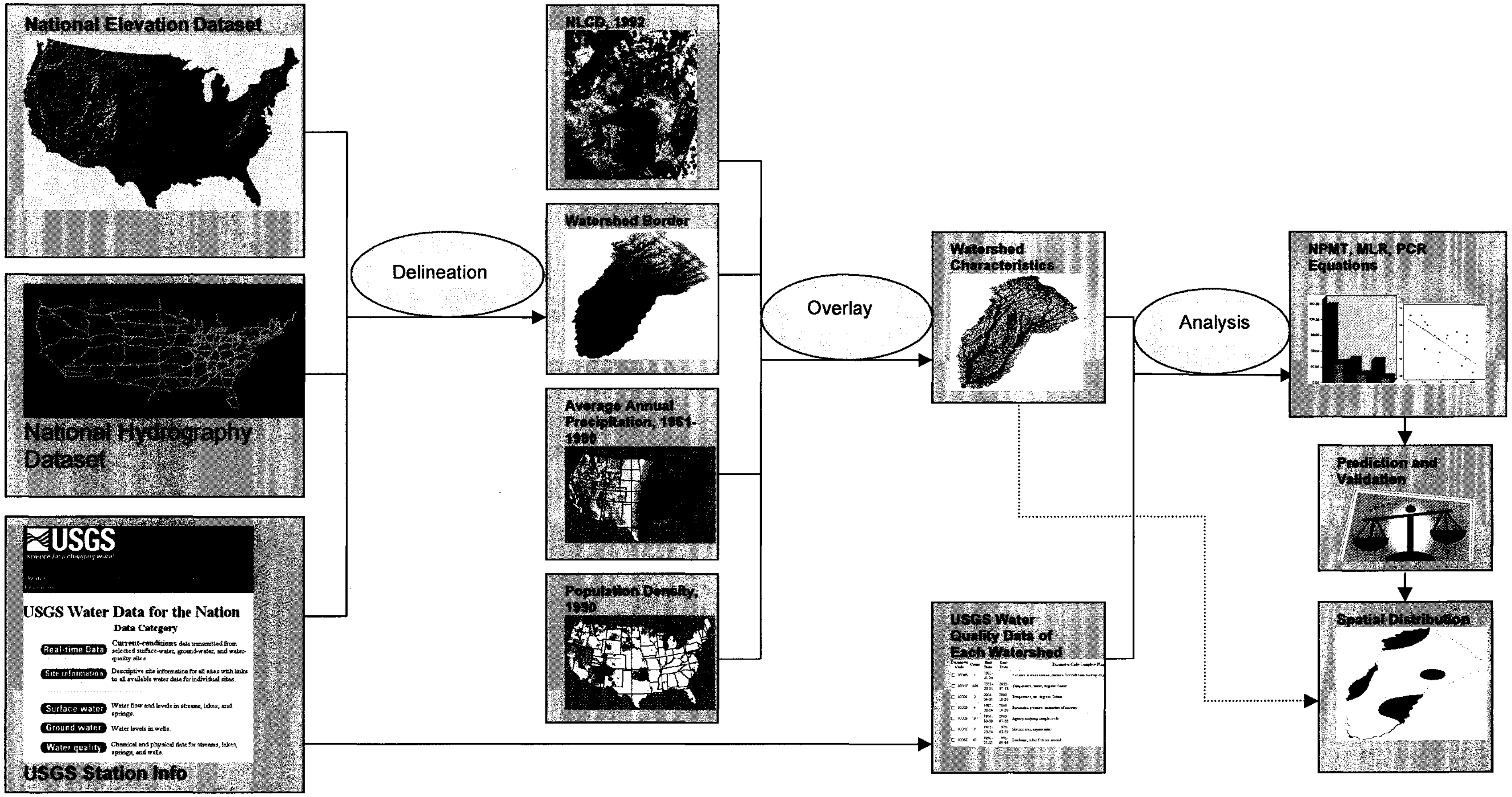


**Figure 1. Methodology used for the study.**

*(NLCD: National Land Cover Dataset; NPMT: Nonparametric test; MLR: Multiple linear regression; PCR: Principal component regression)*

sedimentary rocks) and it is composed primarily of easily weathered minerals of calcium and magnesium; sandstone consists of silicate minerals, which are more resistant to weathering than carbonates; crystalline rock constitutes 75 percent by volume of the Earth's crust and occupies the widest range in mineral texture and composition (crystalline was chosen for contrast with the other two types). Of the 47 basins selected, 19 were limestone, 11 were sandstone, and 17 were crystalline. Locations of USGS selected gage stations and rock types of the drainage basin are presented in Figure 2.

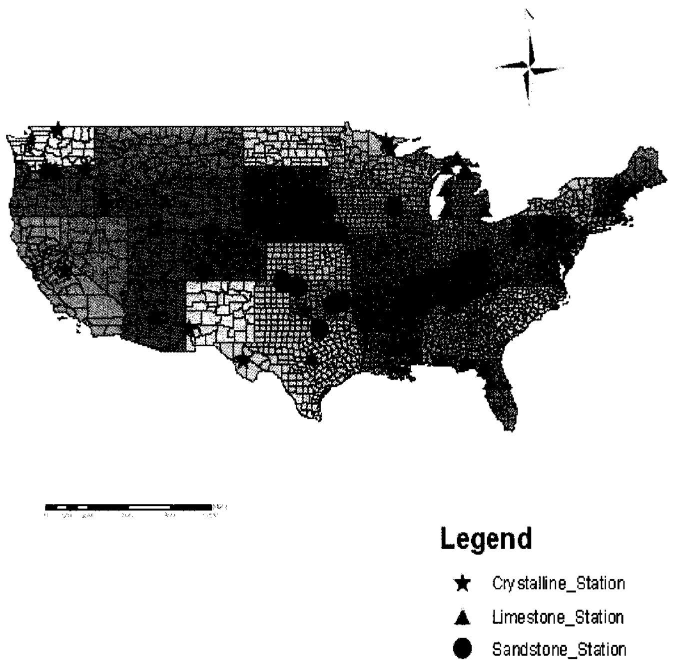


**Figure 2. Locations of the selected USGS gaged stream stations for drainage watersheds underlain by single bedrock types in the United States.**

Since there were many large basins containing a mixture of rock types, the conceptual determination of each bedrock type was based on Peters (1984) rationales (1) if a basin included more than 50 percent of the bedrock containing carbonate, then it was designated as limestone basin; (2) if the carbonate minerals were virtually absent, the basin was designated as sandstone; (3) a basin would be designated as crystalline if it contained more than 80 percent metamorphic or igneous rock.

The selected stations and corresponding complied data representing their basin characteristics (latitude, longitude, HUC, and drainage area in square miles) are summarized in Table 2 and in Appendix 1.

## 2. Sources and periods of selected data

Based on available data accessible via the internet, and to make the study consistent and compatible among the data types, the data were designed to compile as close to the same source and the same period as possible (Table 1). For metadata information, please go to the corresponding website (Table 1 and in the list of Literature Cited). These data were originally collected for multiple purposes, and this study uses them as secondary data based on EPA's Information Quality Guidelines (IQG) (EPA, 2003).

**Table 1. Data sources for the study.**

| | **Basin stations And Bedrock Types** | **Basin and sub-basin border** | **National Elevation Dataset** | **Land Cover** | **Average Annual Precipitation (inch)** | **Population Density (People/square mile)** | **Water Quality Constituents** |
|---|---|---|---|---|---|---|---|
| **Data Source** | Peters, N., 1984<br>U.S. Geological Survey | National Hydrography Dataset | U.S. Geological Survey (1 Arc Second) | U.S. Geological Survey | U.S. Geological Survey | U.S. Geological Survey<br>National Atlas of the United States | U.S. Geological Survey |
| **Time/ Periods** | ............... | .......... | ......... | 1992 | 1961-1990 | 1990 | 1980-1990* |
| **Internet Address** | http://waterdata.usgs.gov/nwis | http://nhd.usgs.gov/ | http://seamless.usgs.gov/website/seamless/viewer. Php<br>http://www.mrlc.gov | http://seamless.usgs.gov/website/seamless/ viewer.php<br>http://www.mrlc.gov | http://datagateway.nrcs.usda.gov/GatewayHome.html | http://www.nationalatlas.gov | http://water.usgs.gov/waterwatch/<br>http://waterdata.usgs.gov/nwis |

*(*: See Appendix 2 for more details).*

## Table 2. General information of the selected USGS gaged stream stations in the United States.

| Site No. | Station No. | Station name | Lat (NAD27) | Long (NAD27) | HUC | Drainage Area (SQMI) |
|---|---|---|---|---|---|---|
| | | **Limestone Bedrock** | | | | |
| 1 | 02313000 | WITHLACOOCHEE RIVER NR HOLDER, FLA. | 28°59'19" | 82°20'59" | 03100208 | 1,820.00 |
| 2 | 02320500 | SUWANNEE RIVER AT BRANFORD, FLA. | 29°57'20" | 82°55'40" | 03110205 | 7,880.00 |
| 3 | 02329000 | OCHLOCKONEE RIVER NR HAVANA, FLA. | 30°33'14" | 84°23'03" | 03120003 | 1,140.00 |
| 4 | 02368000 | YELLOW RIVER AT MILLIGAN, FLA. | 30°45'10" | 86°37'45" | 03140103 | 624 |
| 5 | 03085000 | Monongahela River at Braddock, PA | 40°23'28" | 79°51'30" | 05020005 | 7,337.00 |
| 6 | 03254000 | LICKING RIVER AT BUTLER, KY | 38°47'22" | 84°21'05" | 05100101 | 3,385.00 |
| 7 | 03290500 | KENTUCKY RIVER AT LOCK 2 AT LOCKPORT, KY | 38°26'20" | 84°57'48" | 05100205 | 6,180.00 |
| 8 | 03425000 | CUMBERLAND RIVER AT CARTHAGE, TN | 36°14'53" | 85°57'19" | 05130201 | 10,690.00 |
| 9 | 03543005 | TENNESSEE RIVER AT WATTS BAR DAM (TAILWATER), TN | 35°37'13" | 84°47'00" | 06010201 | 17,310.00 |
| 10 | 03571850 | TENNESSEE RIVER AT SOUTH PITTSBURG, TN | 35°00'41" | 85°41'51" | 06030001 | 22,640.00 |
| 11 | 03593005 | TENNESSEE RIVER AT PICKWICK LANDING DAM (LL), TN | 35°03'54" | 88°15'08" | 06040001 | 32,820.00 |
| 12 | 04045500 | TAHQUAMENON RIVER NEAR PARADISE, MI | 46°34'30" | 85°16'10" | 04020202 | 790 |
| 13 | 04057004 | MANISTIQUE RIVER ABOVE MANISTIQUE, MI | 45°58'18" | 86°14'35" | 04060106 | 1,445.00 |
| 14 | 04108690 | KALAMAZOO RIVER AT SAUGATUCK, MI | 42°38'50" | 86°11'53" | 04050003 | 2,020.00 |
| 15 | 04126520 | MANISTEE R AT MANISTEE, MI | 44°15'02" | 86°19'09" | 04060103 | 1,928.00 |
| 16 | 04132052 | CHEBOYGAN R (POND) AT LINCOLN AVE AT CHEBOYGAN, MI | 45°38'02" | 84°28'52" | 04070004 | 1,500.00 |
| 17 | 04165500 | CLINTON RIVER AT MORAVIAN DRIVE AT MT. CLEMENS, MI | 42°35'45" | 82°54'32" | 04090003 | 734 |
| 18 | 08103900 | S Fk Rocky Ck nr Briggs, TX | 30°54'41" | 98°02'12" | 12070203 | 33.3 |
| 19 | 13018300 | CACHE CREEK NEAR JACKSON, WY | 43°27'08" | 110°42'12" | 17040103 | 10.6 |
| | | **Sandstone Bedrock** | | | | |
| 1 | 01545600 | Young Womans Creek near Renovo, PA | 41°23'22" | 77°41'28" | 02050203 | 46.2 |
| 2 | 03049625 | Allegheny River at New Kensington, PA | 40°33'52" | 79°46'22" | 05010009 | 11,546.00 |
| 3 | 03237280 | Upper Twin Creek at McGaw OH | 38°38'37" | 83°12'57" | 05090201 | 12.2 |
| 4 | 06452000 | WHITE R NEAR OACOMA,SD | 43°44'54" | 99°33'22" | 10140204 | 10,200.00 |
| 5 | 06478500 | JAMES R NEAR SCOTLAND,SD | 43°11'09" | 97°38'07" | 10160011 | 20,653 |
| 6 | 07026000 | OBION RIVER AT OBION, TN | 36°15'04" | 89°11'33" | 08010202 | 1,852.00 |
| 7 | 07231500 | Canadian River at Calvin, OK | 34°58'40" | 96°14'36" | 11090202 | 27,952 |
| 8 | 07234000 | Beaver River at Beaver, OK | 36°49'20" | 100°31'08" | 11100102 | 7,955 |
| 9 | 07237500 | North Canadian River at Woodward, OK | 36°26'12" | 99°16'41" | 11100301 | 11,589 |
| 10 | 07245000 | Canadian River near Whitefield, OK | 35°15'50" | 95°14'21" | 11090204 | 47,576 |
| 11 | 06438000 | BELLE FOURCHE RIVER NEAR ELM SPRINGS,SD | 44°22'11" | 102°33'56" | 10120202 | 7,210 |
| | | **Crystalline Bedrock** | | | | |
| 1 | 01059400 | ANDROSCOGGIN RIVER AT BRUNSWICK, ME | 43°55'03" | 69°58'25" | 01040002 | 3,434.00 |
| 2 | 01066000 | Saco River at Cornish, Maine | 43°48'28.94" | 70°46'53.54" | 01060002 | 1,293.00 |
| 3 | 04014500 | BAPTISM RIVER NEAR BEAVER BAY, MN | 47°20'15" | 91°12'02" | 04010101 | 140 |
| 4 | 05124480 | KAWISHIWI RIVER NEAR ELY, MN | 47°55'22" | 91°32'06" | 09030001 | 254 |
| 5 | 06623800 | ENCAMPMENT RIVER AB HOG PARK CR, NR ENCAMPMENT, WY | 41°01'25" | 106°49'27" | 10180002 | 72.7 |
| 6 | 07311200 | Blue Beaver Creek near Cache, OK | 34°37'24" | 98°33'48" | 11130203 | 24.6 |
| 7 | 08431700 | Limpia Ck abv Ft Davis, TX | 30°36'48" | 104°00'04" | 13070005 | 52.4 |
| 8 | 09352900 | VALLECITO CREEK NEAR BAYFIELD, CO. | 37°28'39" | 107°32'35" | 14080101 | 72.5 |
| 9 | 09430600 | MOGOLLON CREEK NEAR CLIFF, NM | 33°10'00" | 108°38'57" | 15040001 | 69 |
| 10 | 09508300 | WET BOTTOM CREEK NEAR CHILDS, AZ. | 34°09'39" | 111°41'32" | 15060203 | 36.4 |
| 11 | 11264500 | MERCED R A HAPPY ISLES BRIDGE NR YOSEMITE CA | 37°43'54" | 119°33'28" | 18040008 | 181 |
| 12 | 12447390 | ANDREWS CREEK NEAR MAZAMA, WA | 48°49'23" | 120°08'41" | 17020008 | 22.1 |
| 13 | 13169500 | BIG JACKS CREEK NR BRUNEAU ID | 42°47'06" | 115°59'00" | 17050102 | 253 |
| 14 | 13331500 | MINAM RIVER NEAR MINAM, OR | 45°37'12" | 117°43'32" | 17060105 | 240 |
| 15 | 14048000 | JOHN DAY RIVER AT MCDONALD FERRY, OR | 45°35'16" | 120°24'30" | 17070204 | 7,580.00 |
| 16 | 14103000 | DESCHUTES RIVER AT MOODY, NEAR BIGGS, OR | 45°37'20" | 120°54'16" | 17070306 | 10,500.00 |
| 17 | 14113000 | KLICKITAT RIVER NEAR PITT, WA | 45°45'24" | 121°12'32" | 17070106 | 1,297.00 |

*(Source: http://waterdata.usgs.gov)*

## 3. Watershed/Basin delineation

The watershed delineation process was based on Desktop GIS software (9.1) and the latest Hydrology Modeling Module/Extension made by ESRI (ESRI, 2005) in combination with the reference from basin and sub-basin borders of the National Hydrography Dataset (http://nhd.usgs.gov).

Watershed delineation was the most time-consuming process of the data preparation. The process is summarized as below:

(1) First get the geo-coordinates of the gaged-stream stations, corresponding hydrologic unit code (HUC), corresponding basin drainage area by entering the station number into the "Site information" link shown on the USGS website (http://waterdata.usgs.gov/nwis). Then, summarize this information in a Table (e.g. Table 2).

(2) Get a "draft" basin area by entering the HUC into the National Hydrography Dataset (http://nhd.usgs.gov). For large and very large basins (e.g. Canadian River near Whitefield, OK, station number: 07245000), it is necessary to get the basin area of a lower digit number of the HUC (e.g. 2 or 4 digit HUC) to make sure that the "draft" area is big enough to contain the actual area of the basin wanted.

(3) Get the coordinates of the "draft" border gathered in step 2, and enter them into the USGS website at http://seamless.usgs.gov/website/seamless/viewer.php or http://www.mrlc.gov and download the corresponding National Elevation Dataset (NED)

(4) Use the geo-coordinates of each station to create a corresponding point shapefile in ArcMap.

(5) Use Desktop GIS software and Hydrology Modeling extension to delineate the watershed based on the National Elevation Data gathered in step 3 and the point shapefile feature created in step 4.

The delineated watershed area will be compared with the corresponding area from USGS (in Table 2) by calculating the area variation in percent. The formula for this is expressed as:

$$\textbf{Area Difference (\%)} = \textbf{(Delineated area - USGS area)*100/ (USGS area)} \qquad (1)$$

The watersheds delineated in the process are very important input to get other GIS data such as population, precipitation, and land cover.

## 4. Data calculations for the stream water constituents

The raw stream water ion data were downloaded from a USGS website at http://waterdata.usgs.gov. The data were then calculated for average values per record and complied in a table. For pH, instead of the mean, the median was calculated because of the fact that a change of one pH unit represents a tenfold change in hydrogen ion concentration (Project SEARCH, undated). This means that the pH distribution is right skewed, therefore, a median pH can be more appropriate to be expressed rather than a mean pH.

The formula for converting a measure of a constituent from mg/L to kg/ha/year is expressed as:

**Ion yield of a constituent (kg/ha/yr) = 3.45 x Average number of mg/L of the constituent x Average number of instantaneous discharge (cfs)/ Drainage area of the station (2)**

*(3.45 = 365 x 60 x 60 x 24 x 0.3048 x 10/27878400; 1 square mile = 27878400 square feet; 1 foot = 0.3048 meter); (constant of 10 = 10000/1000)*

Due to lack of available data, dissolved solids were calculated indirectly from the specific conductance by multiplying specific conductance by a constant of 0.6 (Stednick, 1991). Also, bicarbonate was calculated by multiplying alkalinity as $CaCO_3$ by a constant of 1.22 (Stednick, 1991).

## 5. GIS data preparation and calculations for the environmental factors

GIS data layers of land cover, precipitation, population are delineated based on the corresponding delineated watershed using Arctoolbox of Desktop GIS. After those layers are derived, they were exported to *.dbf files, to calculate the data as follows:

### *(1) For land cover data:*

Originally, these data were divided into 21 classes/categories (Appendix 3a_4). However, as stated, all classes may not be represented in a specific state data set. The class numbers representing the digital value of the class in the data set are as the followings: (http://sain.nbii.org/downloads/regional_map/gisdata/nlcd_faq.shtml):

*Water: 11 Open Water; 12 Perennial Ice/Snow*

*Developed: 21 Low Intensity Residential; 22 High Intensity Residential; 23 Commercial/Industrial/Transportation*

*Barren: 31 Bare Rock/Sand/Clay; 32 Quarries/Strip Mines/Gravel Pits; 33 Transitional Vegetated;*

*Natural Forested Upland:*

*41 Deciduous Forest; 42 Evergreen Forest; 43 Mixed Forest*

*Shrubland: 51 Shrubland*

*Non-natural Woody: 61 Orchards/Vineyards/Other*

*Herbaceous Upland: 71 Grasslands/Herbaceous*

*Herbaceous Planted/Cultivated: 81 Pasture/Hay; 82 Row Crops; 83 Small Grains; 84 Fallow; 85 Urban/Recreational Grasses*

*Wetlands: 91 Woody Wetlands; 92 Emergent Herbaceous Wetlands*

The area of each land cover class (in square mile – SQMI) is calculated from the dbf file using the following equation:

**Area of a land cover class (SQMI) = Number of counts of that class x Area of the watershed (SQMI)/Total count of the watershed** **(3)**

One example of area calculation of each land cover class is presented in Table 3 as follows:

**Table 3. Land cover classification and corresponding area of each class for station number 06478500 - JAMES R NEAR SCOTLAND, SD (sandstone).**

| Station number: 06478500 - JAMES R NEAR SCOTLAND,SD (Sandstone) | | | | |
|---|---|---|---|---|
| **ObjectID** | **Value of land cover type** | **Count** | **Area fraction (%)** | **Area (SQMI)** |
| 0 | 11 | 680019 | 0.8 | 176.2 |
| 1 | 21 | 105367 | 0.1 | 27.3 |
| 2 | 22 | 28343 | 0.0 | 7.3 |
| 3 | 23 | 169145 | 0.2 | 43.8 |
| 4 | 31 | 363 | 0.0 | 0.1 |
| 5 | 32 | 1462 | 0.0 | 0.4 |
| 6 | 33 | 24 | 0.0 | 0.0 |
| 7 | 41 | 685083 | 0.8 | 177.5 |
| 8 | 42 | 7194 | 0.0 | 1.9 |
| 9 | 43 | 819 | 0.0 | 0.2 |
| 10 | 51 | 41613 | 0.1 | 10.8 |
| 11 | 71 | 12931117 | 16.0 | 3350.3 |
| 12 | 81 | 21736287 | 26.9 | 5631.6 |
| 13 | 82 | 39021698 | 48.2 | 10110.1 |
| 14 | 83 | 2633738 | 3.3 | 682.4 |
| 15 | 84 | 43169 | 0.1 | 11.2 |
| 16 | 85 | 18262 | 0.0 | 4.7 |
| 17 | 91 | 22358 | 0.0 | 5.8 |
| 18 | 92 | 2825651 | 3.5 | 732.1 |
| **Total** | | **80951712** | **100.0** | **20973.7** |

Since 21 classes of land cover are not practical for regression and model selection processes, the classes were recombined into 4 categories only, namely ***Water, Developed, Forest,*** and ***Agriculture***, in which:

***Water*** = land cover values of (11 + 12 + 91 + 92).

***Developed*** = land cover values of (21 + 22 + 23 + 31 + 32 + 33).

***Forest*** = land cover values of (41 + 42 + 43 + 51) (note: 51 is shrubland but it is more reasonable to add this class into Forest than into other categories).

***Agriculture*** = land cover values of (61 + 71 + 81 + 82 + 83 + 84 + 85)

One example of the recombination of land cover classes into new categories is illustrated in Table 4 as below:

**Table 4. Recombination of land cover classes into 4 categories for station number 06478500 - JAMES R NEAR SCOTLAND, SD (sandstone).**

| Station number: 06478500 (sandstone, square mile) | | | | |
|---|---|---|---|---|
| Water | Developed | Forest | Agriculture | Total area (SQMI) |
| 914.1 | 78.9 | 190.4 | 19790.3 | 20973.7 |

The unit of area is in square mile (SQMI).

***(2) For precipitation data:***

These average data were compiled based on the dbf data file exported from the precipitation layer of each watershed using ***area weighted fractions*** and the range of precipitation for that area.

**Annual Average Precipitation = $\Sigma$ Area of range(i) x Precipitation of range(i) /Total area, (4)**

**( i runs from 1 to n).**

The unit of average annual precipitation is in inch (in).

One example of precipitation calculation is indicated in Table 5 as below:

**Table 5. Average annual precipitation for basin station number 06478500 (sandstone).**

| Station number: 06478500 - JAMES R NEAR SCOTLAND,SD (sandstone) | | | |
|---|---|---|---|
| **Range of Precipitation (In)** | **Area in square degree** | **Area fraction** | **Precipitation (in)** |
| 21 | 1.3 | 0.2 | 4.3 |
| 19 | 3.2 | 0.5 | 9.9 |
| 17 | 1.2 | 0.2 | 3.2 |
| 23 | 0.6 | 0.1 | 2.0 |
| **Total** | **6.2** | **1** | **19.4** |

***(3) For population data:***

These data were compiled using the spatial summary of the GIS population layer (or can be calculated in the same way as precipitation data were. The two results are almost the same).

The unit of population is an average of people/square mile (people/SQMI).

***(4) For slope data:***

The average slope of each watershed is calculated by using the slope function of the Spatial Analysis in ArcGIS.

The unit of average slope is in degrees ($^{o}$).

***(5) For watershed area data:***

These data are available from delineated watersheds using area function of ArcGIS. The unit of watershed area is in square mile (SQMI).

## 6. Building regression models for estimating constituent yield

Empirical relationships between ion yield and the selected environmental characteristics (independent variables) were examined. Various multivariate techniques are available to quantify the relation between the ion yields and those independent variables, the linear and multiple linear regressions were preferred for this study. These techniques were considered most suitable to provide quantitative information on the relationship between the environmental characteristics and stream ion yields (Lystrom et al., 1978; Peters, 1984).

The general form of a multiple linear regressions (after Peters, 1984) can be expressed as:

$$Y_i = a + b_1X_1 + b_2X_2 + .... + b_nX_n ; \quad (5)$$

Where

$Y_i$ = the annual yield of constituent,

$X_1, X_2 \ldots, X_n$ = the basin characteristics.

All regressions and other related procedures were performed by using SPSS version13.0 (SPSS, 2004).

As tested in this study, most distributions of stream ion concentrations and the environmental factors are non-normal with a p-value less than 0.05. These results are consistent with the research findings of Davis (1973, as cited in Peters, 1984). Hydrologic data are positively skewed, meaning that values smaller than the mean are more common than values greater the mean (Gray, 1970, as cited in Peters, 1984).

The distribution of log-transformed data is closer to normal than the distribution of original data, and the linear regression model works better with normal variables (SPSS, 2004). Therefore, these data were transformed into the common logarithm transformation (base 10) (Peters, 1984). In addition, for any new recombined categories of the land cover data with a value of zero, a value of 0.000347 SQMI is replaced to make the log transform possible (the original spatial resolution of the Landsat-5 TM images, which the land cover data were derived, was 30 meters or 30 x 30 = 900 square meters, or equal to 0.003474 SQMI). Therefore, any areas less than that value are not considered due to unclear spectrals influenced by other land cover types and objects. Bedrock types were treated as class/dummy variables in the multiple linear regression analysis (Peters, 1984).

