BACTERIAL TOTAL MAXIMUM DAILY LOAD (TMDL): DEVELOPMENT AND EVALUATION OF A NEW CLASSIFICATION SCHEME FOR IMPAIRED WATERBODIES OF TEXAS

A Dissertation

by

SABU PAUL

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

DOCTOR OF PHILOSOPHY

December 2003

Major Subject: Biological and Agricultural Engineering

BACTERIAL TOTAL MAXIMUM DAILY LOAD (TMDL): DEVELOPMENT AND EVALUATION OF A NEW CLASSIFICATION SCHEME FOR IMPAIRED WATERBODIES OF TEXAS

A Dissertation

by

SABU PAUL

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

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Patricia K. Haan (Co-Chair of Committee)	Saqib Mukhtar (Co-Chair of Committee)
Raghavan Srinivasan (Member)	Clyde Munster (Member)
Gerald Riskowski (Head of Department)	

December 2003

Major Subject: Biological and Agricultural Engineering

ABSTRACT

Bacterial Total Maximum Daily Load (TMDL): Development and Evaluation of a New
Classification Scheme for Impaired Waterbodies of Texas. (December 2003)
Sabu Paul, B.Tech., Kerala Agricultural University, India;
M.Tech., Indian Institute of Technology, India
Co-Chairs of Advisory Committee: Dr. Patricia K. Haan
Dr. Saqib Mukhtar

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) listed 110 stream segments with pathogenic bacteria impairment in 2000. The current study was conducted to characterize the watersheds associated with the impaired waterbodies. The main characteristics considered for the classification of waterbodies were designated use of the waterbody, land use distribution, density of stream network, average distance of a land of a particular use to the closest stream, household population, density of on-site sewage facilities (OSSF), bacterial loading due to the presence of different types of farm animals and wildlife, and average climatic conditions. The availability of observed in-stream fecal coliform bacteria concentration data was evaluated to obtain subgroups of data-rich and data-poor watersheds within a group. The climatic data and observed in-stream fecal coliform bacteria concentrations were analyzed to find out seasonal variability of the water quality. The watershed characteristics were analyzed using the multivariate statistical analysis techniques such as factor analysis/principal component analysis, cluster analysis, and discriminant analysis. Six groups of watersheds were formed as result of the statistical analysis. The main factors that differentiate the clusters were found to be bacterial contribution from farm animals and wildlife, density of OSSF, density of households connected to public sewers, and the land use distribution. Two watersheds were selected each from two groups of watersheds. Hydrological

Simulation Program-FORTRAN (HSPF) model was calibrated for one watershed within each group and tested for the other watershed in the same group to study the similarity in the parameter sets due to the similarity in watershed characteristics. The study showed that the watersheds within a given cluster formed during the multivariate statistical analysis showed similar watershed characteristics and yielded similar model results for similar model input parameters. The effect of parameter uncertainty on the in-stream bacterial concentration predictions by HSPF was evaluated for the watershed of Salado Creek, in Bexar County. The parameters that control the HSPF model hydrology contributed the most variance in the in-stream fecal coliform bacterial concentrations corresponding to a simulation period between 1 January 1995 and 31 December 2000.

Dedicated to…

- my family
- my teachers

ACKNOWLEDGEMENTS

I am indebted to Dr. Patricia K. Haan for her help during various stages of this research. She has been a great inspiration and support since she agreed to serve as my advisor and co-chair of my dissertation committee. I greatly appreciate all her help, especially for taking the time to review the dissertation manuscript many times. I am grateful to Dr. Raghavan Srinivasan for his invaluable help during this project. He is due a lot of credit for my success in my research and in securing my future job. I would like to express my sincere appreciation for all the help by my co-chair, Dr. Saqib Mukhtar. My thanks go to Dr. Clyde Munster for serving as a member on my dissertation committee.

Special thanks are due to Dr. Joaquin Sanabria of Blackland Research Center for helping me with cluster analysis and SAS. I would like to express my sincere thanks to Kerry Neimann and Judy Allen of Texas Commission on Environmental Quality for helping me with data collection. I am grateful to Paul Cocca of the Environmental Protection Agency and Tom Jobes of Aqua Terra Consultants for helping me overcome the difficulties with the HSPF model. I would also like to recognize the help from Dr. John Doherty of Watermark Numerical Computing.

Another person who greatly influenced me after coming to Texas A&M University is my earlier advisor Dr. Marty D. Matlock. Special thanks are due to my teachers at the Indian Institute of Technology, Dr. R. Singh, Dr. S. N. Panda, and Dr. N. Raghuwanshi for inspiring me to come to Texas A&M University.

I would like to thank my parents, my brother and his family, and my sister and her family for their guidance and support. Nothing would have been possible without your help.

I greatly appreciate the help and support from my colleagues Balaji Narasimhan, Jeff Ullman and Matthew Murawski. My friends Kiran, Preetam, Ketan, Aman and Sapna made my life in College Station beautiful. Also my roommates Sanjeev, Arjun, Jayant and Abhay made my last few days in College Station really enjoyable.

All the people mentioned above and many others helped me directly and indirectly at various stages during the past few years. But, without the blessing of my Lord my God, nothing would have been possible and for that I am indebted to Almighty.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

According to the Code of Federal Regulations (CFR), Title 40, Part 131, all States, Territories, and authorized Tribes of the United States must update the impaired or threatened waterbodies list under their jurisdiction, once in two years. The above list is called the Clean Water Act (CWA) §303(d) list and should be prepared and submitted to the Environmental Protection Agency (EPA) for their approval (USEPA, 2001c, 1998a). These authorities must also set the priority rankings for the listed waterbodies considering the severity of the pollution and the designated uses of the waterbodies. Once the list is prepared and the priority rankings are set, the next step is to develop Total Maximum Daily Loads (TMDLs) for each one of these waterbodies in order to restore the health of these impaired waterbodies.

USEPA (2001c) defines Total Maximum Daily Load (TMDL) as "the sum of the individual waste load allocations for point sources and load allocations for nonpoint sources and natural background with a margin of safety". The following equation describes TMDL as:

$$TMDL = \sum WLA + \sum LA + MOS \tag{1.1}$$

where, WLA is the waste allocation from existing or future point sources, LA is the load allocation from existing or future nonpoint sources and natural background, and MOS is the margin of safety. The units of a TMDL are usually expressed in terms of mass per time.

This dissertation follows the style and format of *Transactions of the ASAE.*

Among the many pollutants that require the development of TMDLs, fecal coliform is included because the presence of high concentration of fecal coliform bacteria indicates the presence of pathogens. Fecal coliforms are a group of bacteria that primarily live in the lower intestines of all warm-blooded animals and humans. Many water-borne diseases, like gastroenteritis and cholera, are associated with certain strains of *E.Coli*, which is but one category of fecal coliforms. Thus, a high concentration of the bacteria in a body of water normally indicates a serious potential health risk. Fecal coliforms are only associated with humans and animals. The presence of fecal coliform bacteria in ambient water is a result of the overflow of domestic sewage or nonpoint sources of human and animal waste and hence an indication of contamination of water with the fecal material of human or other animals.

Under the CWA program the Texas Commission on Environmental Quality listed 110 stream segments in the year 2000 having pathogenic bacteria concentrations more than those permitted for their designated uses (TNRCC, 2000). The next logical step is to verify impairment and if sufficient evidence is present then develop TMDLs for each of these impaired waterbodies. However, developing TMDLs for each and every one of these stream segments will require an enormous amount of input, both in terms of capital and human labor. A case study conducted by EPA showed that the cost of a single TMDL study varied between $4,039 and $1,023,531 (USEPA, 1996c). It was pointed out that on an average 32% of the total expense was allotted for the modeling component of the TMDL studies.

In general, the factors that affect the prioritization of waterbodies for TMDL development are the severity of risk to human health and the aquatic community, impairment to the waterbody, and resource value of the waterbody to the public (USEPA, 1995). In the 1980s the Oklahoma Conservation Commission used a numerical index method for ranking 300 watersheds delineated for nonpoint source assessment based on beneficial use factor, human use factor, and ecological value of the waterbody (USEPA, 1995). The current study classifies impaired waterbodies in Texas into different groups based on land use distribution, density of stream network, average distance of a particular land use to the closest stream, household population, density of on-site sewage facilities (OSSFs), bacterial loading rates due to the presence of different types of animals, and average climatic conditions. The availability of observed in-stream fecal coliform bacteria concentration data is also evaluated to obtain subgroups of data-rich and data-poor watersheds within a group.

Potential pathogen sources can be either point sources or nonpoint sources. Major point sources of pathogens are the discharges from waste water treatment plants (WWTPs), combined sewer overflows (CSOs), confined animal feeding operations, slaughterhouses and meat processing facilities, tanning, textile, and pulp and paper factories; and fish and shellfish processing facilities (USEPA, 2001c). Nonpoint sources of pathogens include urban litter, contaminated refuse, domestic pet and wildlife excrement, failing sewer lines in urban and suburban areas and confined animal operations, excrement from barnyards, pastures, rangelands, feedlots and uncontrolled manure storage areas in rural or agricultural areas. Parker and Lahlou (2001) found that point sources tend to have the most profound influence on receiving waterbodies during dry conditions. They also pointed out that nonpoint sources like runoff from urban and agricultural areas had the most effect during a storm event. Bacterial densities also have a high correlation with the human population, density of housing, and land development (Young and Thackston, 1999).

Water quality criteria differ based on the designated uses of the waterbody. The numerical water quality criteria for waterbodies are set based on whether the waterbodies are used for contact recreation, shellfish harvesting or as a public drinking water source (USEPA, 1986). Criterion for contact recreation is also dependent on whether the waterbody is used for primary contact recreational activities like swimming or for secondary recreational activities like boating. In addition, prioritization for TMDL development for a given stream segment will be driven by its designated use. It is therefore appropriate to separate listed stream segments into different groups based on their designated use.

One of the main steps in a TMDL study is to have a proper link between the water quality targets and pollutant sources. Even though real-time monitoring of water quality constituents is the best way to assess the water quality of a stream network, this method is very expensive and laborious. An alternative to this approach is to model the environment to predict pollutant loads and in-stream concentrations. There are a number of models available for water quality studies, but their ability to model bacteria under different situations varies widely (USEPA, 1991b; 1997a). The selection of an appropriate model for a given group of stream segments is a difficult task and depends on many factors such as their ability to handle different types of land uses and the pollutant sources under consideration. The selection of such a model also depends on factors such as simplicity in using the model, data requirements, accuracy in system representation, and cost of the model.

All TMDL studies involve prediction of water quality with the help of water quality models. The selection of the model for TMDL development will be influenced by the frequency of water quality violation. This is based on the fact that different sources have different effects on the timing of the water quality impairment. Water quality violations occurring during low-flow, dry weather conditions points to continuous loading point sources (USEPA, 2001c). If the water quality violation is associated with storm events then there is a high probability that it is the result of nonpoint sources. Water quality violations associated with storm events can be modeled using an event-based model, whereas frequent water quality violation can only be addressed with the help of a more complex continuous simulation model.

A one-dimensional model can accurately represent a waterbody like a well- mixed flowing river. If the waterbody under consideration is a lake or a deep reservoir, where there is vertical stratification and mixing, then a two-dimensional model should be considered. Similarly, three-dimensional models should be the choice if the waterbody is an estuary or near the coast and influenced by tides (USEPA, 1997a). Therefore the location of the waterbodies is considered during the classification process.

The prediction of water quality will be valid only when the model is well calibrated with observed concentrations of the constituent of interest. Once the model is properly calibrated we could assume that the water quality targets are properly linked to the pollutant sources. However, the validity of a TMDL study will be at stake if the number of water quality observations is not sufficiently large to have a proper comparison between model predictions and observations. In addition, data requirements for the calibration of different models vary widely.

There are a number of water quality models available at varying spatial scales that are generally used for TMDL development studies. Some of these models can be used at a watershed level, whereas others can be used at a receiving waterbody level. Some of the watershed scale models, which are capable of addressing bacterial pollution, are AUTO-QI, HSPF, SLAMM, and SWAT (USEPA, 2001c; 1997a; 1991b). Similarly CE-QUAL-ICM, CE-QUAL-RIVl, CE-QUAL-W2, QUAL2E, and WASP are examples of some receiving water models that can be used for bacterial studies. The capabilities of these models in handling various sources and media are quite different. Models such as AUTO-QI, SLAMM and SWMM are generally used for urban land uses, while SWAT is efficient in dealing with agricultural watersheds. Models like HSPF can address both urban and agricultural land uses. Similarly, HSPF and CE-QUAL-RIVl can model well mixed flowing rivers quite efficiently, but they are incapable of handling waterbodies like deep lakes and reservoirs. Water quality models like CE-QUAL-W2, and CE-QUAL-ICM are good for simulating the processes that take place in lakes.

In general water quality modeling tools can be either used to characterize the magnitude of the problem or to analyze the problem. The selection of a model is dependent on the magnitude of the problem. Nix (1990) gives criteria for selection of model, including, hardware availability, availability of trained personnel, long-term usability of the model, experience of other modelers in using the model, technical support available, and commitment to the modeling process. Nix (1990) also points out that the use of a maladapted model may lead to complications of the problem because of misleading results. Though complex models tend to be effective in reproducing the processes of interest, they are highly dependent on the input data. Because of this dependency, usage of these complex models is highly limited by data availability.

In recent years USEPA has emphasized the importance of incorporating variability and uncertainty in risk assessment (USEPA, 1997b). They pointed out that probability analysis techniques like Monte Carlo analysis are useful tools in adequately quantifying variability and uncertainty (Chang, 1999).

Hession et al. (1996a and 1996b) used a two-phase Monte Carlo procedure to propagate uncertainty in modeling natural processes, based on knowledge uncertainty and stochastic variability. According to the authors a parameter was assumed to have knowledge uncertainty when the actual value of that parameter is not known prior to modeling, hence a range of possible values were obtained from the literature and used in the model. Stochastic variability was considered when the parameter values have known temporal or spatial variability and they follow a specific probability distribution function. Chang (1999) in his review of the plans of US EPA's Office of Emergency and Remedial Response (OERR) to implement Probabilistic Risk Assessment (PRA) described a concept of choosing a range of plausible values for the input parameters instead of considering a point estimate. Variability and uncertainty of the parameters are explicitly handled in the PRA technique by providing a means to obtain risk estimates for the individual parameter values and also to quantify the level of uncertainty in these risk estimates. Suter (1999) presented a strategy for creating conceptual models that are useful in risk assessment studies, which included components such as explicit mechanistic models, compartments based on functional groups, exposure-response relationships, hierarchies of details and modular components. Crosetto et al. (2000) pointed out the importance of performing uncertainty analysis and sensitivity analysis for the improvement of modeling in general and GIS-based models in particular.

Among the different models that could be used for modeling fecal coliform, HSPF (Bicknell et al., 1997) has successfully been used for many fecal coliform TMDL studies (USEPA, 1997a). Therefore, in the current study HSPF was used in a case study to evaluate the uncertainty in predicted coliform bacteria concentrations due to uncertainty in the most influential model parameters. The watershed corresponding to Salado Creek in the San Antonio River basin, upstream of USGS gauging station 08178800, was used as the case study area. Sensitivity analysis was used to find the input parameters that when changed have the greatest effect on HSPF predicted fecal coliform concentrations. The sensitivity of a parameter is the ratio of the change in the model output resulting from a change in an input parameter to the change in the input parameter, all other parameters being unchanged. All parameters that affect fecal coliform concentration or mortality were examined to find their sensitivities. First Order Approximation and Monte Carlo simulation were used to determine the effect of uncertainty in the most sensitive parameters on the uncertainty in HSPF in-stream bacterial concentration predictions.

Objectives and hypothesis

The current study focuses on the development of a classification scheme for stream segments based on their watershed's characteristics, the possible sources of pollution and the extent of water quality data collected. This scheme will be helpful in reducing the cost of restoration of water quality by restricting the development of TMDLs to one or two representative streams under a single group of stream segments. The specific objectives of this dissertation are:

1. Develop a method to classify the Texas waterbodies listed for bacterial quality violation under CWA§303(d) into groups having similar characteristics
2. Select an appropriate water quality prediction model for each of these groups and validate the selection of the model
3. Evaluate the parameter uncertainty in the water quality model on predictions of coliform bacterial concentration

The hypothesis of the research is: *Impaired waterbodies in Texas can be sub-divided into groups based on similarities of the waterbodies for which a single TMDL can be developed.*

Significance of the research

It is possible that many of the waterbodies considered for TMDL development listed under the current CWA §303(d) for Texas may be grouped based on their watershed characteristics and the possible sources of pollution. Such a grouping scheme will be helpful in reducing the cost of restoration of water quality by restricting the development of TMDLs for only one or two representative waterbodies under a single group and applying the knowledge to other waterbodies in the same group. Since the modeling of in-stream bacterial concentration is one of the major components of a bacterial TMDL, the selection of an appropriate model for a particular TMDL study will greatly enhance the validity of the specific TMDL. The inclusion of Monte Carlo simulation technique will help in quantifying the risk associated with parameter uncertainty in a better way than the current practice of adding an arbitrary margin of safety (MOS). The results of Monte Carlo Simulation will be useful in predicting the percent chance of exceeding a given level of in-stream bacterial concentrations for a given type of land use distribution, animal management scenario, and a population pattern. A better understanding of the effect of parameter uncertainty on the model results will be helpful in the decision making process.

CHAPTER II

COMPARISON BETWEEN IN-STREAM BACTERIAL CONCENTRATIONS DURING RAINFALL EVENTS AND BASEFLOW PERIODS IN TEXAS

Synopsis

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) currently lists 110 stream segments with pathogenic bacteria impairment. A study was conducted to characterize the watersheds associated with these impaired waterbodies based on characteristics such as designated use of the waterbody, sources of pathogens, frequency of water quality violation, location and type of the waterbody, and size of the watershed. In-stream bacterial concentration data obtained from the TCEQ corresponding to 1900 water quality stations, stream flow data from 165 US Geological Survey (USGS) gauging stations and precipitation data from 903 National Climatic Data Center (NCDC) weather stations for a period between 1985 and 2000 were statistically analyzed to evaluate the effect of rainfall within the watersheds on in-stream bacterial concentrations. In-stream bacterial concentration data were separated for baseflow and stormflow periods. Partitioning of the data was either based on flow conditions in the streams or based on the average rainfall over the area contributing flow to the water quality stations. A pooled T test was used to evaluate significant differences in in-stream bacterial concentrations between baseflow and stormflow periods. In general, 67 stream segments showed higher mean bacterial concentrations during stormflow periods and 21 stream segments showed higher mean bacterial concentrations during baseflow periods. Eleven stream segments showed no significant difference in the means of bacterial concentrations during stormflow and baseflow periods. There was not enough data for analysis for 11 stream segments. The results of the analysis were used to determine the nature of the

source of bacterial pollution. A consistently higher concentration of bacteria during low-flow periods is an indication of loading from a continuous point source, whereas an increase in bacterial concentration during rainfall events indicates loading from various nonpoint sources.

Introduction

According to the Code of Federal Regulations (CFR), Title 40, Part 131, all States, Territories, and authorized Tribes of the United States must update the impaired or threatened waterbodies list under their jurisdiction, once in two years. The list of impaired waterbodies is called the Clean Water Act (CWA) §303(d) list and is prepared and submitted to the Environmental Protection Agency (EPA) for their approval (USEPA, 2001c). Regulatory authorities must also set priority rankings for the waterbodies based on the severity of the pollution and the designated use. Once priority rankings are given to waterbodies in the list, the next step is to develop Total Maximum Daily Loads (TMDLs) for each of these waterbodies in order to restore the health of the waterbodies. A TMDL is the maximum pollutant load a waterbody can receive without violating water quality standards (Hession et al., 1996c). In other words it is the sum of wasteload allocations for all point sources, load allocations for all nonpoint sources, background loadings from natural sources, and a margin of safety to ensure achievement of the water quality standard (USEPA, 1991a).

Among the many pollutants that require the development of TMDLs, fecal coliform is included because it indicates a serious potential health risk. Fecal coliforms are a group of bacteria that primarily live in the lower intestines of warm-blooded animals and humans. The presence of high concentrations of fecal coliform bacteria indicates the presence of dangerous pathogens. Under the CWA program the Texas Commission on Environmental Quality (TCEQ) listed 110 waterbodies (Figure 2.1, stream segments listed in Table A1 in Appendix) in the year 2000 having pathogenic bacteria concentrations more than those permitted for their designated uses (TNRCC, 2000). The development of TMDLs for these waterbodies seems to be the best solution for the problem. However, developing TMDLs for each and every one of these stream segments will require an enormous amount of input, both in terms of capital and human labor. A case study conducted by the EPA showed that the cost of a single TMDL study varied between $4,039 and $1,023,531 (USEPA, 1996c). They pointed out that on average 32% of the total expense was allotted for the modeling component of the TMDL studies. It is possible that many of the waterbodies considered for TMDL development listed under the current CWA §303(d) for Texas may be grouped based on their watershed characteristics and the possible sources of pollution. Such a grouping scheme will be helpful in reducing the cost of restoration of water quality by restricting the development of TMDLs for only one or two representative waterbodies under a single group and applying the knowledge to other waterbodies in the same group.

One of the characteristics analyzed during the grouping process was the nature of the bacterial water quality violation. High concentrations of bacteria during low-flow, baseflow weather conditions usually indicate continuous loading from point sources (USEPA, 2001c). At the same time, if concentrations of bacteria are higher during storm events, then there is a high probability that the water quality violation is the result of nonpoint sources. The correlation between seasonal rainfalls and the fecal coliform concentration in watersheds was reported in the literature (Whitlock et al., 2002). They also reported a very close correlation between the average in-stream fecal coliform concentration and the cumulative rainfall measurement for 2 days prior to sample collection.

The objectives of the current study are to separate bacterial concentration observations for stormflow and baseflow periods at the watershed level and to analyze the data for the two periods to find out whether the means of concentrations during the two periods show a significant difference.

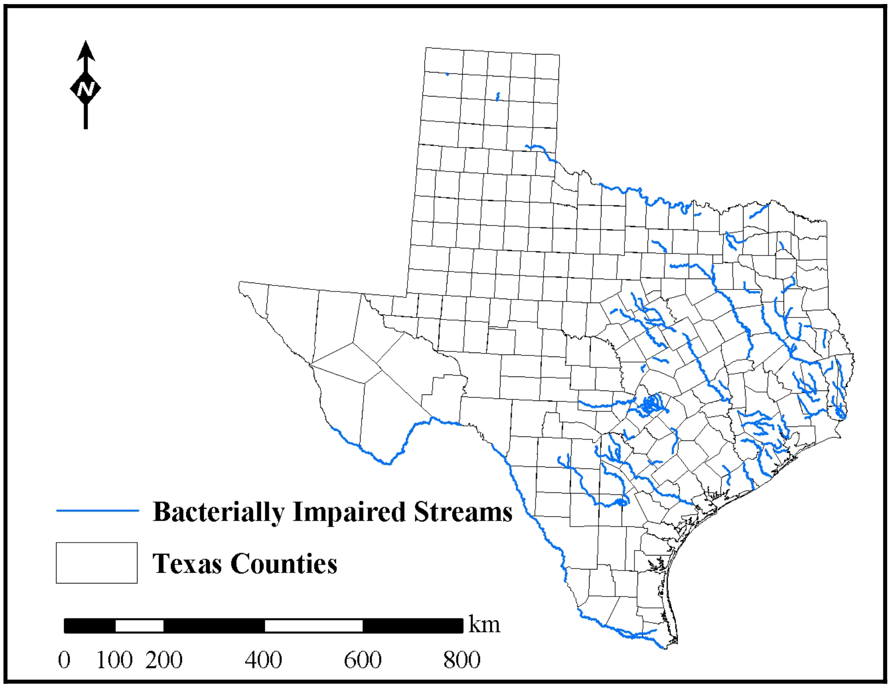

Figure 2.1 Stream segments in Texas listed for bacterial impairment.

Methodology

Overview

The effect of nonpoint sources of bacteria on the timing of water quality impairment is different from that of point sources. High concentrations of bacteria during low-flow, baseflow weather conditions usually indicate continuous loading from point sources (USEPA, 2001c). At the same time, if the concentrations of bacteria are higher during storm events, then there is a high probability that the water quality impairment is the result of nonpoint sources. Observed water quality data obtained from stations were compared with stream flow data from a USGS gauging station located at the same site. The bacterial parameter used for analysis was fecal coliform concentrations. When there was not a USGS gauging station located at the site of the water quality station, daily flow at the water quality station location was estimated using the flow from a nearby USGS gauging station. The flow at the USGS gauging location was transformed to the water quality station location by considering the hydrologic properties such as drainage area, and mean Curve Number (CN) and mean precipitation over the drainage area at both locations. The hydrologic parameter (product of drainage area, mean CN, and mean annual precipitation upstream of the location) was calculated for all the water quality stations and USGS stations. The ratio of the hydrologic parameter of the water quality station and the hydrologic parameter of the closest USGS station was calculated. When the hydrologic parameters were found to be different the approach was determined inappropriate and the precipitation data over the contributing area of the water quality station was used to determine the nature of flow at the water quality station location. The flow at the location of water quality station was produced using the flow transferring technique only if the above ratio was between 0.05 and 20, assuming that hydrological properties of the two locations were comparable. The range (0.05, 20) was selected since most of the water quality stations within this range were located very close to the selected USGS station and the flow pattern at the location was believed to be similar to that at the USGS gauging location. When the ratio value was outside the range of (0.05, 20) the separation of water quality observations into high flow-low flow periods were carried out based on the average precipitation over the contributing area.

Flow data within the period from 1985 to 2000 were statistically analyzed and baseflow was separated using the technique developed by Arnold and Allen (1999). The measured flow for each day was compared to that of the separated baseflow to determine whether the stream flow on that day was due to a rainfall event or due to the baseflow contribution. If the stream flow was found to be more than 1.1 times that of the baseflow, then it was assumed to be the result of a rainfall event, otherwise stream flow was assumed to be contributed only by baseflow. The period during which stream flow was contributed only by baseflow was designated the low-flow period or baseflow period. The period during which stream flow was more than 1.1 times baseflow was designated the high-flow period or stormflow period. Water quality observations were made using grab samples collected by the TCEQ during the period from 1985 to 2000. On each day a water quality observation was available, the stream flow and baseflow were analyzed to find out whether the water quality observation occurred during a high-flow or low-flow period. The water quality data were separated into those associated with baseflow periods and those associated with rainfall events.

When a USGS gauging station was not located near the water quality station the statistical comparison between observed in-stream fecal coliform bacteria concentrations and precipitation was similar to the comparison between observed in-stream fecal coliform bacteria concentrations and stream flow data. The periods for which mean rainfall over the contributing area of a water quality station were greater than 1mm were designated high-flow periods. If the mean rainfall was less than 1mm then that period was designated a low-flow period. The threshold value of 1mm precipitation was selected because this was large enough to produce a reasonable change in stream flow regime and small enough to be sure that no rainfall events are missed. Based on the precipitation information obtained from the National Climatic Data Center (NCDC), the water quality data were separated into days belonging to the rainfall events and days without rainfall.

The water quality stations within each watershed were identified with the help of ArcView (ESRI, 1999) Geographic Information System (GIS). Based on the flow pattern or the precipitation over the contributing area, the water quality observations for the individual stations were separated into baseflow periods (periods for which the flow at the gauging location was assumed be due to the baseflow) and stormflow periods (periods for which the flow at the gauging location was assumed be due to the a recent rainfall event). The data were compiled for all the water quality stations that fell within a single watershed. The Student's pooled T test (Milton and Arnold, 1995) was conducted on the watershed level dataset to find out whether there was a significant difference between means of bacterial concentration during rainfall related and baseflow related periods. A visual analysis using boxplots of bacterial concentration during both periods was done. This process was repeated for all the watersheds.

Water Quality Stations

A database that contains the location information for all the water quality stations in Texas that measure bacterial concentration and are close to an impaired waterbody segment was obtained from the TCEQ. Using latitude and longitude, the GIS layer for the water quality observation stations was created. TCEQ currently collects data corresponding to 7252 water quality stations available within Texas. Since the current study was limited to bacterial impairment, the stations that were located within the watersheds corresponding to the impaired stream segments were selected. Thus the current study utilized the data available for about 1900 such stations. Figure 2.2 shows the locations of the selected water quality observation stations in Texas. The water quality observations for each of these stations were also obtained from the TCEQ database and were used for the statistical analysis. The in-stream fecal coliform concentration observations for all the water quality stations within each watershed were compiled for the period from 1985 to 2000. Figure 2.3 shows the distribution of mean concentrations for 110 impaired stream segments. The values of the mean in-stream bacteria concentrations and total number of observations for the impaired stream segments are in Appendix A (Table A.1).

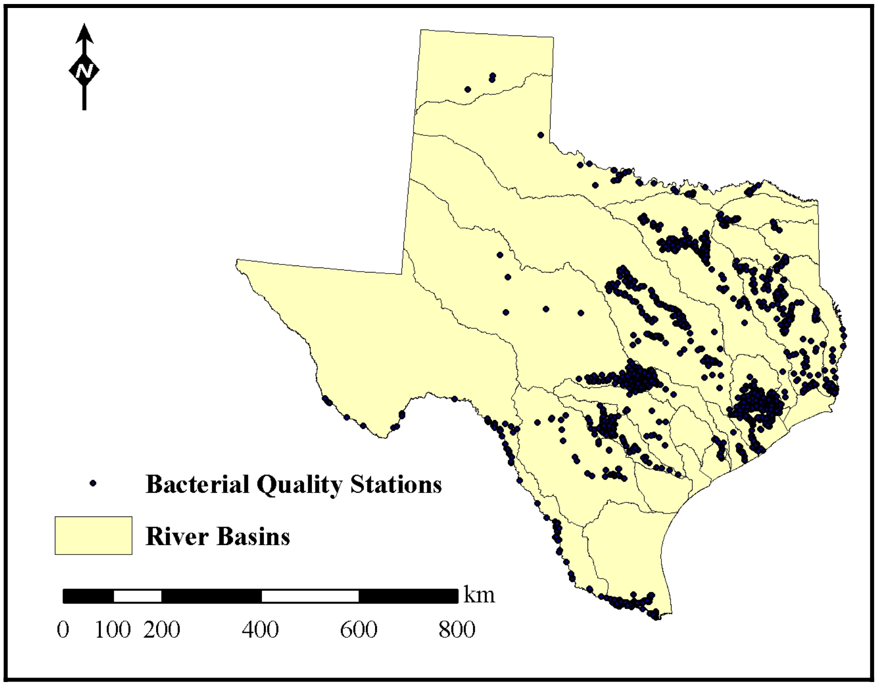

Figure 2.2 Location of selected water quality observation stations in Texas.

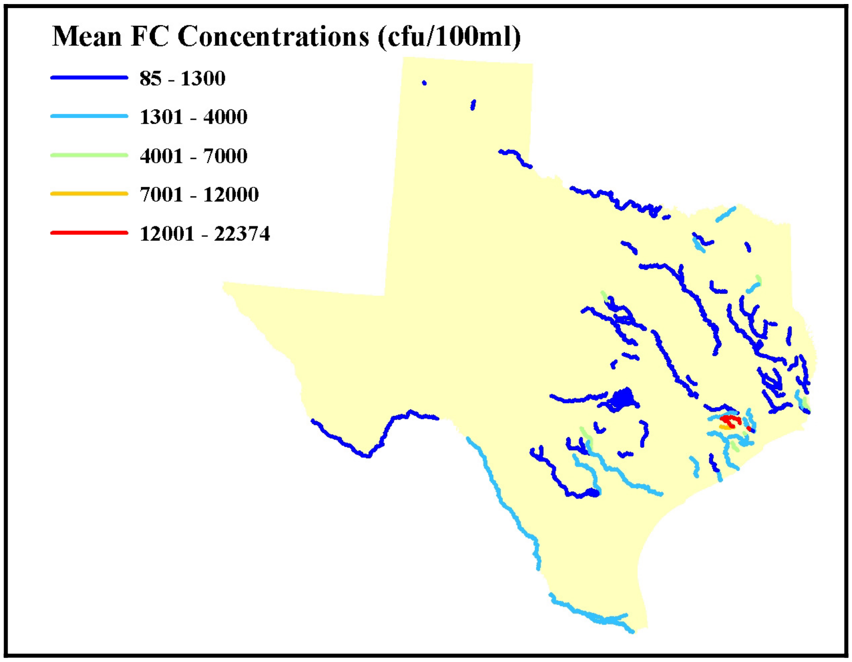

Figure 2.3 Mean fecal coliform bacterial concentrations (cfu/100ml) for impaired stream segments for duration 1985-2000.

