DEVELOPMENT AND APPLICATION OF THE SPATIALLY EXPLICIT LOAD

ENRICHMENT CALCULATION TOOL (SELECT) TO DETERMINE

POTENTIAL *E. coli* LOADS IN WATERSHEDS

A Thesis

by

KENDRA JEAN RIEBSCHLEAGER

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

MASTER OF SCIENCE

August 2008

Major Subject: Biological and Agricultural Engineering

DEVELOPMENT AND APPLICATION OF THE SPATIALLY EXPLICIT LOAD

ENRICHMENT CALCULATION TOOL (SELECT) TO DETERMINE

POTENTIAL *E. coli* LOADS IN WATERSHEDS

A Thesis

by

KENDRA JEAN RIEBSCHLEAGER

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

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Raghupathy Karthikeyan
Committee Members,	Bruce Lesikar
	Raghavan Srinivasan
	Ralph Wurbs
Head of Department,	Gary Riskowski

August 2008

Major Subject: Biological and Agricultural Engineering

ABSTRACT

Development and Application of the Spatially Explicit Load Enrichment Calculation
Tool (SELECT) to Determine Potential *E. coli* Loads in Watersheds. (August 2008)
Kendra Jean Riebschleager, B.S., Texas A&M University
Chair of Advisory Committee: Dr. Raghupathy Karthikeyan

According to the USEPA National Section 303(d) List Fact Sheet, bacterial pathogens
are the leading cause of water quality impairments in Texas. The automated Spatially
Explicit Load Enrichment Calculation Tool (SELECT) uses spatially variable factors
such as land use, soil condition, and distance to streams to characterize pathogen sources
across a watershed. The results support development of Total Maximum Daily Loads
(TMDLs) where bacterial contamination is of concern. SELECT calculates potential *E.
coli* loads by distributing the contributing source populations across suitable habitats,
applying a fecal production rate, and then aggregating the potential load to the
subwatersheds. SELECT provides a Graphical User Interface (GUI), developed in
Visual Basic for Applications (VBA) within ArcGIS 9.X, where project parameters can
be adjusted for various pollutant loading scenarios.

A new approach for characterizing *E. coli* loads resulting from on-site wastewater
treatment systems (OWTSs) was incorporated into the SELECT methodology. The
pollutant connectivity factor (PCF) module was created to identify areas potentially
contributing *E. coli* loads to waterbodies during runoff events by weighting the influence
of potential loading, runoff potential, and travel distance.

Simulation results indicate livestock and wildlife are potentially contributing large
amounts of *E. coli* in the Lake Granbury Watershed in areas where these contributing
sources are not currently monitored for *E. coli*. The bacterial water quality violations

iv

near Lake Granbury are most likely the result of malfunctioning OWTSs and pet waste in the runoff.

The automated SELECT was verified by characterizing the potential *E. coli* loading in the Plum Creek Watershed and comparing to results from a prior study (Teague, 2007). The *E. coli* potential load for the watershed was lower than the previous study due to major differences in assumptions. Comparing the average ranked PCF estimated by physical properties of the watershed with the statistical clustering of watershed characteristics provided similar groupings.

SELECT supports the need to evaluate each contributing source separately to effectively allocate site specific best management practices (BMPs). This approach can be used as a screening step for determining areas where detailed investigation is merited. SELECT in conjunction with PCF and clustering analysis can assist decision makers develop Watershed Protection Plans (WPPs) and determine TMDLs.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

viii

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1. Introduction

Bacterial pathogens (fecal coliform and *Escherichia coli (E. coli)*) are the leading cause
of water quality impairments in the United States (USEPA, 2008). Pathogens are also the
principal cause of impairments in Texas. The Total Maximum Daily Load (TMDL)
program, mandated by the Clean Water Act (CWA) Section 303, is a process to develop
pollutant specific management plans integrating water quality assessment for protection
of impaired watersheds. A stream segment is classified as impaired due to pathogens if
25% of its samples exceed 394 cfu/100mL or if the geometric mean of the samples
exceeds 126 cfu/100mL (TCEQ, 2000) for the indicator organism *Escherichia coli* (*E.
coli*). The goal of the CWA is to restore and maintain the chemical, physical, and
biological integrity of the nation's waters. To meet the criteria of these mandates, models
are often developed to study the current status of water quality and the impacts of
various management plans (Chen et al., 1999; Zeckoski et al., 2005).

A representative watershed-scale water quality model is needed to address bacterial
pollution (primarily fecal coliform and *E. coli*) issues and help decision makers evaluate
contamination problems and determine the appropriate course of action. A
comprehensive model will: i) characterize the production and distribution of waste and
the associated microorganisms; ii) simulate transport of microorganisms from land
surfaces to streams; and iii) route microorganisms through existing stream networks
(Jamieson et al., 2004).

In the United States alone, the estimated cost of waterborne illness ranges from $269 to
$806 million for medical costs and $40 to $107 million in lost work and productivity

---

This thesis follows the style of Environmental Modelling & Software.

(Payment and Hunter, 2001).When a waterbody is impaired, it impacts the local economy due to loss of the designated use, whether as drinking water supply or recreational activities. The cost of TMDL development and implementation may average $1 million per impaired watershed (Houck, 1997). Spatially explicit analysis is needed to investigate the location of the sources of a specific contaminant. By spatially referencing *E. coli* sources, the potential load resulting from different locations in a watershed can be determined. The load distribution throughout the watershed can then be combined with a watershed model to determine the amount of *E. coli* that will be transported by surface runoff to the stream. Using this information, best management practices (BMPs) can specifically target areas and prominent sources that significantly contribute to stream contamination. As an automated tool within ArcGIS, SELECT can be used by stakeholders as a preliminary assessment tool for a Watershed Protection Plan (WPP) or in TMDL development. This tool is flexible and can be coupled with any hydrologic modeling program.

According to a clustering analysis study by Paul et al. (2006), understanding the characteristics that are closely related to contaminants and their sources can aid in the decision making process for bacterial TMDLs by developing approaches based on the watershed characteristics of greatest influence. Typical TMDL development approaches use watershed models such as SWAT and HSPF or the load-duration curve method. The load-duration method for developing TMDLs provides an overall representation of water quality and the needed improvement (Cleland, 2002). However, intra-watershed contributions must be determined through further sampling or modeling that relates hydrologic response and land use to bacterial concentrations in waterbodies. Hydrologic simulation models are needed to predict the fate and transport of non-point source pollutants in watersheds since overland flow is the primary driving force. Watershed models allow the user to describe a watershed and the expected pollutant loading both spatially and temporally. The main drawbacks of most hydrologic simulation models are data intensive, user input, and cumbersome calibration. In addition, these models are

limited in their ability to describe life cycles of bacteria adequately or to simulate concentrations during extreme climate conditions (Benham et al., 2006). These models often require a total bacteria load for each subwatershed; however, these models do not have a comprehensive bacteria load assessment tool. Characterization of non-point source pollutants is inherently difficult, and research is needed to develop a tool that can spatially and temporally distribute non-point source pollutants more efficiently. Once this goal is achieved, the fate and transport processes can more accurately be modeled (Benham et al., 2006).

Improved Risk Assessment for Water Distribution Systems (IRA-WDS) is a GIS-based risk analysis tool for evaluating the risk of deteriorating water distribution systems developed using C^{++} and ArcView 3.2. The software was integrated into GIS using a tight-coupling approach. A graphical user interface (GUI) enables data for the models to be retrieved through dynamic-link library (DLL) calling functions within GIS. The outputs are converted into tables and thematic maps displaying the risk of contaminant intrusion into the water distribution system and describe the main factors that contribute to the risk. This software is helpful in understanding the model structure for use in risk assessment (Vairavamoorthy et al., 2007). This approach allows user flexibility and could be modified for the development of other risk assessment tools tightly coupled within GIS using a GUI.

Another approach for identifying risk 'hot spots' uses a screening model (Mitchell, 2005). This screening model was used to support sustainable urban drainage planning by mapping hazard from urban non-point source pollution. This approach uses a semi-distributed stochastic GIS-model to map basin-wide loadings of various stormwater pollutants in small basins. Information on surface water quality objectives were combined with pollutant load maps which identify diffuse emission 'hot spots'. This tool can be used in planning and strategic management of urban pollution (Mitchell, 2005).

The identification of 'hot spots' can be incorporated into other modeling approaches for further evaluation and investigation of BMPs.

The Italian Environmental Protection Agency has developed the potential non-point pollution index (PNPI), a GIS-based watershed-scale tool (Munafo et al., 2005). PNPI is a simple method designed to inform decision makers about the potential environmental impacts of different land management scenarios. This tool identifies and displays areas that are likely to produce pollution due to their land use, geo-morphology, and location with respect to the stream network. This approach uses expert knowledge to generalize the relationship between the land cover indicator (LCI), run-off indicator (ROI), and the distance indicator (DI). This is a new approach which focuses on the driving forces of pollution instead of the impacts (Munafo et al., 2005). The use of a simplified qualitative approach similar to the PNPI can aid the initial stages of TMDL development by concentrating efforts in the appropriate locations within the watershed as well as addressing the appropriate sources.

The Center for TMDL and Watershed Studies at Virginia Tech has developed a software tool, the Bacteria Source Load Calculator (BSLC), to support the bacterial source characterization process of the TMDL and automate the creation of input files for water quality modeling (Zeckoski et al., 2005). The BSLC uses a systematic process that includes inventorying bacterial sources, estimating loads generated from these sources, distributing estimated loads to the land as a function of land use and source type, and generating bacterial load input parameters for watershed scale simulation models for source characterization. The BSLC uses externally generated loadings over land and hourly bacterial stream loadings. This program was developed using Visual Basic for Applications (VBA) in Microsoft Excel (Zeckoski et al., 2005). This loosely-coupled model will become a spatially distributed one only if tied to a GIS-based model. To spatially distribute the loadings, a watershed is divided into subwatersheds and source populations are assigned to each subwatershed. However, within each subwatershed the

loads are not spatially allocated in the BSLC application. In addition, the data for source populations are often available by county, not by subwatersheds. Consequently, the user has to redistribute the data to calculate bacteria loadings on a subwatershed basis. The Hydrologic Simulation Program in Fortran (HPSF) was used with the BSLC tool to simulate accumulation and die off of *E. coli* (Zeckoski et al., 2005). This model does not provide maps, charts, or any other visual aid for decision making in the TMDL process.

The US Environmental Protection Agency (EPA) developed the Bacterial Indicator Tool (BIT) as a spreadsheet that estimates the bacteria contribution from multiple sources (USEPA, 2000). Output from this tool is used as input to WinHSPF and the HSPF water quality model within BASINS. This tool calculates monthly accumulation of fecal coliform resulting from four land uses: cropland, forest, built-up, and pastureland; on up to 10 subwatersheds. The user inputs land use, agricultural animals, and wildlife density distribution for each subwatershed, the number of on-site wastewater treatment systems (OWTSs), number of people on OWTS, and malfunction rate of OWTS systems for the study area (US EPA, 2000). This approach does not account for spatial variability within subwatersheds.

A spatially explicit watershed-scale water quality model focused on pathogen contamination (primarily fecal coliform and *E. coli*) issues is needed. Understanding the governing factors closely related to fate and transport of *E. coli* and contributing sources can aid in the decision making process for bacterial TMDLs by developing approaches based on the watershed characteristics. It is essential to develop a model that can spatially and temporally distribute non-point source pollutants more effectively. Only after accurate characterization of potential pathogen loading, fate and transport processes can be simulated accurately.

Geographic Information Systems (GIS) can aid in the difficult task of characterizing non-point source pollution in a watershed. Teague, 2007 developed the Spatially Explicit

Load Enrichment Calculation Tool (SELECT) utilizing GIS to assist in the source characterization component of the TMDL process and within Watershed Protection Plans (WPPs) where bacterial contamination is a concern. This approach was not automated. Automation of SELECT will reduce processing errors and analysis time as well as offer program flexibility. This tool in conjunction with a fate and transport watershed model can be used to determine the actual bacterial loads resulting in streams.

1.2. Research Objectives

The major objective of the research was to develop the pathogen load assessment component of a watershed-scale water quality model using spatially variable governing factors such as land use, soil condition, and distance to streams. Associated with this objective was a thorough characterization of the production and distribution of waste and the associated pathogens.

To achieve the research objectives, an automated tool was developed to spatially identify and assess pathogen sources in a watershed. This spatially explicit tool characterizes non-point pollutant sources such as wildlife, livestock, on-site sewage facilities (OSSFs), and pets as well as point sources from Wastewater Treatment Plants (WWTPs). Also, a Pollutant Connectivity Factor (PCF) component was developed based on indicative factors for contamination which include potential loading, runoff potential, and distance to waterbodies. Simulation results using SELECT along with PCF component provide stakeholders and decision makers 'hot spots' in a watershed vulnerable to bacterial contamination without using a complex water quality model. This approach can also be used to identify the appropriate locations of water quality monitoring stations in a watershed.

The following two chapters present the results of this research and are written according to corresponding journal format. The last chapter presents the conclusions of this study.

CHAPTER II

AUTOMATION OF THE SPATIALLY EXPLICIT LOAD ENRICHMENT

CALCULATION TOOL (SELECT) TO CALCULATE POTENTIAL *E. coli*

LOADS IN LAKE GRANBURY WATERSHED, TEXAS

The Spatially Explicit Load Enrichment Calculation Tool (SELECT) was automated to characterize waste and the associated pathogens from various sources within a watershed. SELECT assesses pathogen loads in a watershed using spatially variable governing factors such as land use, soil condition, and distance to streams to support development of Total Maximum Daily Loads (TMDLs) where bacterial contamination is a concern. A new approach for characterizing *E. coli* loads resulting from malfunctioning On-site Wastewater Treatment Systems (OWTSs) was incorporated into the spatially explicit load enrichment calculation methodology along with the pollutant connectivity factor (PCF) module. The PCF component will help identify areas contributing *E. coli* loads during runoff events by weighting the influence of potential *E. coli* loading, runoff potential, and travel distance to waterbodies. SELECT results prove the need to evaluate each contributing source separately to effectively allocate site specific best management practices (BMPs). It also serves as a powerful screening tool for determining areas where detailed investigation is merited. Simulation results indicated livestock and wildlife are potential *E. coli* contributing sources in the Lake Granbury Watershed. The areas in which these sources are contributing are not currently monitored for *E. coli*. The bacterial water quality violations seen around Lake Granbury are most likely the result of malfunctioning OWTSs and pet wastes.

2.1. Introduction

Bacterial pathogens (fecal coliform and *Escherichia coli* (*E. coli*)) are the leading cause of water quality impairments in the United States (USEPA, 2008). The Total Maximum Daily Load (TMDL) program, mandated by the Clean Water Act (CWA) Section 303, is a process to develop pollutant specific management plans integrating water quality

assessment for protection of impaired watersheds. The goal of the CWA is to restore and maintain the chemical, physical, and biological integrity of the nation's waters. To meet the criteria of these mandates, models are often developed to study the current status of water quality and the impacts of various management plans (Borah and Bera, 2004). The Soil and Water Assessment Tool (SWAT) and Hydrologic Simulation Program Fortran (HSPF) are both watershed hydrologic simulation models used for evaluating Best Management Practices (BMPs) and characterizing pollutant sources. For example, SWAT was used to evaluate the effects of BMPs related to dairy manure management and municipal waste-water treatment plant effluent for TMDL development in the North Bosque River Watershed in Texas (Santhi et al., 2001). The United States Geological Survey (USGS) used the Hydrologic Simulation Program Fortran (HSPF) with bacterial source tracking as part of the source characterization component for the TMDL program for Accotink Creek, Fairfax County, Virginia (Moyer and Hyer, 2003).

Others (e.g. Chen et al., 1999; Zeckoski et al., 2005) have developed new models specifically for calculating TMDLs. Chen et al. (1999) developed a decision support system for calculating TMDLs that employs stakeholder involvement along with watershed models. The decision support system includes its own watershed simulation model, database, consensus building module, and a TMDL module with a calculation worksheet. The system generates various combinations of waste load and non-point load allocations to meet the water quality criteria. Ferguson et al. (2003) provide a review of problems with modeling fate and transport of indicator bacteria and highlight that more research is needed to determine the reaction of *E. coli* in aquatic environments. Benham et al. (2006) have modeled bacteria fate and transport to support TMDL development.

A popular approach currently in the TMDL development process to identify whether point or non-point sources are contributing to contamination is the Load Duration Curve (LDC) method (USEPA, 2006a). The LDC approach was developed for assessing nutrient loading in streams (Cleland, 2002). This graphical approach combines daily

stream flow with water quality data for the pollutant to be evaluated. It is assumed that point sources are a constant loading that are present during all flow regimes where as non-point source loadings are present in streams during high flows due to runoff events (Cleland, 2003). This approach relies solely on the field data available, thus the LDC determines load reductions for the flow conditions at which measurements were taken (Li and Guo, 2003). Load reduction is based on the percent exceedance above the maximum allowable load line. In Texas, the water quality standards enforced when using *E. coli* as the pathogen indicator are i) geometric mean concentration of 126 cfu/100mL or ii) 394 cfu/100mL for grab samples (TCEQ, 2000). Based on these load reductions, best management practices (BMPs) are applied broadly across the entire watershed for all sources. The only break down in percent reduction is based on whether the source is point or non-point and the percent of time the stream is exceeding the standard during various flow ranges. The New Jersey Department for Environmental Protection (NJDEP) used the LDC approach to determine fecal coliform load reductions for the Whippany River (Al-Ebus and Jacobson, 2002). In comparison, a watershed model computes loading across all flow regimes and uses the field data for calibration (Li and Guo, 2003).

The Center for TMDL and Watershed Studies at Virginia Tech has developed a software tool, the Bacteria Source Load Calculator (BSLC), to support the bacterial source characterization process of the TMDL and automate the creation of input files for water quality modeling (Zeckoski et al., 2005). The BSLC uses a systematic process that includes inventorying bacterial sources, estimating loads from these sources, distributing estimated loads across the landscape as a function of land use and source type, and generating bacterial load input parameters for watershed scale simulation models for source characterization. The BSLC uses externally generated watershed loadings and hourly bacterial stream loadings. This program was developed using Visual Basic for Applications (VBA) in Microsoft Excel (Zeckoski et al., 2005). This loosely-coupled model will become a spatially distributed only if tied to a GIS-based model. To spatially

distribute the loadings, the watershed is divided into subwatersheds and source populations are assigned to each subwatershed. However, within each subwatershed the loads are not spatially allocated in the BSLC application. In addition, the data for source populations are often available by county, not by subwatersheds. Consequently, the user has to redistribute the data on a subwatershed basis. The Hydrologic Simulation Program in Fortran (HPSF) is used with the BSLC tool to simulate accumulation and die off of *E. coli* (Zeckoski et al., 2005). This model does not provide maps, charts, or any other visual aid for decision making in the TMDL process.

