2007

DETERMINING FACTORS THAT CONTRIBUTE TO NONPOINT SOURCE POLLUTION IN THE LOWER KENTUCKY WATERSHED

James A. Wermuth
University of Kentucky, jawerm0@yahoo.com

ABSTRACT OF THESIS

DETERMINING FACTORS THAT CONTRIBUTE TO NONPOINT SOURCE POLLUTION IN THE LOWER KENTUCKY WATERSHED

The water quality in the United States has greatly improved since the implementation of the Clean Water Act (CWA) in the early 1970's. Unfortunately, the Clean Water Act only addresses one kind of water pollution, point source pollution. The major problem that is present in the degradation of today's water quality has to deal with nonpoint source pollution. Agriculture is commonly regarded as the leading contributor to nonpoint source pollution in the United States. This study uses two analytical tools to try to determine the significant factors in the transport of pollutants in the Lower Kentucky Watershed, located in central Kentucky. Spatial analysis (GIS) coupled with the statistical analysis (SAS), allowed for significant factors to be identified within a small proximity of sampling sites throughout the watershed. The results suggest that although agriculture is commonly regarded as the largest contributor to nonpoint source pollution, other factors outside of agriculture were also found to be significant, such as resident land use and rainfall. The results generated from this study suggest that land managers in communities throughout the watershed should analyze agricultural factors, as well as, factors outside of agriculture, in an effort to protect their communities' water quality.

KEYWORDS: Nonpoint Source Pollution (NPS), Nutrients, Geographic Information System (GIS), Siltation, Pathogens

James A. Wermuth

12-12-06

DETERMINING FACTORS THAT CONTRIBUTE TO NONPOINT SOURCE POLLUTION IN THE LOWER KENTUCKY WATERSHED

By

James Allen Wermuth

Angelos Pagoulatos
Director of Thesis

David Freshwater
Director of Graduate Studies

12-12-06

RULES FOR USE OF THESIS

Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

<u>Name</u> <u>Date</u>

THESIS

James A. Wermuth

The Graduate School

University of Kentucky

2006

DETERMINING FACTORS THAT CONTRIBUTE TO NONPOINT SOURCE POLLUTION IN THE LOWER KENTUCKY WATERSHED

THESIS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the College of Agriculture at the University of Kentucky

By

James A. Wermuth

Lexington, KY

Director: Dr. Angelos Pagoulatos, Professor of Agricultural Economics

Lexington, KY

2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FILES

CHAPTER ONE: Introduction

The water quality of the waterways in the United States has been an important issue, dating back to the early 1970's and the creation of the Clean Water Act (CWA). This is one of the pioneer and most influential pieces of environmental legislation to ever be passed in the United States, and has been relatively successful over the years. The passing of the Clean Water Act has been tremendously successful in the reduction of chemicals present in waterways from point sources, but the reduction of pollutants from nonpoint source pollution has not been nearly successful. The Environmental Protection Agency (EPA) has reported over one third of the river miles, lake acres, and estuary square miles are impaired (USDA / ERS).

Though there are many forms of nonpoint source pollution, agriculture is generally recognized as the leading contributor to nonpoint source pollution in the United States (USDA / ERS). Some of the major sources of nonpoint source pollution from agriculture have been classified as nutrients, pathogens, and sediment. The difficulty of dealing with nonpoint source pollution from agriculture (or any other nonpoint source) is that nonpoint source pollution does not enter waterways at a specific location that is easily identifiable like point source pollution.

Throughout the literature, many studies have focused on finding new spatial and statistical methods to try to determine significant factors in the transport of pollutants into our nation's waterways. Many regions throughout the United States have been analyzed, in an effort to try to determine relationships that are present in different geographical regions. Performing studies in different regions of the United States to find significant factors in the transport of pollutants into waterways is vital research that needs to be performed, due to the fact that the significant factors to pollutant transport will differ according to the characteristics of the research area.

The purpose of this study is to try to determine sampling loads at individual sampling sites in the Lower Kentucky Watershed by looking at adjacent land use(s) and other explanatory variables, in an effort to try to determine the factors that are significant in the transport of pollutants into waterways. If significant factors can be identified before the pollutants have a chance to enter the waterway, then new policy recommendations can be developed to help with the reduction of common pollutants from agriculture and other sources. This study allows for nonpoint source pollution to be viewed in an aspect of correcting the problem before it has occurred and could have a major impact on the new kind of policies and land management techniques that are employed to improve water quality.

This study will require the use of two analytical tools to spatially and statistically correlate land use(s) and other explanatory variables to the water quality problems that exist from sediment, nutrients, and pathogens. The first analytical tool that needs to be used in the study is a Geographical Information Systems (GIS), such as ArcMap. ArcMap will be used to create a spatial correlation between individual sampling sites' loads and the explanatory variables that are relevant to the specific location. The use of ArcMap will also allow for areas to be calculated that will aid in data manipulation on the explanatory variables that are present within the buffer zones surrounding each sampling site.

The areas that are calculated by ArcMap will then be able to be analyzed and manipulated with the second analytical tool that is necessary for this study, Statistical Analysis Software. Statistical Analysis Software (SAS) is the analytical tool that will be used to run regressions on the data present in the model. After SAS has performed the regression, significant variables can be identified and conclusions and policy recommendations can be made for future reference.

CHAPTER TWO: Background Information

Throughout the history of the United States agriculture has been a sector of the economy that has been able to thrive, due to the necessity of its products. The United States, with its abundance of arable land, has continued to play a major role in the world economy. Even though the agriculture commodity prices have not kept up with other United State's industry prices, allowing for cheap food prices, the United States had farm receipts that totaled over $240 billion in 2004 (ERS, 2006). Over $62 billion worth of agricultural products are exported each year, allowing agricultural commodities to help offset the United States' trade deficits from other sectors of the economy (FATUS, 2006). Other industries that perform value added activities, such as the agribusiness industry, have thrived in the United States due to the economic importance of the agriculture industry. The economic importance of the agricultural industry is shown in the United States by its portion of the U.S. gross domestic product ($1.26 trillion in 2000) and by the amount of individuals that it employs throughout the U.S. (just over one-eighth of the U.S. civilian labor force or 24.1 million workers) (EPA, 2006). The financial benefits that are created by the United State's agriculture industry are outstanding, but along with the monetary benefits come costs in the form of negative impacts on our environment.

In the past and still currently, the government has given farmers aid in the United States based on production, where the more that they could produce the better. This impression has been made by the government in the form of subsidies, where farmers received these subsidies based on output produced. The government has provided more than $113 billion in specific commodity subsidies to farmers around the United States from 1995-2004 (EWG). With the production subsidies in place, many of the United States' natural resources have been degraded due to excessive measures taken by farmers in an effort to receive as many subsidies as possible from their increased production. This degradation that impacts our environment is one of the externalities that occur from intense agricultural production. Externalities occur when impacts, either positive or negative, are felt by individuals that are not part of a transaction. In this case negative impacts on our environment, felt by others in an area, are externalities that are associated with agricultural production. In particular, our nation's waterways have been negatively affected from some of the reckless practices that have been used in the past to try to maximize output. Negative externalities from intense production and cultivation on farmland and pastureland have been quantified in the amount of $419.4 million per year (Tegtmeier &

Duffy, 2004). After years of output related subsidies, it was evident that actions needed to be taken to promote more environmentally friendly practices from farmers across the country. The outcome would be new legislation and programs that were put into place in an effort to protect our nation's natural resources.

In the late 1960's and early 1970's growing public awareness and concern for controlling water pollution led to enactment of the Federal Water Pollution Control Act Amendments of 1972 (EPA). As amended in 1977, this law became commonly known as the Clean Water Act (CWA). The Clean Water Act established the basic structure for regulating discharges of pollutants into the waters of the United States (EPA). This is the first and the main piece of legislation that is used to guide water quality legislation and programs around the United States. The Clean Water Act embodies a philosophy of federal/state partnership in which the federal government (U.S. Environmental Protection Agency) sets the agenda and standards for pollution abatement, while the states carry out day-to-day activities such as: authority to issue discharge permits to industries and municipalities, to enforce permits, and to establish water quality standards (KY DOW). Also, according to the Clean Water Act Section 305(b) the states are required to assess and report current water quality conditions to EPA every two years. The data collected is analyzed to inspect and see what waterways are supporting their designated use(s) and what waterways are classified as impaired (EPA).

With the negative impacts from intensive farming techniques, it was evident that new incentive programs needed to be created to entice farmers to use new techniques, rather than the techniques used in past years that helped contribute to the current water quality situation. The result was the creation of new subsidy programs, sometimes referred to as "green programs", which were instituted to help encourage farmers to practice friendlier techniques in their cultivation practices. If new techniques were voluntarily adopted by farmers in selected regions throughout the United States, the government would then reward the land owners with a subsidy payment for their environmental enhancement. These kind of programs have began to accumulate over the last several years as a new monetary support system used by the government to get funds to land owners. Commodity subsidies are still the dominant form of subsidy that farmers receive from the government, but a transition is needed to reduce the negative externalities that have been felt from the previous production techniques used by farmers. The new conservation form of subsidies are a new, more socially appropriate way to get farmers funds, which in the past have only been given to them based on the production of certain commodities.

CHAPTER THREE: Problem Statement

In May 2002, President George W. Bush signed the Farm Bill, providing up to $13 billion for conservation programs for six years. This Farm Bill represents an 80 percent increase above current levels of funding available for conservation programs designed to prevent polluted runoff (EPA). There have been several programs created to help protect our natural resources, mainly our waterways, and allow them to support their previous uses by providing financial and technical assistance to land owners. Some of the most well-known programs that are used around the country to promote conservation by farmers are the Conservation Reserve Program (CRP), the Environmental Quality incentive Program (EQIP), and the Conservation Security Program (CSP). Each of these programs were instituted to provide different kinds of assistance to farmers, with the goal of creating a healthier environment. The Conservation Reserve Program provides technical and financial assistance to eligible farmers and ranchers to address soil, water, and related natural resource concerns on their lands in an environmentally beneficial and cost-effective manner (NRCS). The Environmental Quality Incentive Program (EQIP) was authorized to provide a voluntary conservation program for farmers and ranchers that promote agricultural production and environmental quality as compatible national goals (NRCS). The Conservation Security Program is a program that provides assistance to promote the conservation and improvement of soil, water, air, energy, plant and animal life, and other conservation purposes on Tribal and private working lands (NRCS).

The present condition of water quality in the United States is better than it was almost forty years ago, due to legislation and programs that have been instituted over that same time span. Unfortunately, there is still a substantial amount of pollution from multiple sources in our waterways causing impairments. It is important that measures are taken to control what substances enter our rivers, streams, and lakes, providing individuals with safe drinking water, recreational uses, and a safe ecosystem environment for aquatic organisms.

In trying to control pollution, we must first know some basic information on how and what substances enter our waterways. Pollution is discharged into water from two distinct sources: point sources and nonpoint sources. Point sources discharge effluent directly into water resources through an identifiable pipe, ditch, or other conveyance. Industrial and municipal discharges fall into this category (AER-782). Point sources have been identified as contributing nearly 1.5 million metric tons of nitrogen and 330,000 metric tons of phosphorus per year into the United State's surface waters (Revenga & Mock, 2000). Even though these numbers can look astounding, point source pollution has been the form of pollution that has been reduced over the years. Water quality improvements have been largely due to reductions in toxic and organic chemical loadings from point sources, where toxic pollutant reductions have been reduced by an estimated one billion pounds per year (AER-782). These reductions have been mainly due to the fact of the provisions that were provided in the Clean Water Act, making it unlawful for any person to discharge any pollutant from a point source into navigable waters, without a permit obtained under its provisions. The permit system and water quality standards created by the Clean Water Act have been vital in the pollution reductions that have been seen over the past several years, since its institution.

The real problem with water pollution today is captured by the second form of water pollution, nonpoint source pollution. Nonpoint source pollution enters water diffusely in the runoff or leachate from rain or melting snow, and is often a function of land use(s) (AER-782). Nonpoint source pollution is much more difficult to trace back to the original source of the pollutants because it enters waterways from many points across an area of land, where point sources enter waterways at an identifiable point. Nonpoint source pollution from cropland and pastures alone have been identified as contributing almost 3.5 million metric tons of nitrogen and 710,000 metric tons of phosphorus per year to the United State's surface waters (Revenga & Mock, 2000). Without an identifiable source, like point source pollution, the provisions set forth by the Clean Water Act have had a much smaller impact on nonpoint source pollution. Provisions that were set forth by the Clean Water Act, such as water quality sampling, have not been applied to nonpoint source pollution in the manner that they have been applied to point source pollution. This is shown by a lack of sampling information of water quality data from one state to another. In 2000, only 19% of the total streams and river miles were assessed and only 43% of the total acres of lakes across the United States were assessed (EPA).

One thing is consistent throughout literature, and that is that agriculture is recognized as the largest contributor to nonpoint source pollution in the United States with pollutants such as sediments, nutrients (mainly nitrogen and phosphorus), and fecal material (pathogens) being some of the most reported pollutants (AER-782). A U.S. Geological Survey study of agricultural lands in watersheds with poor water quality estimated that 71 percent of U.S. cropland (nearly

300 million acres) is located in watersheds where the concentration of at least one of the four common surface-water contaminants (dissolved nitrate, total phosphorus, fecal coliform bacteria, and suspended sediment) exceeds criteria for supporting water-based recreation (AER-782).

Siltation is the leading pollution problem in U.S. rivers and streams (AER-782). Sediment enters waterways through erosion, and the damages from agricultural erosion have been estimated to be between $2 billion and $8 billion per year (AER-782). Sediment that is carried off of agricultural land is often referred to as a “double edge sword”, in that the sediment often carries other sources of nonpoint source pollution (e.g. nutrients). Nutrients are the leading cause of impairments in lakes and the second leading cause of impairments in rivers (AER-782). Nutrients enter surface water through runoff, which transports pollutants over the soil surface by rainwater, melting snow, or irrigation water that does not soak into the soil. Nutrients move from fields to surface water while dissolved in runoff water or absorbed to eroded soil particles (AER-782). Excessive nutrients (nitrogen & phosphorus) found in a water way can cause a over production of algae and other plankton choking the waterway, as these organisms die they use up vital oxygen in the water causing a condition known as eutrophication or over-fertilization (KY River Assessment). Pathogens are the third leading cause of impairments in rivers across the United States (AER-782). Pathogens from agriculture enter waterways through animal waste in surface water runoff or by direct contact with a waterway from an animal. Microorganisms in livestock waste can cause several diseases through direct contact with contaminated water or consumption of contaminated drinking water (AER-782). The presence of fecal contamination is an indicator that a potential health risk exists for individuals exposed to the water (KY River Assessment).

Selecting a Proper Region to Investigate

Agriculture is the leading cause of impairments from nonpoint source pollution in rivers and lakes across the nation (AER-782), where more than one third of the nation’s waterways are impaired. It is evident that new economic and educational tools need to be developed to help assist government agencies in finding ways for the reduction of agriculture’s nonpoint source pollution. Many states offer funding to farmers for the implementation of best management practices (BMP’s) on a voluntary basis, but additional measures must be taken to ensure the quality of our nation’s waters. The purpose of this study is to determine what affect explanatory variables have on the sampling data that was captured individual sampling sites throughout a selected region.

There are several criteria that need to be met, in order to select an area that will have the proper characteristics needed to perform the study. Dealing with a water related topic would suggest that the area to be analyzed would be a watershed. A watershed is defined as "a geographic area in which all water running off the land drains to a specific location: creek, river, or stream" (KY DOW). The United States is broken down into over 2,000 eight digit watersheds or hydrological unit code (HUC), where the average watershed size is 748 square miles (KY DOW). A hydrologic unit code (HUC) is identified by a unique code consisting of two to fourteen digits based on the levels of classification in the hydrologic unit system. Hydrological unit codes (HUC's) were developed for the purpose of water-resources planning and data management. A proper watershed to be selected needs to be at least the average watershed size in square miles, but choosing a larger, more substantial watershed would allow for a larger area to be analyzed using the same model. Choosing a larger watershed or HUC would allow for water-resource planning to be performed on a larger scale, allowing the results to be utilized over a larger region.

Several similar studies have been performed in portions of Ohio in the Maumee and Sandusky River basins (Moog & Whiting, 2002), which have tried to identify which significant climatic, hydrological, and agricultural variables, explain the variation in nitrate, phosphorous, and total suspended solids. The Maumee and Sandusky River basins are categorized as primarily agricultural areas that feed into Lake Erie (Moog & Whiting 2002). The fact that these previous studies have been conducted in watersheds that are classified primarily as agricultural watersheds would lead to another criteria that must be met to find a sufficient area to study. The selected area must be a watershed that is classified as agricultural to meet the condition that present in these previous studies.

The Lower Kentucky watershed (05100205), located in central Kentucky, is a watershed that meets the two necessary qualifications. The Lower Kentucky watershed is located in the Bluegrass physiographic region of Kentucky, which is characterized by "hilly or undulating terrain, medium to very rapid rates of surface runoff, and slow to medium groundwater drainage." (KRB Mgt. Plan 2002) It consists of all or a portion of 23 counties in central Kentucky and covers 3,200 square miles (USGS 2006), classifying it as a much larger than the average watershed across the United States. In selecting a large watershed, the results from the study can then be employed over a much larger geographical area.

Table 3.1 Counties Contained In the Lower Kentucky Watershed

Anderson	Gallatin	Mercer
Boone	Garrard	Owen
Boyle	Grant	Rockcastle
Carroll	Henry	Scott
Casey	Jessamine	Shelby
Clark	Kenton	Trimble
Fayette	Lincoln	Woodford
Franklin	Madison	

Figure 3.1 Lower Kentucky Watershed

* *http://www.ky.nrcs.usda.gov/programs/CSP/Watershed1.html*

The land use statistics of the Lower Kentucky watershed show that nearly three-fourths of the land (70.25%) is characterized as agricultural land (KY River Assessment). The land use statistics for the watershed would classify it as primarily agricultural, even though the landscapes of central Kentucky and north western Ohio are very different. The land of north western Ohio is defined by relatively flat land that is ideal for row crops, where the land in the central Kentucky is described as rolling hills that is used mainly as pastureland for livestock. Examining different geographical landscapes will allow for new policy implications to be created and those implications to be applied to a wider variety of geographical areas across the nation.