The form of the log-transformed regression was:

$$\mathbf{Log\ Y_i = log\ a + b_1D_1 + b_2D_2 + b_3 logX_1 + b_4 logX_2 + .... + b_n logX_n} \qquad \textbf{(6)}$$

Where:

Y = Average annual ion yield,

$D_1$ and $D_2$ = rock types (since there were 3 bedrock types, only two dummy variables need to be entered, the other is implied to be the control).

and X1, etc. = remaining basin characteristics.

## 7. Multicollinearity diagnostics

Multicollinearity can occur when there are near-constant linear functions of two or more of the predictor variables. If multicollinearities exist, the variances of the estimated regression coefficients may become very large. That may lead to unstable estimates of the regression equation. Several approaches have been proposed to overcome the multicollinearity problem. For example, the equation does not contain multicollinearities if only a subset of the predictor variables are used (Jolliffe, 2002).

In this study, multicolinearity among the independent variables can be detected using the correlation matrix and the regression collinearity diagnostics based on the "Enter" method of the regression model selection (Appendices 4, 5, 6). In more detail, this can be based on the following indicators (SPSS, 2004):

***(1) Based on the tolerance:***

Tolerance equals to $1 - R^2_i$, "where $R^2_i$ is the squared multiple correlation of the i^th^ variable with other independent variables" (Liu et al., 2002). In other words, "the tolerance is the percentage of the variance in a given predictor that cannot be explained by the other predictors". If the tolerances are close to 0, there is high multicollinearity and the standard error of the regression coefficients will be inflated." (SPSS, 2004).

### *(2) Based on the Variance Inflation Factor (VIF):*

It is usually considered problematic if variance inflation factor greater than 2 (SPSS, 2004). In fact, VIF is reciprocal of tolerance. The variables with low tolerance tend to have large VIF values. Variables with low tolerance and large VIF, therefore, suggest that they have a collinearity (Liu et al., 2002).

### *(3) Based on the initial total eigenvalues:*

If the initial total eigenvalues in the collinearity diagnostics are close to 0, the predictors are highly intercorrelated and "small changes in the data values may lead to large changes in the estimates of the coefficients" (SPSS, 2004).

### *(4) Based on the Condition Indices:*

"The condition indices are computed as the square roots of the ratios of the largest eigenvalue to each successive eigenvalue" (SPSS, 2004). If condition index value is greater than 15, there is a possible problem with collinearity. When condition index value is greater than 30, it is considered a serious problem with collinearity (SPSS, 2004).

There are two ways to minimize the muticollinearity problem used in this study. First, use the stepwise method of model selection to include only the most useful variables in the model (SPSS, 2004) and therefore, multicollinearity among independent variables can be reduced.

Second, apply principal component regression. In a principal component regression, the independent variables are, not separate variables per se, but the components of all original input independent variables. The principal component regression analysis does not only make collinearity diagnosis for each independent variable, but also solve the collinearity problem (Liu et al., 2002).

To do a principal component regression (PCR), it is necessary to extract components by implementing the principal component analysis (PCA). Since the measurement scale of the independent variables in this study is different, a correlation matrix (not covariance matrix) was used for the extraction analysis. In addition, for ease of interpretation, a rotated component matrix was performed. The rotated component matrix is useful to determine what the components represent (SPSS, 2004).

In a principal component analysis, initial total eigenvalues is the expression of the amount of variance in the original variables accounted for each component (SPSS, 2004). "For the initial solution, there are as many components as variables, and in a correlations analysis, the sum of the initial total eigenvalues equals to the number of components." (SPSS, 2004). Since one of the aims of this study was to model the most effective components making the most considerable contributions on the variation of the dependent variables (ion yields), only initial eigenvalues over 1.0 were extracted and taken into account for the regression performance.

For compatibility with the multiple linear regressions, principal component regressions also used the log transformed data and the stepwise method of the model selection. The form of this regression model is as below:

$$\mathbf{Log\,Y_i = \mathit{a} + b_1C_1 + b_2C_2 + \ldots. + b_nC_n} \tag{7}$$

$$\mathbf{C_i = \Sigma a_{ij}\, Z(X'_j)} \qquad \mathbf{(i = 1,\ldots., n\,;\, j = 1,\ldots.., n)} \tag{8}$$

$$\mathbf{Z(X'_j) = (X_j - X^{-}{}_j)/Sx_j} \qquad \mathbf{(j = 1,\ldots.., n)} \tag{9}$$

Where,

$Y$ = Average ion yield,

$b_i$ = Slope coefficient of the i$^{th}$ principal component

$C_i$ = The i$^{th}$ principal component,

n = Number of independent variables,

$Z(X'_j)$ = The jth standardized independent variable (Z-score) of the jth independent log-transformed variable (Xj ),

$X^{-}{}_j$ = The mean of the jth log-transformed independent variable,

$Sx_j$ = The standard deviation of the jth log-transformed independent variable,

$a_{ij}$ = The coefficient of the principal component coefficient matrix.

Once the regression in equation (7) is created, it is possible to transform those Ci back to $Z(X'_j)$ and then to the original $X_j$

## 8. Determining the most effective factors on constituent yields

The factors with significant slope coefficients in the regression equation are more important than those not. The higher the standardized slope coefficient (beta), the more the variation of the dependent variable is accounted for (SPSS, 2004). The adjusted R square, part, partial, and statistical significance (p) are also useful to determine the importance of the environmental factors. Since regression coefficient R square is affected by the number of independent variables and the corresponding sample size of the model, the adjusted R square (not R square) is usually used to compare the goodnesses of fit between different linear models (Liu et al., 2002).

## 9. Examining the effects of bedrock types on ion yields

One way to examine the effects of bedrock types on ion yields is to use comparison tests. If ion yields of a same constituent for any two bedrock types are significantly different, then the difference of the bedrocks can be accounted for the difference of that constituent.

Since the sample size of each bedrock type is not large (limestone -19, sandstone – 11, and crystalline – 17) and the distribution of the data is non-normal, a non parametric comparison using Mann Whitney U Test in SPSS (2004) was conducted to examine the differences between constituents of any two bedrock types.

Another way to examine the effects of bedrock types on ion yields is to use bedrock types as class (dummy) variables in the multiple linear regression analysis (Equation 6). However, equation 6 is compatible for the assumption that the slope coefficients of the same independent variables are the same in the two equations. This assumption is not appropriate in many circumstances. Therefore, that model is good for testing the significant differences of the intercepts only.

Since most of the differences among constituents of the three bedrock types were significant, equation (6) was used to fit for each type of bedrock separately, meaning that the class variable was not taken into account. However, if a same independent variable was detected in any two equations (of different bedrock types) for the same constituent, a comparison test was performed to test the significance of that coefficient slope and the intercept. This is another way to test the effects of bedrock types on ion yields. The general form for this regression can be expressed as:

$$\mathbf{Log\ Y_i = \log(b_0) + b_1D_1 + b_2D_2 + b_{3(1)}\log X_1 + b_{4(1)}\log X_2 + \dots + b_{n(1)}\log X_n + b_{3(2)}D_1(\log X_1) + b_{4(2)}D_1(\log X_2) + \dots + b_{n(2)}D_1(\log X_n) + b_{3(3)}D_2(\log X_1) + b_{4(3)}D_2(\log X_2) + \dots + b_{n(3)}D_2(\log X_n)} \qquad \mathbf{(10)}$$

Where

$Y$ = Average ion yield,

$D_1$ and $D_2$ = rock types (since there were 3 bedrock types, only two dummy variables need to be entered, the other is implied to be the control one).

and

$X_1$, etc. = remaining basin characteristics.

In equation (10), if the $b_1$ and or $b_2$ coefficients are significant, they are due to the effects of bedrock types on the ion yields. If any slope coefficient for terms with bedrock dummy variables (e.g. $b_iD_j$) is significant, the corresponding independent variable is interpreted to be different between/among the bedrock types. Otherwise, it is not.

## 10. Stream water ion yield prediction, model validation, and spatial distribution of the predicted ion yields.

The models of principal component regression and the traditional multiple linear regression were compared to each other. Then, the most significant regression equation models were selected for ion yield prediction. The ArcGIS-based models were used to display and classify the areas of different affects on stream water ion yields.

The predicted values are compared with actual data of the average control observed constituent concentrations to determine which model might better represent prediction for the dependent responses. The model verification and validation procedures are based on the following criteria:

- High regression correlation ($R^2 >= 0.64$)
- No or low collinearity
- Low Average Percent Error (APE) for each model:

$$\text{APE (\%)} = (\text{actual value} - \text{predicted value})/\text{actual value} \qquad (11)$$

In general, the percentage in equation (11) is expected to be as small as possible (e.g. less than or equal to 5%). That was also known as the "acid test". However, depending on each application and situation, higher APE may be also accepted.

Due to the lack of data, only 4 sandstone stations were available for the prediction and validation of the models.

# IV. RESULTS AND DISCUSSION

## 1. Descriptions of the project watersheds

### *1. 1. Watershed delineation and corresponding areas:*

Watershed delineations were implemented based on the National Elevation Dataset, ArcGIS and the Hydrology Modeling Extension. The delineations of 47 watersheds can be found in Appendix 8. Figure 3 shows an example of a sandstone watershed delineated from National Elevation Dataset.

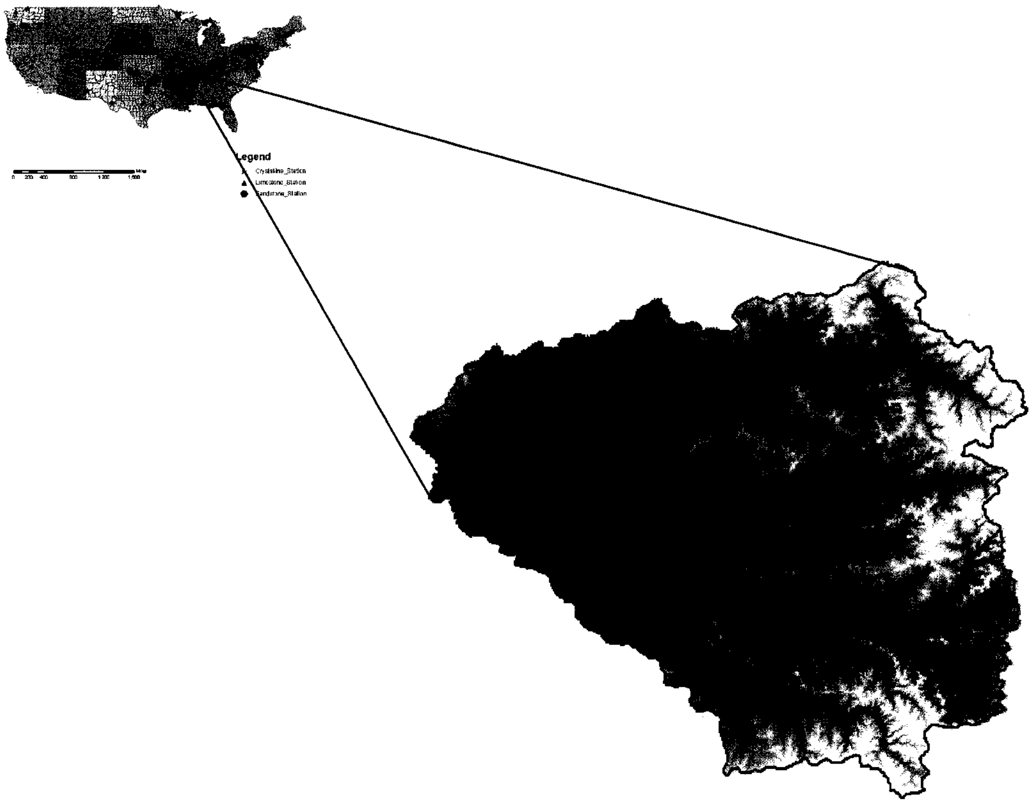


**Figure 3. A sandstone watershed delineated from National Elevation Dataset.**

It can be noted that the watershed shapes are very diverse and irregular (Appendix 8, Figure 3). The areas (SQMI) of those watersheds are also variable (Table 6 and Table 7a). They range from small (10.8 SQMI, limestone) to large (47223.1 SQMI, sandstone) (Table 7a). Furthermore, the area differences between delineated and USGS watersheds are very small although most delineated watersheds are slightly smaller than those of USGS data (indicated by negative sign, Table 7a). The mean area difference of sandstone bedrock is 1.4%, of limestone bedrock is 1.9%, and of crystalline is 3.5% (Table 6). Those numbers indicate that almost delineated watersheds were acceptable when compared to those of USGS (the differences < 5 %). It should also be noted that, however, there were some watersheds with a larger difference (e.g. 18.6% (USGS 08103900, limestone), 13.3% (USGS 04014500, crystalline)) (Table 6 and Appendix 3).

**Table 6. Descriptive statistics of the delineated watershed areas and the area differences in Table 7a (in square mile, SQMI).**

| | n | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Limestone_Delineated (SQMI) | 19 | 10.8 | 32784.1 | 6314.9 | 8889.0 |
| Limestone_Difference (%) | 19 | .0 | 18.6 | 1.9 | 4.1 |
| Sandstone_Delineated (SQMI) | 11 | 12.1 | 47223.1 | 13241.4 | 14085.9 |
| Sandstone_Difference (%) | 11 | .2 | 3.0 | 1.4 | .9 |
| Crystalline_Delineated(SQMI) | 17 | 22.2 | 10004.1 | 1473.6 | 2925.4 |
| Crystalline_Difference (%) | 17 | .1 | 13.3 | 3.5 | 4.0 |

**Table 7a. Differences between the delineated and USGS areas of the project watersheds.**

| | Area (SQMI) | | |
|---|---|---|---|
| | Delineated | USGS | Difference (%) |
| Basin Station | Limestone Bedrock | | |
| USGS 02313000 | 1804.5 | 1820.0 | -0.9 |
| USGS 02320500 | 7698.4 | 7880.0 | -2.3 |
| USGS 02329000 | 1145.9 | 1140.0 | 0.5 |
| USGS 02368000 | 628.8 | 624.0 | 0.8 |
| USGS 03085000 | 7332.3 | 7337.0 | -0.1 |
| USGS 03254000 | 3371.3 | 3385.0 | -0.4 |
| USGS 03290500 | 6168.3 | 6180.0 | -0.2 |
| USGS 03425000 | 10690.2 | 10690.0 | 0.0 |
| USGS 03543005 | 17266.2 | 17310.0 | -0.3 |
| USGS 03571850 | 22627.5 | 22640.0 | -0.1 |
| USGS 03593005 | 32784.1 | 32820.0 | -0.1 |
| USGS 04045500 | 771.8 | 790.0 | -2.3 |
| USGS 04057004 | 1461.5 | 1445.0 | 1.1 |
| USGS 04108690 | 2023.0 | 2020.0 | 0.1 |
| USGS 04126520 | 1902.6 | 1928.0 | -1.3 |
| USGS 04132052 | 1537.6 | 1500.0 | 2.5 |
| USGS 04165500 | 718.5 | 734.0 | -2.1 |
| USGS 08103900 | 39.5 | 33.3 | 18.6 |
| USGS 13018300 | 10.8 | 10.6 | 1.9 |
| Sandstone Bedrock | | | |
| USGS 01545600 | 44.8 | 46.2 | -3.0 |
| USGS 03049625 | 11445.1 | 11546.0 | -0.9 |
| USGS 03237280 | 12.1 | 12.2 | -0.9 |
| USGS 06452000 | 9913.5 | 10200.0 | -2.8 |
| USGS 06478500 | 20973.7 | 20653.0 | 1.6 |
| USGS 07026000 | 1848.7 | 1852.0 | -0.2 |
| USGS 07231500 | 27609.2 | 27952.0 | -1.2 |
| USGS 07234000 | 7826.6 | 7955.0 | -1.6 |
| USGS 07237500 | 11696.9 | 11589.0 | 0.9 |
| USGS 07245000 | 47223.1 | 47576.0 | -0.7 |
| USGS 06438000 | 7061.8 | 7210.0 | -2.1 |
| Crystalline Bedrock | | | |
| USGS 01059400 | 3462.5 | 3434.0 | 0.8 |
| USGS 01066000 | 1287.5 | 1293.0 | -0.4 |
| USGS 04014500 | 121.4 | 140.0 | -13.3 |
| USGS 05124480 | 237.2 | 254.0 | -6.6 |
| USGS 06623800 | 72.2 | 72.7 | -0.7 |
| USGS 07311200 | 22.2 | 24.6 | -9.9 |
| USGS 08431700 | 52.0 | 52.4 | -0.8 |
| USGS 09352900 | 78.2 | 72.5 | 7.8 |
| USGS 09430600 | 73.4 | 69.0 | 6.4 |
| USGS 09508300 | 36.5 | 36.4 | 0.1 |
| USGS 11264500 | 179.9 | 181.0 | -0.6 |
| USGS 12447390 | 22.6 | 22.1 | 2.3 |
| USGS 13169500 | 244.2 | 253.0 | -3.5 |
| USGS 13331500 | 237.1 | 240.0 | -1.2 |
| USGS 14048000 | 7629.7 | 7580.0 | 0.7 |
| USGS 14103000 | 10004.1 | 10500.0 | -4.7 |
| USGS 14113000 | 1291.3 | 1297.0 | -0.4 |

To assure that the delineated watershed area and corresponding USGS area are not different, a nonparametric test was conducted. The results (Table 7b) indicate that they are not statistically different, the P-values range from 0.847 (sandstone watersheds) to 0.919 (crystalline watersheds) and to 0.977 (limestone watersheds) (Table 7b).

**Table 7b. Nonparametric test for the differences between delineated watershed areas and USGS watershed areas.**

| | Area_Limestone | Area_Sandstone | Area_Crystalline |
|---|---|---|---|
| Mann-Whitney U | 179.0 | 57.0 | 141.0 |
| Wilcoxon W | 369.0 | 123.0 | 294.0 |
| Z | -.044 | -.230 | -.121 |
| Asymp. Sig. (2-tailed) | .965 | .818 | .904 |
| Exact Sig. [2*(1-tailed Sig.)] | .977 | .847 | .919 |

## *1. 2. Basin characteristics (environmental factors):*

Land cover, population, and precipitation were clipped for each watershed separately using the delineated results from the National Elevation Dataset. These results are described in the following sections:

### 1. 2.1. Land cover:

The 21 land cover classes of the watersheds in each bedrock type and the recombined 4 land cover categories were presented in Appendix 3a_4.

The descriptive statistics of the 4 new categories of land cover of each watershed are displayed in Table 8. The land cover types range considerably from zero SQMI (e.g. ***Water*** of crystalline bedrock) to approximately 36,000 SQMI (***Agriculture*** of sandstone bedrock).

**Table 8. Descriptive statistics of the 4 land cover categories of the project watersheds (in square mile, SQMI).**

| | n | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| | **Limestone Bedrock** | | | | |
| Water | 19 | 0 | 1923.6 | 369.3 | 471.8 |
| Developed | 19 | .2 | 1050.8 | 231.9 | 310.4 |
| Forest | 19 | 8.9 | 23514.5 | 4383.5 | 6782.3 |
| Agriculture | 19 | 1.8 | 7159.9 | 1330.2 | 1750.0 |
| | **Sandstone Bedrock** | | | | |
| Water | 11 | 0 | 914.1 | 213.8 | 276.9 |
| Developed | 11 | 0 | 563.9 | 147.8 | 173.8 |
| Forest | 11 | 12.0 | 10206.0 | 2668.9 | 3909.3 |
| Agriculture | 11 | .1 | 35934.6 | 10210.8 | 11125.0 |
| | **Crystalline Bedrock** | | | | |
| Water | 17 | 0 | 271.8 | 35.2 | 70.0 |
| Developed | 17 | 0 | 194.9 | 29.5 | 51.8 |
| Forest | 17 | 5.5 | 8707.1 | 1271.6 | 2555.7 |
| Agriculture | 17 | .5 | 989.5 | 137.3 | 297.8 |

In general, average ***Water*** and ***Developed*** lands of each bedrock type occupied only very small percentages (1.0% to 13.6%) while ***Forest*** and ***Agriculture*** lands are major categories occupying up to 79.6% (crystalline) and 62.3% (sandstone), respectively (Table 9).

**Table 9. Percentage of the 4 land cover categories of the project watersheds (%).**

| | ***Water*** | ***Developed*** | ***Forest*** | ***Agriculture*** |
|---|---|---|---|---|
| Limestone | 13.6 | 4.8 | 57.0 | 24.7 |
| Sandstone | 1.8 | 1.0 | 34.9 | 62.3 |
| Crystalline | 4.2 | 3.5 | 79.6 | 12.7 |
| Average | 6.5 | 3.1 | 57.2 | 33.2 |

### 1. 2. 2. Population:

The average population density (people/SQMI) of the watersheds of each bedrock type was analyzed to get the descriptive statistics (Table 10).

**Table 10. Average population of each watershed (people/SQMI, 1990).**

| Site No. | Limestone | Sandstone | Crystalline |
|---|---|---|---|
| 1 | 146.7 | 39.0 | 51.9 |
| 2 | 41.3 | 169.8 | 42.7 |
| 3 | 74.5 | 122.4 | 5.0 |
| 4 | 62.1 | 4.1 | 4.5 |
| 5 | 225.7 | 8.1 | 1.3 |
| 6 | 49.4 | 55.2 | 103.0 |
| 7 | 104.2 | 18.3 | 1.0 |
| 8 | 55.8 | 5.9 | 14.1 |
| 9 | 112.4 | 5.8 | 2.7 |
| 10 | 113.2 | 27.0 | 9.2 |
| 11 | 105.3 | 5.7 | 22.1 |
| 12 | 11.1 | | 6.0 |
| 13 | 8.0 | | 1.0 |
| 14 | 186.1 | | 7.2 |
| 15 | 37.2 | | 3.1 |
| 16 | 31.2 | | 11.9 |
| 17 | 1285.6 | | 25.4 |
| 18 | 22.0 | | |
| 19 | 3.0 | | |
| **Descriptive Statistics** | | | |
| Minimum | 3.0 | 4.1 | 1.0 |
| Maximum | 1285.6 | 169.8 | 103.0 |
| **Mean** | **140.8** | **41.9** | **18.4** |
| Std. Deviation | 283.9 | 55.0 | 26.3 |

Average population density ranges from 1.0 people/SQMI (crystalline) to 1285.6 people/SQMI (limestone). The mean population density of each bedrock type also varied considerably from 18.4 (crystalline) to 140.8 (limestone). Limestone bedrock had the largest mean, however, the standard deviation was also highest (283.9) (Table 10).

### 1. 2. 3. Precipitation:

The lowest average annual precipitation (in/yr) was found in crystalline bedrock (11.5 in) while the highest amount belonged to limestone bedrock (60.3 in) (Table 11). Sandstone bedrock had the lowest mean annual precipitation (28.4 in) and limestone bedrock had the highest (43.7 in). Also, the standard (std.) deviation of each bedrock type was not very different from each other (from 10.2 to 13.7 in).

**Table 11. Average annual precipitation of each watershed (in/year, 1961-1990).**

| Site No. | Limestone | Sandstone | Crystalline |
|---|---|---|---|
| 1 | 51.3 | 41.8 | 45.7 |
| 2 | 51.6 | 43.9 | 50.1 |
| 3 | 53.2 | 41.0 | 29.0 |
| 4 | 60.3 | 17.8 | 29.0 |
| 5 | 46.2 | 19.4 | 46.5 |
| 6 | 45.0 | 53.0 | 29.8 |
| 7 | 47.0 | 19.6 | 20.3 |
| 8 | 53.3 | 17.5 | 40.6 |
| 9 | 52.4 | 18.9 | 26.7 |
| 10 | 53.9 | 22.3 | 22.0 |
| 11 | 54.7 | 17.1 | 43.4 |
| 12 | 33.0 |  | 34.2 |
| 13 | 32.5 |  | 11.5 |
| 14 | 35.8 |  | 46.4 |
| 15 | 32.5 |  | 18.4 |
| 16 | 31.7 |  | 20.5 |
| 17 | 31.1 |  | 34.6 |
| 18 | 31.0 |  |  |
| 19 | 34.6 |  |  |
| **Descriptive Statistics** |  |  |  |
| Minimum | 31.0 | 17.1 | 11.5 |
| Maximum | 60.3 | 53.0 | 50.1 |
| **Mean** | **43.7** | **28.4** | **32.3** |
| Std. Deviation | 10.2 | 13.5 | 11.7 |

### 1. 2. 4. Average slope:

The average slope (degree) of each watershed was low, ranging from 0.7 degrees (limestone) to 27.5 degrees (crystalline) (Table 12). The means of the average slope of each bedrock type were also low and very close to each other, ranging from 5.2 (sandstone) to 6.4 (limestone) and to 13.2 (crystalline). However, the standard deviations of the mean slope of each bedrock type was large, ranging from 5.1 (sandstone) to 6.2 (limestone) and to 8.1 (crystalline) (Table 12). This suggests that slope may not affect the average annual stream ion yields.

**Table 12. Average slope of each watershed (degree).**

| Site No. | Limestone | Sandstone | Crystalline |
|---|---|---|---|
| 1 | 0.7 | 12.3 | 7.7 |
| 2 | 0.8 | 7.0 | 8.5 |
| 3 | 1.7 | 17.2 | 2.7 |
| 4 | 2.7 | 4.0 | 3.0 |
| 5 | 11.1 | 1.0 | 12.6 |
| 6 | 9.6 | 2.5 | 6.2 |
| 7 | 13.3 | 3.8 | 14.0 |
| 8 | 11.3 | 1.0 | 26.8 |
| 9 | 13.6 | 1.2 | 27.5 |
| 10 | 12.7 | 2.9 | 20.0 |
| 11 | 10.4 | 4.8 | 17.8 |
| 12 | 1.1 | | 21.2 |
| 13 | 0.9 | | 5.7 |
| 14 | 1.6 | | 22.9 |
| 15 | 2.6 | | 12.8 |
| 16 | 2.3 | | 6.8 |
| 17 | 1.1 | | 8.5 |
| 18 | 2.8 | | |
| 19 | 21.7 | | |
| **Descriptive Statistics** | | | |
| Minimum | .7 | 1.0 | 2.7 |
| Maximum | 21.7 | 17.2 | 27.5 |
| Mean | **6.42** | **5.24** | **13.23** |
| Std. Deviation | 6.23 | 5.14 | 8.10 |

### *1. 3. Stream water ion yields (kg/ha/yr):*

Stream water ion yields were calculated (Appendix 3). The average annual ion yields (kg/ha/yr) varied dramatically among watersheds and among bedrock types (Table 13a). In general, total dissolved solids (TDS) was the highest for the three bedrock types, with TDS of limestone (722.6), sandstone (301.3), crystalline (230.3). These differences were significantly different ($p < 0.05$, Table 16). The lowest values were for orthophosphate and phosphorus, being just less than 1.0 for all three bedrocks. The next lowest values were for total nitrogen, ranging from 2.5 (sandstone) to 4.3 (limestone). Potassium ranged from 4.6 (sandstone) to 7.0 (limestone); magnesium from 9 (crystalline) to 36 (limestone); sodium from 21.4 (crystalline) to 43.6 (limestone); dissolved oxygen from 28.3 (sandstone) to 50.4 (crystalline); calcium from 40.2 (crystalline) to 144.0 (limestone); chloride from 16.0 (crystalline) to 61.3 (limestone); sulfate from 30.8 (crystalline) to 123.0 (limestone); bicarbonate from 89.5 (sandstone) to 441.3 (limestone). These results are similar to the earlier study (Peters 1984).