The Italian Environmental Protection Agency has developed the potential non-point pollution index (PNPI), a GIS-based watershed-scale tool (Munafo et al., 2005). PNPI is a simple method designed to inform decision makers about the potential environmental impacts of different land management scenarios. This tool helps the user detect and display areas that are likely to produce pollution due to their land use, geo-morphology, and location with respect to the stream network. This approach uses expert knowledge to generalize the relationship between the land cover indicator (LCI), run-off indicator (ROI), and the distance indicator (DI) to study the driving forces of pollution instead of impacts (Munafo et al., 2005).

A representative watershed-scale water quality model is needed to address microbial pollution (primarily fecal coliform and *E. coli*) issues. A comprehensive model can aid decision makers evaluate multifaceted problems and determine the appropriate course of action (Jamieson et al., 2004). Geographic Information Systems (GIS) can aid in the difficult task of characterizing non-point source pollution in a watershed. A spatial semi-qualitative approach can aid the initial stages of TMDL development by concentrating efforts in the appropriate locations within the watershed as well as addressing the appropriate sources. The Spatially Explicit Load Enrichment Calculation Tool (SELECT) methodology was developed to assist in the source characterization component of the TMDL development process and Watershed Protection Plans (WPPs)

where bacterial contamination is a concern (Teague, 2007). The SELECT is a pathogen load assessment component of a watershed-scale water quality model using spatially variable governing factors such as land use, soil condition, and distance to streams to support TMDLs and WPPs. This tool can be used to determine the actual contaminant loads resulting in streams when used in conjunction with a fate and transport watershed model. SELECT can simulate potential pathogen loading in a watershed for various management scenarios. Application of SELECT will help stakeholders identify the areas potentially contributing to pathogen contamination of waterbodies without using complex hydrologic models. An additional pollutant connectivity factor (PCF) component was developed based on three indicative factors for contamination: a) potential pollutant loading, b) runoff potential, and c) travel distance to streams and other waterbodies. The PCF component of SELECT offers stakeholders a less expensive, less time consuming, and easier approach for evaluating best management practices.

SELECT provides a Graphical User Interface (GUI), developed in Visual Basic for Applications (VBA) within ArcGIS 9.X, where project parameters can be adjusted for various pollutant loading scenarios. From the visual output of the program a decision maker or stakeholder can identify areas of greatest concern for contamination contribution and incorporate that information while developing the WPP or the TMDL development. Details of the model development and results of applying SELECT to the Lake Granbury Watershed in Texas are presented in this paper.

2.2. Methodology

The approach for characterizing the *E. coli* sources is similar to the methodology developed by Teague (2007), with the exception of on-site sewage facilities (OWTS systems). The approach outlined here for SELECT has been expanded, revised, and automated for extending its application to diverse watersheds.

2.2.1. Spatially Explicit Approach

To develop a spatially distributed approach for bacterial TMDL development spatially variable factors that have the greatest influence on impairment should first be identified. This primarily comes from common logic, literature from other TMDL assessments, and communication with agricultural and wildlife experts as well as stakeholders. Land use is the factor that has the greatest effect on potential *E. coli* loading because the type of land use / land cover dictates whether the area is suitable for pollutant contribution. For example, it can be assumed that cattle will be confined to pasture and grazing lands and will not be found in cultivated cropland or residential neighborhoods. Quantifying the extent of influence from variables such as soil type and distance to streams is not always accurate and assumptions need to be based on the best knowledge available.

To characterize the production and distribution of waste and associated pathogens, sources which are contributing to contamination should be determined. This can be achieved by looking at the agricultural census information provided by National Agriculture Statistics Service (NASS), talking to the local extension agents and wildlife experts, obtaining permitted Waste-Water Treatment Plants discharges from the EPA Envirofacts Data Warehouse, and researching previous pathogen TMDLs. The fecal production rates for the various sources can be calculated using the EPA Protocol for Developing Pathogen TMDLs (USEPA, 2001) which includes a summary of source-specific pathogen and fecal indicator concentrations.

To integrate SELECT into a hydrologic simulation model, the potential loading on a daily time scale is needed. This is achieved by estimating the source populations, distributing the sources uniformly across suitable habitats, applying fecal production rates, and then aggregating to the level of interest (often the subwatersheds) for analysis.

2.2.2. GIS Modeling Framework

The development of the automated tool started with using the Model Builder application within GIS to conceptualize the file processing and determining appropriate input parameters for each type of source assessment (livestock, wildlife, on-site wastewater treatment systems, pets, and waste-water treatment plants). A Graphical User Interface (GUI) was developed in Visual Basic for Application (VBA) to create a tightly-coupled model within ArcGIS 9.X. The GUI was used to create the watershed project setup, add layers to the map, and input parameters such as appropriate habitats, source populations, and fecal production rates. The next step was to process the spatial files using the inputs from the GUI. The map processing code was written using ArcObjects relationship classes and divided into several modules.

A central module processes information from the GUI and then initializes the appropriate subroutines within the various modules in an ordered sequence of events. The remaining modules contain subroutines for determining the potential loading from both point (Waste-Water Treatment Plants (WWTPs)) and non-point (livestock, wildlife, and domestic) sources. The livestock module has separate subroutines for cattle, dairy, sheep/goats, horses, and swine. The wildlife module calculates potential loading for deer, feral hogs, and two generic (Other1 and Other2) sources. Subroutines for on-site wastewater treatment systems (OWTS) and pets are part of the domestic module. The urban module has a subroutine for calculating *E. coli* contributions from WWTPs. Lastly; the pollutant connectivity module is a set of subroutines for weighting the driving forces of pollutant contributions reaching waterbodies to create the Pollutant Connectivity Factor (PCF).

2.2.3. Watershed Description

Lake Granbury is a man-made lake within the Middle Brazos-Palo Pinto watershed. The Lake Granbury Watershed was delineated into 34 subwatersheds (Figure 2.1) using ArcSWAT (SWAT, 2005). This watershed is maintained by the Brazos River Authority

(BRA). The city of Granbury is located in north-central Texas approximately 20 miles southwest of Fort Worth, Texas. This is a diverse watershed characterized by multiple landuse classifications (Figure 2.2). This lake is used for recreation and is a water source for municipalities, industries, and agriculture. This popular area is rapidly growing with a large number of people populating the areas around the lake.

Lake Granbury is currently under scrutiny for its rising levels of bacteria within the coves of the lake. According to a recent water quality study (Espey Consultants, Inc., 2007) there are four coves nearing bacteria impairments and one already impaired. In addition, four coves exceed the dissolved oxygen standard, eight exceed the chloride standard, and one is approaching the nitrogen screening level. Currently, the main body of the lake is not impaired due to bacteria, but if conditions continue to worsen in the coves it is possible the lake, a drinking water for over 250,000 people and 15 cities (TWRI, 2007), could be contaminated. There are few centralized sewage systems and new residential areas have OWTSs near the coves of the lake. Unfortunately, much of the soil around the lake is not suitable for traditional septic tank and gravity trench soil treatment areas. The BRA plans to work with the Texas Commission on Environmental Quality (TCEQ), local entities, and federal and state agencies to develop a Watershed Protection Plan (WPP) with a focus to reduce bacterial contamination. SELECT has been applied to assist in the characterization of sources and the potential loading of bacteria for the Lake Granbury Watershed.

2.2.4. SELECT Initiation

The first step for automating SELECT is to set the database structure by indicating the project directory, the desired model output, and appropriate analysis properties. Under the SELECT menu is the dropdown for Project Setup. Here the user can start a new project or edit the current project. The Project Setup user form is displayed with tabs for setting up the Project Directory, Sources to be evaluated, and the Project Properties for analysis. The sources SELECT evaluates include livestock, wildlife, domestic sources,

and waste-water treatment plants. The Project Properties form gives options for setting
the analysis cell size, number of counties in the study area, and the shapefile indicating
the delineation of interest of the watershed (zonal aggregation file). To select the file for
zonal aggregation, the file must already be in the map frame. The Add Data/Layers
button displays a dropdown menu for the minimal files necessary for the analysis such as
land use, county delineation, and urban areas. Other necessary files depend upon sources
to be evaluated. Table 2.1 lists the appropriate files and formats for SELECT input.

The pathogen sources selected for Lake Granbury were Beef Cattle, On-site Wastewater
Treatment Systems (OWTS) malfunction, Deer, and Wastewater Treatment Plants
(WWTPs). The conversion factor from fecal coliform to *E. coli* is set at the default value
of 0.5. The project properties such as analysis cell size of 30m x 30m, the file indicating
the level of aggregation (subwatershed delineation), and the number of counties in the
analysis area (for Granbury there are two, Hood and Parker) are fixed.

2.2.5. Model Simulation
The default fecal production rates are the highest from the range of values provided in
the EPA Protocol for Developing Pathogen TMDLs (USEPA, 2001) for all *E. coli*
sources in the Lake Granbury Watershed (Table 2.2).

2.2.5.1. Potential E. coli Sources in Lake Granbury Watershed
SELECT simulated potential *E. coli* load resulting from cattle, deer, pets, malfunctioning
OWTS, and Waste-Water Treatment Plants.

Livestock
All livestock populations (beef cattle, dairy cattle, sheep/goats, swine, and horses) were
determined from the National Agricultural Statistics Service (NASS) inventory on a per
county basis. The user chooses the appropriate land use classification (indicated within
the interface). The program clips the landuse file to create a landuse grid for each county

and then creates a raster from the indicated land use for each county and reclassifies the grid into suitable (value of 1) and non-suitable (0). Next the population density grid is created by multiplying the suitable habitat grid times the population and divided by the number of cells. The population density grids for each county are combined using the mosaic operation into one population density grid. Finally, the population density grid is multiplied by the fecal coliform production rate indicated in the user form and converted into an *E. coli* production rate using a conversion factor of 0.5 (Doyle and Erikson, 2006), an adjustable parameter in the project setup. Finally, a zonal sum is performed to aggregate the resultant load for each zone of interest.

The cattle populations for Hood and Parker counties were 30,059 and 71,601 cattle, respectively. The cattle population was distributed uniformly on grasslands (2001 NLCD Classification 71) and pasture/hay (NLCD Classification 81), since cattle graze mainly on these land uses.

Wildlife

SELECT attempts to account for wildlife contributions by distributing population estimates across suitable habitats as determined by consultation with wildlife experts. The first step in calculating wildlife pollutant loading is to identify the types of wildlife most likely contributing the most significant amounts of pollution and ignore the sources that only minimally contribute. This was achieved by consulting wildlife experts such as the Texas Parks and Wildlife Department (TPWD), thorough literature review, and applying expert knowledge. It is also important to identify the landuse types wildlife prefer/need for survival, along with population estimates. Many agencies such as the TPWD have published studies that address these issues. Currently, SELECT provides the option to evaluate pollutant loading of *E. coli* from deer, feral hogs, and two other generic sources. The program allows for two methods. In the first method the user inputs a suitable habitat shapefile and then the program assumes the wildlife will graze only in these areas. In the second approach the user indicates appropriate landuse and whether or

not to include urban areas, then the program determines the suitable habitat based on the indicated landuse and other assumed parameters (for deer at least 20 acres of contiguous terrain is available and a buffer from streams for feral hogs). Once the suitable habitat is created the program applies fecal production rates multiplied by the population density and then aggregates the total loading for the source to each zone of interest.

The population density of 13.25 deer/1000 acre is estimated from the Lockwood (2000) report. This report was a study the Texas Parks and Wildlife Department (TPWD) performed to track white tail deer populations. The deer population density was determined as the average of Resource Management Unit (RMU) 22 and RMU 24 for the Lake Granbury Watershed. It was assumed that deer roam in forest (41, 42, and 43) and shrubland (52). The model also assumes the deer need continuous suitable habitat of at least 20 acres. Urban areas were removed from the suitable habitat.

On-site Wastewater Treatment Systems (OWTSs)

Another need for bacteria load assessment is an improved understanding of when OWTSs malfunction, how much these systems contribute to contamination, and how to reasonably predict such occurrences. For evaluating the potential *E. coli* loading from malfunctioning OWTSs a new approach different from Teague (2007) was developed. Clark et al. (2001) indicated that the age of OWTS, soil condition, and vicinity to water bodies have the greatest influence on contamination due to OWTSs. Methods for developing a sewage pollution risk assessment have been developed and were used as a guideline (Kenway and Irvine, 2001). Combining this methodology for OWTS risk assessment with soil landscape mapping can assess the individual system contribution to the cumulative risk of sewage pollution (Chapman et al., 2004). Two methods for OWTS malfunction prediction have been created for the SELECT. The first method can be used when detailed OWTS permit information is available. The second method relies only on readily available public data sources. The primary function of SELECT is to provide a total potential *E. coli* loading before fate and transport mechanisms are

incorporated. Therefore, the distance component when predicting contribution from malfunctioning OWTSs is not included in the load assessment.

Method 1:

This method was developed based on the age of subdivisions and the OWTS absorption field limitation ratings (slight, moderate, and severe) provided with National Resource Conservation Service (NRCS) SSURGO soils data (USDA-NRCS, 2004). The user inputs the appropriate OWTS shapefile and indicates the 'fields' within the attribute table containing the number of permits and the average estimated age of the subdivision/OWTSs in each polygon. The number of systems contributing to potential is determined from the number of permitted homes on OWTSs multiplied by the expected percent malfunction. The percent malfunction is a reclassification of the OWTS suitability rating for a given area. The suitability rating is calculated as:

$$SuitabilityRating = 0.7 \times SoilRate + 0.3 \times AgeRate \qquad (1)$$

The NRCS provides limitation ratings based on geophysical factors such as soil classification, depth to bedrock, and slope (Table 2.3). The program creates an age rating for the OWTS shapefile (Table 2.4), and a soil rating based on the SSURGO soil limitation ratings of severely limited (3), somewhat limited (2), and slightly limited (1). The soil file with the suitability rating is intersected with the age rate and then weighted with 70% to soil rate and 30% to the age rating to create a new OWTS malfunction index. This weighting scheme is based on the assumption that soil treatment capability has the greatest role in contribution, followed by malfunction due to limited maintenance (related to age of system) (Lesikar, 2007). Areas missing soil or age information are assigned index ratings of -99. In this case the higher the suitability rating, the less effluent the system can treat. A malfunction index based on the suitability rating is converted to a raster file and then reclassified into percent malfunctioning (contributing to load potential) (Table 2.5). After determining the number of homes contributing, a flow rate (gal / person × day), effluent rate (cfu/100mL), the average population per

home, and necessary conversion factors are applied to estimate the potential *E. coli* loading in cfu/day.

Method 2:

The second method is conceptually similar to Method 1, however, using only publicly available information. To determine the number of OWTS without detailed permit information, the number is estimated using the U.S. Census Bureau census block shapefile with demographics and then creating a raster grid. Areas using municipal sewage were removed, determined from the Texas Commission on Environmental Quality (TCEQ) shapefile with Certificates of Convenience and Necessity service areas (CCN) (TCEQ, 2008a), by creating a 'not sewered' grid and then multiplying by the number of homes grid. The potential loading is then determined in the same manner as in Method 1 except the Suitability Rating is simply the SSURGO Soil Rate.

Method 1 for predicting OWTS *E. coli* contributions was applied to the Lake Granbury Watershed. OWTS information was obtained from county permit records (Hood County Appraisal District). The population density, 1.94 people per home, was estimated from the year 2000 Hood County Census (U. S. Census Bureau). SSURGO soil shapefiles for each county and the associated soil properties tables were obtained from the NRCS Soil Datamart.

Pets

Generally, dogs are the primary pet allowed to defecate outside the home and most often the defecated waste is not cleaned up. Cats and other pets are primarily kept in homes and waste disposed of directly to solid waste management so these contributions will be neglected. The assumption of a constant 0.8 dogs per home for Texas (AVMA, 2002) is an adjustable model parameter included in SELECT. The program creates a raster that represents the number of homes from the census block demographics table joined to the census block shapefile. Again the program applies the fecal production rate and then

aggregates the potential load to zones of interest. Census block shapefiles are needed for each county. The associated census block demographics table for the state of Texas is indicated in the GUI as well as the appropriate field for the number of homes in each census block.

Wastewater Treatment Plants (WWTPs)

To assess point sources SELECT evaluates the contribution from Wastewater Treatment Plants (WWTPs). Within the GUI, the user indicates the shapefile with the permitted outfall locations ensuring unrelated outfalls (i.e. cooling plants or any other non-pathogenic discharges) removed. The file should include permitted discharges in the units of millions of gallons per day (MGD) as a field within the shapefile. The default (adjustable within the GUI) value of 126 cfu/100mL effluent standard is assumed. The loading is calculated by simply multiplying the effluent by the discharge and applying conversion factors to determine the loading in cfu/day. For this study, wastewater outfall locations were obtained from TCEQ GIS files (TCEQ, 2008b). The permitted flows were obtained from the EPA Envirofacts Data Warehouse (USEPA, 2006b).

Once all individual source inputs are selected and fed into the model a summation from all sources is carried out. Thus, potential loading in a watershed, here Lake Granbury, are spatially distributed.

2.2.5.2. Pollutant Connectivity Module

The pollutant connectivity factor (PCF) was developed to weigh the influence of the driving forces of contamination with the total pollution present. The PCF indicates areas within the watershed vulnerable to contributing bacteria to waterbodies. This module determines the curve number, which directly relates to runoff potential, and the distance to streams, which directly relates to fate and transport. The total pollutant connectivity factor will be calculated using a weighted combination of the normalized potential loading, curve number grid, and the inverse of the normalized flow length to streams

(Figure 2.3). This will provide stakeholders and decision-makers useful information to implement mitigation efforts in areas of greatest concern for water quality impairment. The flow length is derived from a digital elevation model (DEM) using ArcHydro Tools within ArcGIS. The curve number grid is created from intersecting the SSURGO soils hydrologic soil grouping (HSG) and the NRCS 2001 land use classification and then using a NRCS Curve Number Lookup Table. The resulting PCF is a ranking of potential contribution from subwatershed without considering any detailed fate and transport processes in the watershed. The following is the weighted overlay expression for determining the pollutant connectivity factor (PCF):

$$PCF = W_P \times P_I + W_R \times R_I + W_D \times 1/D_I \qquad (2)$$

Where,

PCF = Pollutant Connectivity Factor

W_P = weighting factor for the pollutant indicator, P_I

P_I = pollutant indicator, normalized pollutant load on scale from 0 to 100

W_R = weighting factor for the runoff indicator, R_I

R_I = runoff indicator, curve number

W_D = weighting factor for the distance indicator, D_I, and

D_I = distance indicator, normalized flow length on scale from 0 to 100

Appropriate weighting should be based on best knowledge available or expert opinion. Alternatively, sensitivity of weighting factors can be determined by running multiple trials of the pollutant connectivity factor with various weighting schemes (Table 2.6). If a particular subwatershed consistently is determined to be a 'hot spot' for contributing to contamination, then it is likely this subwatershed is of great concern and should be more readily addressed. On the other hand if a particular watershed is consistently rated low, then this watershed should not be of concern when determining management practices.