USGS Flow Gauges

In 2002, there were 747 USGS stations located in Texas. Initially, 211 gauging stations located within the watersheds of the impaired water bodies were selected. Because of the lack of flow data during the period between 1985 January 1 to 2002 December 31, some of the USGS gauge stations were discarded. The total number of stations with full data availability was 165. Figure 2.4 shows the locations of the USGS gauging stations selected for the flow-concentration comparisons.

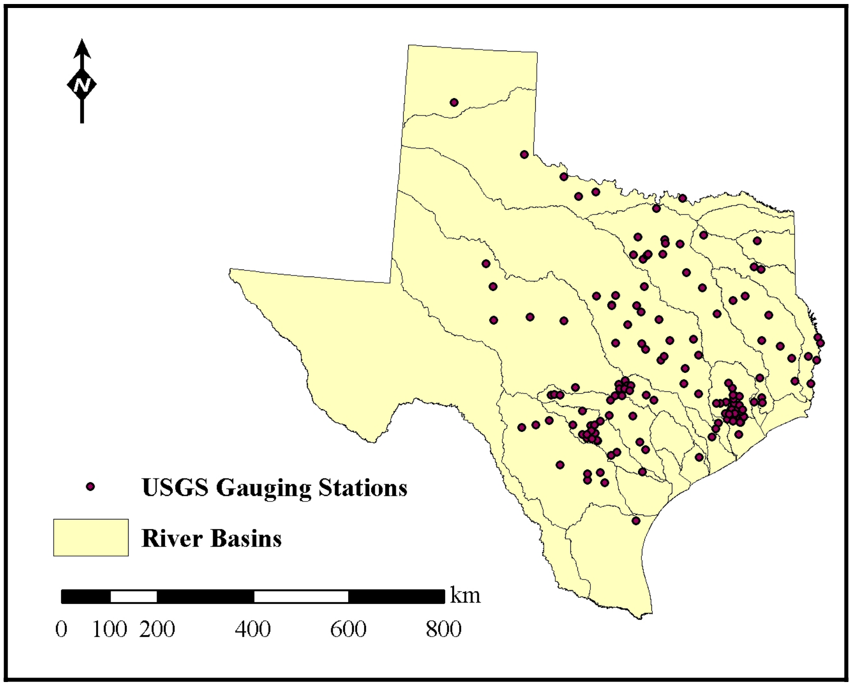

Figure 2.4 Location of selected USGS gauging stations in Texas.

Precipitation Gauges

The geographic information (latitude and longitude) of 903 weather stations were obtained from National Climatic Data Center (NCDC) and a GIS layer was created. Figure 2.5 shows the locations of the NCDC stations available within Texas in 2002. Daily rainfall data for the period between 1985 January 1 to 2000 December 31 was obtained for all of these gauge locations.

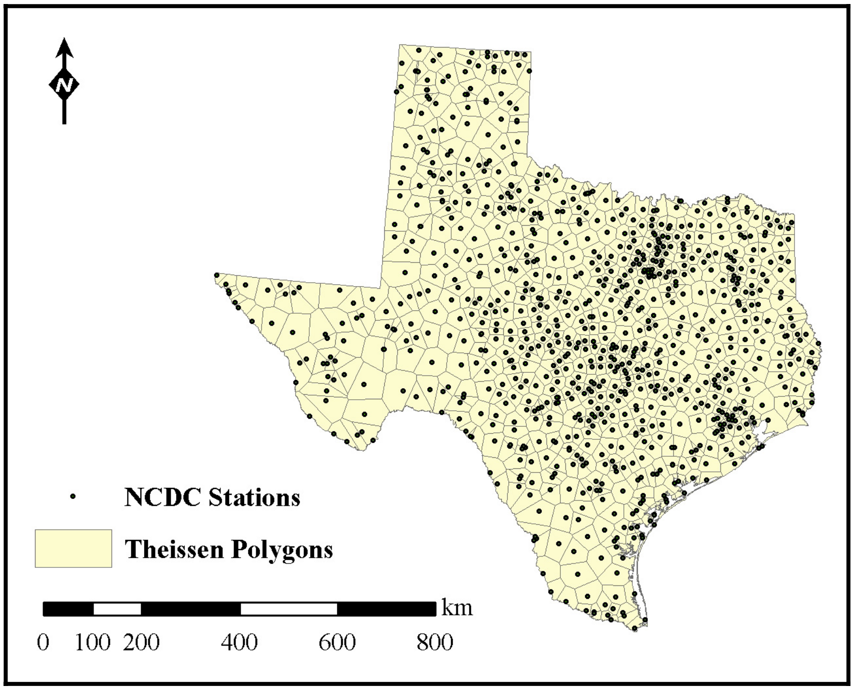

Figure 2.5 Location of NCDC raingage stations in Texas.

Estimation of Daily Streamflow at Ungauged Locations

Estimates of daily streamflow were required at the water quality observations stations which did not coincide with a USGS gauging station. The daily flows for these stations were estimated using a proportional relationship of discharge to change in hydrologic parameters with a nearby USGS gauging station (Hoffpauir, 2002). The proportional relationship between a gauged location and an ungauged location can be written as:

$$Q_U = Q_G * \frac{A_U}{A_G} * \frac{CN_U}{CN_G} * \frac{P_U}{P_G} \tag{2.1}$$

where Q is the mean daily discharge, A is the drainage area upstream, CN is the average curve number of the drainage area, P is the mean annual precipitation over the drainage area and subscripts G and U correspond to gauged and ungauged locations respectively. The assumption in using the proportional method was that the gauged and ungauged locations were close enough on the stream network so that there was a linear relationship between the incremental change in their discharge and the change in their hydrologic parameters.

Estimation of Baseflow

There are a variety of methods available to separate baseflow from stream flow (Nathan and McMahon, 1990a; Arnold et. al., 1995; Arnold and Allen, 1999). The current study used a program developed by Blackland Research Center at Temple, Texas based on a digital filter technique (Nathan and McMahon, 1990a) that was originally used in signal analysis and processing. The digital filter uses the equation,

$$q_t = \beta q_{t-1} + (1-\beta)/2 * (Q_t - Q_{t-1}) \tag{2.2}$$

where q is the filtered surface runoff, Q is the original streamflow, β is the filter parameter and t is the time step (one day). Baseflow is calculated as,

$$b_t = Q_t - q_t \tag{2.3}$$

where b_t is the baseflow at time, t.

The inputs for the baseflow separation program are initial Julian day, beginning year, drainage area (sq. miles) and the daily streamflow (cfs). The filter is passed over the streamflow data three times (forward, backward and forward).

Using the baseflow separation program, the surface runoff and baseflow were calculated for all the flow gauge locations. The actual streamflow for each day was compared to the baseflow on the same day, and the streamflow was considered to be the result of a rainfall event if $Q_t > 1.1\ b_t$.

Drainage Area Calculation

The drainage area of the gauged (USGS gauges) and ungauged locations (water quality stations that were not located at any USGS gauge) were calculated using the Digital Elevation Model (DEM) obtained from USGS. A DEM is a matrix of equally sized square cells of elevation values over an area. As an example, the DEM GRID for the Neches River Basin is shown in Figure 2.6. The flow accumulation GRID, which gives the number of cells contributing flow to any given point on the study area, was calculated using DEM GRIDs. The drainage area at a given point is equal to the flow accumulation values multiplied by the area of a single DEM cell. The flow accumulation GRIDs for each of the river basins in Texas were generated using ArcView GIS software.

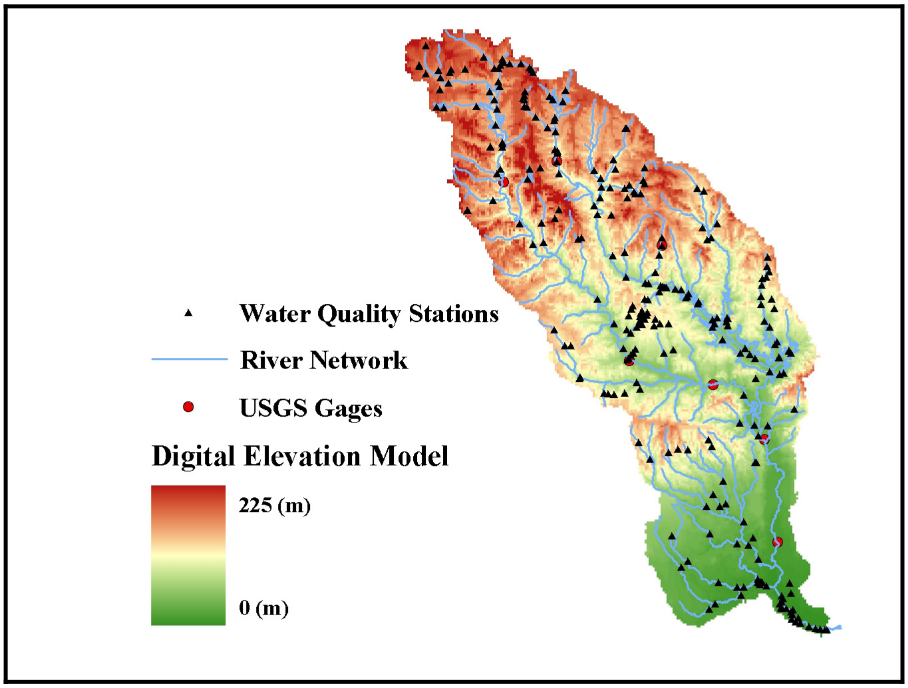

Figure 2.6 DEM, USGS gauging stations, water quality stations and river network for the Neches River Basin.

Mean Curve Number

The mean curve number (CN) over a drainage area can be calculated as the average CN, weighted by the drainage areas (A):

$$Mean\,CN_{location} = \frac{\sum_i CN_i A_i}{\sum_i A_i} \tag{2.4}$$

This calculation can be carried out in ArcView using a weighted flow accumulation function that was slightly modified as below in equation 2.5 for obtaining the CN of the current cell when the flow accumulation value is zero. This will avoid a division by zero error.

$$Mean\,CN = \frac{flowaccumulation(flowdirection, CN) + CN}{flowaccumulation(flowdirection) + 1} \tag{2.5}$$

The CN GRID for the State of Texas was obtained from the Blackland Research Center in Temple, TX. The cell resolution of the CN GRID was 250m, which was different from that of the DEM (85m, approximately). Therefore the CN GRID was resampled to the size of the DEM and used for mean CN calculations. The mean CN GRIDs for each of the river basins in Texas were generated using ArcView GIS software. Figure 2.7 shows an example of a mean CN map generated corresponding to the Neches River Basin.

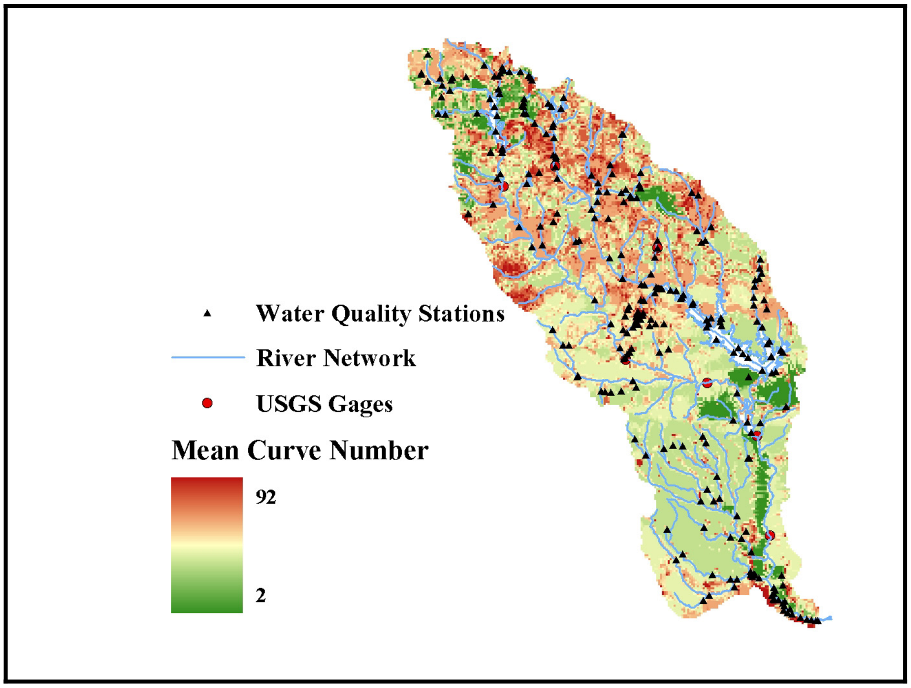

Figure 2.7 The mean CN for the Neches River Basin.

Mean Annual Precipitation

The mean annual precipitation was obtained similarly to the calculation of mean CN as:

$$Mean\ Ppn = \frac{flowaccumulation(flowdirection, Ppn) + Ppn}{flowaccumulation(flowdirection) + 1} \tag{2.6}$$

where Ppn is the average annual precipitation. The GRID layer corresponding to the annual average precipitation over Texas was obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) web site. Using the above equation, the mean precipitation GRIDs for each of the river basins in Texas were generated using ArcView GIS software. Figure 2.8 shows an example of a mean precipitation map generated corresponding to the Neches River Basin.

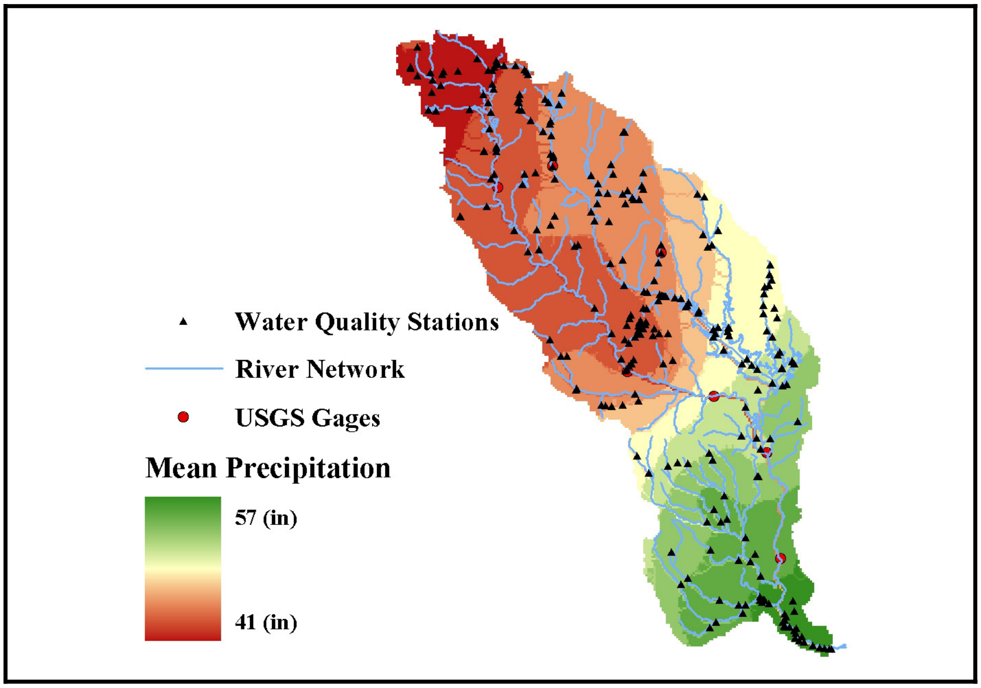

Figure 2.8 The mean precipitation for the Neches River Basin.

Comparison of Means: Statistical Tests (Kanji, 1999; Milton and Arnold, 1995)

The t-test assesses whether the means of two groups are statistically different from each other. The null hypothesis, H_0, states that the population means of bacterial concentration during baseflow period and after rainfall events are the same. In other words, the difference between the population means of bacterial concentration during the baseflow period and after rainfall events is zero. This can be written as:

$$H_0 : \mu_b = \mu_r \tag{2.7}$$

Alternative hypothesis:

$$H_\alpha : \mu_b \neq \mu_r \tag{2.8}$$

where μ_b is mean bacterial concentration during the baseflow period and μ_r is the mean bacterial concentration after a rainfall event.

The observed value of the test statistic is calculated by (Kanji, 1999; Milton and Arnold, 1995):

$$t_{stat} = \frac{(\overline{X}_b - \overline{X}_r) - (\mu_b - \mu_r)}{\sqrt{S_p^2 (1/n_b + 1/n_r)}} \tag{2.9}$$

where $\overline{X}_b$ and $\overline{X}_r$ are means of the observed bacterial concentrations during baseflow period and after a rainfall event, respectively, $(\mu_b - \mu_r)$ is the hypothesized difference in population means (in current study the value of hypothesized difference is zero), n_b and n_r are the number of observations during the baseflow period and after a rainfall event, respectively and S_p^2 is the pooled variance given by:

$$S_p^2 = \frac{(n_b - 1)S_b^2 + (n_r - 1)S_r^2}{n_b + n_r - 2} \tag{2.10}$$

where S_b^2 and S_r^2 are the sample variances of bacterial concentrations during the baseflow period and after a rainfall event, respectively. The p-value is the probability of observing a difference equal to or higher than the observed difference if the null hypothesis is true. A very small p-value means the difference between the sample means is unlikely to be a coincidence. Thus the null hypothesis can be rejected when the p-value is very small.

Results and discussion

The stream flow data were analyzed to identify baseflow periods and rainfall events. Using the baseflow filter program the baseflow was separated from the stream flow and the stream flow was compared to the calculated baseflow for each day. For example, Figures 2.9 and 2.10 show the daily observed stream flow and the calculated baseflow for the USGS station 08180800. The local peaks in the flow variable in Figure 2.9 are the result of recent rainfalls and they represent the period of high-flow. A careful observation of Figure 2.10 will show that stream flow during 1999 June 14 to 1999 July 31 was considerably higher compared to the baseflow for the same period. The hydrograph also exhibits multiple local peaks during this period. Hence the flow during 1999 June 14 to 1999 July 31 can be interpreted as the result of recent rainfall events.

For water quality stations without any associated flow data and without any nearby USGS gauging stations, the average precipitation over the contributing area was calculated to find out the flow pattern at those water quality stations. Both one-day mean and three-day mean precipitation over the area were calculated. For example Figure 2.11 shows the mean of one-day and three-day rainfalls over the contributing area of water quality station 17606.

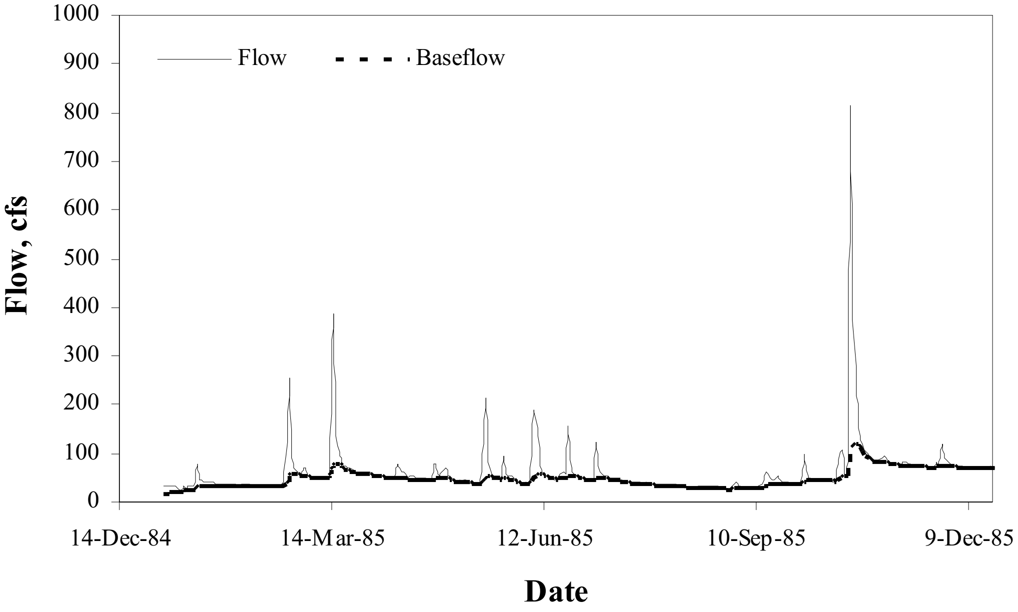

Figure 2.9 Baseflow separation results for USGS station 08180800 for the year 1985.

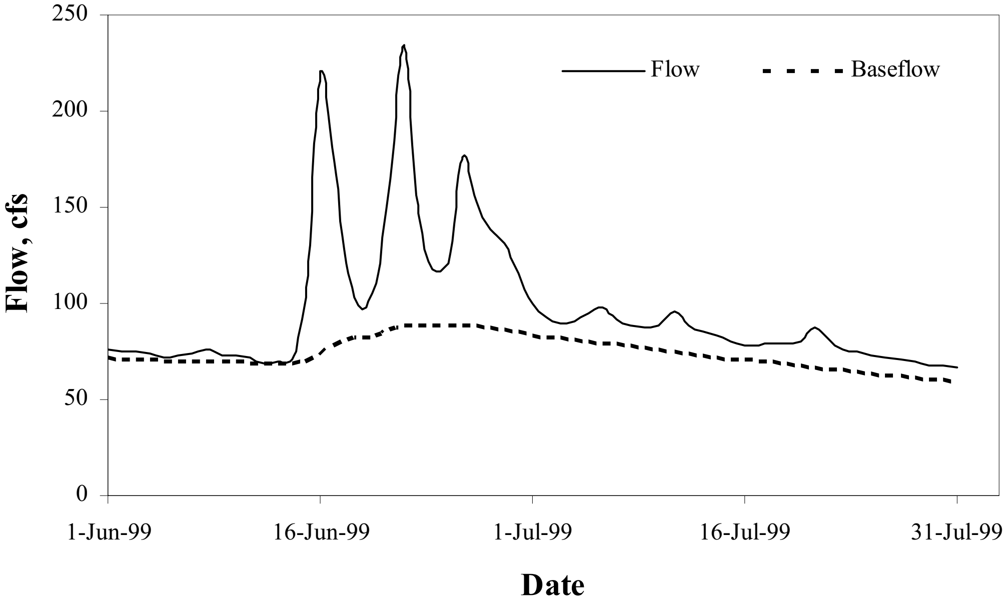

Figure 2.10 Baseflow separation results for USGS station 08180800 for June and July 1999.

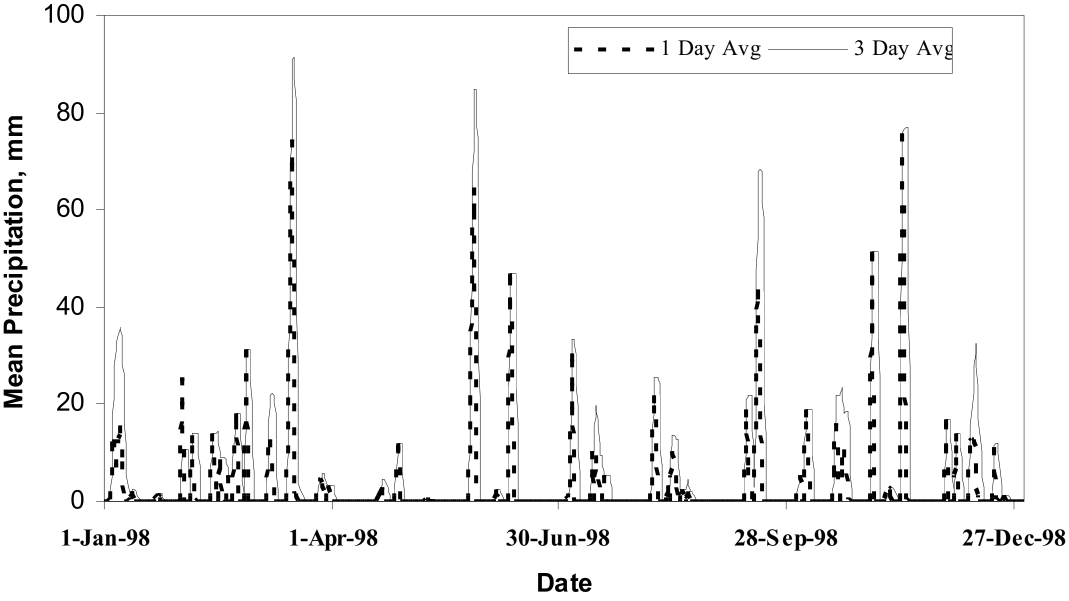

Figure 2.11 Plot of 1 day average and 3 day average precipitation over the contributing area of water quality station 17606 for the year 1998.

Figures 2.12 and 2.13 show the means of fecal coliform bacteria concentrations during baseflow and stormflow periods, respectively, for the period from 1985 to 2000. The high means of bacterial concentration during the baseflow period are of great importance since that indicates the presence of continuous point sources. From Figures 2.12 and 2.13 it can be noticed that in general the mean bacterial concentration during the stormflow period was higher than the mean bacterial concentration during the baseflow period.

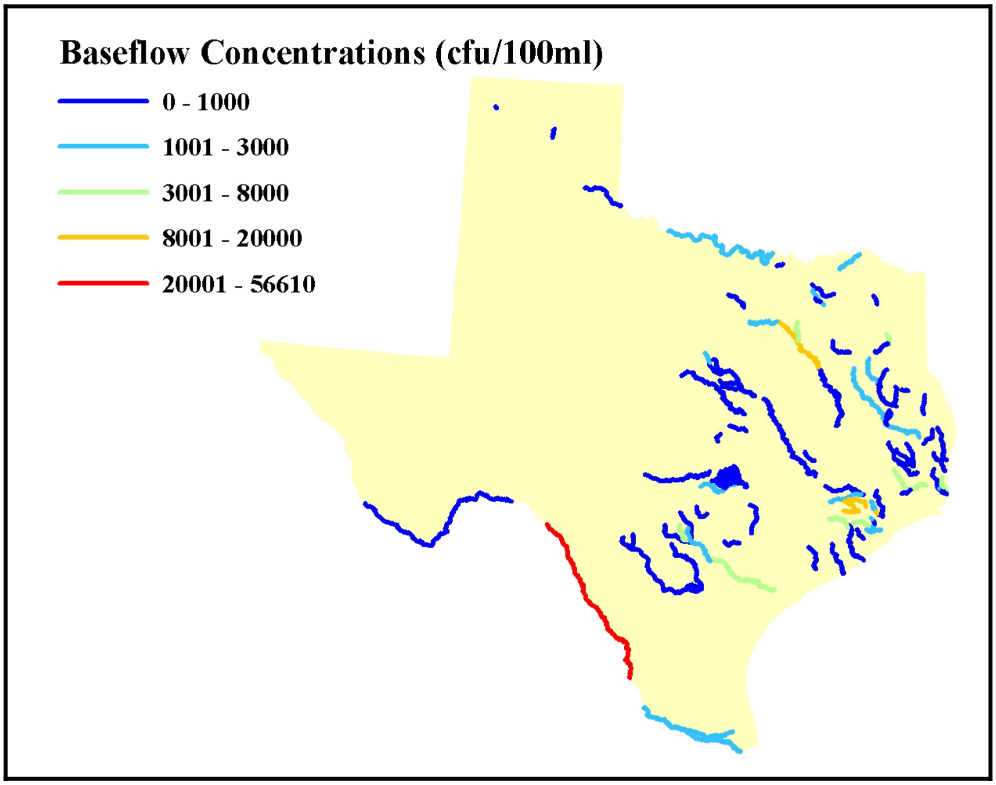

Figure 2.12 Mean fecal coliform bacterial concentrations (cfu/100ml) for impaired stream segments during the baseflow period for duration 1985-2000.

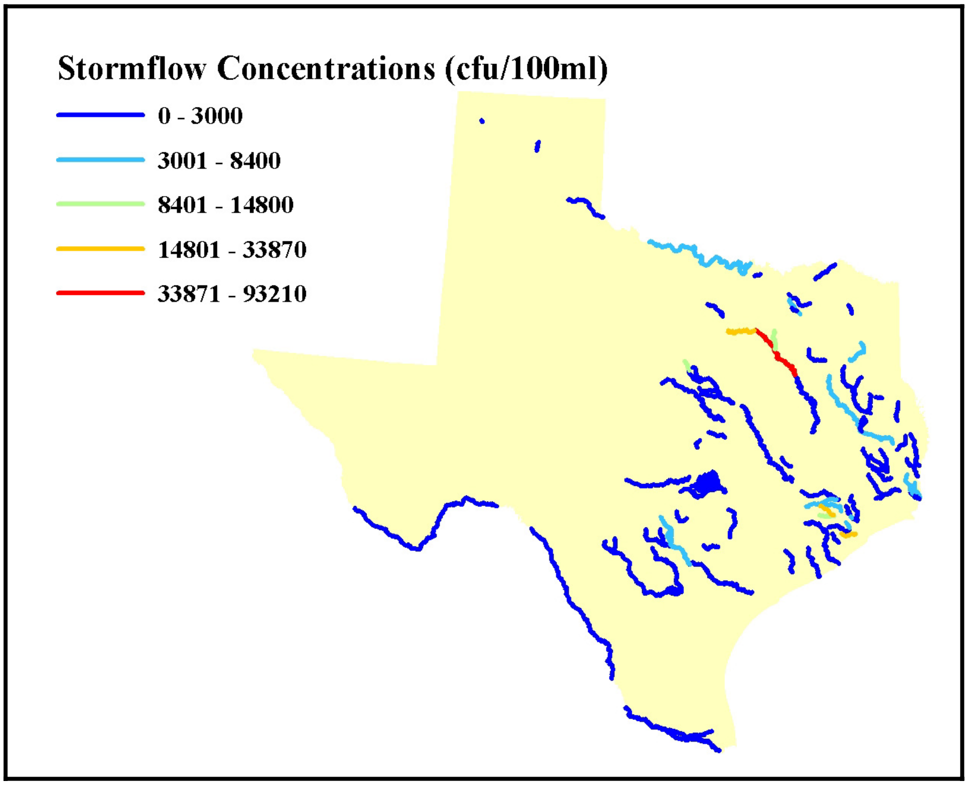

Figure 2.13 Mean fecal coliform bacterial concentrations (cfu/100ml) for impaired stream segments during the stormflow period for duration 1985-2000.

In order to get a clear idea of the distribution of bacterial concentration values during baseflow periods and rainfall periods, the log-transformed data was plotted graphically using boxplots. The boxplots of bacterial concentrations corresponding to the watersheds of stream segments 1218 (Nolan Creek/South Nolan Creek), 1910 (Salado Creek) and 0611C (Mud Creek), respectively are shown in Figures 2.14 through 2.16. From Figure 2.14 it can be interpreted that the bacterial concentrations during the stormflow and baseflow periods are comparable (means of the untransformed data were 408 cfu /100ml and 443 cfu/100ml, respectively), but there are some extreme values of concentration following a rainfall event. A careful observation of Figure 2.15 shows that the mean in-stream bacterial concentration for Salado Creek during the stormflow period is higher than that during the baseflow period. Also it can interpreted that the observations during the stormflow period for Salado Creek tend to be higher than the observations during the baseflow, with extreme observations during the stormflow period being very high compared to the extreme during the baseflow period. However, the distributions of bacterial observations in Mud Creek shown by Figure 2.16 tend to be higher during the baseflow period. The mean value during the baseflow period is also high compared to the mean during the stormflow period. Though the boxplots are good for visual interpretation deriving conclusions from these plots could not be possible. The pooled T-test was carried out to see if there is statistically significant difference between means during the two periods of observation.