The values of pH among different bedrock types were very close to each other, ranging from 7.5 crystalline to 7.7 (limestone, and sandstone).

In general, when compared to Peters 1984, most average constituent yields of the current research appear higher, except for potassium, magnesium, calcium, and sulfate in limestone watersheds (Table 13b and Appendix 7) although they are not statistically different ($p > 0.05$, Table 13c). The slight differences are partly due to the fact that the delineated watershed areas are a little smaller than those of USGS data as shown in Table 7a.

**Table 13a. Descriptive statistics of average annual stream water ion yields of each bedrock type (kg/ha/yr).**

| Water constituent | *n* | *Minimum* | *Maximum* | *Mean* | *Std. Deviation* |
|---|---|---|---|---|---|
| | **Limestone Bedrock** | | | | |
| TDS | 19 | 199.6 | 1415.5 | 722.6 | 384.4 |
| DO | 19 | 8.9 | 64.0 | 40.6 | 15.6 |
| pH | 19 | 7.0 | 8.4 | 7.7 | 0.4 |
| Total_Nitrogen | 16 | 0.6 | 11.7 | 4.3 | 3.2 |
| Orthophosphate | 19 | 0.1 | 2.6 | 0.7 | 0.7 |
| Phosphorus | 19 | 0.0 | 0.7 | 0.2 | 0.2 |
| Ca | 19 | 17.3 | 270.4 | 144.0 | 72.8 |
| Mg | 19 | 6.7 | 71.9 | 36.0 | 20.1 |
| Na | 19 | 7.1 | 178.0 | 43.6 | 43.4 |
| K | 19 | 0.9 | 17.7 | 7.0 | 4.9 |
| Chloride | 19 | 4.2 | 305.1 | 61.3 | 71.6 |
| Sulfate | 19 | 14.8 | 618.7 | 123.0 | 143.8 |
| Bicarbonate | 14 | 87.1 | 830.0 | 441.3 | 255.1 |
| | **Sandstone Bedrock** | | | | |
| TDS | 11 | 55.4 | 1352.6 | 301.3 | 355.8 |
| DO | 9 | 0.5 | 98.2 | 28.3 | 35.0 |
| pH | 11 | 6.7 | 8.3 | 7.7 | 0.6 |
| Total_Nitrogen | 9 | 0.1 | 8.9 | 2.5 | 3.1 |
| Orthophosphate | 11 | 0.0 | 0.9 | 0.2 | 0.3 |
| Phosphorus | 9 | 0.0 | 0.4 | 0.1 | 0.1 |
| Ca | 11 | 6.7 | 218.5 | 42.0 | 59.5 |
| Mg | 11 | 2.0 | 58.8 | 15.0 | 16.0 |
| Na | 11 | 5.8 | 102.8 | 34.8 | 27.0 |
| K | 11 | 0.3 | 15.3 | 4.6 | 4.6 |
| Chloride | 11 | 4.1 | 132.3 | 37.1 | 42.2 |
| Sulfate | 11 | 16.8 | 563.9 | 119.8 | 160.0 |
| Bicarbonate | 10 | 11.7 | 326.8 | 89.5 | 92.6 |
| | **Crystalline Bedrock** | | | | |
| TDS | 17 | 28.4 | 1397.3 | 230.3 | 308.5 |
| DO | 17 | 3.6 | 171.4 | 50.4 | 42.4 |
| pH | 17 | 6.6 | 8.3 | 7.5 | 0.5 |
| Total_Nitrogen | 17 | 0.2 | 19.7 | 3.3 | 4.5 |
| Orthophosphate | 16 | 0.0 | 13.5 | 1.0 | 3.3 |
| Phosphorus | 17 | 0.0 | 3.2 | 0.3 | 0.8 |
| Ca | 17 | 4.0 | 271.6 | 40.2 | 61.1 |
| Mg | 17 | 1.0 | 39.9 | 9.0 | 8.9 |
| Na | 17 | 3.4 | 105.5 | 21.4 | 24.9 |
| K | 17 | 0.9 | 57.3 | 6.6 | 13.2 |
| Chloride | 17 | 1.8 | 92.4 | 16.0 | 25.1 |
| Sulfate | 17 | 0.3 | 197.4 | 30.8 | 45.8 |
| Bicarbonate | 14 | 37.8 | 247.3 | 128.5 | 62.8 |

The area differences are not statistically different (Table 7b). Nevertheless, these results suggest that there may be a trend of increasing water constituent yields from the period of 1970-1980 to 1980-1990 in the project watersheds in the United States.

### Table 13b. Average annual ion yields (kg/ha/yr) in watersheds of each bedrock type from Peters 1984 and the current study (complied for the corresponding watersheds only).

| | Na | K | Mg | Ca | Chloride | Sulfate | Bicarbonate | Dissolved Solids |
|---|---|---|---|---|---|---|---|---|
| | | | | Limestone | | | | |
| This study | 43.6 | 6.99 | 36.0 | 144.0 | 61.3 | 123.0 | 441.3 | 722.6 |
| Peters 1984 | 41.1 | 7.01 | 36.1 | 145.6 | 59.1 | 123.5 | 439.1 | 894.5 |
| Changes (%) | 6.0 | -0.3 | -0.1 | -1.1 | 3.7 | -0.4 | 0.5 | -19.2 |
| | | | | Sandstone | | | | |
| This study | 34.7 | 4.9 | 14.4 | 42.1 | 40.2 | 109.3 | 96.0 | 306.3 |
| Peters 1984 | 19.7 | 4.3 | 12.9 | 30.1 | 23.6 | 76.9 | 61.0 | 227.6 |
| Changes (%) | 76.4 | 14.6 | 11.9 | 40.2 | 70.7 | 42.2 | 57.3 | 34.6 |
| | | | | Crystalline | | | | |
| This study | 21.4 | 6.6 | 9.0 | 40.2 | 16.0 | 30.8 | 128.5 | 230.3 |
| Peters 1984 | 15.9 | 3.2 | 6.8 | 26.1 | 11.8 | 21.2 | 122.7 | 198.4 |
| Changes (%) | 34.4 | 102.8 | 32.6 | 54.2 | 35.6 | 45.4 | 4.8 | 16.1 |

### Table 13c. Nonparametric test for the differences between stream water ion yields of Peters 1984 and the current study.

| | Limestone | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Na | K | Mg | Ca | Chloride | Sulfate | Bicarbonate | Dissolved Solids |
| Mann-Whitney U | 158.50 | 175.0 | 169.0 | 178.5 | 160.5 | 167.0 | 94.0 | 148.0 |
| Wilcoxon W | 348.50 | 365.0 | 359.0 | 368.5 | 350.5 | 357.0 | 214.0 | 338.0 |
| Z | -.642 | -.161 | -.336 | -.058 | -.584 | -.394 | -.480 | -.949 |
| Asymp. Sig. (2-tailed) | .521 | .872 | .737 | .953 | .559 | .693 | .631 | .343 |
| Exact Sig. [2*(1-tailed Sig.)] | .525 | .885 | .751 | .954 | .563 | .708 | .652 | .354 |
| | Sandstone | | | | | | | |
| Mann-Whitney U | 28.5 | 39.0 | 37.5 | 32.0 | 39.0 | 26.0 | 29.0 | 34.0 |
| Wilcoxon W | 83.5 | 94.0 | 92.5 | 87.0 | 94.0 | 81.0 | 74.0 | 89.0 |
| Z | -1.626 | -.832 | -.945 | -1.361 | -.832 | -1.814 | -1.015 | -1.209 |
| Asymp. Sig. (2-tailed) | .104 | .406 | .345 | .173 | .406 | .070 | .310 | .226 |
| Exact Sig. [2*(1-tailed Sig.)] | .105 | .436 | .353 | .190 | .436 | .075 | .340 | .247 |
| | Crystalline | | | | | | | |
| Mann-Whitney U | 119.0 | 125.0 | 128.0 | 131.50 | 137.0 | 130.0 | 91.5 | 118.5 |
| Wilcoxon W | 272.0 | 278.0 | 281.0 | 284.5 | 290.0 | 283.0 | 196.5 | 271.5 |
| Z | -.878 | -.672 | -.568 | -.448 | -.259 | -.499 | -.299 | -.896 |
| Asymp. Sig. (2-tailed) | .380 | .501 | .570 | .654 | .796 | .617 | .765 | .370 |
| Exact Sig. [2*(1-tailed Sig.)] | .394 | .518 | .586 | .658 | .812 | .634 | .769 | .375 |

The average annual yield of sulfate in sandstone watersheds and of potassium in crystalline watersheds are good examples of demonstrating the increasing trend of some stream water ion yields in the project watersheds (Figure 4 and Figure 5). These increases might be due to the effects of human and industrial activities such as waste, fertilizers, smoke, and water treatment plants.

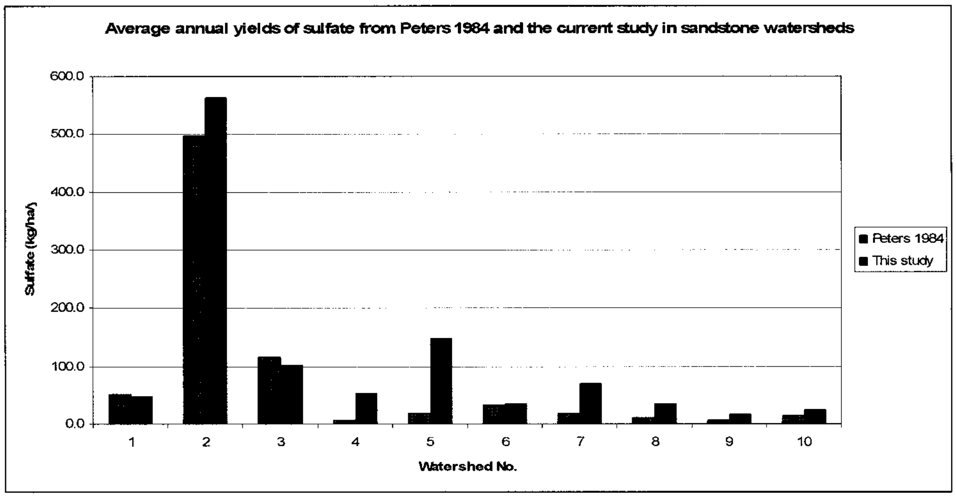


**Figure 4. Comparison of average annual sulfate yields (kg/ha/yr) between Peters 1984 and the current research in limestone watersheds.**

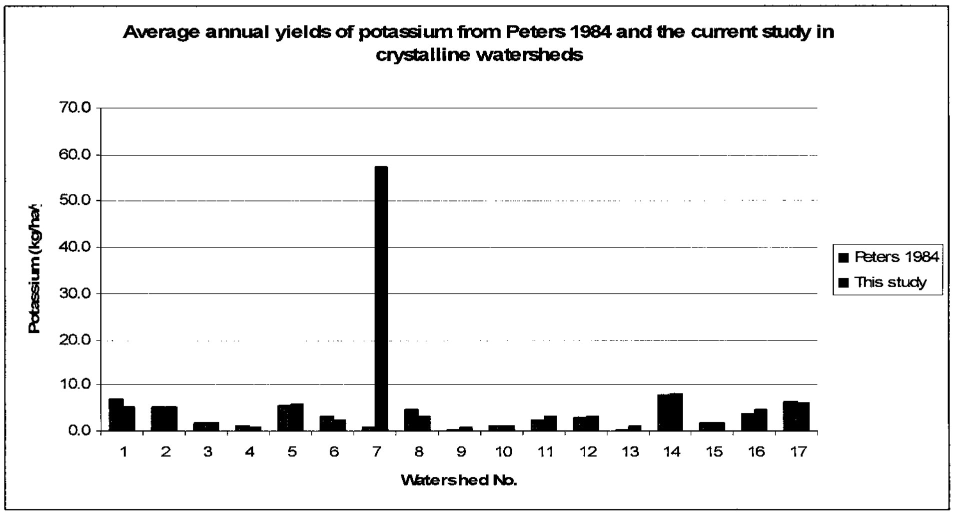


**Figure 5. Comparison of average annual potassium yields (kg/ha/yr) between Peters 1984 and the current research in crystalline watersheds.**

### *1. 4. Kolomogorov-Smirnov Tests:*

The One-Sample Kolmogorov-Smirnov (KS) test is used to test the null hypothesis that a sample comes from a particular distribution (e.g. normal, uniform, Poisson, or exponential). In this study, the null hypothesis was that a sample comes from a normal distribution. Since the sample size (the number of watersheds) of each bedrock type was small, the KS test was conducted for the combined data set of 47 watersheds.

For the stream water ion yield, the results (Table 14) indicated that almost all stream water ion yields were non-normal, except for pH ($p = 0.758$) and DO ($p = 0.509$). Magnesium, sodium, and bicarbonate were almost close to normality (p approximately equals to 0.05). However, since the data size of 47 was still small, the normality test might be biased.

**Table 14. Kolomogorov-Smirnov Test for the stream ion yields combined from the three bedrocks (limestone, sandstone, and crystalline).**

| | | TDS | DO | pH | Total_ Nitrogen | Ortho-phosphate | Phos-phorus | Ca | Mg | Na | K | Cl | $SO_4$ | Bicar-bonate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | | 47 | 45 | 47 | 42 | 46 | 45 | 47 | 47 | 47 | 47 | 47 | 47 | 38 |
| Normal Parameters (a,b) | Mean | 446.0 | 41.8 | 7.7 | 3.5 | .7 | .2 | 82.6 | 21.3 | 33.5 | 6.3 | 39.2 | 88.9 | 233.4 |
| | Std. Deviation | 415.0 | 32.3 | .5 | 3.7 | 2.0 | .5 | 82.2 | 19.9 | 34.8 | 8.7 | 54.9 | 127.9 | 229.0 |
| Most Extreme Differences | Absolute | .25 | .122 | .098 | .213 | .368 | .362 | .257 | .187 | .193 | .247 | .248 | .259 | .216 |
| | Positive | .25 | .122 | .069 | .213 | .347 | .362 | .257 | .187 | .185 | .234 | .224 | .259 | .216 |
| | Negative | -.16 | -.100 | -.098 | -.181 | -.368 | -.343 | -.169 | -.154 | -.193 | -.247 | -.248 | -.244 | -.166 |
| Kolmogorov-Smirnov Z | | 1.72 | .82 | .67 | 1.38 | 2.50 | 2.43 | 1.76 | 1.28 | 1.33 | 1.69 | 1.70 | 1.78 | 1.33 |
| Asymp. Sig. (2-tailed) | | **.006** | .509 | .758 | **.044** | **.000** | **.000** | **.004** | .074 | .060 | **.007** | **.006** | **.004** | .058 |

a Test distribution is Normal.
b Calculated from data.

For the watershed characteristics normality test (Table 15), only ***Precipitation*** and ***Slope*** were normal (p = 0.651 and p = 0.212, respectively). All other factors were significant statistically, meaning that they are not normal.

**Table 15. Kolomogorov-Smirnov Test for the environmental factors combined from the three bedrocks (limestone, sandstone, and crystalline).**

| | | Water | Developed | Forest | Agriculture | Precipitation | Population | Slope | Area | Bedrock |
|---|---|---|---|---|---|---|---|---|---|---|
| N | | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 | 47 |
| Normal Parameters (a,b) | Mean | 212.1 | 139.0 | 2856.7 | 2977.2 | 36.0 | 73.4 | 8.6 | 6185.0 | 2.0 |
| | Std. Deviation | 356.7 | 230.7 | 5049.1 | 6689.4 | 13.1 | 188.8 | 7.5 | 9856.3 | .9 |
| Most Extreme Differences | Absolute | .276 | .273 | .301 | .328 | .107 | .351 | .155 | .266 | .265 |
| | Positive | .233 | .258 | .301 | .321 | .107 | .291 | .155 | .259 | .265 |
| | Negative | -.276 | -.273 | -.286 | -.328 | -.098 | -.351 | -.146 | -.266 | -.243 |
| Kolmogorov-Smirnov Z | | 1.893 | 1.874 | 2.065 | 2.250 | .736 | 2.405 | 1.059 | 1.820 | 1.817 |
| Asymp. Sig. (2-tailed) | | **.002** | **.002** | **.000** | **.000** | .651 | **.000** | .212 | **.003** | **.003** |

a Test distribution is Normal.

## 2. Effects of bedrock types on stream ion yields

The descriptive statistics of average annual stream water ion yields of each bedrock type (kg/ha/yr) are displayed (Table 13a and Appendix 3). Stream water ion yields of limestone bedrock are quite different from those of the other two (Table 16), including total dissolved solids, orthophosphate, calcium, magnesium, and bicarbonate ($p < 0.05$), and consistent with Peters 1984. However, sandstone and crystalline ion yields are not much different, except for chloride (p = 0.033) and sulfate (p = 0.002). These results in combination with the results in Table 13a, suggest that yield of total dissolved solids of limestone watersheds is about 2.4 times greater than sandstone watersheds and 3.1 times greater than crystalline watersheds.

The ratio of ion yields in limestone watersheds to those in crystalline watersheds and in sandstone watersheds is about: bicarbonate, 5:1.4:1; orthophosphate, 3.5:5:1; calcium, 3.4:1:1, and magnesium, 2.4:0.6:1 (Table 13a). These ratio results are consistent with Peters 1984.

Chloride and sulfate concentrations from limestone and sandstone watersheds are significantly different from those of crystalline watersheds, and the yield of chloride from limestone bedrock is about 4 times greater than that from crystalline bedrock, and sandstone is about 1.4 times greater than crystalline bedrock. The yields of sulfate from limestone and sandstone bedrocks are 4 times greater than that for crystalline bedrock (Table 13a). This may be related to both climate and farming practices (Peters, 1984). When compared to the crystalline watersheds, the limestone watersheds generally have a higher population density and higher annual precipitation (Table 10 and Table 11), which consequently bring in larger quantities of sulfate either by atmospheric deposition or fertilizer application and other farming activities (Peters, 1984). However, the sulfate yields from limestone watersheds are not statistical different from those of sandstone watersheds like Peters 1984. That might be due to other factors affecting sulfate, such as, developed land, forest, and agriculture.

The difference test between bicarbonate concentrations from sandstone and crystalline watersheds resulted in a p-value of 0.064 (Table 16). Dissolved oxygen is not different between limestone and crystalline watersheds ($p = 0.684$) while there is just only slight difference between dissolved oxygen yields of limestone and sandstone ($p = 0.105$) or sandstone and crystalline watersheds ($p = 0.085$). Potassium is almost different between limestone and sandstone (p = 0.103) or limestone and crystalline watersheds (p = 0.061). Sodium yields of limestone and crystalline or of sandstone and crystalline bedrocks are also almost different (p = 0.057 and 0.100, respectively). The statistical significance of total nitrogen yields of limestone and sandstone watersheds is slightly larger than 0.05 (p = 0.074).

**Table 16. Nonparametric Test Statistics for the ion yield differences between two bedrock types of the three bedrock studied.**

| | **TDS** | **DO** | **pH** | **Total Nitrogen** | $PO_4$ | **Phosphorus** | **Ca** | **Mg** | **Na** | **K** | **Cl** | **Sulfate** | **Bicar-bonate** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Limestone vs. Sandstone** | | | | | | | | | | | | |
| Mann-Whitney U | 28.0 | 52.0 | 88.5 | 40.0 | 39.0 | 45.0 | 23.0 | 32.0 | 97.0 | 66.0 | 73.0 | 91.0 | 9.0 |
| Wilcoxon W | 94.0 | 97.0 | 278.5 | 85.0 | 105.0 | 90.0 | 89.0 | 98.0 | 163.0 | 132.0 | 139.0 | 157.0 | 64.0 |
| Z | -3.29 | -1.648 | -.691 | -1.812 | -2.819 | -1.99 | -3.507 | -3.1 | -.323 | -1.657 | -1.356 | -.581 | -3.572 |
| Asymp. Sig. (2-tailed) | .001 | .099 | .490 | .070 | .005 | .046 | .000 | .002 | .747 | .098 | .175 | .561 | .000 |
| Exact Sig. [2*(1-tailed Sig.)] | **.001** | .105 | .497 | .074 | **.004** | **.048** | **.000** | **.001** | .767 | .103 | .185 | .582 | **.000** |
| | **Limestone vs. Crystalline** | | | | | | | | | | | | |
| Mann-Whitney U | 26.0 | 148.0 | 128.0 | 102.0 | 67.0 | 106.0 | 34.0 | 33.0 | 101.0 | 102.0 | 52.0 | 45.0 | 20.0 |
| Wilcoxon W | 179.0 | 338.0 | 281.0 | 255.0 | 203.0 | 259.0 | 187.0 | 186.0 | 254.0 | 255.0 | 205.0 | 198.0 | 125.0 |
| Z | -4.294 | -.428 | -1.065 | -1.225 | -2.815 | -1.759 | -4.04 | -4.1 | -1.917 | -1.885 | -3.470 | -3.692 | -3.584 |
| Asymp. Sig. (2-tailed) | .000 | .669 | .287 | .221 | .005 | .079 | .000 | .000 | .055 | .059 | .001 | .000 | .000 |
| Exact Sig. [2*(1-tailed Sig.)] | **.000** | .684 | .300 | .231 | **.004** | .081 | **.000** | **.000** | .057 | .061 | **.000** | **.000** | **.000** |
| | **Sandstone vs. Crystalline** | | | | | | | | | | | | |
| Mann-Whitney U | 66.0 | 44.0 | 69.5 | 60.0 | 72.0 | 67.0 | 91.0 | 65.0 | 58.0 | 90.0 | 48.0 | 30.0 | 38.0 |
| Wilcoxon W | 219.0 | 89.0 | 222.5 | 105.0 | 138.0 | 112.0 | 157.0 | 218.0 | 211.0 | 156.0 | 201.0 | 183.0 | 93.0 |
| Z | -1.294 | -1.752 | -1.132 | -.889 | -.790 | -.512 | -.118 | -1.3 | -1.670 | -.165 | -2.140 | -2.987 | -1.874 |
| Asymp. Sig. (2-tailed) | .196 | .080 | .257 | .374 | .430 | .609 | .906 | .180 | .095 | .869 | .032 | .003 | .061 |
| Exact Sig. [2*(1-tailed Sig.)] | .208 | .085 | .264 | .396 | .451 | .634 | .926 | .191 | .100 | .890 | **.033** | **.002** | .064 |

Phosphorus yield of sandstone is the same as that of crystalline, but is close to the difference level between limestone vs. crystalline ($p = 0.081$) and is significantly different between limestone vs. sandstone watersheds ($p = 0.048$). This indicates that the yield of phosphorus of limestone watersheds is 2 times greater than that of sandstone watersheds.

The pH values of the three project bedrocks showed no difference between bedrock type.

Ion yields for limestone and sandstone bedrocks are more closely related to the environmental factors than those in crystalline bedrock (Table 17). Most of the ion yields in crystalline watersheds have no relation with the environmental factors, except for DO, pH, and total nitrogen. In the three bedrock types, the present of significant environmental factors of the same ion yield (and the corresponding adjusted R square) is almost different from this bedrock compared to the other two. For example, in limestone bedrock, bicarbonate is related to precipitation, in sandstone, bicarbonate is related to population and water, and in crystalline, bicarbonate is not related to any factor (Table 17). Precipitation is significant in the relation with pH in all three bedrocks. It is also significant in the relation with dissolved oxygen and total nitrogen in sandstone and crystalline bedrocks (Table 17). In addition, the slope of a regression of an ion yield on any specific environmental factors (if significant) is the same of the three bedrocks. pH, for example, has a negative relation with precipitation for all three bedrocks (Table 17).

**Table 17. Summary of the multiple linear regression results of the relationship between Log of ion yield and Log of the environmental factors in the three bedrocks.**

| Ion yield | Limestone | Sandstone | Crystalline |
|---|---|---|---|
| TDS | ND | Population(+), Water(+); (.530) | ND |
| DO | Slope(+); (.385)* | Precipitation(+); (.821) | Precipitation(+); (.447) |
| pH | Precipitation(-); (.611) | Precipitation(-); (.932) | Precipitation(-); (.379) |
| Total Nitrogen | ND | Precipitation(+); (.761) | Precipitation(+); (.290) |
| Orthophosphate | Developed(+), Water(-); (.388) | ND | ND |
| Phosphorus | Developed(+); (.261) | Precipitation(+); (.447) | ND |
| Ca | ND | ND | ND |
| Mg | Precipitation(-), Slope(+), Agriculture(+); (.834) | Population(+); (.307) | ND |
| Na | Population(+), Slope(+), Precipitation(-); (.740) | Water(+); (.392) | ND |
| K | Slope(+), Population(+); (.530) | Population(+); (.540) | ND |
| Chloride | Population(+); (.585) | ND | ND |
| Sulfate | Forest(+), Precipitation(-); (.354) | ND | ND |
| Bicarbonate | Precipitation(-); (.311) | Population(+), Water(+); (.636) | ND |

*Notations in Table 17:*

***ND: No regression coefficient detected to be significant***

***Precipitation(-): Significant negative relation (between the corresponding Log of ion yield and Log of precipitation)***

***Population(+): Significant positive relation (between the corresponding Log of ion yield and Log of population)***

***Significant level (Alpha) = 0.05***

***(.385): Adjusted R square of the regression***

To know how bedrock types affect the relation between the environmental factors and the selected ion yields, it is also necessary to distinguish the slope coefficient of the same predictor (environmental factor) for the same ion yield regression among bedrock types. As mentioned earlier, precipitation proved to be significant in relation to pH in all three bedrocks, and with dissolved oxygen and total nitrogen in sandstone and crystalline bedrocks.

To test for effects of bedrocks on significant coefficient slope of precipitation on pH in the three bedrock types, the data were recompiled (Table 18) and regression analysis prepared (Table 19).