2.3. Results and Discussion

Potential *E. coli* loadings from livestock, wildlife, and domestic sources in the Lake Granbury Watershed were calculated by SELECT. The loadings from the individual

sources were combined and aggregated on a subwatershed basis (Figure 2.4). The potential loading component of SELECT can help identify source contributions spatially distributed across the watershed. However, this is only a snapshot of the amount of *E. coli* present in the watershed. The Pollutant Connectivity Factor (PCF) applied weighting to important fate and transport factors such as runoff capabilities and travel distance to provide helpful information to determine whether *E. coli* from various sources potentially contaminate the waterbodies. This weighting scheme when based on expert knowledge (Figure 2.5a) provides a screening tool to indicate the areas of highest concern for *E. coli* contamination. For the Lake Granbury Watershed, PCF analyses was based on applying multiple weighting schemes and then ranking the subwatersheds (Figure 2.5b) for potential water quality problems due to bacteria. The results from SELECT and the PCF were compared with water quality data to help decision makers and stakeholders develop a spatially explicit WPP or determine TMDLs.

2.3.1. Daily Potential E. coli Loading in Lake Granbury Watershed

The potential *E. coli* loading can be broken into two classes for analyses; non-point (Figure 2.6) and point sources (Figure 2.7). For each of these classifications it is important to consider how potential loads can be compared to actual *E. coli* concentrations in waterbodies, as measured at water quality monitoring locations (Figure 2.8).

2.3.1.1. Non-Point Sources

High potential *E. coli* load resulting from cattle (Figure 2.6a) occurs in the northern-most subwatersheds 26 and 34 as well as in subwatersheds 14 and 30 (Figure 2.1). These subwatersheds have a landscape dominated by grasslands with a mixture of pasture/hay (Figure 2.2). The middle of the watershed has lower loads mainly due to higher human population. Subwatershed 14 is an area of potential concern due to its close proximity to the lake with highest *E. coli* potential load. Further analysis using the PCF was applied to verify this concern (Figure 2.5b). During a runoff event the highest ranked 'hot spots'

are the most likely to significantly contribute to contamination in the waterbodies. The same subwatersheds with high potential loads were determined to be the three highest ranked, by PCF, areas likely to be contributing to contamination in the waterbodies. The highest average PCF ranking was subwatershed 34. Water quality data could be used to verify the PCF results; however, the subwatersheds with high loading resulting from cattle are not monitored for *E. coli* concentrations (Figure 2.8).

The highest potential *E. coli* loading resulting from deer (Figure 2.6b) can be seen in the northern portions of the watershed where human population is less dense. The subwatersheds with the highest potential loading (6, 18, 23, 26, and 34, (Figure 2.1)) have large amounts of forest landuse. The second highest group of potential loading tends to have significant amounts of forests but these areas are more scattered and broken up by streams and intermixed with open range and grass lands. The southern half of the watershed generally has lower potential loads resulting from deer mainly due to the influence of higher human populations. When these loads are compared with the PCF ranking, again subwatersheds 26 and 34 are among the areas of high concern. Subwatersheds 6, 18, and 23 are in the middle range of PCF ranking (fourth through eighth). Unfortunately all of the subwatersheds with high loading resulting from deer are not monitored for *E. coli* concentrations.

Potential *E. coli* loading resulting from malfunctioning OWTSs (Figure 2.6c) was calculated for Hood County only where descriptive permit data was gathered to create a spatial subdivision OWTS file by the Brazos River Authority from the Hood County Appraisal District. This information has not been gathered for Parker County (Morgan, 2008). This does not pose a significant problem since the northern portion of the watershed in Parker County is much further from the waterbodies of concern. In addition, the only areas with significant populations are on the north-eastern edge of the watershed where the populations are quite dense and most likely on combined sewer networks. Method 2 for OWTS malfunction potential loading without detailed permit

information could be run to verify this assumption. Subwatersheds 1 and 3 are located across the main section of Lake Granbury and have the highest potential *E. coli* loads resulting from malfunctioning OWTSs. Subwatershed 1 is characterized by significant developed, low intensity landuse classification which generally includes single-family housing units. Subwatershed 3 includes significant developed, medium and high intensity, landuse which includes single-family housing units with higher percent impervious land cover and areas where people reside or work in high numbers. The second highest potential loading group is located west of the lake and characterized by residential development scattered amongst undeveloped grasslands, forests, and pastures. The areas potentially contributing significant *E. coli* loadings resulting from malfunctioning OWTSs range from a PCF ranking of three to ten. Water quality monitoring data for *E. coli* in subwatersheds 1 and 3 indicate several stations where from 23 to 43% of observations at these locations exceed the maximum concentration standard of 126 cfu/100mL (Figure 2.8).

The potential *E. coli* loading resulting from pets (Figure 2.6d) is highest in subwatershed 26 in the northern portion of the watershed, subwatershed 8 along the southeastern edge, and in subwatersheds 2 and 3 around Lake Granbury (Figure 2.1). This is explained by significant low and medium intensity developments within these subwatersheds. These are popular residential areas because of the lake in the southern portion of the watershed and the close proximity to the Fort Worth metropolitan area in the northeast. The PCF ranking incorporated driving forces of pollutant fate and transport. The subwatersheds with highest potential *E. coli* resulting from pets are ranked using the average PCF over several weighting schemes as first, fourth, eighth and tenth. The next highest subwatersheds have a PCF ranking ranging from fourth to tenth. As noted earlier, subwatershed 26 (Figure 2.1) is not currently monitored for *E. coli* contamination (Figure 2.8). Several water quality monitoring stations are located in subwatershed 8, but the data does not indicate significant violations in water quality due to *E. coli* (Figure

2.8). Again subwatersheds 1 and 3 do indicate high *E. coli* concentrations from 23 to 43% out of all observations.

2.3.1.2. Point Sources

There are seven wastewater treatment plant facilities operating within the watershed (Figure 2.7). The highest *E. coli* loading occurs in subwatershed 8 (Figure 2.1) on the south-eastern edge of the watershed. These facilities contribute large amounts of treated effluents and could impact the environment if improper/inefficient treatment of wastewater were to occur. When localities are considering consolidating on-site wastewater treatment systems into municipal sewage systems, the local officials should take into account the amount of pollutants, such as *E. coli* and nutrients, that would be discharged as a direct point source (with virtually zero travel time or attenuation).

2.3.1.3. Combined Loading from All Sources

The highest total *E. coli* loads (Figure 2.4) occur in subwatersheds 14, 26, 30, and 34 (Figure 2.1). Subwatersheds 30 and 34 have land uses appropriate for cattle and deer. Hence, it can be concluded that major *E. coli* contributors in these subwatersheds are cattle and deer. Subwatershed 14 is ranked as the third highest area of concern based on the PCF due to the combined effects of potentially higher loading from cattle and a potentially high load from deer and OWTSs. Subwatershed 26 has the greatest likelihood to contribute to bacterial contamination in waterbodies based on the PCF ranking. This particular subwatershed is characterized by grasslands, pastures, and forests in the majority of the region and with significant development on the northern edge. It can be concluded that the potential *E. coli* loading in this subwatershed with diverse landuse is a result of combined contributions from cattle, deer, and pets.

The SELECT results including the PCF analysis indicate that across the entire watershed cattle is the largest contributor to *E. coli* loading followed by deer, pets, OWTS, and then WWTPs (Figure 2.5b). Comparing the SELECT results with actual *E. coli*

concentrations measured at water quality monitoring stations (Figure 2.8) indicates that malfunctioning OWTS are potentially a major concern followed by pets. Currently, bacterial water quality is not monitored where SELECT predicts high potential *E. coli* loads in the Lake Granbury Watershed (Figure 2.4 and Figure 2.8).

2.3.2. Versatility of SELECT

When potential *E. coli* loads simulated by SELECT are combined with the PCF module, decision makers can identify *E. coli* sources and areas of potential concern in a watershed. This will ultimately help decision makers choose cost effective BMPs to alleviate contamination issues in an impaired watershed. Once BMPs have been chosen, PCF analysis can be performed in order to determine the spatially explicit locations to implement source specific BMPs. The PCF results can also be used to determine the locations for water quality monitoring. Ideally, these locations should be in potential *E. coli* contributing areas and in areas where BMPs have been implemented to measure the success of the *E. coli* load reductions.

The current approach for many WPPs target load reductions from all sources applied uniformly across the watershed. It is evident from the geographical representation provided by SELECT that this is not practical and enforcement of pollutant reduction should only be in areas of greatest concern and should address each source separately. This will save both time and money by effectively developing BMPs that will preserve vital water resources.

It is very possible that the water quality data will indicate a different scenario than the simulated loads using SELECT. In this case a more thorough investigation is imperative. It will be necessary to apply a more advanced hydrologic simulation model to route the pollutants through the watershed to more accurately predict pollutant loads reaching the waterbodies. Differences between water quality data and the simulated results could also be an indication that input data into SELECT should be adjusted. For example, if the

WWTPs are not treating effluent properly or are discharging pollutants more than the permitted concentration, this actual amount should be used in SELECT simulations. Unfortunately this data is not readily available. For the Lake Granbury Watershed, most of the high *E. coli* measurements were taken on days of significant precipitation or immediately preceding the day of measurement (BRA, 2008; NCDC, 2008). There are a few incidences where high *E. coli* concentrations were measured at water quality monitoring locations with no recent precipitation events (BRA, 2008; NCDC, 2008). This indicates that point source discharges either from WWTPs or illicit discharges were causing *E. coli* contamination on these days.

Bacteria loading in a watershed can have seasonal variability due to migratory patterns of wildlife and grazing rotations for livestock. SELECT can easily simulate this temporal variability of *E. coli* with appropriate assumptions and input data. The simulated potential *E. coli* loads can be fed into a comprehensive water quality model to predict *E. coli* loads at different spatial scales. This is important because some hydrologic simulation models use subwatersheds while others such as SWAT use hydrologic response units (HRUs). This tool is able to integrate into a wide range of simulation models. Also, the SELECT approach can be used to determine potential loads of other contaminants such as nutrients by using appropriate source inputs.

The benefit of the automated SELECT is its ability to generate various scenarios to simulate potential contaminant loads with minimizing the errors inherent in manual approaches. The automated approach takes about five minutes to incorporate input files and parameters and 20 minutes to do the simulations for a watershed of 1100 km^2 evaluating five contaminant sources. Prior to the initial application some preprocessing of data is necessary, and then subsequent simulations are simple and fast.

2.4. Conclusions

The Spatially Explicit Load Enrichment Calculation Tool (SELECT) was developed and automated to characterize the production of pathogens from various pollutant sources across a watershed. SELECT was applied to the Lake Granbury Watershed in Texas. Based on simulation results for Lake Granbury, Best Management Practices (BMPs) are recommended to decrease *E. coli* loads from pets and OWTSs near the lake. Further investigation using watershed-scale water quality models such as SWAT or HSPF is needed to determine the influence of various *E. coli* sources across the watershed. Travel time from the subwatersheds with high potential loading should be determined to characterize the amount of *E. coli* reaching the waterbodies after a rainfall event. It is also recommended that water quality monitoring should be carried out in northern and western portions of the Lake Granbury watershed to monitor *E. coli* concentrations in the watershed. This will ultimately help in protecting Lake Granbury from contamination due to pathogenic bacteria.

SELECT is a user-friendly tool to conduct spatial analysis under different land use scenarios. In addition to this, maps and tables resulting from SELECT can be used for technical and educational communication. This approach proves the need to evaluate each contaminant source separately to effectively allocate site specific BMPs and serves as a powerful screening tool for determining areas where detailed investigation is merited.

CHAPTER III

**PREDICTING POTENTIAL *E. coli* LOADS IN PLUM CREEK WATERSHED,
TEXAS: USING THE AUTOMATED SPATIALLY EXPLICIT LOAD
ENRICHMENT CALCULATION TOOL (SELECT)**

3.1. Introduction

Bacterial pathogens (fecal coliform and *Escherichia coli* (*E. coli*)) are the leading cause of water quality impairments in the United States (USEPA, 2008). The Total Maximum Daily Load (TMDL) program, mandated by the Clean Water Act (CWA) Section 303, is a process to develop pollutant specific management plans integrating water quality assessment for protection of impaired watersheds. The goal of the CWA is to restore and maintain the chemical, physical, and biological integrity of the nation's waters. To meet the criteria of these mandates, models are often developed to study the current status of water quality and the impacts of various management plans (Borah and Bera, 2004).

Models are developed to represent actual systems with the best knowledge available. A parsimonious model describes the system accurately while keeping the complexity of the model at a minimum. When modeling environmental impacts it is important to remember that the world is diverse, continually evolving and has interdependence between the natural, built, economic, and social, cultural, and legal environments (Brimicombe, 2003). Models require revision as knowledge about the environment and technologies change. The revision efforts should be focused on improving the quality of results as well as the ability of users to interpret the results.

The Spatially Explicit Load Enrichment Calculation Tool (SELECT) was revised and automated for the Lake Granbury Watershed (Chapter II) to characterize *E. coli* production and distribution across the watershed. SELECT provides a Graphical User Interface (GUI), developed in Visual Basic for Applications (VBA) within ArcGIS 9.X, where project parameters can be adjusted for various pollutant loading scenarios. From

the visual output of the program a decision maker or stakeholder can identify areas of greatest concern for contamination contribution and incorporate that information while developing the Water Protection Plan or the TMDL development.

In a prior study (Teague, 2007) methodology for the Spatially Explicit Load Enrichment Calculation Tool (SELECT) for the characterization of *E. coli* sources was developed and applied to Plum Creek Watershed in Texas to aid Watershed Protection Plan (WPP) development. The draft of the WPP has been completed by the Plum Creek Watershed Partnership (PCWP, 2008). The major objective of this study was to compare the potential *E. coli* loads resulting from various sources in Plum Creek Watershed, Texas calculated using the automated approach with a manual approach developed by Teague (2007). The secondary objective of this study was to show improvement in load estimation through refinement, modification of assumptions, and automation.

3.2. Methodology

The approach for characterizing the *E. coli* sources is similar to the methodology developed by Teague (2007), with the exception of on-site wastewater treatment systems (OWTS). The approach outlined here for SELECT has been expanded, revised, and automated for extending its application to other watersheds.

The ultimate goal in spatially distributing potential *E. coli* loads is determining the suitable habitat for *E. coli* contributing sources, distributing the source populations, applying a fecal production rate, and then aggregating the potential load to the subwatersheds. Combining the potential load with environmental factors directly related to fate and transport of contaminant helps describe the connection between pollution across the watershed and the capability to reach the waterbodies. This is achieved through the pollutant connectivity factor (PCF) module of SELECT. The subwatersheds are ranked for potential contamination using the average PCF over multiple weighting scenarios for the influencing factors.

Two phases are involved in the model verification procedure. The first was to determine potential *E. coli* loads using SELECT with the same input sources as used in the previous manual approach. The simulated potential loads were then compared with the results from the 'manual' approach. The second phase evaluated the factors that influence contamination with the pollutant connectivity factor (PCF) component of SELECT. The PCF is a weighted combination of the Curve Number, distance to streams, and potential loading. The resulting ranked average PCF was compared to the results of the prior cluster analysis.

3.2.1. Watershed Description

The Plum Creek Watershed is a part of the Guadalupe River Basin and is located in Central Texas. It encompasses a drainage area of 397 square miles stretching across Hays, Caldwell, and Travis Counties. Plum Creek has a length of 52 river miles and joins the San Marcos River and eventually the Guadalupe River. Within the watershed there are several rapidly growing towns including Lockhart, Kyle, and Luling. Land use varies from urban to agriculture and oil field activities. The northern part of the watershed is primarily urban whereas the southern section has crop and animal agriculture along with oil wells. The landscape is characterized as rolling hills of pasture and cropland surrounded by scrub oak forest (GBRA, 2006). A section of Plum Creek is classified as impaired due to bacteria (USEPA, 2008). The watershed has been delineated using SWAT into 35 subwatersheds (Figure 3.1).

3.2.2. Model Simulation

The fecal production rates (Table 3.1) are provided in the EPA Protocol for Developing Pathogen TMDLs (USEPA, 2001). The lowest values of fecal coliform counts from the range of values were used to calculate potential *E. coli* loads. Each pollutant source population was distributed across the watershed in their appropriate habitats. The source populations were multiplied by a daily average fecal coliform excretion rate (USEPA,

2001) and then multiplied by 0.5, assuming a conversion estimate that fifty percent of fecal coliform (FC) are *E. coli* (Doyle and Erikson, 2006).

3.2.2.1. Potential E. coli Sources in Plum Creek Watershed

The potential *E. coli* sources evaluated for the Plum Creek watershed were cattle, sheep/goats, horses, feral hogs, deer, pets, OWTS malfunction, and wastewater treatment plants (WWTPs). Potential load from urban runoff is not calculated in the automated version of SELECT whereas it was calculate in the manual approach. Pets and other waste sources combined with OWTS sources contribute to *E. coli* loading in urban runoff. To compute the potential *E. coli* loading from urban areas a runoff volume is required. Since SELECT accounts for pet and human waste and has the ability to assess other sources (assuming available data), potential loading resulting from urban runoff is not included to reduce over prediction and double accounting of sources.

Livestock

The livestock (cattle, sheep/goats, and horses) were distributed as in the previous study (Teague, 2007) except areas inside of city limits were not excluded for distributing livestock populations. This is justified because livestock can be on pasture/rangeland within the jurisdiction of the city limits particularly in rural watersheds.

The livestock populations (Table 3.2) for Caldwell, Hays, and Travis counties were determined from county NASS agricultural inventory data. The cattle and sheep/goats population was distributed uniformly on grasslands and pasture/hay, since this land use is typically used for grazing. Horses are distributed only across pasture/hay landuse since horses are not allowed to roam as freely as other livestock.

Wildlife

To calculate *E. coli* potential load resulting from feral hogs, the same approach was applied as the previous study (Teague, 2007). The regional feral hog density is estimated

at 5 hogs/km^2 resulting in a total feral hog population of 5,141 hogs for the Plum Creek watershed. The feral hog population is redistributed on to all undeveloped land within a 100 m buffer of the streams.

According to a white-tail deer survey for the state of Texas (Lockwood, 2005) the Plum Creek Watershed intersects Resource Management Units (RMUs) 19, 20, and 7. The previous study (Teague, 2007) applied three different RMU specific deer population densities to determine the potential *E. coli* loading in the Plum Creek Watershed. Wildlife surveys are estimates with limited reliability and should be used only as a guideline for population densities in the areas studied. To estimate the potential *E. coli* loading resulting from deer the automated approach applied an area weighted population density of 0.0363 deer/ha uniformly across the watershed.

Pets

The approach for calculating *E. coli* potential load resulting from dogs was the same as the previous Plum Creek Watershed study (Teague, 2007). The assumption of a constant 0.8 dogs per home for Texas (AVMA, 2002) is an adjustable model parameter included in SELECT. The automated SELECT initiates ArcGIS to create a raster that represents the number of homes from the census block demographics table joined to the census block shapefile. The program allows the user to indicate the fecal production rate and then aggregates the potential load to zones of interest. Census block shapefiles are needed for each county and the associated census block demographics table for the state of Texas is indicated along with the appropriate attribute table field with the number of homes in each census block.

On-site Wastewater Treatment Systems (OWTSs)

For quantifying *E. coli* contributions from OWTS malfunction an entirely new approach was developed. Gathering detailed OWTS permit information from the county appraisal district is a tedious process as experienced in the Lake Granbury study (Chapter II).