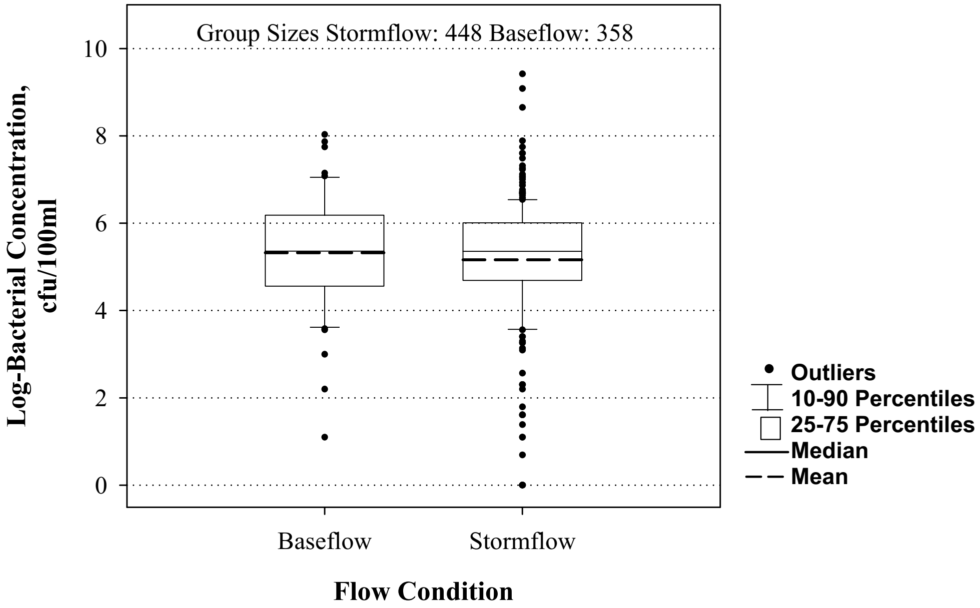

Figure 2.14 Boxplot showing distribution of bacterial concentration (cfu/100ml) during the two periods of flow corresponding to Nolan Creek/South Nolan Creek (Segment 1218).

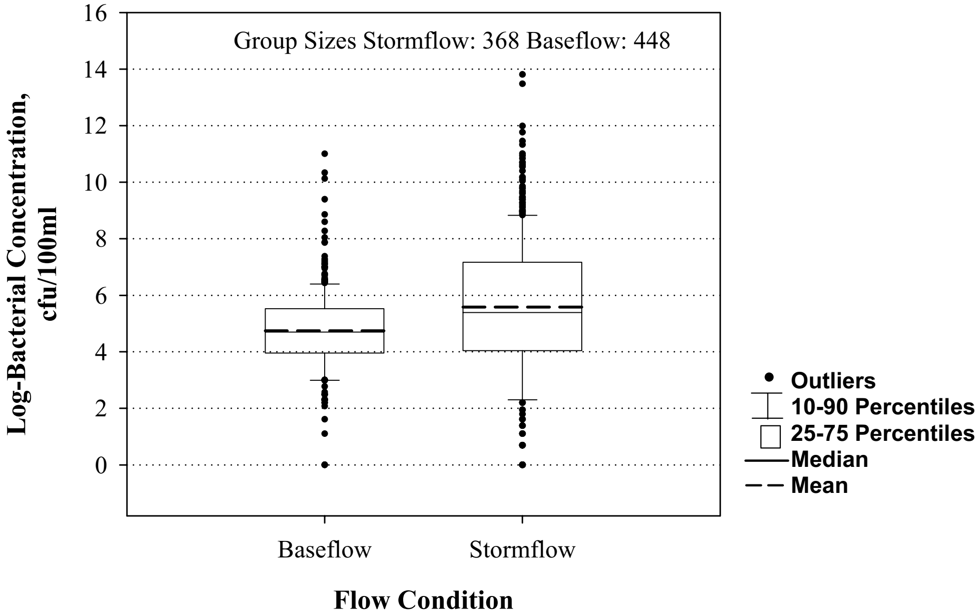

Figure 2.15 Boxplot showing distribution of bacterial concentration (cfu/100ml) during the two periods of flow corresponding to Salado Creek (Segment 1910).

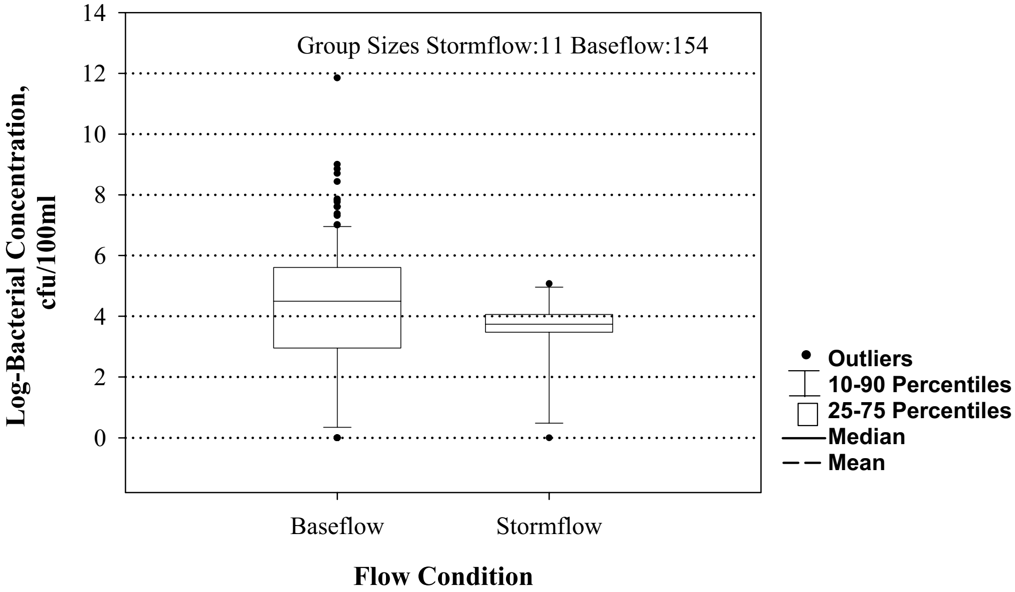

Figure 2.16 Boxplot showing distribution of bacterial concentration (cfu/100ml) during the two periods of flow corresponding to Mud Creek (Segment 0611C).

Statistical Comparisons of Means

The means of water quality observations during the baseflow period (low-flow) and following a rainfall event (high-flow) for all the watersheds were analyzed. The analyses were conducted using the raw observation values after transforming the data using the natural logarithm. Based on the earlier studies it can be assumed that bacterial concentration values generally follow a lognormal distribution. Transforming the data helped to overcome the problem of outliers. This was also helpful in visual interpretation of boxplots. The pooled T-test was carried out to see if there is a statistically significant difference between means during the two periods of observation, using both the raw and the log transformed datasets. Observations within the watersheds corresponding to 44 stream segments showed statistically significant high bacterial concentrations during the stormflow period, 10 during the baseflow period, and watersheds corresponding to 45 stream segments showed no significant difference between the two periods (at $\alpha = 0.05$). For 11 stream segments the number of observations during either stormflow or baseflow period was very little (less than five observations), thus the analysis was inappropriate. The results of the statistical analysis are shown in Figure 2.17.

For the 45 stream segments with no significant difference between the mean concentrations for stormflow and baseflow periods, the mean and median concentrations for the two periods were compared. If any of the following conditions was true, then two concentrations were assumed different from one another:

$$\frac{\overline{X}_r - \overline{X}_b}{\min.(\overline{X}_r, \overline{X}_b)} > 1 \tag{2.11}$$

or

$$\frac{X_{0.5r} - X_{0.5b}}{\min.(X_{0.5r}, X_{0.5b})} > 0.5 \tag{2.12}$$

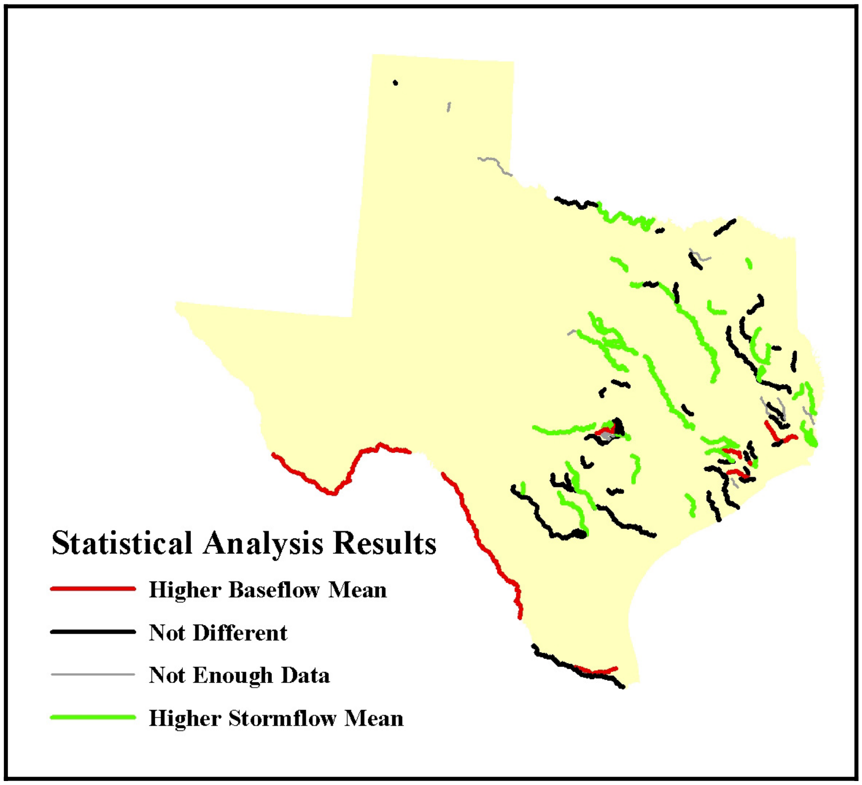

Figure 2.17 Map showing the inferences based on the Student's T test for the duration 1985-2000.

where $\overline{X}_r$ is the mean of bacterial observations during the stormflow period, $\overline{X}_b$ is the mean of bacterial observations during the baseflow period, $X_{0.5r}$ is the median of bacterial observations during the stormflow period, and $X_{0.5b}$ is the median of bacterial observations during the baseflow period.

Based on the comparisons described using equations 2.11 and 2.12 it was found that few of the watersheds showed differences between the means of bacterial concentrations during the stormflow and baseflow periods. When mean and median comparisons were combined with the T test, 67 stream segments showed higher concentrations during stormflow period, 21 stream segments showed higher means during the baseflow period and 11 stream segments had no significant difference between means during the two periods. Figure 2.18 shows the results based on T test and mean and median comparisons. Table 2.1 lists all the stream segments considered to have higher means during the stormflow period. Few of these stream segments showed no significant difference between the means during stormflow and baseflow periods based on the pooled T-test (Statistically Significant value = "No"). Also a careful observation will show that for some of the stream segments the mean of the bacterial concentration during the baseflow period is reasonably high, indicating the presence of both nonpoint and point sources of coliform bacteria.

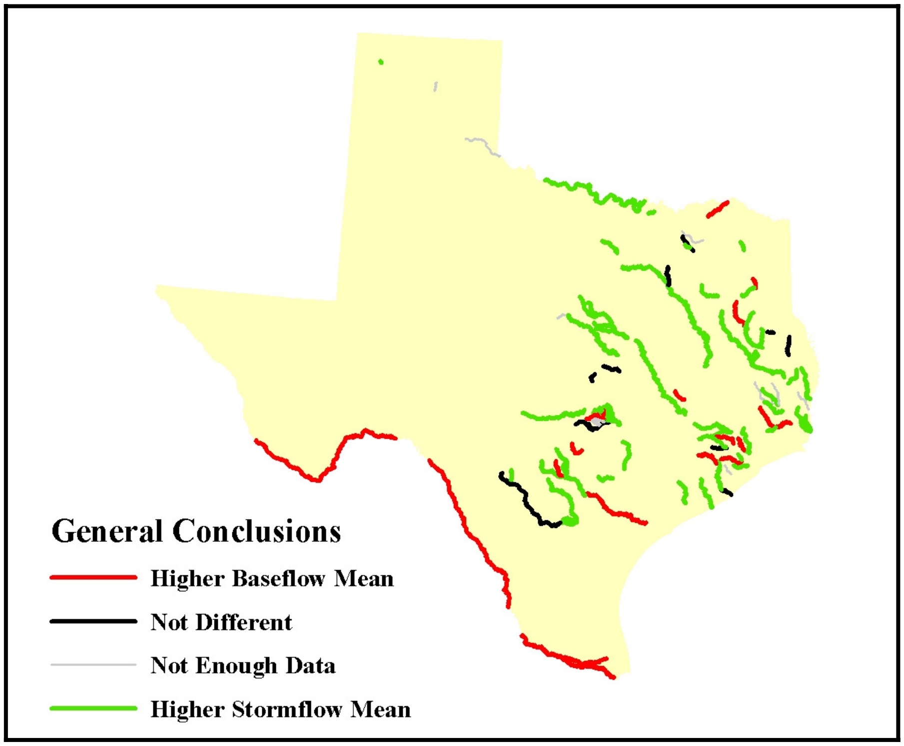

Figure 2.18 Map showing the inferences based on the statistical analysis for the duration between 1985 and 2000.

Table 2.1 Water quality statistics for stream segments with higher means during stormflow period.

Segment ID	Overall number of observations	Overall mean	Minimum count	Baseflow period mean	Stormflow period mean	Statistically Significant	Number of Observations >400 (cfu/100ml)
0805	335	46980	131	17290	93210	Yes	120
0841	191	17320	94	1288	33870	No	49
1103	83	3125	5	1752	24550	No	43
1104	83	3125	5	1752	24550	No	43
0806	223	12910	98	1200	22090	Yes	55
1017	699	17260	341	14180	20500	Yes	603
1255	567	4661	174	2189	10250	Yes	333
0508B	44	5472	11	3958	10010	Yes	23
1113A	11	4875	5	1937	8400	No	9
1910	816	4247	368	601	7242	Yes	261
1113	59	2853	21	631	6873	No	23
0508A	368	5445	92	5032	6683	Yes	192
0511C	50	2327	11	1140	6537	Yes	24
0508	352	5498	87	5172	6492	Yes	182
0505D	33	1832	8	659	5499	No	7
0604	189	2922	78	1642	4743	Yes	14
1911	2601	3480	1261	2191	4694	Yes	1301
1009	1046	3520	324	1251	4538	Yes	599
0511A	165	1762	42	924	4217	Yes	43
0205	171	3100	62	2283	4085	No	38
0204	175	977	60	1437	3191	Yes	28
1803A	44	1231	15	376	2886	No	18
2107	75	1773	31	738	2517	Yes	47
1804B	65	851.1	10	555	2479	Yes	20
1110	44	1260	20	655	1985	No	16
1008	495	1219	205	173	1958	Yes	162
1502	300	1635	55	345	1924	Yes	95
1242	419	389	191	568	1888	Yes	183
0804	485	1296	166	179	1877	Yes	59
0901	34	1101	14	99	1802	Yes	6
1428B	126	1222	39	380	1599	No	19
1304	99	999.1	49	453	1557	No	24
0507B	20	1001	8	222	1520	No	8
0604B	38	1171	10	1068	1462	Yes	19
1903	462	416.66	161	597	1435	No	102
0611B	80	871.4	23	648	1425	Yes	28
0511	595	1075	188	914	1421	Yes	148

Table 2.1 Continued.

Segment ID	Overall number of observations	Overall mean	Minimum count	Baseflow period mean	Stormflow period mean	Statistically Significant	Number of Observations >400 (cfu/100ml)
0604A	131	877.9	29	735	1382	No	45
0603A	33	441.1	5	294	1266	Yes	9
0404B	114	717.9	53	285	1216	Yes	15
0608C	55	488.1	7	390	1161	No	13
1226	2037	849.4	970	572	1154	Yes	483
1226A	135	447.3	45	97	1148	Yes	18
2116	160	798.33	65	560	1147	No	52
0810	387	368.6	125	30	1079	Yes	36
0605A	64	909.9	20	834	1076	Yes	19
0511B	416	942.7	146	879	1061	Yes	95
0513	173	498.8	54	263	1019	Yes	43
0604C	47	528.5	8	438	970	Yes	15
1428C	61	673.9	9	628	937	No	17
1428	1875	875.4	843	872	879	Yes	319
0611	250	468.7	41	390	872	Yes	74
0902	32	896	5	173	713	No	8
1414	426	391.3	98	301	693	Yes	40
2426	411	435.1	204	208	659	Yes	63
0608D	107	475	5	467	646	No	26
1403A	802	454.4	326	392	546	Yes	92
1304A	36	473.3	10	291	544	No	10
1221	418	326.4	136	249	487	Yes	119
0203A	26	358.4	8	328	427	No	9
1403	1873	164.5	624	76	342	Yes	136
2110	84	211.6	31	73	342	Yes	4
0607C	34	270.6	6	263	307	No	5
1226D	251	192.4	124	86	296	Yes	12
1429B	466	279.9	130	277	288	Yes	46
1226C	224	163.4	81	154	180	Yes	13
0105	30	98.5	8	98	99	No	3

Note: Overall mean, baseflow period mean and stormflow period mean are in cfu/100ml

Stream segments that had higher means during the baseflow period are listed in Table 2.2. Eleven of the stream segments showed no significant difference between the means during stormflow and baseflow periods based on the pooled T-test. A high mean concentration during the baseflow period can be considered as an indication of continuous loading of bacteria into the stream from point sources. A higher mean concentration during the baseflow period together with a relatively low mean concentration during the stormflow period may indicate a dilution effect as a result of a heavy rainfall and also indicate the absence of high loading from nonpoint sources.

Table 2.2 Water quality statistics for stream segments with higher means during baseflow period.

Segment ID	Overall number of observations	Overall mean	Minimum count	Baseflow period mean	Stormflow period mean	Statistically Significant	Number of Observations >400 (cfu/100ml)
2304	1344	54610	48	56610	709	Yes	514
2429	1749	16100	296	18340	5112	Yes	1240
1016	737	13620	126	15250	5747	Yes	627
1906	393	2381	149	4495	1090	No	100
1901	543	2849	251	4362	1548	No	165
607	319	3338	83	3752	2160	Yes	62
1245	286	3192	54	3633	1296	No	119
0505B	30	3394	8	3522	3044	No	18
1102	320	2671	75	3184	994	Yes	269
1101	706	2687	48	2766	1601	Yes	361
2202	179	2562	8	2636	966	Yes	122
0202D	70	2005	12	2303	562	No	24
0611C	135	1482	7	1561	53	No	17
1001	1346	1312	179	1415	640	No	228
2302	949	1167	118	1194	980	No	372
1429A	292	641.8	101	814	551	Yes	50
1811A	191	684.9	26	718	479	No	24
1209C	141	569.5	12	594	309	No	60
2306	246	536.5	25	561	323	Yes	65
1430	757	379.9	232	424	280	Yes	57
1428A	25	141.6	11	206	60	No	2

Note: Overall mean, baseflow period mean and stormflow period mean are in cfu/100ml

Stream segments with no difference between the means of bacterial observations during baseflow and stormflow periods are given in Table 2.3. Though these stream segments showed no significant difference between means of concentration during stormflow and baseflow periods, the mean concentrations during both stormflow and baseflow periods were high for a few of these stream segments. High means of bacterial concentration during both stormflow and baseflow periods may be because of large contributions from both point and nonpoint sources. In fact, the number of observations with values greater than 400 cfu/100ml for two stream segments (1013 and 1014) was found to be very high.

Table 2.3 Water quality statistics for stream segments without significant difference between the means during stormflow and baseflow periods.

Segment ID	Overall number of observations	Overall mean	Minimum count	Baseflow period mean	Stormflow period mean	Number of Observations >400 (cfu/100ml)
1013	931	13320	279	12690	14800	762
1014	846	11800	305	11470	12380	672
0819	138	9517	58	7270	12620	60
0507A	78	2808	38	2042	3614	38
1427	579	1190	163	1336	817	72
1109	83	905.7	38	668	1188	22
2117	139	556.6	56	601	491	43
0612B	14	373.9	6	449	274	4
1218	332	413.2	50	443	408	96
0610A	91	357.3	29	304	471	25
1217A	46	215.9	18	166	248	1

Note: Overall mean, baseflow period mean and stormflow period mean are in cfu/100ml

Stream segments without enough data to compare the means during baseflow and stormflow periods are given in Table 2.4. Without collecting more data, deriving any conclusions regarding potential sources of pollution for these stream segments is not feasible.

Table 2.4 Water quality statistics for stream segments without sufficient data.

Segment ID	Overall number of observations	Overall mean	Minimum count	Baseflow period mean	Stormflow period mean	Number of Observations >400 (cfu/100ml)
1108	19	1018.8	1	45	1076.12	8
0502A	12	904	0	904		4
1427B	162	863.4	0	863.4	--	26
0306	36	771.3	4	808.7	322.5	7
0608B	82	629.35	4	530.9	2548.5	19
0207A	30	467.47	3	446.22	658	12
0608F	53	449.7	4	376.6	1344.8	14
0101A	47	384.4	0	384.42	--	18
1222A	1	288	0	288	--	0
1427A	26	264.8	2	176.38	1327	4

Note: Overall mean, baseflow period mean and stormflow period mean are in cfu/100ml

Conclusions

In general, 67 stream segments showed higher mean bacterial concentrations during stormflow periods and 21 stream segments showed higher mean bacterial concentrations during baseflow periods. Eleven stream segments showed no significant difference in the means of bacterial concentrations during stormflow and baseflow periods. There was not enough data for analysis for 11 stream segments. Based on a pooled T-test, 44 stream segments showed significantly higher bacterial concentrations during the stormflow period, 10 during the baseflow period, and 45 stream segments showed no significant difference between the two periods. It may be concluded that there was a considerable increase in in-stream bacterial concentrations for 44 stream segments due to runoff, while the effect of rainfall on the in-stream bacterial concentrations of 45 stream segments was negligible.

The waterbodies with high mean bacterial concentration during the baseflow period and relatively low mean bacterial concentrations during the stormflow period may have continuous bacterial loading coming from point sources. The reduction in bacterial concentrations for these stream segments during the stormflow period could be due to the dilution effect of the heavy rainfall events.

Lower concentrations of bacteria during the baseflow period can be interpreted as the absence of continuous point sources. An increase in concentration of bacteria during stormflow periods may be due to the increase in loading of bacteria into the streams due to washoff of bacteria from runoff due to the rainfall. These can be interpreted as the evidence of nonpoint sources of bacterial pollution.

Higher concentrations during both stormflow and baseflow periods may be the result of both point and nonpoint source loading. However, for a two waterbodies (2304 and 2306) there were no known point sources and the densities of livestock and wildlife were high. Hence, the reason for having a high concentration of bacteria during baseflow period for these stream segments may be the presence of animals within accessible distance of the streams.

Four stream segments (2202, 2302, 0205 and 0204) had reasonably high concentrations during both stormflow and baseflow periods, but there was no evident source within the watersheds. Further research is required to understand these differences in the effect of rainfall.

CHAPTER III

CLUSTERING BACTERIALLY IMPAIRED WATERSHEDS THROUGH ANALYSIS OF WATERSHED CHARACTERISTICS

Synopsis

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) listed 110 stream segments in the year 2000 with pathogenic bacteria impairment. The next logical step is to verify impairment and if sufficient evidence is present then to develop a Total Maximum Daily Load (TMDL) for each of these impaired waterbodies. A study was conducted to characterize the watersheds associated with these impaired waterbodies. The primary aim of the study was to explore the possibility of clustering the waterbodies into groups having similar watershed characteristics, studying them as a group, and choosing models for TMDL development based on their characteristics. This approach will reduce the number of required TMDLs and thereby will help in reducing the effort required for restoring the health of the impaired waterbodies in Texas. The main characteristics considered for the classification of waterbodies were designated use of the waterbody, land use distribution, density of stream network, average distance of a land of a particular use to the closest stream, household population, density of on-site sewage facilities (OSSF), bacterial loading due to the presence of different types of farm animals and wildlife, and average climatic conditions. The availability of observed in-stream fecal coliform bacteria concentration data was evaluated to obtain subgroups of data-rich and data-poor watersheds within a group. The climatic data and observed in-stream fecal coliform bacteria concentrations were analyzed to evaluate seasonal variability of the water quality. The grouping of waterbodies was carried out using multivariate statistical techniques, factor analysis/principal component analysis, cluster analysis, and discriminant analysis. The multivariate statistical analysis resulted in six clusters of waterbodies. The main factors that differentiate the clusters were found to be bacterial contribution from farm animals and wildlife, density of OSSF, density of households connected to public sewers, and land use distribution.

Introduction

According to the Code of Federal Regulations (CFR) Title 40, Part 131, all States, Territories, and authorized Tribes of the United States must update the impaired or threatened waterbodies list under their jurisdiction, once in two years. The list of impaired waterbodies is called the Clean Water Act (CWA) §303(d) list and is prepared and submitted to the Environmental Protection Agency (EPA) for their approval (USEPA, 1998a). Also the regulatory authorities set the priority rankings for TMDL development for the waterbodies based on the severity of the pollution and the designated use. Once the waterbodies are prioritized, the next step is to develop the Total Maximum Daily Load (TMDL) for each of these waterbodies in order to restore the health of the waterbodies.

Among the many pollutants that require the development of TMDLs, fecal coliform is included because it indicates a serious potential health risk. Fecal coliforms are a group of bacteria that primarily live in the lower intestines of warm-blooded animals, including humans. The presence of high concentration of fecal coliform bacteria indicates the presence of dangerous pathogens. Under the CWA program the Texas Commission on Environmental Quality (TCEQ) listed 110 waterbodies (Figure 3.1, stream segments are listed in Table A1 in Appendix) in the year 2000 with indicator bacteria concentrations more than those permitted for their designated uses (TNRCC, 2000). Once the impairment is verified, the development of TMDLs for these waterbodies seems to be the best solution for the problem. But, developing a TMDL for each and every one of these stream segments will require an enormous amount of input, both in terms of capital and human labor. A case study conducted by the EPA showed that the cost of a single TMDL study varied between $4,039 and $1,023,531 (USEPA, 1996c). It was pointed out that on average 32% of the total expense was allotted for the modeling component of the TMDL studies. Many of the waterbodies considered for TMDL development listed under the current CWA §303(d) for Texas may be grouped based on their watershed characteristics and possible sources of pollution. Such a grouping scheme would be helpful in reducing the cost of restoration of water quality by restricting the development of the TMDL to one or two representative waterbodies under a single group and applying the knowledge to other waterbodies in the same group.

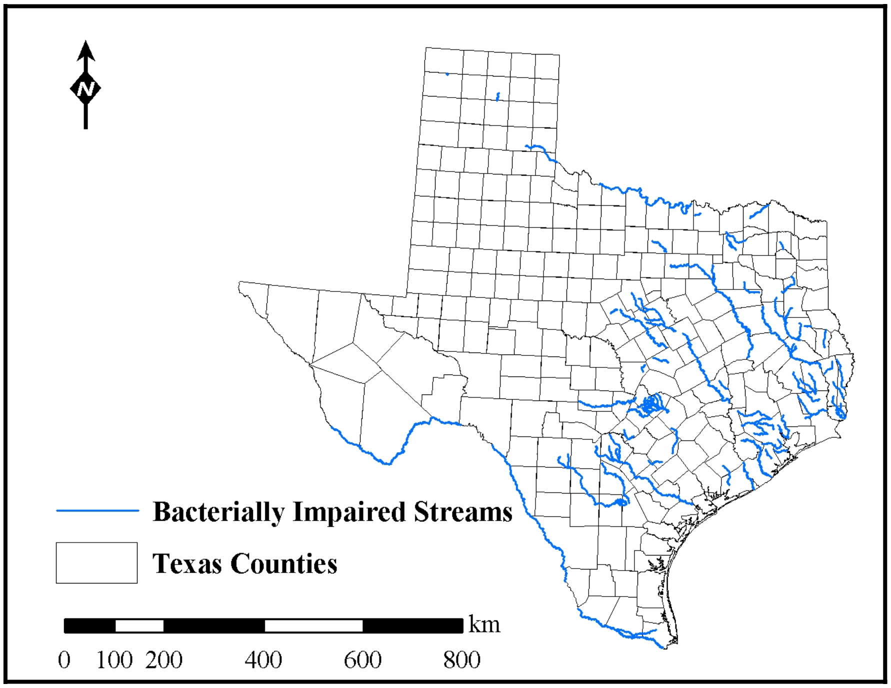

Figure 3.1 Network of stream segments listed for bacterial impairment in Texas in the year 2000 (TNRCC, 2000).

The current study focuses on the development of a method for classifying the Texas waterbodies listed for bacterial quality violation under CWA §303(d) into groups having similar watershed characteristics using multivariate statistical techniques.

Multivariate analysis is a branch of statistics dealing with pattern of relationships between several variables simultaneously. The various multivariate analysis methods employed in the current study are factor analysis (FA), principal component analysis (PCA), cluster analysis (CA) and discriminant analysis (DA). A brief description of FA, PCA, CA and DA are given in the following sections.

Factor analysis (FA) is one of the most common multivariate statistical techniques used to reduce the dimensionality of large sets of variables (Karson, 1982). Factor analysis is used to analyze the interrelationships among different variables and to find common factors, thus to condense the information contained in a large number of variables into a smaller set of factors without sacrificing much information. The two main types of FA are PCA and common factor analysis.

Principal component analysis is used to create linear combinations of the original variables into a smaller set of new variables, the principal components (PC), which explain the maximum amount of variance possible (Karson, 1982). These PC are orthogonal to each other thus they are uncorrelated. Successive PC account for decreasing proportions of the total variances of the original variables. Factor analysis/PCA has been used in many water quality assessment studies (Vega et al. 1998; Helena et al. 2000). Alberto et al. (2001) used the FA/PCA technique in a study to evaluate the spatial and temporal changes of water quality in Suquía River Basin, Argentina.

Cluster analysis, also known as unsupervised pattern recognition is a set of statistical techniques for exploratory data analysis and is used to classify a set of observations into multiple groups based on multivariate properties (Karson, 1982). The groups are formed in such a way that the observations are highly internally homogenous and highly externally heterogeneous, meaning, the members of a group are similar to one another and are different from members of other groups. Different methods of CA produce entirely different results and they are interpreted based on the particular need.

Discriminant analysis is used to determine the variables that can discriminate between different groups. Discriminant analysis is used when membership of different observations to a given group is known a priori. A discriminant function is constructed for each group and is used to divide the observations into different regions in the data space (Karson, 1982). Discriminant analysis has been used in many studies to identify the sources of fecal pollution in aquatic systems using antibiotic resistance patterns (Wiggins et al., 1999; Parveen et al., 1999; Choi et al., 2003; Whitlock et al., 2002). Discriminant analysis has also been used to verify the efficiency of grouping schemes produced by CA (Alberto et al., 2001; Clucas, 1997).

Jenerette et al. (2002) used a combination of PCA, CA, and DA to evaluate the effectiveness of delineating aquatic systems based on the ecoregion approach. A similar approach was used by Alberto et al. (2001) in a study to evaluate the spatial and temporal changes of water quality in Suquía River Basin, Argentina using a combination of FA/PCA, CA, and DA.

Methodology

Overview

The sources of fecal coliform bacteria are divided into point and nonpoint sources. The main nonpoint sources of fecal bacteria are wildlife, livestock, poultry, humans, domestic pets, OSSF and migratory birds. The point sources of fecal bacteria are confined animal feeding operations, discharges from waste water treatment plants, and fish and shellfish processing facilities (USEPA, 2001c). In the current study normalized data from the impaired watersheds were initially analyzed with FA/PCA to find the main PC. Cluster analysis was conducted using the PC obtained from the PCA to group the watersheds with similar characteristics. After obtaining the clusters of watersheds, the data were analyzed using descriptive DA to find the set of variables that were important in discriminating the groups. Discriminant analysis was also used to find the percentage of error in the clustering approach.

Factors Considered in Multivariate Analysis

The factors used to cluster the waterbodies impaired by bacteria are:

1. The possible sources of bacteria within the watershed stated above.
2. Proximity of the sources to the reach network.
3. Seasonal variability of bacterial observations.
4. Climatic conditions.
5. Land use distributions and.
6. Density of stream network.