**Table 18. Data for the test of effects of bedrocks on significant coefficient slope of Precipitation on pH of the three bedrock types (except for Sandstone, all other data are in log-transformed)**

| Site No. | Station No. | pH | Precipitation | Limestone | Sandstone | Limestone x Precipitation | Sandstone x Precipitation |
|---|---|---|---|---|---|---|---|
| 1 | 02313000 | 0.88 | 1.71 | 1 | 0 | 1.71 | 0 |
| 2 | 02320500 | 0.89 | 1.71 | 1 | 0 | 1.71 | 0 |
| 3 | 02329000 | 0.85 | 1.73 | 1 | 0 | 1.73 | 0 |
| 4 | 02368000 | 0.85 | 1.78 | 1 | 0 | 1.78 | 0 |
| 5 | 03085000 | 0.88 | 1.66 | 1 | 0 | 1.66 | 0 |
| 6 | 03254000 | 0.89 | 1.65 | 1 | 0 | 1.65 | 0 |
| 7 | 03290500 | 0.89 | 1.67 | 1 | 0 | 1.67 | 0 |
| 8 | 03425000 | 0.88 | 1.73 | 1 | 0 | 1.73 | 0 |
| 9 | 03543005 | 0.88 | 1.72 | 1 | 0 | 1.72 | 0 |
| 10 | 03571850 | 0.88 | 1.73 | 1 | 0 | 1.73 | 0 |
| 11 | 03593005 | 0.87 | 1.74 | 1 | 0 | 1.74 | 0 |
| 12 | 04045500 | 0.89 | 1.52 | 1 | 0 | 1.52 | 0 |
| 13 | 04057004 | 0.89 | 1.51 | 1 | 0 | 1.51 | 0 |
| 14 | 04108690 | 0.91 | 1.55 | 1 | 0 | 1.55 | 0 |
| 15 | 04126520 | 0.9 | 1.51 | 1 | 0 | 1.51 | 0 |
| 16 | 04132052 | 0.91 | 1.5 | 1 | 0 | 1.5 | 0 |
| 17 | 04165500 | 0.91 | 1.49 | 1 | 0 | 1.49 | 0 |
| 18 | 08103900 | 0.9 | 1.49 | 1 | 0 | 1.49 | 0 |
| 19 | 13018300 | 0.92 | 1.54 | 1 | 0 | 1.54 | 0 |
| 20 | 01545600 | 0.84 | 1.62 | 0 | 1 | 0 | 1.62 |
| 21 | 03049625 | 0.87 | 1.64 | 0 | 1 | 0 | 1.64 |
| 22 | 03237280 | 0.85 | 1.61 | 0 | 1 | 0 | 1.61 |
| 23 | 06452000 | 0.91 | 1.25 | 0 | 1 | 0 | 1.25 |
| 24 | 06478500 | 0.91 | 1.29 | 0 | 1 | 0 | 1.29 |
| 25 | 07026000 | 0.83 | 1.72 | 0 | 1 | 0 | 1.72 |
| 26 | 07231500 | 0.91 | 1.29 | 0 | 1 | 0 | 1.29 |
| 27 | 07234000 | 0.92 | 1.24 | 0 | 1 | 0 | 1.24 |
| 28 | 07237500 | 0.91 | 1.28 | 0 | 1 | 0 | 1.28 |
| 29 | 07245000 | 0.89 | 1.35 | 0 | 1 | 0 | 1.35 |
| 30 | 06438000 | 0.92 | 1.23 | 0 | 1 | 0 | 1.23 |
| 31 | 01059400 | 0.85 | 1.66 | 0 | 0 | 0 | 0 |
| 32 | 01066000 | 0.82 | 1.7 | 0 | 0 | 0 | 0 |
| 33 | 04014500 | 0.88 | 1.46 | 0 | 0 | 0 | 0 |
| 34 | 05124480 | 0.84 | 1.46 | 0 | 0 | 0 | 0 |
| 35 | 06623800 | 0.89 | 1.67 | 0 | 0 | 0 | 0 |
| 36 | 07311200 | 0.86 | 1.47 | 0 | 0 | 0 | 0 |
| 37 | 08431700 | 0.86 | 1.31 | 0 | 0 | 0 | 0 |
| 38 | 09352900 | 0.89 | 1.61 | 0 | 0 | 0 | 0 |
| 39 | 09430600 | 0.89 | 1.43 | 0 | 0 | 0 | 0 |
| 40 | 09508300 | 0.9 | 1.34 | 0 | 0 | 0 | 0 |
| 41 | 11264500 | 0.83 | 1.64 | 0 | 0 | 0 | 0 |
| 42 | 12447390 | 0.88 | 1.53 | 0 | 0 | 0 | 0 |
| 43 | 13169500 | 0.92 | 1.06 | 0 | 0 | 0 | 0 |
| 44 | 13331500 | 0.88 | 1.67 | 0 | 0 | 0 | 0 |
| 45 | 14048000 | 0.92 | 1.27 | 0 | 0 | 0 | 0 |
| 46 | 14103000 | 0.91 | 1.31 | 0 | 0 | 0 | 0 |
| 47 | 14113000 | 0.89 | 1.54 | 0 | 0 | 0 | 0 |

**Table 19. Model summary and the analysis of coefficients of the relationship between pH and precipitation in the combined data of the three bedrock types.**

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .631(a) | .398 | .385 | .022 |
| 2 | .799(b) | .638 | .622 | .017 |

a Predictors: (Constant), Precipitation
b Predictors: (Constant), Precipitation, Limestone

Coefficients(a)

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.033 | .028 | | 37.395 | .000 |
| | Precipitation | -.098 | .018 | -.631 | -5.454 | .000 |
| 2 | (Constant) | 1.086 | .024 | | 45.623 | .000 |
| | Precipitation | -.142 | .016 | -.909 | -8.715 | .000 |
| | Limestone | .031 | .006 | .564 | 5.401 | .000 |

a Dependent Variable: pH

Excluded Variables(c)

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Tolerance | VIF | Minimum Tolerance |
| 1 | Limestone | .564(a) | 5.401 | .000 | .631 | .756 | 1.323 | .756 |
| | Sandstone | -.149(a) | -1.210 | .233 | -.179 | .877 | 1.140 | .877 |
| | Limestone_Precipitation | .569(a) | 5.314 | .000 | .625 | .727 | 1.375 | .727 |
| | Sandstone_Precipitation | -.180(a) | -1.516 | .137 | -.223 | .926 | 1.080 | .926 |
| 2 | Sandstone | .037(b) | .356 | .724 | .054 | .772 | 1.296 | .665 |
| | Limestone_Precipitation | -.473(b) | -.377 | .708 | -.057 | .005 | 187.435 | .005 |
| | Sandstone_Precipitation | .015(b) | .143 | .887 | .022 | .794 | 1.260 | .648 |

a Dependent Variable: pH
b Predictors in the Model: (Constant), Precipitation, Limestone
c Dependent Variable: pH

As shown in Table 19, *constant, and coefficients of Precipitation* and *Limestone* are significant while *Sandstone, Limestone x Precipitation,* and *Sandstone x Precipitation* are not significant ($p > 0.05$). Thus, they are not presented in the output of the "stepwise" model selection). This means that only pH in limestone is different from crystalline.

However, the coefficient of limestone, which is small (0.031), can be considered negligible. Since the slopes of the product of *Limestone x Precipitation* and of *Sandstone x Precipitation* (as shown in Table 18) are not significant, no effect of bedrocks on the predictor (precipitation) for pH is found.

Second, to test for effects of bedrocks on significant coefficient slope of precipitation on dissolved solids, pH, and on total nitrogen in between sandstone and crystalline bedrock types, the data were recompiled as shown in Table 20. The results of this regression analysis are expressed in Table 21, 23, and 24, for dissolved solids, pH, and nitrogen, respectively.

**Table 20. Data for the test of effects of bedrocks on significant coefficient slopes of precipitation on DO, pH, and on total nitrogen in sandstone and crystalline bedrock types (except for Sandstone, all other data are in log-transformed).**

| Site No. | Station No. | pH | DO | total nitrogen | Precipitation | Sandstone | Sandstone x Precipitation |
|---|---|---|---|---|---|---|---|
| 1 | 01545600 | 0.84 | 1.81 | 0.51 | 1.62 | 1.00 | 1.62 |
| 2 | 03049625 | 0.87 | 1.99 | 0.95 | 1.64 | 1.00 | 1.64 |
| 3 | 03237280 | 0.85 | 1.54 | 0.31 | 1.61 | 1.00 | 1.61 |
| 4 | 06452000 | 0.91 | 0.6 | -0.07 | 1.25 | 1.00 | 1.25 |
| 5 | 06478500 | 0.91 | 0.4 | -0.51 | 1.29 | 1.00 | 1.29 |
| 6 | 07026000 | 0.83 | 1.66 | 0.78 | 1.72 | 1.00 | 1.72 |
| 7 | 07231500 | 0.91 | 0.6 | -0.24 | 1.29 | 1.00 | 1.29 |
| 8 | 07234000 | 0.92 | Nodata | Nodata | 1.24 | 1.00 | 1.24 |
| 9 | 07237500 | 0.91 | -0.29 | -0.88 | 1.28 | 1.00 | 1.28 |
| 10 | 07245000 | 0.89 | Nodata | Nodata | 1.35 | 1.00 | 1.35 |
| 11 | 06438000 | 0.92 | 0.27 | -0.31 | 1.23 | 1.00 | 1.23 |
| 12 | 01059400 | 0.85 | 1.68 | 0.56 | 1.66 | .00 | .00 |
| 13 | 01066000 | 0.82 | 1.98 | 0.46 | 1.7 | .00 | .00 |
| 14 | 04014500 | 0.88 | 1.7 | 0.42 | 1.46 | .00 | .00 |
| 15 | 05124480 | 0.84 | 1.49 | 0.12 | 1.46 | .00 | .00 |
| 16 | 06623800 | 0.89 | 1.8 | 0.55 | 1.67 | .00 | .00 |
| 17 | 07311200 | 0.86 | 1.14 | 0 | 1.47 | .00 | .00 |
| 18 | 08431700 | 0.86 | 2.23 | 1.29 | 1.31 | .00 | .00 |
| 19 | 09352900 | 0.89 | 1.68 | 0.69 | 1.61 | .00 | .00 |
| 20 | 09430600 | 0.89 | 1 | -0.36 | 1.43 | .00 | .00 |
| 21 | 09508300 | 0.9 | 0.93 | -0.37 | 1.34 | .00 | .00 |
| 22 | 11264500 | 0.83 | 1.9 | 0.52 | 1.64 | .00 | .00 |
| 23 | 12447390 | 0.88 | 1.82 | 0.65 | 1.53 | .00 | .00 |
| 24 | 13169500 | 0.92 | 0.56 | -0.63 | 1.06 | .00 | .00 |
| 25 | 13331500 | 0.88 | 1.93 | 0.53 | 1.67 | .00 | .00 |
| 26 | 14048000 | 0.92 | 1.01 | -0.26 | 1.27 | .00 | .00 |
| 27 | 14103000 | 0.91 | 1.36 | 0.04 | 1.31 | .00 | .00 |
| 28 | 14113000 | 0.89 | 1.69 | 0.52 | 1.54 | .00 | .00 |

### Table 21. Model summary and the analysis of coefficients of the relationship between DO and precipitation in the combined data of sandstone and crystalline bedrock types.

**Model Summary**

| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** |
|---|---|---|---|---|
| 1 | .754(a) | .568 | .550 | .433 |
| 2 | .832(b) | .693 | .666 | .373 |
| 3 | .871(c) | .759 | .726 | .338 |

a Predictors: (Constant), Precipitation
b Predictors: (Constant), Precipitation, Sandstone
c Predictors: (Constant), Precipitation, Sandstone, Sandstone_Precipitation

**Coefficients(a)**

| **Model** | | **Unstandardized Coefficients** | | **Standardized Coefficients** | **t** | **Sig.** |
|---|---|---|---|---|---|---|
| | | **B** | **Std. Error** | **Beta** | | |
| 1 | **(Constant)** | -2.558 | .696 | | -3.673 | .001 |
| | **Precipitation** | 2.652 | .472 | .754 | 5.619 | .000 |
| 2 | **(Constant)** | -2.201 | .611 | | -3.601 | .002 |
| | **Precipitation** | 2.521 | .409 | .716 | 6.163 | .000 |
| | **Sandstone** | -.473 | .155 | -.355 | -3.055 | .006 |
| 3 | **(Constant)** | -1.105 | .712 | | -1.551 | .135 |
| | **Precipitation** | 1.779 | .479 | .506 | 3.715 | .001 |
| | **Sandstone** | -3.166 | 1.109 | -2.378 | -2.855 | .009 |
| | **Sandstone_Precipitation** | 1.852 | .757 | 2.028 | 2.448 | .023 |

a Dependent Variable: DO

In Table 21 (model 3), *constant, Precipitation, Sandstone*, and the product of *Sandstone x Precipitation* are significant. This indicates that both the log-transformed value of DO in sandstone and the coefficient slope of precipitation of sandstone are different from those in crystalline bedrock. The differences between DO in sandstone and crystalline bedrocks are not significant (Table 16). That is contrary with the result in Table 21. The reason for that may be due to the collinearity of *sandstone* and *precipitation* with the product of *sandstone x precipitation.*

**Table 22. Model summary and the analysis of coefficients of the relationship between pH and precipitation in the combined data of sandstone and crystalline bedrock types.**

**Model Summary**

| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** |
|---|---|---|---|---|
| 1 | .797(a) | .635 | .621 | .019 |

a Predictors: (Constant), Precipitation

**Coefficients(a)**

| **Model** | | **Unstandardized Coefficients** | | **Standardized Coefficients** | **t** | **Sig.** |
|---|---|---|---|---|---|---|
| | | **B** | **Std. Error** | **Beta** | | |
| 1 | **(Constant)** | 1.082 | .030 | | 35.849 | .000 |
| | **Precipitation** | -.139 | .021 | -.797 | -6.723 | .000 |

a Dependent Variable: pH

The results (Table 22 and Table 23) indicate that there are no differences of pH and total nitrogen and the significant slope coefficients between sandstone and crystalline bedrocks because the slopes of *Sandstone* and *Sandstone x Precipitation* are not significant ($p > 0.05$) (therefore *Sandstone* and *Sandstone x Precipitation* are not presented in Table 22 and Table 23).

**Table 23. Model summary and the analysis of coefficients of the relationship between total nitrogen and precipitation in the combined data of sandstone and crystalline bedrock types.**

**Model Summary**

| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** |
|---|---|---|---|---|
| 1 | .712(a) | .506 | .486 | .383 |

a Predictors: (Constant), Precipitation

**Coefficients(a)**

| **Model** | | **Unstandardized Coefficients** | | **Standardized Coefficients** | **t** | **Sig.** |
|---|---|---|---|---|---|---|
| | | **B** | **Std. Error** | **Beta** | | |
| 1 | **(Constant)** | -2.826 | .615 | | -4.592 | .000 |
| | **Precipitation** | 2.069 | .417 | .712 | 4.961 | .000 |

a Dependent Variable: Total_Nitrogen

The effects of bedrock types on stream water ion yields and their relationship with other environmental factors are discussed in later sections.

## 3. Effects of the other environmental factors on stream water ion yields

### *3.1. Effects of land cover:*

The 21 class National Land Cover Dataset is grouped into 4 new categories, including ***Water, Developed, Forest,*** and ***Agriculture***. Therefore, the effects of land cover are analyzed in effects of the 4 categories separately.

### 3.1.1. Effects of Water:

As listed in Table 9, ***Water*** of each bedrock type occupied only very small percentages (1.8% in sandstone to 13.6% in limestone), averaging at 6.5 %. However, there are still some slope coefficients of this land cover type significant in regressions of orthophosphate and magnesium in limestone, in regression of total dissolved solids, sodium, and bicarbonate in sandstone bedrock. In more details, the equations for these relations are found in Table 24:

**Table 24. The presence of significant slope coefficients of *Water* in the regression equations in limestone and sandstone bedrocks ($p < 0.05$).**

| **Limestone** | | **$R^2$** | **Adjusted $R^2$** |
|---|---|---|---|
| Orthophosphate = $10^{(-0.696)} \times (\text{Developed})^{(0.417)} \times (\text{Water})^{(-0.20)}$ | (12) | 0.456 | 0.388 |
| **Sandstone** | | | |
| TDS = $10^{(1.59)} \times (\text{Population})^{(0.489)} \times (\text{Water})^{(0.091)}$ | (13) | 0.624 | 0.53 |
| Na = $10^{(1.272)} \times (\text{Water})^{(0.118)}$ | (14) | 0.453 | 0.392 |
| Bicarbonate = $10^{(0.777)} \times (\text{Population})^{(0.655)} \times (\text{Water})^{(0.129)}$ | (15) | 0.717 | 0.636 |

As shown in Table 24 and in Appendix 4 and 5, the slope coefficient of ***Water*** and its standardized coefficient are always smaller than those for other factors in a same equation. This indicates that ***Water*** is a less important predictor when compared to the others, yet more important than those factors found not significant in those equations.

In equation (12), when ***Water*** increases, orthophosphate decreases. This might be due to the effects of sorption by wetlands and/or the effects of dilution water on orthophosphate ion yields. In equations (13), (14), and (15) ***Water*** is directly proportional to TDS, Na (Figure 6), and bicarbonate, respectively. Those relations may be attributed to the contributions of ice/snow from high elevations and organic materials in the wetlands.

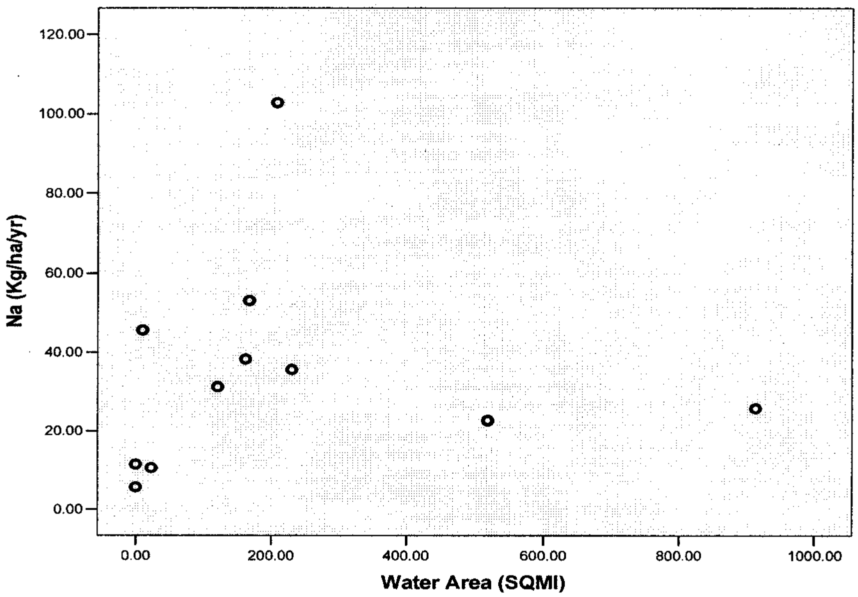


**Figure 6. Bivariate plot of annual yield of sodium (kg/ha/yr) and Water land cover (SQMI) in watersheds underlain by sandstone bedrock in the United States.**

### 3.1.2. Effects of Developed land cover:

As shown in Table 9, ***Developed*** land cover occupies even smaller percentages than ***Water,*** ranging from 1.0% in sandstone to 4.8% in limestone, and averaged at 3.1%. The slope coefficients of this land cover type are found to be significant in limestone bedrock only (Table 25).

**Table 25. The presence of significant slope coefficients of *Developed* in the regression equations in limestone bedrock ($p < 0.05$).**

| Limestone | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| Orthophosphate $=10^{(-0.696)} \times (\text{Developed})^{(0.417)} \times (\text{Water})^{(-0.20)}$ | (16) | 0.456 | 0.388 |
| Phosphorus $= 10^{(-1.202)} \times (\text{Developed})^{(0.186)}$ | (17) | 0.302 | 0.261 |

In equations (16) and (17), ***Developed*** is in direct relationship to orthophosphate and phosphorus (Figure 7). In equation (16), ***Developed*** has the highest standardized coefficient (Beta = 1.07) and partial or part correlations (0.666, or 0.658 respectively) (Appendix 4), meaning that ***Developed*** is the most important predictor for the prediction of orthophosphate in limestone bedrock. In equation (17) only ***Developed*** is a significant positive predictor for phosphorus. In both equations (16 and 17) when ***Developed*** increases, the corresponding dependent variable also increases, possibly due to human sewage surface disturbalances causing increased rates of erosion and accumulated downstream sediment which carries phosphate.

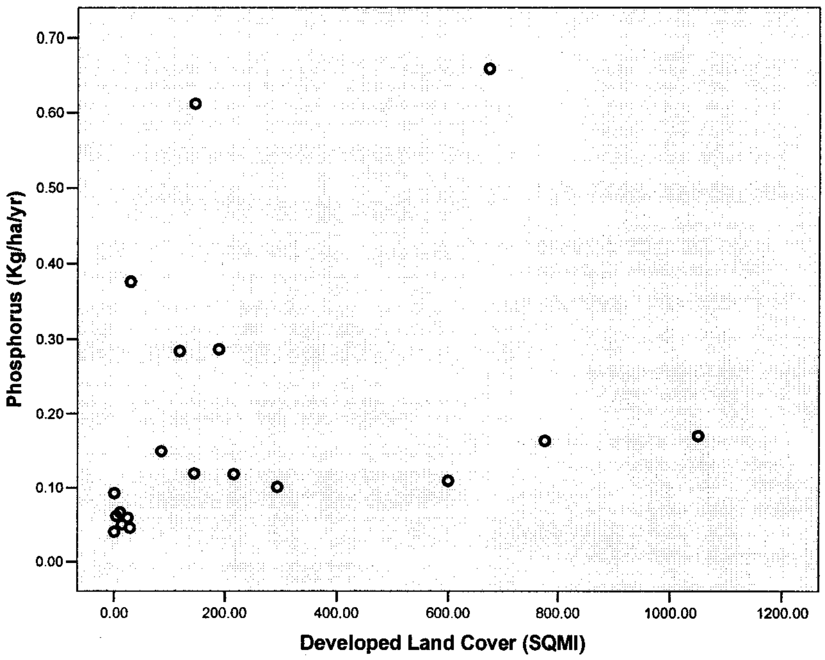


**Figure 7. Relationship between annual yield of phosphorus (kg/ha/yr) and *Developed* land cover (SQMI) in limestone watersheds in the United States.**

### 3.1.3. Effects of Forest:

***Forest*** cover is a highest percentage of all land cover types, ranging from about 35% in sandstone to approximately 80% in crystalline and averaged at 57.2% in all bedrock types (Table 9). However, this land cover is found to be just significant for a regression of sulfate in limestone bedrock (Table 26).

**Table 26. The presence of significant slope coefficients of *Forest* in the regression equations in limestone bedrock ($p < 0.05$).**

| Limestone | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| Sulfate = $10^{(4.251)} \times (\text{Forest})^{(0.372)} \times (\text{Precipitation})^{(-2.136)}$ | (18) | 0.426 | 0.354 |

***Forest***'s standardized coefficients and the corresponding partial or part correlations in equations (18) are also higher than those of ***Precipitation*** in a same equation (Appendix 4), indicating that ***Forest*** is the most important predictor for prediction of sulfate in limestone watersheds.

The role of ***Forest*** area in equation (18) in increasing yields of sulfate may be due to the increase of organic matter containing sulfate from the forest. Also shown in Appendix 4, ***Forest*** is highly correlated positively with ***Agriculture*** (R = 0.90). ***Forest*** and ***Agriculture*** plants can contribute a significant amount of sulfate into the stream system (because the organic matters come from forest and agriculture products can contain sulfate). Also, since ***Forest*** is correlated with ***Precipitation*** (R = 0.61, Appendix 4), the effect of acid rain (containing $SO_4$) is another possible reason for the increase of sulfate in the stream water.

### 3.1.4. Effects of Agriculture:

***Agriculture*** is the second largest size of the 4 land cover categories, varying from 24.7% in limestone to 62.3% in crystal, and averaged at 33.2% (Table 9). However, ***Agriculture*** is not very significant in ion yield predictions, except for magnesium in limestone bedrock (Table 27). In equation (19), the beta value of ***Agriculture*** is 0.50, which is smaller than that of ***Slope*** (0.63) and ***Precipitation*** (-1.13) (Appendix 4). This means that ***Agriculture*** is the least important predictor for the prediction of Mg in limestone bedrock.

**Table 27. The presence of significant slope coefficients of *Agriculture* in the regression equations in limestone bedrock ($p < 0.05$).**

| Limestone | $R^2$ | Adj. $R^2$ |
|---|---|---|
| $Mg = 10^{(6.403)} \times (\text{Precipitation})^{(-3.437)} \times (\text{Slope})^{(0.397)} \times (\text{Agriculture})^{(0.170)}$ (19) | 0.862 | 0.834 |

In summary, land cover types may have some profound effects on ion yields, especially for orthophosphate, phosphorus, and magnesium in limestone bedrock. These effects are somehow reduced in sandstone and are not significant in crystalline bedrocks. The reasons for those phenomena are not clear and need more study.

### 3.2. *Effects of Population:*

The effects of ***Population*** are defined and quantified by the influence of population density (people/SQMI). In crystalline watersheds, ***Population*** is found of no significant contribution, but in sandstone it is significant for estimation of TDS, magnesium, potassium, and bicarbonate. In limestone watersheds, it is significant for estimation of sodium, potassium, and chloride (Table 28).

***Population*** is always in direct proportion to the dependent variable corresponding to its contribution (Figure 8 and Figure 9). These might be due to human activities of road salt releases, chemical uses or fertilizer application, and farming activities.

**Table 28. The presence of significant slope coefficients of *Population* in the regression equations in limestone and sandstone bedrocks ($p < 0.05$).**

| **Limestone** | $R^2$ | Adj. $R^2$ |
|---|---|---|
| $Na = 10^{(2.645)} \times (Population)^{(0.619)} \times (Slope)^{(0.388)} \times (Precipitation)^{(-1.533)}$ (20) | 0.783 | 0.740 |
| $K = 10^{(-0.021)} \times (Slope)^{(0.399)} \times (Population)^{(0.299)}$ (21) | 0.582 | 0.530 |
| $Chloride = 10^{(0.435)} \times (Population)^{(0.634)}$ (22) | 0.608 | 0.585 |
| **Sandstone** | | |
| $TDS = 10^{(1.59)} \times (Population)^{(0.489)} \times (Water)^{(0.091)}$ (23) | 0.624 | 0.530 |
| $Mg = 10^{(0.386)} \times (Population)^{(0.468)}$ (24) | 0.376 | 0.307 |
| $K = 10^{(-0.416)} \times (Population)^{(0.672)}$ (25) | 0.586 | 0.540 |
| $Bicarbonate = 10^{(0.777)} \times (Population)^{(0.665)} \times (Water)^{(0.129)}$ (26) | 0.717 | 0.636 |

In equations (20), the beta value of ***Population*** is 0.877 which is almost double from that of ***Slope*** (0.476) and ***Precipitation*** (-0.390) (Appendix 4). Likewise, the values of partial and part correlations of ***Population*** are also the highest compared to those of ***Slope*** and ***Precipitation*** (Appendix 4). In equation (23), the values of beta, partial, and part correlations of ***Population*** (0.84, 0.78, and 0.763, respectively) are higher than those of ***Water*** (Appendix 5). Those comparisons are also likely to occur in equation (26). These results suggest that ***Population*** is the most important predictor in those regressions.