Another approach was developed here for OWTS malfunction predictions using publicly available data and expert knowledge.

The number of homes on OWTSs was assumed to be equal to the number of homes not on municipal sewage. The Texas Commission on Environmental Quality (TCEQ) provides a GIS shapefile with water and sewer Certificates of Convenience and Necessity (CCN) service areas. This file indicates the areas that are currently served by public municipalities. The areas under public sewage collection were created into a raster grid with a value of 1 if within the CCN area and 0 outside for the extent of the study area. This grid was then converted to a "not-sewered" grid by reclassifying the 0 values to 1 and vice versa. The number of homes in the suitable area is determined from the census block shapefile, distributed on a per cell basis, and then multiplied by the "not-sewered" grid. The block-averaged number of people per home is also determined from creating a raster from the census block demographics. A soil rating file was created by processing SSURGO soils attribute table as described in the Lake Granbury Watershed study (Chapter II). The septic index combines the influence of soil type and age of the OWTS. For Plum Creek Watershed only limited subdivision age data is available, thus the septic index cannot be created. To estimate the percent of OWTS malfunctioning and potential contribution to pollution in a runoff event the soil rating was reclassified (Table 3.3) based on the assumption that up to 20% of all OWTSs malfunction and the estimated amount of effluent the soil is capable of treating.

The total potential *E. coli* loading was then determined by multiplying the number of homes per cell not on public sewer, by the number of people per home, the percent of effluent available, the discharge rate (70 gal/person/day), the effluent concentration of fecal coliform (10^6 cfu/100mL), and conversion factors. This resulted in potential daily *E. coli* load per cell which was then aggregated to the subwatershed level to yield the potential *E. coli* load in cfu/day on a subwatershed basis.

Wastewater Treatment Plants (WWTPs)

Within the GUI, the user selects the shapefile with the permitted outfall locations ensuring unrelated outfalls (i.e. cooling plants or any other non-pathogenic discharges) are removed. The file should include permitted discharges in the units of millions of gallons per day (MGD) as a field within the shapefile. The default (adjustable within the GUI) value of 126 cfu/100mL, the permitted limit for *E. coli*, was assumed. The loading was calculated by multiplying the discharge rate and the effluent concentration and then applying conversion factors to determine the loading in cfu/day. For this study, wastewater outfall locations were obtained from TCEQ GIS files (TCEQ, 2008b). The permitted flows were obtained from the EPA Envirofacts Data Warehouse (USEPA, 2006b). The automated approach is identical to the manual approach of the previous study (Teague, 2007).

Once all individual source inputs were selected and fed into SELECT a summation of potential *E. coli* loading from all sources was carried out. Thus, potential *E. coli* loading in a watershed, here Plum Creek, is spatially distributed.

3.2.2.2. Pollutant Connectivity Factor

The Pollutant Connectivity Factor (PCF) component of SELECT was applied to the Plum Creek watershed to identify potential areas contaminating waterbodies. The PCF was developed (Chapter II) to weigh the influence of the driving forces of contamination with the total pollution present. This application helps provide more information as to the areas with the greatest likelihood to contribute to contamination in waterbodies based on physical processes. The total PCF is a weighted combination of the normalized potential loading, a curve number grid, and the inverse of the normalized flow length to streams. The curve number is a direct relationship to runoff potential, and the flow length directly relates to fate and transport. A variety of weighting schemes were used to determine the ranking of the watersheds over a range of scenarios. The results from the PCF analysis were compared with the results from the clustering analysis in the prior

study (Teague, 2007) for the Plum Creek Watershed to determine the appropriate management practices for the watershed.

3.3. Results and Discussion

The resultant total potential *E. coli* loading and cluster analysis of SELECT parameters were previously determined (Figure 3.2). The total potential *E. coli* loading from the automated SELECT (Figure 3.3a) was calculated by adding all sources. For sources where the assumptions were almost identical to the previous application of SELECT, a percent difference from the prior to the current application was calculated. For sources where the approach was modified a comparison and explanation is provided. It is important to note when comparing the results of the prior and current study that the potential *E. coli* load determined by the manual approach is not a "true" value and the actual spatially distributed *E. coli* loading in the watershed is unknown.

3.3.1. Potential E. coli Loads in Plum Creek Watershed: Manual SELECT Approach

The identified point sources are WWTPs discharging effluent into Plum Creek and tributaries. Non-point sources in the watershed included urban runoff, OWTS malfunction, pets, livestock, and wildlife. Feral hogs and deer were the only wildlife sources characterized with SELECT because they are the primary populations of concern with available data. The livestock production within the study area included cattle, horses, sheep, and goats. The subwatersheds with high estimated potential *E. coli* loads are spatially distributed throughout the watershed (Figure 3.2a).

The results from the previous study indicate the highest potential contribution resulted from cattle, with 41% of the total average potential *E. coli* load. The second highest potential daily contributor was urban runoff with 27% of the total potential load. Dogs and feral hogs each had a potential of approximately 10.5% of the total potential *E. coli* load, and malfunctioning OWTSs comprised approximately 6.5% of the total potential. All other sources contributed less than five percent to the total potential load. The

percent contributions were also calculated for each subwatershed. It was observed that where urban runoff was present it was the predominant potential source. Furthermore, although cattle were the overall largest contributor, this source was more dispersed.

A cluster analysis was performed in the prior study to identify similar clusters of the subwatersheds of the Plum Creek Watershed based on the identification of distinguishing variables. Knowledge of the influencing factors through factor and principal component analysis allows for optimal planning efforts. The watershed was spatially characterized, by cluster analysis, into groups allowing for targeted efforts as determined by the identified unique features. This was accomplished through factor analysis, cluster analysis, and discriminant analysis. Plum Creek Watershed was statistically characterized into four distinct clusters, grouping the subwatersheds into management areas (Figure 3.2b). One cluster was high density urban; another was characterized as urban growth, another with range and forest lands, and lastly one cluster with no distinguishing characteristics. The set of variables used to characterize the subwatershed was reduced to factors that captured 80% of the variability. Furthermore, variables describing dog and cattle population were found to account for the majority of the variability within the watershed.

3.3.2. Potential E. coli Loads in Plum Creek Watershed: Automated SELECT Results

The resultant potential *E. coli* load from the automated SELECT provides a different characterization of the Plum Creek Watershed (Figure 3.3a) from the prior study results. The areas with the lowest potential *E. coli* load are in the northwest and central regions of the watershed. In contrast to the manual approach, the urban areas are no longer the highest potential subwatersheds. The ranked average PCF from multiple indicator weighting schemes can be used as a screening tool to determine which subwatersheds are potentially contributing to contamination (Figure 3.3b). Subwatershed 33 (Figure 3.1) has the highest potential to contribute to *E. coli* contamination followed by

Subwatersheds 20 and 30, respectively (Figure 3.3b). Subwatersheds 1 and 26 (Figure 3.1) have the lowest potential contribution of *E. coli* to waterbodies (Figure 3.3b).

3.3.3. Potential E. coli Loading Comparison

The differences in the approaches were compared (Table 3.4) to identify potential processing errors and understand the effect of assumptions on determining potential *E. coli* loads in Plum Creek Watershed. Variation in resultant potential *E. coli* loading was expected in comparison with the manual approach of the Teague (2007) study due to differences in processing of the spatial data. For example, calculations were performed using raster grids in the automated version of SELECT; whereas, in the previous application calculations were performed within feature polygon files and then the potential *E. coli* loads per polygon were converted to raster grids where a zonal sum was applied.

The total *E. coli* potential loading for the Plum Creek watershed using the automated SELECT was 32% lower than the previous study. This is a result of modified assumptions for all sources except feral hogs, wastewater treatment plants, and dogs. Another reason for this difference is the exclusion of calculating potential loading from urban runoff, which previously accounted for 27% of the total estimated load. The general trends for each source type, with the exception of OWTS, followed the same pattern of spatial distribution of potential loading (Figures 3.4, 3.5, and 3.7).

3.3.3.1. Cattle

The methodology for calculating potential loading from livestock sources is similar to the previous approach except livestock were not excluded within city limits. It is not appropriate to compare differences from the manual and automated approach on a subwatershed basis for livestock because the population distribution was modified in the automated approach. Theoretically the total potential loading across the watershed should be the same for both approaches since the same populations from livestock

inventories were assumed. The potential load resulting from cattle determined by the automated approach (Figure 3.4 b) across the watershed was 3% higher than the potential load determined by the manual approach (Figure 3.4a). The potential sheep/goat loading was 11% lower for the entire watershed for the automated results compared to the manual approach. The potential *E. coli* load from horses calculated using the manual approach was 38% lower than the automated SELECT results. The cause of this error is unknown since discrepancies in the code or processing steps could not be found. It was unexpected for the percent differences for the various livestock potential loading to be so disparaging since the subroutines followed the same processing steps. As mentioned earlier, an advantage of automating map processing is it is easy to determine if and when errors occur in processing while it is difficult to examine results of a manual approach.

3.3.3.2. Feral Hogs

Plum Creek Watershed was characterized with a similar spatial distribution of potential *E. coli* loading for feral hogs from the results of the manual and automated SELECT approaches (Figure 3.5a, b). The percent difference of potential *E. coli* loading for feral hogs from the prior study to the current ranges from 0 to 12 % for the subwatersheds (Figure 3.6). The highest percent difference occurs in areas with more stream segments. These areas have a greater chance for computational error since there is more information to be discretized into raster grid form.

3.3.3.3. Deer

The potential *E. coli* loading resulting from deer was determined in a similar manner as in the previous study. The number of deer was estimated using a population density of 0.0363 deer / ha distributed across the appropriate habitat. The assumed deer population was different while the suitable habitat was the same as the previous study. Thus, it is appropriate to compare the spatial trend of the deer *E. coli* potential load distribution (Figure 3.5c, d). Applying a uniform population distribution for deer resulted in the total

potential loading across the watershed to be 35% lower than the prior study. To compare the distribution trends, both resultant loading grids were classified into three categories based on a quantile distribution (Figure3.5c, d). The potential loads from the two approaches resulted in similar trends with only four subwatersheds having different classifications. The results from changes in methodology show the sensitivity of the input data. Three different deer population densities were used in the prior study and yet similar results were produced using only one population density. This indicates the importance of considering the reliability of data when determining the appropriate population distribution. For livestock it was appropriate to apply county specific population densities because the census data is more reliable. For wildlife populations, an area weighted average for the watershed is more appropriate due to inherent inaccuracies in data collection. The results from the two approaches indicate the suitable habitat, based primarily on landuse, has the greatest influence on the spatial distribution of potential load.

3.3.3.4. On-site Wastewater Treatment Systems (OWTSs)

The potential *E. coli* loading resulting from OWTSs using the automated SELECT approach was very low compared to the previous study (Figure 3.7). A low potential load was determined because the areas of the watershed with high human populations are on consolidated sewer systems. The Plum Creek Watershed is rural with low human population outside of Kyle, Lockhart, and Luling. Outside of these cities, the potential contribution of *E. coli* loading from human sources was almost negligible.

3.3.3.5. Wastewater Treatment Plants (WWTPs) and Pets

As expected, the WWTPs did not deviate from the previous study since outfall locations are represented with a point shapefile, thus discrepancies are not introduced when creating raster grids. The potential *E. coli* loading for dogs was 87% lower than the previous study. Errors in code or processing steps were not evident. Differences could be introduced when discretizing the continuous polygons into grids. Some of the census

blocks are small, irregular areas and the creation of raster cells may not be as representative as the density per unit area (900 m^2) as calculated in the previous application of SELECT. However, this only explains some of the expected difference. The disadvantage of manual steps is it is arduous to trace computational errors. In contrast, automation using programming code allows for debugging of conceptual and processing errors. Also, if the problem seems to be in the forming of grids from shapefiles, the cell size can be adjusted accordingly to minimize loss of information and then the program can be run again.

3.3.4. Susceptible Areas for Potential E. coli Loading: PCF vs. Cluster Analysis

The results from the Pollutant Connectivity Factor (PCF) analysis (Figure 3.3b) can be compared with the statistical clustering (Figure 3.2b) to establish the relationship between the physical processes and the statistical analysis. A similar grouping scheme emerges with the PCF ranking as the Cluster Analysis. PCF accounts for runoff potential and travel distance to streams, while clustering places similar variables such as landuse classifications into groups. Together, these representations can be used as a supplement in Watershed Protection Planning. The Plum Creek Watershed has four distinct clusters as determined by clustering and factor analysis (Teague, 2007).

 Cluster 1 membership is characterized by forest and rangeland (Figure 3.2b). The same subwatersheds belonging to Cluster 1 were ranked using the PCF as the highest potential contributors to bacterial pollution (Figure 3.3b). This indicates that the load reductions determined by the stakeholders should focus on forest and rangeland management practices. Cluster 2 membership is characterized as high density urban. These subwatersheds include the cities of Kyle and Lockhart. The PCF ranking for both of these subwatersheds was eleventh, signifying similar loading and flow characteristics of these areas. The ranking is moderately low, indicating a lower 'risk' or potential for contributing to contamination than the subwatersheds in Cluster 1. Cluster 3 membership includes subwatersheds without distinct characteristics. In other words, no particular

source or landuse is dominating the results for determining the potential load. These subwatersheds had a wide range of rankings from fifth to fourteenth with most of the subwatersheds having a moderate to moderately low potential to contribute to contamination. Lastly, Cluster 4 membership is distinguished as urban growth areas. The subwatersheds belonging to Cluster 4 were ranked from ninth to seventeenth with the majority ranking relatively low compared to other clusters. This comparison portrays that urban growth is currently not a major concern for contributing to *E. coli* contamination in waterbodies. These areas are typically ranked lower than the high density urban areas primarily due to the influence of lower human population resulting in less contribution from domestic sources such as pets and OWTS malfunction. As these areas do experience more growth, however, they will likely become a concern unless they are included in consolidated sewer systems or people are educated about proper installation and maintenance of on-site wastewater treatment systems and pet waste management.

3.4. Conclusions

The SELECT methodology was automated and refined (Chapter II) to apply this tool to more diverse watersheds. The automated tool was verified by assessing the potential *E. coli* loading in the Plum Creek Watershed and then comparing to previous results from the manual approach. As expected the potential loading was not the same as previously determined because of changes in methodology. The total *E. coli* potential load for the watershed was also significantly lower than the previous study due to major differences in assumptions such as the exclusion of urban runoff. The Pollutant Connectivity Factor provides helpful aid in grouping areas with similar physical characteristics such as load potential which is a direct relation to landuse, runoff capability as determined with the curve number, and the ability for pollutants to reach waterbodies using the flow length. Comparing the results of the PCF analysis estimated by physical properties of the watershed with the statistical clustering of watershed characteristics provided similar groupings. When PCF and Cluster Analysis techniques are used in conjunction with the

results of SELECT, these approaches can facilitate the determination of load reductions through implementing spatially explicit, source specific best management practices.

SELECT can assist decision makers and stakeholders develop Watershed Protection Plans and determine TMDLs. The applicability of this approach depends on the quality of data available and the ability of the user to interpret the results. The SELECT methodology may need to be revised in the future as further knowledge becomes available such as a more accurate account of OWTS malfunction or suitable habitats and population densities for wildlife. The PCF and clustering analysis are additional components to supplement the decision makers with more information on interactions of the source pollutants with their environment.

CHAPTER IV
CONCLUSIONS

4. 1. Conclusions of this Research

The following are the conclusions of this research work:

1. The Spatially Explicit Load Enrichment Calculation Tool (SELECT) was developed and automated to calculate potential *E. coli* loads from various sources across a watershed.

2. SELECT was applied to the Lake Granbury Watershed in Texas. Based on simulation results for Lake Granbury, Best Management Practices (BMPs) are recommended to decrease *E. coli* loads from pets and OWTSs near the lake. Further investigation using watershed-scale water quality models such as SWAT or HSPF is needed to determine the influence of various *E. coli* sources across the watershed. Travel time from the subwatersheds with high potential loading should be determined to characterize the amount of *E. coli* reaching the waterbodies after a runoff event.

3. It is also recommended that water quality monitoring should be carried out in northern and western portions of the Lake Granbury watershed to monitor *E. coli* concentrations in the watershed. This will ultimately help in protecting Lake Granbury from contamination due to pathogenic bacteria.

4. SELECT is a user-friendly tool to conduct spatial analysis under different land use scenarios. In addition to this, maps and tables resulting from SELECT can be used for technical and educational communication. This approach proves the need to evaluate each contaminant source separately to effectively allocate site specific BMPs. This methodology also serves as a powerful screening tool for determining areas where detailed investigation is merited.

5. The automated SELECT methodology was verified by assessing the potential *E. coli* loading in the Plum Creek Watershed and comparing to previous manual approach results (Teague, 2007). As expected the potential loading was not the same as previously determined because of changes in methodology. The total *E. coli* potential load for the watershed was also significantly lower than the previous study due to major differences in assumptions such as the exclusion of urban runoff.

6. The Pollutant Connectivity Factor (PCF) module provides helpful aid in grouping areas with similar physical characteristics such as load potential which is a direct relation to landuse, runoff capability as determined with the curve number, and the ability for pollutants to reach waterbodies using the flow length. Comparing PCF estimated by physical properties of the watershed with the statistical clustering of watershed characteristics provided similar groupings in the Plum Creek Watershed. In conjunction with SELECT results, PCF and Cluster Analysis can help facilitate the determination of load reductions through implementing spatially explicit, source specific best management practices.

7. SELECT is a tool that can assist decision makers and stakeholders develop Watershed Protection Plans and determine TMDLs. The SELECT methodology may need to be revised in the future as further knowledge such as a more accurate account of OWTS malfunction or suitable habitats and population densities for wildlife are available.

8. SELECT with PCF and clustering analysis will provide the decision makers more information of the interactions of source pollutants with their environment.

4.2. Future Recommendations

SELECT is an assessment tool to determine potential *E. coli* loads in watersheds and evaluate contributing sources. In order to create a more comprehensive tool further research is recommended. The OWTS Malfunctioning Rate, developed from expert opinion, should be incorporated into the GUI as a model parameter to adjust for varying assumptions. To obtain a more accurate assessment of potential *E. coli* contributions from OWTSs studies should be conducted to determine malfunction rates in relation to the age of OWTSs and soil conditions. Another program option that would help more accurately predict the distribution of *E. coli* loads across watersheds would weigh the landuse classifications to distribute source populations (i.e. 70% of cattle graze on grassland and 30% on pasture). In order to investigate the source of discrepancies between the manual and automated approach it is recommended to distribute census block demographics data on a density per unit area (equal to the cell size) prior to creating a raster from the shapefiles. The average ranking for PCF using a sensitivity analysis based on expert opinion is not fully automated. The weighting scheme could be implemented with a user input table and then the program would produce the average ranking for PCF. Research into the appropriate weighting scheme using statistical techniques such as Bayesian statistics or expert opinion surveys such as with the Delphi approach would further validate the PCF approach.