The process of clustering starts with the delineation of the watershed that contributes flow to each of the waterbodies listed for bacterial impairment. The watersheds were analyzed using a Geographic Information System (GIS) to obtain the characteristics such as the land use distribution, the distance of land of a particular use to the nearest stream, the relationship between flow and water quality, the number of potentially failing OSSF, number of wildlife, and number of livestock. A data matrix containing the characteristics of all the watersheds was compiled. This matrix was subjected to different multivariate statistical techniques and the waterbodies with similar characteristics were grouped using the cluster analysis technique. The GIS analysis was done using ArcView GIS software (ESRI, 1999). The statistical analysis was done using SAS (SAS, 1999).

Watershed Delineation

Watershed delineation is the process of separating the geographic area that drains water, sediments and other dissolved substances through a common outlet as concentrated drainage. The common outlet could be a lake, stream, estuary or ocean. The delineation process was done using the GIS data layers including the fourteen-digit hydrologic unit code (HUC) boundaries, the National Hydrography Dataset (NHD) stream network developed by the US Geological Survey (USGS), the USGS Digital Elevation Model (DEM), and the EPA Reach File Version 3 (Rf3). The watersheds for each of the 110 waterbodies were delineated and saved as ArcView shape files. The watershed data for each of the stream segments was collected based on the watershed the stream segment occurred in. The watershed delineation process for an impaired stream segment started with identifying the NHD streams that contribute flow to that stream segment. Next, the 14-digit HUC boundaries that intersected with the NHD streams were selected to create the watershed boundary layer. Based on an initial cluster analysis the watersheds were eventually modified to include only the 14-digit HUCs directly connected to a given impaired stream segment, discarding those 14-digit HUCs that contributed flow but were not directly connected to the impaired stream segment . The 14-digit HUC boundaries were developed by the Blackland Research

Center located at Temple, Texas in collaboration with Texas State Soil and Water Conservation Board. The creation of 14-digit HUC boundaries were not yet completed for the entire state of Texas, however, they were available for most of the study area. For the part of the study area where the 14-digit HUCs were not available the watersheds were delineated based on the USGS DEM. The process was carried out with the help of the automatic watershed delineation tool available within BASINS (USEPA, 2001a).

Land Use Distribution

Identification of the possible sources of bacteria within a watershed requires knowledge of the location and distribution of land uses. The type of the land use affects loading and accumulation rates of fecal coliform bacteria on land surface. The land use distribution was determined using the National Land Cover Data (NLCD) obtained from the USGS. The NLCD data were developed from Landsat satellite Thematic Mapper data acquired by the Multi-resolution Land Characterization (MRLC) Consortium. The percentage of each type of land use was calculated for each of the impaired watersheds.

Drainage Density and Distance Factor

The transport of pathogens from the land surface to streams is thought to have a major role in the process that ultimately results in stream water pollution. The transport rate of pathogens from different type of land use is a function of the loading rate of the pathogens on the land surface and the ease with which they are carried to the stream. Since the nonpoint source loading rate is dependent on the transport rate of the pollutants with overland flow, the density of the stream network and the distance of the pollutant source from the stream influence the contribution of pollutants from a nonpoint source to the stream. It is assumed that as the drainage density increases, the possibility of a stream being polluted during a rainfall event also increases. Conversely, an increase in the average distance between the land area acting as the nonpoint pollutant source and the nearest stream will decrease the possibility of a stream being polluted during a rainfall event, especially during a low-intensity, low-duration rainfall event. The drainage density was determined by dividing the total length of the NHD stream network within the watershed by the total watershed area. The average distance between source areas based on their land use to the nearest stream was determined using the NLCD land use layer and NHD stream network. The distance between each land use polygon and the closest NHD stream segment was calculated first. Then the weighted mean distance for each type of land use was determined using the area of the individual polygons as the weight. These weighted mean distances of land use to stream were used in the multivariate analysis.

Flow-Bacteria Concentration Relationship

The effect of nonpoint sources of bacteria on the timing of water quality impairment is different from that of point sources. High concentrations of bacteria during low-flow, dry weather conditions usually indicate continuous loading from point sources (USEPA, 2001c). At the same time, if the concentration of bacteria is higher during storm events, then there is a high probability that impairment is the result of nonpoint sources. The observed water quality data corresponding to available water quality stations were compared with the flow data from USGS gage stations close to the water quality stations. Figure 3.2 and 3.3 show the locations of USGS gauging stations and water quality observation stations, respectively. First, the flow data within the period 1985 to 2000 were collected and the baseflow were separated using the technique developed by Arnold and Allen (1999). The total measured flow for each day was compared to the baseflow to determine whether the stream flow on that day was due to a rainfall event or due to baseflow contribution. If the stream flow was more than 1.1 times that of the baseflow, then streamflow was assumed to be the result of a rainfall event. Otherwise streamflow was assumed to be contributed only by baseflow.

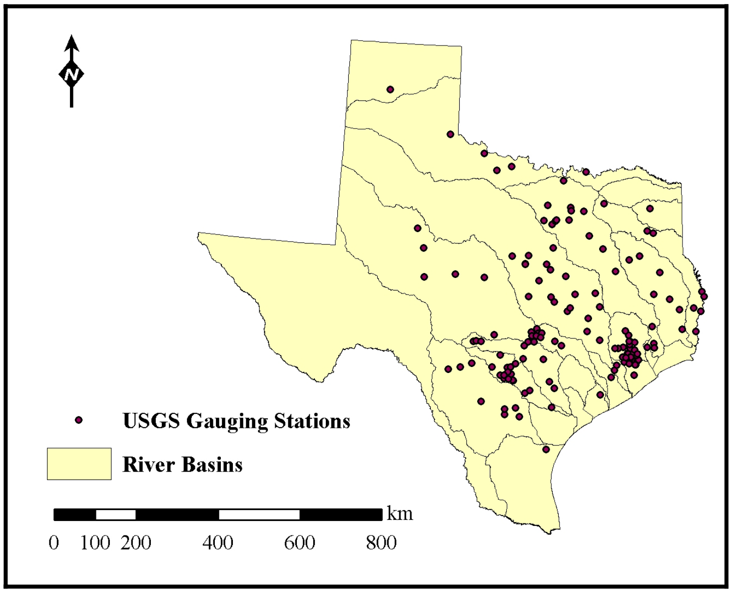

Figure 3.2 Locations of the USGS gauging stations.

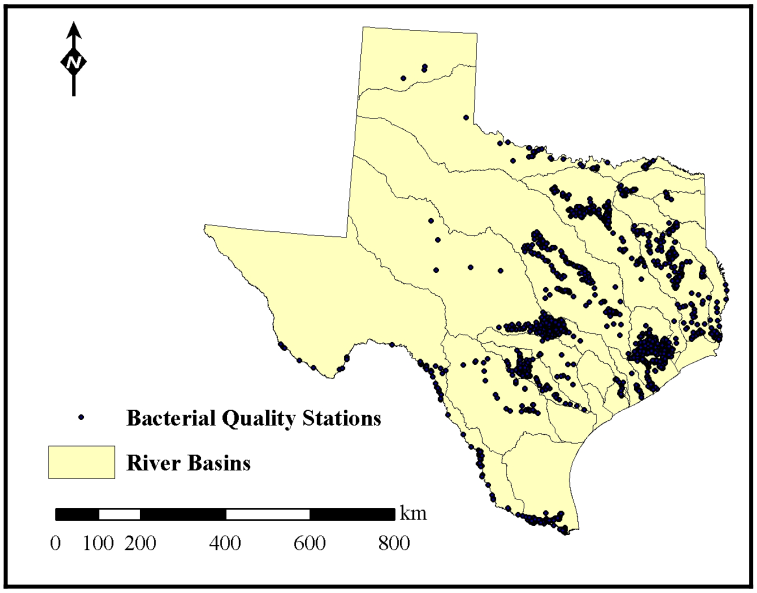

Figure 3.3 Locations of the water quality stations.

Water quality observations were from grab samples collected by the TCEQ from 1985 to 2000. The flow and the baseflow were recorded for everyday a water quality observation was available. The water quality data were then separated into days where streamflow was only from baseflow period and days where streamflow resulted from rainfall events.

Rainfall- Bacteria Concentration Relationship

When there was no USGS station close to the water quality station the data were compared using the average daily rainfall over the area contributing flow to the water quality station. The precipitation data from weather stations maintained by the National Climatic Data Center (NCDC) were obtained for this comparison. Figure 3.4 shows the locations of the NCDC weather stations. Based on the precipitation information obtained from NCDC, the water quality data were grouped into days having non-zero rainfall events (> 1mm) and days without rainfall.

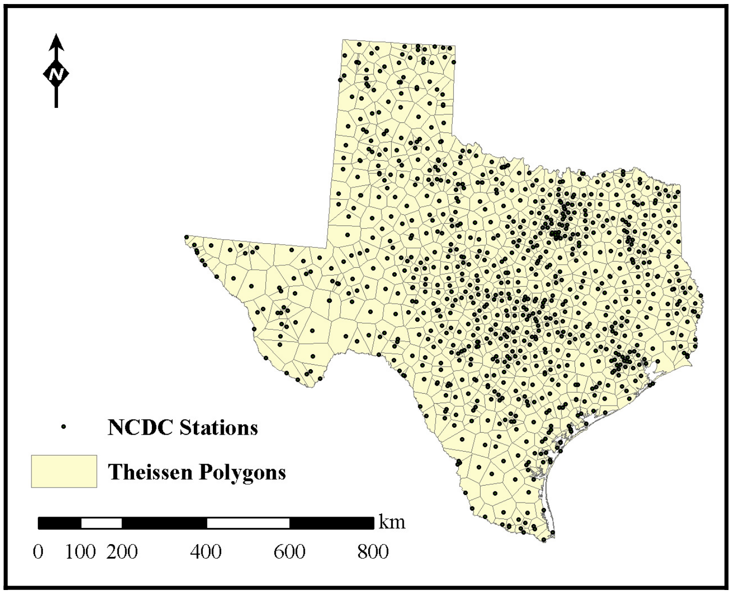

Figure 3.4 Locations of the NCDC precipitation stations in Texas.

The data were compiled for all the water quality stations that fell within a single watershed. A pooled T-test (Milton and Arnold, 1995) was used to test the hypothesis that there was a significant difference in mean in-stream fecal coliform concentration during periods with rainfall and periods without rainfall.

Septic Systems

The contribution of fecal coliform bacteria from humans through the discharge from failed septic systems is undisputable (USEPA, 2001c). Although an accurate calculation of bacterial loading from human sources is not possible, it can be assumed that the rate of loading would be proportional to the population and number of households within the contributing watershed of an impaired stream segment.

The population and number of households within a watershed were calculated by summing data from census blocks intersecting the watershed under consideration. The population data for each census block was obtained from the US Census Bureau (US Census Bureau, 2000). The number of households connected to public sewer systems and the number of OSSF installed before 1990 were obtained from the 1990 US Census. The number of OSSF installed after 1990 was obtained from the TCEQ. GIS layers were used to identify the geographical locations, population, and number of households that utilize OSSF or public sewer systems. It was assumed that the OSSF were generally present only outside major cities. The assumption is that households located within major cities or Metropolitan Statistical Areas (MSA) are generally connected to the public sewer systems. The GIS data layer corresponding to these MSA was obtained from Spatial Science Laboratory, Texas A & M University. The number of OSSF within the watershed, NSS_{ws}, was calculated as:

$$NSS_{ws} = \sum_{c=1}^{C} \frac{HH_{ws} \times NSS_c}{HH_c} \qquad (3.1)$$

where, HH_{ws} is the number of households within the watershed, HH_c is the number of households within the county and NSS_c is the number of OSSF in the county. If the watershed is located completely within an MSA, the number of OSSF is equal to zero. The number of OSSF is calculated only for the area that fell outside the MSA, in watersheds that were only partially within an MSA.

Wildlife Population

One of the major sources of fecal coliform bacteria is wildlife. The largest relevant group of wildlife in Texas is the white-tailed deer. Availability of information regarding wildlife was limited to the number of white-tailed deer by county in Texas. The estimated number of deer present in all Texas counties was obtained from the Texas Parks and Wildlife Department for the last 5 years. Deer were assumed to be present only in land use areas of forestland, barrenland and pastureland. The number of wildlife in a watershed (ND_{ws}) was estimated using the number of white-tailed deer reported for the county and the land use distribution within the county (ND_c) and the watershed using the following equation:

$$ND_{ws} = \sum_{c=1}^{C} ND_c \times \frac{Forest_{ws} + Barren_{ws} + Pasture_{ws}}{Forest_c + Barren_c + Pasture_c} \tag{3.2}$$

where, *Forest* is the area of forestland, *Barren* is the area of barrenland, *Pasture* is the area of pastureland, and *ws* and *c* represent watershed and county, respectively. The areas corresponding to the different land uses were calculated using the NLCD GIS layers. The calculation of the bacterial loading from wildlife is given in Appendix B.

Livestock and Poultry

The major source of fecal coliform bacteria from agricultural land and pastureland is through manure application or through direct contribution from livestock and poultry. The numbers of different types of livestock and poultry within the watershed was estimated based on data from the US Department of Agriculture (USDA) Agricultural Statistics Database and data from Concentrated Animal Feeding Operations (CAFOs). The data collected included the number of cattle, swine, sheep, goats, and poultry. The cattle data were available for each county, while poultry data were limited to the agricultural district level. Goat, swine and sheep data were available partially on the county level and partially on the agricultural district level. The location and numbers of CAFOs were obtained from the TCEQ. The calculation of the total number of animals in each watershed, NA_{ws}, was based on the USDA database and the CAFO data layer. Livestock and poultry were assumed to be present only in pastureland and cropland. The number of animals present in CAFO layers within a county or an agricultural district was subtracted from the total number of animals reported in the USDA database. The result was the total number of animals that were not in CAFOs. Then the number of animals within the county or the agricultural district was multiplied by the ratio of pastureland within the watershed to the pastureland within the given county or agricultural district, to obtain (non-CAFO) animals within the watershed. The number of animals within a watershed that were part of a CAFO was obtained directly from the CAFO data layer. Thus, the total number of animals within a watershed would be the sum of animals present in the CAFOs and the number of animals that were not part of any CAFOs. The calculations are as shown below:

$$NA_{ws} = \sum_{c=1}^{C} (NA_c - CA_c) \times \frac{Pasture_{ws}}{Pasture_c} + CA_{ws} \qquad (3.3)$$

or

$$NA_{ws} = \sum_{c=1}^{D} (NA_d - CA_d) \times \frac{Pasture_{ws}}{Pasture_d} + CA_{ws} \tag{3.4}$$

where *NA* is the number of given animal, *CA* is the number of animals reported in the CAFO data layer, *Pasture* is the area under pastureland, and *ws, c* and *d* represent watershed, county, and district, respectively. The calculation of the bacterial loading from livestock is given in Appendix B.

Though we calculated the numbers of poultry, the initial cluster analysis results showed that the use of the poultry data would lead to erroneous conclusions. According to expert's opinion, the litter and slurry manure from poultry industries is stored for a considerably long time in the manure and waste water retention facilities or barns before is applied on the land, resulting in die-off of most of the fecal coliform bacteria before application (A. Jones, personal communication, 29 April 2003). Thus, calculating the bacterial loading directly from the number of chicken present within a watershed would be inappropriate. Therefore, the poultry data was excluded from the dataset.

Migratory Birds

Another source of fecal coliform bacteria in streams is excrement from wild birds. Different types of migratory birds constitute the major population of birds in Texas. Data availability regarding the number of migratory birds was limited to the numbers of different species of birds within each natural-regional boundary. The number of birds in a given watershed was estimated based on the area of the impaired watershed that lay within certain natural region boundaries. Favorable habitat conditions were also used to find areas where the migratory birds will most likely to concentrate. Initial analysis of the data obtained for birds showed that the use of the migratory bird numbers was inappropriate because of the unavailability of location specific data.

Though the consideration of habitat might provide reasonable estimates of the numbers, the migratory bird dataset was excluded from the multivariate statistical analysis.

Domestic Pets

The excrement from domestic pets is another possible source of fecal coliform bacteria. However, the data available were the statewide or nationwide percentages of households that owned different types of pets. These data were obtained from American Veterinary Medical Association [AVMA] (AVMA, 2002). The Center for Information Management of AVMA reports the percentage of Texas households that own dogs and cats and the average number of these pets in those households. The data pertaining to birds and horses were limited to nationwide numbers. Since the information on the pet population was limited to calculation of numbers of pets based on the number of households this information was considered redundant and thus excluded from further analysis.

Location or Type of Waterbody

Another factor considered for clustering the impaired waterbodies is the location or type of waterbody. A one-dimensional model can accurately represent a waterbody like a well-mixed flowing river. If the waterbody under consideration is a lake or a deep reservoir where there is vertical stratification and mixing, then a two-dimensional model should be considered. Similarly three-dimensional models should be used for estuaries or areas near the coast and influenced by tides (USEPA, 1997a). Therefore, the location of each waterbody was identified and was considered during the clustering process.

Multivariate Analysis

The various multivariate techniques used to analyze the data in the current study were factor analysis (FA) using principal component analysis (PCA), hierarchical cluster analysis (CA) and discriminant analysis (DA).

Normalization of Data

Many statistical tests are based on the assumption of normality. Hence, the distributions of the data for each of the variables were tested to see if they fit a normal distribution using a Kolmogorov-Smirnov goodness of fit test (Haan, 2002). Since the data for many of the variables were not found to be normally distributed, the data were transformed using the Box-Cox family of transformations given by (Box and Cox, 1964):

$$T(X) = \left(X^{\lambda} - 1\right) / \lambda \tag{3.5}$$

where X is the original variable and λ is the transformation parameter. For $\lambda = 0$, the data is transformed using the natural log. In the current study different values of λ were tried for each variable, and the transformed data were tested using Kolmogorov-Smirnov goodness of fit test. The λ was selected for which the transformed data was found to be distributed normally.

Factor Analysis/Principal Component Analysis

Principal component analysis under FA (Srivastava and Carter, 1983) was used to identify the factors most important for clustering the watersheds from the group of factors known to affect in-stream fecal coliform concentrations. Each of these selected principal components is a linear combination of some of the original variables. The number of principal components considered for CA is based on the percentage of variance explained by the factors. The criteria used to select the number of factors retained were the Kaiser criterion (Kaiser, 1960) and the Scree test (Cattell, 1966). According to the Kaiser criterion a factor is retained only if the eigenvalue is greater than 1. Essentially this means a factor is selected only if it extracts at least as much variance as the equivalent of one original variable. The Scree test is visual test where the eigenvalues are plotted against the number of factors. The general rule is to select the number of factors corresponding to a point beyond which the curve becomes approximately horizontal.

Cluster Analysis

The watersheds were grouped using hierarchical CA (Nathan and McMahon, 1990b; SAS, 1999). The hierarchical method uses a sequential method for forming clusters, starting with the most similar pair of objects, and then forming higher clusters in a step-wise fashion. The similarity measure generally used is the Euclidean distance which is given as:

$$d_{ij} = \left(\sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^2 \right)^{1/2} \tag{3.6}$$

where d_{ij} is the distance between the i^{th} and j^{th} observations, X_{ik} is the value of i^{th} observation for the k^{th} variable of p variables.

The most frequently used hierarchical cluster analysis method is Ward's minimum variance method (Kalkstein et al., 1987). The clustering is carried out by minimizing the within-cluster sum of squares, W, which is given as:

$$W = \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{N_k} \left(X_{ijk} - X_{.jk} \right)^2 \tag{3.7}$$

where K is the number of clusters, X_{ijk} is the value of the j^{th} variable for the i^{th} observation in the k^{th} cluster, J is the total number of variables, N_k is the number of observations in k^{th} cluster and $X_{.jk}$ is the k^{th} cluster sample mean of j^{th} variable.

Cluster Mean Comparisons

The means of the variables were compared across the clusters using Duncan's multiple range test. Results were compared at the 95% confidence interval ($\alpha = 0.05$). Based on the results from the mean comparisons the clusters were labeled as high, medium or low for the individual variables. This information along with the graphical plots of the means were used to determine the characteristics of the clusters.

Discriminant Analysis

To test the effectiveness of the clustering method and to determine the important parameters that discriminate the clusters, the results from the CA were analyzed using a DA technique. The stepwise DA adds one variable at each step, starting with no variable, and examines the model to check for variables failing to meet the criterion to remain part of the model. If all variables in the model meet the criterion, a new variable that contributes the most to the discriminatory power of the model is entered. When all variables in the model meet the criterion to remain and none of the other variables meet the criterion to enter, the stepwise selection process stops. Discriminant analysis generates a function called discriminant function similar to multiple regression, to determine the group membership criteria. Based on the discriminant function, DA produces a classification matrix which can be used to determine the effectiveness of a given classification scheme.

The DA technique generates a discriminant function, which is created as a linear combination of discriminating (independent) variables as given by Johnson and Wichern (2002),

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} . p_{ij} \tag{3.8}$$

where i is the number of groups (G), k_i is the constant inherent to each group (clusters in this case), n is the number of discriminating variables, w_i is the weight coefficient assigned by the DA to a variable and p_i is the analytical value of the variable. In the current study, based on the discriminant function, the DA will analyze the differences between groups and help to reassign to the appropriate cluster, waterbodies that have been wrongly assigned to a cluster by the CA based on the resemblance of the data.

Results and discussion

Water Quality Analysis

The list of water quality sampling points and the observed in-stream bacterial concentrations were obtained from the TCEQ. The USGS flow data for all the USGS gauging stations within the impaired watersheds were downloaded from the USGS website. The precipitation data corresponding to the NCDC weather stations were obtained from NCDC website. First, the bacterial stations closer to the USGS stations were identified and the observed bacterial concentrations were separated for periods belonging to baseflow and to rainfall events. Then, the concentration data for all the bacterial stations within a particular watershed were compiled together with the identifier that specified whether the data belonged to a rainfall event or a baseflow period. Then the means of observed in-stream fecal coliform bacteria concentrations for the two groups were compared to find out whether there was any statistically significant difference between them. This process was repeated for all the watersheds. The details of the results are given in Chapter II.

Table 3.1 Parameter λ used for Box-Cox transformation of variables affecting in-stream fecal coliform bacteria concentrations.

Variable	λ for Box-Cox Transformation
Percent Forest	0.44
Percent Cropland	0.12
Percent Water	Log Normal
Percent Residential	0.175
Percent Commercial	-0.1
Percent Wetland	0.2-0.3
Percent Barrenland	0.1
Percent Pasture	0.5
Distance Factor Forest	0.35
Distance Factor Cropland	0.2
Distance Factor Water	0.3
Distance Factor Residential	0.35
Distance Factor Commercial	0.4
Distance Factor Wetland	0.33
Distance Factor Barren	0.228
Distance Factor Pasture	0.02
Density of Households	0.005
Population Density	-0.05
Density of Other Septic systems	-0.1
Density of Public Sewers	0.01
Density of Conventional Septic Systems	-0.02
Loading from Deer	0.5
Loading from Farm Animals	0.35

Normalization of Data

The descriptive statistics of the watershed characteristics are given in Appendix B (Table B.1). The distributions of each variable were tested to see if they fit a normal distribution using a Kolmogorov-Smirnov goodness of fit test. Since the data for many of the variables were not conforming to normal distributions, the data were transformed using the Box-Cox family of transformations. The λ values used for each variable are listed in Table 3.1. For the percent water a log-normal transformation was found to be the best.

Principal Component Analysis

Based on the initial cluster analysis the parameters related to the distance factors were excluded from the FA/PCA. The results of the PCA are reported in Table 3.2 and Table 3.3. Based on the Kaiser Criterion and the scree plot as shown in Figure 3.5, six factors were retained for the cluster analysis. Table 3.2 shows the Varimax rotated factor loadings. The cumulative variance explained by the six factors was 97%. The parameters that have a magnitude of 0.6 (those underlined in the table) or more were considered to be contributing significantly to a particular factor. Thus factor 1 had 5 parameters with magnitudes greater than 0.6. These parameters are related to the human, thus might be termed as human factor 1. The second factor had two parameters with loadings greater than 0.6 in magnitude, percent wetland and average precipitation. The third factor could be termed as the second human factor or septic factor, since both OSSF density and density of the alternative septic system had a high positive magnitude for this factor. The rate of bacterial loading from livestock was highly correlated to the fourth factor. The fifth factor was found to have high magnitudes of forestland and cropland. The main component of sixth factor was the average temperature. Fourteen parameters were included to account for 97% of the overall variance. Other parameters had relatively low magnitudes on any of the factors retained for the analysis. Factor analysis/PCA shows the parameters having the greatest contribution to the overall variability of the dataset and also helps to identify the structure of the dataset.

Table 3.2 Varimax rotated factor loading for the first six factors[+].

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Density of Households	0.98	0.07	0.14	-0.01	-0.05	0.07
Population Density	0.97	0.10	0.14	0.00	0.01	0.12
Density of Public Sewers	0.97	0.06	0.13	-0.09	-0.03	0.04
Percent Residential	0.90	0.11	0.10	-0.22	-0.14	0.07
Percent Commercial	0.86	0.19	0.14	-0.03	-0.08	0.16
Percent Wetland	-0.05	0.83	-0.01	0.11	0.02	0.07
Average Precipitation	0.11	0.78	0.13	0.30	-0.25	0.09
Bacterial Loading from Deer	-0.30	-0.51	-0.10	0.46	0.05	0.01
Percent Barrenland	-0.36	-0.76	-0.12	0.03	-0.08	0.22
Density of Other Septic systems	0.26	0.10	0.92	0.15	-0.05	0.19
Density of OSSF	0.32	0.17	0.85	0.28	0.02	-0.02
Average Age of Households	0.04	-0.04	0.70	-0.30	-0.12	-0.10
Bacterial Loading from Farm Animal	-0.14	0.13	0.10	0.67	-0.20	-0.05
Percent Forestland	-0.03	0.26	0.01	0.58	-0.47	0.17
Percent Pastureland	-0.10	0.10	0.03	-0.05	0.69	-0.28
Percent Cropland	-0.15	-0.24	-0.14	-0.24	0.61	0.03
Percent Water	0.13	0.29	-0.21	-0.12	0.32	0.20
Average Temperature	0.18	0.00	0.04	0.01	-0.14	0.60

[+] Underlined values are considered to have significant loading for a magnitude greater than 0.6

Table 3.3 Variance explained by the factors.

Factors	Eigenvalues	Cumulative % of variance
1	4.91	36.50
2	2.48	54.97
3	2.26	71.77
4	1.42	82.33
5	1.36	92.44
6	0.66	97.37

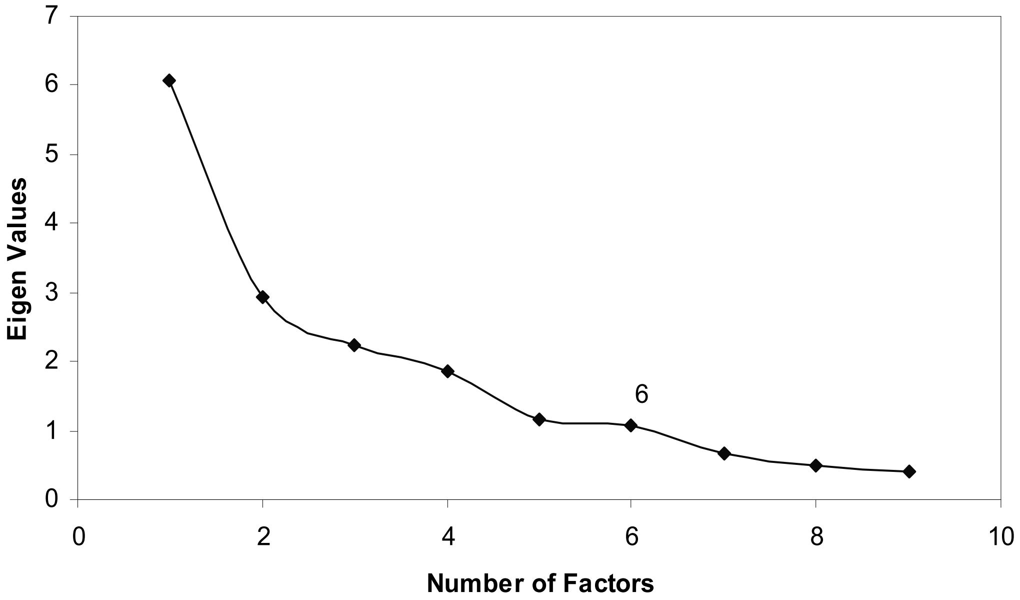

Figure 3.5 Scree plot for determining the number of factors to be retained.

Cluster Analysis

A CA was performed with the factors obtained during the FA/PCA step using Ward's minimum variance method. This method was selected after several trials using different methods. Ward’s method has been used in many cluster analysis studies (Alberto et al., 2001; Vega et al., 1998; Helena et al. 2000). The determination of the number of clusters was a difficult one and one without a perfect solution. After comparing the values for different criteria such as pseudo t^2 statistic, pseudo F statistic and cubic clustering criterion (CCC), the number of clusters decided on was six. The idea is to select the number of clusters corresponding to the local peak value in the pseudo F statistic and the CCC combined with a small value of the pseudo t^2 statistic and a larger pseudo t^2 for the next cluster fusion. In the current study there was no clear guidance from any of the criteria. After some initial analysis it was decided to obtain 6 clusters of watersheds.

Discriminant Analysis

The cluster analysis (CA) results were analyzed using the discriminant analysis (DA) technique. CA produced six different clusters and the cluster membership information was added to the original data matrix. Initially a stepwise DA was performed to obtain the discriminating variables. The summary result from the stepwise DA is given in Table 3.4. Some of the variables used in the CA procedure were excluded by the DA. Out of all the variables listed in Table 3.4, the variable that required particular attention is the F-value. The F-value for the variable in Table 3.4 indicates the statistical significance in the discrimination of clusters, or in other words the contribution by a variable in prediction of the cluster membership. In each step of the discriminant analysis the variable with the highest F-value will be selected. The Wilk’s Lambda value in Table 3.4 is the fractional amount of within-cluster variance, relative to the between cluster variance, that remains unaccounted for after each variable is entered or selected in each step of the discriminant analysis. As more variables are selected the value of Wilk's lambda decreases. The average squared canonical correlation (ASCC) is the proportion of the variance accounted for by the selected independent variables. The larger the value of ASCC, the better is the ability of the variables to capture the overall variance in the data matrix. As new variables are added, the value of ASCC is also increased. The results from stepwise DA were used to generate a classification matrix based on a discriminant criterion.

Table 3.4 Stepwise selection summary of DA.

Number		Parameter Entered	Partial R-Square	F Value	Pr>F	Wilk's Lambda	Pr< Lambda	Average Squared Canonical Correlation	Pr> ASCC
Step	In								
1	1	Age of Households	0.90	188.6	<.0001	0.099	<.0001	0.18	<.0001
2	2	Average precipitation	0.76	65.36	<.0001	0.024	<.0001	0.33	<.0001
3	3	Loading Rate from Farm Animal	0.63	35.35	<.0001	0.009	<.0001	0.45	<.0001
4	4	Households	0.47	18.14	<.0001	0.005	<.0001	0.53	<.0001
5	5	Forest Land	0.60	30.48	<.0001	0.002	<.0001	0.62	<.0001
6	6	Wetland	0.39	12.65	<.0001	0.001	<.0001	0.65	<.0001
7	7	Distance Factor-Pasture	0.26	6.79	<.0001	0.001	<.0001	0.69	<.0001
8	8	Alternate Septic Systems	0.26	6.76	<.0001	0.001	<.0001	0.70	<.0001
9	9	Pastureland	0.28	7.31	<.0001	0.000	<.0001	0.72	<.0001
10	10	Average Temperature	0.20	4.70	0.0007	0.000	<.0001	0.73	<.0001
11	11	Loading Rate from Deer	0.19	4.42	0.0012	0.000	<.0001	0.74	<.0001
12	12	Cropland	0.14	3.05	0.0135	0.000	<.0001	0.75	<.0001
13	13	Distance Factor - Water	0.15	3.12	0.012	0.000	<.0001	0.75	<.0001
14	14	Distance Factor-Forest	0.12	2.54	0.0337	0.000	<.0001	0.76	<.0001
15	15	Distance Factor-Residential	0.11	2.24	0.0472	0.000	<.0001	0.76	<.0001

The classification matrix is shown in Table 3.5 and the corresponding error statistics is given in Table 3.6. Table 3.5 shows the number of watersheds placed by DA into a specified cluster compared to the number of watersheds assigned in that cluster during CA. The larger the numbers in the diagonal elements in the matrix, the better is the grouping scheme based on the discriminating variables. It can be seen from the two tables that assignment of two waterbodies in the third cluster by CA was not in agreement with the assignment by DA, accounting for a 2% error rate. Also the discriminating parameters were found to be different from those selected for the CA. Hence, it was decided to rerun the CA using only the discriminating parameters. Priors indicate the probabilities of a particular item being grouped under a given cluster if the assignment was done randomly.