The Lower Kentucky watershed meets the two main criteria to be selected as the study area, but it also has several other important policy issues surrounding it that make it an important selection area. The three main loads that will be examined in this study are sediment (total suspended solids), pathogens (fecal coliform), and nutrients (nitrate-nitrogen and total recoverable phosphorus), which are all placing a burden on both recreational and biological uses of Kentucky's waters. This area continually receives sampling data results that represent a continuing nutrient problem in the region (KWRI, 2002). The entire state of Kentucky has been categorized as having suspended solids as one of its biggest water quality problems (KWRI, 2002). Lastly, the presence of fecal contamination in water poses a threat to any organism or individual that comes into contact with the contaminated waterway (KWRI, 2002). It is important that this water quality data is analyzed to see what suggestions can come from the results, allowing for new policies to be implemented in land use management that can help reverse the common trends of the region.

CHAPTER FOUR: Previous Literature

Numerous studies have been conducted throughout the United States to try to determine the relationship between water quality and explanatory variables, where many different techniques have been utilized and many explanatory variables have been examined. There are three major tools employed to try to examine and analyze water quality issues, which include the following: statistical analysis, the use of a Geographical Information System (GIS), or some combination of statistical analysis and a GIS.

Langley (2004) employed the use of a GIS and statistical analysis to see what connections were evident between land use and water quality in the Kentucky River Basin. Langley performed similar techniques that will be employed in this study, but several key steps are done differently. The use of buffer zones are employed in Langley's study to create a buffer zone around the impaired reaches (streams) throughout the mapped 14-digit HUC's to look for relationships of land use and water quality in the "immediate vicinity". A similar technique is performed in this study, where 3km buffer zones are created around each of the sampling sites (instead of creating buffers around the reach itself) to only capture land use and other explanatory variables that are in the "immediate vicinity" of the sampling site. In Langley's study each buffer was created and then they were assigned a 1 (impaired) or a 0 (non-impaired). Due to correlations between the independent variables included in the model factor analysis was performed, according to the principal components method to group factors that described the largest patterns of land use variation within the 14-digit HUC's (Langley, 2004). After grouping the land use values into factors, factors with eigenvalues greater than one were kept to be included in the factor analysis. The results of Langley's univariate regressions revealed that nutrients, pathogens, and organic enrichment generated significant results (organic enrichment will not be analyzed in this study.) (Langley, 2004) Nutrient loadings were contributed to by an increase in percent of pasture land, a loss of deciduous forest, high density residential land use, percent of urban / recreational grasses, and low density residential areas, where the coefficients on the residential and urban land uses were much larger than the coefficients for the agricultural land uses. Pathogens were not correlated with agricultural land uses, but were found to be extremely correlated with urban and residential land uses (Langley, 2004). The coefficients of the urban and residential land uses were much greater than the coefficients relating agricultural land uses to violations. The principal components method yielded two factors that were characterized as (1) urban and residential and (2) deciduous forest and pasture. Langley determined through a multivariate regression of the independent variables that agricultural land uses are not helpful in predicting pathogen contamination. According to Langley's result, urban and residential land uses are to blame for pathogen contaminations, not agriculture. In general, Langley found that urban and residential land uses reveal more information about a larger number of pollution types and are a better predicting tool for contamination. Tong and Chen (2002) attempted to use statistical, spatial, and hydrologic modeling (BASINS) analysis to examine relationships to inspect hydrologic effects of land use at regional and local scale. The use of statistical and spatial analyses will be included in this study, but hydrologic modeling will not be utilized. The non-parametric statistical tests (due to data that was not distributed normally) and analysis of variance were performed and "the results revealed that there were significant relationships between land use and nitrogen, phosphorus, and fecal coliform," which are three of the sampling parameters that will be analyzed in this study (Tong & Chen, 2002). Tong and Chen's results revealed that there is a strong correlation between phosphorus and agricultural and urban land uses. Their findings suggest that "more conservation efforts need to be targeted at reducing phosphorus levels and not just reducing nitrogen levels" (Tong & Chen, 2002). They also found that total nitrogen, total phosphorus, and fecal coliform all had positive relationships with commercial, residential, and agricultural land uses, and as expected they had a negative relationship with forested land uses. The analysis of variance revealed that mean values of total nitrogen and total phosphorus are much higher in agricultural watersheds than in urban watersheds, which are much greater than in forested watersheds (Tong & Chen, 2002). GIS was utilized by Tong and Chen to indicate watersheds throughout Ohio that had high level of pollutants and high percentages of agricultural and urban land uses.

Numerous studies throughout the literature have utilized statistical analysis as the primary analysis tool to examine relationships surrounding water quality. Moog and Whiting (2002) utilized statistical analysis to aid in identifying the climatic, hydrological, and agricultural variables that best explained the variations that were present in nitrate, phosphorus, and total suspended solids over a period of time in two agricultural watersheds in northern Ohio (Maumee and Sandusky). Moog and Whiting's study is similar to this study in that similar categories of variables are included, as well as, the inclusion of other variables, to see what correlations can be found between loadings and different explanatory variables. There will be differences in some variables, both climatic and agricultural, that are included because of the differences in geographic location, causing some variables to be added and some to be left out. Also, in trying to complete a model that looks outside of the variables that have been historically included in previous models in the literature to explain nutrient loads, the addition of several other categories of variables (socioeconomic and land use) have been included. The inclusion of other variables will allow further conclusions to be made on what factors (outside of agriculture) are significant in water quality. The statistical model that was created and included in the research by Moog and Whiting included four types of data: stream loads, climate, stream flow, and agriculture, which include several similar variable categories. Including other data categories allows insight into if other variables cause load variations outside of agriculture. They conducted their model by setting up a stepwise linear regression model, where stream loadings were regressed against the selected explanatory variables (Moog & Whiting, 2002). An important aspect of their study to point out is that they modeled streamflow rate separately from other explanatory variables because of the how streamflow rate dominated the variation shown from preliminary investigations (Moog & Whiting, 2002). It is also important to mention that the relationship between load and discharge differed by nutrient, where SRP, TSS, TP were reasonably linear with constant variance while "NO_{2+3} exhibited substantial downward curvature." The final selection of variables to be included in the model included: precipitation, streamflow rate, temperature, snow depth, snowfall, Conservation Reserve Program (CRP) enrollment, nitrogen fertilizer deliveries, phosphorus fertilizer deliveries, nitrogen from manure, phosphorus from manure, conservation tillage area, rainfall, snow cover fraction, snow melt, freezing day fraction, rain-to-snow days, and rain-on-bare ground days. Due to differences in location many climatic variables will not be included in this study, because they are not relevant given the Lower Kentucky Watershed's location.

Univariate regression with forward selection was selected for use because it offered several advantages over a multiple regression model, including alleviating colinearity problems from closely related explanatory variables and it allowed for relationships between secondary terms to be revealed that would have been obscured if earlier correlated terms were not first removed (Moog & Whiting, 2002). The model was split into two modeling time periods due to a lack of agricultural data for a small time period and each time period was modeled for each month or "season" (5 "seasons" included). After further examination the model was solved by using groupings of months that were based on consistency among months and consistency among explanatory variables based on individual month regressions. The model was run once for a group of unique combinations characterized by season in each watershed.

Moog and Whiting concluded that numerous statistical relationships were found, dominated by the significance of streamflow rate. Nitrate plus nitrite was negatively correlated by previous year precipitation and rainfall, which suggests that “wet weather from preceding years decreases nitrate plus nitrite in the soil and loads in successive years” (Moog & Whiting, 2002). Very few agricultural variables showed significance for nitrate plus nitrite (partly due to the lack of agriculture data), but nitrogen fertilizer deliveries did show some significance in the Maumee basin having a negative correlation between nitrate plus nitrite and current-year nitrogen deliveries. Significance from agricultural variables were present in both watersheds from SRP, where CRP enrollment, conservation tillage, and phosphorus from manure were well-correlated to load-discharge residuals in time frames throughout the year in both the Maumee and Sandusky basins. TSS loads were associated with wet conditions (similar to nitrate plus nitrite), where wet conditions in recent preceding three month spans had explanatory power. Differences were present between the two loads of streamflow and maximum rainfall, implying that “extreme wet conditions in winter and spring affected TSS loads more than dissolved-constituent transport.” (Moog & Whiting, 2002) TP showed very little correlations with the explanatory variables included in the model except for a negative correlation to conservation tillage and streamflow in the three previous months and a positive correlation to mean temperature.

Nitrate plus nitrite was mainly explained by climatic variables, SRP was mainly explained by agricultural variables, TSS was similar to nitrate plus nitrite (except for negative correlation to CRP enrollment and conservation tillage), and TP generally did not meet the criteria (p-value) needed to be included in the model (moog & Whiting). The different loads were affected by the different explanatory variables that were included in their model, showing that statistical analysis is a good tool to use to explain load variations in agricultural watersheds.

Moog and Whiting (2002a) performed a similar study on the Maumee and Sandusky basins in northern Ohio looking for climatic and agricultural contributions to changing loads, in an effort to complete two objectives. “First, they wanted to solve and estimate trends and changes in nutrient loads over their selected time frame. Second, they wanted to solve for predicted changes in loads from explanatory variables and compare them with the actual changes in loads." (Moog & Whiting, 2002a) Several trends have been observed throughout the region over the study's time span, such as reductions in fertilizer use and sediment loss. By analyzing all the trends, as well as, other climatic and streamflow variables allows for comparison of significant factors affecting nutrient exports in agricultural watersheds.

To estimate and detect the trends in climatic and streamflow data Moog & Whiting employed two techniques: linear regression on time and the nonparametric Mann-Kendall test. To detect and estimate trends of explanatory variables on time Moog and Whiting used linear regression to emphasize the changes that actually occurred. The explanatory variables that were used were the same as the variables that were included in their previous study above, where the independent variables could be grouped into three main groups: climatic, streamflow, and agricultural. Results from their previous study were used to compare changes in loads to changes in variables. An expression for time derivatives from the previous statistical model was developed, where the time derivatives were the slopes computed using the linear least-square regression (Moog & Whiting, 2002a). As performed in the preceding article, months were grouped to develop a system where five "seasons" were utilized to gain more knowledge on the dynamics of climatic, agriculture, and loads.

The results indicate that TSS and TP did not meet the conditions needed to be included in the model (which supports the hypothesis that TP and TSS are closely related due to TP absorption into soil particles (Moog & Whiting)), while nitrate plus nitrite and SRP were the only two loads that were included in the model. Nitrate plus nitrite had increases in the adjusted loads, which were larger than expected, but were in the right direction. Load contributions continued to be lower than the total load change suggesting that the model that was developed by Moog and Whiting "is not a comprehensive explanation of variations in nitrate plus nitrite loads." The inclusion of new explanatory variables could make the model developed in this study able to explain more of the variation from nitrogen. SRP had a "significant downward trend" (Moog & Whiting, 2002a), dominated by decreases in phosphorus fertilizer deliveries and phosphorus from manure. SRP decreases are shown from changes in agricultural practices, namely conservation tillage and CRP enrollment.

Geographical Information Systems and other analytical tools have been developed that are able to aid in soil and water conservation, allowing for the reduction in environmental degradation and water pollution from agriculture. The use of these analytical tools are very evident in the literature as a way to help with “precision conservation”, defined by Berry (2003) as “a set of spatial technologies and procedures linked to mapped variables directed to implement conservation management practices that take into account spatial and temporal variability across natural and agricultural systems.”

Numerous studies are present in the literature where “precision conservation” has been applied to investigate different scenarios that occur in agriculture. Goddard (2005) reported on the need to integrate landscape position to implement conservation practices. The application of computer models to assess the nutrient or erosion scenarios across spatial variability will be an approach to “precision conservation.” (Goddard, 2005) Other studies by Schumacher et al. (2005) reported that we can use ^{137}Cs to identify spatial patterns of erosion for use in “precision conservation”. In a similar study, Kitchen (2005) and Lerch (2005) determined that we can use “precision conservation” at a field level to improve soil and water conservation practices. Similar studies have been performed on irrigation water (Sadler et al. 2005), spatial variability of residual soil NO_3-N and leaching (Delgado 1999, 2001), and the effectiveness of best management practices (Renschler & Lee 2005).

CHAPTER FIVE: Theoretical Model

In order to develop a model to try to achieve the objective of this study, it is essential to know what variables need to be included in the selected model. There are two types of variables that are included in the theoretical model, independent or explanatory and dependent. The relationship between the two variables is that the independent or explanatory variables are the variables that are manipulated by the researcher to try to determine their significance on the observed or dependent variable. In this particular study, the dependent variable is described as the individual sampling site reading for total suspended solids (TSS), fecal coliform, nitrate-nitrogen, and total recoverable phosphorus (TRP) and the independent variables could include any set of variables. In this case, a set of independent variables that could be affecting the transport of pollutants into waterways will be included in the model, allowing for the significance of independent variables to be tested.

Past literature suggests that the major sets of variables that are tested should include the following: land use (Langley (2004), Tong & Chen (2002)), agriculture (Moog & Whiting 2002), weather and climate (Moog & Whiting 2002), topographic features (Calhoun, Baker, & Slater 2002), soil properties (Calhoun, Baker, & Slater 2002), and hydrology (Moog & Whiting 2002). Table 5.1 gives specific examples of variables that were included in previous studies that have concentrated on factors affecting water quality.

Table 5.1 Examples of Independent Variables Included in Previous Studies

Agriculture	Climatic	Topo.	Soil	Hydrology	Land Use
Fertilizer	Rainfall	Physiography	Drainage	Streamflow	Residential
Manure	Snowfall	Parent	Sand %	Dis. Oxygen	Agricultural
Crop Est.	Temperature	Material		pH	Commercial
Livestock		Slope %		Water Temp	Forest

There are two different methods that we can use to try to analyze the independent variable data that we collect to see its affect on the dependent variable, both of which deal with the classification of the dependent variable as being either continuous or discrete. A continuous dependent variable can take on any given amount of values, while a discrete dependent variable can only take on a certain amount of values. The classification between the two types of dependent variables is important in identifying what model is the most appropriate to use when setting up a regression. This study will classify the dependent variable as a continuous variable that can take on any value. When setting up a model to use for a continuous dependent variable the ordinary least squares (OLS) method can be used to set up the regression model that minimizes the sum of squared errors, where in this case the dependent variable would be continuous and could take on any value as the individual site's sampling data. Past studies in the literature have used continuous dependent variables to look at water quality (Moog & Whiting 2002), where they looked at the continuous variable as being the sampling data. The sampling data (load) was then regressed against the collected information that will be used as the explanatory variables in the model to see which variables significantly affected the sampling data.

Loads = f (agriculture, climatic, topographic, soil, hydrology, landuse)

The variables that have been included in past literature offer a great starting point from which to begin to investigate sediment, nutrient, and pathogen loads on water quality, but other variables and variable categories could have an affect on the amount of pollutants present in a waterway. In an effort to take a look at some other variable categories that could have an affect on pollutant levels at specific sites, the inclusion of the variable category of socioeconomic variables will provide information on what affects they have on the amount of pollutants present at the individual sampling sites. With the addition of this new variable category, the functional form would look like the following:

Loads = f (agriculture, climatic, topographic, soil, hydrology
landuse, socioeconomic)

The variable categories that are above and included in the theoretical model include several different kinds of variables. Variables that fall into the agriculture variable category include data such as livestock numbers present within an area, crop estimates, and other relevant data. This variable category will allow for multiple criteria and data from agriculture to be tested in the model. The climatic variable category is a category that includes precipitation data, temperature data, and other relevant data. The topographic variable category that is included in the theoretical model includes data that deals with the slope of the area surrounding the waterways that are examined, as well as, the physiography of the land. The soil variable category is made up of variables that deal with the composition of the soil, in an effort to see if the composition of the soil has any significant affect on the quality of the water criteria sampled at each individual sampling site. The hydrology variable category includes data that deals with the physical and chemical conditions of the; examples of the hydrology variable category include pH, water temperature, flow conditions, etc… The land use variable category refers to the way that the land parcels are being utilized, with examples being residential, agriculture, commercial, etc… The socioeconomic variable category is a category that allows for multiple forms of data from an economic and demographic standpoint to be analyzed in the model. Examples from the socioeconomic variable category include income, poverty status, education, etc…

It is not always possible to include all the variables that were utilized in previous studies into this study, due to differences in availability of data and differences in geographic location. The creation of the final model will be discussed in further detail in the next chapter.

CHAPTER SIX: Empirical Model

The theoretical model was constructed using variables that were included in models from previous studies in the literature. Each of the loads (nitrate-nitrogen, total recoverable phosphorus, total suspended solids, and fecal coliform) will be regressed against the independent variables that have been suggested from previous literature. However, not all variables from previous studies are relevant or have available data to be included in this study. The construction of the empirical model allows for all variables that are relevant and available to the study to be included into a final model.

The theoretical model that was discussed earlier was as follows:

Loads = f (agriculture, climatic, topographic, soil, hydrology,
landuse, socioeconomic).

There are several variable categories that are included in the theoretical model that will not be included in the empirical model. The following variable categories will be left out of the final model: topographic, soil properties, and hydrology. Topographic variables were left out of the model due to the fact that there was no relevant data that was associated with the topographic features to be included in the regression.

Soil properties are considered an important aspect when it comes to looking for relationships that exist between land use and water quality, but it was not included in the model based on findings from Calhoun, Baker, and Slater (2002). Calhoun, Baker, and Slater (2002) conducted a study to examine soil properties and watershed size on water quality in northern Ohio. The Maumee (1,620, 516 hectares) and Sandusky (320,248 hectares) basins were used in this study, where they tested four different sizes of watershed scales to see if any water quality and soil interrelationships exist. The four watershed scales utilized in their research were large, intermediate, small, and plot size. The preferred method that was used in the study was to start with the data form the large watershed size (Maumee and Sandusky) and compare them on a unit area basis with the smaller watersheds.