To provide stakeholders and decision makers more information for WPP and TMDL support an application could be incorporated within SELECT to determine potential loads after applying common management practices (i.e. restricting cattle to certain distances from streams). Another goal of SELECT is to apply a fate and transport mechanism and incorporate into a watershed model. Model calibration requires water quality data. Unfortunately, historical bacteria water quality samples measured fecal coliform whereas current standards require *E. coli* monitoring. Pathogen production from animal waste was also studied for fecal coliform rather than *E. coli*. Therefore, the appropriate conversion factor from fecal coliform observations to *E. coli* data is needed.

REFERENCES

Al-Ebus, M. and Jacobson, B. A., 2002. Protocol for Developing Fecal Coliform TMDLs in the State of New Jersey. In: Proceedings of National TMDL Science and Policy 2002 Specialty Conference, 13-16 NAl-Ebusov., Phoenix, AZ.

AVMA, 2002. US Pet Ownership and Demographics Source Book. Center for Information Management, American Veterinary Association. Schaumburg, IL.

Benham, B.L., Baffaut, C., Zeckoski, R. W., Mankin, K. R., Pachepsky, Y.A., Sadeghi, A. M., Brannan, K. M., Soupir, M. L., Habersack, M. J., 2006. Modeling Bacteria Fate and Transport in Watersheds to Support TMDLs. Transactions of American Society of Agricultural and Biological Engineers. 49 (4) 987-1002.

Brimicombe, A., 2003. GIS, Environmental Modeling and Engineering. Taylor & Francis, London.

Bora, D. K., and Bera, M., 2004. Watershed-Scale Hydrologic and Nonpoint-Source Pollution Models: Review of Applications. Transactions of the American Society of Agricultural Engineers. 47 (3) 789-803.

BRA, 2008. Brazos River Watershed Information Center. Brazos River Authority: Waco, Texas. Available at: http://crpdata.brazos.org/. Accessed on: 12 April 2008.

Chapman, G., Gray, J., Irvine, R., Barry, M., 2004. Using Soil Landscape Mapping for On-Site Sewage Risk Assessment. SuperSoil 2004: 3rd Australian New Zealand Soils Conference, University of Sydney, Australia. Available at: www.regional.org.au/au/asssi/. Accessed on: 15 April 2008.

Chen, C. W., Herr, J., Ziemelis, L., Goldstein, R. A., Olmsted, L., 1999. Decision Support System for Total Maximum Daily Load. Journal of Environmental Engineering. ASCE 125 (7) 653-659.

Clark, M. K., Heigis, W. S., Douglas, B. F., Hoover, J. B., 2001. Decentralized Wastewater Management Needs Assessment: A Small Community's Approach to On-Site Treatment. In: Proceedings of the Ninth National Symposium On

Individual and Small Community Sewage Systems, American Society of Agricultural Engineers, Warren, VT. Pp. 427-434.

Cleland, B., 2002. TMDL Development from the "Bottom Up" – Part II: Using Duration Curves to Connect the Pieces. National TMDL Science and Policy 2002 – WEF Specialty Conference, America's Clean Water Foundation: Washington, D.C., Pp. 7-14

Cleland, B., 2003. TMDL Development from the "Bottom Up" – Part III: Duration Curves and Wet-weather Assessments. America's Clean Water Foundation: Washington, D.C. Available at: http://www.tmdls.net/tipstools/docs/TMDLsCleland.pdf. Accessed on: 15 May 2007.

Doyle, M. and Erikson, M., 2006. Closing the Door on the Fecal Coliform Assay. Microbe 1 (4) 162-163.

Espey Consultants, Inc., 2007. Lake Granbury Water Quality Modeling Project PHASE 1 Draft Report – Data Trend Analysis, Modeling Overview and Recommendations. Prepared for Brazos River Authority. Espey Consultants: Austin, TX.

Ferguson, C., de Roda Husman, A.M., Altavilla, N., Deere, D., Ashbolt, N., 2003. Fate and Transport of Surface Water Pathogens in Watersheds. Critical Reviews in Environmental Science and Technology. 33 (3) 299-361.

GBRA, 2006. Guadalupe River Basin: Basin Highlights Report, Guadalupe-Blanco River Authority, Seguin, TX. Available at: http://www.gbra.org/Documents/CRP/BDA/2006BasinHighlightsReport.pdf. Accessed on: 22 August 2006.

Houck, O. A., 1997. TMDLs, Are We There Yet?: The Long Road Toward Water Quality-Based Regulation under the Clean Water Act. Environmental Law Review. 27 (8) 10401.

Jamieson, R., Gordon, R., Joy, D., Lee, H., 2004. Assessing Microbial Pollution of Rural Surface Waters: A Review of Current Watershed Scale Modeling Approaches. Agricultural Water Management. 70, 1-17.

Kenway, S. and Irvine, R., 2001. Sewage Pollution Risk Assessment for Environmental Health. In: Conference Proceedings - Environmental Health Conference 2001, 11-12 September, Bathurst, Australia. 1-8.

Lesikar, B., 2007. Personal Communication with Dr. Bruce Lesikar; Professor, Extension Specialist, Associate Department Head & Extension Program Leader, Texas A&M University, College Station, Texas. 7 December 2007.

Li, C., Guo, Q., 2003. Integrated Point and NonPoint Source Modeling of Fecal Coliform in Whippany Watershed, New Jersey. In: Conference Proceedings of the World Water & Environmental Resources Congress, Philadelphia, PA. ASCE 22-26 June 2003.

Lockwood, 2005. White-tailed Deer Population Trends. Federal Aid in Fish and Wildlife Restoration. Project. W-127-R-14. Texas Parks and Wildlife Department. Austin TX.

Mitchell, G. 2005. Mapping Hazard from Urban Non-Point Pollution: A Screening Model to Support Sustainable Urban Drainage Planning. Journal Environmental Management. 74, 1-9.

Morgan, T., 2008. Personal Communication. Watershed Project Coordinator, Brazos River Authority, Waco, TX.

Moyer, D. and Hyer, K.E., 2003. Use of the Hydrologic Simulation Program - FORTRAN and Bacterial Source Tracking for Development of the Fecal Coliform Total Maximum Daily Load (TMDL) for Accotink Creek, Fairfax County, Virginia. Water Resources Investigation Report 03-4160. United States Geological Survey and Virginia Department of Conservation and Recreation: Richmond, Virginia.

Munafo, M., Cecchi, G., Baiocco, F., Mancini, L., 2005. River Pollution from Non-Point Sources: A New Simplified Method of Assessment. Journal of Environmental Management. 77, 93-98.

NCDC, 2008. National Climate Data Center. U.S. Department of Commerce - NOAA: Washington, D.C. Available at: http://www.ncdc.noaa.gov/oa/mpp/freedata.html. Accessed on: 12 April 2008.

Paul, S., Srinivasan, R., Sanabria, J., Haan, P., Mukhtar, S., Neimann, K., 2006. Groupwise Modeling Study of Bacterially Impaired Watersheds in Texas: Clustering Analysis. Journal of the American Water Resource Association. August: 1017-1031.

Payment, P., Hunter, P.R., 2001. Endemic and Epidemic Infectious Intestinal Disease and Its Relationship to Drinking Water. In: World Health Organization (WHO). Water Quality: Guidelines, Standards and Health, IWA Publishing: London, UK.

PCWP, 2008. Plum Creek Watershed Protection Plan: Draft. Plum Creek Watershed Partnership. Texas AgriLife Extension, College Station, TX. Available at: http://pcwp.tamu.edu/wpp.html. Accessed on: 7 April 2008.

Santhi, C., Arnold, J., Williams, J., Huack, L., Dugas, W., 2001. Application of a Watershed Model to Evaluate the Management Effects on Point and Nonpoint Source Pollution. Transactions of ASAE. 44(6) 1559-1570.

SCS, 1986. Urban Hydrology for Small Watersheds, Technical Release 55, U.S. Dept. of Agriculture, Soil Conservation Service, Washington, D.C., Chapter 2: pp. 2-5.

SWAT, 2005. Soil and Water Assessment Tool. Texas A&M University, College Station, Texas and USDA Agricultural Research Service: Grassland, Soil and Water Research Laboratory, Temple, Texas. Available at: http://www.brc.tamus.edu/swat/soft.html Accessed on: 10 March 2008.

TCEQ, 2000. Revisions to Chapter 307 – Texas Surface Water Quality Standards. Austin, TX. TCEQ. Available at: http://www.epa.gov/waterscience/standards/wqslibrary/tx/tx-wqs-20061215.pdf. Accessed on: 21 May 2007.

TCEQ, 2008a. Regulatory/Administrative Boundaries. Texas Commission on Environmental Quality: Austin, Texas. Available at: http://www.tceq.state.tx.us/gis/boundary.html. Accessed on: 20 March 2008.

TCEQ, 2008b. Site Layers. Texas Commission on Environmental Quality: Austin, Texas. Available at: http://www.tceq.state.tx.us/gis/sites.html. Accessed on: 20 March 2008.

Teague, A. E., 2007. Spatially Explicit Load Enrichment Calculation Tool and Cluster Analysis for Identification of *E. coli* Sources in Plum Creek Watershed, Texas. Unpublished MS thesis. Texas A&M University, Department of Biological and Agricultural Engineering, College Station, Texas.

TWRI, 2007. Water Quality Program for Lake Granbury, Texas. Texas A&M AgriLife: College Station, Texas. Available at: http://lakegranbury.tamu.edu/index.php. Accessed on: 5 February 2007.

USCB, 2000. Census 2000 TIGER/Line® Files. Washington, D.C.: US Census Bureau. Available at: http://www.census.gov/geo/www/tiger/index.html. Accessed on: 13 October 2007.

USDA-NRCS, 2004. Soil Survey Geographic (SSURGO) Database. Washington, D.C.: United States Department of Agriculture National Resource Conservation Service. Available at: http://soils.usda.gov/survey/geography/ssurgo/. Accessed on: 19 December 2006.

USEPA, 2000. Bacterial indicator tool: User's guide. EPA-823-B-01-003. USEPA Office of Water: Washington, D.C.

USEPA, 2001. Protocol for Developing Pathogen TMDLs: Source Assessment. 1st ed. EPA841-R-00-002. Ch. 5 pp 1-18. Washington, D.C.:USEPA Office of Water.

USEPA, 2006a. Draft: An Approach to Using Load Duration Curves in Developing TMDLs. United States Environmental Protection Agency: Office of Wetlands, Oceans, and Watersheds, Washington, D.C.

USEPA, 2006b. Envirofacts Data Warehouse. U.S. Environmental Protection Agency. Washington, D.C. Available at: Accessed on: 8 February 2008.

USEPA, 2008. Total Maximum Daily Loads: National Section 303(d) List Fact Sheet: Top 100 Impairments. U.S. Environmental Protection Agency: Washington, D.C. Available at: http://oaspub.epa.gov/waters/national_rept.control#TOP_IMP. Accessed on: 8 February 2008.

Vairavamoorthy, K., Yan, J., Galgale, H., and Gorantiwar, S. D., 2007. IRA-WDS: A GIS-Based Risk Analysis Tool for Water Distribution Systems. Environmental Modelling & Software. 22, 951-965.

Zeckoski, R. W., Benham, B. L., Shah, S. B., Wolfe, M. L., Brannan, K. M., Al-Smadi, M., Dillaha, T. A., Mostaghimi, S., Heatwole, C. D., 2005. BSLC: A Tool for Bacteria Source Characterization for Watershed Management. Applied Engineering in Agriculture. ASAE. 21(5) 879-889.

APPENDIX A

FIGURES
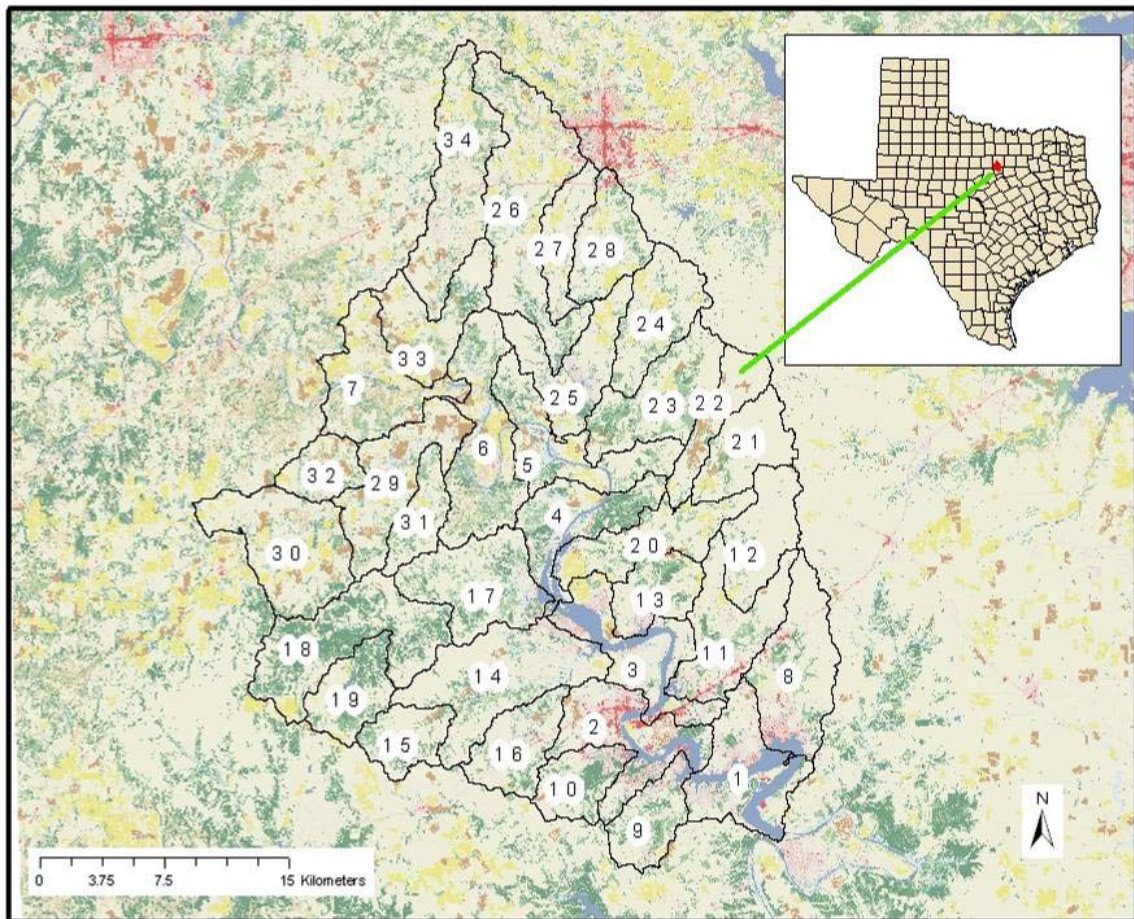
Produced by: Kendra Riebschleager
Date: April 2008