Table 3.5 Number of observations and percent classified into cluster.

Cluster	Quantity	1	2	3	4	5	6	Total
1	Number	32	0	0	0	0	0	32
	Percentage	100	0	0	0	0	0	100
2	Number	0	19	0	0	0	0	19
	Percentage	0	100	0	0	0	0	100
3	Number	1	1	28	0	0	1	31
	Percentage	3.23	3.23	90.32	0	0	3.23	100
4	Number	0	0	0	6	0	0	6
	Percentage	0	0	0	100	0	0	100
5	Number	0	0	0	0	10	0	10
	Percentage	0	0	0	0	100	0	100
6	Number	0	0	0	0	0	12	12
	Percentage	0	0	0	0	0	100	100
Total	Number	33	20	28	6	10	13	110
	Percentage	30	18.18	25.45	5.45	9.09	11.82	100
Priors		0.17	0.17	0.17	0.17	0.17	0.17	

Table 3.6 Error count estimates for cluster.

Cluster	1	2	3	4	5	6	Total
Rate	0	0	0.10	0	0	0	0.02
Priors	0.17	0.17	0.17	0.17	0.17	0.17	

Final Clusters

A CA using Ward's minimum variance method was performed with the factors obtained during the FA/PCA after selecting only the variables obtained as a result of the DA. Varimax rotated factor loadings corresponding to the factor analysis considering only the parameters selected during DA is given in Appendix B (Table B.2). Six clusters of waterbodies were obtained. The number of factors to be retained after the FA/PCA procedure was set to six. The number of clusters was also selected to be six. Once the six clusters were formed, the means of different watershed parameters were statistically compared among the clusters using Duncan's multiple range test with $\alpha = 0.05$. The results of the statistical test are given in Table 3.7. The values within the same parenthesis show the cluster numbers with means were not significantly different from each other. The clusters shown in different parentheses are significantly different, if they do not appear together in any of the other parentheses. For example, based on the Duncan's test the mean of the percentage forest land for cluster numbers 3 and 2 were significantly different from the means of all other clusters. At the same time mean of the percentage forest land for cluster 1 was not significantly different from that of clusters 5 or 4, where as it was different from cluster 3, 2 and 6. The mean of the percentage forest land for cluster 4 and cluster 6 were not significantly different from each, but mean of the percentage forest land for cluster 6 was different from all other cluster means. Thus, based on Duncan's multiple range test the mean of percentage forest land was not significantly different between cluster 1, 5 and 4, and cluster 4 and 6. The cluster number corresponding to the highest mean for a particular variable appears at the left most position in the row and means decrease from left to right.

Table 3.7 Results of cluster mean comparisons.

Parameters/Method	Duncan
Percent Forest Land	(3)(2)(1,5,4)(4,6)
Percent Cropland	(6)(4,1,2,3)(1,2,3,5)
Percent Urban Land	(5)(1,2)(3,6,4)
Percent Pasture	(6,1,4)(2,5,3)
Percent Residential	(5)(2,1)(3,6,4)
Percent Commercial	(5)(1,2)(3,6,4)
Population Density	(5)(2,1)(3,6,4)
Density of Households	(5)(2,1)(1,3)(3,6,4)
Density of OSSF	(1,3,2,6,4,5)
Age of Households	(5)(6,4)(4,3,1)(3,1,2)
Density of Public Sewer	(5)(2,1)(3,6,4)
Stream Density	(1,5,3,4,6,2)
Average precipitation	(3)(1,5)(2,6,4)
Average Temperature	(5,2,1,3,6,4)
Loading Rate from Deer	(4,2)(3,1,6,5)
Loading Rate from Farm Animal	(4)(3,1)(1,5,2)(6)

The means of different watershed parameters were plotted graphically for comparison. Figures 3.6-3.9 show the mean plots for the important watershed parameters. Analyzing the results from the mean comparison tests and the graphical plots some general conclusions were derived for different clusters and are given in Table 3.8. Table 3.8 explains the relative ranking of different clusters when compared using individual variables. For example, consider the means of percent forestland. Cluster 3 had the highest means compared to other clusters; the mean of cluster 2 followed that; and the means of clusters 1, 4, 5 and 6 were not significantly different from one another but were low compared to the mean of cluster 3 and 2.

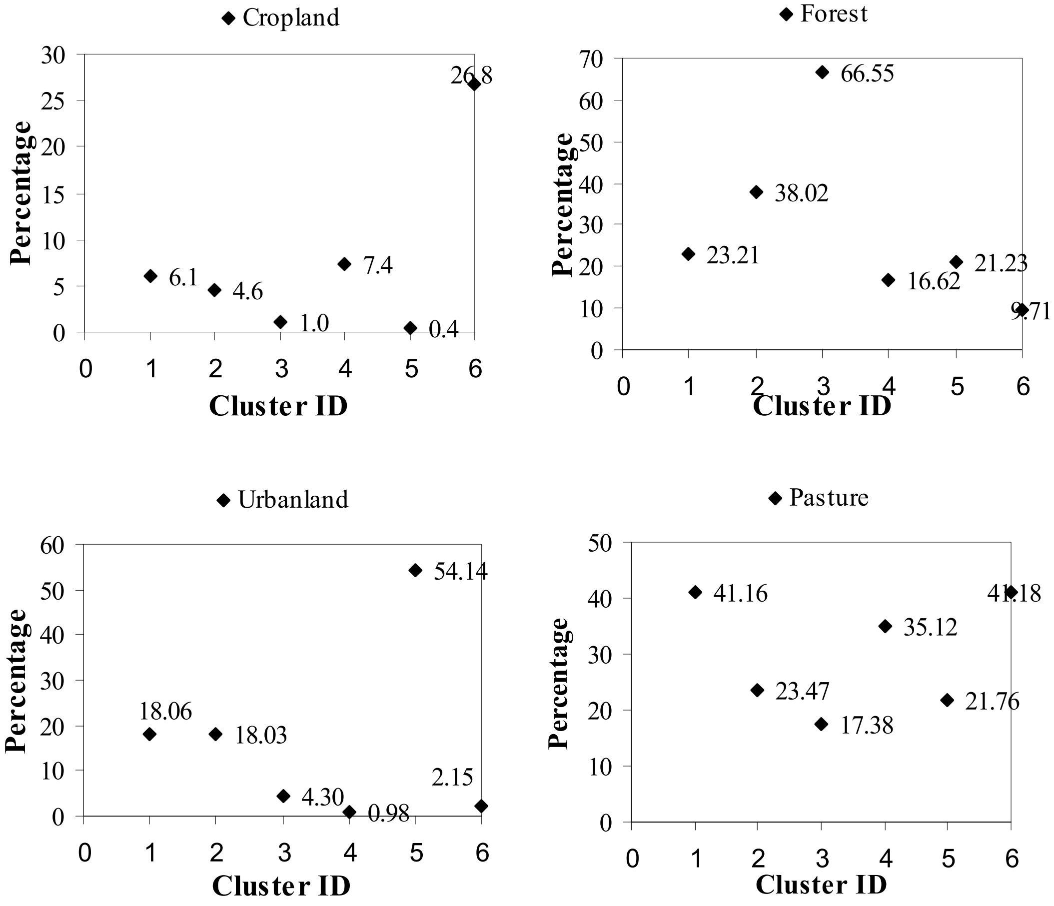

Figure 3.6 Mean percentages of different land uses within each cluster.

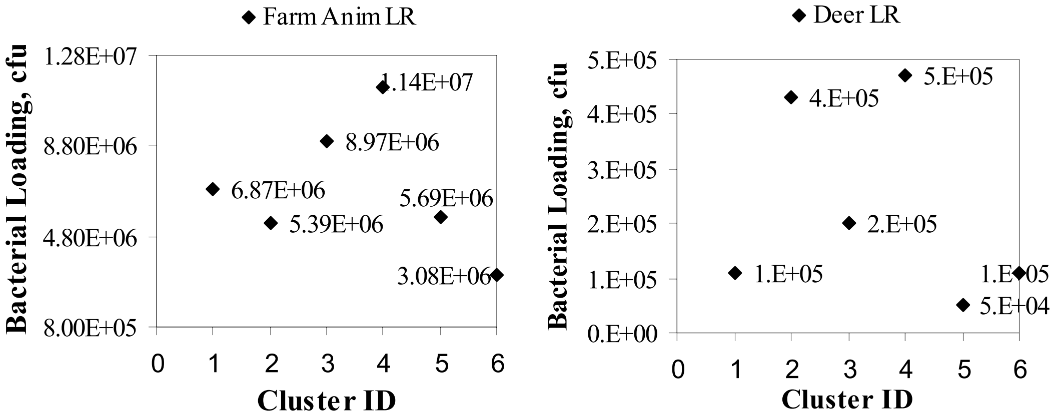

Figure 3.7 Mean bacterial loading rates from different animals for each cluster.

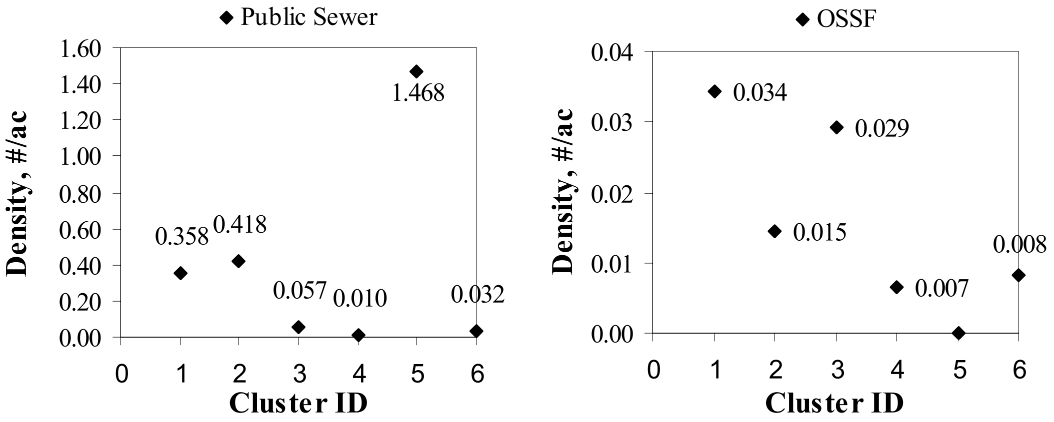

Figure 3.8 Mean densities of households under public sewer systems and OSSFs for each cluster.

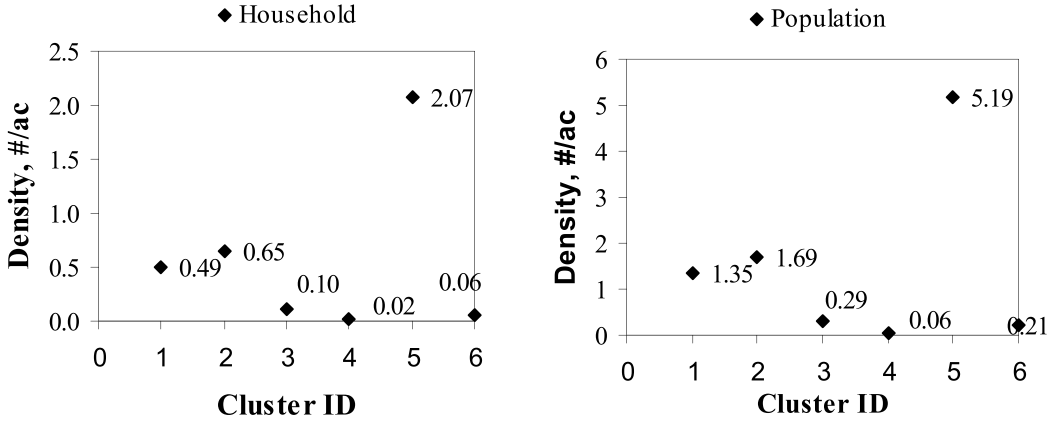

Figure 3.9 Mean population densities and densities of households.

Table 3.8 Comparison of important watershed characteristics among clusters.

Variable/Cluster	1	2	3	4	5	6
Frequency	39	18	24	12	6	11
Percent Forest	Low	Medium	High	Low	Low	Low
Percent Cropland	Medium	Medium	Low	Medium	Low	High
Percent Urban Land	Medium	Medium	Low	Low	High	Low
Percent Pastureland	High	Low	Low	High	Low	High
Population Density	Medium	Medium	Low	Low	High	Low
Density of Households	Medium	Medium	Low	Low	High	Low
Density of OSSFs	High	Low	Medium	Low	None	Low
Density of Public Sewers	Medium	Medium	Low	Low	High	Low
Density of Other Septic	High	Low	High	Low	Low	Low
Age of Households	Medium	Low	Medium	High	No OSSF	High
Loading from Deer	Low	High	Low	High	Low	Low
Loading from Farm Animals	Low	Low	Medium	High	Low	Low

Cluster 1

The first cluster contains 39 impaired waterbodies with relatively high densities of OSSF. Relative here and in the discussion of the clusters throughout, means in comparison to the other clusters. The major land use within these watersheds is pastureland. This cluster of waterbodies shows low bacterial loading from both farm animals and wildlife and relatively low public sewer use. The location of the waterbodies falling under the cluster 1 is shown in Figure 3.10. Table 3.9 shows means of in-stream fecal coliform concentrations during stormflow and baseflow periods for the stream segments in cluster 1. Table 3.9 also includes the conclusions of the statistical comparison of means during the two periods testing whether they are statistically different. Based on the statistical comparison of means, three stream segments showed higher means during baseflow periods and 16 stream segments showed higher means during stormflow periods. Though, there was no significant difference in the means for 13 of the stream segments, careful observation shows that the means during both stormflow and baseflow periods were higher than 400 cfu/100ml. A few of the streams, stream segment 0805 for example, showed higher means during the stormflow period, but also showed high mean bacterial concentrations during the baseflow period. This may indicate a relatively high contribution of bacteria from nonpoint sources, together with a noticeable contribution from point sources. There was no known facility which was permitted to discharge fecal coliform bacteria into stream segments. However, there is possibility of accidental discharge from Waste Water Treatment Plants (WWTP). The total discharge capacity of Waste Water Treatment Plants within the watersheds of the stream segment is given in Table 3.9. From Table 3.9 it is clear that except for three stream segments the total discharge capacity of WWTP was relatively low or negligible. This fact, along with a high mean concentration during baseflow periods can be interpreted as being the result of either a high background bacterial concentration in the streams, chronic failure of OSSF systems close to the streams, or a high density of animals within the accessible distance of the waterbody.

Table 3.9 Water quality characteristics of cluster 1 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means[+]	WWTP[++] capacity (mgd)
1101	43161	2687.0	2766.00	1601.00	Baseflow	2
1102	56669	2671.0	3184.00	994.00	Baseflow	15
2429	91969	16100.0	18340.00	5112.00	Baseflow	1
0202D	120445	2005.0	2303.00	562.00	ND	0
0203A	66439	358.4	328.00	427.00	ND	0
0505B	26002	3394.0	3522.00	3044.00	ND	0
0507A	90053	2808.0	2042.00	3614.00	ND	0
0507B	7764	1001.0	222.00	1520.00	ND	0
0611C	280526	1482.0	1561.00	53.00	ND	0
0819	171737	9517.0	7270.00	12620.00	ND	26
0841	129261	17320.0	1288.00	33870.00	ND	1237
1014	140163	11800.0	11470.00	12380.00	ND	80
1103	32466	3125.0	1752.00	24550.00	ND	0
1104	32466	3125.0	1752.00	24550.00	ND	0
1109	94240	905.7	668.00	1188.00	ND	0
1110	138429	1260.0	655.00	1985.00	ND	5
1209C	37438	569.5	594.00	309.00	ND	0
1245	167572	3192.0	3633.00	1296.00	ND	9
1304	63979	999.1	453.00	1557.00	ND	7
1304A	120646	473.3	291.00	544.00	ND	0
1803A	200120	1231.0	376.00	2886.00	ND	0
1901	487408	2849.0	4362.00	1548.00	ND	4
1108	57476	1018.8	45.00	1076.12	NED	0
0404B	25079	717.9	285.00	1216.00	Stormflow	0
0508	17019	5498.0	5172.00	6492.00	Stormflow	3
0508B	7792	5472.0	3958.00	10010.00	Stormflow	0
0511B	22473	942.7	879.00	1061.00	Stormflow	0
0604B	15317	1171.0	1068.00	1462.00	Stormflow	0
0605A	183995	909.9	834.00	1076.00	Stormflow	0
0804	923235	1296.0	179.00	1877.00	Stormflow	2978
0805	538659	46980.0	17290.00	93210.00	Stormflow	30
0806	172432	12910.0	1200.00	22090.00	Stormflow	0
0810	232524	368.6	30.00	1079.00	Stormflow	3
0901	28242	1101.0	99.00	1802.00	Stormflow	40
1009	142310	3520.0	1251.00	4538.00	Stormflow	28
1017	90835	17260.0	14180.00	20500.00	Stormflow	13
1502	149443	1635.0	345.00	1924.00	Stormflow	0
1911	315273	3480.0	2191.00	4694.00	Stormflow	3362
2426	21841	435.1	208.00	659.00	Stormflow	13

[+] ND = means of two periods are not significantly different, NED = Not enough data to compare; [++] WWTP – Waste Water Treatment Plant

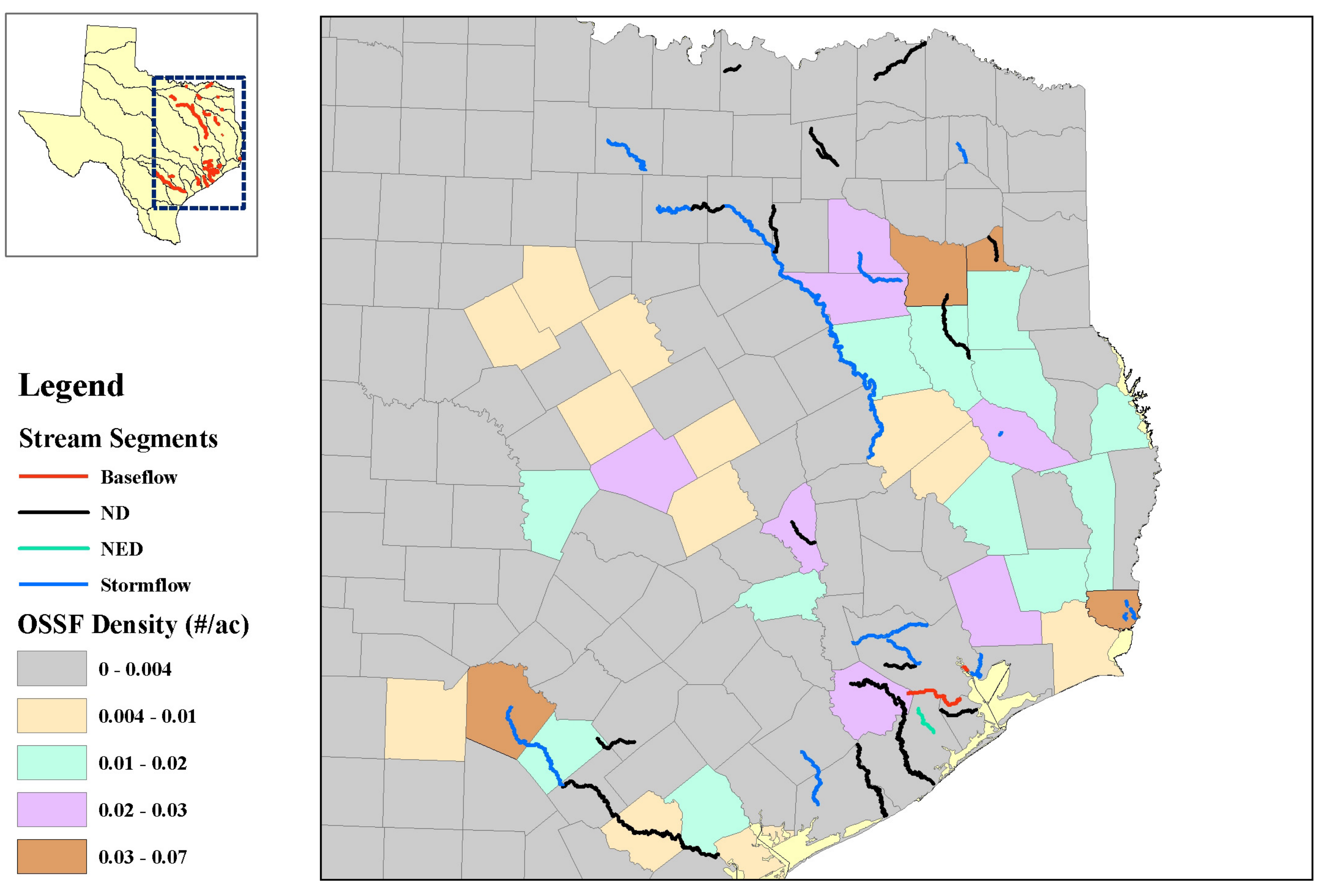

Figure 3.10 Locations of stream segments belonging to cluster 1.

For one of the stream segments, 1108, the data was not adequate to test the difference in means between the two flow conditions.

Cluster 2

Cluster 2 is a group of 18 impaired waterbodies with well mixed land use distribution within the watersheds. The contribution of bacteria loading from wildlife is high compared to other clusters. These waterbodies had low bacteria loading from farm animals and relatively low densities of OSSF and public sewers. Table 3.10 shows some important water quality characteristics of the waterbodies within this cluster. Figure 3.11 shows the locations of the waterbodies belonging to cluster 2.

Table 3.10 Water quality characteristics of cluster 2 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means+	WWTP capacity (mgd)
1430	78649	379.9	424	280	Baseflow	1
1218	72758	413.2	443	408	ND	6
1427	195978	1190.0	1336	817	ND	1
1428B	35821	1222.0	380	1599	ND	0
1428C	46190	673.9	628	937	ND	0
1811A	83133	684.9	718	479	ND	0
1903	260613	416.66	597	1435	ND	56
1906	105374	2381.0	4495	1090	ND	14
0101A	32509	384.4	384.4	--	NED	0
1427A	17951	264.8	176.38	1327	NED	0
1427B	34585	863.4	863.4	--	NED	0
1427C	32459	--	--	--	NED	0
1403A	20924	454.4	392	546	Stormflow	0
1414	791996	391.3	301	693	Stormflow	3
1428	344199	875.4	872	879	Stormflow	904
1429B	8818	279.9	277	288	Stormflow	0
1910	112734	4247.0	601	7242	Stormflow	1
2110	135550	211.6	73	342	Stormflow	0

+ND = means of two periods are not significantly different, NED = Not enough data to compare

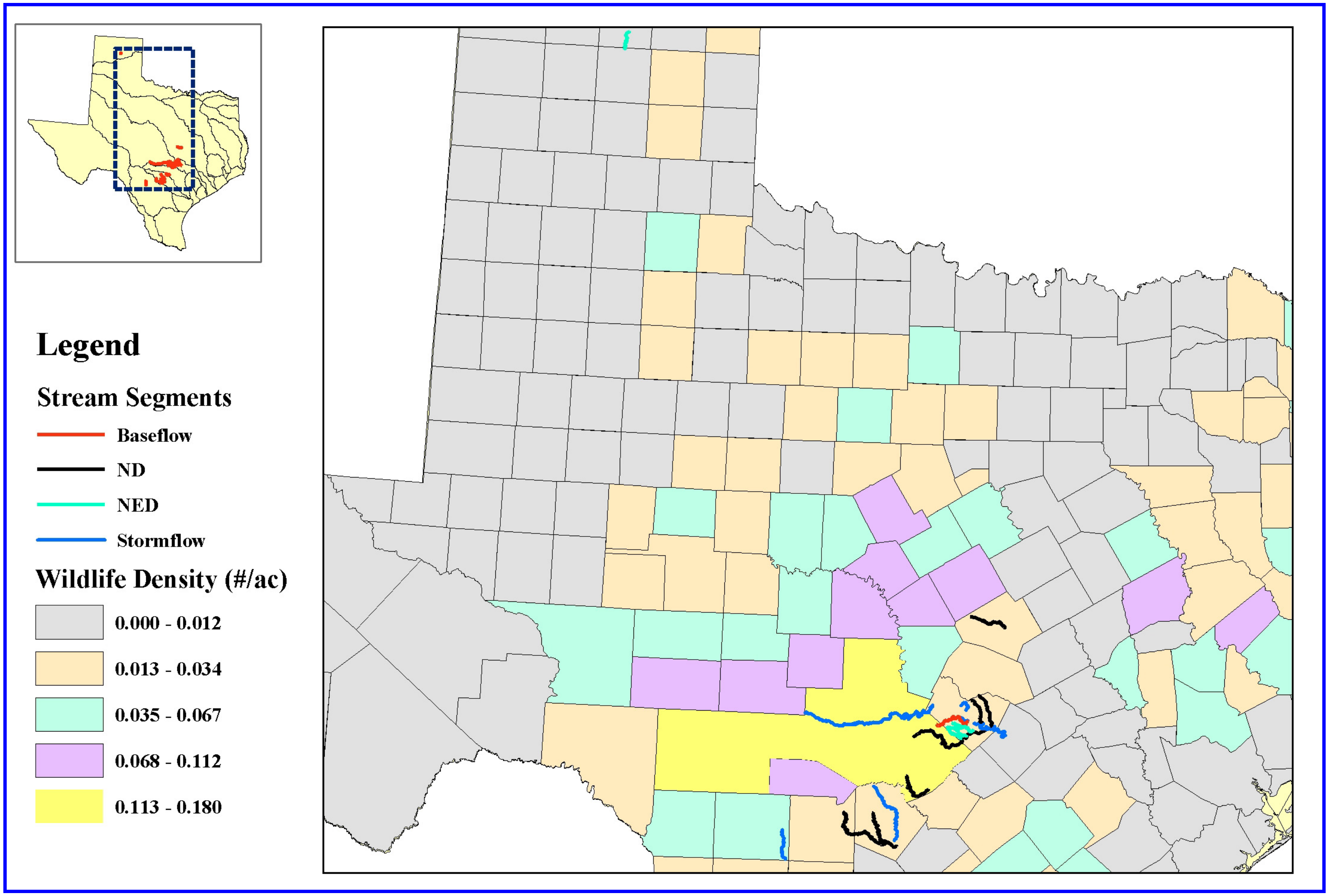

Figure 3.11 Locations of stream segments belonging to cluster 2.

One of the waterbodies showed a higher mean concentration of bacteria during baseflow periods than during stormflow periods and six waterbodies showed higher mean bacteria concentrations during stormflow periods. Although stream segments such as 1427, 1428C, 1903 & 1906 showed no significant difference between the means during stormflow and baseflow periods, they did show a reasonably high mean concentration during both stormflow and baseflow periods. Overall, the waterbodies in cluster 2 showed lower mean bacterial concentrations during both stormflow and baseflow periods than the other clusters. Higher concentrations during stormflow periods may be attributed to the high contribution of bacterial load from the wildlife in this cluster.

Cluster 3

Cluster 3 contains 24 impaired waterbodies with a high density of OSSF and high bacterial loading from farm animals. The major land use within these watersheds is forestland. This cluster of waterbodies has low bacterial loading from wildlife and relatively low public sewer use. The locations of the impaired waterbodies falling into cluster 3 are shown in Figure 3.12. Table 3.11 shows some important water quality characteristics of the waterbodies within cluster 3. One of the waterbodies had a higher mean concentration of bacteria during baseflow periods compared to stormflow periods and 12 waterbodies had higher mean concentrations during stormflow periods. The higher concentrations during stormflow periods may be attributed to the high contribution of bacteria from failed OSSFs and farm animals. The data was not adequate to test for a significant difference of the mean concentrations between the two flow conditions for three of the waterbodies. Eight waterbodies had no significant difference between the means during stormflow and baseflow periods and showed relatively low mean concentrations during both periods.

Table 3.11 Water quality characteristics of cluster 3 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means+	WWTP capacity (mgd)
0607	364631	3338.0	3752	2160	Baseflow	2
0505D	94455	1832.0	659	5499	ND	0
0604	1084810	2922.0	1642	4743	ND	5
0608C	127517	488.1	390	1161	ND	0
0608D	95996	475.0	467	646	ND	0
0610A	89394	357.3	304	471	ND	0
0612B	19368	373.9	449	274	ND	0
1001	59651	1312.0	1415	640	ND	20
0502A	74347	904	904	--	NED	0
0608B	140397	629.35	531	2549	NED	0
0608F	69710	449.7	377	1345	NED	0
0508A	47340	5445.0	5032	6683	Stormflow	0
0511	67670	1075.0	914	1421	Stormflow	24
0511A	73789	1762.0	924	4217	Stormflow	0
0511C	42852	2327.0	1140	6537	Stormflow	0
0513	174096	498.8	263	1019	Stormflow	1
0603A	38617	441.1	294	1266	Stormflow	0
0604A	75884	877.9	735	1382	Stormflow	0
0604C	19090	528.5	438	970	Stormflow	0
0607C	60223	270.6	263	307	Stormflow	0
0611	502576	468.7	390	872	Stormflow	599
0611B	52821	871.4	648	1425	Stormflow	0
1008	162812	1219.0	173	1958	Stormflow	16
1403	223646	164.5	76	342	Stormflow	1

+ND = means of two periods are not significantly different, NED = Not enough data to compare

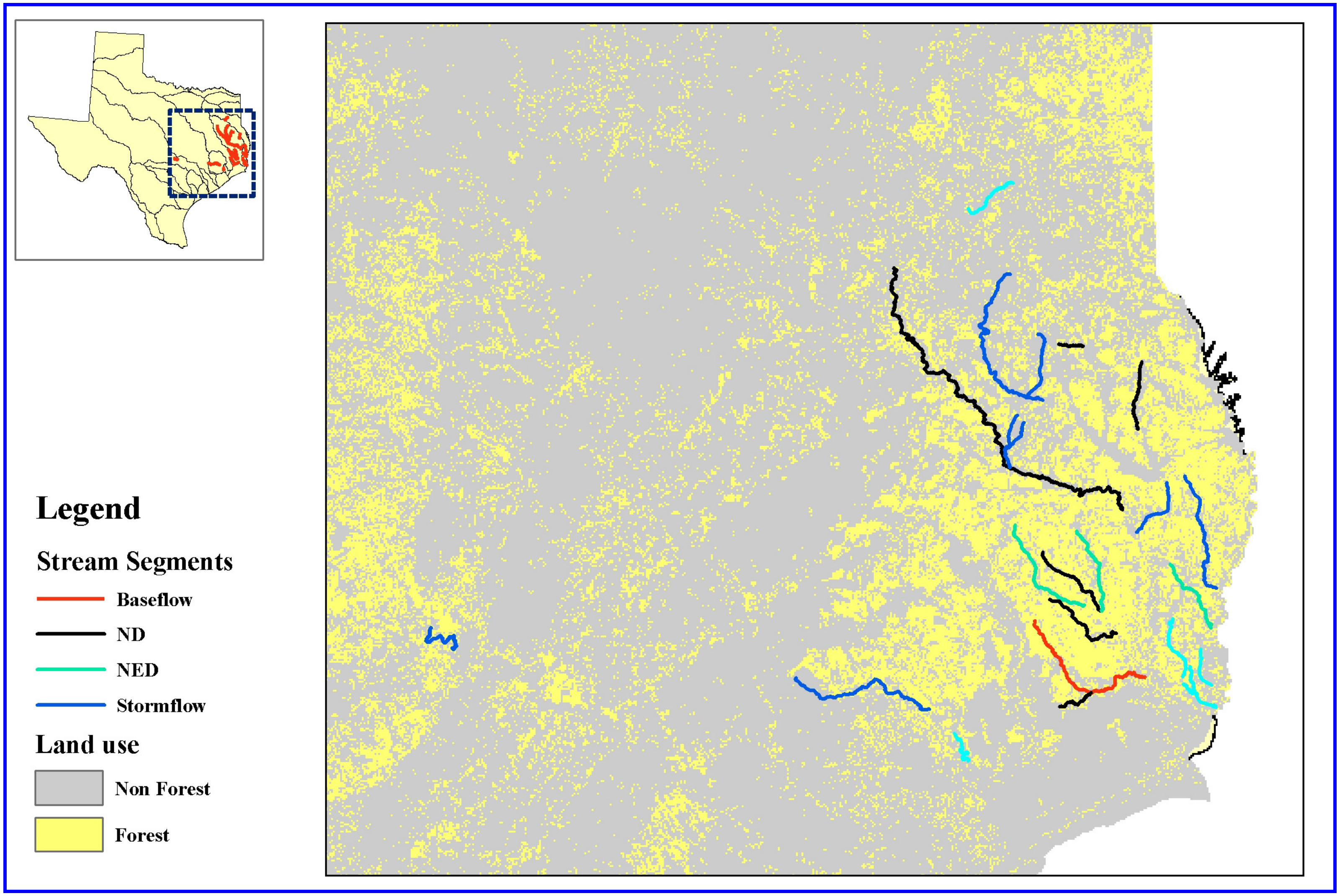

Figure 3.12 Locations of stream segments belonging to cluster 3.