The positive effects of population density on sodium and chloride (equation 20 and equation 22) in this study are compatible with Peters 1984. However, the regression correlations in this study, which are 0.74 (between population and sodium) and 0.78 (between population and chloride) (Appendix 4) are much higher than the corresponding regression correlations of Peters 1984, which were only 0.36 and 0.25, respectively.

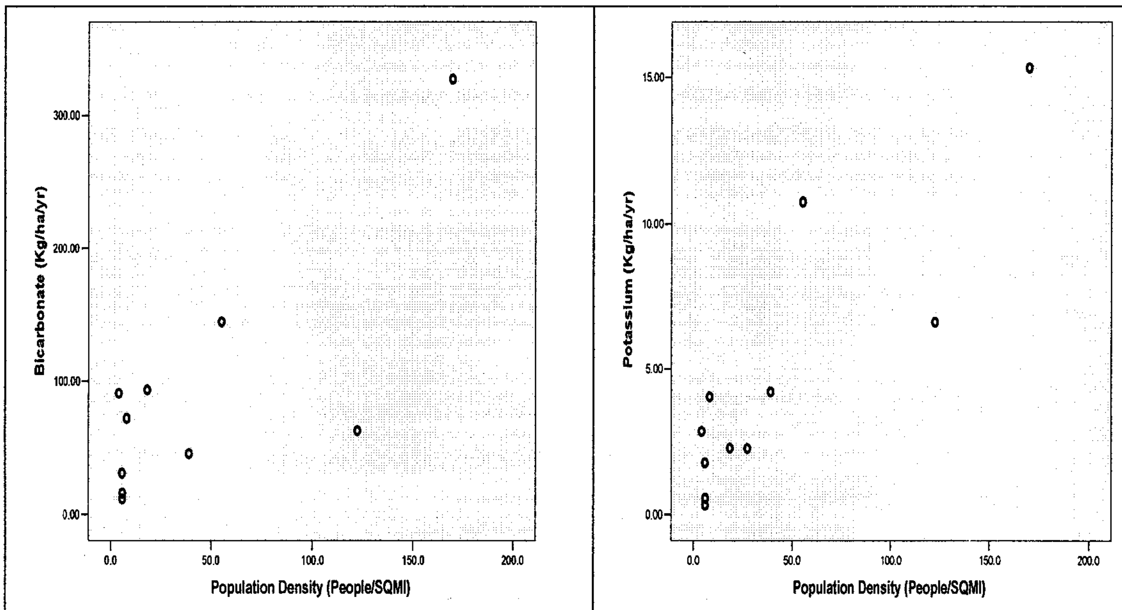


**Figure 8. Effects of population density (people/SQMI) on annual yields of bicarbonate and potassium (kg/ha/yr) in sandstone watersheds in the United States.**

The effects of population density on ion yields in this study are not consistent with what was found in Peters 1984. Population density was an ineffective indicator of all regressions of constituents except for sodium and chloride (Peters, 1984).

In this study, population density is significant in the regressions not only for sodium and chloride (in limestone watersheds), but also for total dissolved solids, magnesium, potassium, and bicarbonate (in sandstone watersheds) (Table 28). Furthermore, the regression correlations in these equations are much higher than those of

Peters 1984, ranging from 0.6 to 0.77 (Table 28 and Appendix 4, Appendix 5). Again, these results indicate that population density is important for estimations of these ion yields.

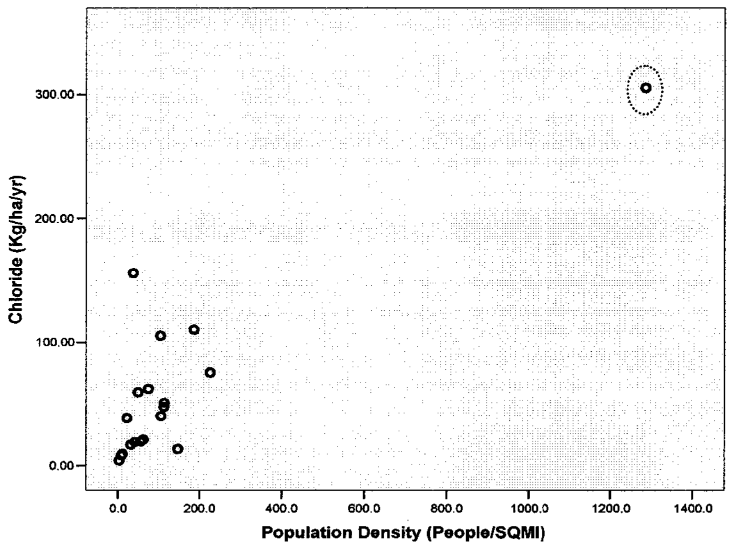


**Figure 9. Effects of population density (people/SQMI) on annual yields of chloride (kg/ha/yr) in limestone watersheds in the United States.**

*Regression equation with the farthest point resulted in a slight larger* $R^2$ *(0.61) compared to not taking that point into account (0.50). Since that point does not affects regression very much, it is kept in the regression model (equation 22, Table 28).*

### *3.3. Effects of Precipitation:*

Unlike land cover types and ***Population, Precipitation*** is statistically significant in all three bedrocks, especially for the estimations of DO, pH, and total nitrogen (Table 17). The significant contributions of ***Precipitation*** to the regressions of dependent variables are shown in Table 29.

**Table 29. The presence of significant slope coefficients of *Precipitation* in the regression equations in limestone, sandstone, and crystalline bedrocks ($p < 0.05$).**

| **Limestone** | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| pH = $10^{(1.139)}$ x (Precipitation)$^{(-0.155)}$ | (27) | 0.633 | 0.611 |
| Mg = $10^{(6.403)}$ x (Precipitation)$^{(-3.437)}$ x (Slope)$^{(0.397)}$ x (Agriculture)$^{(-0.170)}$ | (28) | 0.862 | 0.834 |
| Na = $10^{(2.645)}$ x (Population)$^{(0.619)}$ x (Slope)$^{(0.388)}$ x (Precipitation)$^{(-1.533)}$ | (29) | 0.783 | 0.740 |
| Sulfate = $10^{(4.251)}$ x (Forest)$^{(0.372)}$ x (Precipitation)$^{(-2.136)}$ | (30) | 0.426 | 0.354 |
| Bicarbonate = $10^{(5.407)}$ x (Precipitation)$^{(-1.737)}$ | (31) | 0.364 | 0.311 |
| **Sandstone** | | | |
| DO = $10^{(-4.271)}$ x (Precipitation)$^{(3.631)}$ | (32) | 0.843 | 0.821 |
| pH = $10^{(1.137)}$ x (Precipitation)$^{(-0.176)}$ | (33) | 0.939 | 0.932 |
| Total nitrogen = $10^{(-3.790)}$ x (Precipitation)$^{(2.678)}$ | (34) | 0.791 | 0.761 |
| Phosphorus = $10^{(-4.22)}$ x (Precipitation)$^{(2.033)}$ | (35) | 0.516 | 0.447 |
| **Crystalline** | | | |
| DO = $10^{(-1.105)}$ x (Precipitation)$^{(1.779)}$ | (36) | 0.482 | 0.477 |
| pH = $10^{(1.036)}$ x (Precipitation)$^{(-0.180)}$ | (37) | 0.418 | 0.379 |
| Total nitrogen = $10^{(-2.083)}$ x (Precipitation)$^{(1.599)}$ | (38) | 0.334 | 0.290 |

In equation (28), when compared to ***Slope*** and ***Agriculture***, ***Precipitation*** has the highest values of beta (-1.13), partial and part correlations (-0.914 and -0.837) (Appendix 4). In equation (29), ***Precipitation*** has values of beta, partial, and part coefficients less than the corresponding values of ***Population*** and ***Slope*** (Appendix 4). In equation (30), the coefficients of ***Precipitation*** are also lower than those of ***Forest*** (Appendix 4).

The reason for the negative relation between pH and precipitation on all three bedrocks (Figure 10 and Figure 11) might be attributed to acid rain. Another reason for that is due to the increase of precipitation increasing the decay and decomposition of the organic materials in the soil and on the land surface. In turn, these processes release more carbondioxide into the flow network system to the stream. In addition, the effective carbonic acid weathering in humid regions may be more closely related to the increase of dissolved solutes with increasing mean annual precipitation than in dry regions (Peters, 1984).

The atmospheric composition might be one of the main sources explaining the contribution of ***Precipitation*** to total nitrogen yields, phosphorus, magnesium, sodium, sulfate, and bicarbonate (Table 29). That conclusion is comparable with what was stated earlier (Peters 1984). ***Precipitation*** is also directly related to dissolved oxygen (Table 29) because of rainfall increasing streamflow and resulting in more mixing in the streams.

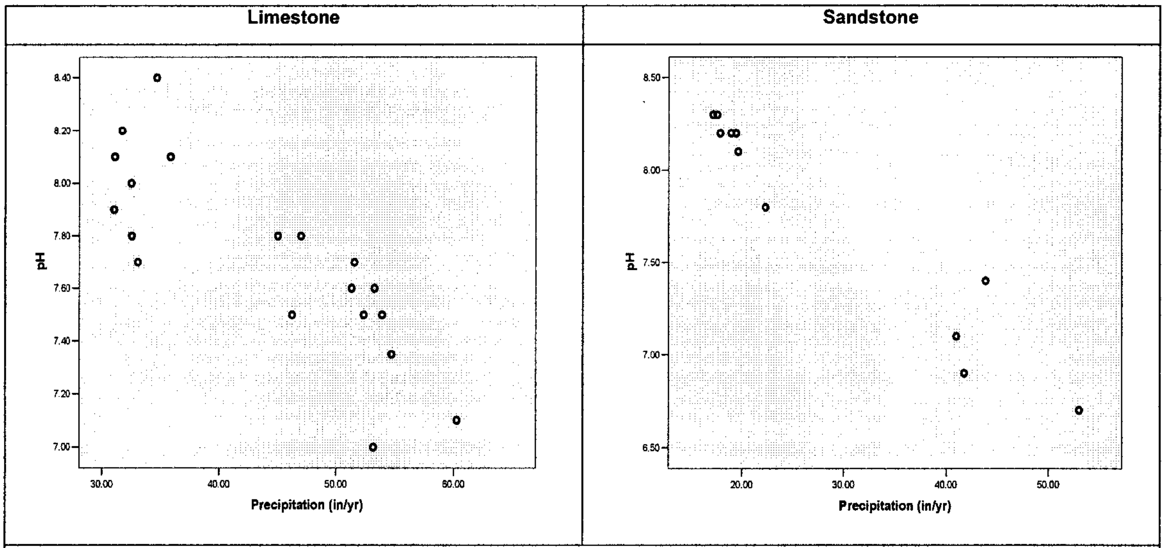


**Figure 10. Effects of average annual precipitation (in/yr) on annual median of pH in the drainage watersheds underlain by limestone and sandstone bedrock types in the United States.**

*In this Figure, pH seems to have a linear relation to Precipitation. However, as tested, an exponential function resulted in a better fit for this relationship.*

Precipitation is not significant in the regressions of total dissolved solids, calcium, potassium, and chloride like the results of Peters 1984. However, in this study, precipitation is statistically significant in many regressions for other ion yields, such as bicarbonate, pH, phosphorus, and total nitrogen (Table 29).

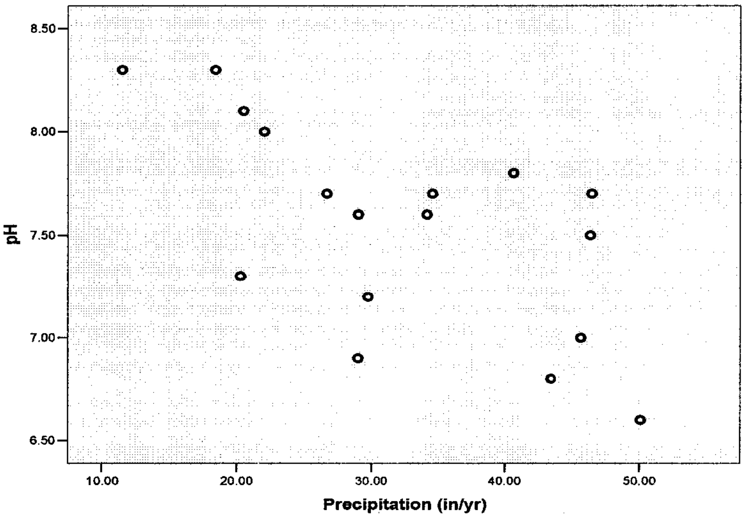


**Figure 11. Effects of average annual precipitation (in/yr) on annual median of pH in the drainage watersheds underlain by crystalline bedrock types in the United States.**

### *3.4. Effects of Slope:*

***Slope*** is found to be significant only in limestone watersheds. However, the effects of ***Slope*** are different in the regressions of different constituent yields. In the regressions with DO, ***Slope*** is the unique predictor (Table 30) while in the regressions of magnesium and sodium (equations 40 and 41, Table 30) ***Slope*** is significant in the company of other watershed characteristics (e.g. ***Precipitation***, ***Population***, and ***Agriculture***). In the regression of potassium (equation 42), Slope has higher values of beta, partial, and part correlations compared to ***Population*** in the same equation (Appendix 4).

**Table 30. The presence of significant slope coefficients of *Slope* in the regression equations in limestone and crystalline bedrocks ($p < 0.05$).**

| Limestone | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| $DO = 10^{(1.397)} \times (Slope)^{(0.296)}$ | (39) | 0.419 | 0.385 |
| $Mg = 10^{(6.403)} \times (Precipitation)^{(-3.437)} \times (Slope)^{(0.397)} \times (Agriculture)^{(-0.170)}$ | (40) | 0.862 | 0.834 |
| $Na = 10^{(2.645)} \times (Population)^{(0.619)} \times (Slope)^{(0.388)} \times (Precipitation)^{(-1.533)}$ | (41) | 0.783 | 0.740 |
| $K = 10^{(-0.021)} \times (Slope)^{(0.399)} \times (Population)^{(0.299)}$ | (42) | 0.582 | 0.530 |

In limestone bedrock, in equation 39 (Table 30), ***Slope*** is directly proportional to dissolved oxygen (Figure 12). This may be due to the fact that when slope increases, the velocity of the surface flow increases resulting in more mixing in the streams. ***Slope*** is also in direct proportion to magnesium (equation 40), sodium (equation 41), and potassium (equation 42) (Table 30). This might be attributed to the effects of slope in increasing erosion and sediment transport.

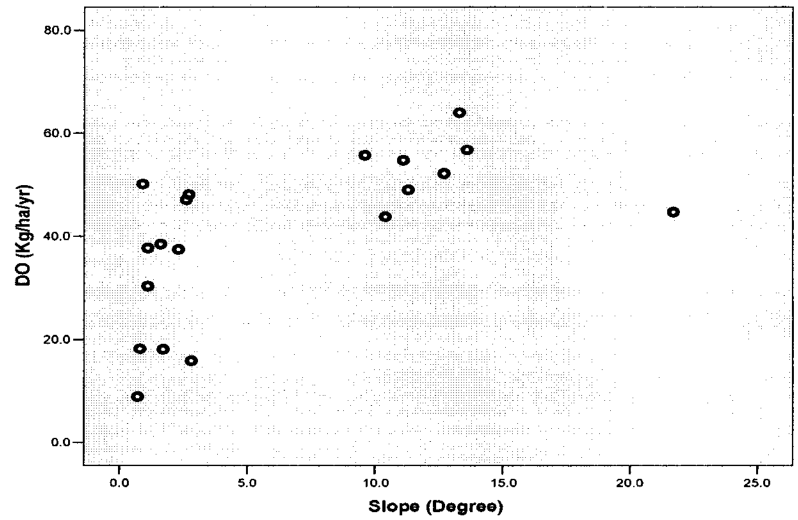


**Figure 12. Annual average yield of dissolved oxygen in relation to slope degree of drainage watersheds underlain by limestone bedrock in the United States.**

### *3.5. Effects of Area:*

The ***Area*** variable is not significant in any regression with the stream water ion yields when the significant level of alpha equals to 0.05. This indicates that ***Area*** is not an important variable in estimating stream water ion yields for this study. This can be explained with two reasons. The watershed sizes taken into account in this study might be too large (over the effective threshold) so that their area effects appear not discriminated from one to the others on a given water ion yield.

## 4. Effects of principal components of the environmental variables on stream water ion yields

As mentioned earlier in methods, principal component regression (PCR) is conducted based on principal component analysis using principal components extraction of the log transforms of the 8 environmental variables, which is rotated for ease of interpretation, including ***Water, Developed, Forest, Agriculture, Precipitation, Population, Slope,*** and ***Area*** of each watershed. Only components with eigenvalues greater than 1 are used for the principal component regression procedures.

### *4.1. Limestone watersheds:*

#### 4.1.1. Principal component analysis:

"The extraction communalities are estimates of the variance of each variable accounted for by the components in a principal component analysis" (SPSS, 2004). In this study, almost all of the extraction communities of the principal components of the log values of the 8 environmental variables in this bedrock are high (except ***Population*** and ***Precipitation***), ranging from 0.87 to 0.95 (Table 31).

**Table 31. Extraction communities of the principal components of the log transforms of the 8 environmental variables in limestone watersheds.**

| Communalities | | |
|---|---|---|
| | **Initial** | **Extraction** |
| **Water** | 1.000 | .867 |
| **Developed** | 1.000 | .945 |
| **Forest** | 1.000 | .904 |
| **Agriculture** | 1.000 | .949 |
| **Precipitation** | 1.000 | .614 |
| **Population** | 1.000 | .473 |
| **Slope** | 1.000 | .931 |
| **Area** | 1.000 | .942 |

Extraction Method: Principal Component Analysis.

The Rotation Sums of Squared Loadings maintains the cumulative percentage of variation explained by the extracted components (SPSS, 2004) (82.8 %, Table 32). However, that variation is a little bit spread more over the components (after SPSS, 2004) (from 65.3 and 17.48 % to 63.5 and 19.3 %) (Table 32).

**Table 32. Total variance explained for principal components in limestone watersheds.**

Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | **5.2** | **65.3** | **65.3** | **5.2** | **65.3** | **65.3** | **5.1** | **63.5** | **63.5** |
| 2 | **1.4** | **17.5** | **82.8** | **1.4** | **17.5** | **82.8** | **1.5** | **19.3** | **82.8** |
| 3 | .721 | 9.009 | 91.828 | | | | | | |
| 4 | .508 | 6.356 | 98.184 | | | | | | |
| 5 | .068 | .856 | 99.039 | | | | | | |
| 6 | .050 | .620 | 99.659 | | | | | | |
| 7 | .026 | .322 | 99.982 | | | | | | |
| 8 | .001 | .018 | 100.000 | | | | | | |

Extraction Method: Principal Component Analysis.

Together, the first two extracted components explain nearly 83% of the variability in the log transforms of the original 8 variables. Thus, we can reduce the complexity of the data set by using these two components (SPSS, 2004).

The rotated component matrix indicates what the components represent (SPSS, 2004) (Table 33). The first component is most highly correlated with ***Developed*** (component coefficient = 0.970), ***Agriculture*** (0.948), and ***Area*** (0.946). It is possible to estimate each variable based on the Rotated Component Matrix (Table 33). For example, ***Developed = 0.970 x Component 1 + 0.07 x Component 2.*** However, ***Developed*** is the best representative for component 1, because it is less correlated with component 2 (0.970 compared to 0.070). The second component is most highly correlated with ***Slope*** (component coefficient = 0.964) and ***Precipitation*** (0.505). Nevertheless, ***Slope*** is a better representative for the second component because it is also much less correlated with the first component (Table 33).

It should note that using the two saved components (as shown in Table 32 and Table 33) is more preferable than using each individual variable such as ***Developed***, ***Agriculture,*** and ***Slope*** because the two components are representative of all log transforms of the 8 original variables, and they are not linearly correlated with each other (after SPSS, 2004).

**Table 33. The Rotated Component Matrix of the log transforms of the 8 variables in limestone watersheds.**

**Rotated Component Matrix(a)**

| | Component | |
|---|---|---|
| | 1 | 2 |
| **Water** | .875 | -.320 |
| **Developed** | .970 | .070 |
| **Forest** | .868 | .387 |
| **Agriculture** | .948 | .221 |
| **Precipitation** | .599 | .505 |
| **Population** | .682 | -.089 |
| **Slope** | -.048 | .964 |
| **Area** | .946 | .217 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

The Regression Factor Scores (REGR factor score, Table 35) is computed by multiplying the case's standardized variable values by the component's score coefficients (SPSS, 2004) (Table 34).

**Table 34. Component Score Coefficient Matrix of the log transforms of the 8 variables in limestone watersheds.**

**Component Score Coefficient Matrix**

| | Component | |
|---|---|---|
| | 1 | 2 |
| **Water** | .217 | -.310 |
| **Developed** | .198 | -.049 |
| **Forest** | .145 | .182 |
| **Agriculture** | .178 | .058 |
| **Precipitation** | .076 | .291 |
| **Population** | .153 | -.130 |
| **Slope** | -.107 | .675 |
| **Area** | .178 | .055 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

### Table 35. Regression Factor Scores for the first 2 principal components of the log transforms of the 8 variables in limestone watersheds.

| Station No. | REGR factor score 1 | REGR factor score 2 |
|---|---|---|
| USGS 02313000 | 0.520 | -1.090 |
| USGS 02320500 | 0.905 | -0.805 |
| USGS 02329000 | 0.183 | -0.413 |
| USGS 02368000 | -0.201 | 0.163 |
| USGS 03085000 | 0.564 | 0.788 |
| USGS 03254000 | -0.016 | 0.857 |
| USGS 03290500 | 0.320 | 1.055 |
| USGS 03425000 | 0.523 | 1.086 |
| USGS 03543005 | 0.831 | 1.109 |
| USGS 03571850 | 0.952 | 1.106 |
| USGS 03593005 | 1.156 | 0.989 |
| USGS 04045500 | -0.657 | -1.110 |
| USGS 04057004 | -0.315 | -1.273 |
| USGS 04108690 | 0.296 | -0.917 |
| USGS 04126520 | -0.160 | -0.542 |
| USGS 04132052 | -0.191 | -0.719 |
| USGS 04165500 | 0.214 | -1.567 |
| USGS 08103900 | -2.026 | -0.103 |
| USGS 13018300 | -2.895 | 1.385 |

*(**REGR factor score 1** = 0.217 * **ZWater** + 0.198 * **ZDeveloped**+ 0.145 * **ZForest** + 0.178 * **ZAgriculture** + 0.076 * **ZPrecipitation** + 0.153 * **ZPopulation** + - 0.107 * **ZSlope** + 0.178 * **ZArea**;*

***REGR factor score 2** = -0.310 * **ZWater** + -0.049 * **ZDeveloped** + 0.182 * **ZForest** + 0.058 * **ZAgriculture** + 0.291 * **ZPrecipitation** + -0.130 * **ZPopulation** + 0.675 * **ZSlope** + 0.055 * **ZArea** )*

*(**ZWater** = (*log transformed *value of **Water** (i) – average* log transformed *value of all* log transformed ***Water** values)/std. deviation of all* log transformed ***Water** values). This formula is applied similarly for Zscores of the other variables).*

### 4.1.2. Principal component regression:

There are no significant regressions between the first two selected principal components with total dissolved solids, total nitrogen, orthophosphate, calcium, magnesium, potassium, chloride, sulfate, and bicarbonate (Appendix 4). The PCRs of dissolved oxygen, pH, and phosphorus are shown in Table 36. However, the squared correlations are not very high, ranging from 0.217 to 0.284. It is noted that component 1 is significant in the regressions of pH because as mentioned earlier in equation (27) (Table 29), pH is negatively related to ***Precipitation*** and the component coefficient of ***Precipitation*** in component 1 is higher than that in component 2 (0.60 compared to 0.51, Table 33). Thus, the reason that pH is also negatively related to component 2 because of the effect of ***Precipitation*** on component 2 and ***Precipitation***'s relation to pH. Similarly, the reason that phosphorus is positively related to component 1 because it is positively related to ***Developed*** as shown in equation (17) (Table 25) while ***Developed*** also has the highest correlation with component 1 as shown in Table 33 (the component correlation coefficient equals to 0.970). Likewise, dissolved oxygen is positively related to component 2 because it is positively related to S***lope*** as shown in equation (39, Table 30) while ***Slope*** also has the highest correlation with component 2 (the component correlation coefficient equals to 0.964) (Table 33).

**Table 36. Significant principal component regressions in limestone watersheds ($p < 0.05$).**

| Limestone | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| $DO = 10^{[1.562 + 0.123 \times (\text{REGR factor score 2})]}$ | (43) | .284 | .242 |
| $pH = 10^{[0.887 - 0.01 \times (\text{REGR factor score 1})]}$ | (44) | .256 | .212 |
| $\text{Phosphorus} = 10^{[-0.882 + 0.17 \times (\text{REGR factor score 1})]}$ | (45) | .217 | .171 |

It is necessary to understand that although component 1 has correlations with pH and phosphorus lower than that of component 2 (Table 36), as shown in Table 32, component 1 is much more important because it is accounted for 63.5 % of the total variation of the 8 variables while component 2 is accounted for 19.3 % of the total variation only. In addition, the squared correlations of the principal component regressions in Table 36 are lower than those of the corresponding multiple linear regressions in equation (39) (Table 30), equation (27) (Table 29), and in equation (17) (Table 25).

## *4.2. Sandstone watersheds:*

### 4.2.1. Principal component analysis:

The extractions of the principal components of the environmental variables for sandstone watersheds are higher than those of limestone watersheds, ranging from 0.71 to 0.99 (Table 37). Therefore, the extracted components represent the variables well (SPSS, 2004).

**Table 37. Extraction communities of the principal components of the log transforms of the 8 environmental variables in sandstone watersheds.**

| Communalities | | |
|---|---|---|
| | **Initial** | **Extraction** |
| **Water** | 1.000 | .944 |
| **Developed** | 1.000 | .974 |
| **Forest** | 1.000 | .906 |
| **Agriculture** | 1.000 | .982 |
| **Precipitation** | 1.000 | .841 |
| **Population** | 1.000 | .936 |
| **Slope** | 1.000 | .707 |
| **Area** | 1.000 | .989 |
| Extraction Method: Principal Component Analysis. | | |

The cumulative percentage of variation explained by the extracted components is maintained of the rotation loadings method (approximately 91%) (Table 38). However, that variation is spread more over the components (53.9 and 37.1 instead of 71.6 and 19.4) (after SPSS, 2004) (Table 38).