Figure 2.1. Location of Lake Granbury with Subwatersheds Delineated Using SWAT.
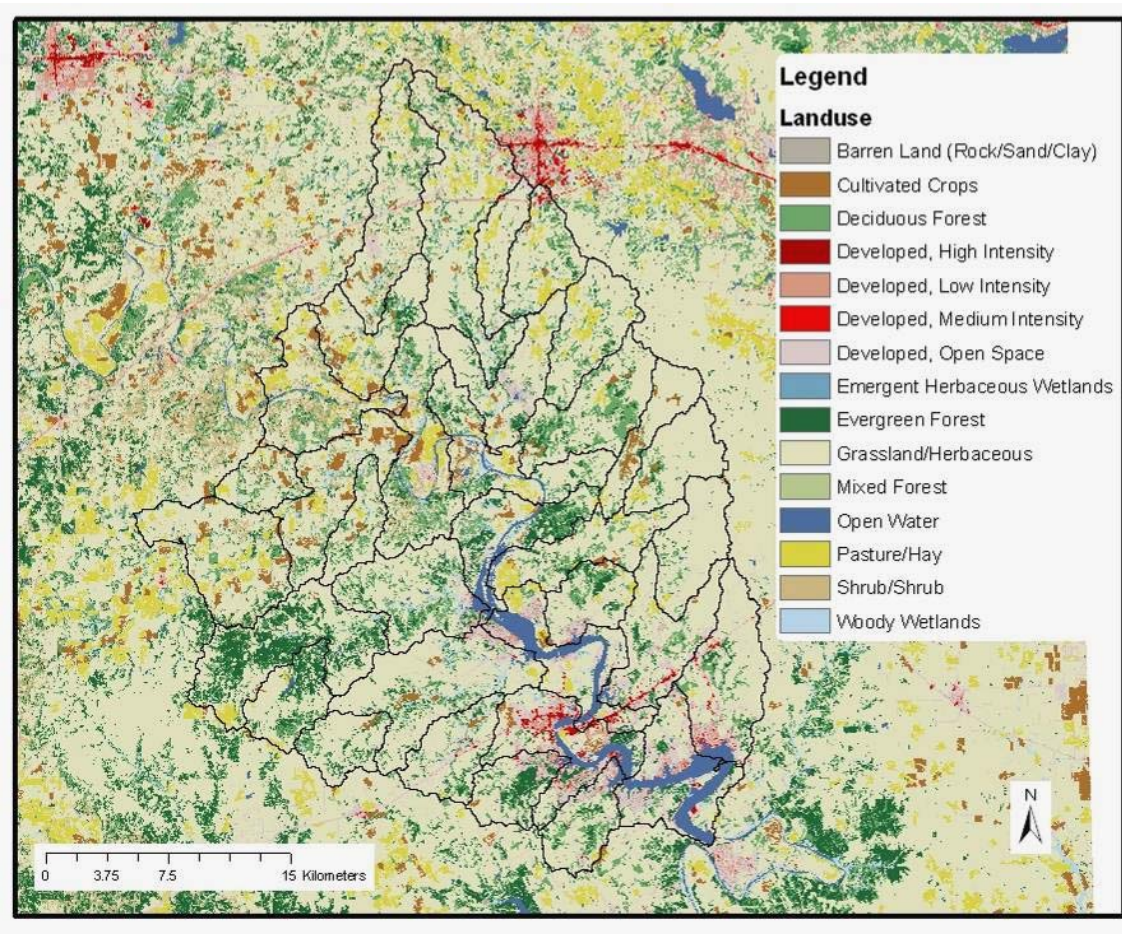
Figure 2.2. Landuse Classification (2001 NLCD) of Lake Granbury Watershed.
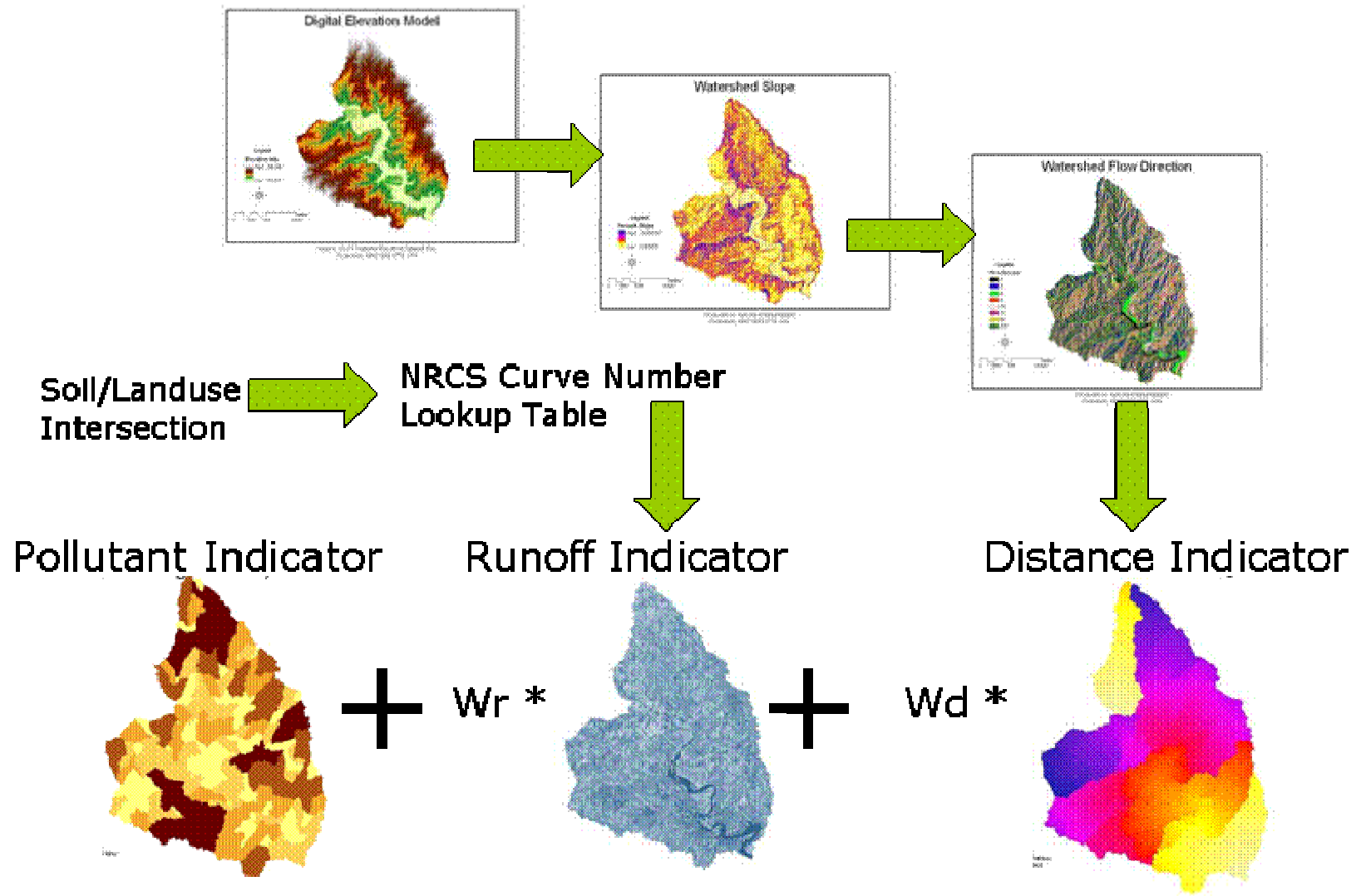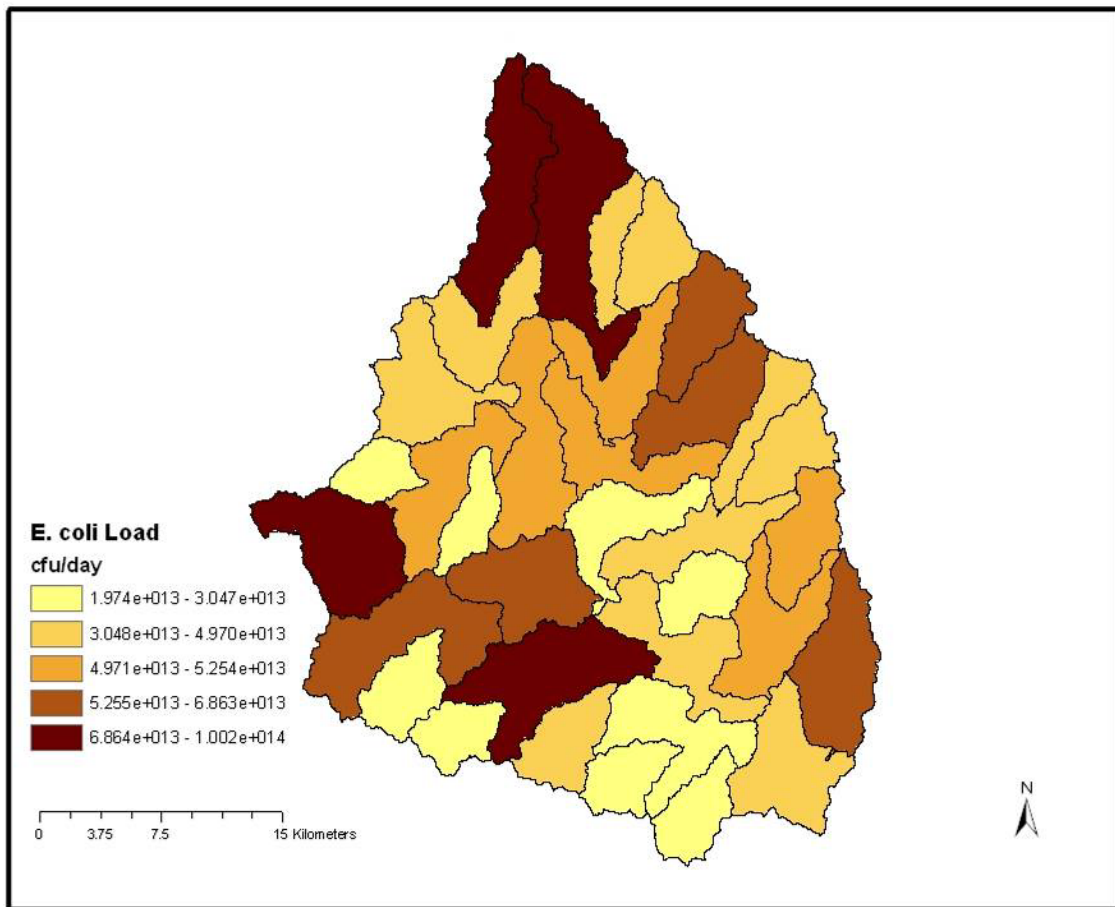
Figure 2.3. Spatial and Hydrologic Processes to Determine the Pollutant Connectivity Factor.

Produced by: Kendra Riebschleager
Date: April 2008

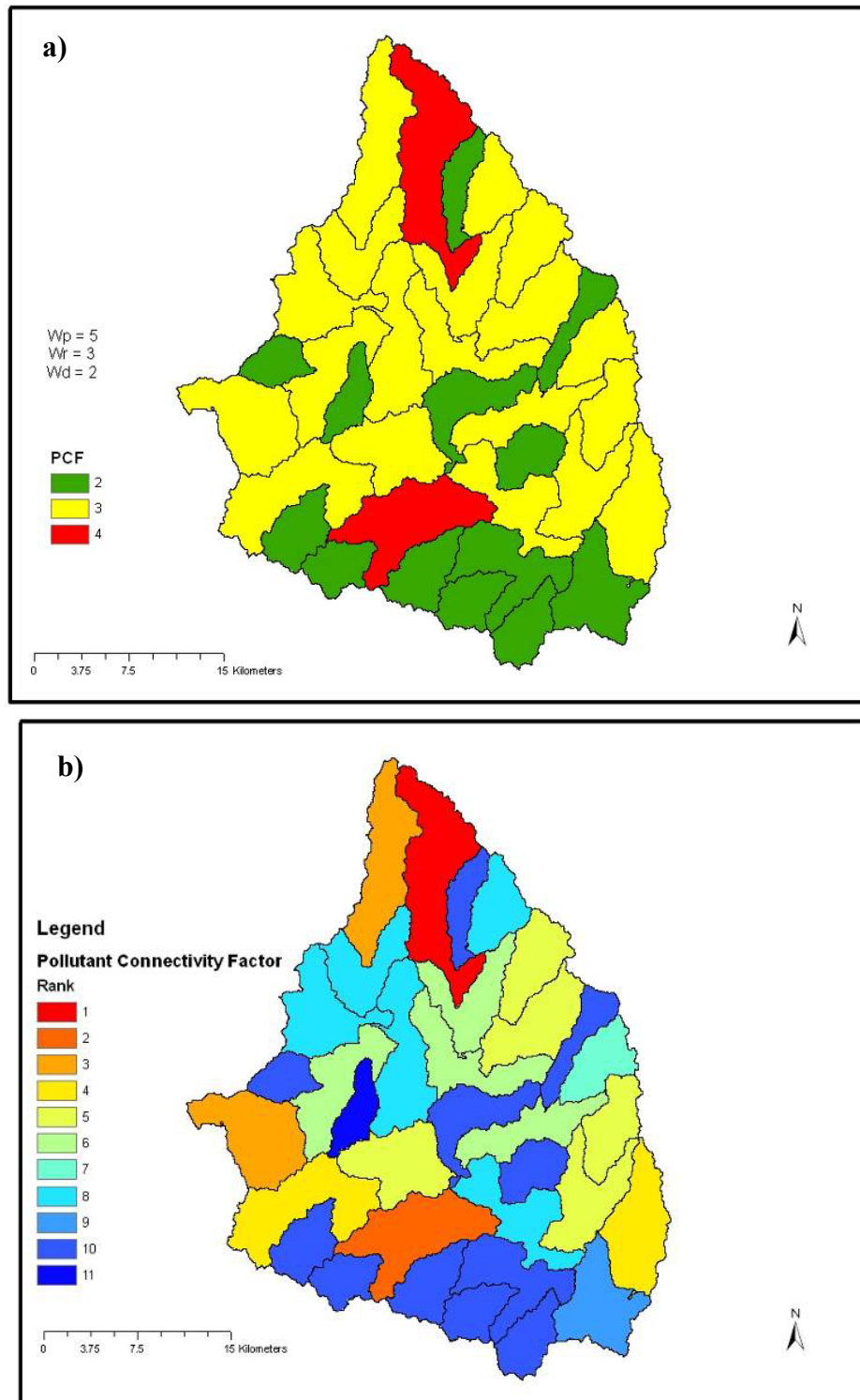**Figure 2.4. Total Potential *E. coli* Load from All Sources in Lake Granbury Watershed.**

Figure 2.5. Pollutant Connectivity Factor for Total *E. coli* Potential Load Determined by a) Expert Knowledge Weighting and b) Ranked Subwatersheds Averaged Over Multiple Weighting Scenarios.
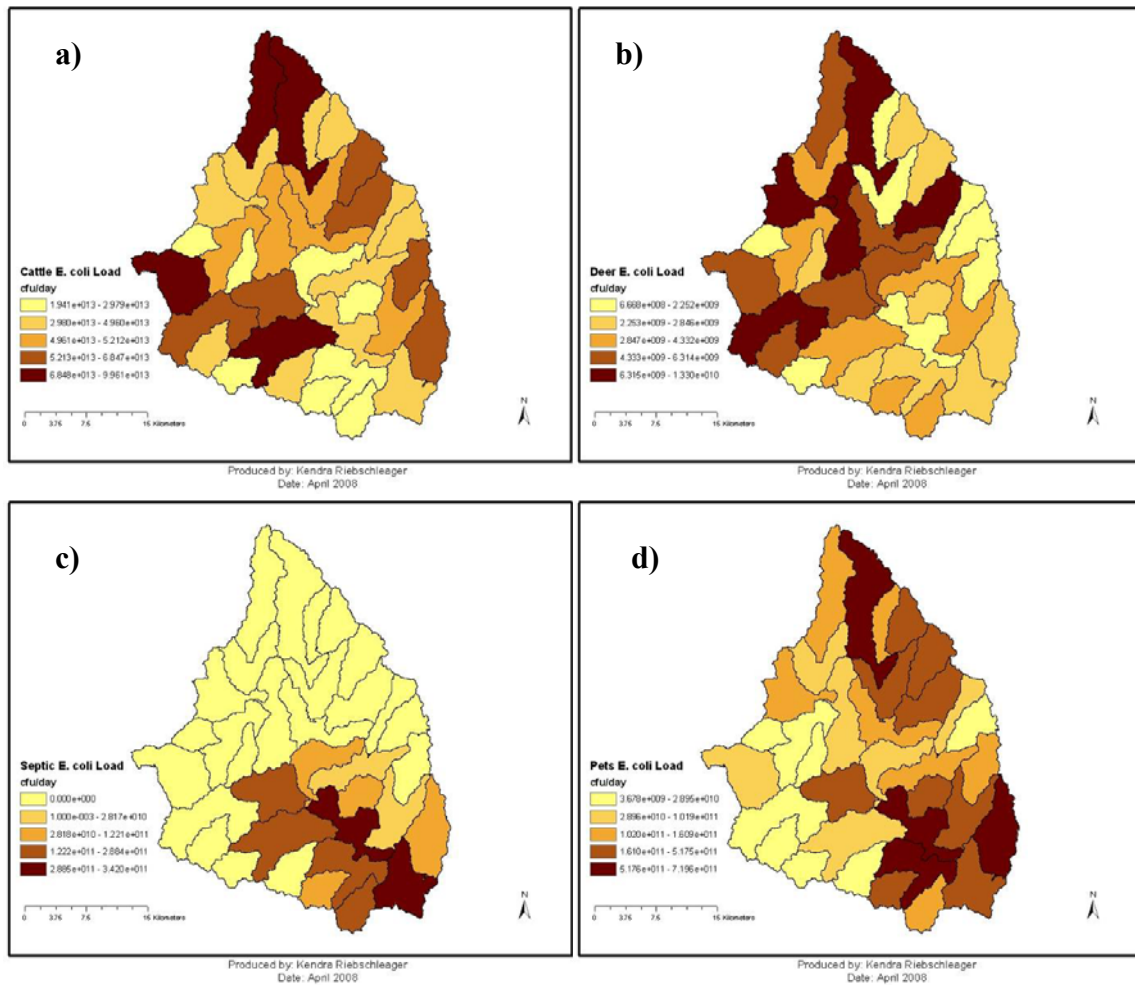
59

Figure 2.6. Potential *E. coli* Load in Lake Granbury Watershed Resulting from Various Non-Point Sources: a) Cattle, b) Deer, c) Pets, and d) On-site Wastewater Treatment Systems (OWTS).
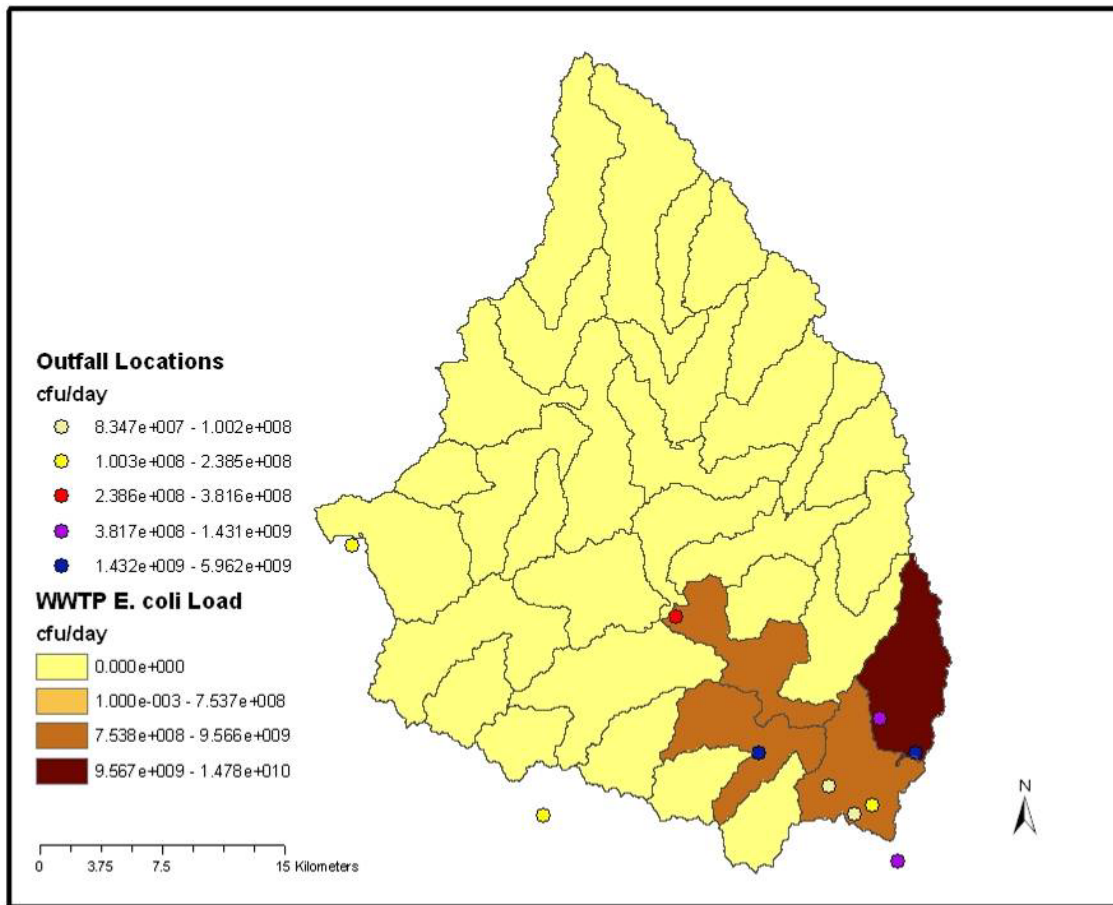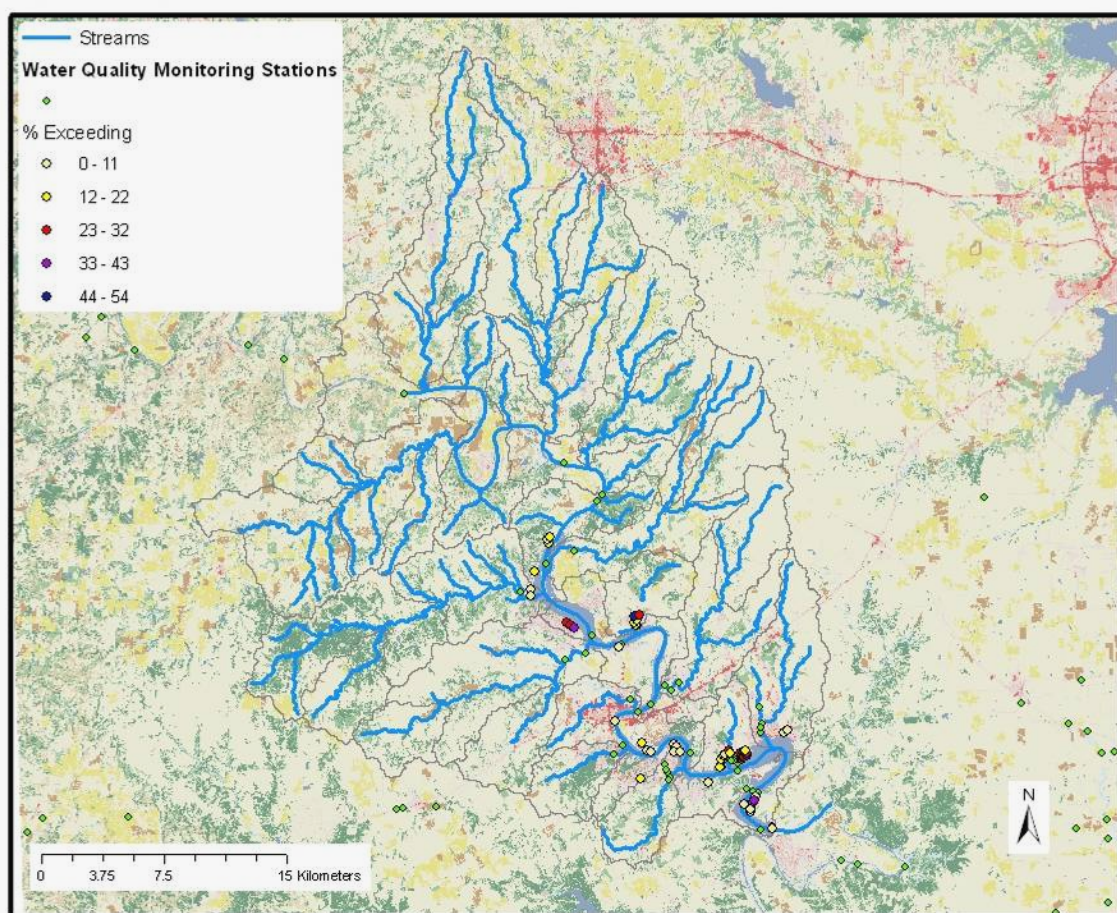
Figure 2.7. Potential *E. coli* Loading from Wastewater Treatment Plants.