The study area used in this research is a large watershed which contains 828,590.65 hectares of land and falls into the large watershed category. The results from the Calhoun, Baker, and Slater (2002) study on large watersheds "found that pollutant levels and soil conditions (soil properties and slope) are vague and contradictory at a large watershed size," (Calhoun, Baker, and Slater, 2002) allowing for soil conditions to be excluded from the model used in this paper. After concluding their study, Calhoun, Baker, and Slater also concluded that the assumption of slope being a significant factor of pollutants should be reexamined.

Hydrology data was collected at particular sites across the region at three different times periods during the year (May, July, and September), but they were left out of the final model because of a lack of data that was available for each sampling site. The large amounts of missing data would cause problems in running the regression due to missing values in the data set, and therefore will not be included in the model. Also, it is important to note that the climatic variable category has been altered because the only variable that will be included in the model from this category will be three month (May, July, and September) total rainfall averages, which is when the sampling data was collected. The data that was available for rainfall was a county average that was generated by GIS Model output from a central point in the respective county (Ag Weather Center, 2006).

After reviewing past literature and collecting data from variable categories that are relevant to the study, the final model will be set up as follows:

Loads = f (agriculture, rainfall, landuse, socioeconomic).

The regression will be performed as an Ordinary Least Squares (OLS) standard regression where the stream loads or sampling site data is the dependent variable and the independent variables will be the data collected for the agriculture, rainfall, landuse, and socioeconomic variable categories. Below is a list of the final variables that will be included as the independent variables, along with a brief description of each independent variable.

Table 6.1 Independent Variable List to be Included in the Models

Variable Name	Variable Description
resident	Residential + Other Urban Built-Up + Transitional Areas (% of buffer)
forest	Evergreen + Deciduous + Mixed Forest Land (% of buffer)
crop_past	Cropland & Pastureland (% of buffer)
rain	May + July + September Averaged Rainfall Totals
beef	Beef Cattle Present in 3 km Buffer Zone(head)
tobacco	Burley Tobacco Present in 3 km Buffer Zone (in lbs.)
dcows	Dairy Cows Present in 3 km Buffer Zone (head)

In the initial steps of analyzing the data it became evident that two main problems were going to be encountered with both the dependent and independent variables. The first problem that was discovered was shown clearly by looking at a spatial overview of the sampling data within the region. The original maps that were created in ArcMap showed instances where there were extreme amounts of variation in the sampling data (loads) present within the region due to sampling variability. The extreme variation that was present within the watershed would most likely be explained by looking at the sampling site location. The sample itself is a function of where the sampling site is located (relative to pollution sources) and the characteristics of the environment at the time that the sample is taken (e.g. rain amount, temperature, etc..). By correcting for the sample variability that was present throughout the clusters, the assumption of these location and characteristics of the environment affecting the sample is eliminated. The location relative to pollution sources will play major role as into how high a particular reading will be for a particular pollutant (e.g. the closer the sample is taken to a pollution source, the higher that particular sample will be.)

Also, the sampling sites were not located evenly throughout the watershed; instead they were grouped into clusters throughout different sections of the watershed. These clusters caused in many instances different measurements to be present within a very small distance. The differences in measurements were so severe that there was no possibility that a regression model would be able to explain the variation that was present (Fig 6.1, Fig 6.3, Fig 6.5, & Fig 6.7).

Figure 6.1 Nitrate-Nitrogen Data Without Sampling Variability Corrected

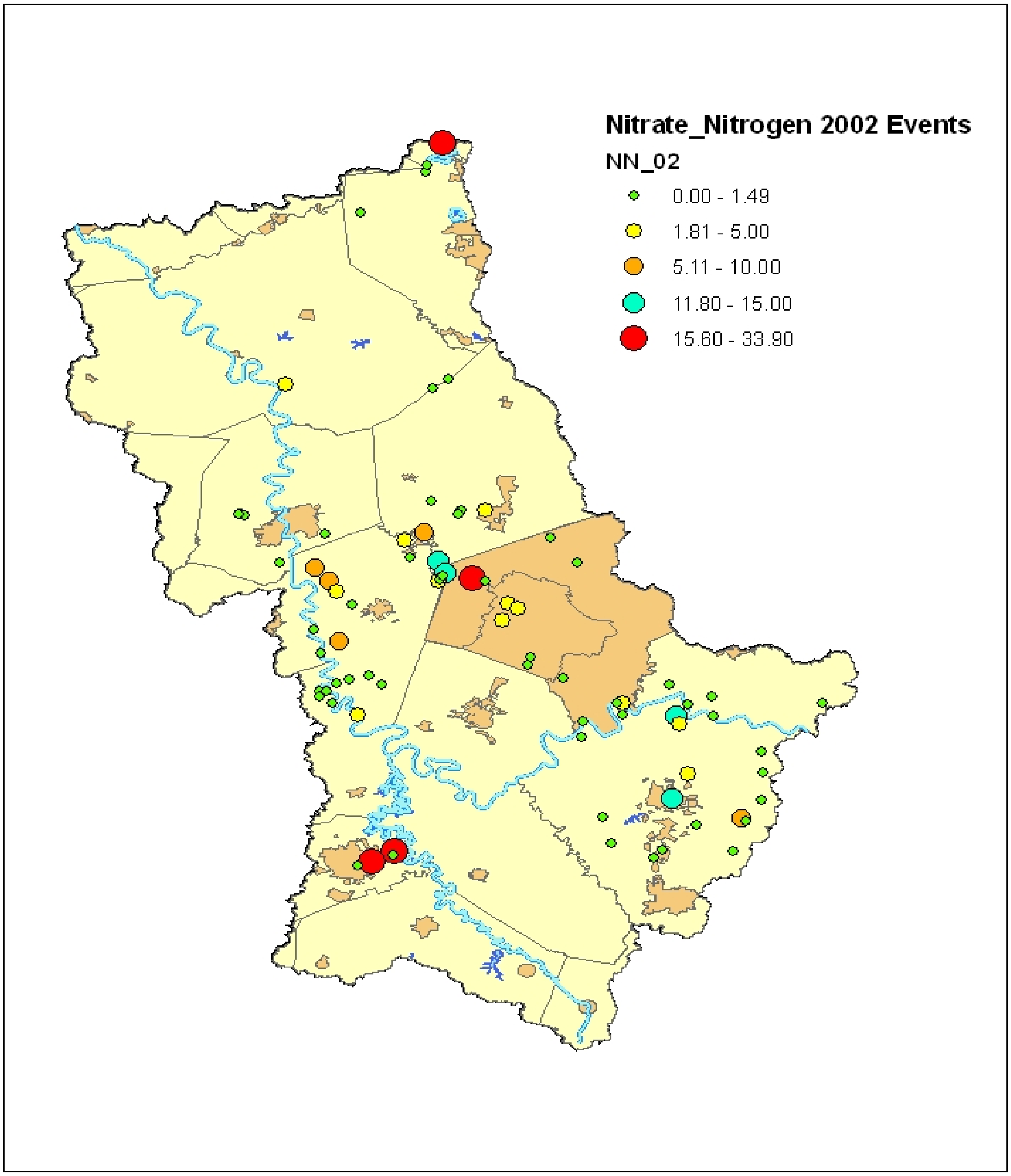


Figure 6.2 Nitrate-Nitrogen Data After Corrected for Sampling Variability

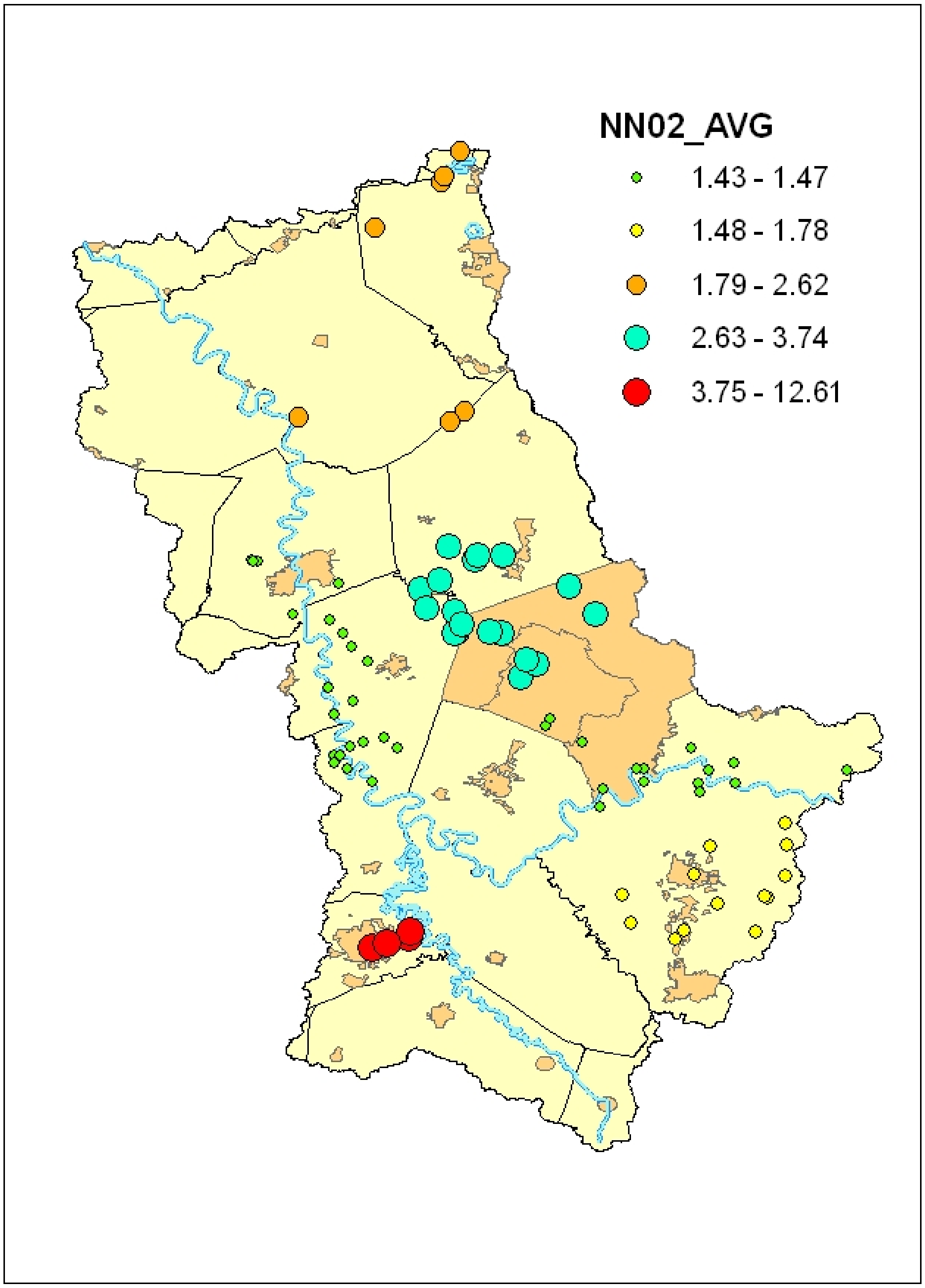


Figure 6.3 Total Recoverable Phosphorus Data Without Sampling Variability Corrected

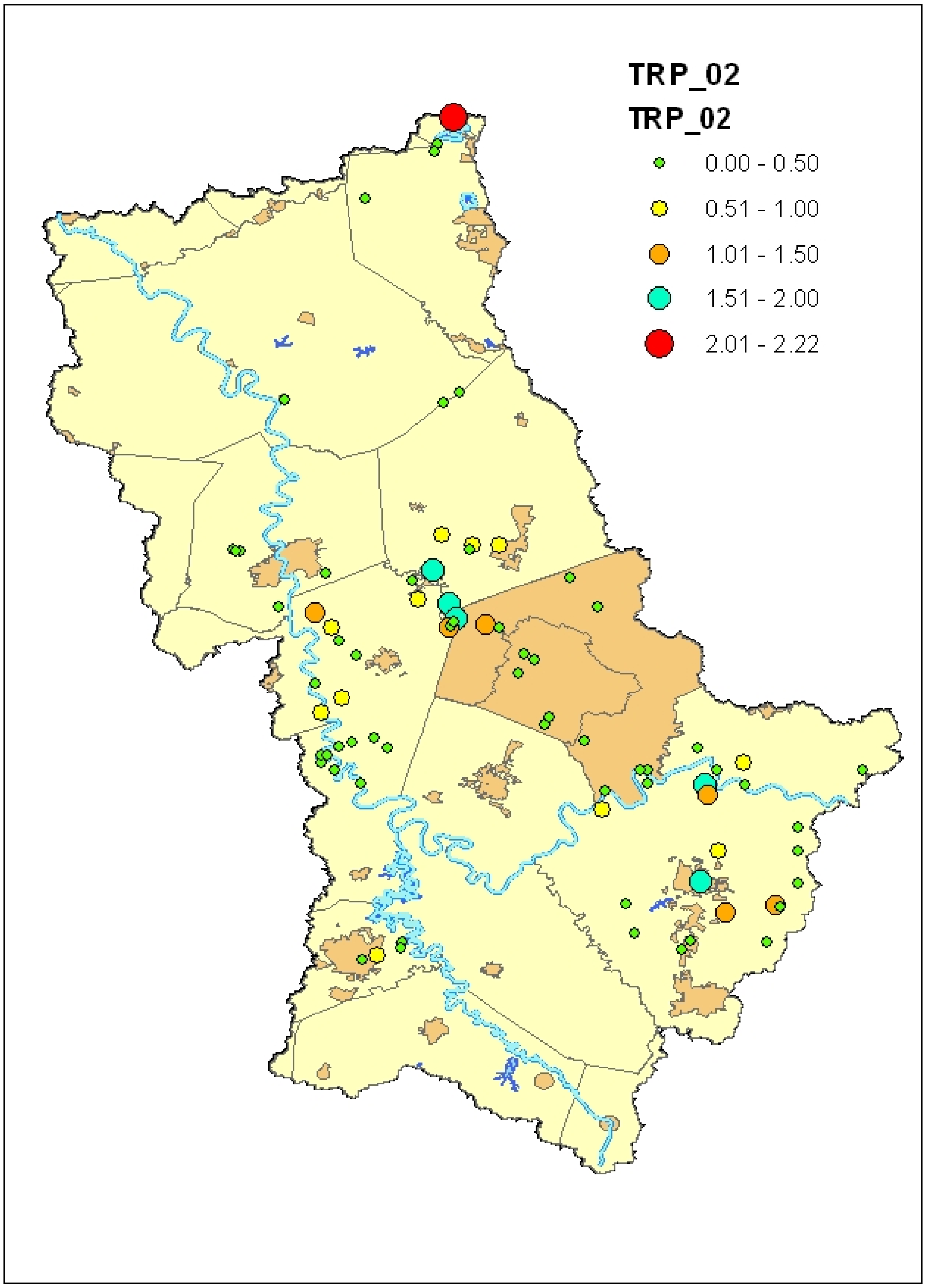


Figure 6.4 Total Recoverable Phosphorus Data After Corrected for Sampling Variability