**Table 38. Total variance explained for principal components in sandstone watersheds.**

| Total Variance Explained | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | **5.7** | **71.6** | **71.6** | **5.7** | **71.6** | **71.6** | **4.3** | **53.9** | **53.9** |
| 2 | **1.6** | **19.4** | **91.0** | **1.6** | **19.4** | **91.0** | **3.0** | **37.1** | **91.0** |
| 3 | .494 | 6.174 | 97.157 | | | | | | |
| 4 | .140 | 1.755 | 98.912 | | | | | | |
| 5 | .054 | .669 | 99.581 | | | | | | |
| 6 | .020 | .246 | 99.828 | | | | | | |
| 7 | .013 | .158 | 99.986 | | | | | | |
| 8 | .001 | .014 | 100.000 | | | | | | |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |

We can also use the scree plot, which presents eigenvalue of each component in the initial solution, to determine the optimal number of components (Figure 13). Generally, the components on the steep slope are extracted, because the components on the shallow slope contribute little to the solution (SPSS, 2004). As shown in Figure 13, the last big drop occurs between the second and the third components. Thus, using the first two components is a better option (SPSS, 2004).

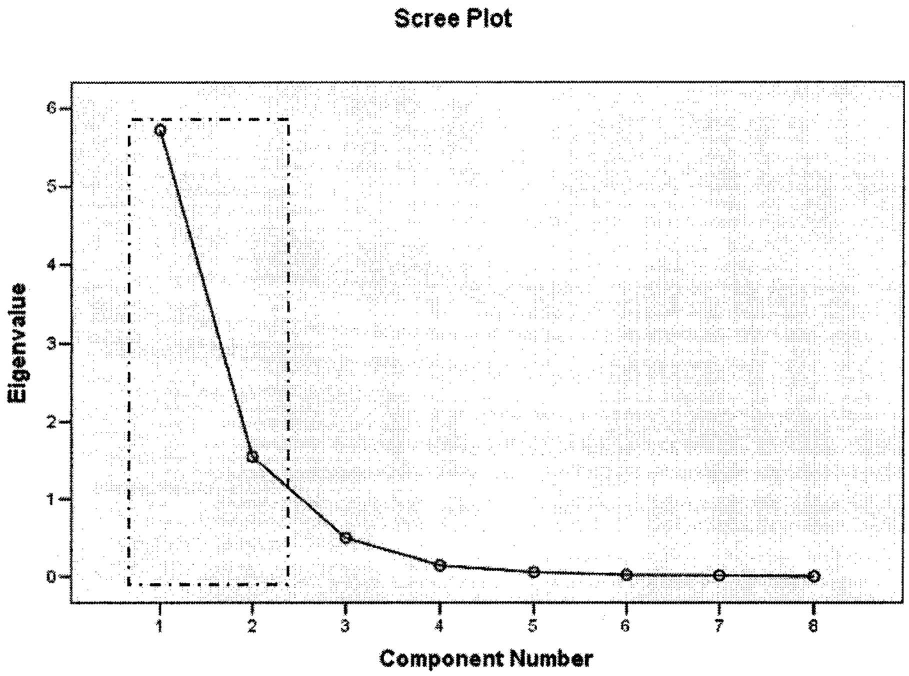


**Figure 13. Scree plot of the initial eigenvalue of each component in sandstone watersheds.**

In the rotated component matrix (Table 39), the first component is most highly correlated with ***Forest***, ***Developed,*** and ***Water***. However, ***Forest*** is the best representative, because it is less correlated with the second component. The second component is most highly correlated with ***Population*** and ***Precipitation*** followed by ***Area and Agriculture.*** Nevertheless, ***Population*** is a better representative for the second component because it is also much less correlated with the first component (Table 39) (component correlation coefficient of ***Population*** with component 2 is 0.964 while the coefficient with the first component is only -0.086).

**Table 39. The Rotated Component Matrix of the log transforms of the 8 variables in sandstone watersheds.**

Rotated Component Matrix(a)

| | Component | |
|---|---|---|
| | 1 | 2 |
| **Water** | .910 | -.341 |
| **Developed** | .921 | -.354 |
| **Forest** | .947 | .098 |
| **Agriculture** | .847 | -.515 |
| **Precipitation** | -.235 | .887 |
| **Population** | -.086 | .964 |
| **Slope** | -.417 | .730 |
| **Area** | .888 | -.449 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

The two regression factor scores (Table 41) used for the PCR procedures were calculated by multiplying component score coefficients (Table 40) by the corresponding standardized values (Zscores).

**Table 40. Component Score Coefficient Matrix of the log transforms of the 8 variables in sandstone watersheds.**

Component Score Coefficient Matrix

| | Component | |
|---|---|---|
| | 1 | 2 |
| **Water** | .228 | .037 |
| **Developed** | .229 | .033 |
| **Forest** | .337 | .258 |
| **Agriculture** | .168 | -.062 |
| **Precipitation** | .118 | .378 |
| **Population** | .185 | .448 |
| **Slope** | .023 | .261 |
| **Area** | .196 | -.021 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

**Table 41. Regression Factor Scores for the first 2 principal components of the log transforms of the 8 variables in sandstone watersheds.**

| Station No. | REGR factor score 1 | REGR factor score 2 |
|---|---|---|
| USGS 01545600 | -1.497 | 0.737 |
| USGS 03049625 | 1.260 | 1.724 |
| USGS 03237280 | -1.994 | 1.001 |
| USGS 06452000 | 0.004 | -0.861 |
| USGS 06478500 | 0.103 | -1.031 |
| USGS 07026000 | 0.394 | 0.886 |
| USGS 07231500 | 0.806 | 0.052 |
| USGS 07234000 | -0.326 | -1.229 |
| USGS 07237500 | -0.057 | -1.009 |
| USGS 07245000 | 1.109 | 0.267 |
| USGS 06438000 | 0.198 | -0.538 |

### 4.2.2. Principal component regression:

There is no regression between the first two selected principal components with orthophosphate, phosphorus, Mg, chloride, and sulfate. The PCRs of total dissolved solids, dissolved oxygen, pH, total nitrogen, calcium, potassium, sodium, and bicarbonate are shown in Table 42. As shown in Table 42, the *REGR factor score 2* is significant positively in most of those regressions (except for the PCR of pH) while *REGR factor score 1* is significant positively in the regression of sodium only.

In general, the squared correlations of the principal component regressions expressed in Table 42 are higher than those in limestone watersheds (Table 36), ranging from 0.39 to 0.83.

**Table 42. Principal component regressions in sandstone watersheds ($p < 0.05$)**

| Sandstone | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| $TDS = 10^{[2.334 + 0.213 \times (\text{REGR factor score 2})]}$ | (46) | .408 | .343 |
| $DO = 10^{[0.873 + 0.735 \times (\text{REGR factor score 2})]}$ | (47) | .861 | .841 |
| $pH = 10^{[0.888 - 0.029 \times (\text{REGR factor score 2})]}$ | (48) | .673 | .637 |
| $\text{Total nitrogen} = 10^{[0.002 + 0.551 \times (\text{REGR factor score 2})]}$ | (49) | .834 | .810 |
| $Ca = 10^{[1.426 + 0.242 \times (\text{REGR factor score 2})]}$ | (50) | .406 | .340 |
| $Na = 10^{[1.422 + 0.262 \times (\text{REGR factor score 1})]}$ | (51) | .539 | .488 |
| $K = 10^{[0.448 + 0.402 \times (\text{REGR factor score 2})]}$ | (52) | .640 | .600 |
| $\text{Bicarbonate} = 10^{[1.776 + 0.291 \times (\text{REGR factor score 2})]}$ | (53) | .495 | .432 |

*(**REGR factor score 1** = 0.228 * **ZWater** + 0.229 * **ZDeveloped**+ 0.337 * **ZForest** + 0.168 * **ZAgriculture** + 0.118 * **ZPrecipitation** + 0.185 * **ZPopulation** + - 0.023 * **ZSlope** + 0.196 * **ZArea**;*

***REGR factor score 2** = 0.037 * **ZWater** + 0.033 * **ZDeveloped** + 0.258 * **ZForest** + -0.062 * **ZAgriculture** + 0.378 * **ZPrecipitation** + 0.448 * **ZPopulation** + 0.261 * **ZSlope** + -0.021 * **ZArea** )*

*(**ZWater** = (*log transformed *value of **Water** (i) – average* log transformed *value of all* log transformed ***Water** values)/std. deviation of all* log transformed ***Water** values).*

The reason that total dissolved solids, potassium, and bicarbonate are positively related to component 2 because they are proportional related to ***Population*** (equations 23 through 26, Table 28), which has the highest component correlation with component 2 (0.964) while ***Population***'s correlation with component 1 is only -0.086 (Table 39). Similarly, the reason that dissolved solids, pH, total nitrogen, and phosphorus are significant related to component 2 because they are related to ***Precipitation*** (equations 32 - 35, Table 29), which has the second highest component correlation with component 2 (0.887) while ***Precipitation***'s correlation with component 1 is only -0.235 (Table 39). Likewise, the reason that sodium is positively related to component 1 because it is positively related to ***Water*** (equation 14, Table 24), which has much higher component correlation (0.910) with component 1 compared to (-0.341) of component 2 (Table 39).

Note that while almost all principal component regressions in Table 42 have higher squared correlations than those of the corresponding multiple linear regressions, some of them (e.g. regressions of dissolved oxygen, pH, and bicarbonate) have lower squared correlations when compared to those of equations (23) and (26) (Table 28), and equation (33) (Table 29) . In addition, although principal component regression gives higher correlations, it is not necessary that it give better predictions because most of the regressions in Table 42 are significant with component 2, which is accounted for only 37% of the total variation of the 8 independent variables.

## *4.3. Crystalline watersheds:*

### 4.3.1. Principal component analysis:

The extractions of the principal components of the log values of the 8 environmental variables in crystalline bedrock are also higher than those of limestone watersheds, ranging from 0.71 to 0.98 (Table 43).

**Table 43. Extraction communities of the principal components of the log transforms of the 8 environmental variables in crystalline watersheds.**

| Communalities | | |
|---|---|---|
| | Initial | Extraction |
| Water | 1.000 | .921 |
| Developed | 1.000 | .873 |
| Forest | 1.000 | .906 |
| Agriculture | 1.000 | .848 |
| Precipitation | 1.000 | .864 |
| Population | 1.000 | .705 |
| Slope | 1.000 | .977 |
| Area | 1.000 | .952 |
| Extraction Method: Principal Component Analysis. | | |

Unlike the results for limestone and sandstone watersheds, crystalline watersheds have three components containing eigenvalues greater than 1. Also, the variation explained by the extracted components is more evenly spread over the rotated components (Table 44). Together, the first three extracted components explain nearly 88% of the variability in the log transforms of the original 8 variables.

**Table 44. Total variance explained for principal components in crystalline watersheds.**

| Total Variance Explained | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4.2 | 53.0 | 53.0 | 4.2 | 53.0 | 53.0 | 3.8 | 47.6 | 47.6 |
| 2 | 1.6 | 20.4 | 73.5 | 1.6 | 20.4 | 73.5 | 1.7 | 21.6 | 69.3 |
| 3 | 1.2 | 14.6 | 88.1 | 1.2 | 14.6 | 88.1 | 1.5 | 18.8 | 88.1 |
| 4 | .554 | 6.924 | 95.004 | | | | | | |
| 5 | .284 | 3.556 | 98.560 | | | | | | |
| 6 | .072 | .904 | 99.463 | | | | | | |
| 7 | .042 | .522 | 99.985 | | | | | | |
| 8 | .001 | .015 | 100.000 | | | | | | |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |

As shown in the rotated component matrix (Table 45), the first component is most highly correlated with ***Forest, Agriculture***, and ***Developed*** land covers (0.933, 0.920, and 0.854, respectively, Table 45). However, ***Agriculture*** and **Forest** might be the best representatives for component 1 because they are less correlated with the other components (Table 45). The second component is most highly correlated with ***Precipitation*** and ***Population*** (0.913 and 0.810, Table 45). ***Precipitation*** is a very good representative for the second component because it is also much less correlated with the first component and the third one. The third component is highly correlated with ***Slope*** (-0.983, Table 45), and ***Slope*** is also much less correlated with the other two components (Table 45). Therefore, ***Slope*** is the best representative for component 3.

**Table 45. The Rotated Component Matrix of the log transforms of the 8 variables in crystalline watersheds.**

**Rotated Component Matrix(a)**

| | **Component** | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| **Water** | .663 | .294 | .629 |
| **Developed** | .854 | .378 | -.033 |
| **Forest** | .933 | -.071 | .173 |
| **Agriculture** | .920 | .036 | .032 |
| **Precipitation** | -.026 | .913 | -.173 |
| **Population** | .138 | .810 | .174 |
| **Slope** | -.074 | .071 | -.983 |
| **Area** | .949 | -.010 | .229 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 4 iterations.

The highest component score coefficient in component 1 is ***Agriculture*** (0.285) while in component 2 it is ***Precipitation*** (0.554), and in component 3, it is ***Slope*** (-0.745) (Table 46).

**Table 46. Component Score Coefficient Matrix of the log transforms of the 8 variables in crystalline watersheds.**

| Component Score Coefficient Matrix | | | |
|---|---|---|---|
| | Component | | |
| | 1 | 2 | 3 |
| **Water** | .067 | .133 | .371 |
| **Developed** | .244 | .147 | -.174 |
| **Forest** | .273 | -.127 | -.042 |
| **Agriculture** | .285 | -.065 | -.146 |
| **Precipitation** | -.063 | .553 | -.104 |
| **Population** | -.066 | .484 | .131 |
| **Slope** | .153 | .023 | -.745 |
| **Area** | .262 | -.090 | .000 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

The regression factor scores (Table 47) for the PCR analysis were calculated by multiplying the component score coefficients (Table 46) by the corresponding standardized values (Zscores).

**Table 47. Regression Factor Scores for the first 3 principal components of the log transforms of the 8 variables in crystalline watersheds.**

| Station No. | REGR factor score 1 | REGR factor score 2 | REGR factor score 3 |
|---|---|---|---|
| USGS 01059400 | 1.152 | 1.184 | 0.493 |
| USGS 01066000 | 0.621 | 1.299 | 0.421 |
| USGS 04014500 | -0.713 | -0.027 | 1.864 |
| USGS 05124480 | -0.766 | -0.205 | 1.952 |
| USGS 06623800 | -0.374 | 0.109 | -0.361 |
| USGS 07311200 | -1.279 | 1.027 | 0.985 |
| USGS 08431700 | -0.542 | -1.344 | -0.614 |
| USGS 09352900 | -0.049 | 0.870 | -1.075 |
| USGS 09430600 | -0.587 | -0.655 | -1.322 |
| USGS 09508300 | -1.472 | -0.651 | -0.502 |
| USGS 11264500 | 0.048 | 1.017 | -0.589 |
| USGS 12447390 | -0.931 | 0.115 | -1.219 |
| USGS 13169500 | 0.090 | -2.156 | 0.743 |
| USGS 13331500 | 0.196 | 0.688 | -0.967 |
| USGS 14048000 | 1.859 | -1.249 | -0.368 |
| USGS 14103000 | 1.855 | -0.585 | 0.440 |
| USGS 14113000 | 0.892 | 0.563 | 0.118 |

### 4.3.2. Principal component regression:

In crystalline bedrocks, there are only 2 principal component regressions are significant, that is, the PCR of dissolved oxygen and PCR of pH (Table 48). *REGR factor score 1* and *REGR factor score 3* are not significant in any of these 2 regressions while *REGR factor score 2* is positively significant in the relation with dissolved oxygen and negatively significant in the relation with pH. The reason for that is because of the significant relations between ***Precipitation*** and dissolved oxygen, and between ***Precipitation*** and pH as shown in equations (36) and (37), Table 29 while ***Precipitation*** has the highest component correlation with component 2 but is much less correlated with component 1 and component 3 (Table 45).

**Table 48. Significant principal component regressions in crystalline watersheds ($p < 0.05$).**

| Crystalline | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| $DO = 10^{[1.525 + 0.246 \times (\text{REGR factor score 2})]}$ | (54) | .296 | .249 |
| $pH = 10^{[0.876 - 0.02 \times (\text{REGR factor score 2})]}$ | (55) | .450 | .413 |

In general, the squared correlations of the principal component regressions expressed in Table 48 are higher than those in limestone watersheds (Table 36) but lower than those in sandstone watersheds (Table 42). In addition, the squared correlations in Table 48 are lower than those of the corresponding multiple linear regressions in equations (36) and (37) (Table 29).

## 5. Summary of the principal component regression analysis

Principal component regression (PCR) is based on principal component analysis and the multiple regression technique. In a principal component analysis, all variables are used to compute components. This technique reduces the number of variables and therefore, gives a simple descriptive if the principal components are easily interpreted. In addition, principal components are nearly not collinearity. Unlike the multiple regression technique, PCR composites variables rather than eliminates variables in the regression processes.

In this study, generally speaking, PCR equations have lower correlation (except for the regressions of total dissolved solids, pH, and bicarbonate in sandstone watershed, Table 49) and approximately equal Standard Error of the Estimate when compared to those of multiple regression technique using stepwise method of model selection (Table 49). Besides that, using multiple regressions can detect more relationship than PCR (Table 49). In addition, most principal component regressions in Tables 37, 43, and 49 are significant with component 2 which is accounted for only 20 to 37% of the total variation of the 8 independent variables. Those clues indicate that using PCR equations for ion yield estimations might not be as good as using the corresponding multiple regression equations. This inference will be tested again in the next step of **Stream water ion yield prediction, model validation, and spatial distribution of the predicted ion yields.**

So why do most PCR equations have lower correlations than those of multiple regression technique using stepwise method of model selection? This can be explained that PCR takes into account all independent variables although some of them are not actually effective predictors (e.g. ***Slope*** and ***Area***) while multiple regression selects only those predictors most affecting the ion yield responses into an equation (e.g. ***Precipitation*** and ***Population***).

**Table 49. Comparison of Multiple Regression and Principal Component Regression using Correlation and Standard Error of the Estimate for each bedrock type (more details are presented in Appendix 4, 5, 6).**

| | | | | |
|---|---|---|---|---|
| **Limestone** | | | | |
| | **Multiple Regression Using Log Scale** | | **Principal Component Regression Using Log Scale** | |
| | **$R^2$** | **Std. Error of the Estimate** | **$R^2$** | **Std. Error of the Estimate** |
| TDS | No sig. reg. | | No sig. reg. | |
| DO | .419 | 0.181 | .284 | 0.201 |
| pH | .633 | 0.013 | .256 | 0.018 |
| Total Nitrogen | No sig. reg. | | No sig. reg. | |
| Orthophosphate | .456 | 0.329 | No sig. reg. | |
| Phosphorus | .302 | 0.313 | .217 | 0.332 |
| Ca | No sig. reg. | | No sig. reg. | |
| Mg | .862 | 0.130 | No sig. reg. | |
| Na | .783 | 0.210 | No sig. reg. | |
| K | .582 | 0.234 | No sig. reg. | |
| Chloride | .608 | 0.305 | No sig. reg. | |
| Sulfate | .426 | 0.332 | No sig. reg. | |
| Bicarbonate | .364 | 0.249 | No sig. reg. | |
| **Sandstone** | | | | |
| | **Multiple Regression using log scale** | | **Principal Component Regression using log scale** | |
| | $R^2$ | **Std. Error of the Estimate** | $R^2$ | **Std. Error of the Estimate** |
| TDS | .624 | 0.228 | .408 | 0.270 |
| DO | .843 | 0.342 | .861 | 0.322 |
| pH | .939 | 0.009 | .673 | 0.021 |
| Total Nitrogen | .791 | 0.300 | .834 | 0.268 |
| Orthophosphate | No sig. reg. | | No sig. reg. | |
| Phosphorus | .516 | 0.430 | No sig. reg. | |
| Ca | No sig. reg. | | .406 | .309 |
| Mg | .376 | 0.363 | No sig. reg. | |
| Na | .453 | 0.278 | .539 | 0.255 |
| K | .586 | 0.341 | .640 | 0.317 |
| Chloride | No sig. reg. | | No sig. reg. | |
| Sulfate | No sig. reg. | | No sig. reg. | |
| Bicarbonate | .717 | 0.262 | .495 | 0.327 |
| **Crystalline** | | | | |
| | **Multiple Regression using log scale** | | **Principal Component Regression using log scale** | |
| | **$R^2$** | **Std. Error of the Estimate** | **$R^2$** | **Std. Error of the Estimate** |
| TDS | No sig. reg. | | | |
| DO | .482 | 0.337 | .31 | 0.39 |
| pH | .418 | 0.023 | .63 | 0.020 |
| Total Nitrogen | .334 | 0.412 | No sig. reg. | |
| Orthophosphate | No sig. reg. | | No sig. reg. | |
| Phosphorus | No sig. reg. | | No sig. reg. | |
| Ca | No sig. reg. | | No sig. reg. | |
| Mg | No sig. reg. | | No sig. reg. | |
| Na | No sig. reg. | | No sig. reg. | |
| K | No sig. reg. | | No sig. reg. | |
| Chloride | No sig. reg. | | No sig. reg. | |
| Sulfate | No sig. reg. | | No sig. reg. | |
| Bicarbonate | No sig. reg. | | No sig. reg. | |

## 6. Stream water ion yield prediction, model validation, and spatial distribution of the predicted ion yields.

There are only 4 stations available for the testing data in sandstone bedrock (see Appendices 1, 2, 3 and the websites listed in "References for additional (testing) sandstone stations" at the end of Literature Cited section). This data set is too small for an actual validation. However, this step is necessary for a completed conceptual methodology. Based on the data available for the 4 stations, and based on high correlations and a compatible comparison between multiple regression and PCR approaches, the following equations in sandstone watersheds were selected for predictions (Table 50).

**Table 50. Selected equations for predictions and validations of stream water ion yields in sandstone watersheds.**

| **Multiple Regression** | | $R^2$ | Adj. $R^2$ |
|---|---|---|---|
| $DO = 10^{(-4.271)} \text{ x } (\text{Precipitation})^{(3.631)}$ | (32) | 0.843 | 0.821 |
| $pH = 10^{(1.137)} \text{ x } (\text{Precipitation})^{(-0.176)}$ | (33) | 0.939 | 0.932 |
| $\text{Total nitrogen} = 10^{(-3.790)} \text{ x } (\text{Precipitation})^{(2.678)}$ | (34) | 0.791 | 0.761 |
| $K = 10^{(-0.416)} \text{ x } (\text{Population})^{(0.672)}$ | (25) * | 0.586 | 0.540 |
| **Principal Component Regression** | | $R^2$ | Adj. $R^2$ |
| $DO = 10^{[0.873 + 0.735 \text{ x } (\text{REGR factor score 2})]}$ | (47) | 0.861 | 0.841 |
| $pH = 10^{[0.888 - 0.029 \text{ x } (\text{REGR factor score 2})]}$ | (48) | 0.673 | 0.637 |
| $\text{Total nitrogen} = 10^{[0.002 + 0.551 \text{ x } (\text{REGR factor score 2})]}$ | (49) | 0.834 | 0.810 |
| $K = 10^{[0.448 + 0.402 \text{ x } (\text{REGR factor score 2})]}$ | (52) * | 0.640 | 0.600 |

*(*: Although the regressions of potassium yield does not have both multiple regression and PCR square coefficients >= 0.640, these equations are taken into account for prediction just for a comparison between the 2 regression techniques).*

The predicted values of ion yields are shown in Table 51 and the corresponding Average Percentage Errors are shown in Table 52.

**Table 51. Predicted Ion Yields Using Multiple Regression, Principal Component Regression, and GIS techniques in sandstone watersheds.**

| | DO | pH | Total Nitrogen | K |
|---|---|---|---|---|
| **Multiple Regression** | ***kg/ha/yr*** | ***units*** | ***kg/ha/yr*** | ***kg/ha/yr*** |
| USGS 08053500 | 21.8 | 7.3 | 2.2 | 6.2 |
| USGS 06810000 | 17.2 | 7.4 | 1.9 | 4.2 |
| USGS 06485500 | 5.7 | 7.8 | 0.8 | 4.0 |
| USGS 05413500 | 17.5 | 7.4 | 1.9 | 4.7 |
| | | | | |
| **PCR** | | | | |
| USGS 08053500 | 13.9 | 7.5 | 1.6 | 3.9 |
| USGS 06810000 | 9.1 | 7.7 | 1.2 | 3.1 |
| USGS 06485500 | 4.6 | 7.9 | 0.7 | 2.1 |
| USGS 05413500 | 12.6 | 7.6 | 1.5 | 3.7 |
| | | | | |
| **Actual Values** | | | | |
| USGS 08053500 | 6.2 | 8.1 | NoData | 2.7 |
| USGS 06810000 | 20.9 | 7.9 | NoData | 7.2 |
| USGS 06485500 | 19.6 | 8.3 | 5.7 | 12.0 |
| USGS 05413500 | NoData | 8.3 | NoData | 8.1 |
| | | | | |
| **PCR Using GIS** | | | | |
| USGS 08053500 | 15.1 | 7.6 | 1.6 | 3.9 |
| USGS 06810000 | 13.2 | 7.6 | 1.5 | 3.7 |
| USGS 06485500 | 4.8 | 8.0 | 0.7 | 2.0 |
| USGS 05413500 | 19.2 | 7.5 | 2.0 | 4.6 |
| | | | | |
| **Multiple Regession Using GIS** | | | | |
| USGS 08053500 | 21.9 | 7.3 | 2.2 | 5.3 |
| USGS 06810000 | 17.3 | 7.4 | 1.9 | 4.1 |
| USGS 06485500 | 5.9 | 7.8 | 0.8 | 3.5 |
| USGS 05413500 | 17.5 | 7.4 | 1.9 | 4.7 |

**Table 52. Average Percent Error (APE, in fraction) for Predicted Ion Yields Using Multiple Regression, Principal Component Regression, and GIS techniques in sandstone watersheds.**

| | DO | pH | Total Nitrogen | K |
|---|---|---|---|---|
| **Multiple Regression** | *(APE, in faction)* | *(APE, in faction)* | *(APE, in faction)* | *(APE, in faction)* |
| USGS 08053500 | -2.52 | 0.10 | NoData | -1.30 |
| USGS 06810000 | 0.18 | 0.06 | NoData | 0.42 |
| USGS 06485500 | 0.71 | 0.06 | 0.86 | 0.67 |
| USGS 05413500 | NoData | 0.11 | NoData | 0.42 |
| *General Average Percent Error (GAPE)** | ***1.13*** | ***0.08*** | ***0.86*** | ***0.70*** |
| **PCR** | | | | |
| USGS 08053500 | -1.24 | 0.07 | NoData | -0.44 |
| USGS 06810000 | 0.56 | 0.03 | NoData | 0.57 |
| USGS 06485500 | 0.77 | 0.05 | 0.88 | 0.83 |
| USGS 05413500 | NoData | 0.08 | NoData | 0.54 |
| *General Average Percent Error (GAPE)** | ***0.86*** | ***0.06*** | ***0.88*** | ***0.60*** |
| **PCR Using GIS** | | | | |
| USGS 08053500 | -1.4 | 0.1 | NoData | -0.4 |
| USGS 06810000 | 0.4 | 0.0 | NoData | 0.5 |
| USGS 06485500 | 0.8 | 0.0 | 0.9 | 0.8 |
| USGS 05413500 | NoData | 0.1 | NoData | 0.4 |
| *General Average Percent Error (GAPE)** | ***0.85*** | ***0.06*** | ***0.88*** | ***0.55*** |
| **Multiple Regession Using GIS** | | | | |
| USGS 08053500 | -2.53 | 0.10 | NoData | -0.96 |
| USGS 06810000 | 0.17 | 0.06 | NoData | 0.43 |
| USGS 06485500 | 0.70 | 0.06 | 0.86 | 0.71 |
| USGS 05413500 | NoData | 0.11 | NoData | 0.42 |
| *General Average Percent Error (GAPE)** | ***1.13*** | ***0.08*** | ***0.86*** | ***0.63*** |

*General Average Percent Error (GAPE)* is calculated by absolute values of APE from each watershed because we are interested in understanding how far the predicted results are from the actual values using different methods.*

**1- For multiple regression and PCR not using GIS layers: dissolved oxygen, pH, total nitrogen, potassium, and bicarbonate are predicted.**

**a.** For multiple regression, pH has the smallest Average Percent Error (APE, 0.08) followed by potassium (0.70), total nitrogen (0.86), and dissolved oxygen (1.13) (Table 52)

**b.** For PCR, pH has the smallest APE (0.06) followed by potassium (0.60), dissolved oxygen (0.86), and total nitrogen (0.88) (Table 52).