Figure 2.8. Water Quality Monitoring Stations Located within the Lake Granbury Watershed with Percent of Observations Exceeding *E. coli* Standard (126 cfu / 100mL).
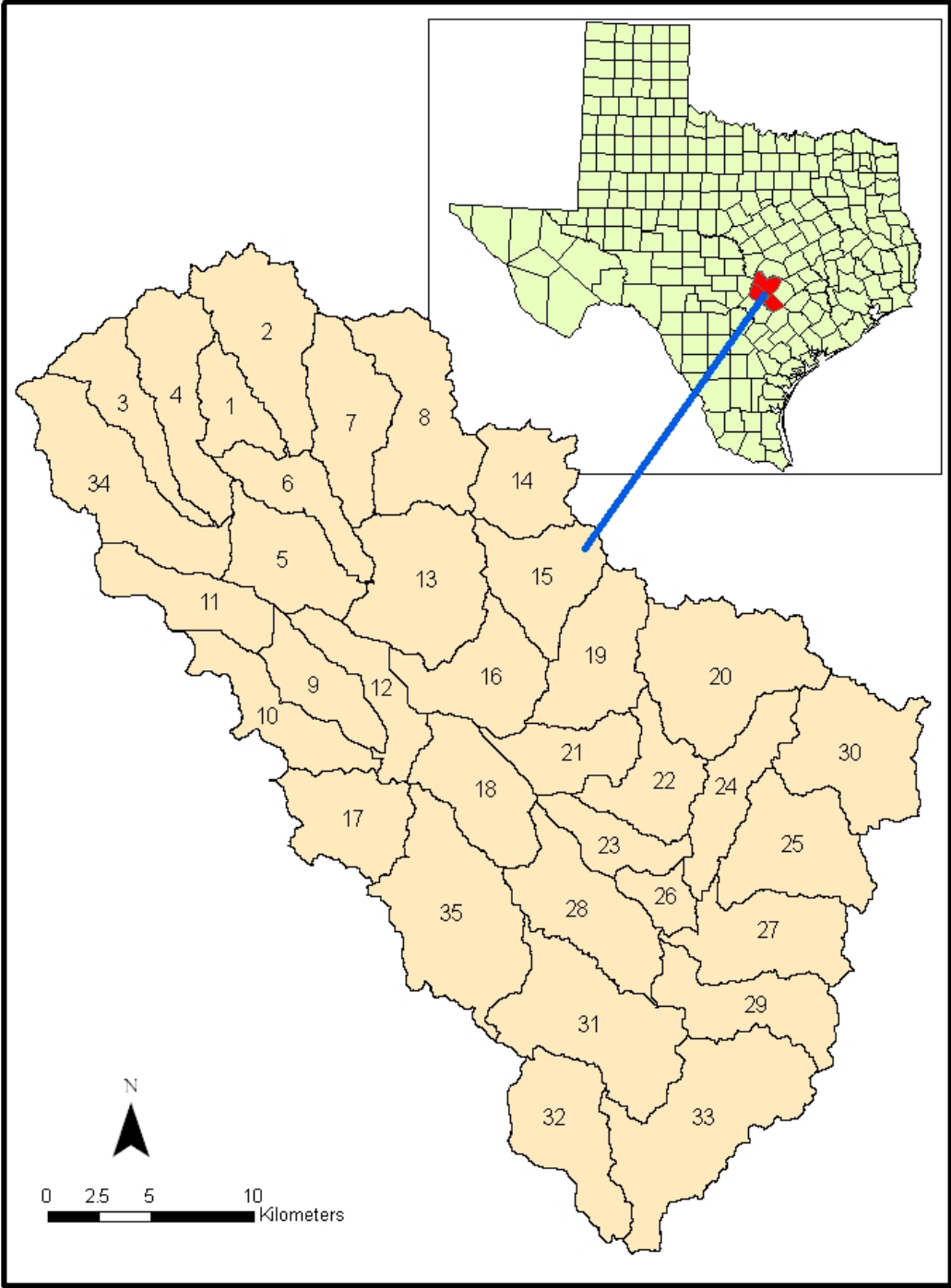
Figure 3.1. Plum Creek Watershed Divided into 35 Subwatersheds Delineated Using SWAT.
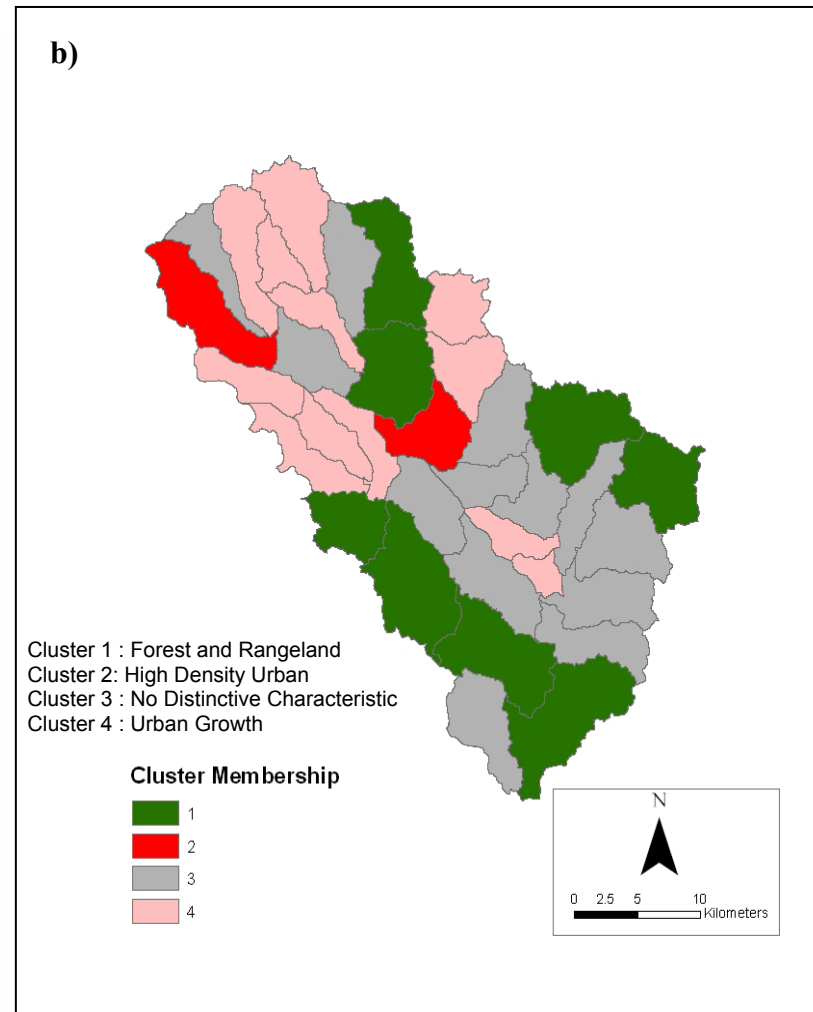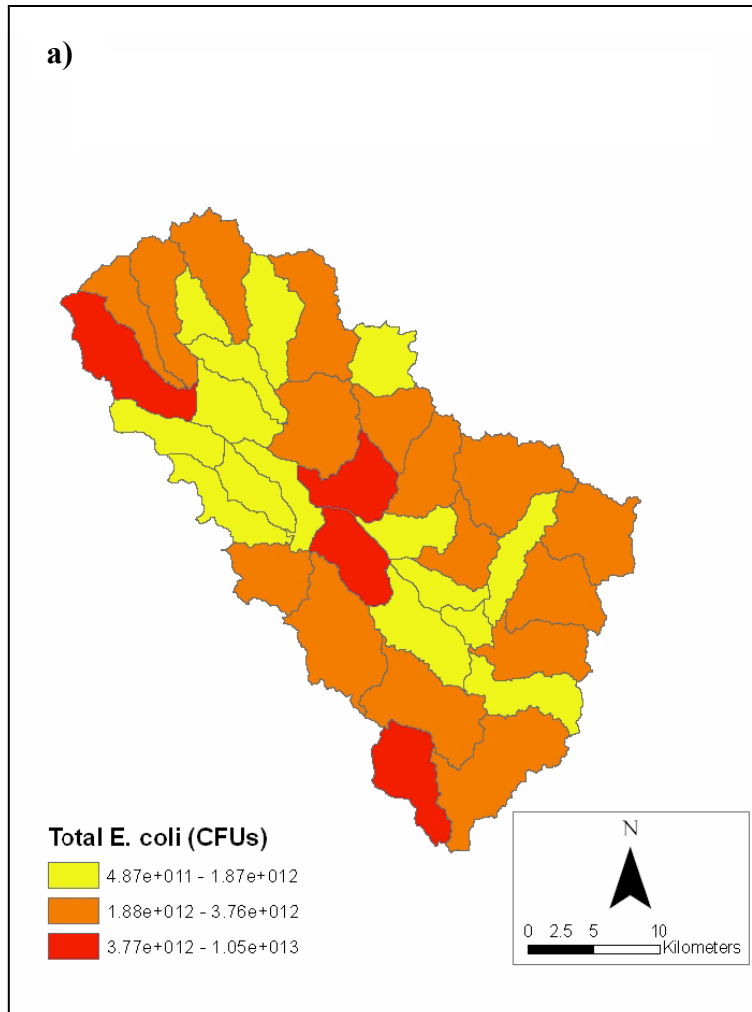
Figure 3.2. a) Total Potential Daily *E. coli* Load and b) Clusters of Subwatersheds with Unique Features (Teague, 2007).
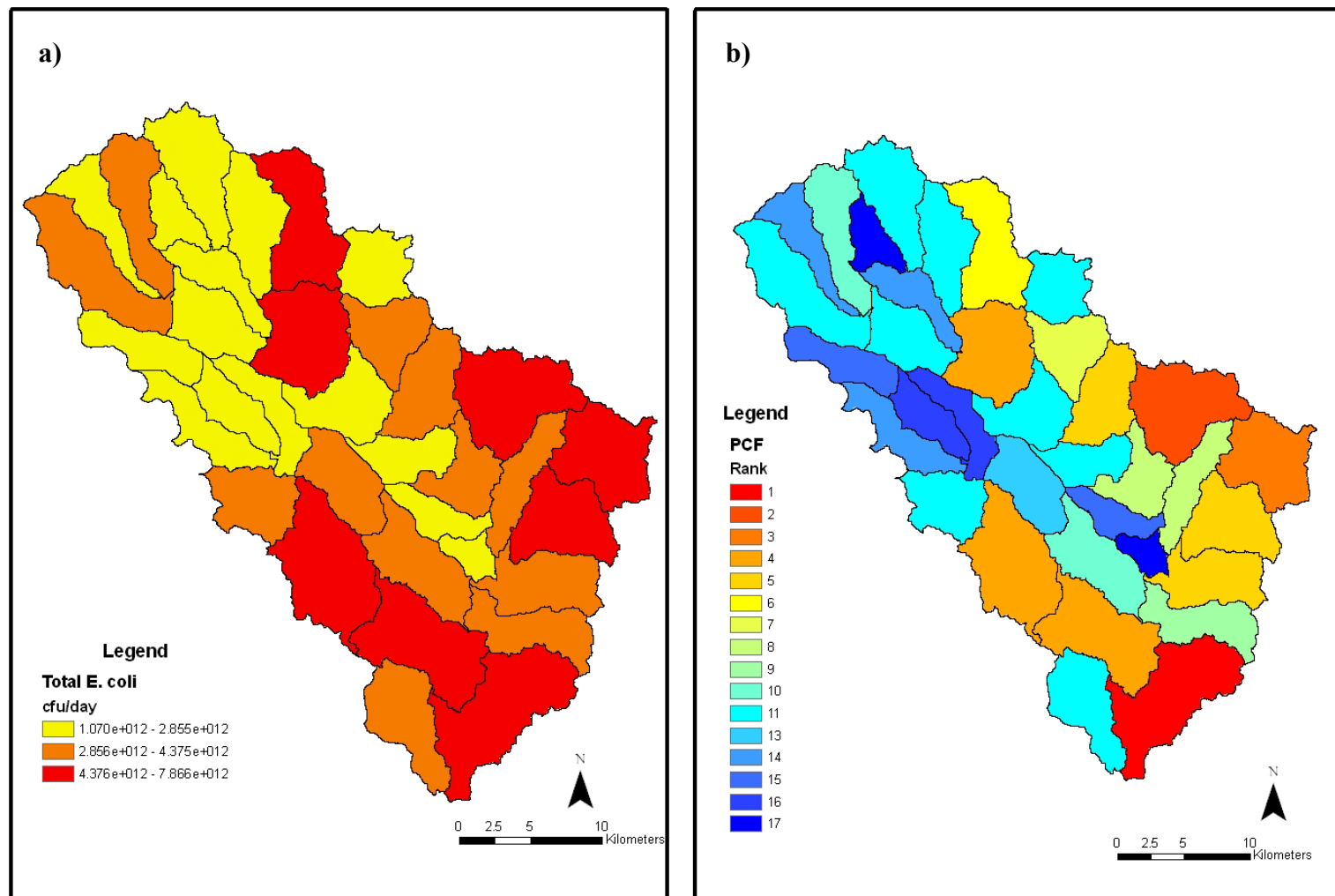
Figure 3.3. a) Daily *E. coli* Potential Loading from All Sources and b) Ranking of Pollutant Connectivity Factors Averaged from Multiple Weighting Scenarios Using Automated SELECT.
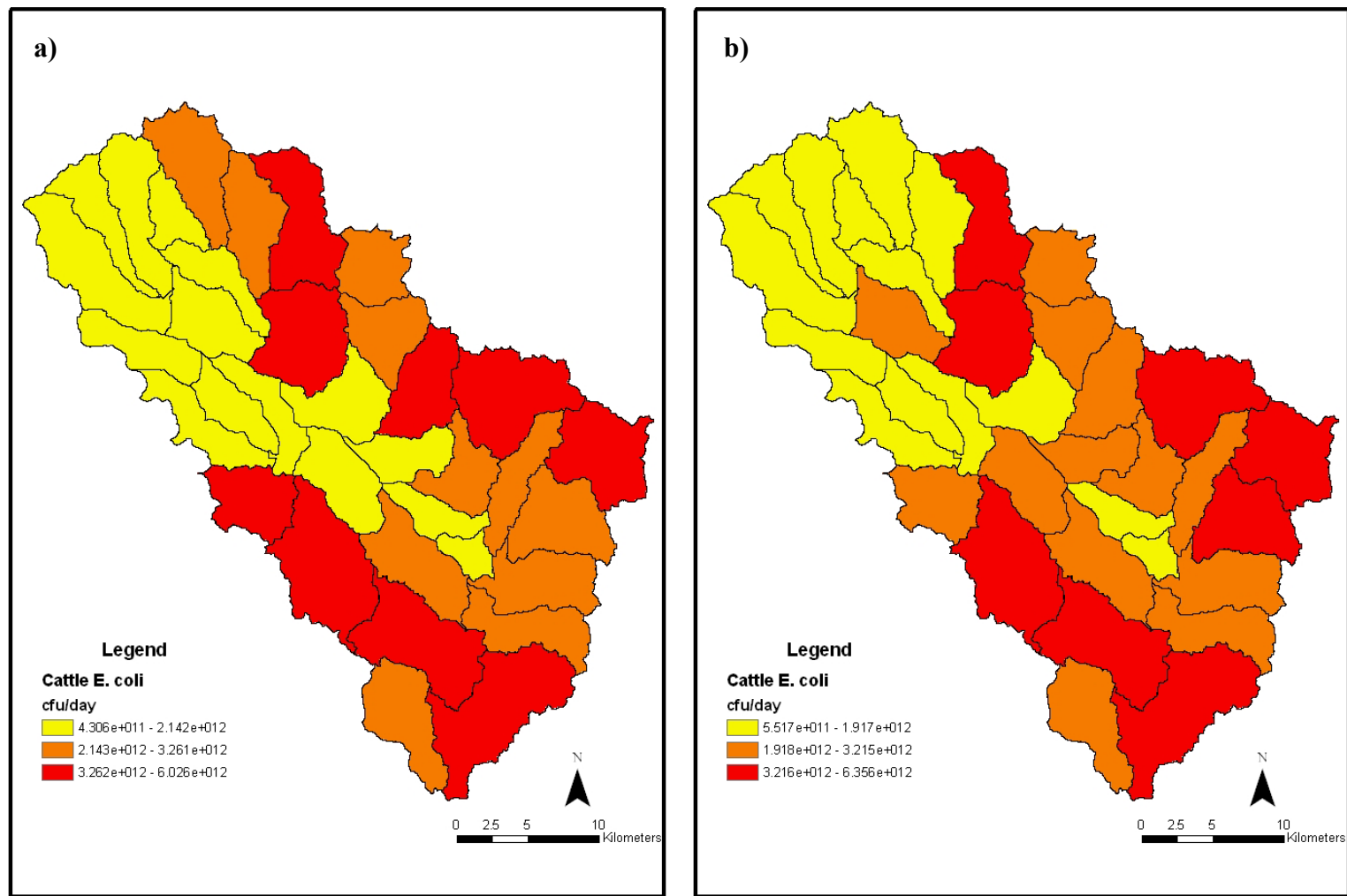
Figure 3.4. Difference in Potential *E. coli* Load Resulting from Cattle: a) Prior Study (Teague, 2007) and b) Automated.
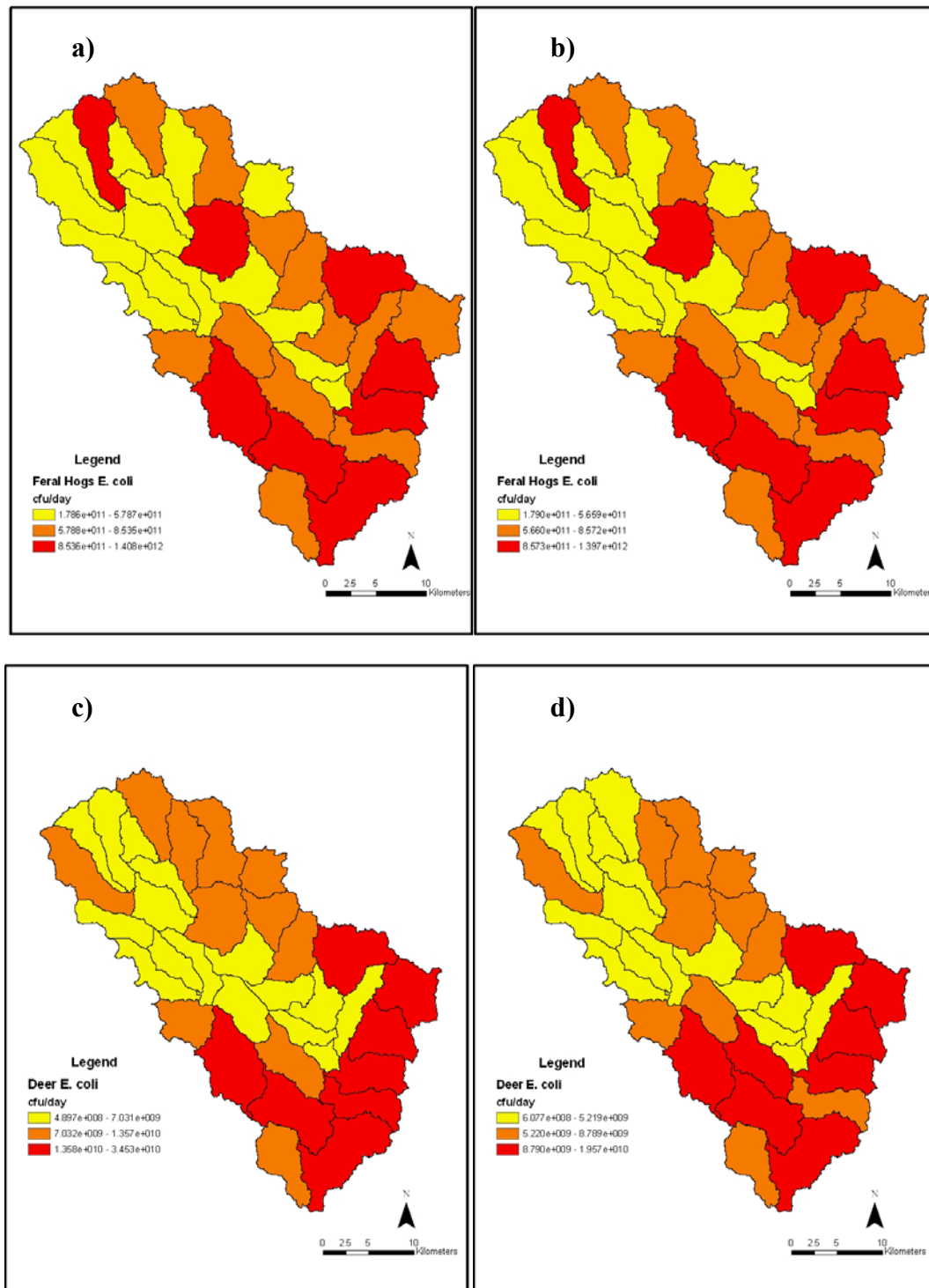
Figure 3.5. Difference in Potential *E. coli* Load Resulting from Wildlife a) Feral Hogs (Teague, 2007), b) Feral Hogs (This study), c) Deer (Teague, 2007), d) Deer (This study).
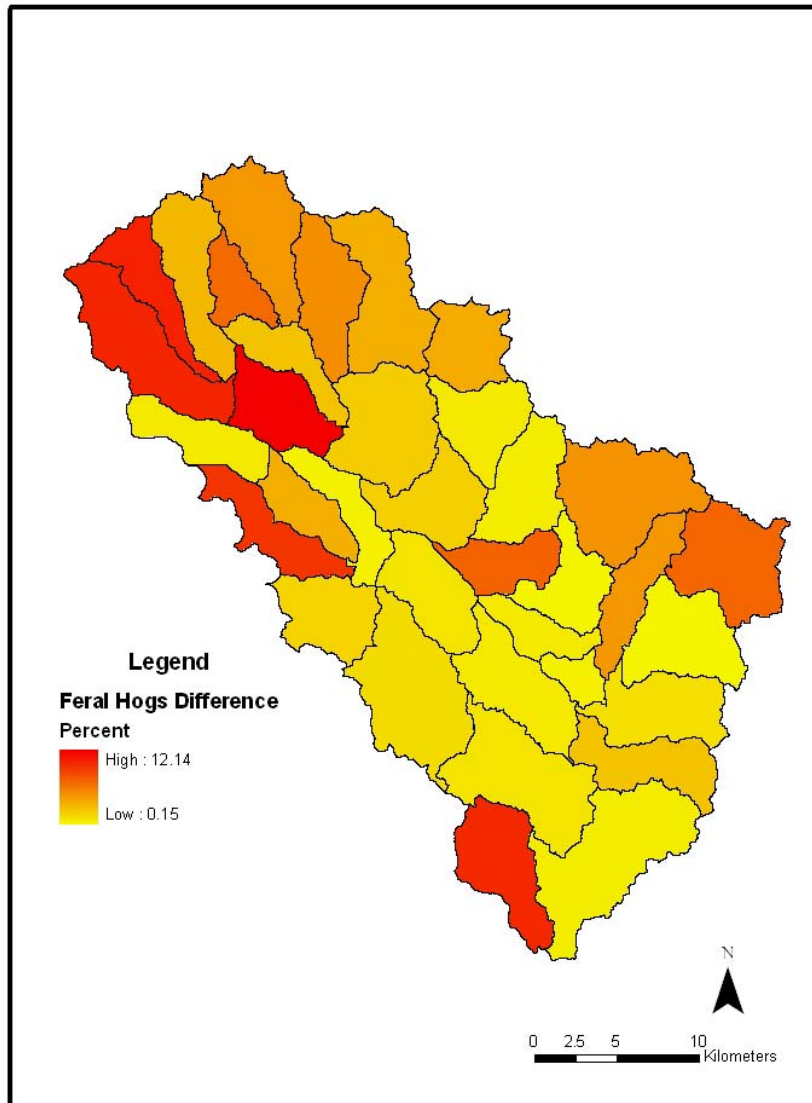
Produced by K. Riebschleager
April 2008