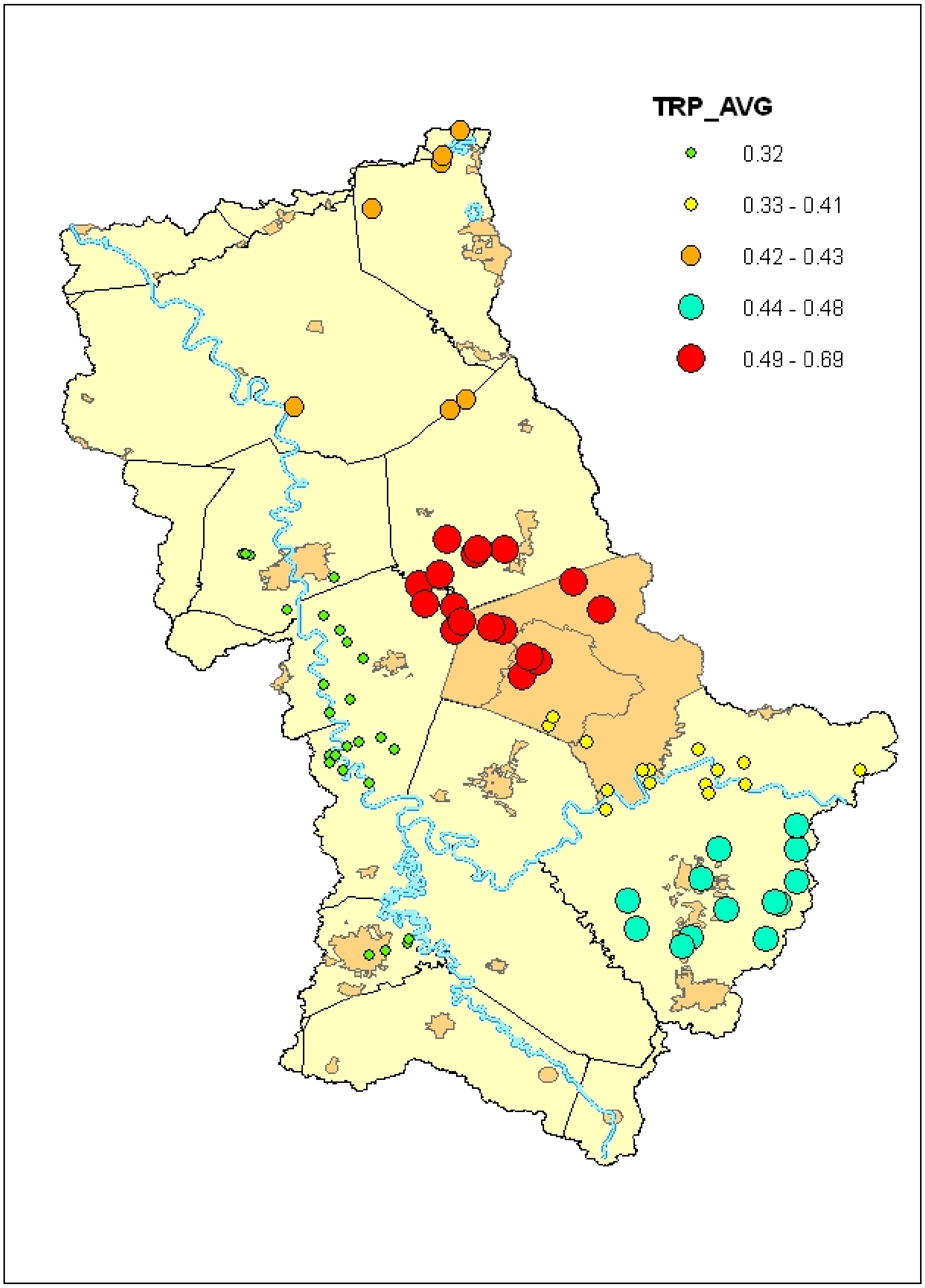


Figure 6.5 Fecal Coliform Sampling Without Sampling Variability Corrected

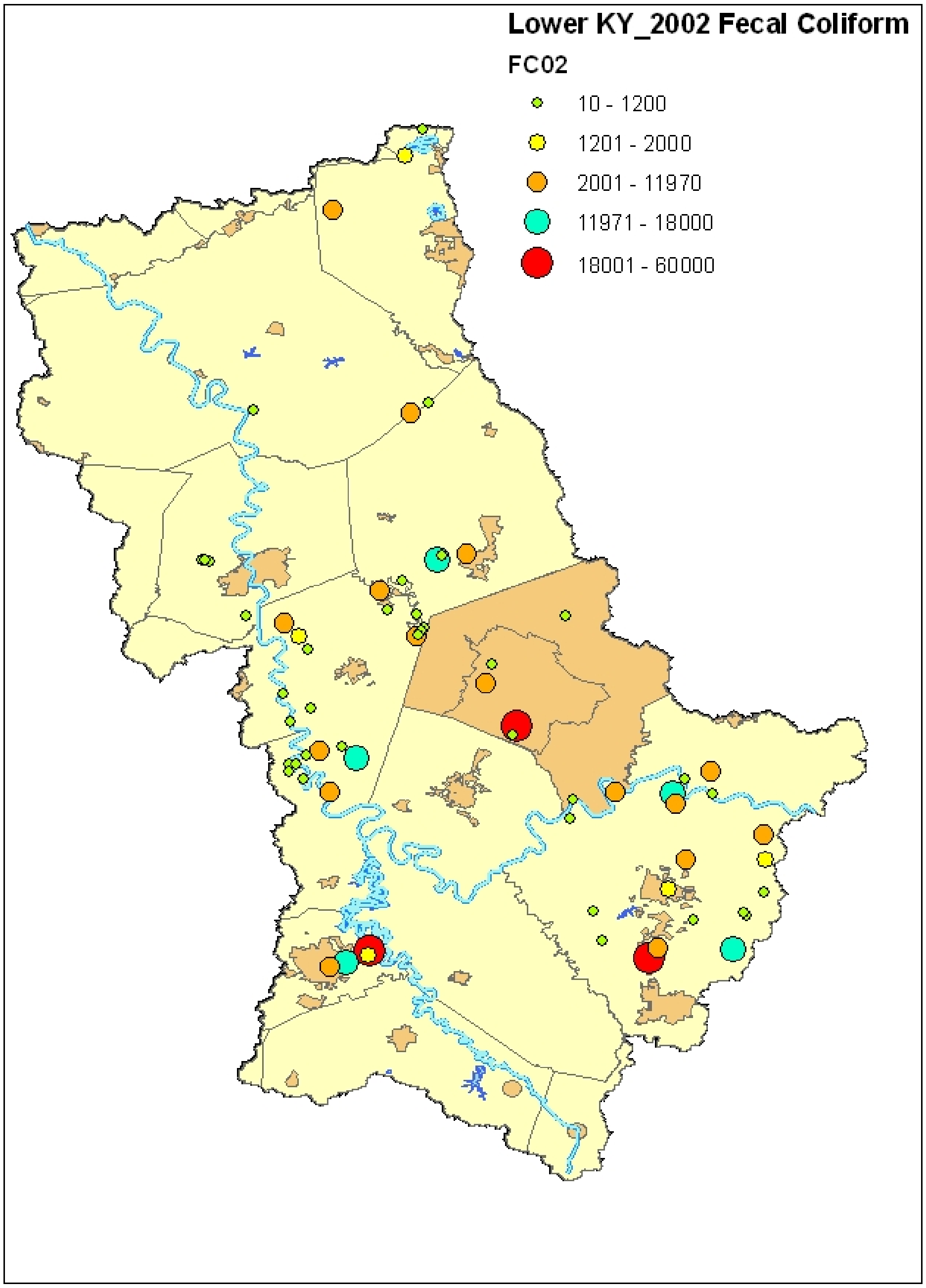


Figure 6.6 Fecal Coliform Data After Corrected for Sampling Variability

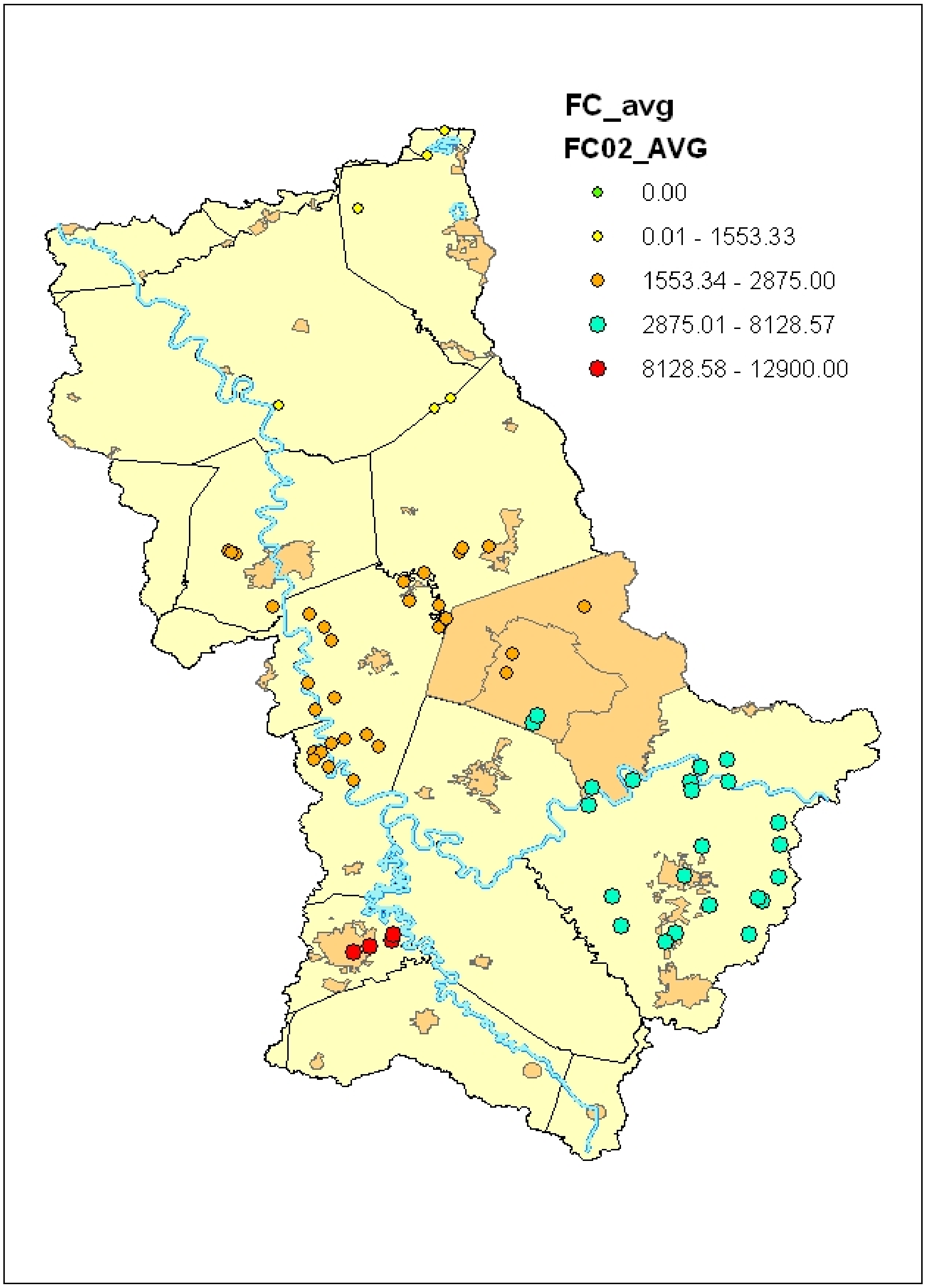


Figure 6.7 Total Suspended Solids Data Without Sampling Variability Corrected

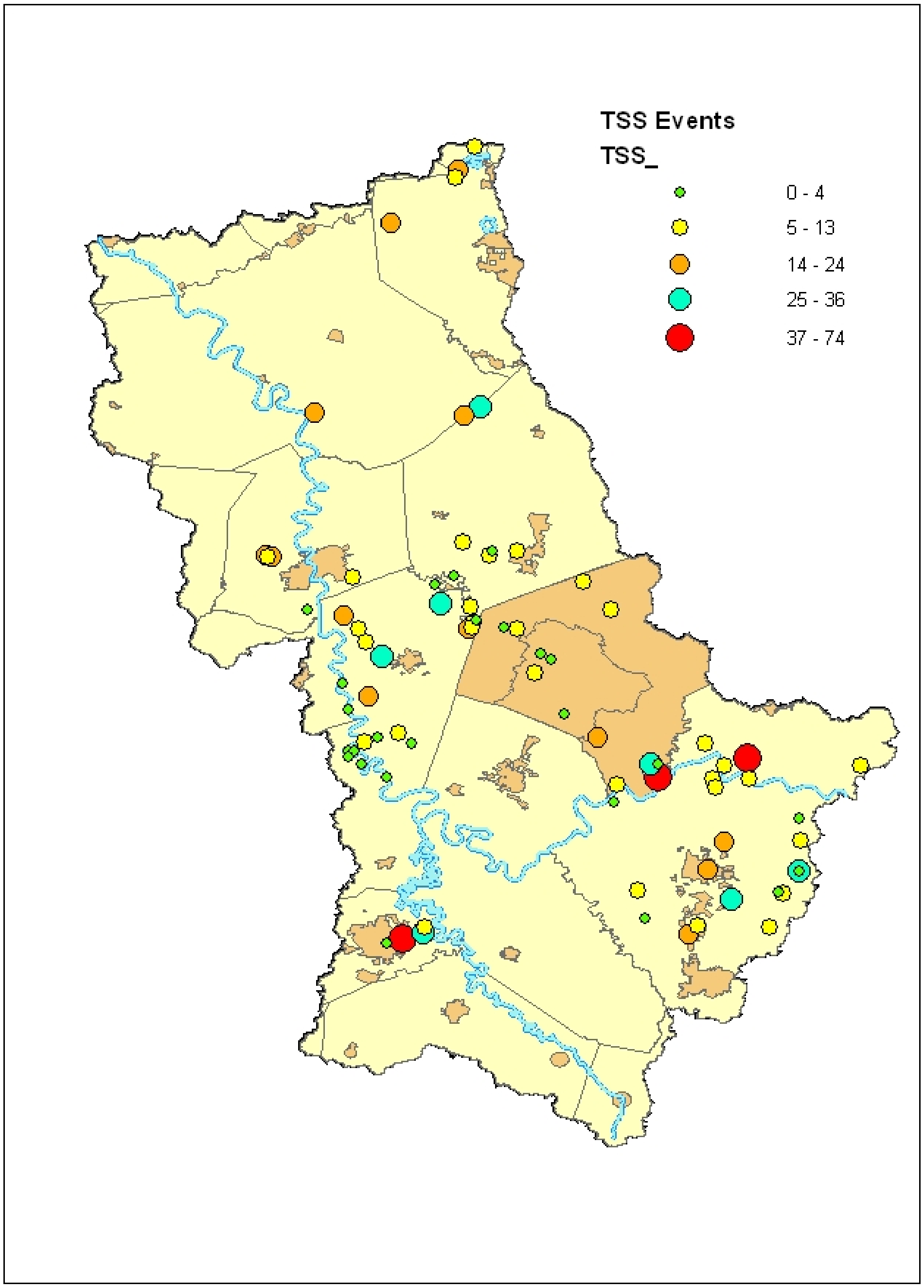
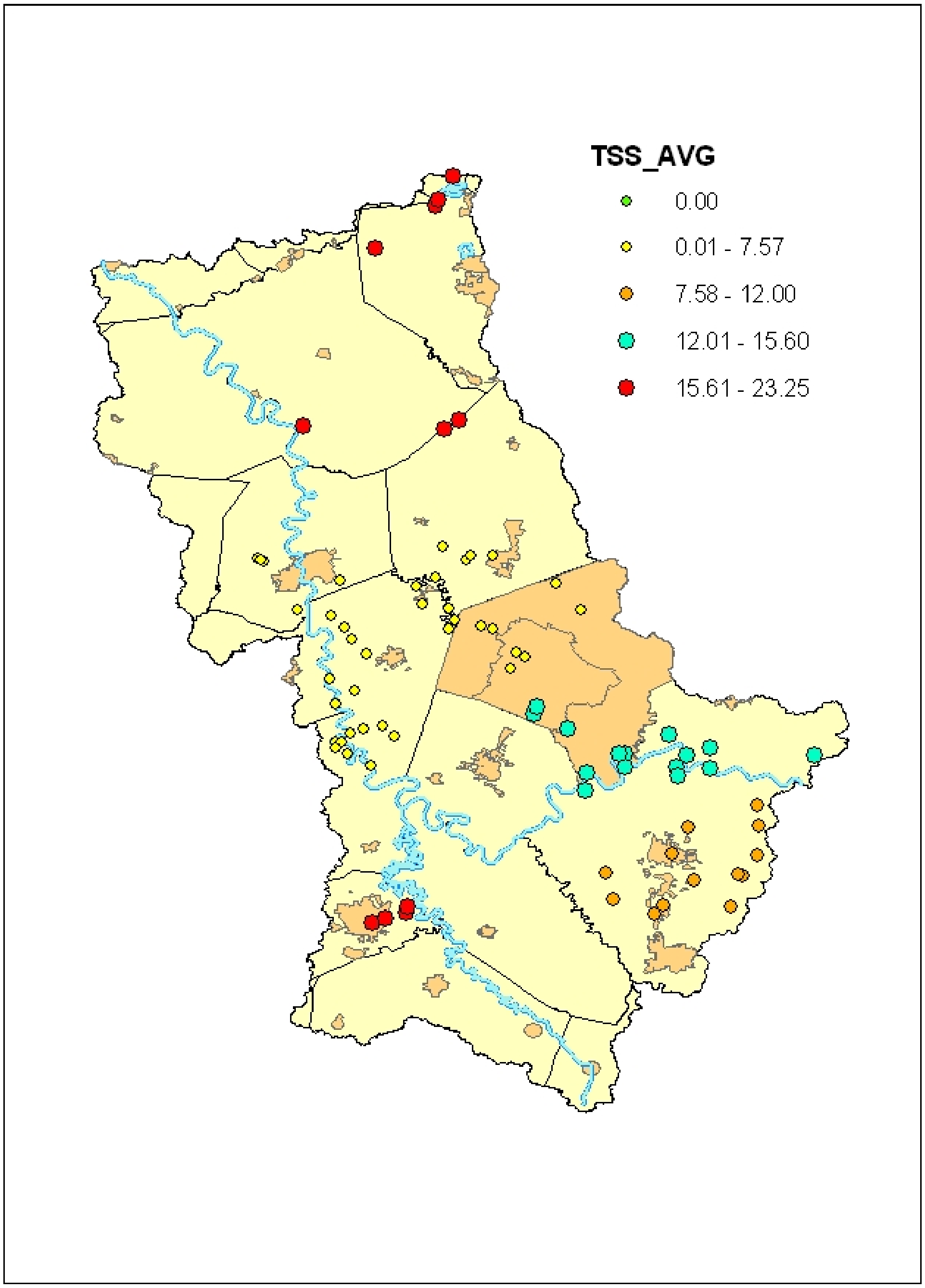


Figure 6.8 Total Suspended Solids Data After Corrected for Sampling Variability



In order to get the dependent variables (sampling loads) in a manner that could be utilized in the regression model and keep the same number of observations, the watershed was broken down into different "clusters". An expected value was then calculated from the sampling sites' data from each "cluster" to help correct for the sampling variability that lead to extreme variation in the sampling data between sampling sites in close proximity (Fig 6.2, Fig 6.4, Fig 6.6, & Fig 6.8). This technique allowed for the sampling variability that was present in the data to be corrected and legitimate results to be calculated, as opposed to results that were produced from initial regression models that did not correct for the sampling error that was present in the dependent variables.

An important statement to make at this time is that all the original socioeconomic variables will be left out of the analysis due to their high level of correlation to resident land use. A correlation matrix (Table 6.3) was constructed that revealed the high level of correlation that was found between resident land use and all the socioeconomic variables. This matrix reveals that a high level of multicollinearity is present between these independent variables, allowing for the explanatory power of the socioeconomic variables to be shown through the resident land use variable.