**2- For multiple regression and PCR using GIS layers: dissolved oxygen, pH, total nitrogen, and potassium are predicted.**

**a.** For multiple regression, pH has the smallest APE (0.08) followed by potassium (0.63), total nitrogen (0.86), and dissolved oxygen (1.13) (Table 52).

**b.** For PCR, pH has the smallest APE (0.06), followed by potassium (0.55), dissolved oxygen (0.85), and total nitrogen (0.88) (Table 52).

From (1) and (2), it is possible to conclude that the ion yield predictions made by multiple regression and PCR are not very much different although using PCR has lower APE for pH and potassium. Practically speaking, multiple regression is much preferred because it requires a smaller number of independent variable data. Besides that, the use of GIS models gives almost the same average predicted results as calculated by the corresponding regressions without using GIS. While the non-GIS models just use average values for each ion yield of each watershed (lumped/aggregated), GIS models take into account spatial distribution of the input data (disaggregated) which may be far from those average values taken to build the regression models. Therefore, using the GIS approach may introduce more standard deviation (Table 53) and larger range of the results than non-GIS method.

In addition, the GIS models of PCR give larger standard deviation and larger range of the predictions when compared to GIS models of multiple regressions, except for potassium (Table 53, and Figure 14 – Figure 16). For example, as shown in Figure 14, the range of dissolved oxygen predicted by PCR using GIS is from 0 – 61.3 (kg/ha/yr) while the range of dissolved oxygen predicted by multiple regression using GIS is only from 3.3 – 11.0 (kg/ha/yr). The reason for this might be due to the fact that PCR takes into account all variables for prediction so that the range of the input data is also larger than some selected variables used in the multiple regressions, and consequently, PCR using GIS introduces more uncertainty.

Nevertheless, using multiple regressions with GIS-based models is a better choice for ion yield prediction is this study because this approach not only gives the same average predicted results as using the regression models without GIS, but also gives a visualization depicting the spatial effects. Therefore, we can see and delineate different areas of different levels of ion yields (Figure 14 – Figure 16). As shown, the shape and pattern of the map of predicted ion yield reflects the shape and the pattern of the corresponding predictor. For example, the map of potassium predicted (Table 14) using multiple linear regression and GIS is similar to the map of population density because potassium depends on population density, which was mapped by counties. However, the map of potassium predicted using PCR and GIS (Table 14) is much smoother compared to that of using multiple linear regression due to effects of other variables with higher resolution that were significant in the model.

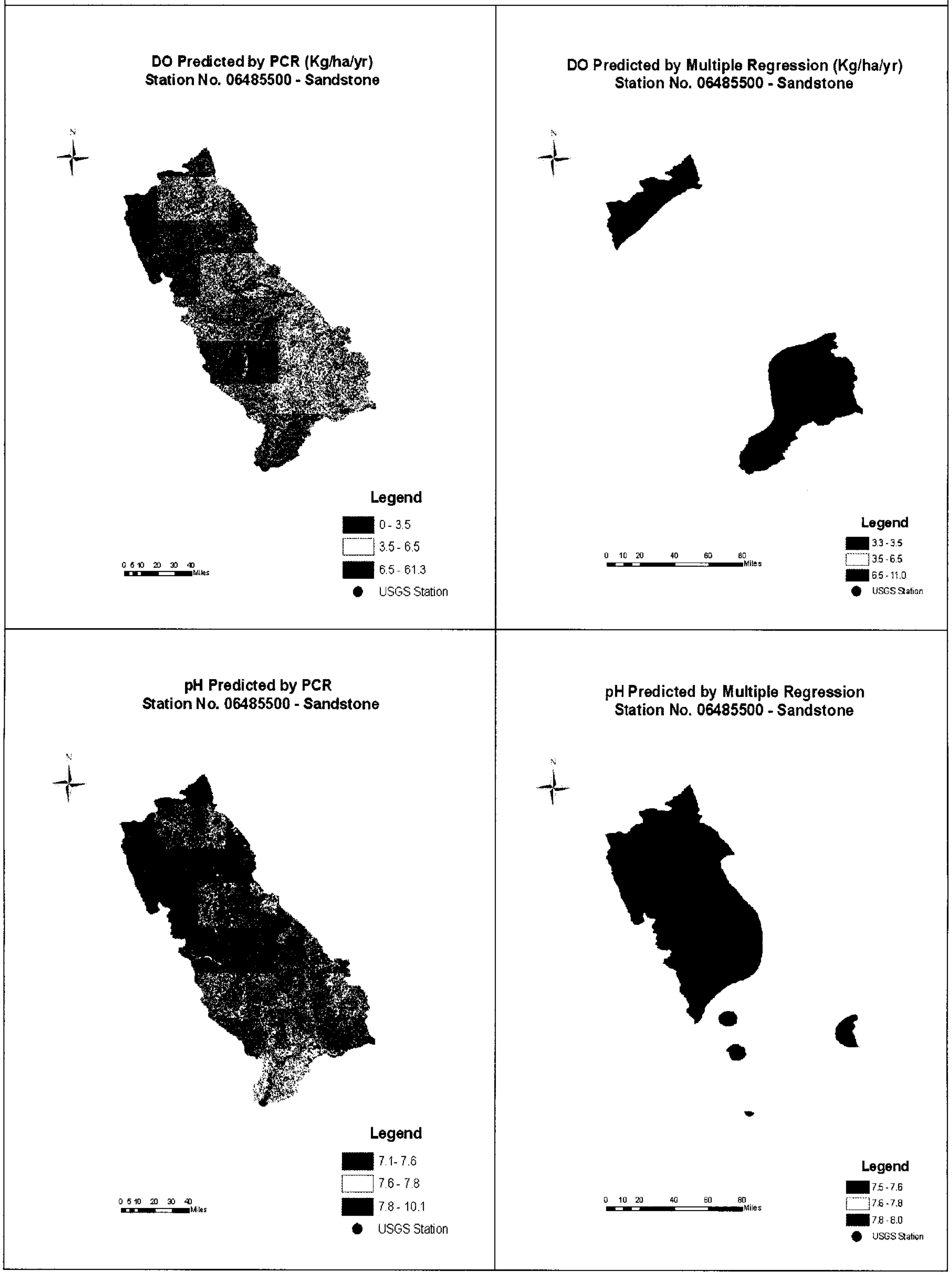
**Table 53. Standard deviation for predictions of multiple regression and PCR using GIS layers.**

| Station No. | DO | | pH | | Total Nitrogen | | K | |
|---|---|---|---|---|---|---|---|---|
| | Multiple Re. | PCR | Multiple Re. | PCR | Multiple Re. | PCR | Multiple Re. | PCR |
| USGS 08053500 | 3.10 | **11.8** | 0.05 | **0.2** | 0.23 | **0.9** | 4.59 | **1.5** |
| USGS 06810000 | 1.83 | **6.8** | 0.04 | **0.2** | 0.15 | **0.6** | 1.86 | **1.0** |
| USGS 06485500 | 1.74 | **4.3** | 0.11 | **0.3** | 0.18 | **0.4** | 2.61 | **0.9** |
| USGS 05413500 | 0.00 | **6.2** | 0.00 | **0.1** | 0.00 | **0.5** | 0.00 | **0.9** |

Generally speaking, as indicated in this study, the regression models for predicting pH are most valid because they have the smallest APE (0.06 – 0.08) (Table 52), and regression models for predicting dissolved oxygen are least valid because of highest APE (1.13, Table 52). However, those models are still very useful for ion yield predictions of those watersheds where no water quality data are available.

For multiple regression perspective, most of the significant regression equations in this study have higher percentage variance of each constituent explained by the related predictors, compared to those of Peters 1984. For example, in Peters 1984 most regression squared correlations were just from 0.4 to 0.7 while in this study, most regression squared correlations are from 0.5 to 0.93 (Table 17, Table 42, and Table 48).

**Figure 14. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 06485500 - Big Sioux R At Akron, IA.**



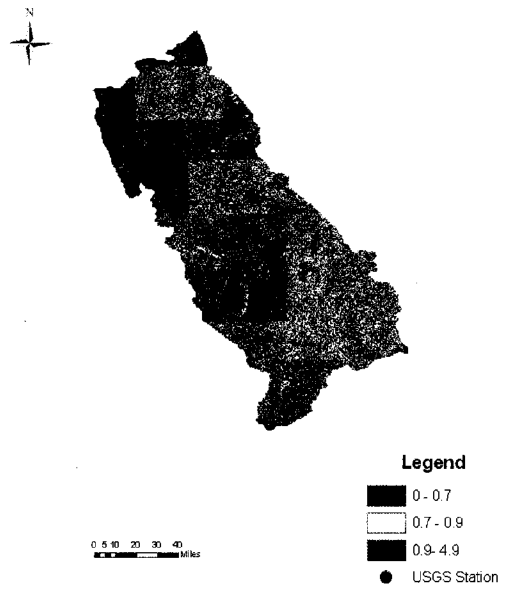
DO Predicted by PCR (Kg/ha/yr)
Station No. 06485500 - Sandstone

DO Predicted by Multiple Regression (Kg/ha/yr)
Station No. 06485500 - Sandstone

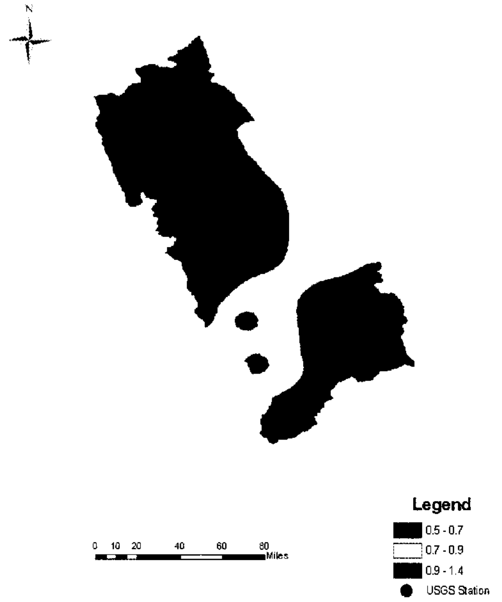
pH Predicted by PCR
Station No. 06485500 - Sandstone

pH Predicted by Multiple Regression
Station No. 06485500 - Sandstone

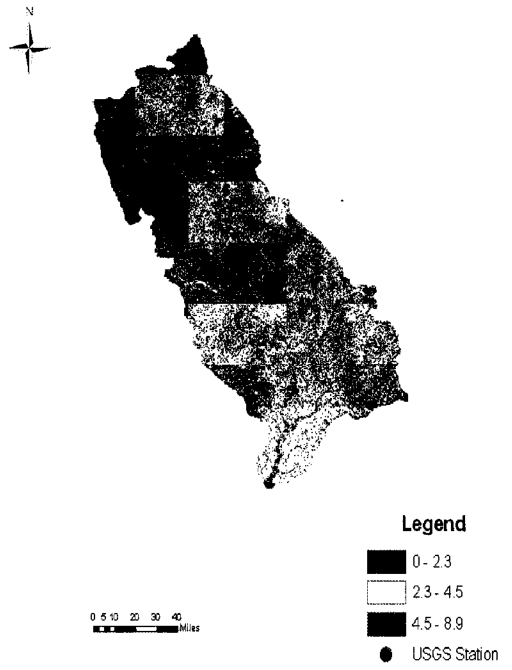
**Total Nitrogen Predicted by PCR (Kg/ha/yr)**
**Station No. 06485500 - Sandstone**



**Total Nitrogen Predicted by Multiple Regression (Kg/ha/yr**
**Station No. 06485500 - Sandstone**



**K Predicted by PCR (Kg/ha/yr)**
**Station No. 06485500 - Sandstone**



**K Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 06485500 - Sandstone**

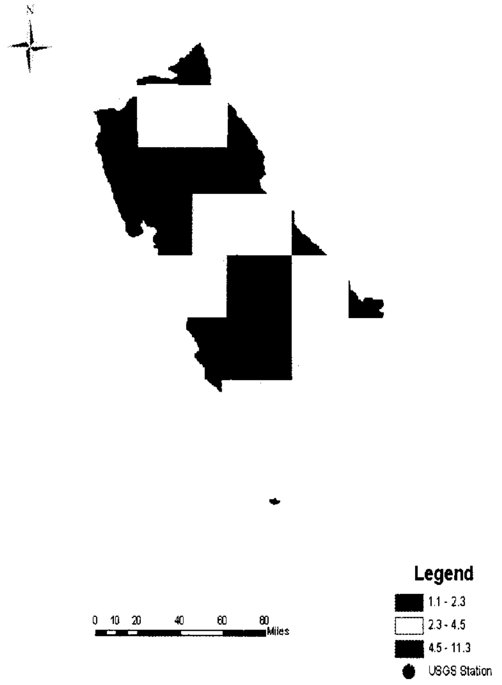
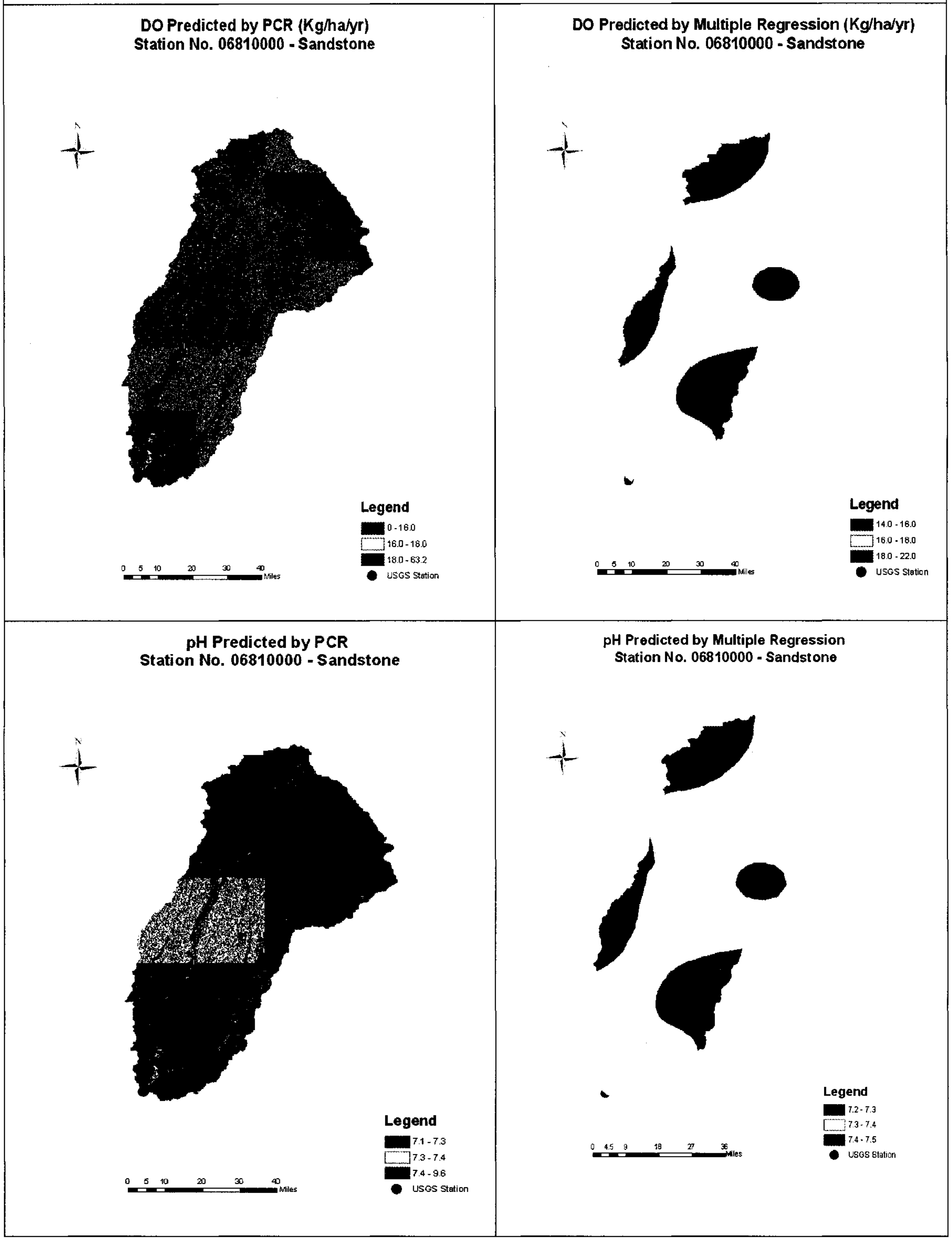


**Figure 15. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 06810000 - Nishnabotna River above Hamburg, IA.**



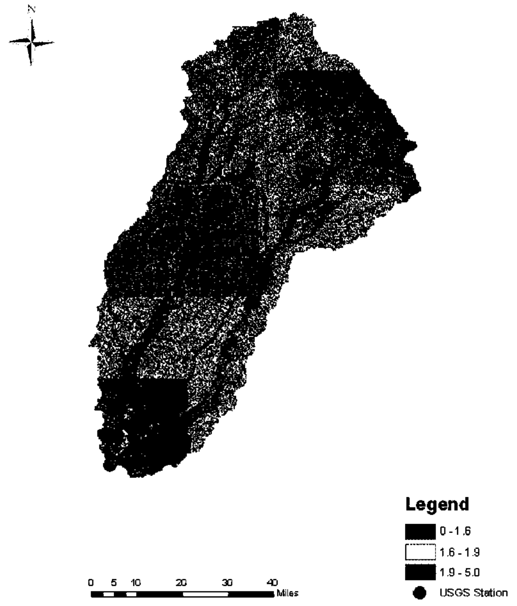
**DO Predicted by PCR (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**

**DO Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**

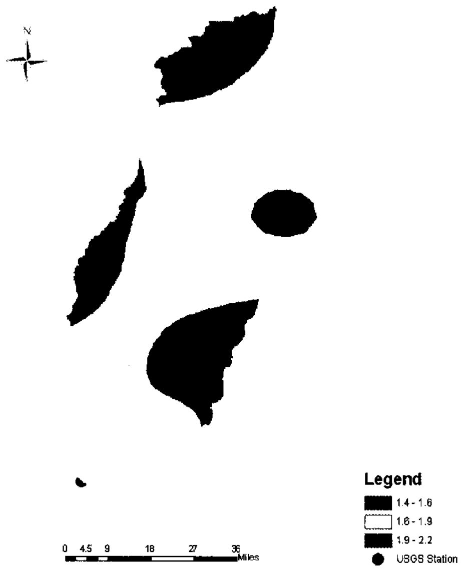
**pH Predicted by PCR**
**Station No. 06810000 - Sandstone**

**pH Predicted by Multiple Regression**
**Station No. 06810000 - Sandstone**

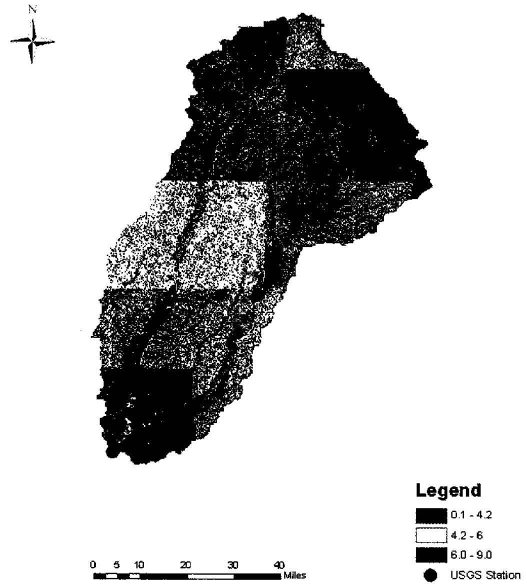
**Total Nitrogen Predicted by PCR (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**



**Total Nitrogen Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**



**K Predicted by PCR (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**



**K Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 06810000 - Sandstone**

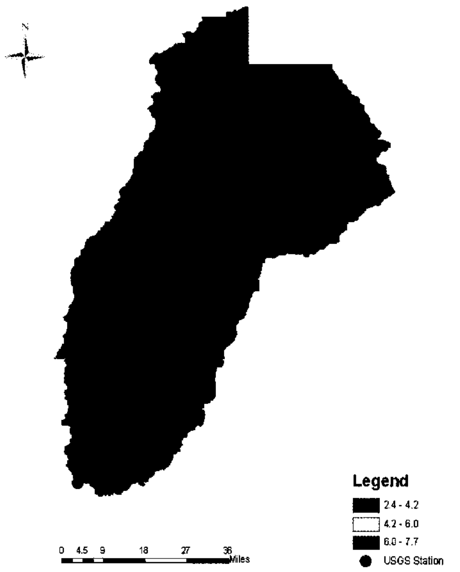
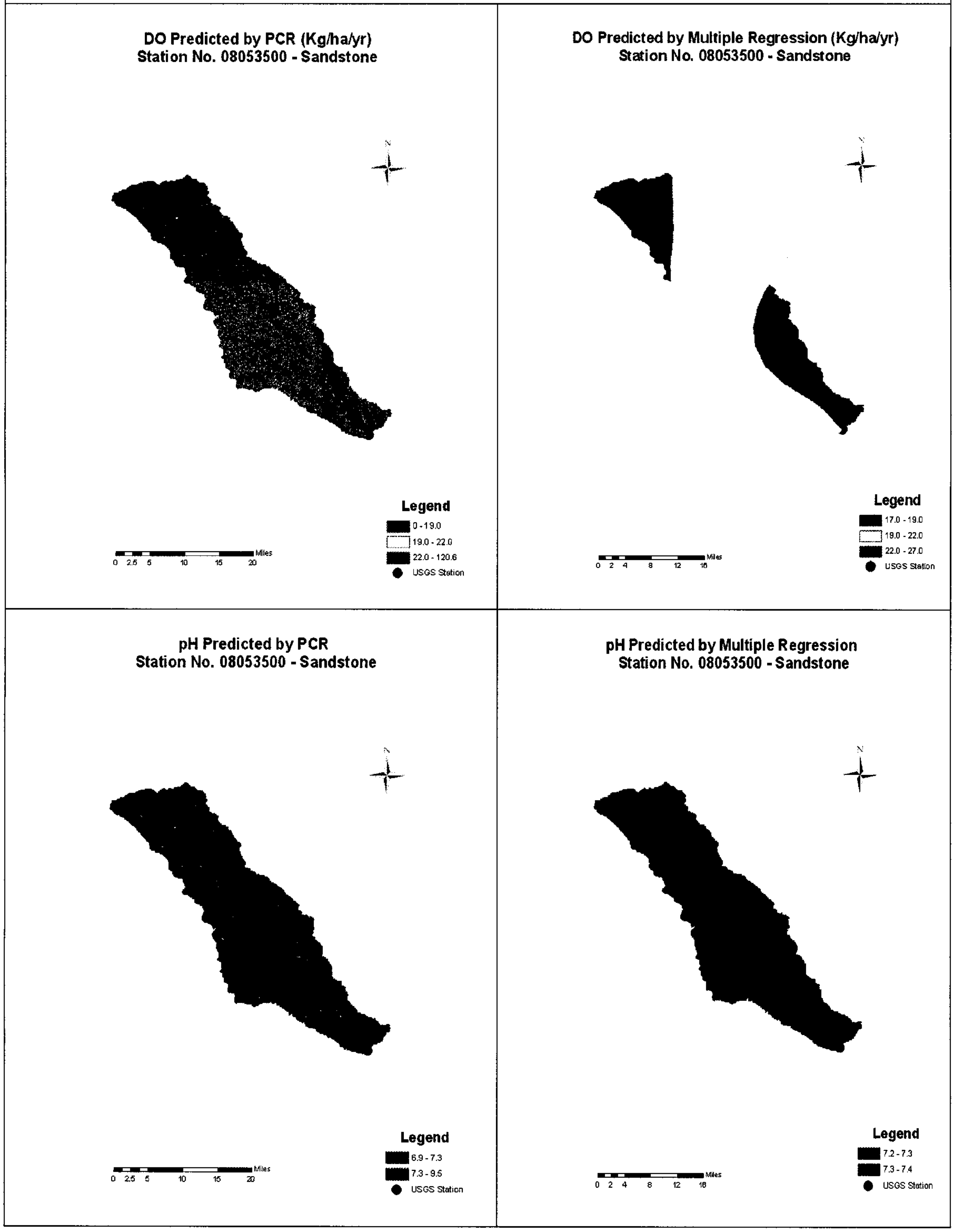
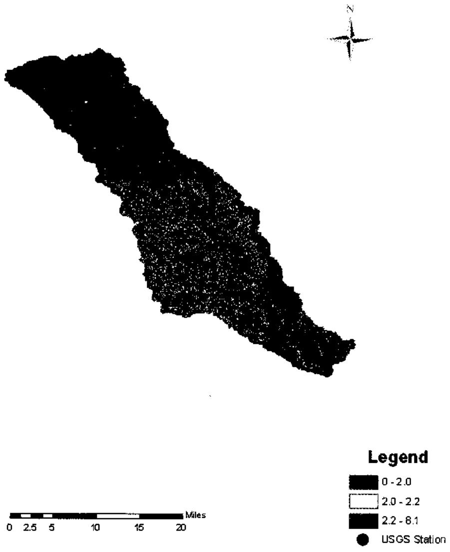


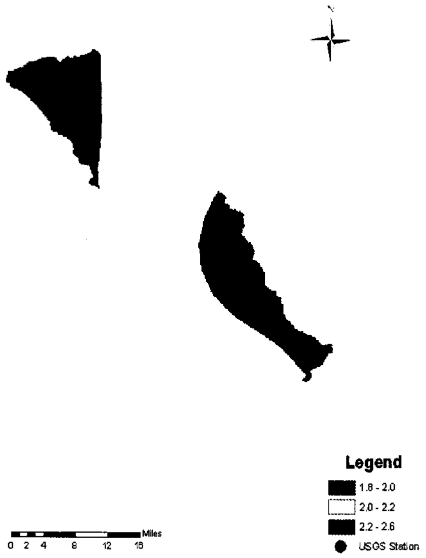
Figure 16. Predicted values of DO, pH, Total Nitrogen, and K using multiple regression and PCR equations for station No. USGS 08053500 - Denton Ck nr Justin, TX.