Cluster 4

Twelve impaired waterbodies fall into cluster 4. These waterbodies have high bacterial loading from both farm animals and wildlife. The main land uses in these watersheds are pastureland and cropland. The densities of OSSF and public sewers are low compared to waterbodies in the other clusters. The locations of the impaired waterbodies falling into cluster 4 are shown in Figure 3.13. Table 3.12 shows the mean of in-stream fecal coliform concentrations during stormflow and baseflow periods for the stream segments in cluster 4 and the conclusions based on statistical comparison of means during the two periods.

Table 3.12 Water quality characteristics of cluster 4 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means$^+$	WWTP capacity (mgd)
2304	2345780	54610.0	56610	709	Baseflow	30
2306	4482980	536.5	561	323	Baseflow	0
2116	132120	798.33	560	1147	ND	0
1217A	122866	215.9	166	248	ND	0
1222A	19262	288.0	288	--	NED	0
1255	85146	4661.0	2189	10250	Stormflow	1
1804B	399467	851.1	555	2479	Stormflow	0
1226	505641	849.4	572	1154	Stormflow	2
1226A	57937	447.3	97	1148	Stormflow	0
1221	765811	326.4	249	487	Stormflow	2
1226D	87565	192.4	86	296	Stormflow	0
1226C	119715	163.4	154	180	Stormflow	0

$^+$ND = means of two periods are not significantly different, NED = Not enough data to compare

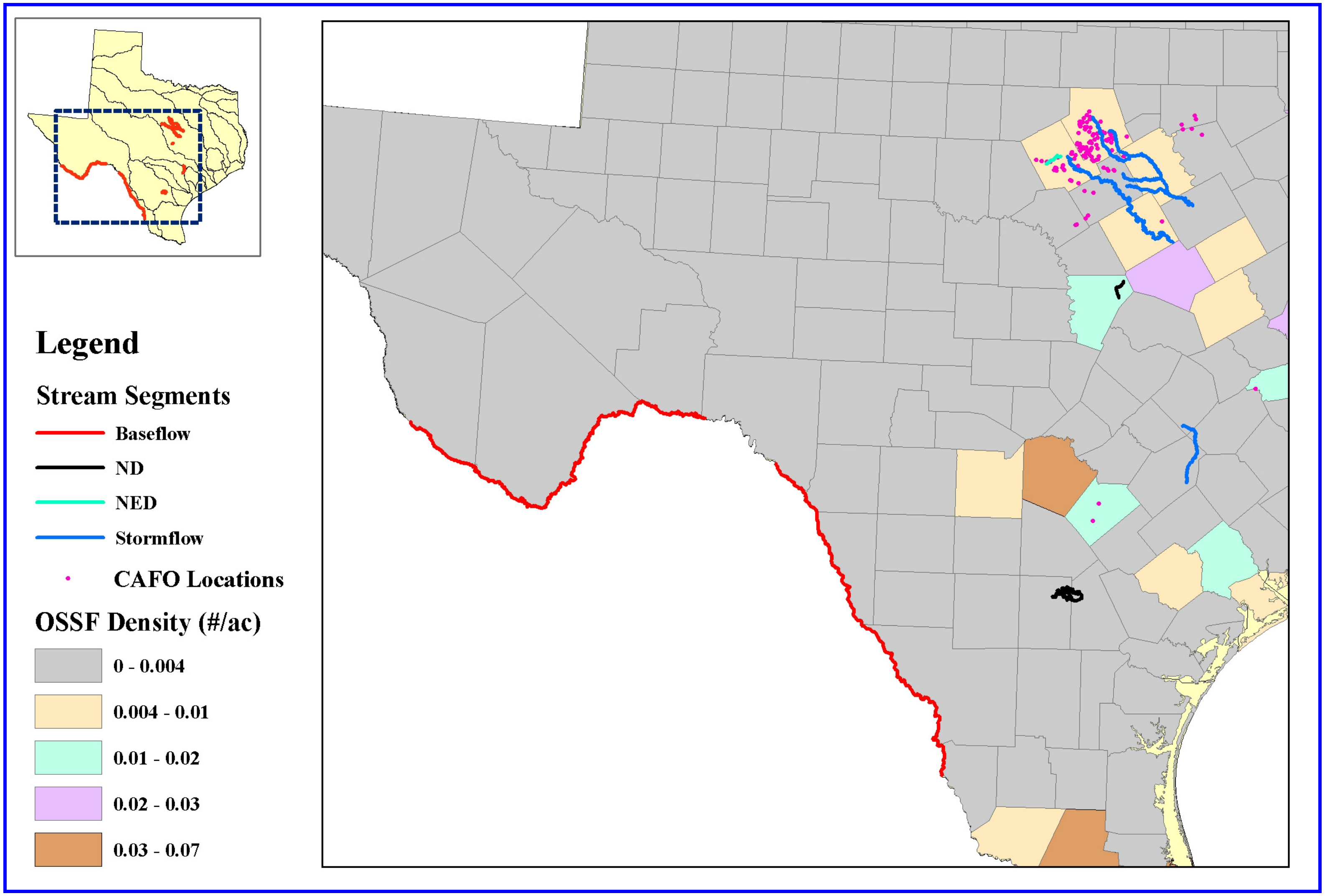

Figure 3.13 Locations of stream segments belonging to cluster 4.

Based on a test of significant difference between the means, two stream segments showed higher mean concentrations during baseflow periods and seven stream segments showed higher mean concentrations during stormflow periods. Although stream segment 1255 showed relatively higher mean concentrations during stormflow periods, it also exhibited high mean concentrations during baseflow periods. This may be an indicator of the presence of farm animals within accessible reach of the waterbody. The total discharge capacity of WWTP was relatively low or negligible for the waterbodies in cluster 4. One of the stream segments had insufficient data to test for a significant difference of mean bacterial concentrations between the two flow conditions.

Cluster 5

Cluster 5 consists of six impaired waterbodies in highly urbanized watersheds. The density of households connected to public sewers is the highest in this cluster compared to the other clusters. One characteristic of this cluster that separates it from other clusters is the absence of OSSF within the watersheds. Figure 3.14 shows the locations of the impaired waterbodies in cluster 5. These watersheds are located completely within major urban areas. Table 3.13 shows some important water quality characteristics of the waterbodies in cluster 5. Three of the stream segments (1016, 1013, and 1113A) showed very high bacteria concentrations during both stormflow and baseflow periods. Two stream segments had significantly higher mean in-stream fecal coliform concentrations during baseflow period. However, all the stream segments in this cluster had a relatively high mean concentration during baseflow periods. Since the watersheds of these stream segments fall completely within major urban areas of Texas, the higher concentration of bacteria during baseflow periods can be an indicator of point sources from public sewage systems.

Table 3.13 Water quality characteristics of cluster 5 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means+	WWTP capacity (mgd)
1016	99441	13620.0	15250.00	5747.00	Baseflow	37
1429A	10118	641.8	814.00	551.00	Baseflow	0
1013	61612	13320.0	12690.00	14800.00	ND	0
1113	9958	2853.0	631.00	6873.00	ND	0
1113A	8197	4875.0	1937.00	8400.00	ND	0
1428A	9927	141.6	206.00	60.00	ND	0

+ND = means of two periods are not significantly different, NED = Not enough data to compare

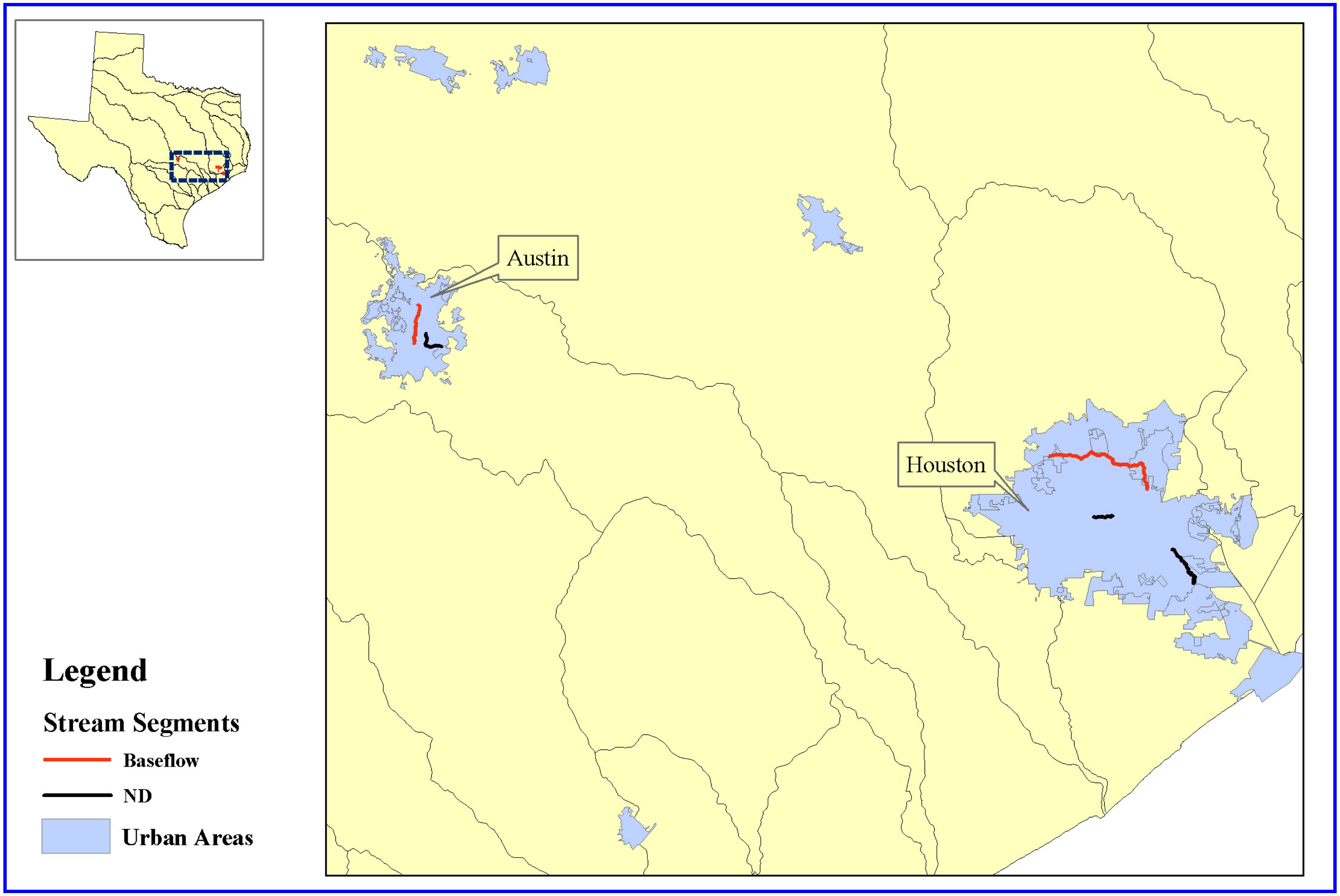

Figure 3.14 Locations of stream segments belonging to cluster 5.

Cluster 6

Cluster 6 contains 11 impaired waterbodies with low contributions of bacteria from any source. The watersheds are predominantly pastureland and cropland. The locations of the waterbodies in cluster 6 are shown in Figure 3.15. Table 3.14 shows the water quality comparison results for the stream segments in cluster 6. Four stream segments showed reasonably high mean bacteria concentrations during both stormflow and baseflow periods, though the contribution of bacteria from any source is not evident.

Table 3.14 Water quality characteristics of cluster 6 stream segments.

Segment ID	Watershed area (acres)	Overall mean (cfu/100ml)	Baseflow period mean (cfu/100ml)	Stormflow period mean (cfu/100ml)	Period of larger means⁺	WWTP capacity (mgd)
2202	198136	2562.0	2636.00	966.00	Baseflow	40
0105	66962	98.5	98.00	99.00	ND	0
0205	551863	3100.0	2283.00	4085.00	ND	6000
0902	83833	896.0	173.00	713.00	ND	5
2117	412677	556.6	601.00	491.00	ND	1
2302	2719860	1167.0	1194.00	980.00	ND	3
0207A	234645	467.47	446.22	658.00	NED	0
0306	143204	771.3	808.70	322.50	NED	0
0204	1094790	977.0	1437.00	3191.00	Stormflow	0
1242	1087540	389.0	568.00	1888.00	Stormflow	1440
2107	438802	1773.0	738.00	2517.00	Stormflow	2

⁺ND = means of two periods are not significantly different, NED = Not enough data to compare

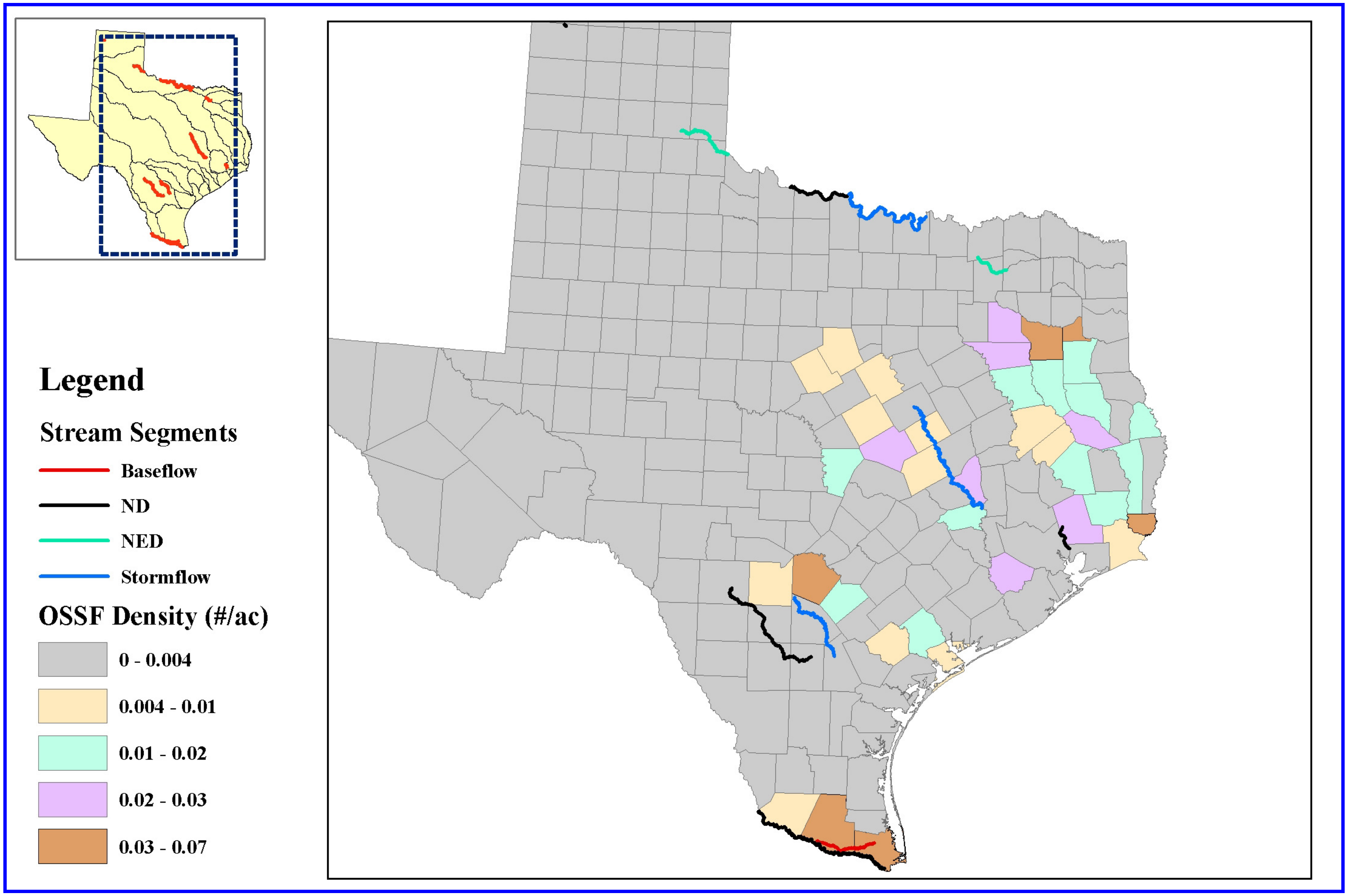

Figure 3.15 Locations of stream segments belonging to cluster 6.

Conclusions

The Texas waterbodies listed for bacterial water quality violation under CWA §303(d) were clustered based on their watershed characteristics. A pooled T-test was used to test for a significant difference in baseflow and stormflow bacterial concentrations. The results of the T-test were used to help identify the potential sources of bacterial pollution in each watershed. The impaired waterbodies were grouped into six homogenous clusters based on their watershed characteristics using the multivariate statistical techniques of factor analysis/principal component analysis, cluster analysis, and discriminant analysis. The conclusions derived from the current study are summarized below:

- The primary watershed characteristics that differentiate the clusters are bacterial contribution from farm animals and wildlife, density of OSSF, density of households connected to public sewers, and the land use distribution.
- Few of the watersheds found to share a border with other states or another country (Mexico). The data collection was done only within the boundaries of the state of Texas. This may have left some of the potential sources of bacteria out of the analysis. The effect of potential sources across the boundaries on the in-stream water quality in these watersheds should be studied.
- The presence of point and nonpoint sources within the watershed boundaries was apparent for many watersheds regardless of their membership to a certain cluster.
- Higher concentrations during both stormflow and baseflow periods may indicate both point and nonpoint sources of bacteria. However, in a few of these waterbodies there were no known point sources and the density of livestock and/or wildlife were high in the contributing watersheds. It may be appropriate to believe that the livestock has direct access to many of the bacterially impaired stream segments.

- It was found that the sample size of bacterial observations for eleven of the stream segments was not enough to carry out the statistical comparison between stormflow and baseflow flow conditions.
- The currently available information on domestic pets and migratory birds was insufficient to be incorporated into the multivariate analysis.

The use of GIS was found to be very useful in disaggregating the data available at County level or State level into watershed level for analysis. However, an extensive data collection at watershed level will greatly improve the results. The incorporation of location specific knowledge on the application of manure on land surface would increase the accuracy of the results. Similarly, collecting the numbers of failing septic systems on a watershed basis would improve the results.

CHAPTER IV

SELECTION OF MODEL FOR BACTERIAL TMDL

Synopsis

Under the Clean Water Act program the Texas Commission on Environmental Quality listed 110 stream segments in Texas with pathogenic bacteria impairment in the year 2000. In order to test the hypothesis of grouping the stream segments based on watershed characteristics and conducting a modeling study using a group-wise approach, the impaired stream segments within five river basins Brazos, Neches, Nueces-Rio Grande Coastal, Sabine and San Antonio were selected based on their association with expected point and nonpoint sources. Using different multivariate statistical techniques such as, factor analysis/principal component analysis, cluster analysis, and discriminant analysis the waterbodies were clustered into five groups having similar watershed characteristics. In the current study 20 different water quality models were reviewed to determine the appropriateness of using them to predict in-stream fecal coliform concentrations for stream segments in any of the five clusters. HSPF was found to be the only water quality model that can be used for all the clusters of stream segments.

Introduction

The transmission of human pathogenic agents via source water and treated water has been reported extensively in the literature (Barwick et al, 2000; Cruz et al., 1990; Carter et al., 1987). It is reasonable to believe that human population growth and anthropogenic activities such as intensive animal rearing and feeding operations may be partially responsible for the contamination of water bodies. Given that a variety of microbial pathogen contamination sources do exist and will probably continue to exist in the future, modern society has to identify those water bodies that are contaminated or potentially vulnerable to contamination, and develop management and remediation strategies for those systems. However, evidence is mounting that these microbial organisms survive and proliferate in some ecosystems under specific climatic conditions. The primary sources of pollution to the waters of the US are urban and agricultural runoff (USEPA, 1998b). The most common pollutants from these non-point sources are nutrients, bacteria, and silt (USEPA, 1998b). The persistence of potential microbial pathogens from wastes in soil and water is a constant concern.

Though fecal coliforms are not pathogens themselves, they are usually used as indicators of potential risk for water contamination. Fecal coliforms are a group of bacteria that primarily live in the lower intestines of warm-blooded animals and humans. Fecal coliforms are associated with warm-blooded mammals and are rare or absent in unpolluted waters. Many water-borne diseases, such as dysentery and cholera, are associated with certain strains of *E. Coli*, which is but one category of fecal coliforms. Because of the serious potential health threat associated with certain strains of this general type of bacteria, the presence of fecal coliform is a very important indicator of the health risk associated with human contact with a body of water.

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) listed 110 stream segments in the year 2000 with pathogenic bacteria impairment. The next logical step is to verify impairment and if sufficient evidence is present then to develop Total Maximum Daily Loads (TMDLs) for each of these impaired waterbodies. But the development of such a large number of TMDLs will require a colossal resource of both money and man power. There is a probability that many of the waterbodies considered for TMDL development listed under the current CWA §303(d) for Texas may be grouped based on their watershed characteristics and the potential sources of pollution. Such a grouping scheme will be helpful in reducing the cost of restoration of water quality by restricting the development of TMDLs for only one or two representative waterbodies under a single group and applying the knowledge to other waterbodies in the same group. In order to test this hypothesis the impaired stream segments within five river basins Brazos, Neches, Nueces-Rio Grande Coastal, Sabine and San Antonio were selected based on expected point and nonpoint sources. Using different multivariate statistical techniques such as, factor analysis (FA)/principal component analysis (PCA), cluster analysis (CA), and discriminant analysis (DA) the waterbodies were clustered into five groups having similar watershed characteristics. These five groups are different from those mentioned in Chapter III. The first group of stream segments showed a high density of public sewer systems and a relatively high percentage of urban land use compared to the other clusters of stream segments. The second group had a high density of On-site Sewage Facilities (OSSFs), and a relatively high contribution of bacterial loading from livestock. The third group of stream segments was located in forested areas with a relatively high contribution of bacterial loading from livestock, relatively low bacterial loading from wildlife and low densities of OSSFs. The fourth group of stream segments had a high percentage of pastureland but low bacterial loading from animals and low densities of OSSFs. The fifth group of stream segments was located in areas with high percentages of pastureland and cropland and had high bacterial loading from both livestock and wild animals. The primary aim of the current study is to select appropriate water quality models for each of these groups of stream segments so that the in-stream bacterial concentration in the stream segments can be predicted.

The restoration of water quality of an impaired stream starts with acquiring knowledge about the system, including the amount and sources of pollutant loading. One of the essential components of developing a TMDL is to establish this linkage between the sources and numeric indicators used to measure the attainment of uses. This linkage is most often made by using a combination of monitoring data and modeling tools.

Though there are number of models that can be used to predict fecal coliform concentrations, the selection of the appropriate model for a particular stream segment depends on many factors. Nix (1990) gives criteria for selection of an appropriate model, including, hardware availability, availability of trained personnel, long-term usability of the model, experience of other modelers in using the model, technical support available, and commitment to the modeling process. Nix (1990) also points out that the use of a maladapted model may lead to complications of the problem because of misleading results. Though complex models tend to be effective in reproducing the processes of interest, they are highly dependent on the input data. Since complex models require more input data, usage of such models are highly limited by the availability of the required data.

For the current study the selection of an appropriate model is based on the following criteria:

- Simulation time step: Daily/Sub-daily, Fixed/Flexible
- Simulation mode: Continuous/storm event/both
- Land use addressed: Agricultural/urban/both
- Sources addressed: Nonpoint/point/both
- Location of waterbodies: Flowing rivers/shallow lakes/deep lakes or reservoirs

Water quality modeling

The fate of chemicals in an aquatic system is determined by their reactivity and the rate of their physical transport through the system (Schnoor, 1996*)*. In general, water quality parameters are measured in mass quantities or concentration units. The mathematical models used to predict the concentrations of a chemical or a pollutant are representations of the fate and transport of the pollutant within the aquatic system. Most often they use a mass balance approach to calculate the concentration of a pollutant within the system. Pollutant transport through water depends primarily on two phenomena: advection and dispersion. Advection is the movement of dissolved material in any direction due to the flow of water, while dispersion is the mixing of substances within the water column. Based on the principle of conservation of mass and Fick's law, the basic equation describing advection and dispersion can be written as (Schnoor, 1996):

$$\frac{\partial C}{\partial t} + \sum_{i=x,y,z} \left(u_i \frac{\partial C}{\partial x_i} \right) = \sum_{i=x,y,z} \left(\frac{\partial}{\partial x_i} E_i \frac{\partial C}{\partial x_i} \right) - R \tag{4.1}$$

where, C is the concentration of the pollutant, E is the dispersion coefficient, t is the time, u_i is the average velocity in the i direction, x_i is distance in the i direction, and R is the reaction transformation rate. For one-dimensional flow in rivers under nonsteady flow conditions the equation becomes (Schnoor, 1996):

$$\frac{\partial (AC)}{\partial t} + \frac{\partial (QC)}{\partial x} = \frac{\partial}{\partial x} \left(EA \frac{\partial C}{\partial x} \right) - AR \tag{4.2}$$

where, Q is the volumetric flow rate, and A is the cross-sectional area.

General hydrologic/hydraulic equations

Though there are many water quality models available, the backbone of all such models is the governing hydrologic and hydraulic equations. A few of the hydraulic and hydrologic equations are described in the following section since they are common to many of the models reviewed under the current study.

Methods for calculating runoff

Rational Method (Bras, 1990)

The rational method used to calculate the peak discharge rate of runoff from a watershed is given as:

$$Q_p = CIA \tag{4.3}$$

where, Q_p is the peak discharge rate (cfs), C is the dimensionless runoff coefficient, A is the drainage area (acres), and I is the rainfall intensity having a duration equal to or larger than the time of concentration of the drainage basin (in/hr).

Unit Hydrograph Method

One method to calculate the runoff from the land surface due to a certain quantity of excess rainfall is the unit hydrograph method. Rainfall excess is the amount of rainfall available for runoff. Unit hydrograph is defined as the discharge produced by a unit volume of rainfall excess of a given duration (Bras, 1990). The unit hydrograph method assumes that the river basin responds linearly to the rainfall excess.

Soil Conservation Service (SCS): Curve Number Method

The SCS curve number is a common method used by many models to calculate the total runoff due to rainfall excess. The equation is given as (Bras, 1990):

$$R = \frac{(P - 0.2S)^2}{P + 0.8S} \tag{4.4}$$

where, $S = \frac{1000}{CN} - 10$ (4.5)

R is the accumulated runoff volume, P is the accumulated precipitation, S is the soil water retention parameter, and CN is the curve number. The SCS curve number is a function of the soil's permeability, land use and antecedent soil moisture conditions. The CN values are generally available for different land uses and hydrologic soil groups.

Methods for channel routing

St. Venant's Equations for Gradually Varied Flow

In general, flow of water in a river can be described using the principles of conservation of mass and momentum. The pair of equations for the conservation of mass and momentum is called Saint Venant's equations (Chow et al., 1988) and are given as:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} - q = 0 \tag{4.6}$$

$$\frac{1}{A}\frac{\partial Q}{\partial t} + \frac{1}{A}\frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + g\frac{\partial y}{\partial x} - g\left(S_0 - S_f\right) = 0 \tag{4.7}$$

where, *A* is the effective flow area, *Q* is the volumetric flow rate, *t* is time coordinate, *x* is the space coordinate, y is the depth of flow, *q* is the lateral inflow or sources/sinks, g is the acceleration due to gravity, S_0 is the channel bed slope, and S_f is the friction slope.

Muskingum Method (Chow et al., 1988)

Another method commonly used for channel routing is the Muskingum method given by:

$$S = K\lfloor xI + (1 - x)O \rfloor \tag{4.8}$$

where, S is the storage of water in the channel reach, K is the storage constant, I and O are the inflow to the reach and outflow from the reach, respectively, and x is the constant that expresses the relative importance of inflow and outflow in determining storage.

Modeling of bacteria - factors affecting fecal coliform kinetics

There are a number of factors, which affect the fate of bacteria in general and fecal coliform in particular. These factors can be divided into physical, physicochemical, and biochemical and biological factors. Some of the important factors and the way they affect the fate of fecal coliform are given in Table 4.1.

A review of available modeling tools

There are a number of mathematical models used for modeling fecal coliform. Some of the models used are Agricultural Runoff Management II: Animal Waste Version (ARM II) model (Overcash et al., 1983); the Utah State (UTAH) model (Springer et al., 1983); the MWASTE model (Moore et al., 1988); and the COLI model (Walker et al., 1990). All of these models calculate the bacterial die-off using the first order exponential decay expressed as Chick's Law (Crane and Moore, 1986) directly or with some modifications. According to Chick's law, the die-off of fecal coliform bacteria follows a first order decay rate given by the equation:

Table 4. 1 Factors affecting the fecal coliform die-off rate (Paul et al., 2002).

Category	Factors	Effect
Physical	Photo-oxidation	Light increases the mortality of fecal coliform
	Adsorption	Affects fecal coliform mortality, but inconclusive data available
	Flocculation	Affects fecal coliform mortality, but inconclusive data available
	Coagulation	Affects fecal coliform mortality, but inconclusive data available
	Sedimentation	May decrease the mortality rate by depositing the fecal coliforms to the bottom of the stream bed
	Temperature	This is the most important factor affecting the fate of bacteria. Other than directly affecting the mortality rate, temperature affects other factors which affect the mortality rate of bacteria
	Salinity	Salinity has an inverse effect on *E. Coli* survival
Physicochemical	pH	Generally E. Coli survives longer in lower pH
	Chemical toxicity	In general the presence of heavy metals reduces the bacterial concentration
	Redox potential	The higher the redox potential the higher is the mortality rate of bacteria when heavy metals are present
	Nutrient level	Increase in nutrient level may increase amount of in-stream fecal coliform
Biochemical-biological	Presence of organic substance	May decrease mortality rate
	Predators	May increase the mortality rate
	Algae	In general, detrimental to bacteria because of production of toxic substance along with algal boom.
	Presence of fecal matter	Increases the concentration of fecal coliform.

$$\frac{N_t}{N_0} = 10^{-kt} \tag{4.9}$$

where, N_0 and N_t are number of coliform bacteria at time 0 and at time t days, respectively and k is the first order decay rate.

Though there are a large number of water quality models developed by different agencies and institutions, not all of them are appropriate for modeling bacteria. Different models are developed to address different types of aquatic systems. Among the models available, the following water quality models were selected for a detailed review:

- QUAL2E-UNCAS
- HSPF
- QQS
- STORM
- SWMM
- MIKE-SWMM
- WQRRS
- WASP
- EFDC
- CE-QUAL-RIV
- CE-QUAL-W2
- AUTO_QI
- PLUMES
- MIKE-BASIN
- PLOAD
- SLAMM
- CORMIX
- SWAT

A detailed description of each of these models is given in the following section.