Table 6.2 Socioeconomic Variable List (Present in 3 km Buffer Zone)

Variable Name	Variable Description
population	Total population
education	Persons 25+ years old who have an associate degree but no bachelors degree
unemploy	Persons 16+ years old in the civilian labor force and unemployed
poverty	Proportion of total persons below the poverty level last year
income	Median family income last year

**All data was created using the Neighborhood Change Database (NCDB) 1970-2000 Tract Data* CD
***All data from year 2000*

Table 6.3 Correlation Matrix for Resident Land Use and Socio Economic Variables

	Resident_	population	education	unemploy	poverty	income
Resident_ Resident	1.00000	0.91827 <.0001	0.84480 <.0001	0.80876 <.0001	0.79178 <.0001	0.91418 <.0001
population population	0.91827 <.0001	1.00000	0.96551 <.0001	0.88365 <.0001	0.80693 <.0001	0.99095 <.0001
education education	0.84480 <.0001	0.96551 <.0001	1.00000	0.75612 <.0001	0.63262 <.0001	0.94582 <.0001
unemploy unemploy	0.80876 <.0001	0.88365 <.0001	0.75612 <.0001	1.00000	0.96521 <.0001	0.89710 <.0001
poverty poverty	0.79178 <.0001	0.80693 <.0001	0.63262 <.0001	0.96521 <.0001	1.00000	0.83396 <.0001
income income	0.91418 <.0001	0.99095 <.0001	0.94582 <.0001	0.89710 <.0001	0.83396 <.0001	1.00000

After steps had been taken to correct problems that were found through preliminarily regression models, data and spatial analysis, and the construction of a correlation matrix, the final regression model for each dependent variable was constructed. Each regression model was then ran in SAS in three different functional forms: complete log transformation (both dependent and independent variables logged), semi-log transformation (dependent variable logged only), and the original basic functional form. By running multiple functional forms on each dependent variable, the model that explained the most variation would then be able to be selected and significant variables identified. The following functions (Table 6.4) were then analyzed in SAS to check for statistical significance on the independent variables.

Table 6.4 Basic Functional Form, Semi-Log, and Log Regression Models

Nitrate-Nitrogen

1) nn_avg = f (resident, forest, crop_past, rain, beef, dcows)
2) log nn_avg = f (resident, forest, crop_past, rain, beef, dcows)
3) log nn_avg = f (log resident, log forest, log crop_past, log rain, log beef, log dcows)

Total Suspended Solids

4) tss_avg = f (forest, crop_past, rain, beef, tobacco, dcows)
5) log tss_avg = f (forest, crop_past, rain, beef, tobacco, dcows)
6) log tss_avg = f (log forest, log crop_past, log rain, log beef, log tobacco, log dcows)

Fecal Coliform

7) fc_avg = f (resident, forest, crop_past, rain, beef, tobacco)
8) log fc_avg = f (resident, forest, crop_past, rain, beef, tobacco)
9) log fc_avg = f (log resident, log forest, log crop_past, log rain, log beef, log tobacco)

Total Recoverable Phosphorus

10) trp_avg = f (resident, forest, crop_past, rain, beef, tobacco, dcows)
11) log trp_avg = f (resident, forest, crop_past, rain, beef, tobacco, dcows)
12) log trp_avg = f (log resident, log forest, log crop_past, log rain, log beef, log tobacco, log dcows)

Table 6.5 Expected Sign of Parameter Estimates

	nn_avg	tss_avg	fc_avg
resident	positive	positive	positive
forest	negative	negative	negative
crop_past	positive	positive	positive
rain	positive	positive	positive
beef	positive	positive	positive
tobacco	positive	positive	negative
dcows	positive	negative	positive

Analytical Tools Procedure

The use of analytical tools was an important aspect of this study, where they were vital in creating buffer zones and calculating areas for each buffer zone (ArcMap) and manipulating the data and running regression models to check for variable significance (SAS). The use of these tools involved many steps to get the data manipulated down to the correct study area, and following is a list of the steps that were needed for each of the analytical tools.

Steps Used for Converting Data in ArcMap:

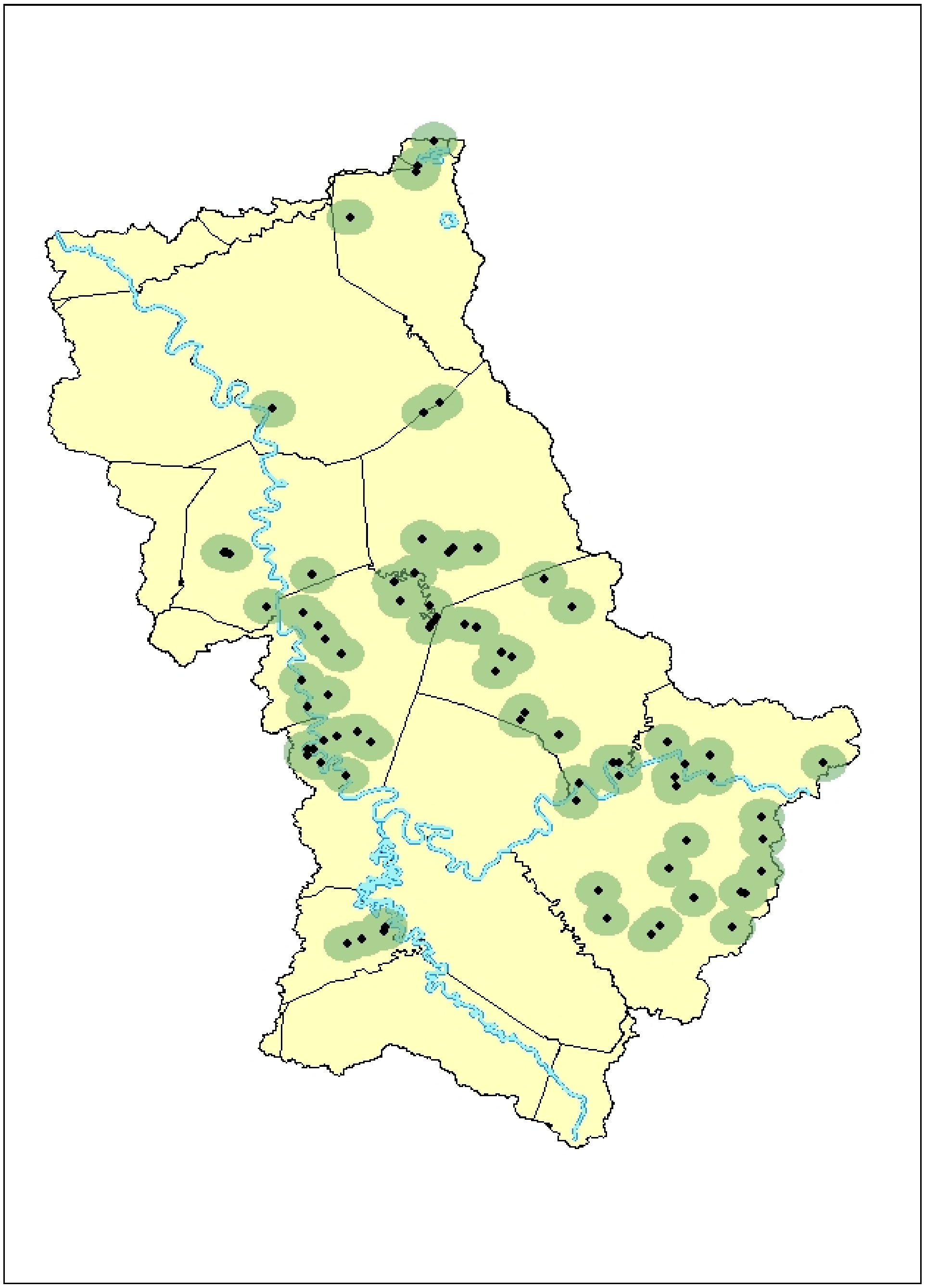
1. All layers were brought into GIS, including National Land Cover (NLC) 1992 data, sampling sites data including the parameter data of TRP, NN, FC, & TSS (which were created by Add X&Y data function in ArcMap), agricultural data (data was available from the University of Kentucky's College of Agriculture and editing was done to the attribute table to include data that collected for this study), census tracts information created by using the *Neighborhood Change Database (NCDB) 1970-2000 Tract Data* CD that allowed for selected features to be mapped in ArcMap.

2. All layers were then clipped down to the Lower Kentucky Watershed area by using the "clip tool" located in Analysis Tools located in the ArcToolbox.

3. Next, *Buffer Wizard* was used to create 3 km buffers that would be produced around each of the sampling sites. By creating these 3 km buffers, the focus will be on the specific surrounding land use(s), agricultural data, and census information to see what affects they have on a site's individual sampling data (Figure 6.9).

4. The toughest part of working with watersheds is that their boundaries are not drawn similarly to other types of boundaries such as counties, census tracts, etc… and they rarely pay any attention to these types of boundaries. ArcMap was the tool that enabled the calculation of the percentage of data categories that were present inside the 3 km buffer, but several steps were needed to be able to calculate these values.

Figure 6.9 3 km Buffer Zones Surrounding Sampling Sites

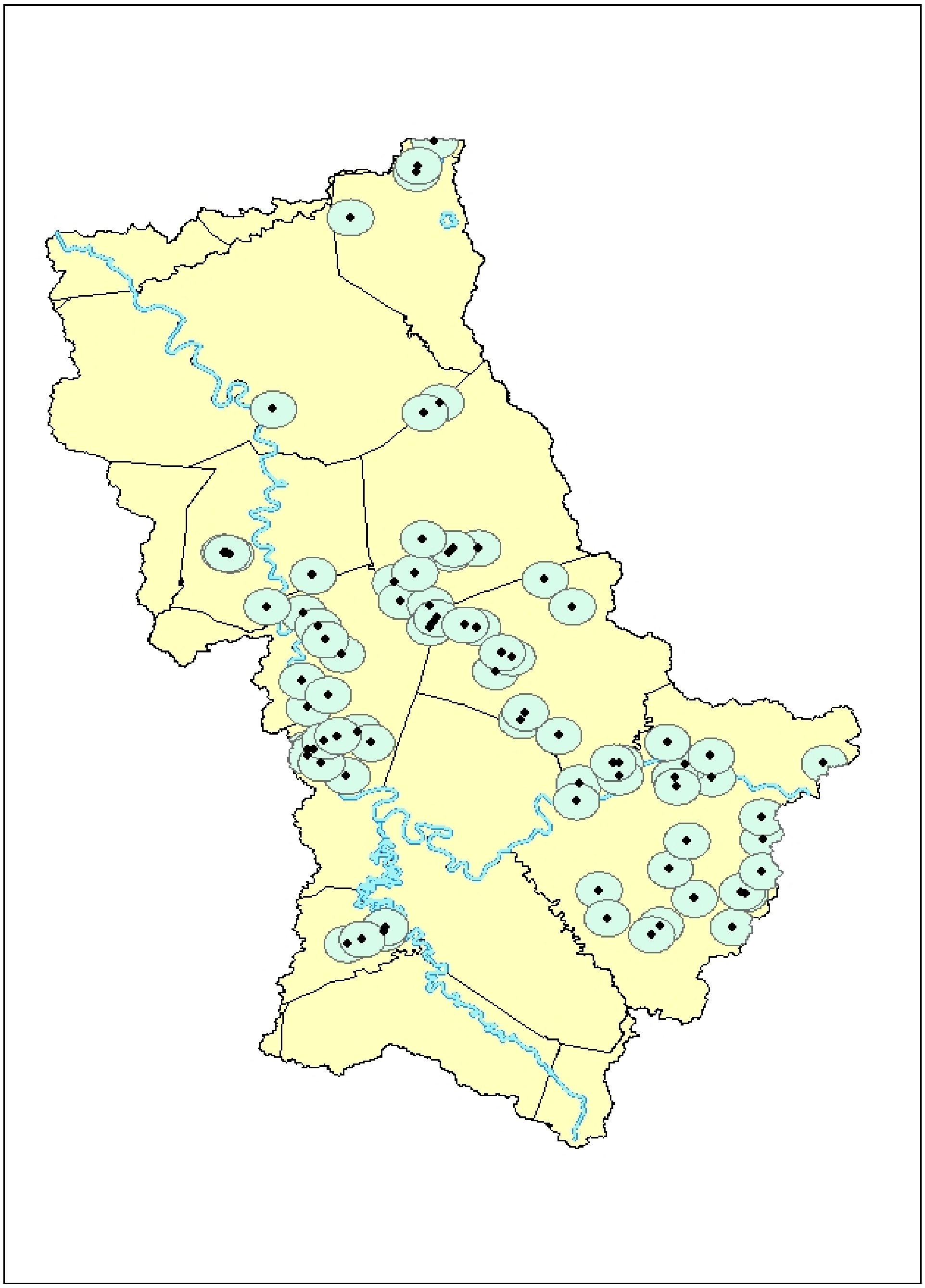


5. After creating the buffers using the *Buffer Wizard* it was time to use the "intersect tool" that was located in the Analysis Tools of the ArcToolbox. The "intersect tool" is similar to the "clip tool", but it transfers the attributes from the clipped layers. By doing this all the data that intersected the buffer was able to be captured in an associated attribute table of each sampling site within the Lower Kentucky Watershed. At this step it is not possible to tell the exact amount present of each of the variable categories in the 3 km buffer zone, but instead it is possible to simply tell what data in the individual buffers are "intersected" by a particular land use, county, or census tract. Later calculations will be done to try to accomplish the previously stated objective.

6. After the use of the "intersect tool" it is possible to tell which of the land use(s), counties, or census tracts variables intersect the buffer zone, then it was time to calculate the area that was present of each land use, county, or census tract in each 3 km buffer. By using the "calculate area tool" located in the Spatial Statistics Tools in the ArcToolbox, the areas were calculated within the individual 3 km buffer zones. The majority of the buffers had an area of .0029 degrees, while some had smaller areas due to the fact that portions of their 3 km buffer zones lied outside the studied watershed. Any area outside the watershed is an area that is not of our concern for this study. By clipping the buffers by the 8-digit watershed, investigation will only occur on the aspects that will contribute to water degradation within the selected 3 km buffer zone surrounding the sampling sites (Figure 6.10). By calculating the area of the intersection between the buffer zones and the layers that had the associated land use(s), county, and census tract data, the percentage of the each particular land use(s), county, or census tract variables that lied within the buffer zone will then be calculated.

7. The calculated data will have one main assumption associated with it. The assumption of equal distribution of data across the collection areas was used to aid in the calculation of the variables by the second analytical tool, SAS.

Figure 6.10 3 km Buffer Zones "Clipped" by Watershed



Steps Used for Converting Data in SAS:

1. To do the data manipulation to the information created in ArcMap, SAS was utilized to make calculations to the areas that were calculated to help determine the values of the variables that lied within the buffers. Having the areas calculated in ArcMap, SAS was used to transpose land use from the area calculations or determine the formula that was needed to calculate the values for the entire buffer (remember the assumption presented before of equal distribution throughout the county or census tract!)

2. The first layer that was converted was the land use area contained within the buffer zone, which did not have to have a formula calculation needed. The reason that no formula calculation was necessary was due to the fact that the land use data was inputted into ArcMap in a form that had each separate land use as its own polygon (unlike the county and census tract data.) By using the calculated area that was calculated in ArcMap by using the "calculate area tool", the percentage of each buffered land use for each individual category of land use(s) that was present in the buffers by the total buffer area. This calculation was done by the following formula:

   (buffered land use area / buffer area).

   This allowed for the total number of individual land use(s) present in the buffers to be calculated, which were then summed up using the ***proc means*** statement in SAS. After summed up, the different land uses were then transposed in SAS by using the ***proc transpose*** statement. By transposing the data, SAS manipulated it so what was row data then became column data, where each individual land use had only one value which represents the total percentage of the land use within each 3 km buffer zone.

3. Agricultural variables and rainfall data were the next set of variables that were used in SAS to be manipulated down to the 3 km buffer zone. The calculations of the agricultural and rainfall variables (as well as the socioeconomic variables) included some extra steps from the land use calculations. The agricultural variables are all at the county

level, which need to be distributed equally across the county, to do this the following formula was calculated:

((agriculture variable / county area) * buffered county area).

The first part of the formula allowed for the equal distribution of each variable across the entire county, which was then multiplied by the particular buffer area to get a final calculation of the total number of each variable within the buffer. The rainfall calculation was very similar to the agricultural variables calculation, except that the rainfall data was already on a county average basis. There is no reason to distribute it equally across the county, which leaves the formula as follows:

(monthly rainfall) * (buffered county area / county area).

Using the ***proc means*** statement on both the agricultural and the precipitation data allowed for the total values to be calculated for each individual variable within the buffers!

4. The calculations for the socioeconomic variables were very similar to the agricultural variables' calculations because the agricultural variables were on a county basis, while the socioeconomic variables were on a census tract basis. The formula that was used to calculate the socioeconomic variables for each buffer is as follows:

((socioeconomic variable / census tract area) * buffered census tract area).

The first part of the formula allowed for the equal distribution of each variable across the entire census tract, which was then multiplied by the particular buffer area to get a final calculation of the total number of each variable within the buffer. Using the ***proc means*** statement on socioeconomic data allowed for summations for each individual variable within the buffer.

5. After all of the land use, agricultural, rainfall, and socioeconomic data was manipulated to produce only the values for each that were contained inside the 3 km buffer zone, it was then necessary to get them into one data set that also included the sampling data. This data was all merged (by the sampling sites' Buffer ID) into one data set in SAS using the ***merge*** statement in SAS, where they were all merged into the new data set by a water buffer identifier that was created earlier.

6. After all of this data manipulation was completed, all of the existing data was now at buffer level, as opposed to county or census tract level, and ready for regressions to be performed in SAS.

CHAPTER SEVEN: Results

The results from the OLS regressions that were analyzed in SAS for nitrate-nitrogen (nn_avg), total suspended solids (tss_avg), and fecal coliform (fc_avg) are summarized in Table 7.1 and Table 7.2. It is important to note that total recoverable phosphorus (TRP) is not discussed in the results section, due to the fact that the results from the regressions for TRP showed that very little of the variation in TRP could be explained by the independent variables that were included in any of the models (all three models had an r^2 of less than $r^2 = 0.13$). Even though the results are not included in this chapter, the results from the TRP regressions are included in the appendix. Included in the results is one dependent variable for each of the three main causes of nonpoint source pollution throughout the United States and Kentucky (nutrients, siltation, and pathogens).