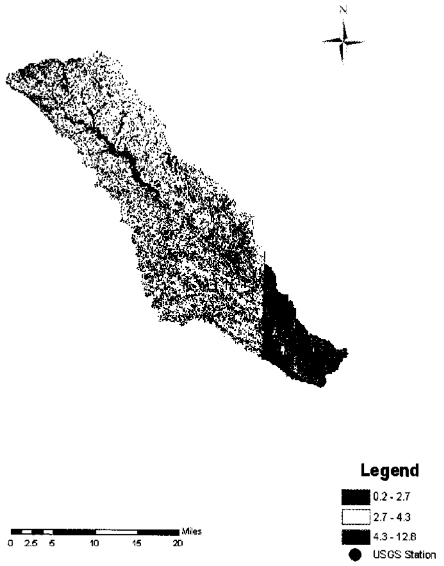
**Total Nitrogen Predicted by PCR (Kg/ha/yr)**
**Station No. 08053500 - Sandstone**



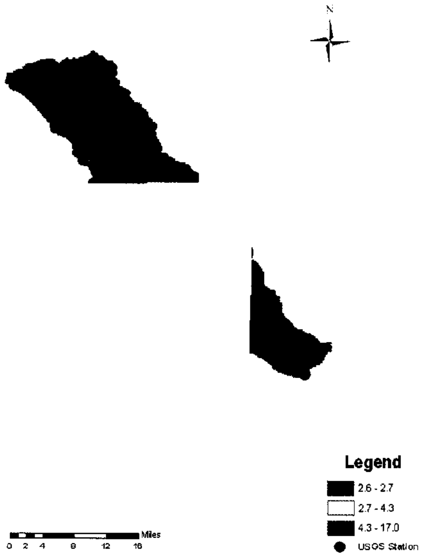
**Total Nitrogen Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 08053500 - Sandstone**



**K Predicted by PCR (Kg/ha/yr)**
**Station No. 08053500 - Sandstone**



**K Predicted by Multiple Regression (Kg/ha/yr)**
**Station No. 08053500 - Sandstone**



# V. CONCLUSIONS

In this study, 13 stream water constituents – total dissolved solids, dissolved oxygen, pH, total nitrogen, orthophosphate, phosphorus, calcium, magnesium, sodium, potassium, sulfate, chloride, and bicarbonate from 47 watersheds in the United States were analyzed in relation to the environmental factors, including bedrock types (limestone, sandstone, and crystalline), land cover type, average annual precipitation, population density, watershed slope, and watershed area.

The results from this study indicate that there is a strong exponential relationship among environmental factors and selected stream water ion yields for selected watersheds. The most important factors affecting annual stream water ion yields are average annual precipitation and bedrock type. Stream water ion yields of limestone bedrock are much different from those of the other two, including total dissolved solids, orthophosphate, calcium, magnesium, and bicarbonate. Sandstone and crystalline ion yields are not very much different, except for chloride and sulfate. For example, some average annual ion yields of limestone watersheds are 2 to 5 times larger than those of sandstone or crystalline watersheds, including total dissolved solids (723 kg/ha/yr), bicarbonate (441 kg/ha/yr), magnesium (36 kg/ha/yr), chloride (61 kg/ha/yr), sulfate (123 kg/ha/yr), and calcium (144 kg/ha/yr). However, calcium is not related to any environmental factor of all three bedrocks except for the principal component regression of calcium in sandstone watershed ($R^2 = 0.406$).

The yields of dissolved oxygen, potassium, sodium, and total nitrogen were not different among the three bedrock types nor was the median pH values. Ion yields in limestone and sandstone bedrocks are more closely related to the environmental factors than those in crystalline bedrock. Most of the ion yields in crystalline bedrock have no relation to the environmental factors, except for dissolved oxygen, pH, and total nitrogen.

For the stepwise multiple regression of this study shows that out of 22 significant equations of all watersheds in the three bedrock types, average annual precipitation appears in 11 equations, including regressions of dissolved oxygen (sandstone-$R^2 = 0.84$ and crystalline-$R^2 = 0.48$ ), pH (in all three bedrocks-$R^2$ ranges from 0.42 to 0.63 and 0.94), total nitrogen (sandstone-$R^2 = 0.79$ and crystalline-$R^2 = 0.33$), phosphorus (sandstone-$R^2 = 0.52$), magnesium (limestone-$R^2 = 0.37$), sodium (limestone-$R^2 = 0.12$), sulfate (limestone-$R^2 = 0.18$), and bicarbonate (limestone-$R^2 = 0.36$).

In the significant regressions, average annual precipitation is either a unique predictor (in 9 equations) or among the most significant predictors. These significant regressions and the corresponding squared regression coefficient values indicate that average annual precipitation in company with other predictors is the most important factor affecting stream water ion yields, accounting for 12 to 94 percent of the variance of ion yields of sodium, sulfate, bicarbonate, magnesium, phosphorus, total nitrogen, dissolved oxygen and pH. The validation of this study indicates that it is applicable to use average annual precipitation to predict pH in all three bedrock types.

Followed by bedrock type and average annual precipitation, population density is the next most important factor in most ion yield prediction. Out of the 22 significant equations of all watersheds in the three bedrock types, population density is found to be in 7 equations, including regressions of total dissolved solids (sandstone-$R^2 = 0.37$), sodium (limestone-$R^2 = 0.54$), magnesium (sandstone-$R^2 = 0.31$), potassium (limestone-$R^2 = 0.26$, sandstone-$R^2 = 0.54$), chloride (limestone-$R^2 = 0.61$), and bicarbonate (sandstone-$R^2 = 0.41$). Like average annual precipitation, in these regressions, population density is either a unique predictor or among the most significant predictors having highest beta values. The significant squared regression coefficient values indicate that population density, in company with other factors, accounted for 26 to 61 percent of the variance of ion yields of potassium, magnesium, total dissolved solids, bicarbonate, sodium, and chloride.

Nevertheless, population is not present in the models of crystalline watersheds. In limestone watersheds, population affects sodium and chloride most, while in sandstone watersheds, it affects total dissolved, magnesium, potassium, and bicarbonate most.

In principal component regression analysis, precipitation and population are also among the highest score coefficient weighted factors, especially for sandstone watersheds. This explains why the predicted yield maps of dissolved oxygen, pH, total nitrogen, and potassium in sandstone watersheds show a similar shape and pattern of population maps.

Land cover type also affects ion yields in limestone and sandstone watersheds. The effects of the four land cover types (***Water, Developed, Forest,*** and ***Agriculture***) on yield prediction are not very different, although ***Developed*** and ***Water*** cover types tend to be more effective. For example, ***Developed*** appears in the regressions of orthophosphate and phosphorus in limestone watersheds ($R^2 = 0.28$), ***Water*** appears in the regressions of orthophosphate in limestone watershed ($R^2 = 0.19$), total dissolved solids ($R^2 = 0.25$), sodium ($R^2$ = 0.45), and bicarbonate ($R^2$ = 0.31) in sandstone watersheds while ***Agriculture*** is found in only a regression of magnesium ($R^2$ = 0.15), and ***Forest*** is found in only a regression of sulfate ($R^2$ = 0.24) in limestone watersheds. These results indicate that effects of land cover types were significant with other factors and accounted for 15 to 45 percent of the variance of ion yields in limestone and sandstone watersheds, including orthophosphate, phosphorus, magnesium, and sulfate (limestone watersheds), total dissolved solids, sodium, and bicarbonate (sandstone watersheds).

Watershed slope was an ineffective predictor, significant only in limestone watersheds in company with other factors, accounted for 12 to 42 percent of variance of ion yields of sodium, potassium, magnesium, and dissolved oxygen. Watershed area was not significant in any ion yield regression analysis.

Multiple linear regression (MLR) provided simpler models but larger R square, smaller standard of error of estimate, and could detect more significant equations than principal component regression (PCR).

The use of GIS-based regression models to predict ion yields introduces more variation and uncertainty because of the larger range of input values, compared to those selected to build the regression models, although the mean predicted values of non-GIS and GIS models are almost the same. This GIS-based approach, however, is very useful for visualization and comparison of different patterns of predicted ion yield.

# VI. RECOMENDATIONS

The digital environmental data obtained by the National Elevation Dataset, National Land Cover Dataset, National Hydrography Dataset, Census data from the National Atlas, and Average Annual Precipitation from Data Gateway have an accuracy level which is most suitable for large areas like state wide or national wide regions, especially for population and land cover datasets. Thus, applying these datasets to obtain data for the small watersheds, such as some of this study, is not very applicable. This explains why land cover types are not effective for the ion yield prediction models in this study.

The frequency of data of this study is based mainly on average annual basis (e.g. precipitation). In fact, constituent yields vary very much in different seasons and even in different months of the year. Thus, shorter time periods of the data is recommended. In addition, the effects of soil types are not considered (due to difficulty of obtaining data), although soil is very important for water quality. Both training and testing sample sizes are small. Point source pollution is not separated from non point sources because of the difficulty of collecting point sources due to the fact that most of the project watersheds are large. The use of mean slope values is not very suitable for large watersheds. These limitations conceal and even can distort the relationship between stream ion yields and the environmental factors. Furthermore, the effects of bedrock types on ion yields are indirectly reflected and are not very explicitly because other factors were not controlled. The GIS-based prediction results were based on a small testing sample size and on a cell-by-cell basis without taking accumulative effects into account. Thus, those results would be used for relative spatial distribution comparison of ion yields only.

The prediction of stream ion yields using 4 selected sandstone watersheds results some high Average Percent Error (APE) values (e.g. APE of total nitrogen). These results might be partly due to some parts of the testing watersheds do not contain sandstones. Nevertheless, prediction is necessary and useful for environmental planning, remediation, and protection. Prediction can be improved by adding soil type into the regression processes, because water quality is very sensitive to soil properties in the drainage system. In addition, the effects of slope and watershed area should be reconsidered, although they are not very effective for ion yield estimation as shown in this study. The use of larger data sets, more accurate, and shorter frequency can also improve the regression models.

GIS-based implementation models are very robust and useful for a spatial delineation of predicted stream water ion yields. However, to reduce the uncertainty introduced by this method, one of the most effective ways is to collect as various ranges of the input data as possible. This would help to assure that the range of the input data in each GIS layer is in the same scale of or not much different from the data ranges used to build regression models.

# LITERATURE CITED

Adamus, C.L. and Bergman, M.J., (1995): *Estimating Nonpoint Source Pollution Loads with a GIS Screening Model.* Water Resources Bulletin 31(4):571-778.

Arabi, M., Govindaraju, R. S., Hantush, M. M., (2004): *Impact of Best Management Practices on Water Quality of Two Small Watersheds in Indiana: Role of Spatial Scale,* as accessed in August 2006 at: http://www.epa.gov/nrmrl/pubs/600r05080/600r05080.pdf

Austin, S. H., (2006). *A Forest Water Quality Literature Review: Abstracted Articles Describing The Water Quality Effects of Forests and Forestry,* as accessed in September 2006 at: http://www.dof.virginia.gov/wq/resources/lit-review-2006_austin.pdf

Brown, D., (2000): *Defining Relationships between Environmental factors and Ion Yields.* MS thesis. Department of Earth Resources, Colorado State University, 2000.

Calder, I.R., (2002). *Forests and Hydrological Services: Reconciling public and science perceptions.* Land Use and Water Resources Research 2 (2002) 2.1–2.12, (http://www.luwrr.com) © Venus Internet

Choi, W. Y. W., Deal, B. M., (2005): *Long-Term Impacts of Land-Use Change on Non-Point Source Pollutant Loads for the St. Louis Metropolitan Area, USA.* Environmental Management Vol. 35, No. 2, pp. 194–205.

Dartiguenave, C. M., (1997): *Water Quality Master Planning for Austin,* Online Publication, CRWR, University of Texas of Austin.

Day, S., Thyne, G., Poeter, E., and J.E. McCray, (2002): *Effects of population growth on water quality in a mountain watershed.* Geol Soc. Amer, as accessed in September 2006 at: http://www.mines.edu/~epoeter/research/TC/Thyne-water-quality-abstr.pdf

Donohue R., Davidson, W. A., Peters, N. E., Nelson S., and Jakowynal, B. (2001). *Trends in total phosphorus and total nitrogen concentrations of tributaries to the Swan–Canning Estuary, 1987 to 1998. Hydrological Process.* 15, 2411–2434 (2001). DOI: 10.1002/hyp.300.

Efron, B. and Tibshirani, R.J., (1991): *Statistical Data Analysis in the Computer Age.* Science 253:390-395.

Endreny, T. A., (2002). *Forest Buffer Strips: Mapping the Water Quality Benefits.* Journal of Forestry. Society of American Foresters. Bethesda, MD (US). Vol. 100. Number 1. January/February 2002.

EPA, (2003). *A Guid to Developing Secondary Information Products: Methods Review and Documentation.* EPA 260-B-01-006, 28 p.

ESRI (2005): Hydrology Modeling, as accessed in October 2005 at: http://edndoc.esri.com/arcobjects/9.0/default.asp?url=/arcobjects/9.0/Samples/SpatialAnalyst/HydrologicModeling/HydrologicModeling.htm

Frankenberger, J.: *Land Use & Water Quality,* as accessed in September 2006 at: http://www.ces.purdue.edu/extmedia/ID/ID-230/ID-230.html

Havel, M., Peters, N.E., and Cerny, J. (1999). *Longitudinal patterns of stream chemistry in a catchment with forest dieback, Czech Republic.* Elsevier Pulishers. Environmental Pollution 104 (1999) 157±167.

Hem, J.D., Demayo, A., and Smith, R.A., (1990): *Hydrogeochemistry of Rivers and Lakes. In: Wolman, M.G., and Riggs, H.C., eds., Surface Water Hydrology: Boulder, Colorado, Geological Society of America*, The Geology of North America, v. O-l. p. 189-231.

Johnson, A. H., and Reynolds, R. C., Jr., (1977): *Chemical character of headwater streams in Vermont and New Hampshire:* Water Resources Research, v. 13, no. 2, p. 469-473.

Jolliffe, I.T, (2002): *Principal Component Analysis.* Second Edition. Springer Series in Statistics, 2002.

Khalil A. F., (2002). Estimation of Runoff and nonpoint source pollution using GIS, as accessed in August 2005 at: http://hydrology.neng.usu.edu/giswr/archive02/akhalil/

Liu, R.X., Kuang, J., Gong, Q., Hou, X. L. (2002): *Principal component regression analysis with SPSS.* Available online 14 May 2002. Computer Methods and Programs in Biomedicine Volume 71, Issue 2 , June 2003, Pages 141-147.

Lystrom, D.J., Rinella, F.A., Rickert, D.A., Zimmerman, L., (1978): *Multiple Regression Modeling Approach for Regional Water Quality Management.* U.S. Environmental Protection Agency EPA-600/7-78-198. 59 p.

Massachusetts Low Impact Development Toolkit, as accessed in October 2006 at:
http://www.mapc.org/regional_planning/LID/PDFs/SiteDesign.pdf
http://www.mapc.org/regional_planning/LID/site_design.html

Mattson, M.D. and Godfrey, P.J., (1994): *Identification of Road Salt Contamination Using Multiple Regression and GIS.* Environmental Management 18(5):767-773.

Melancon, P. A., (1999): *A GIS Based Watershed Analysis System for Tillamook Bay, Oregon*, CRWR, University of Texas, Austin, U.S.A.

Musser, J. J., and Whetstone, G. W., (1964): *Geochemistry of water, in Influences of strip mining on the hydrologic environment of parts of Beaver Creek basin, Kentucky*, 1955-59, C. R. Collier and others, eds.: U.S. Geological Survey Professional Paper 427-B, p. B25-B48.

Naranjo, E., (1997). *A GIS based nonpoint pollution simulation model,* as accessed in May 2004 at: http://gis.esri.com/library/userconf/europroc97/4environment/E2/e2.htm

Osborne, K. G., (2000): *A Water Quality GIS Tool for the City of Austin Incorporating Non Point Sources and Best Management Practices*, CRWR, University of Texas, Austin, U.S.A.

Peters, N.E., (1984): *Evaluation of Environmental Factors Affecting Yields of Major Dissolved Ions of Streams in the United States.* U. S. Geological Survey Water Supply Paper 2228. 39 p.

Peters, N. E. (2004). *Preliminary analysis of the water quality variability of urban streams, Atlanta, Georgia, May 2003-October 2004. Proceedings of the 2005 Georgia Water Resources Conference,* held April 25–27, 2005, at The University of Georgia. Kathryn J. Hatcher, editor, Institute of Ecology, The University of Georgia, Athens, Georgia.

Peters, N. E., Donohue, R., (2001). *Nutrient transport to the Swan–Canning Estuary, Western Australia. Hydroogical Process.* 15, 2555–2577 (2001). DOI: 10.1002/hyp.304.

Peters, N. E., Meybeck, M. (2000). *Water Quality Degradation Effects on Freshwater Availability: Impacts of Human Activities.* International Water Resources Association. Water International, Volume 25, Number 2, Pages 185–193, June 2000.

Peters, N. E, and Turk, J. T., (1981): *Increases in sodium and chloride in the Mohawk River, New York from 1950's to 1970's attributed to road salt:* Water Resources Bulletin.

Project SEARCH: Water Quality Monitoring (and more) for Secondary Schools. Water Quality Testing Manual for the Project SEARCH Monitoring Program. Chapter 7: Chemical Parameters, as accessed in July 2006 at: http://www.sciencecenterct.org/projectsearch/download.htm

Quenzer, A. M., (1988): *A GIS Assessment of the Total Loads and Water Quality in the Corpus Christi Bay System*, CRWR, University of Texas, Austin, U.S.A.

Renard, K. G., Foster, G. R., Weesies, G. A, McCool, D. K., and Yoder, D. C., Coordinators, (1996). *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE).* Agriculture Handbook Number 703, 407 p. United States Department of Agriculture, 1996.

Saunders, W. K., Maidment, D . R., (1996): *A GIS Assessment of Nonpoint Source Pollution in the San Antonio - Nueces Coastal Basin*, CRWR, University of Texas, Austin, U.S.A.

SPSS Inc., SPSS Base 13.0 User's Guide, SPSS Inc., USA, 2004.

SPSS 13.0 for Windows. Released 13.0 (1 Sep 2004). Copyright (c) SPSS Inc., 1989 – 2004. All rights reserved.

Stednick, J.D., (1991): *Wildland Water Quality Sampling and Analysis.* Academic Press, Inc., San Diego, CA. 217 p.

U.S. Geological Survey, (1993): *National Water Summary 1990-1991: Hydrologic Events and Stream Water Quality.* U.S. Geological Water-Supply Paper 2400.

U.S. Environmental Protection Agency, (2000): *National Water Quality Inventory, 2000 Report to Congress.* U.S. Environmental Protection Agency, 207 p, as accessed in June 2005 at: http://www.epa.gov/305b/2000report/

U.S. G.P.O., (1988). *A Legislative History of the Water Quality Act of 1987 (Public Law 100-4)*

Vose, J., Clinton, B., (2006). *Even a Mile of Forest Makes a Difference in Water Quality.* USDA Forest Service Southern Research Station (SRS). 2006, as accessed in October 2006 at:http://www.brightsurf.com/news/headlines/22828/Even_a_Mile_of_Forest_Makes_a_Difference_in_Water_Quality.html

Wong, K. M., Strecker, E. W., Stenstrom, M. K., (1997): *GIS to Estimate Storm-Water Pollutant Mass Loadings*, Journal of Environmental Engineering, Vol. 123, No. 8

**Websites referenced for Land Cover and National Elevation Dataset:**

http://seamless.usgs.gov/website/seamless/viewer.php

http://edcftp.cr.usgs.gov/pub/data/landcover/states/

http://www.mrlc.gov/

http://gisdata.usgs.net/website/MRLC/viewer.php

http://www.mrlc.gov/scripts/mapserv.exe?map=d%3A%5CInetpub%5Cwwwroot%5Clccp%5Cmrlc2k%5Czones%5Czones.map

**Websites referenced for the accuracy of land cover:**

http://landcover.usgs.gov/accuracy/

**Websites referenced for Population Density of the US in 1990:**

http://www.nationalatlas.gov

**Websites referenced for Precipitation and Hydrologic Units:**

http://datagateway.nrcs.usda.gov/GatewayHome.html

**Websites referenced for Hydrologic Units and Flowlines:**

http://nhd.usgs.gov/

**Websites referenced for Water Quality Data:**

http://water.usgs.gov/waterwatch/

http://waterdata.usgs.gov/nwis

**Websites referenced for additional (testing) sandstone stations:**

http://digicoll.library.wisc.edu/cgi-bin/EcoNatRes/EcoNatRes-idx?type=div&did=ECONATRES.0027.0280.0010&isize=M

http://72.14.203.104/search?q=cache:CEfXiJCFhYsJ:www.seta.iastate.edu/publicservices/water/plan/chapter3/+sandstone+of+of+cambrian-ordovician+bedrock&hl=en&gl=us&ct=clnk&cd=9

http://72.14.203.104/search?q=cache:iWHCtloK_VoJ:courses.unt.edu/hwilliams/lab_7.htm+sandstone+bedrock+and+water&hl=en&gl=us&ct=clnk&cd=9

http://www.igsb.uiowa.edu/gsbpubs/pdf/OFM-1997-1.pdf

http://www.igsb.uiowa.edu/gsbpubs/pdf/OFM-2003-1.pdf

http://plantsci.sdstate.edu/woodardh/Geology/Eastern/General_Geology/general_geology.htm

http://www.northern.edu/natsource/EARTH/Aquife1.htm

http://www.beg.utexas.edu/UTopia/images/pagesizemaps/physiography.pdf#search=%22bedrock%20map%20of%20texas%22

http://www.esi.utexas.edu/outreach/groundwater/resources/maps/trinity.pdf#search=%22antlers%20formation%20of%20texas%22

http://www.cretaceousfossils.com/formations/trinity_group/paluxy_sandstone.htm

# APPENDICES

All appendices are recoded in a CD attached in the inner side of the back cover of the dissertation. They can be open, printed, and copied. Appendices 1, 2, 3, and 7 are in Excel Worksheets. Appendices 4, 5, 6, and 8 are in Microsoft Word Documents.

**Appendix 1:** Station Information.

**Appendix 2a.** Water Quality Data and Corresponding Number of Records Compiled from http://waterdata.usgs.gov/nwis for the Project Watersheds in Limestone Bedrock.

**Appendix 2b.** Water Quality Data and Corresponding Number of Records Compiled from http://waterdata.usgs.gov/nwis for the Project Watersheds in Sandstone Bedrock.

**Appendix 2c.** Water Quality Data and Corresponding Number of Records Compiled from http://waterdata.usgs.gov/nwis for the Project Watersheds in Crystalline Bedrock.

**Appendix 3a_1.** Land Cover, Precipitation, Population, Slope, and Area Data in the Project Watersheds in Limestone Bedrock.

**Appendix 3a_2.** Water Qulaity Data in the Project Watersheds in Limestone Bedrock (kg/ha/yr).

**Appendix 3a_3.** Combined Land Cover Classes in the Project Watersheds in Limestone Bedrock (% & SQMI).

**Appendix 3a_4.** Explaination of Land Cover codes and Combined Land Cover Classes.

**Appendix 3b_1.** Land Cover, Precipitation, Population, Slope, and Area Data in the Project Watersheds in Sandstone Bedrock.

**Appendix 3b_2.** Water Qulaity Data in the Project Watersheds in Sandstone Bedrock (kg/ha/yr).

**Appendix 3b_3.** Combined Land Cover Classes in the Project Watersheds in Sandstone Bedrock (% & SQMI).

**Appendix 3c_1.** Land Cover, Precipitation, Population, Slope, and Area Data in the Project Watersheds in Crystalline Bedrock.

**Appendix 3c_2.** Water Qulaity Data in the Project Watersheds in Crystalline Bedrock (kg/ha/yr).

**Appendix 3c_3.** Combined Land Cover Classes in the Project Watersheds in Crystalline Bedrock (% & SQMI).

**Appendix 4_1.** Correlation matrix and Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and "Enter" Method of Model Selection to see a Multicollinearity Diagnosis for Watersheds in Limestone Bedrock.

**Appendix 4_2.** Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and "Stepwise" Method of Model Selection for Watersheds in Limestone Bedrock ($\alpha = 0.05$).

**Appendix 4_3.** Factor Analysis and Principal Component Regression among the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Stepwise” Method of Model Selection for Watersheds in Limestone Bedrock ($\alpha = 0.05$).

**Appendix 5_1.** Correlation matrix and Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Enter” Method of Model Selection to see a Multicollinearity Diagnosis for Watersheds in Sandstone Bedrock.

**Appendix 5_2.** Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Stepwise” Method of Model Selection for Watersheds in Sandstone Bedrock ($\alpha = 0.05$).

**Appendix 5_3.** Factor Analysis and Principal Component Regression among the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Stepwise” Method of Model Selection for Watersheds in Sandstone Bedrock ($\alpha = 0.05$).

**Appendix 6_1.** Correlation matrix and Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Enter” Method of Model Selection to see a Multicollinearity Diagnosis for Watersheds in Crystalline Bedrock.

**Appendix 6_2.** Multiple Linear Regression between the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and “Stepwise” Method of Model Selection for Watersheds in Crystalline Bedrock ($\alpha = 0.05$).

**Appendix 6_3.** Factor Analysis and Principal Component Regression among the Predefined Environmental Factors and Stream Water Ion Yields (kg/ha/yr) Using Log-Transformed Values and "Stepwise" Method of Model Selection for Watersheds in Crystalline Bedrock ($\alpha = 0.05$).

**Appendix 7.** Comparison between average annual ion yields (kg/ha/yr) from Peters's research in 1984 and the current research.

**Appendix 8.** Watersheds Delineated in the Three Bedrock Types.