QUAL2E-UNCAS (Brown and Barnwell, 1987)

The current version of QUAL2E-UNCAS is a result of many years of evolution. The original version of program, QUAL-1 was developed by F.D. Masch and Associates and the Texas Water Development Board (TWDB, 1970) in the 1960s. The QUAL2E model (Brown and Barnwell, 1987) first released in 1985 was developed by Tufts University and the USEPA. QUAL2E is a stream water quality model with the capability of simulating the interaction of up to 15 water quality constituents, including dissolved oxygen, biochemical oxygen demand, temperature, algae as Chrolophyll *a*, organic nitrogen as N, ammonia as N, nitrite as N, nitrate as N, organic phosphorus as P, dissolved phosphorus as P, coliforms, an arbitrary nonconservative constituent and three conservative constituents.

Modeling Concept

QUAL2E is a one-dimensional, steady-state model that assumes that a stream is composed of a number of completely mixed computational elements of equal length that are connected sequentially to one another. Each of the computational elements are assumed to have the same hydrogeometric properties such as stream slope, channel cross section, channel roughness, and biological rate constants such as BOD decay rate, benthic source rate, algae settling rates. The hydrologic balance can be written as:

$$Q_i = Q_{i-1} + \left(Q_x\right)_i \tag{4.10}$$

where, $\left(Q_x\right)_i$ is the sum of the external inflows and/or withdrawals to that element and i is the element index.

The model simulates solute transport by solving the one-dimensional advection-dispersion mass transport equation for each water quality constituent. The equation can be written as:

$$\frac{\partial C}{\partial t}=\frac{\partial\left(A_x D_L \frac{\partial C}{\partial x}\right)}{A_x.\partial x}-\frac{\partial\left(A_x \bar{u} C\right)}{A_x.\partial x}\frac{dC}{dt}+\frac{s}{V} \qquad (4.11)$$

where, C is the concentration of the constituent (ML^{-3}), A_x is the cross-sectional area of the computational element (L^2), D_L is the dispersion coefficient (L^2T^{-1}), $\bar{u}$ is the mean velocity (LT^{-1}), x is the distance along the stream (L), t is the time (T), s are external sources or sinks (MT^{-1}), and V is (A_x . dx) the incremental volume (L^3). The equation is solved using a finite difference method (the classical backward difference method).

In simulating coliform bacteria, QUAL2E assumes that coliform die-off follows a first order decay function, which can be written as:

$$\frac{dC}{dt}=-kC \qquad (4.12)$$

where, C is the coliform concentration (ML^{-3}), t is the time (T) and k is the coliform die-off rate (T^{-1}). The die-off rate, k is dependent on temperature. Hence the model accepts the values of k at 20°C as input and is then modified to temperature computed by the model using a Streeter-Phelps type formulation:

$$k_T = k_{20}\theta^{(T-20)} \qquad (4.13)$$

where, k_T and k_{20} are die-off rates at the local temperature T and at 20°C, respectively, and θ is the temperature correction factor.

Temperature is modeled by using a heat balance on each computation element. The equation can be written as:

$$H_n = H_{sn} + H_{an} - H_b - H_c - H_e \tag{4.14}$$

where, H_n is the net energy flux passing the air-water interface ($HL^{-2}T^{-1}$), H_{sn} is the net short wave solar radiation($HL^{-2}T^{-1}$), H_{an} is the net long wave radiation($HL^{-2}T^{-1}$), H_b is the outgoing long-wave back radiation flux($HL^{-2}T^{-1}$), H_c is the convective energy flux passing back and forth between the interface and the atmosphere($HL^{-2}T^{-1}$), and H_e is the energy loss by evaporation ($HL^{-2}T^{-1}$).

Input Requirements

- Stream network
- Dispersion and shear velocity parameters
- Headwater source data such as flow, temperature, and concentration
- Incremental flow and concentration information

Model Outputs

- Hydraulics information such as flows, velocities, travel time, depths, and cross-sectional areas along each reach
- Reaction coefficients for simulated constituents
- Constituent concentrations along a reach
- A summary of temperature calculations may also be included

Limitations

The dimensional limitations imposed in the current version of QUAL2E are: a maximum of 50 reaches, no more than 20 computational elements per reach or a total of 500 computational elements, a maximum of 10 headwater elements, a maximum of 9 junction elements, a maximum of 50 point source and withdrawal elements.

HSPF (Bicknell et al., 1997)

The Hydrological Simulation Program -FORTRAN (HSPF) was developed by EPA in the mid-1970's. The model was built on the Stanford Watershed Model and incorporates the concepts from several related models. HSPF is considered to the most comprehensive and flexible model of watershed hydrology and water quality available (Zoppou, 2001). The HSPF model is also widely used for bacterial TMDL studies across United States.

Modeling Concepts

HSPF is a continuous hydrological modeling software that can be used to simulate a comprehensive range of hydrologic and water quality processes. In HSPF, modules are divided into pervious land (PERLND*)*, impervious land (IMPLND), and reaches (RCHRES*)*. Each land segment is considered as a lumped catchment and simulations are based on an hourly time step.

The hydrologic simulation in HSPF is based on a mass balance approach. On pervious land surfaces the amount of precipitation is divided into the following components; direct runoff, direct evaporation from the land surface, surface storage followed by evaporation, surface storage followed by interflow, and infiltration to the subsurface zone. The subsurface zone is divided further into upper zone, lower zone and deep groundwater zone. Water reaching the subsurface layer either remains in storage, percolates to the zone below the current zone or evaporates. Water that percolates deep below the groundwater zone is considered to be lost from the system. The amount of precipitation over the impervious land surface contributes to overland flow, evaporates directly, or remains in surface detention storage. The surface runoff calculation is based on Chezy-Manning's equation and an empirical relationship between the outflow depth and detention storage. The evapotranspiration (ET) is calculated based on the user supplied potential ET and moisture available in surface and subsurface zones to meet the potential ET demand.

The water quality simulation in HSPF assumes constituents are associated with one or more of the following; sediment, overland flow, interflow, and groundwater flow. Most frequently, fecal coliform is assumed to be flow associated. Flow associated constituents are assumed to be accumulated on the land surface until the occurrence of a rainfall event. The contribution of fecal coliform from the land surface to the stream is calculated using the following equation:

$$SOQO = SQO\left[1.0 - e^{(SURO \times WSFAC)}\right] \tag{4.15}$$

where, SOQO is the washoff of fecal coliform from the land surface (colony forming unit [cfu] acre^{-1} day^{-1}) , SQO is the storage of fecal coliform on the surface (cfu acre^{-1}), SURO is the surface outflow of water (in. acre^{-1}), and WSFAC is the susceptibility of the quality constituent to washoff (in.$^{-1}$). The amount of storage on the land surface depends on the accumulation rate of bacteria and the maximum limit of storage depending upon the landuse type. The contribution of fecal coliform through interflow or groundwater flow is calculated using constant concentrations in these flows. The in-stream fecal coliform dynamics are calculated based on Chick's law.

Input Requirements

- Stream geometry
- Land use distribution
- Hourly precipitation, potential evapotranspiration, and lake evaporation
- Land use dependent pollutant build-up and washoff information
- If snow is simulated, additional climatic data such as wind speed, solar radiation, air temperature and cloud cover are required

Model Outputs

- Timeseries of both water quantity and quality
- Optional frequency and duration analysis results

QQS (Geiger and Dorsch, 1980)

The Quantity-Quality Simulator (QQS) model is a comprehensive mathematical model that can perform both continuous and single event simulations using a five minute time step. The model is used to simulate unsteady state runoff over the urban drainage area and flow in the storm and combined sewer system and the receiving water bodies. The flow is routed through the main and interceptor sewers and other structures, such as branches, overflows, basins, pump stations, control gates, and treatment facilities. The QQS is capable of simulating water quality constituents such as BOD, COD, total suspended solids, settleable solids, total N, total P and fecal coliform. Limited information is available about the application of QQS for fecal coliform. It was used to simulate the fecal coliform and some other pollutants for the combined sewer systems of the City of Rochester, NY (Geiger and Dorsch, 1980)

Modeling Concepts

Runoff from pervious and impervious drainage area is calculated separately using unit hydrograph approach. Hence, the study area should be subdivided into small drainage areas of size ranging from 30 to 100 ha to meet the unit hydrograph linearity assumptions.

The runoff ordinate after a certain rainfall event is calculated using the equation:

$$Q(t) = \int_0^t u(t-\tau) I_{eff}(\tau) d\tau \tag{4.16}$$

where, $Q(t)$ is the runoff hydrograph ordinate value at time t, $u(t-\tau)$ is the transformation relation, I_{eff} is the effective precipitation, and τ is the duration of the unit hydrograph. The effective precipitation for impervious areas is calculated as:

$$I_{eff}(t) = I_{tot}(t) - I_{st}(t) \tag{4.17}$$

where I_{tot} is the total gaged precipitation and I_{st} is the amount of initial abstraction, which includes the wetting loss and depression storage. The effective precipitation for pervious areas is calculated as:

$$I_{eff}(t) = I_{tot}(t) - I_{st}(t) - I_{si}(t) \tag{4.18}$$

where, I_{si} is the soil infiltration loss.

Pollutant discharge to the sewers is calculated by modifying the unit hydrograph to include pollutant related parameters. The pollutant washoff rate at time t, $P(t)$ is calculated as:

$$P(t) = \int_0^t RDI(t)\, DTI(d)\, \phi(t-\tau)\, I_{eff}(\tau)\, d\tau \tag{4.19}$$

where, d is the time of day, $DTI(d)$ is the function of diurnal variation of pollutant washoff due to land use, $RDI(t)$ is the function for decrease in pollutant washoff with increasing rainfall duration, and $\phi(t-\tau)$ is the pollutant unit impulse response. The pollutant is assumed to be built-up on the land and is calculated as:

$$P_{up}(z) = \left(1 - \frac{c_1 z + 1}{c_2 z^3 + c_3 z^2 + 1}\right) P_{up,100} \tag{4.20}$$

where, $P_{up}(z)$ is the amount of pollutant accumulated after z number of time steps (of 5 minutes), $P_{up,100}$ is the maximum pollution build-up (typical value for fecal coliform is between $5\text{x}10^{12}$ and $5\text{x}10^{13}$ ha^{-1}), and c_1, c_2, and c_3 are the coefficients in the range of 10^{-4} to 10^{-7} depending upon the pollutant.

Flow in the sewer systems and receiving water bodies is simulated using an implicit finite difference approximation of the kinematic wave equations. The energy equation used is given as:

$$S_f = S_0 - \frac{\partial H}{\partial x} + \frac{\partial}{\partial x}\frac{(v^2)}{2g} + \frac{1}{g}\frac{\partial v}{\partial t} \tag{4.21}$$

and the continuity equation is given as:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = 0 \tag{4.22}$$

where, A is the flow cross-sectional area, g is the acceleration due to gravity, H is the flow depth, Q is the flow rate, S_f is the frictional slope, S_0 is the slope of the channel bottom, t is the time and x is the coordinate. The baseflow of receiving waters is entered as table of mean diurnal discharge values for the simulation period and are superimposed on the flow from sewer systems and the resulting amount is routed using the above equations.

Pollutant transport is treated as plug flow and is calculated at systems nodes using the continuity equation:

$$\sum_{i=1}^{3} P_{c,i} + \sum_{i=1}^{3} P_{b,i} = \sum_{i=1}^{2} P_{a,i} \tag{4.23}$$

where, P_a is the pollutant load at the upper end of the downstream element, P_b is the pollutant load at the lower end of the upstream element and P_c is the pollutant load from the drainage area. The background pollution of the receiving water bodies are supplied and are added to the pollutant loading from the sewer systems.

Input Requirements

- Precipitation intensities at five minute intervals and dry spells between the individual events
- Geometrical data defining the sewer network and the receiving water systems, and drainage area characteristics
- Unit flow hydrographs and data such as initial abstraction and soil infiltration required to compute effective rainfall
- Pollutant load unit hydrograph and their modifying functions

Model Outputs

Model outputs can be separated for single event simulations and continuous simulations. The outputs from the single event simulations are:

- Hydrographs and pollutographs for each node
- Water surface elevations, flows, pollution loads and concentrations for each time interval for each node
- Total runoff and pollutant loads washed off from different drainage areas

Outputs for continuous simulation are given as statistics and they include:

- Annual and monthly frequencies
- Cumulative frequencies
- Durations

The above statistics are available for variables such as, discharge and overflow volumes, peaks, averages, intensities, and associated pollution values.

STORM (HEC, 1977)

The original version of STORM model was developed by Water Resources Engineers, Inc. in 1973 under a contract with the Hydrologic Engineering Center (HEC). The model is designed to model urban watersheds and is capable of calculating loads and concentrations of water quality parameters such as suspended and settleable solids, biochemical oxygen demand, total nitrogen, orthophosphate, and total coliform. STORM is also capable of calculating land surface erosion. STORM is used to aid in sizing of storage and treatment facilities to control the quantity and quality of storm water runoff and land surface erosion. A continuous simulation model, STORM requires hourly precipitation data to model the seven storm water elements such as rainfall/snowmelt, runoff, dry weather flow, pollutant accumulation and washoff, land surface erosion, treatment rates, and detention reservoir storage. Dust and dirt and the associated pollutants are washed off from the watershed by the rainfall. The runoff is routed to the treatment-storage facilities and the effect of treatment is calculated. Runoff in excess of treatment plant capacity is stored and treated later except for the quantity in excess of storage, which is waste untreated and becomes overflow directly into the receiving waters.

Modeling Concepts

The runoff of water is computed by one of the three methods, coefficient method, the SCS Curve Number technique, or a combination of the two. By the coefficient method runoff volume is given by:

$$Q = C(P - f) \tag{4.24}$$

$$C = C_p + (C_I - C_p)\sum_{i=1}^{L} X_i F_i \tag{4.25}$$

where, Q is the runoff (in.), P is the rainfall/snowmelt over the area in (in.), f is the available depression storage (in.), C is composite runoff coefficient, C_p and C_i are the runoff coefficients for pervious and impervious surfaces, respectively, X_i is area in land use i as a fraction of total urban watershed area, F_i is the fraction of land use i that is impervious, and L is the total number of land uses.

The model computes a soil moisture balance at the beginning of each time increment by the following equation:

$$S_t = S_{t-1} - IN \times \Delta t + A \times EV \times \Delta t + B \times MP \times \Delta t \tag{4.26}$$

where $A = 0.7((SM - S_{t-1}) / SM)^v$ (4.27)

$$B = ((SM - S_{t-1}) / SM)^p \tag{4.28}$$

where, S is the soil moisture capacity for storage of water (in.), IN is the maximum infiltration rate from initial abstractions (in./hr), EV is the pan evaporation rate (in. /hr), MP is the maximum soil percolation rate (in./hr), SM is the maximum soil moisture capacity for storage of water (in.), t is the time, Δt is time increment (1 hr), v is the exponent regulating evapotranspiration, and p is the exponent regulating percolation.

Dry weather flow in the combined sewer systems is computed by specifying either the total waste water flow and infiltration flow (mgd), domestic, commercial, industrial, and infiltration flow separately or the coefficients required to compute the individual flows based on population, and areas under commercial and industrial land.

The STORM model calculates the pollutant loadings based on either dust and dirt method or the daily pollutant accumulation method. The dust and dirt method calculates pollutants as fractions of the dust and dirt for each land use. The amount of the dust and dirt is calculated based on accumulation rate specified in terms of pounds/100 feet of gutter length /day for each land use. The factors such as the intensity of rainfall, rate of runoff, the accumulation of dust and dirt on the watershed, and the frequency and efficiency of street sweeping operations affects the amount of pollutants entering the storm drains and the treatment facilities or the receiving waters. The initial quantity of a pollutant on a particular land use at the beginning of a storm is computed as:

$$P_p = \begin{cases} F_p DD_L N_D + P_{po}; \quad \mathit{if}\ N_D < N_s \\ P_{po}(1-E)^n + N_s DD_L F_p[(1-E)^n + (1-E)^{n-1} + ...(1-E)] \\ \qquad +DD_L F_p(N_D - nN_s);\ \mathit{if}\ N_D < N_s \end{cases} \tag{4.29}$$

where, P_p is the total pollutant p on land use L at the beginning of the storm (pounds), F_p is the pollutant p per unit mass of dust and dirt (pounds), N_D is the number of days without runoff since the last storm, P_{po} is the pollutant remaining on land use L at the end of the last storm, N_s is the number of days between street sweeping, n is the number of times the street was swept since the last storm, and E is the efficiency of the street sweeping expressed as a fraction.

The coliform wash-off from the watershed is calculated as:

$$M_{coli} = P_{coli}(1 - e^{-KR_I}) \tag{4.30}$$

where, R_I is the runoff rate from the impervious surface for the coefficient method or total runoff for the SCS method and combination method (in./hr) and K is the wash-off decay coefficient.

Dry weather coliform loading in the combined sewer systems is computed similar to the flow calculation during the same period by specifying either the total waste water flow and infiltration flow (mgd), domestic, commercial, industrial, and infiltration flow separately or the coefficients required to compute the individual flows based population, areas under commercial and industrial land.

Input Requirements

The input data requirements of the STORM is relatively less because of its simpler hydrologic and water quality routines compared to other continuous simulation models. The general inputs required are:

- Runoff coefficients
- SCS parameters
- Hourly precipitation

Model Outputs

- Runoff volume
- Summaries of storage and treatment utilization
- Total overflow loads and concentrations
- Statistical information on quantity and quality of washoff and overflow
- Pollutographs for individual storm events

SWMM (Huber et al., 1984; Huber and Dickinson, 1988; Roesner et al., 1988)

An urban storm water model, Storm Water Management Model (SWMM) was developed in 1969-71 by a consortium of Metcalf and Eddy, Inc., Water Resources Engineers, Inc., and the University of Florida (Donigian and Huber, 1991). Over the years the model went through many developments and the latest version is Version 4. The SWMM is capable of simulating pollutants such as; suspended solids, settleable solids, BOD, COD, total nitrogen, total phosphorous, oil/grease, total coliforms, and an arbitrary pollutant.

Modeling Concepts

The model consists of many modules, also called blocks, capable of simulating both water quality and quantity processes in the urban storm water runoff and combined sewer overflow. SWMM is both a continuous and single event model. Surface runoff is calculated using the rainfall intensities and antecedent moisture conditions, land use and topography. Using a simple nonlinear reservoir storage, Runoff Block simulates rainfall-runoff processes and the snowmelt processes, including infiltration depression storage, evaporation and surface runoff. The Extran Block simulates backwater, surcharging, looped sewer connections and a variety of hydraulic structures. Storage/Treatment Block is used to simulate storage-indication flow routing. There are different options available for water quality simulation. Pollutant accumulation over time can be calculated using both linear and nonlinear accumulation method. A simple relationship of runoff or first-order decay method can be used to obtain the washoff of the pollutants. Sewer flows are generated using land use, population density and other factors. The routing of flows and pollutants through the sewer system is based on the modified kinematic wave approximation and the assumption of complete mixing.

Input Requirements

- Information on watershed area, imperviousness, slope, roughness, depression storage and infiltration characteristics
- Channel or pipe data such as shapes, dimensions, slopes, and roughness
- Build-up coefficients if the quality constituent is modeled using build-up/washoff formulation
- Precipitation data in the form of hyetographs for individual storm events, or long-term hourly or 15-minute magnitudes

Model Outputs

- Hydrographs and pollutographs at any point in system on time step or longer basis
- Removal quantities in storage/treatment units, generated sludge quantities
- Summaries of volumes and pollutant loads for simulation period
- Daily, monthly, annual and total summaries for continuous simulation
- Statistical analysis of continuous/ single event output

MIKE-SWMM (DHI, 2003a)

The MIKE-SWMM is a combination of Danish Hydraulic Institute *(DHI)'s* MIKE 11(Havno et al., 1995) and SWMM (Huber et al., 1984; Huber and Dickinson, 1988; Roesner et al., 1988). The coupling gives the strengths of MIKE 11 in one-dimensional unsteady flow modeling, which solves the shallow water wave equations using an implicit finite difference scheme, replacing the Extran Block in SWMM. The MIKE-SWMM model can simulate hydrology, hydraulics and water quality of storm water and waste water drainage systems such as waste water treatment plants and water quality control devices. The runoff from the single storm events are simulated using a unit hydrograph approach.

The MIKE-SWMM model can simulate water quality parameters such as: total coliform, total phosphorous, total nitrogen, DO, temperature, ammonia, nitrate, heavy metals, suspended sediments and bed sediments, and BOD as both dissolved and attached to suspended sediments. The model can be interfaced with other DHI models.

Model Concepts

Runoff is calculated using a mass balance approach for simulating surface detention, lower soil storage, and upper and lower groundwater storage. Runoff includes both overland and baseflow. The pollutant transport is calculated using the one-dimensional advection-diffusion equation.

WQRRS (HEC, 1978)

The Water Quality for River-Reservoir Systems (WQRRS) is a product brought out by combining a reservoir simulation developed by Chen and Orlob and a river simulation model developed by Norton, to model the entire basin of the Trinity River System in Texas. Later HEC developed a preprocessor to simplify the input preparation and the program was then called WQRRS (HEC, 1978).

Modeling Concepts

The streamflows are dynamically routed using either the St.Venant equations, Kinematic Wave, Muskingum, or Modified Puls routing methods. The model consists of a reservoir module, a stream hydraulic module, and a stream quality module, where first two modules can be run independently, while the stream quality module needs the hydraulic data file created by the hydraulic module.

Reservoir Module

In the reservoir module a reservoir or a lake is represented by a series of one dimensional, fully mixed, homogenous horizontal slices, each having characteristics such as an area, thickness and volume. The water movement and advection are governed by the location of inflow to, and outflow from, the reservoir. The allocation of withdrawals through an outlet gate to individual reservoir elements is calculated using either Debler-Craya method or the WES method, while the inflow is determined by a modified Debler-Craya method. The thickness of the flow field is calculated when water is withdrawn from a stratified zone.

$$D = 2.88\left(\frac{Q}{W}\sqrt{\frac{\rho}{g\beta}}\right)^{1/2} \tag{4.31}$$

where, D is the thickness of the flow field (m), Q is the withdrawal rate (m^3/s), W is the effective width of reservoir at the withdrawal level (m), β is the density gradient at the withdrawal location (kg/m^4), ρ is the water density at the outlet location (kg/m^3) and g is the acceleration due to gravity (m/s^2).

When water is withdrawn from a convective mixing zone, the maximum amount of flow, called as Craya's critical flow, that will remain contained in the convective zone without encroaching into the stratified zone, is calculated as:

$$Q = C W D^{3/2} \Delta\rho^{1/2} \tag{4.32}$$

where, Q is the Craya's critical flow (m^3/s), C is the empirical constant (0.074 for withdrawal from the surface element and 0.151 for subsurface element), W is the effective width of the reservoir at the withdrawal level (m), D is the thickness of the convective mixing zone (m), and $\Delta\rho$ is the maximum water density difference between convective mixing zone and the stratified zone (kg/m^3). If the actual rate of withdrawal from the reservoir is less than Craya's critical flow, the withdrawal is allocated throughout the mixing zone. If the withdrawal is more than Craya's critical flow then the excess withdrawal is allocated to the stratified zone.

In the WES withdrawal allocation method the average velocity through the orifice is calculated using the following equation:

$$V_0 = \frac{Z^2}{A_0}\sqrt{\left(\frac{\Delta\rho'}{\rho_0}\right)gZ} \tag{4.33}$$

where, V_0 is the average velocity through the orifice (m/s), Z is the vertical distance from the elevation of the orifice center line to the upper or lower limit of the zone of withdrawal (m), A_0 is the area of the orifice opening (m^2), $\Delta\rho'$ is the density difference of fluid between the elevations of the orifice center line to the upper or lower limit of the zone of withdrawal (kg/m^3), ρ_0 is the fluid density at elevations of the orifice center line (kg/m^3), and g is the acceleration due to gravity (m/s^2).

The release concentrations of the water quality constituents are computed by:

$$R_c = \frac{\sum_{p=1}^{N_p}\left(\Phi_{cp}Q_p\right)}{\sum_{p=1}^{N_p} Q_p}; \qquad c = 1, N_c \tag{4.34}$$

where, R_c is the release concentration for constituent c, Φ_{cp} is the concentration of constituent c at port p, Q_p is the flow rate through port p, N_p is the number of open ports, and N_c is the number of constituents.

Stream Hydraulic Module

A stream is represented by a linear network of segments. The characteristics of the linear elements are length, width, cross-section, and certain other parameters. The hydraulic computation in a stream hydraulic module is carried out by using either by St. Venant equations, stage-flow relationship, Muskingum hydrologic routing, or modified Puls hydrologic routing method.

The water quality module of WQRRS is capable of modeling biological and chemical constituents such as; fish, aquatic insects, benthic animals, zooplankton, phytoplankton, benthic algae, detritus, organic and inorganic sediments, inorganic suspended solids, dissolved phosphate and phosphorus, total inorganic carbon, dissolved ammonia, nitrites, nitrates all as nitrogen, dissolved biochemical oxygen, coliform bacteria, total alkalinity as calcium carbonate, total dissolved solids, pH, and unit toxicity. The model is capable of simulating the chemical and biological processes that take place under an aerobic environment and not the processes that take place in an anaerobic condition. The model assumes that coliform bacteria dynamic follows temperature dependent decay rates and basic process of transport is computed using the advection-diffusion equations. The dynamics of heat and any materials is modeled using the following equation:

$$V\frac{\partial C}{\partial t}=\Delta z.Q_z\frac{\partial C}{\partial z}+\Delta z.A_z D\frac{\partial^2 C}{\partial z^2}+Q_i C_i-Q_o C\pm VS \tag{4.35}$$

where, C is the thermal energy or the constituent concentration in the reservoir or stream, V is the volume of the fluid element, t is the time coordinate, z is the space coordinate, Q_z is the vertical advection, A is the element surface area normal to the direction of flow, D is the effective diffusion coefficient, Q_i is the later inflow, C_i is the inflow thermal energy or the constituent concentration, Q_o is the later outflow, and S is the source or sink.

Input Requirements

The data requirements of the reservoir module are:

- Reservoir geometry, dispersion characteristics, inflow and withdrawal location data, and the table of reservoir elevation versus surface area and width at the withdrawal location
- Coliform die-off rate and temperature coefficient
- Dry and wet bulb dew point temperatures, cloud cover, wind speed, and atmospheric pressure

The data requirements of the stream hydraulic module are:

- STORM generated hydrograph data
- Inflow and withdrawal data
- Stream geometry such as length, channel cross-section, and channel bottom elevations
- Boundary conditions such as flow, stage, and/ or stage versus flow relationship depending upon the method of hydraulic computation
- Non-point inflow and withdrawal like groundwater inflow and outflow, and agricultural returns

For simulating bacteria, additional data such as die-off rate coefficient and temperature correction coefficient are also required.

Limitations

Limitations of the model include maximum of 10 reaches, maximum of 100 volume elements, maximum of 105 nodes, and maximum of 10 inflows and 5 withdrawals.

WASP (Wool et al., 2002)

The Water Quality Analysis Simulation Program (WASP) was developed by USEPA for modeling contaminant fate and transport in surface waters. The current version of the model WASP6 contains 1) a user-friendly Graphical User Interface, 2) a pre-processor to assist modelers in the processing of data into a format that can be used in WASP, 3) high-speed WASP eutrophication and organic chemical model processors, and 4) a graphical postprocessor for the viewing of WASP results and comparison to observed field data.

Modeling Concepts

The WASP6 system consists of the hydrodynamics program DYNHYD5 and the water quality program, WASP6, with the capability of running them in conjunction or separately. The DYNHYD5 simulates the movement of water, while WASP6 simulates the movement and interaction of pollutants within the water. WASP6 also consists of two kinetic sub-models EUTRO to simulate conventional pollution (involving dissolved oxygen, biochemical oxygen demand, nutrients and eutrophication) and TOXI to simulate toxic pollution (involving organic chemicals, metals, and sediment). Though modeling of coliform bacteria is not directly mentioned in WASP manual, bacteria can be modeled as another chemical with an appropriate exponential biodegradation rate (USEPA, 2001c).

The basic principle of the sub-models is the conservation of mass. The hydrodynamics program also conserves momentum, or energy, throughout time and space. A mass balance equation for dissolved constituents in a body of water must account for all the material entering and leaving through direct and diffuse loading; advective and dispersive transport; and physical, chemical, and biological transformation. The mass balance equation around an infinitesimally small fluid volume is:

$$\begin{aligned}\frac{\partial C}{\partial t} = &-\frac{\partial}{\partial x}(U_x C) - \frac{\partial}{\partial y}(U_y C) - \frac{\partial}{\partial z}(U_z C) + \frac{\partial}{\partial x}(E_x \frac{\partial C}{\partial x}) + \frac{\partial}{\partial y}(E_y \frac{\partial C}{\partial y}) + \\ &\frac{\partial}{\partial z}(E_z \frac{\partial C}{\partial z}) + S_L + S_B + S_K\end{aligned} \qquad (4.36)$$

where, C is the concentration of the water quality constituent, t is time, Ux, Uy, and Uz are longitudinal, lateral, and vertical advective velocities, Ex, Ey, and Ez are longitudinal, lateral, and vertical diffusion coefficients, respectively, SL is the direct and diffuse loading rate, SB is the boundary loading rate (including upstream, downstream, benthic, and atmospheric), SK is the total kinetic transformation rate; positive is source, negative is sink.

Input Requirements

- Segment geometry
- Advective and dispersive coefficients
- Boundary concentrations
- Point and diffuse source waste loads
- Kinetic parameters, constants, and time functions
- Initial concentrations

Model Outputs

The model outputs include the timeseries of depth and pollutant concentrations for different segments. A powerful post-processor provides the option of producing x-y plots showing the different outputs against time, or spatial plots showing the two-dimensional display over different segments. There is also an option of exporting the spatial grid plots (2D plots) into different formats including ArcView shape file format.

EFDC (Hamrick, 1992)

The Environmental Fluid Dynamics Code (EFDC), a general-purpose modeling package was developed at the Virginia Institute of Marine Science. The model is designed to simulate the three-dimensional flow, transport, and biochemical processes in surface water systems such as rivers, lakes, reservoirs, wetlands, estuaries, and the coastal oceans. The EFDC model has been tested for more than 60 modeling studies (Ji et al., 2001). The model was used for simulations of pollutants and pathogen fate and transport from various types of sources (Hamrick, 1996). Some examples of other applications of the model include discharge dilution, shoreline modification, and shellfish larvae transport studies in the James and York Rivers, Virginia; large-scale wetland simulation in the Everglades; sediment transport simulations at Vero Beach, Florida and Morro Bay, California; and simulation of circulation and temperature in Conowingo Pond on the Susquehanna River in Pennsylvania and Maryland (Hamrick and Mills, 2000).