Agriculture has been classified as the leading cause of impairments from nonpoint source pollution in rivers and lakes across the nation (AER-782), where more than one third of the nation’s waterways are impaired. The results that are summarized in Table 7.1 and Table 7.2 conclude that this is at least partly true, but their results also show that other variables outside of agriculture are also significant in the transport of pollutants into waterways in the Lower Kentucky Watershed.

Nitrate-Nitrogen

The results for the three regressions analyzed in SAS on the nitrate-nitrogen revealed that the complete log-transformation of nitrate-nitrogen was able to achieve the highest r^2 (0.6017) of the regression results. The highest r^2 on the log-transformation model form showed that this set of independent variables was able to account for 60.17% of the variation in the dependent variable of the log of nitrate-nitrogen. The complete log-transformation model is given as follows:

lnn_avg = f (lresident, lforest, lcrop_past, lrain, lbeef, ldcows).

The results from the complete log transformation model found that the variables of resident land use (lresident) and dairy cows (ldcows) were statistically significant at the 95% confidence level. Both of the estimates for resident (0.14692) and dairy cows (0.55748) had the correct expected signs on their parameter estimates. The parameter estimate on resident shows that a 1% increase in resident, would lead to a 0.147 increase in the nitrate-nitrogen sample reading that is present at the sampling sites located in the 3 km buffer zones. Due to the high correlation between resident land use and the omitted socio economic variables, it is correct to state that the socioeconomic variables would also be considered significant at either the 90% or 95% level. Buffered dairy cows were also found to be statistically significant at the 95% level, where the estimate showed that an increase in buffered dairy cows by one head would cause an increase in nitrate-nitrogen sample reading of 0.55748 that is present at the sampling sites located in the 3 km buffer zones.

Total Suspended Solids

The results of the three regression models analyzed in SAS for the siltation variable total suspended solids (TSS), revealed that the highest r^2 (0.4743) was achieved by the semi-log transformation model. The highest r^2 on the semi-transformation model form showed that this set of independent variables was able to account for 47.43% of the variation in the dependent variable of the log of total suspended solids. The semi-log transformation model is given as follows:

ltss_avg = f (forest, crop_past, rain, beef, tobacco, dcows).

The results from the semi-log transformation model revealed that rainfall was statistically significant at the 90% level, and beef cattle and burley tobacco were statistically significant at the 95% level. Rainfall (0.566580) and beef cattle (0.000366) both had the correct expected sign on their coefficients. Burley tobacco (-0.000003) had a negative sign where a positive sign was expected, but the coefficient being essentially zero shows little influence on total suspended solids from burley tobacco present in the 3 km buffer zone. Rainfall was significant at the 90% level and the calculated elasticity for rainfall showed that a one percent increase in rainfall would increase total suspended solids by 51.663% present at the sampling site located in the 3 km buffer zones. Buffered beef cattle was statistically significant at the 95% level, and the calculated elasticity showed that a one percent increase in buffered beef cattle would lead increase total suspended solids by 82.126% present at the sampling site located in the 3 km buffer zones. Burley tobacco was found to be statistically significant at the 95% level and the calculated elasticity for burley tobacco showed that a one percent increase in burley tobacco would lead to a decrease in total suspended solids by 180.689% present at the sampling site located in the 3 km buffer zone.

Fecal Coliform

The results of the three regression models analyzed in SAS for the pathogen variable fecal coliform, revealed that the highest r^2 (0.3312) was achieved by the basic model form. It is important to note that through spatial and data analysis would lead to the belief that fecal coliform would have been described by one of the logged models, but the variation of fecal coilform was best described by the independent variables that were included in the basic model form. The basic model form is given below:

fc_avg = f (resident, forest, crop_past, rain, beef, tobacco).

The results from the basic model found that buffered beef cattle (beef) and buffered burley tobacco (tobacco) are statistically significant at the 90% level and that resident land use and cropland and pastureland (crop_past) are statistically significant at the 95% level. Buffered beef cattle (4.86047), buffered burley tobacco (-0.01404), resident (46,587), and cropland and pastureland (25,624) all had the correct expected signs on their parameter estimates. Buffered beef cattle was significant at the 90% level and the calculated elasticity showed that a one percent increase in buffered beef cattle would lead to a 144.379% increase in the fecal coliform sample reading that is present at the sampling site located in the 3 km buffer zones. Buffered burley tobacco was significant at the 90% level and the calculated elasticity showed that a one percent increase in buffered burley tobacco would lead to a 122.319% decrease in the fecal coliform sample reading that is present at the sampling site located in the 3 km buffer zones. Resident land use was found to be statistically significant at the 95% level and the calculated elasticity showed that a one percent increase in resident land use would lead to a 67.555% increase in the fecal coliform sample reading that is present at the sampling site located in the 3 km buffer zone. Cropland and pastureland was found to be statistically significant at the 95% level and the calculated elasticity showed that a one percent increase in cropland and pastureland would lead to a 424.208% increase in the fecal coliform sample reading that is present at the sampling site located in the 3 km buffer zone.

Table 7.1 OLS Regression Results for Nitrate-Nitrogen, Total Suspended Solids, and Fecal Coliform.

NN_AVG				LNN_AVG				LNN_AVG			
	Estimate	Std. Error	t-value		Estimate	Std. Error	t-value		Estimate	Std. Error	t-value
resident	4.42657	4.44597	1	*resident*	0.48011	1.04677	0.46	*lresident***	0.14692	0.05746	2.56
*forest***	-5.66168	2.95628	-1.92	*forest***	-2.01525	0.69603	-2.9	*lforest*	-0.05846	0.04628	-1.26
crop_past	0.64289	2.66375	0.24	*crop_past*	0.08649	0.62716	0.14	*lcrop_past*	0.45746	0.3171	1.44
*rain***	5.18296	1.73908	2.98	*rain***	0.83593	0.40945	2.04	*lrain*	0.1477	0.17318	0.85
*beef**	0.00136	0.000818	1.67	*beef*	-0.00015	0.00019	-0.75	*lbeef*	0.07375	0.23047	0.32
*dcows**	0.03408	0.01299	2.62	*dcows***	0.00662	0.00306	2.16	*ldcows***	0.55748	0.08944	6.23
R-Square =	0.4197				0.3072				0.6017		
TSS_AVG				LTSS_AVG				LTSS_AVG			
	Estimate	Std. Error	t-value		Estimate	Std. Error	t-value		Estimate	Std. Error	t-value
forest	3.06404	3.48998	0.88	*forest*	0.40035	0.27624	1.45	*lforest*	-0.01463	0.03419	-0.43
crop_past	-0.2054	2.78236	-0.07	*crop_past*	-0.02537	0.22023	-0.12	*lcrop_past*	-0.17297	0.30387	-0.57
*rain***	11.0947	4.39579	2.5	*rain**	0.56658	0.34794	1.63	*lrain*	0.00789	0.21854	0.04
*beef***	0.00377	0.00162	2.33	*beef***	0.000366	0.0001282	2.86	*lbeef*	0.18475	0.21414	0.86
*tobacco***	-0.00003	0.0000058	-5.98	*tobacco***	-0.000003	5E-08	-5.99	*ltobacco***	-0.70467	0.24541	-2.87
dcows	-0.02777	0.02707	-1.03	*dcows*	-0.00291	0.00214	-1.36	*dcows*	-0.02518	0.06976	-0.36
R-Square =	0.4293				0.4743				0.2808		
FC_AVG				LFC_AVG				LFC_AVG			
	Estimate	Std. Error	t-value		Estimate	Std. Error	t-value		Estimate	Std. Error	t-value
*resident***	46587	13275	3.51	*resident**	7.31756	4.21378	1.74	*lresident*	0.26303	0.35897	0.73
forest	13473	9160.038	1.47	*forest*	0.07272	2.90768	0.03	*lforest*	-0.20401	0.27728	-0.74
*crop_past**	25624	9185.018	2.79	*crop_past*	3.80275	2.9156	1.3	*lcrop_past*	-0.14217	2.09577	-0.07
rain	-5296.138	5907.566	-0.9	*rain*	-2.76458	1.87524	-1.47	*lrain*	-0.13676	1.27214	-1.08
*beef**	4.86047	2.57374	1.89	*beef**	0.00159	0.000817	1.95	*lbeef*	1.78103	1.42056	1.25
*tobacco**	-0.01404	0.00845	-1.66	*tobacco*	-0.000003	0.0000029	-1.1	*ltobacco*	0.15797	1.44156	0.11
R-Square =	0.3312				0.288				0.2103		

** Denotes statistical siginificance at the .10 level.*

*** Denotes statistical significance at the .05 level.*

Table 7.2 Elasticity Caluclations for Each Functional Form

NN				Semi-log NN				Complete-log NN			
	mean	Estimate	Elasticity		mean	Estimate	Elasticity		mean	Estimate	Elasticity
nn_avg	2.71350	-	-	*log nn_avg*	-0.22914	-	-	*log nn_avg*	-0.22914	-	-
resident	0.06625	4.42657	0.10807	*resident*	0.06625	0.48011	-0.13881	*log resident*	0.06625	0.14692	0.14692
forest	0.14196	-5.66168	-0.29620	*forest*	0.14196	-2.01525	1.24853	*log forest*	0.14196	-0.05846	-0.05846
crop_past	0.73170	0.64289	0.17336	*crop_past*	0.73170	0.08649	-0.27619	*log crop_past*	0.73170	0.45746	0.45746
rain	0.53098	5.18296	1.01421	*rain*	0.53098	0.83593	-1.93710	*log rain*	0.53098	0.14770	0.14770
beef	1305	0.00136	0.65406	*beef*	1305	-0.00015	0.82684	*log beef*	1305	0.07375	0.07375
dcows	17.66779	0.03408	0.22190	*dcows*	17.66779	0.00662	-0.51044	*log dcows*	17.66779	0.55748	0.55748
TSS				Semi-log TSS				Complete-log TSS			
	mean	Estimate	Elasticity		mean	Estimate	Elasticity		mean	Estimate	Elasticity
tss_avg	11.49888	-	-	*log tss_avg*	0.58232	-	-	*log tss_avg*	0.58232	-	-
forest	0.14196	3.06404	0.03783	*forest*	0.14196	0.40035	0.09760	*log forest*	0.14196	-0.01463	-0.01463
crop_past	0.73170	-0.20540	-0.01307	*crop_past*	0.73170	-0.02537	-0.03188	*log crop_past*	0.73170	-0.17297	-0.17297
rain	0.53098	11.00947	0.50838	*rain*	0.53098	0.56658	0.51663	*log rain*	0.53098	0.00789	0.00789
beef	1305	0.00377	0.42785	*beef*	1305	0.00037	0.82126	*log beef*	1305	0.18475	0.18475
tobacco	385414	-0.00003	-1.15434	*tobacco*	385414	0.00000	-1.80689	*log tobacco*	385414	-0.70467	-0.70467
dcows	17.66779	-0.02777	-0.04267	*dcows*	17.66779	-0.00291	-0.08829	*log dcows*	17.66779	-0.02518	-0.02518
FC				Semi-log FC				Complete-log FC			
	mean	Estimate	Elasticity		mean	Estimate	Elasticity		mean	Estimate	Elasticity
fc_avg	4437	-	-	*log fc_avg*	3.12968	-	-	*log fc_avg*	3.12968	-	-
resident	0.06434	46587	0.67555	*resident*	0.06434	7.31756	0.15043	*log resident*	0.06434	0.26303	0.26303
forest	0.14582	13473	0.44278	*forest*	0.14582	0.07272	0.00339	*log forest*	0.14582	-0.20401	-0.20401
crop_past	0.73455	25624	4.24208	*crop_past*	0.73455	3.80275	0.89252	*log crop_past*	0.73455	-0.14217	-0.14217
rain	0.53700	-5296.13818	-0.64098	*rain*	0.53700	-2.76458	-0.47436	*log rain*	0.53700	-0.13676	-0.13676
beef	1318	4.86047	1.44379	*beef*	1318	0.00159	0.66960	*log beef*	1318	1.78103	1.78103
tobacco	386560	-0.01404	-1.22319	*tobacco*	386560	0.00000	-0.36313	*log tobacco*	386560	0.15797	0.15797

CHAPTER EIGHT: Conclusion

The purpose of this study was to discover factors that were significant in the transport of pollutants from nonpoint source pollution in the Lower Kentucky Watershed in central Kentucky. The analysis required the use of two forms of analytical software, a Geographic Information Software (ArcMap) and statistical analysis software (SAS). The main forms of nonpoint source pollution selected to be investigated in this study were nutrients (nitrate-nitrogen and total recoverable phosphorus (TRP)), siltation (total suspended solids), and pathogens (fecal coliform).

ArcMap was utilized to spatially analyze the sampling site data, as well as, create 3 km buffer zones around the sampling sites. After further spatial analysis, it was concluded that the calculation of an expected value was necessary to explain the extreme variation that was present in the sampling site data. After the completion of use with ArcMap, areas of each of the independent variables had been calculated and were ready to be manipulated by the second analytical tool, Statistical Analysis Software (SAS). SAS was used for data manipulation, as well as, performing the regressions (after the final models were selected.)

After final models for each of the dependent variables were created by the use of a correlation matrix, each model was then analyzed in SAS as a basic OLS form model, semi-log transformation model, and complete log transformation model. All dependent variables were included in the final analysis except for total recoverable phosphorus, due to the small amount of variation that was explained by the independent variables included in its models. Land use (resident) and agriculture (buffered dairy cows) variables were found to be significant predictors of nitrate-nitrogen. Rainfall and agriculture (buffered beef cattle and buffered burley tobacco) variables were found to be significant predictors of total suspended solids. Land use (resident and cropland and pastureland) and agriculture (buffered beef cattle and buffered burley tobacco) variables were found to be significant predictors of fecal coliform.

Implications

The results from this study are subject to scrutiny, based on assumptions that have been made, but the results that have been generated still can be a useful tool for state and local agencies and land managers. The results that were generated show that both agricultural factors and factors outside of agriculture are significant in the transport of pollutants to waterways in the

Lower Kentucky Watershed, even though agriculture is generally portrayed as the leading cause of nonpoint source pollution.

Future Research

As mentioned earlier, the results from this study are subject to scrutiny, but offer many opportunities for future research. It is imperative that similar studies are performed in many different locations across the United States, due to the fact that these kinds of studies are very "location specific." With similar studies replicated across many different landscapes, decision makers will be able to make the appropriate decisions dealing with water quality management issues.

Another aspect of this study that can be addressed through future research would be the addition of the other variables that were described in the theoretical model section. The variables that were omitted from the empirical model were based on limited availability, but the addition of the omitted independent variables that were mentioned in the theoretical model section would create a more comprehensive model. Digital Elevation Models (DEM's) can be utilized by future research, in an effort to check for significance of topographic features on the sample loads. It is also appropriate for future research to look at the affect that soil conditions have on sample loads, even though they were found to be insignificance in previous studies. Hydrology data was not included due to a lack of data, which caused many instances of missing values, but future research can examine ways to perform similar data manipulation done to the clusters to the data that is currently available.

In terms of the econometric model used, a different model form would likely lead to different results. The area of the econometric model that can be addressed by future research is the classification of the dependent variable. In this study, the dependent variables are characterized as continuous and classified as the sampling site data for each of the sampling sites, but a different classification of the dependent variable would likely lead to different results.

In dealing with the large amount of multicollinearity that was present between some independent variables (especially resident land use and the socioeconomic variables), the decision was made to only keep one of the variables present in the econometric model. This is only one technique that could have been employed to help correct for the multicollinearity that was present. Future research has the opportunity to look at other solutions for trying to create a model that uses some of the other options for dealing with high levels of multicollinearity, most notably a variable transformation.

The last, but most important aspect of this study that can be addressed by future research is the scale issues that surround the selected buffer zones. The results that were generated by this study are relevant at the selected buffer zone size, but the selection of a different buffer zone size will lead to different results. The selection of new buffer zone scales will allow researchers to see at what distances certain variables are found to be significant in the transport of pollutants. The replication of studies with different buffer zone scales will allow new policies to be implemented that to will allow for improved water quality throughout multiple regions of the United States.