Figure 3.6. Difference in Potential *E. coli* Loads from Feral Hogs Between Previous Study and This Study.

Figure 3.7. Potential *E. coli* Loading from OWTS a) Previous Application (Teague, 2007) and b) Automated SELECT.

APPENDIX B
TABLES

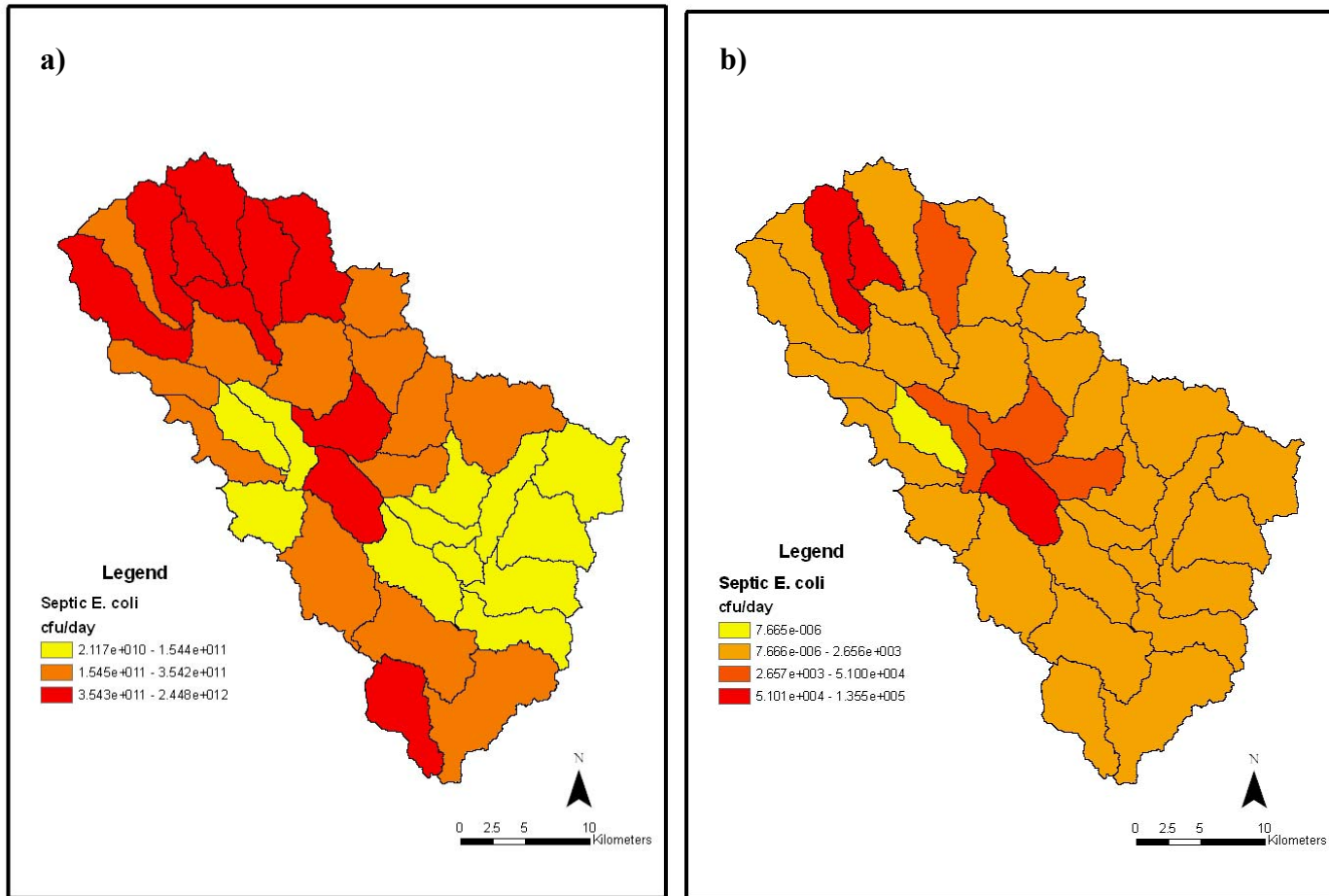**Table 2.1. Data Sources and Format Used in SELECT to Predict Potential *E. coli* Load in Lake Granbury Watershed.**

Pollutant Source	File	Format	Data Source	Comments
Livestock	Counties	Shapefile		Include only needed counties in file
	Ag inventory	Tabular	NASS	Program does not read from file
Wildlife	Suitable habitat	Shapefile	Local wildlife census	Needed for Method 1
	Urban areas	Shapefile	TIGER Census	Method 2 (Optional)
	Streams	Shapefile	NHD plus	Method 2: Feral Hogs
OWTS	Subdivisions	Shapefile	Appraisal District	Method 1: Need Age and No. of Permit Records Fields
	Census Blocks	Shapefile	Tiger Census	Method 2: Merged for all counties
	Demographics	Tabular	Tiger Census	Method 2, state demo. table
	Soils	Shapefile	SSURGO	Separate for Each County
	Soil Properties	Tabular	SSURGO	" "
Pets	Census Blocks	Shapefile	Tiger Census	Separate for Each County
	Demographics	Tabular	Tiger Census	State Census Block Demographics Table
WWTP	Outfall locations	Shapefile	state regulatory agency	Remove non-pathogenic outfalls and inactive permits
	Permitted discharge	Field in Shapefile	EPA Envirofacts Warehouse	Create field in outfall locations file

Table 2.2. Calculation of *E. coli* Loads from Source Populations.

Source	Calculation
Cattle	$EC = \#Cattle * 10 * 10^{10} cfu / day * 0.5$
Deer	$EC = \#Deer * 3.5 * 10^{8} cfu / day * 0.5$
Dogs	$EC = \#Households * \dfrac{0.8 dogs}{Household} * 5 * 10^{9} cfu / day * 0.5$
Malfunctioning OWTS	$EC = \#OWTSs * MalfunctionRate * \dfrac{1x10^{6} cfu}{100mL} * \dfrac{60 gal}{person / day} * \dfrac{Ave\#}{Household}$ $* \dfrac{3758.2mL}{gal} * 0.5$
WWTP	$EC = PermittedMGD * \dfrac{126 cfu}{100mL} * \dfrac{10^{6} gal}{MGD} * \dfrac{3758.2mL}{gal}$

Table 2.3. Interpretative Soil Properties and Limitation Classes for Septic Tank Soil Absorption Suitability (Source: SCS, 1986).

Interpretive Soil Property	Limitation Class		
	Slight	**Moderate**	**Severe**
Total Subsidence (cm)	--	--	>60
Flooding	None	Rare	Common
Bedrock Depth (m)	> 1.8	1-1.8	< 1
Cemented Pan Depth (m)	> 1.8	1-1.8	< 1
Free Water Occurrence (m)	> 1.8	1-1.8	< 1
Saturated Hydraulic Conductivity (µm/s)			
Minimum 0.6 to 1.5 m[a]	10-40	4-10	< 4
Maximum 0.6 to 1 m[a]			> 40
Slope (Pct)	< 8	8-15	> 15
Fragments > 75 mm[b]	< 25	25-50	> 50
Downslope Movement			[c]
Ice Melt Pitting			[c]
Permafrost			[d]

[a]0.6 to 1.5 m pertains to percolation rate; 0.6 to 1 m pertains to filtration capacity

[b]Weighted average to 1 m.

[c]Rate severe if occurs.

[d]Rate severe if occurs above a variable critical depth (see discussion of the interpretive soil property).

Table 2.4. Age Rating for Subdivisions in Lake Granbury Watershed to Calculate OWTS Index.

Age (years)	Age Rate
0 – 15	1
16 – 30	2
> 30	3
No Data	-99

Table 2.5. OWTS Index Reclassification to Percent Malfunction Used in Determining OWTS Malfunction Rates in Lake Granbury Watershed.

Index	Percent Malfunction
< 0	8
0 - 1.5	5
1.5 - 2.5	10
2.5 – 3	15

Table 2.6 Weighting Scheme for Sensitivity Analyses of Pollutant, Runoff, and Distance Indicators for Determining the Pollutant Connectivity Factor (PCF).

Trial Number	W_p	W_r	W_d
1	5	3	2
2	5	2	3
3	4	4	2
4	4	3	3
5	4	2	4
6	3	5	2
7	3	4	3
8	3	3	4
9	3	2	5
10	2	5	3
11	2	4	4
12	2	3	5
13	3.33	3.33	3.33

Table 3.1. Calculation of *E. coli* Loads from Source Populations.

Source	Calculation
Cattle	$EC = \#Cattle * 2.7 * 10^9\, cfu\, /\, day$
Horses	$EC = \#Horses * 2.1 * 10^8\, cfu\, /\, day$
Sheep & Goats	$EC = \#Sheep * 9 * 10^9\, cfu\, /\, day$
Deer	$EC = \#Deer * 1.75 * 10^8\, cfu\, /\, day$
Feral Hogs	$EC = \#Hogs * 4.45 * 10^9\, cfu\, /\, day$
Dogs	$EC = \#Households * \dfrac{0.8 dogs}{Household} * 2.5 * 10^9\, cfu\, /\, day$
Malfunctioning OWTS	$EC = \#OWTSs * MalfunctionRate * \dfrac{5x10^5\, cfu}{100mL} * \dfrac{70gal}{person/day} * \dfrac{Ave\#}{Household} * \dfrac{3758.2mL}{gal}$
WWTP	$EC = PermittedMGD * \dfrac{126cfu}{100mL} * \dfrac{10^6\, gal}{MGD} * \dfrac{3758.2mL}{gal}$

Table 3.2. Livestock Inventory Populations by County

	County Populations		
Source	Caldwell	Hays	Travis
Cattle	50022	26165	31680
Sheep/Goats	945	4154	1869
Horse	1953	2191	2650

Table 3.3. Reclassification of Soil Classification to Percent Malfunction Used in Determining OWTS Malfunction Rates in Plum Creek Watershed.

Soil Classification	Percent Malfunction
Not Rated	8
Slightly Limited	5
Somewhat Limited	10
Very Limited	15

Table 3.4. Summary of Changes in SELECT Approach from Previous Study (Teague, 2007).

Source	Change in Approach	Effect of Changes on *E. coli* Load Prediction
Livestock	Allowed populations to be distributed within city limits	Similar spatial distribution, cattle 3% higher; sheep/goats 11% lower; horses 38% higher
Deer	Applied uniform population density rather than RMU specific	Similar spatial distribution, potential *E. coli* load 35% lower
Feral Hogs	None	Percent differences in load per subwatershed from 0-12%
Dogs	None	87% lower than prior study, unknown error
Malfunctioning OWTS	Malfunction Rate based on soil limitation class, number of OWTS determined from homes not on sewerage services	Much lower potential *E. coli* load predictions. Outside of Kyle, Lockhart, and Luling the potential contribution of *E. coli* loading from human sources was almost negligible.
WWTP	None	None
Urban Runoff	Excluded from assessment	
Total Effect		32% lower than the previous study

VITA

Name: Kendra Jean Riebschleager

Address: 2117 TAMU
                 College Station, TX 77844

Email Address: kjaggie06@tamu.edu

Education: B.S., Agricultural Engineering, Texas A&M University, 2006
                 M.S., Biological and Agricultural Engineering, Texas A&M
                 University, 2008

Presentations:

Teague, A., M. Babbar-Sebens, K. Riebschleager, R. Karthikeyan, R. Srinivasan, J. Jacobs, M. McFarland, N. Dictson, and D. Boellstorf. 2007. Spatially Explicit Load Enrichment Tool (SELECT) and Load Duration Curves for Identification of *E. coli* Sources in Plum Creek Watershed, Texas. ASABE TMDL Conference, San Antonio, Texas

Riebschleager, K., R. Karthikeyan, and R. Srinivasan. 2008. Development of the Spatially Explicit Load Enrichment Tool (SELECT) for Characterizing *E. coli* Potential in a Watershed. AWRA Spring Specialty Conference: GIS & Water Resources V. San Mateo, CA.

Riebschleager, K., R. Karthikeyan, and R. Srinivasan. 2008. Application of the Spatially Explicit Load Enrichment Tool (SELECT) for Characterizing *E. coli* Potential in a Watershed. ASABE Annual International Meeting. Providence, RI