APPENDIX

Nitrate-Nitrogen SAS Log

```
data xxx1;
merge jimmy2.all_avg jimmy2.fc_avg jimmy2.nn_avg jimmy2.trp_avg jimmy2.tss_avg;
by sample_id;
run;

proc corr data=jimmy2.nn_avg;
var nn_avg resident_ forest crop_past _62 rain beef tobacco dcows;
run;

data xxx2;
set jimmy2.nn_avg;
lnn_avg  = log (nn_avg);
*resident_ = log (resident_);
*forest    = log (forest);
*crop_past       = log (crop_past);
*_62       = log (_62);
*rain  = log (rain);
*beef    = log ( beef);
*tobacco     = log (tobacco);
*dcows   = log (dcows);
*if fc02_avg = '.' then delete;
*if fc02_avg = ' ' then delete;
run;

data xxx3;
set jimmy2.nn_avg;
lnn_avg    = log (nn_avg);
lresident_ = log (resident_);
lforest    = log (forest);
lcrop_past       = log (crop_past);
lrain  = log (rain);
lbeef    = log ( beef);
ltobacco     = log (tobacco);
ldcows   = log (dcows);
*if fc02_avg = '.' then delete;
*if fc02_avg = ' ' then delete;
run;


proc reg data=xxx2;
model nn_avg = resident_ forest crop_past rain beef dcows;
run;
```

```
proc reg data=xxx2;
model lnn_avg = resident_ forest crop_past rain beef dcows;
run;
```

Nitrate-Nitrogen SAS Results

The CORR Procedure

9 Variables: NN_avg Resident_ Forest crop_past _62 rain beef tobacco dcows

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
NN_avg	80	2.71350	2.46809	217.08000	1.43000	12.61000	NN_avg
Resident_	80	0.06625	0.11051	5.29966	0	0.54260	Resident
Forest	80	0.14196	0.15084	11.35716	0	0.43942	Forest
crop_past	80	0.73170	0.19754	58.53628	0.07578	0.98450	crop_past
_62	80	0.00107	0.00681	0.08594	0	0.04887	_62
rain	80	0.53098	0.14328	42.47827	0.31906	0.80460	rain
beef	80	1305	313.69684	104365	530.07239	1916	beef
tobacco	80	385414	108287	30833137	156600	590858	tobacco
dcows	80	17.66779	20.01541	1413	6.21904	94.50716	dcows

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	NN_avg	Resident_	Forest	crop_past	_62
NN_avg	1.00000	0.18466	-0.37279	0.15424	-0.06040
NN_avg		0.1010	0.0007	0.1719	0.5946
Resident_	0.18466	1.00000	-0.29178	-0.60146	-0.08664
Resident	0.1010		0.0086	<.0001	0.4448
Forest	-0.37279	-0.29178	1.00000	-0.46170	-0.11808
Forest	0.0007	0.0086		<.0001	0.2969

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
NN_avg	0.37677	0.29294	-0.03027	0.29668
NN_avg	0.0006	0.0084	0.7898	0.0075
Resident_	-0.00543	-0.25825	-0.04999	-0.13437
Resident	0.9619	0.0207	0.6597	0.2347
Forest	0.07984	-0.25237	-0.22480	0.26973
Forest	0.4815	0.0239	0.0450	0.0155

The CORR Procedure

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	NN_avg	Resident_	Forest	crop_past	_62
crop_past	0.15424	-0.60146	-0.46170	1.00000	0.03675
crop_past	0.1719	<.0001	<.0001		0.7462
_62	-0.06040	-0.08664	-0.11808	0.03675	1.00000
_62	0.5946	0.4448	0.2969	0.7462	
rain	0.37677	-0.00543	0.07984	0.07319	-0.23069
rain	0.0006	0.9619	0.4815	0.5188	0.0395
beef	0.29294	-0.25825	-0.25237	0.38079	0.11596
beef	0.0084	0.0207	0.0239	0.0005	0.3057
tobacco	-0.03027	-0.04999	-0.22480	0.34441	-0.16470
tobacco	0.7898	0.6597	0.0450	0.0018	0.1443
dcows	0.29668	-0.13437	0.26973	-0.09336	-0.04252
dcows	0.0075	0.2347	0.0155	0.4101	0.7080

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
crop_past	0.07319	0.38079	0.34441	-0.09336
crop_past	0.5188	0.0005	0.0018	0.4101
_62	0.23069	0.11596	-0.16470	-0.04252
_62	0.0395	0.3057	0.1443	0.7080
rain	1.00000	0.01258	0.57273	0.35690
rain		0.9118	<.0001	0.0012
beef	0.01258	1.00000	0.03094	0.21730
beef	0.9118		0.7853	0.0528
tobacco	0.57273	0.03094	1.00000	-0.08159
tobacco	<.0001	0.7853		0.4718
dcows	0.35690	0.21730	-0.08159	1.00000
dcows	0.0012	0.0528	0.4718	

The REG Procedure
Model: MODEL1
Dependent Variable: NN_avg NN_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	201.95760	33.65960	8.80	<.0001
Error	73	279.26942	3.82561		
Corrected Total	79	481.22702			

Root MSE	1.95592	R-Square	0.4197
Dependent Mean	2.71350	Adj R-Sq	0.3720
Coeff Var	72.08094		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-2.38001	2.71264	-0.88	0.3832
Resident_	Resident	1	4.42657	4.44597	1.00	0.3227
Forest	Forest	1	-5.66168	2.95628	-1.92	0.0594
crop_past	crop_past	1	0.64289	2.66375	0.24	0.8100
rain	rain	1	5.18296	1.73908	2.98	0.0039
beef	beef	1	0.00136	0.00081797	1.67	0.0997
dcows	dcows	1	0.03408	0.01299	2.62	0.0106

The REG Procedure
Model: MODEL1
Dependent Variable: log nn_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	9.10120	1.51687	7.15	<.0001
Error	73	15.48061	0.21206		
Corrected Total	79	24.58181			

Root MSE	0.46050	R-Square	0.3702
Dependent Mean	0.79339	Adj R-Sq	0.3185
Coeff Var	58.04263		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.61303	0.63867	0.96	0.3403
Resident_	Resident	1	0.48011	1.04677	0.46	0.6478
Forest	Forest	1	-2.01525	0.69603	-2.90	0.0050
crop_past	crop_past	1	0.08649	0.62716	0.14	0.8907
rain	rain	1	0.83593	0.40945	2.04	0.0448
beef	beef	1	-0.00014518	0.00019258	-0.75	0.4534
dcows	dcows	1	0.00662	0.00306	2.16	0.0338

The REG Procedure
Model: MODEL1
Dependent Variable: lnn_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	8.51866	1.41978	11.33	<.0001
Error	45	5.63904	0.12531		
Corrected Total	51	14.15770			

Root MSE	0.35399	R-Square	0.6017
Dependent Mean	0.63806	Adj R-Sq	0.5486
Coeff Var	55.47971		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.64304	1.60746	-0.40	0.6910
Log resident	1	0.14692	0.05746	2.56	0.0140
Log forest	1	-0.05846	0.04628	-1.26	0.2131
Log crop_past	1	0.45746	0.31710	1.44	0.1561
Log rain	1	0.14770	0.17318	0.85	0.3982
Log beef	1	0.07375	0.23047	0.32	0.7505
Log dcows	1	0.55748	0.08944	6.23	<.0001

Total Recoverable Phosphorus SAS Log

```
data xxx1;
merge jimmy2.all_avg jimmy2.fc_avg jimmy2.nn_avg jimmy2.trp_avg jimmy2.tss_avg;
by sample_id;
run;

proc corr data=jimmy2.trp_avg;
var trp_avg resident_ forest crop_past _62 rain beef tobacco dcows;
run;

data xxx2;
set jimmy2.trp_avg;
ltrp_avg  = log (trp_avg);
*resident_ = log (resident_);
*forest    = log (forest);
*crop_past       = log (crop_past);
*_62       = log (_62);
*rain  = log (rain);
*beef    = log ( beef);
*tobacco     = log (tobacco);
*dcows   = log (dcows);
run;

data xxx3;
set jimmy2.trp_avg;
ltrp_avg    = log (trp_avg);
lresident_ = log (resident_);
lforest    = log (forest);
lcrop_past       = log (crop_past);
lrain  = log (rain);
lbeef    = log ( beef);
ltobacco     = log (tobacco);
ldcows   = log (dcows);
run;

proc reg data=xxx2;
model trp_avg = resident_ forest crop_past rain beef tobacco dcows;
run;

proc reg data=xxx2;
model ltrp_avg = resident_ forest crop_past rain beef tobacco dcows;
run;
proc reg data=xxx3;
model ltrp_avg = lresident_ lforest lcrop_past lrain lbeef ltobacco ldcows;
run;
```

The CORR Procedure

Total Recoverable Phosphorus SAS Results

```
9  Variables:    TRP_avg   Resident_ Forest    crop_past       _62       rain   beef    tobacco
                   dcows
```

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
TRP_avg	80	0.46238	0.13980	36.99000	0.32000	0.69000	TRP_avg
Resident_	80	0.06625	0.11051	5.29966	0	0.54260	Resident
Forest	80	0.14196	0.15084	11.35716	0	0.43942	Forest
crop_past	80	0.73170	0.19754	58.53628	0.07578	0.98450	crop_past
_62	80	0.00107	0.00681	0.08594	0	0.04887	_62
rain	80	0.53098	0.14328	42.47827	0.31906	0.80460	rain
beef	80	1305	313.69684	104365	530.07239	1916	beef
tobacco	80	385414	108287	30833137	156600	590858	tobacco
dcows	80	17.66779	20.01541	1413	6.21904	94.50716	dcows

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	TRP_avg	Resident_	Forest	crop_past	_62
TRP_avg	1.00000	-0.01282	-0.01838	0.04623	-0.05859
TRP_avg		0.9101	0.8714	0.6839	0.6057
Resident_	-0.01282	1.00000	-0.29178	-0.60146	-0.08664
Resident	0.9101		0.0086	<.0001	0.4448
Forest	-0.01838	-0.29178	1.00000	-0.46170	-0.11808
Forest	0.8714	0.0086		<.0001	0.2969

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
TRP_avg	0.06966	-0.03116	-0.02622	-0.10587
TRP_avg	0.5392	0.7838	0.8174	0.3499
Resident_	-0.00543	-0.25825	-0.04999	-0.13437
Resident	0.9619	0.0207	0.6597	0.2347
Forest	0.07984	-0.25237	-0.22480	0.26973
Forest	0.4815	0.0239	0.0450	0.0155

The CORR Procedure

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	TRP_avg	Resident_	Forest	crop_past	_62
crop_past	0.04623	-0.60146	-0.46170	1.00000	0.03675
crop_past	0.6839	<.0001	<.0001		0.7462
_62	-0.05859	-0.08664	-0.11808	0.03675	1.00000
_62	0.6057	0.4448	0.2969	0.7462	
rain	0.06966	-0.00543	0.07984	0.07319	-0.23069
rain	0.5392	0.9619	0.4815	0.5188	0.0395
beef	-0.03116	-0.25825	-0.25237	0.38079	0.11596
beef	0.7838	0.0207	0.0239	0.0005	0.3057
tobacco	-0.02622	-0.04999	-0.22480	0.34441	-0.16470
tobacco	0.8174	0.6597	0.0450	0.0018	0.1443
dcows	-0.10587	-0.13437	0.26973	-0.09336	-0.04252
dcows	0.3499	0.2347	0.0155	0.4101	0.7080

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
crop_past	0.07319	0.38079	0.34441	-0.09336
crop_past	0.5188	0.0005	0.0018	0.4101
_62	0.23069	0.11596	-0.16470	-0.04252
_62	0.0395	0.3057	0.1443	0.7080
rain	1.00000	0.01258	0.57273	0.35690
rain		0.9118	<.0001	0.0012
beef	0.01258	1.00000	0.03094	0.21730
beef	0.9118		0.7853	0.0528
tobacco	0.57273	0.03094	1.00000	-0.08159
tobacco	<.0001	0.7853		0.4718
dcows	0.35690	0.21730	-0.08159	1.00000
dcows	0.0012	0.0528	0.4718	

The REG Procedure
Model: MODEL1
Dependent Variable: TRP_avg TRP_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.08269	0.01034	0.50	0.8507
Error	71	1.46136	0.02058		
Corrected Total	79	1.54405			

Root MSE	0.14347	R-Square	0.0536
Dependent Mean	0.46238	Adj R-Sq	-0.0531
Coeff Var	31.02813		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.43555	0.20610	2.11	0.0381
Resident_	Resident	1	-0.01292	0.33566	-0.04	0.9694
Forest	Forest	1	-0.00750	0.22339	-0.03	0.9733
crop_past	crop_past	1	0.05962	0.20316	0.29	0.7700
_62	_62	1	-1.00605	2.53491	-0.40	0.6926
rain	rain	1	0.25288	0.16440	1.54	0.1285
beef	beef	1	-0.00000589	0.00006010	-0.10	0.9222
tobacco	tobacco	1	-2.97723E-7	2.117632E-7	-1.41	0.1641
dcows	dcows	1	-0.00145	0.00099505	-1.46	0.1492

The REG Procedure
Model: MODEL1
Dependent Variable: ltrp_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.43438	0.05430	0.63	0.7506
Error	71	6.12723	0.08630		
Corrected Total	79	6.56162			

Root MSE	0.29377	R-Square	0.0662
Dependent Mean	-0.81367	Adj R-Sq	-0.0390
Coeff Var	-36.10406		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-0.85773	0.42202	-2.03	0.0459
Resident_	Resident	1	0.04288	0.68731	0.06	0.9504
Forest	Forest	1	-0.00145	0.45743	-0.00	0.9975
crop_past	crop_past	1	0.15118	0.41600	0.36	0.7174
_62	_62	1	-2.12266	5.19058	-0.41	0.6838
rain	rain	1	0.52389	0.33664	1.56	0.1241
beef	beef	1	-0.00000191	0.00012307	-0.02	0.9876
tobacco	tobacco	1	-7.29509E-7	4.336143E-7	-1.68	0.0969
dcows	dcows	1	-0.00348	0.00204	-1.71	0.0923

The REG Procedure
Model: MODEL1
Dependent Variable: ltrp_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.56130	0.08019	1.01	0.4406
Error	44	3.50956	0.07976		
Corrected Total	51	4.07086			

Root MSE	0.28242	R-Square	0.1379
Dependent Mean	-0.81110	Adj R-Sq	0.0007
Coeff Var	-34.81957		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.34697	2.96354	0.45	0.6517
Log resident	1	0.05126	0.04736	1.08	0.2850
Log forest	1	0.02851	0.03698	0.77	0.4448
Log crop_past	1	0.00684	0.25946	0.03	0.9791
Log rain	1	0.06077	0.18990	0.32	0.7505
Log beef	1	0.07427	0.18525	0.40	0.6904
Log tobacco	1	-0.20764	0.21160	-0.98	0.3318
Log dcows	1	0.09890	0.07649	1.29	0.2028

Total Suspended Solids SAS Log

```
data xxx1;
merge jimmy2.all_avg jimmy2.fc_avg jimmy2.nn_avg jimmy2.trp_avg jimmy2.tss_avg;
by sample_id;
run;

proc corr data=jimmy2.tss_avg;
var tss_avg resident_ forest crop_past _62 rain beef tobacco dcows;
run;

data xxx2;
set jimmy2.tss_avg;
ltss_avg  = log (tss_avg);
*resident_ = log (resident_);
*forest    = log (forest);
*crop_past       = log (crop_past);
*_62       = log (_62);
*rain  = log (rain);
*beef    = log ( beef);
*tobacco     = log (tobacco);
*dcows   = log (dcows);
run;

data xxx3;
set jimmy2.tss_avg;
ltss_avg    = log (tss_avg);
lresident_ = log (resident_);
lforest    = log (forest);
lcrop_past       = log (crop_past);
lrain  = log (rain);
lbeef    = log ( beef);
ltobacco     = log (tobacco);
ldcows   = log (dcows);
run;

proc reg data=xxx2;
model tss_avg = forest crop_past rain beef tobacco dcows;
run;

proc reg data=xxx2;
model ltss_avg = forest crop_past rain beef tobacco dcows;
run;
proc reg data=xxx3;
model ltss_avg = lforest lcrop_past lrain lbeef ltobacco ldcows;
run;
```

Total Suspended Solids SAS Results

The CORR Procedure

9 Variables: TSS_avg Resident_ Forest crop_past _62 rain beef tobacco dcows

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
TSS_avg	80	11.49888	4.98202	919.91000	6.89000	23.25000	TSS_avg
Resident_	80	0.06625	0.11051	5.29966	0	0.54260	Resident
Forest	80	0.14196	0.15084	11.35716	0	0.43942	Forest
crop_past	80	0.73170	0.19754	58.53628	0.07578	0.98450	crop_past
_62	80	0.00107	0.00681	0.08594	0	0.04887	_62
rain	80	0.53098	0.14328	42.47827	0.31906	0.80460	rain
beef	80	1305	313.69684	104365	530.07239	1916	beef
tobacco	80	385414	108287	30833137	156600	590858	tobacco
dcows	80	17.66779	20.01541	1413	6.21904	94.50716	dcows

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	TSS_avg	Resident_	Forest	crop_past	_62
TSS_avg	1.00000	0.03164	0.20006	-0.18476	0.01606
TSS_avg		0.7806	0.0752	0.1009	0.8875
Resident_	0.03164	1.00000	-0.29178	-0.60146	-0.08664
Resident	0.7806		0.0086	<.0001	0.4448
Forest	0.20006	-0.29178	1.00000	-0.46170	-0.11808
Forest	0.0752	0.0086		<.0001	0.2969

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
TSS_avg	-0.14213	0.16758	-0.57445	0.13992
TSS_avg	0.2085	0.1373	<.0001	0.2158
Resident_	-0.00543	-0.25825	-0.04999	-0.13437
Resident	0.9619	0.0207	0.6597	0.2347
Forest	0.07984	-0.25237	-0.22480	0.26973
Forest	0.4815	0.0239	0.0450	0.0155

The CORR Procedure

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	TSS_avg	Resident_	Forest	crop_past	_62
crop_past	-0.18476	-0.60146	-0.46170	1.00000	0.03675
crop_past	0.1009	<.0001	<.0001		0.7462
_62	0.01606	-0.08664	-0.11808	0.03675	1.00000
_62	0.8875	0.4448	0.2969	0.7462	
rain	-0.14213	-0.00543	0.07984	0.07319	-0.23069
rain	0.2085	0.9619	0.4815	0.5188	0.0395
beef	0.16758	-0.25825	-0.25237	0.38079	0.11596
beef	0.1373	0.0207	0.0239	0.0005	0.3057
tobacco	-0.57445	-0.04999	-0.22480	0.34441	-0.16470
tobacco	<.0001	0.6597	0.0450	0.0018	0.1443
dcows	0.13992	-0.13437	0.26973	-0.09336	-0.04252
dcows	0.2158	0.2347	0.0155	0.4101	0.7080

Pearson Correlation Coefficients, N = 80
Prob > |r| under H0: Rho=0

	rain	beef	tobacco	dcows
crop_past	0.07319	0.38079	0.34441	-0.09336
crop_past	0.5188	0.0005	0.0018	0.4101
_62	-0.23069	0.11596	-0.16470	-0.04252
_62	0.0395	0.3057	0.1443	0.7080
rain	1.00000	0.01258	0.57273	0.35690
rain		0.9118	<.0001	0.0012
beef	0.01258	1.00000	0.03094	0.21730
beef	0.9118		0.7853	0.0528
tobacco	0.57273	0.03094	1.00000	-0.08159
tobacco	<.0001	0.7853		0.4718
dcows	0.35690	0.21730	-0.08159	1.00000
dcows	0.0012	0.0528	0.4718	

The REG Procedure
Model: MODEL1
Dependent Variable: TSS_avg TSS_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	841.72333	140.28722	9.15	<.0001
Error	73	1119.09806	15.33011		
Corrected Total	79	1960.82140			

Root MSE	3.91537	R-Square	0.4293
Dependent Mean	11.49888	Adj R-Sq	0.3824
Coeff Var	34.05001		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	14.21165	3.06317	4.64	<.0001
Forest	Forest	1	3.06404	3.48998	0.88	0.3828
crop_past	crop_past	1	-0.20540	2.78236	-0.07	0.9414
rain	rain	1	11.00947	4.39579	2.50	0.0145
beef	beef	1	0.00377	0.00162	2.33	0.0226
tobacco	tobacco	1	-0.00003444	0.00000576	-5.98	<.0001
dcows	dcows	1	-0.02777	0.02707	-1.03	0.3083

The REG Procedure
Model: MODEL1
Dependent Variable: ltss_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	6.32514	1.05419	10.98	<.0001
Error	73	7.01135	0.09605		
Corrected Total	79	13.33649			

Root MSE	0.30991	R-Square	0.4743
Dependent Mean	2.35599	Adj R-Sq	0.4311
Coeff Var	13.15422		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	2.64296	0.24246	10.90	<.0001
Forest	Forest	1	0.40035	0.27624	1.45	0.1515
crop_past	crop_past	1	-0.02537	0.22023	-0.12	0.9086
rain	rain	1	0.56658	0.34794	1.63	0.1078
beef	beef	1	0.00036646	0.00012818	2.86	0.0055
tobacco	tobacco	1	-0.00000273	4.560023E-7	-5.99	<.0001
dcows	dcows	1	-0.00291	0.00214	-1.36	0.1792

The REG Procedure
Model: MODEL1
Dependent Variable: ltss_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2.35733	0.39289	3.25	0.0089
Error	50	6.03801	0.12076		
Corrected Total	56	8.39534			

Root MSE	0.34751	R-Square	0.2808
Dependent Mean	2.47106	Adj R-Sq	0.1945
Coeff Var	14.06300		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.10898	3.33639	3.03	0.0039
Log forest	1	-0.01463	0.03419	-0.43	0.6705
log crop_past	1	-0.17297	0.30387	-0.57	0.5717
log rain	1	0.00789	0.21854	0.04	0.9713
log beef	1	0.18475	0.21414	0.86	0.3924
log tobacco	1	-0.70467	0.24541	-2.87	0.0060
log dcows	1	-0.02518	0.06976	-0.36	0.7197

Fecal Coliform SAS Log

```
data xxx1;
merge jimmy2.all_avg jimmy2.fc_avg jimmy2.nn_avg jimmy2.trp_avg jimmy2.tss_avg;
by sample_id;
run;

*proc sort data=xxx1;
*by sample_id;
*run;

proc corr data=jimmy2.fc_avg;
var fc02_avg resident_ forest crop_past _62 rain beef tobacco dcows;
run;

data xxx2;
set jimmy2.fc_avg;
lfc02_avg  = log (fc02_avg);
*resident_ = log (resident_);
*forest    = log (forest);
*crop_past       = log (crop_past);
*_62       = log (_62);
*rain  = log (rain);
*beef    = log ( beef);
*tobacco     = log (tobacco);
*dcows   = log (dcows);
if fc02_avg = '.' then delete;
if fc02_avg = ' ' then delete;
run;

data xxx3;
set jimmy2.fc_avg;
lfc02_avg  = log (fc02_avg);
lresident_ = log (resident_);
lforest    = log (forest);
lcrop_past       = log (crop_past);
lrain  = log (rain);
lbeef    = log ( beef);
ltobacco     = log (tobacco);
ldcows   = log (dcows);
if fc02_avg = '.' then delete;
if fc02_avg = ' ' then delete;
run;

proc reg data=xxx2;
model fc02_avg = resident_ forest crop_past rain beef tobacco;
```

```
run;

proc reg data=xxx2;
model lfc02_avg = resident_ forest crop_past rain beef tobacco;
run;

proc reg data=xxx3;
model lfc02_avg = lresident_ lforest lcrop_past lrain lbeef ltobacco;
run;
```

Fecal Coliform SAS Results

The CORR Procedure

9 Variables: FC02_avg Resident_ Forest crop_past _62 rain beef tobacco dcows

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
FC02_avg	63	4437	6298	279507	10.00000	31000	FC02_avg
Resident_	66	0.06434	0.11090	4.24616	0	0.54260	Resident
Forest	66	0.14582	0.15407	9.62425	0	0.43942	Forest
crop_past	66	0.73455	0.18761	48.48033	0.15720	0.98450	crop_past
_62	66	0.00130	0.00749	0.08594	0	0.04887	_62
rain	66	0.53700	0.15007	35.44199	0.31906	0.80460	rain
beef	66	1318	311.40395	87004	530.07239	1916	beef
tobacco	66	386560	112906	25512969	156600	590858	tobacco
dcows	66	19.59483	21.52814	1293	6.21904	94.50716	dcows

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	FC02_avg	Resident_	Forest	crop_past	_62
FC02_avg	1.00000	0.26974	-0.32770	0.11856	-0.07474
FC02_avg		0.0325	0.0087	0.3547	0.5605
	63	63	63	63	63
Resident_	0.26974	1.00000	-0.29085	-0.56570	-0.09239
Resident	0.0325		0.0178	<.0001	0.4606
	63	66	66	66	66

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	rain	beef	tobacco	dcows
FC02_avg	-0.16408	0.27050	-0.14593	-0.03352
FC02_avg	0.1988	0.0320	0.2538	0.7942
	63	63	63	63
Resident_	-0.01482	-0.22001	-0.06114	-0.14084
Resident	0.9060	0.0759	0.6258	0.2593
	66	66	66	66

The CORR Procedure

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	FCO2_avg	Resident_	Forest	crop_past	_62
Forest	-0.32770	-0.29085	1.00000	-0.49923	-0.13221
Forest	0.0087	0.0178		<.0001	0.2900
	63	66	66	66	66
crop_past	0.11856	-0.56570	-0.49923	1.00000	0.04010
crop_past	0.3547	<.0001	<.0001		0.7492
	63	66	66	66	66
_62	-0.07474	-0.09239	-0.13221	0.04010	1.00000
_62	0.5605	0.4606	0.2900	0.7492	
	63	66	66	66	66
rain	-0.16408	-0.01482	0.09703	0.08136	-0.25054
rain	0.1988	0.9060	0.4383	0.5161	0.0425
	63	66	66	66	66
beef	0.27050	-0.22001	-0.32347	0.44007	0.12137
beef	0.0320	0.0759	0.0081	0.0002	0.3316
	63	66	66	66	66
tobacco	-0.14593	-0.06114	-0.18290	0.36066	-0.17640
tobacco	0.2538	0.6258	0.1416	0.0029	0.1565
	63	66	66	66	66
dcows	-0.03352	-0.14084	0.27401	-0.11455	-0.05950
dcows	0.7942	0.2593	0.0260	0.3597	0.6351
	63	66	66	66	66

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	rain	beef	tobacco	dcows
Forest	0.09703	-0.32347	-0.18290	0.27401
Forest	0.4383	0.0081	0.1416	0.0260
	66	66	66	66
crop_past	0.08136	0.44007	0.36066	-0.11455
crop_past	0.5161	0.0002	0.0029	0.3597
	66	66	66	66

The CORR Procedure

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	rain	beef	tobacco	dcows
_62	-0.25054	0.12137	-0.17640	-0.05950
_62	0.0425	0.3316	0.1565	0.6351
	66	66	66	66
rain	1.00000	0.03217	0.58128	0.36616
rain		0.7976	<.0001	0.0025
	66	66	66	66
beef	0.03217	1.00000	0.04922	0.22767
beef	0.7976		0.6947	0.0660
	66	66	66	66
tobacco	0.58128	0.04922	1.00000	-0.08123
tobacco	<.0001	0.6947		0.5168
	66	66	66	66
dcows	0.36616	0.22767	-0.08123	1.00000
dcows	0.0025	0.0660	0.5168	
	66	66	66	66

The REG Procedure
Model: MODEL1
Dependent Variable: FCO2_avg FCO2_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	814417293	135736216	4.62	0.0007
Error	56	1644803121	29371484		
Corrected Total	62	2459220414			

Root MSE	5419.54650	R-Square	0.3312
Dependent Mean	4436.62169	Adj R-Sq	0.2595
Coeff Var	122.15480		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-17414	8301.29085	-2.10	0.0405
Resident_	Resident	1	46587	13275	3.51	0.0009
Forest	Forest	1	13473	9160.03828	1.47	0.1469
crop_past	crop_past	1	25624	9185.01792	2.79	0.0072
rain	rain	1	-5296.13818	5907.56579	-0.90	0.3738
beef	beef	1	4.86047	2.57374	1.89	0.0641
tobacco	tobacco	1	-0.01404	0.00845	-1.66	0.1022

The REG Procedure
Model: MODEL1
Dependent Variable: lfc02_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	67.05084	11.17514	3.78	0.0032
Error	56	165.73381	2.95953		
Corrected Total	62	232.78465			

Root MSE	1.72033	R-Square	0.2880
Dependent Mean	7.13478	Adj R-Sq	0.2118
Coeff Var	24.11189		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	4.39134	2.63508	1.67	0.1012
Resident_	Resident	1	7.31756	4.21378	1.74	0.0880
Forest	Forest	1	0.07272	2.90768	0.03	0.9801
crop_past	crop_past	1	3.80275	2.91560	1.30	0.1975
rain	rain	1	-2.76458	1.87524	-1.47	0.1460
beef	beef	1	0.00159	0.00081698	1.95	0.0562
tobacco	tobacco	1	-0.00000294	0.00000268	-1.10	0.2777

The REG Procedure
Model: MODEL1
Dependent Variable: lfc02_avg

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	34.21727	5.70288	1.51	0.2048
Error	34	128.45718	3.77815		
Corrected Total	40	162.67445			

Root MSE	1.94375	R-Square	0.2103
Dependent Mean	7.20265	Adj R-Sq	0.0710
Coeff Var	26.98655		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-8.03042	20.16114	-0.40	0.6929
Log resident	1	0.26303	0.35897	0.73	0.4687
Log forest	1	-0.20401	0.27728	-0.74	0.4669
Log crop_past	1	-0.14217	2.09577	-0.07	0.9463
Log rain	1	-1.36759	1.27214	-1.08	0.2899
Log beef	1	1.78103	1.42056	1.25	0.2185
Log tobacco	1	0.15797	1.44156	0.11	0.9134

Bibliography

2000 National Water Quality Inventory. 2000. Environmental Protection Agency. Rivers and Streams, Lakes and Reservoirs. http://www.epa.gov/305b/2000report/

Berry, J.K. 2003. Map Analysis: Procedures and Applications in GIS Modeling. BASINS Press, Fort Collins, Colorado.

Berry, J.K., 2003. Analyzing Spatial Content. pg. 55-63. In: Analyzing Precision Ag Data: A Hands-On Case Study in Spatial Analysis and Data Mining. J.K. Berry (ed.) Basis Press, Fort Collins, Colorado.

Berry, J.K., Delgado, J.A., Pierce, F.J., and Khosla, R. 2005. Applying Spatial Analysis for Precision Conservation Arcoss the Landscape. Journal of Soil and Water Conservation. 60(6):363-370.

Calhoun, F.G., Baker, B.D., and Slater, B.K. 2002. Soils, water quality, and watershed size: interactions in the Maumee and Sandusky river basins of northwestern Ohio. Journal of Environmental Quality. 31(1):47-53.

Coffey, R., R. Bowman, L. Anderson, J. Cocanougher, B. Crist, R. Fleming, K. Henken, J. Johnson, R. Mitchell, D. Overhults, T. Pescatore, M. Rasnake, S. Samson, and B. Thom. 1999. Assessment for the Potential of Livestock and Poultry Manure to Provide the Nutrients Removed by Crops and Forages in Kentucky.

Delgado, J.A. 1999. NLEAP Simulation of Soil Type Effects on Residual Soil NO_3-N in the San Luis Valley and Potential Use for Precision Agriculture. pg. 1367-1378. In: Proceedings of the 4th International Conference on Precision Agriculture. Precision Agriculture Center, Department of Soil, Water, and Climate, Univ. of Minnesota. St. Paul, MN.

Delgado, J.A. 2001. Use of Simulations for Evaluation of Best Management Practices on Irrigated Cropping Systems. pg. 355-381. In: M.J. Shaffer. L. Ma. And S. Hansen (eds.) Modeling Carbon and Nitrogen Dynamics for Soil Management. Lewis Publishers, Boca Raton, Florida.

Economic Research Service website. www.ers.usda.gov. 2006.

Environmental Protection Agency website. www.epa.gov. 2006.

Environmental Working Group website. 2006. www.ewg.org.

Foreign Agricultural Trade of the United States (FATUS) website. 2006. http://www.ers.usda.gov/data/fatus/.

Goddard, T.W. 2005. An Overview of Precision Conservation in Canada. Journal of Soil and Water Conservation. 60(6):456-461

Kentucky Department of Agriculture—Kentucky Agricultural Statistics Service, 1997-1998 Kentucky Agricultural Statistics.

Kentucky Division of Water website. 2006. www.water.ky.gov.

Kentucky Geological Survey website. 2006. http://www.uky.edu/KGS/.

Kentucky River Authority website. 2006. http://finance.ky.gov/ourcabinet/attached+agencies/kra.htm

Kentucky River Basin Assessment Report. 2000. http://www.uky.edu/WaterResources/Watershed/KRB_AR/INDEX.htm.

Kentucky River Basin Assessment Report. 2000. Background Information for Intrepretation of Watershed Watch Parameters.

Kentucky River Basin Management Plan. 2002.

Kentucky River Basin Watershed Page. 2004. http://www.uky.edu/WaterResources/Watershed/.

Kentucky River Watershed Watch website. 20006. http://www.uky.edu/OtherOrgs/KRWW/Welcome.html.

Kitchen, N.R., K.A. Sudduth, D.B. Myers, R.E. Massey, E.J. Sadler, R.N. Lerch, J.W. Hummel, and H.L. Palm. 2005. Development of a Conservation-Oriented Precision Agriculture System: Crop Production Assessment and Plan Implementation. Journal of Soil and Water Conservation. 60(6):421-430.

Lerch, R.N., N.R. Kitchen, R.J. Kremer, W.W. Donald, E.E. Alberts, E.J. Sadler, K.A. Sudduth, D.B. Meyers, and F. Ghidey. 2005. Development of a Conservation-Oriented Precision Agriculture System: Water and Soil Quality Assessment. Journal of Soil and Water Conservation. 60(6):411-421.

Langley, Meghan. 2004. Assessing the Impact of Land Use Practices on Water Quality in the Kentucky River Drainage Basin.

Moog, D.B., and P.J. Whiting. 2002a. Climatic and Agricultural Contributions to Changing Loads in Two Watersheds in Ohio. Journal of Environmental Quality. 31:83-89.

Moog, D.B., and P.J. Whiting. 2002. Climatic and Agricultural Factors in Nutrient Export from Two Watersheds in Ohio. Journal of Environmental Quality. 31:72-83.

Mueller, T.G., H. Cetin, R.A. Fleming, C.R. Dillion, A.D. Karathanasis, and S.A. Shearer. 2005. Erosion Probability Maps: Calibrating Precision Agriculture Data with Soil Surveys Using Logistic Regression. Journal of Soil and Water Conservation. 57(6):462-468.

Neighborhood Change Database (NCDB) 1970-2000 Tract Data CD

Natural Resource Conservation Service website. 2006. www.nrcs.usda.gov.

Ormsbee, L., and McAlister, M. 2002. Summary Report 2002 Kentucky River Watershed Watch Data Collection Effort. The Kentucky River Watershed Watch.

Renschler, C.S. and T. Lee. 200.5 Spatially Distributed Assessment of Short and Long-Term Impacts of Multiple Best Management Practices in Agricultural Watersheds. Journal of Soil and Water Conservation. 60(6):446-456.

Revenga, Carmen., and Mock, Greg. 2000. Dirty Water: Pollution Problems Persist. Pilot Analysis of Global Ecosystems: Freshwater Systems. earthtrends.wri.org.

Ribaudo, Marc O., Horan, Richard D., and Smith, Mark E. 1999. Current Water Quality Conditions and Government Programs To Protect Water Quality. pg. 3-20.

Sadler, E.J., R.G. Evans, K.C. Stone, and C.R. Camp. 2005. Opportunities for Conservation with Precision Irrigation. Journal of Soil and Water Conservation. 60(6) 371-378:

Schumacher, J.A., T.C. Kaspar, J.C. Ritchie, T.E. Schumacher, D.L. Karlen, E.R. Ventris, G.M. McCarty, T.S. Colvin, D.B. Jaynes, M.J. Lindstrom, and T.E. Fenton. 2005. Identifying Spatial Patterns of Erosion for Use in Precision Conservation. Journal of Soil and Water Conservation. 60(6):355-362.

Tegtmeier, Erin M. and Duffy, Michael D,. 2004. External Costs of Agricultural Production in the United States. International Journal of Agricultural Sustainability. pg. 1-20.

Tong, Susanna., and Chen, Wenli. 2002. Modeling the Relationship Between Land Use and Surface Water Quality. Journal of Environmental Management. 66:377-393.

United States Department of Agriculture website. www.usda.gov. 2006.

United States Geological Survey website. 2006. www.usgs.gov.

University of Kentucky College of Agriculture, Agricultural Weather Center website. 2006. http://wwwagwx.ca.uky.edu/.

Vita

James Allen Wermuth

Born: August 26, 1982 in Cincinnati, OH

Education
Bachelor of Science in Agricultural Economics, Emphasis in Natural Resource Economics. Minor in Business. May 2004.

Experience
Research Assistant. University of Kentucky. May 2005-December 2006.
Teaching Assistant. University of Kentucky. August 2006-December 2006.

Scholastic and professional honors
Gamma Sigma Delta Member. 2006.
The Chancellor's List. 2006.