Modeling Concepts

The four main modules of the model are: (1) a hydrodynamic model, (2) a water quality model, (3) a sediment transport model, and (4) a toxics model. The physical process simulation capabilities of the EFDC model are equivalent to the Blumberg-Mellor model (Blumberg and Mellor, 1987) and U.S. Army Corps of Engineer's Chesapeake Bay Model (Johnson, et al., 1993). The EFDC model solves the three-dimensional, vertically hydrostatic, free surface, turbulent averaged equations of motions for a variable density fluid using a curvilinear and orthogonal coordinate system in the horizontal and stretched in the vertical direction. The momentum and continuity equations after applying Boussinesq approximation becomes:

$$\begin{aligned}&\partial_t(m_x m_y Hu)+\partial_x(m_y Huu)+\partial_y(m_x Hvu)+\partial_z(m_x m_y wu)\\&-(m_x m_y f+v\partial_x m_y-u\partial_y m_x)Hv\\&=-m_y H\partial_x(g\xi+p)-m_y(\partial_x h-z\partial_x H)\partial_z p+\partial_z(mH^{-1}A_v\partial_z u)+Q_u\end{aligned} \tag{4.37}$$

$$\begin{aligned}&\partial_t(m_x m_y Hv)+\partial_x(m_y Huv)+\partial_y(m_x Hvv)+\partial_z(m_x m_y wv)\\&-(m_x m_y f+v\partial_x m_y-u\partial_y m_x)Hu\\&=-m_x H\partial_y(g\xi+p)-m_x(\partial_y h-z\partial_y H)\partial_z p+\partial_z(mH^{-1}A_v\partial_z v)+Q_v\end{aligned} \tag{4.38}$$

$$\partial_z p=-gH(\rho-\rho_0)\rho_0^{-1}=-gHb \tag{4.39}$$

$$\partial_t(m_x m_y H)+\partial x(m_y Hu)+\partial y(m_x Hv)+\partial z\left(m_x m_y w\right)=Q_H \tag{4.40}$$

where, u and v are the horizontal velocities in the curvilinear-orthogonal horizontal coordinates (x,y), m_x and m_y are the square roots of the diagonal components of the metric tensor, w is the vertical velocity in the vertical coordinate z, p is the kinematic excess pressure above the reference density, ρ_0, hydrostatic pressure, ζ is the physical vertical coordinate of the free surface, -h is the physical vertical coordinate of the bottom topography, H is the total depth (h+ ζ), f is the Coriolis parameter, A_v is the vertical turbulent viscosity, Q_u and Q_v are the momentum source-sink terms, and ρ is the density. The source term Q_H represents the direct rainfall, evaporation, groundwater interaction, water withdrawals, and point and nonpoint source discharges. The buoyancy, b, is the normalized deviation of density from the reference value. The generic transport equation in EFDC for a dissolved or suspended constituent C is (Hamrick and Mills, 2000):

$$\begin{aligned}&\partial_t(m_x m_y HC) + \partial_x(m_y HuC) + \partial_y(m_x HvC) + \partial_z(m_x m_y wC)\\ &\quad = m_x m_y \partial_z(H^{-1}K_b \partial_z C + \sigma C) + m_x m_y HR_c + Q_c\end{aligned} \tag{4.41}$$

where, K_b is the vertical turbulent diffusivity, σ is a positive settling velocity, R_c is the reactive source/sinks, and Q_c is the horizontal turbulent diffusion and external sources/sinks associated with volumetric withdrawals and discharges. The model solves the equations of motion using a finite volume-finite difference spatial discretization with an MAC or C grid staggering of the discrete variables. Further details of the equations used and the solution methods are given in Hamrick (1992).

Input Requirements

The input data requirements for the EFDC model include:

- Climatic data such as air temperature, pressure, relative humidity, direct rainfall, wind speed and direction
- Stream flow data
- Grid information such as depth, bottom elevation and bottom roughness

- Time-varying concentrations of water quality constituents

Model Outputs

A wide variety of output options are available. Some of options available are:

- Horizontal plane and vertical plane transect plotting of vector and scalar field at specified time
- Time series of model variables at selected locations and time intervals
- Grab sample simulation at specified times and locations

CE-QUAL-RIV1 (Cole and Buchak, 1995)

CE-QUAL-RIV1 is used to simulate flow and water quality in rivers and run-of-the-river reservoirs. The original version of CE-QUAL-RIV1 was developed by Ohio State University at the request of US EPA, to predict the water quality associated with storm water runoff. Later, the US Army Engineer Waterways Experiment Station (WES) modified the model to include the unsteady flow feature to simulate stream/waterway projects regulated by Army Corps of Engineers.

Modeling Concepts

The two modules of CE-QUAL-RIV1 are RIV1H and RIV1Q, where RIV1H simulates the hydraulic processes, while RIV1Q simulates the water quality. The output from RIV1H is used to drive RIV1Q. The model is used to simulate the one-dimensional riverine systems assuming that there is no vertical stratification for temperature, density, and chemical concentration. The model can be used to simulate both point and nonpoint source discharges but cannot be used for “near field” simulations when point sources are considered because of the fact that mixing may not have occurred sufficiently to satisfy the one-dimensional assumption. The model can be used to address both steady and unsteady flow conditions.

The transport of fluid and the pollutants are governed by the control volume method (Liggett, 1975) together with the laws of conservation of fluid mass, momentum and pollutants. The equations are solved using four-point implicit finite difference method.

Input Requirements

- Climatic data such as cloud cover, wind speed, dry and wet bulb temperatures and atmospheric pressure
- Channel network and geometry
- Timeseries of point inflows (point and lateral) and withdrawals
- Lateral inflows of quality constituents at each segment

Model Outputs

- Timeseries of flow
- Timeseries of in-stream pollutant concentration

CE-QUAL-W2 (Environmental Laboratory, 1995)

CE-QUAL-W2 is a two-dimensional hydrodynamic and water quality model that can be used for modeling rivers, lakes, reservoirs, and estuaries. The original version of CE-QUAL-W2 was developed in Edinger and Buchak and was known as Laterally Averaged Reservoir Model (LARM). The first LARM application was modified to allow for multiple branches from the initial limitation of modeling a reservoir with no branches and also to include estuarine boundary conditions which resulted in the

Generalized Longitudinal-Vertical Hydrodynamics and Transport Model (GLVHT). The Water Quality Modeling Group at the US Army Engineer Waterways Experiment Station (WES) added the water quality algorithms.

Modeling Concepts

CE-QUAL-W2 is a two-dimensional, longitudinal/vertical, hydrodynamic and water quality model. The model is best suited for relatively long and narrow waterbodies exhibiting longitudinal and vertical water because of the lateral homogeneity assumption. The CE-QUAL-W2 model can be applied to rivers, lakes, reservoirs, and estuaries. Version 2.0 is a result of major modifications to the code to improve the mathematical description of the prototype and increase computational accuracy and efficiency. The new capabilities in Version 2.0 are:

- an algorithm that calculates the maximum allowable timestep and adjusts the timestep to ensure hydrodynamic stability requirements are not violated (autostepping)
- a selective withdrawal algorithm that calculates a withdrawal zone based on outflow, outlet geometry, and upstream density gradients
- a higher-order transport scheme (QUICKEST) that reduces numerical diffusion (Leonard, 1979)
- time-weighted vertical advection and fully implicit vertical diffusion
- step function or linear interpolation of inputs
- improved ice-cover algorithm
- internal calculation of equilibrium temperatures and coefficients of surface heat exchange or a term-by-term accounting of surface heat exchange
- variable layer heights and segment lengths
- surface layer extending through multiple layers
- generalized time-varying data input subroutine with input data accepted at any frequency

- volume and mass balances to machine accuracy
- sediment/water heat exchange

Input Requirements

- Channel cross-sections and river segment orientations
- Upstream and downstream flow
- Climatic data such as cloud cover, wind speed/direction, precipitation, air temp
- Water quality data such as upstream and tributary concentrations, and point and nonpoint source concentrations

Model Outputs

- Water surface elevations, velocities, and temperatures
- Timeseries plots of temperature and constituents for inflows, outflows, and withdrawals
- Contour plots of different constituents at user specified timings

AUTO_QI (Illinois Water Survey, 1990)

AUTO_QI is a combination of Arc-Info geographical information system (GIS) interface and a deterministic water quality model Q-ILLUDAS modified to meet the regional level modeling requirements. Q-ILLUDAS is the version of Illinois Urban Drainage Area Simulator (ILLUDAS) capable of modeling urban water quality. AUTO_QI model is capable of simulating suspended sediments, total P, total N, dissolved oxygen demand, trace metal and bacteria.

Modeling Concepts

AUTO_QI is a combination of three programs HYDRO, LOAD, and BMP running in succession along with additional inputs from users. HYDRO is used to perform the continuous simulation of soil moisture based on a daily and hourly rainfall, and to calculate the storm event runoff volumes. LOAD calculates pollutant loadings for each runoff event based on the runoff volume calculated by HYDRO and user supplied pollutant accumulation and washoff information. BMP simulates the effect of user specified best management practices. The model calculates direct runoff from areas directly connected to paved areas. For the supplemental paved areas, the model calculates the initial losses such as initial wetting and depression storage and then the rest of rainfall is assumed to be runoff, while grassed areas will be subjected to infiltration in addition to the initial losses. The infiltration rate is calculated using the Horton equation.

LOAD calculates the accumulation of pollutant during the dry periods using the linear accumulation equation:

$$P_t = P_{t-1}(1 - r) + A \tag{4.42}$$

where, P_t and P_{t-1} are the pollutant load at time, t and $t-1$, respectively, r is the removal rate and A is the accumulation rate of the pollutant. The pollutant washoff following a rainfall event is calculated as:

$$P_0 - P = P_0(1 - e^{-kt}) \tag{4.43}$$

where, P_0 and P are the pollutant amount on the surface at the beginning and end of rainfall, respectively, k is the proportionality constant, and t is the duration of the rainfall. The value of k is calculated as the product of rainfall intensity (in/hr) and a constant B. The default values of B for paved areas and grassed areas are 4.6 and 1.4, respectively.

Input Requirements

- Daily and hourly rainfall data
- Daily accumulation and washoff rate plus the removal rate by any Best Management Practice
- Monthly evaporation and evapotranspiration rates
- Soil infiltration rates
- Land use distribution

Model Outputs

The outputs from the model are the runoff volumes and the pollutant washoff amount for each storm events.

MIKE-BASIN (DHI, 2003b)

MIKE-BASIN is a decision support system developed by Danish Hydraulic Institute (DHI) Water & Environment on an ArcView GIS framework. The model can be used to simulate hydrology and water quality on a watershed level.

Modeling Concepts

In MIKE-BASIN rivers and their main tributaries are represented by branches of a network and the confluences are represented by the nodes. The model simulates the hydrology and water quality based on user specified flexible time steps in terms of days or months and finds stationary solutions for each time step. The water quality simulations are carried out by the WQ module in MIKE-BASIN. The WQ module is capable of simulating transport and degradation of constituents in rivers and reservoirs. The constituents are: ammonia/ammonium, nitrate, dissolved oxygen, total phosphorous, biological and chemical oxygen demands, *E.Coli* and additional solute of interest. It also takes into consideration re-aeration from weirs and phosphate sedimentation in reservoirs. The transport process is modeled using advection without accounting for the dispersion effect. This makes the WS module unsuitable for turbulent and tidal rivers. The degradation of *E.Coli* is assumed to follow a first-order decay rate and the rate constant is adjusted for temperature. The correction for the rate constant is assumed to follow the Arrhenius principle, which states that a 10˚C increase in temperature causes the decay rate to double. The WQ module is also capable of modeling groundwater quality. Since MIKE-BASIN assumes a quasi-steady-state, the net change in storage between the nodes is zero except for reservoirs or groundwater aquifers. MIKE-BASIN is capable of modeling both point and nonpoint source of pollution. The nonpoint source loadings can be input as time series of concentrations or time series of mass fluxes. Also the effective load from the catchments can be calculated by the nonpoint module using landuse themes and numbers of swine, cows, sheep, and chicken within the catchment. Since MIKE-BASIN is a network model, the input of pollutants takes place only at the nodes. Thus the runoff and nonpoint pollution from a catchment is assumed to enter a river only at the catchment node, thus limiting the decay process to take place in the downstream branches and not within the catchment. This makes the comparison of water quality simulation results at the catchment nodes inappropriate. The runoff from the catchments can be supplied as timeseries or can be generated using the MIKE-BASIN RR (rainfall-runoff) module. MIKE-BASIN was applied to a study in Malaysia (Jørgensen, 2002).

Input Requirements

- GIS layers to create watershed boundaries and stream network
- Rainfall data
- Rate coefficients for the pollutants

Model Outputs

- Maps showing flows and concentrations
- Time series of flow and concentrations
- Monthly tables and descriptive statistics any simulation time step in HTML format
- Low-flow statistics computed for every node, for eg. 7Q10, 30Q20, 30Q50 and flow duration tables

PLOAD (USEPA, 2001b)

The PLOAD is the simplest of all the modeling tools available with Environmental Protection Agency's Better Assessment Science for Integrating point and Nonpoint Sources (BASINS). The PLOAD model, developed by CH2M-HILL is a simplified GIS-based model to calculate the annual average pollutant loads.

Modeling Concepts

The PLOAD model calculates the pollutant loads using either the export coefficient or the EPA's Simple Method approach. The export coefficient method calculates the loads for each pollutant type using the equation:

$$L_p = \sum_U (L_{PU} * A_U) \tag{4.44}$$

where, L_p is the total pollutant loads for a given watershed (lbs), L_{PU} is the pollutant loading rate for land use type U (lbs/ac/yr), and A_U is the area of land use type U (ac).

The use of Simple Method is limited to watersheds of less than one square mile in size. When Simple Method is used, first the runoff coefficient for each land use is calculated as:

$$R_{VU} = 0.05 + (0.009 \times I_U) \tag{4.45}$$

where, R_{VU} is the runoff coefficient for land use type U (fraction), and I_U is the percent imperviousness. The pollutant loads are then calculated as:

$$L_P = \sum_U (P \times P_J \times R_{VU} \times C_U \times A_U \times 2.72/12) \qquad (4.46)$$

where P is the precipitation (in/yr), P_J is the ratio of storms producing runoff, A_U is the area of land use type U (ac), and C_U is the Event Mean Concentration (EMC) for land use type U (mm/l)

Input Requirements

- GIS data such as the watershed boundary and the land use maps
- Optional point BMP sites or aerial BMPs represented by the polygons
- Tabular data containing information on pollutant loading rates such as export coefficients and event mean concentrations, impervious factor, and BMP efficiency

Model Outputs

- Total pollutant loads by watershed (map and table)
- Pollutant loads per acre by watershed (map and table)
- Event mean concentration by watershed (map and table)

SLAMM (Pitt and Voorhees, 2000)

The Source Loading and Management Model (SLAMM) was developed in the mid 1970s as a data reduction tool during the EPA's Storm and Combined Sewer Pollution Control Program (Pitt and Voorhees, 2000). Further development was carried out during the EPA's Nationwide Urban Runoff Program (NURP) projects. SLAMM is mostly a planning tool strongly based on actual field observations and mostly focused on small storm hydrology and particulate washoff. It is designed to provide information on the sources of problem pollutants in stormwater and the effectiveness of stormwater management practices in controlling the pollutant at their sources and at outfalls. It calculates the runoff flow volumes and the mass balances for both particulate and dissolved pollutants.

Modeling Concepts

SLAMM is based on an empirical relationship and was developed for the better understanding of relationships between sources of urban runoff pollutants and runoff quality. Contrary to other urban models SLAMM is designed to model small storm hydrology and particulate washoff. It has the capability to incorporate many control practices together such as detention ponds, infiltration devices, porous pavements, grass swales, and street cleaning.

The SLAMM model calculates runoff volumes and suspended solids for each source area and for each rain event, considering the effects of each source area control and the runoff pattern between the areas. Suspended solids washoff and runoff volume from each individual area for each rain event are then summed for the entire drainage system and the effects of the drainage system controls are calculated. Finally, the effects of the outfall controls are evaluated and the pollutant loadings into the receiving waterbody are calculated.

Input Requirements

- Land use distribution (residential, commercial, etc.)
- Runoff coefficients
- Drainage system information
- Control device information

Model Outputs

- Distributions of runoff and pollutants
- Source area contribution of runoff for different storm events
- Source area contribution of pollutants for different storm events

CORMIX (Jirka, 1992; Fischer et al., 1979; Jirka et al., 1996)

The Cornell Mixing Zone Expert System (CORMIX) was developed during the period 1985-1995 under cooperative funding agreements between US Environmental Protection Agency and Cornell University. Though the model was originally developed for steady ambient conditions, the new versions allow application to highly unsteady environments, such as tidal reversal conditions. The model is used to predict the qualitative features and quantitative aspects of the hydrodynamic mixing processes due to different discharges in small streams, large rivers, lakes, reservoirs, estuaries and coastal waters. The CORMIX model is used by the Washington State Department of Health, Shellfish Program to estimate shellfish closure areas around wastewater treatment outfalls. Up until recently, the model was distributed US EPA Center for Environmental Assessment Modeling (CEAM) in Athens, Georgia. Currently CORMIX software and technical support is provided by MixZon Inc. (http://www.mixzon.com/mixzon.html).

Modeling Concepts

The CORMIX model consists of three submodels CORMIX1, CORMIX2, and CORMIX3 that are based on integral solutions of the Eulerian momentum and transport equations. The submodel CORMIX1 analyzes the steady state buoyant submerged discharges from a single port assuming receiving water having a geometry of a rectangular cross sectional shape, flow to be steady and uniform normal to the cross-sectional plane and the ambient flow to be a piecewise linear vertical density structure. CORMIX2 simulates the submerged discharges from multiple ports with the flow environment consistent to CORMIX1. CORMIX3 is used to simulate the surface discharge from a canal or pipe into an ambient flow environment represented by a constant shoreline depth and bottom sloping down away from the shoreline. A number of post-processors are included such as, CORJET for the analysis of the near-field behavior of buoyant jets, FFLOCATR for the far-field delineation of discharge plumes in non-uniform river or estuary environments and CMXGRAPH for plume plotting. The main advantage of CORMIX is the ease of use of its submodels, while the main disadvantage results from assumption of spatial and temporal uniformity of ambient conditions and idealized receiving water geometry. The model can not be applied to non-uniform ambient flow conditions prone to locally recirculating flows, and cases with complicated discharge geometries.

Input Requirements

- Water body depths, flow rate, optional water body width, tidal information, wind speed, roughness coefficient, and density/temperature at the surface and bottom
- Discharge data at the location and distance of the nearest bank, vertical and horizontal angles, port diameter, port height and port area, discharge location and configuration, and discharge cross-sectional data

- Effluent data such as flow rate, velocity, temperature, concentration, heat loss coefficient, and decay rate coefficient
- Mixing zone data such as value of water quality standard, toxicity of pollutant, distance, width, or area of mixing zone, and CMC and CCC for toxic pollutants

PLUMES (Baumgartner et al., 1994, USEPA, 1996b)

PLUMES is a hydrodynamic mixing zone modeling system similar to CORMIX. The PLUMES model has been used by the Rhode Island Department of Environmental Management to predict water quality at the wastewater treatment plant (WWTP) outfalls in order to protect the shellfish growing areas in Narragansett Bay from the impacts of a potential 6-hour failure in the, chlorination process at WWTPs.

Modeling Concepts

The main components of PLUMES are the PLUMES interface, the RSB model, the UM model, a far-field mixing model, and a discharge classification model. The RSB model is a semi-empirical model based on the principles of dynamic similitude and dimensional analysis, while the UM model is based on a Lagrangian formulation. RSB is based on the experimental studies of multiport diffusers in stratified currents and the normalized expression of dilution is given by:

$$\frac{S_m q N}{b^{2/3}} = f\left(\frac{l_m}{l_b}, \frac{s}{l_b}, F, \Theta\right) \tag{4.47}$$

where,

$$l_m = \frac{uq}{b^{2/3}} \tag{4.48}$$

$$l_b = \frac{b^{1/3}}{N} \tag{4.49}$$

u is the jet exit velocity, q is discharge per unit length of diffuser, b is the buoyancy per unit diffuser length, N is the buoyancy frequency (the Brunt-Vaisala frequency), S_m is the minimum initial dilution observed in the vertical plane through the wastefield at the end of the mixing region and Θ is the angle between the current and the diffuser. The model assumes that the receiving water is linearly density-stratified, and flows at a steady velocity.

The main features of the UM model are the Lagrangian formulation and the projected area entrainment hypothesis. The basis of the Lagrangian formulation is the assumptions of mass, horizontal momentum and the energy conservation over time. The projected area entrainment is given by:

$$\frac{dm}{dt} = \rho_a A_p u \tag{4.50}$$

where, dm is the incremental mass entrained in the time increment of dt, A_p is the projected area, ρ_a is the local ambient density, and u is the ambient current speed normal to the projected area.

The farfield algorithm is simpler than the initial dilution model and is given by:

$$S = \frac{S_a}{erf\left[\sqrt{\frac{b^2}{16\alpha b^{4/3} t}}\right]} \tag{4.51}$$

where erf is the error function, S is the centerline dilution in the farfield plume, S_a is the initial dilution, α is a dispersion coefficient, b is the width of the plume field at the end of initial dilution, and t is the time of travel from the point of the end of initial dilution to the point of interest.

Input Requirements

- Ambient data such as water body depth, far-field distance, dispersion coefficient and increment, current speed, density, salinity, temperature, concentration, and average current speed in the far-field
- Outfall structure details such as total diffuser flow, number of ports in the diffuser, spacing between ports, port depth, diameter and elevation, vertical angle, contraction coefficient cell and horizontal diffuser angle
- Effluent characteristics such as density, salinity, temperature, pollutant concentration and first-order decay coefficient

Model Outputs

- CORMIX flow classification
- Pollutant concentration and dilution ratios at various points in the mixing zone

SWAT (Arnold et al., 1998; Neitsch et al., 2001; Sadeghi and Arnold, 2002)

Soil and Water Assessment Tool (SWAT) is developed by modifying Simulator for Water Resources in Rural Basins (SWRRB) and contains features of many models such as Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS), Groundwater Loading Effects on Agricultural Management Systems (GLEAMS), and Erosion-Productivity Impact Calculator (EPIC). A continuous simulation model SWAT has the capability to simulate movement of sediment, nutrients, and pesticides. Recently the SWAT model was modified to add the capability of modeling bacteria, both fecal coliform and *E.Coli* (Sadeghi and Arnold, 2002). Currently the bacterial module is in the testing phase.

SWAT is designed to model two species/strains of pathogens with distinctly different die-off and/or re-growth rates with the intention of including both persistent and less persistent bacteria. The model calculates the fate and transport of bacteria based on manure application, incorporation through tillage, transport through surface runoff and infiltration, removal of bacteria by the use of vegetative filter strips, loading by the point sources, and in-stream transport. Bacteria are assumed to be transported either by the direct washoff or as adsorbed to the minute soil particles.

Bacterial input though manure application is calculated as:

$$bsol_i = bc_{i-1} + conc \times wman \times (1-p) \tag{4.52}$$

$$bsor_i = bc_{i-1} + conc \times wman \times p \tag{4.53}$$

where, *bsol* is the bacterial count in the solution, *bsor* is the adsorbed bacterial count, *conc* is the bacteria in manure, *wman* is the weight of manure applied, *p* is the partition coefficient, and *i* is the day index.

It is assumed that the bacteria that reach the second layer through tillage is no longer available for transport. The amount of bacterial that reach the second layer of soil is calculated as:

$$bc_2 = bc_1 \times (1-\eta_m) \tag{4.54}$$

where η_m is the mixing efficiency of the tillage operation.

The first order decay equation specified by Chick's Law is used to calculate the bacteria quantity following a die-off or re-growth event. Thus the residual bacterial count is calculated as:

$$bc_i = bc_{i-1} \times e^{-(k_d - k_g)} \tag{4.55}$$

where, bc_i and bc_{i-1} are bacterial counts on day i and day $i-1$, respectively, and k_d and k_g are the die-off and re-growth rate constants, respectively.

The bacterial transport from the soil surface due to runoff is determined as:

$$br = bsol \times q \times p / (\rho * D) \tag{4.56}$$

where q is the surface runoff, ρ is the bulk density, and D is the depth of surface soil layer.

The bacterial transport through the sediment due to runoff is determined as:

$$bs = bsor \times Y \times ER \tag{4.57}$$

where, Y is the sediment yield and ER is an enrichment ratio.

Input Requirements

Input requirements specific to SWAT bacteria sub-model are not available at this time. But in general the minimum inputs required by SWAT are:

- Climatic data such as daily precipitation, maximum/minimum air temperature, solar radiation, wind speed and relative humidity (SWAT has a Weather Generator, which can be used to simulate any missing climatic data)
- Geographic Information System (GIS) layers such Digital Elevation Model, land use map and soil map
- Point source loading rates

Model Outputs

In general SWAT outputs include:

- Flow timeseries for daily, monthly or yearly time periods

- In-stream pollutant concentrations for the same time period

General characteristics of the models

Based on the detailed review of the models the general characteristics of the models are summarized and are given in Table 4.2. The selection of a model for in-stream bacteria concentration predictions are dependent on the ability to address both urban and agricultural land uses, capability to handle point and nonpoint sources of bacterial pollution and the ability to simulate un-steady state flow conditions. Other characteristics compared include:

- Types of waterbody – Streams/Reservoir/Estuaries
- Simulation time step – Hourly/Daily/Monthly
- Simulation mode – Event based/Continuous
- Model type – Watershed based/Receiving waterbody
- Availability – Public domain/Commercial

Table 4.2 Comparison of the models.

Model	Type of Waterbody[1]	Dimension[2]	Simulation Time Step[3]	Simulation Type[4]	Simulation Mode[5]	Land use Type [6]
QUAL2E-UNCAS	S	1D	-	S	-	B
HSPF	S	1D	S/H	D	B	B
MIKE-SWMM	S	1D	S/H	D	B	U
QQS	S	1D	S	D	B	U
STORM	S	1D	H	D	C	U
SWMM	S	1D	S/H	D	B	U
WQRRS	S/R/L	1D	H	D	B	B
WASP	S/L/R/E	1D/2D/3D	S/H/D or more	D	-	B
EFDC	S/L/R/E	3D	-	D	-	B
CE-QUAL-RIV	S	1D	H	S/D	-	
CE-QUAL-W2	R/L/S	2D	S/H/D	D	-	B
AUTO_QI	-	0D	D	D	C	U
PLUMES	S/L/E	3D	-	S/D	-	-
MIKE-BASIN	S/R	1D	D/M	S	-	B
PLOAD	-	0D	A	-	-	B
SLAMM	-	0D				U
CORMIX	S/L/R/E	2D/3D	-	S	-	-
SWAT	S	1D	D	D	C	N

Table 4.2 Continued.

Model	Watershed/ Receiving Waterbody[7]	Source of Pollution[8]	Pollutant Predictive Method[9]	GIS Interface Available	Public Domain	Commercial	Source Code Available
QUAL2E-UNCAS	R	P	E	✓	✓		✓
HSPF	W/R	B	E/W/S	✓	✓		✓
MIKE-SWMM	W/R	B	E/W/S	✓		✓	
QQS	W/R	N	E/W				
STORM	W	N	W/S		✓		✓
SWMM	W	N	E/W/S		✓		✓
WQRRS	R	P	E		✓		
WASP	R	B	-		✓		✓
EFDC	R	B	-		✓		✓
CE-QUAL-RIV	R	B	-		✓		✓
CE-QUAL-W2	R	B	-		✓		✓
AUTO_QI	W	N	W	✓	✓		
PLUMES	R	P	-		✓		
MIKE-BASIN	W/R	B	E	✓		✓	X
PLOAD	W	N		✓	✓		✓
SLAMM	W	N	E			✓	
CORMIX	R	P	-			✓	X
SWAT	W/R	B	E/W/S	✓	✓		✓

[1] Stream, open channels (S), lake (L), reservoir (R), or estuary (E) ; [2] 0D, 1D, 2D, or 3D; [3] Subhourly (S), hourly (H), daily (D), monthly (M) or annual (A); [4] steady (S) or dynamic (D); [5] Storm event based (E), continuous (C), or both (B); [6] Urban (U), nonurban (N), or both (B); [7] Watershed level (W) or receiving waterbody (R); [8] Point source (P), nonpoint source (N), or both (B); [9] Empirical (E), Build-up/Washoff (W), or Soil Loss (S)

Model selection for different clusters

Public Sewer Dominated Cluster

The first cluster consists of 6 stream segments with a relatively high density of households on public sewers. The watersheds are highly urbanized compared to other groups. The density of households on OSSF and bacterial loading from livestock are the lowest compared to other groups of segments. There is a moderate rate of bacterial loading from wildlife. Table 4.3 shows some important water quality characteristics of the stream segments within this group. Though only one stream segment showed significantly high mean during the dry period compared to the wet period, the mean bacterial concentration during dry period is relatively high for all the stream segments. This can be an indication of the presence of constant point sources within the watershed. Also the fact that two segments showed significantly higher concentration during the wet period compared to the dry period is an indication of the presence of some nonpoint source within those watersheds. Although the wastewater treatment plants (WWTPs) are not permitted to discharge any bacteria into the streams, their presence within the watershed is worth noticing because of the chance of occasional failure of these facilities. Except for two stream segments, all the stream segments have reasonable number of WWTPs within there watersheds. Considering these facts, the computer model should be able to address both point and nonpoint source pollution from an urban watershed. Of the different models reviewed, QUAL2E, HSPF, and MIKE-SWMM can be used to model in-stream bacterial concentration in these stream segments. Since MIKE-SWMM is a commercial software and is not available in public domain, the selection of models is limited to QUAL2E and HSPF.

Table 4.3 Water quality characteristics of stream segments belonging to public sewer dominated cluster.

Segment ID	Total No. of observations	Observations with >400 (cfu/100ml)	Period of larger means	Wet period mean (cfu/100ml)	Dry period mean (cfu/100ml)	WWTP capacity (mgd)	Watershed area (ac)
1911	2601	1301	Rain	4694	2191	3362.1	315273
1910	816	261	Rain	7242	601	0.74	112734
1906	393	100	Base	1090	4495	14.32	105374
1218	332	96	ND	408	443	6.34	72758
1209C	141	60	ND	309	594	-	37438
0505B	30	18	ND	3044	3522	-	260012

$^{+}$ND = means of two periods are not significantly different, NED = Not enough data to compare

OSSF Dominated Cluster

The second cluster consists of 7 stream segments with very high density of households on OSSF. The watersheds have a moderately high percentage of forestland and pastureland compared to other groups of watersheds. The bacterial loading from livestock is relatively high. The density of households on public sewer and bacterial loading from wildlife are low compared to other groups of watersheds. It was found that means of bacterial concentrations during the wet period were significantly higher than that of the dry period for all watersheds (Table 4.4). At the same time, the mean bacterial concentration during the dry period for all stream segments is also high. Since the number of WWTPs within the watersheds is negligible, the higher concentration during the dry period may be an indication of high background concentration or the presence of wildlife or cattle within accessible distance of the waterbody. Also the watersheds are mostly of non-urban land use type. Hence, the model should be able to address non-urban land uses, and both point and nonpoint sources of pollution. The appropriate models would be HSPF and MIKE-BASIN. Also due to the lack of available monitoring data the modeling of stream segments such as 0604B, 0604C and 0508B may be difficult. Either more data should be collected or the modeling studies should be limited to the use of simple model PLOAD.

Table 4.4 Water quality characteristics of stream segments belonging to OSSF dominated cluster.

Segment ID	Total No. of observations	Observations with >400 (cfu/100ml)	Period of larger means	Wet period mean (cfu/100ml)	Dry period mean (cfu/100ml)	WWTP capacity (mgd)	Watershed area (ac)
0508	352	182	Rain	6492	5172	2.93	17019
0508B	44	23	Rain	10010	3958	0	7792
0511B	416	95	Rain	1061	879	0	22473
0604A	131	45	Rain	1382	735	0	75884
0604B	38	19	Rain	1462	1068	0	15317
0604C	47	15	Rain	970	438	0	19089
0611B	80	28	Rain	1425	648	0	52821

Forested with Livestock Dominated Cluster

The third group of stream segments consists of 15 watersheds that are predominantly forested. The bacterial loading from livestock is relatively high. The densities of households on public sewer and on OSSF are low compared to other groups. Also the bacterial loading from wildlife is relatively low. It was found that the mean of bacterial concentrations during dry periods was significantly higher than that of the wet periods for the watershed corresponding to stream segment 0607. There was no significant difference between two means for watersheds corresponding to stream segments 0607C, 0608D and 0610A. The rest of the watersheds showed higher means during the wet period compared to the dry period. Also the presence of WWTPs within the watershed boundaries was negligible for many of the stream segments except for four streams segments (Table 4.5). A few stream segments showed relatively high mean bacterial concentration during dry period. For detailed modeling

HSPF or MIKE-BASIN can be used. For preliminary modeling for stream segments 0603A, 0607C, and 0502A, PLOAD can be used.

Table 4.5 Water quality characteristics of stream segments belonging to livestock dominated cluster with high forest land cover.

Segment ID	Total No. of observations	Observations with >400 (cfu/100ml)	Period of larger means	Wet period mean (cfu/100ml)	Dry period mean (cfu/100ml)	WWTP capacity (mgd)	Watershed area (ac)
0508A	368	192	Rain	6683	5032	0	47339
0511	595	148	Rain	1421	914	23.83	67670
0607	319	62	Base	2160	3752	1.61	364631
0513	173	43	Rain	1019	263	0.93	174096
0511A	165	43	Rain	4217	924	0	73789
0608D	107	26	ND	646	467	0	95996
0610A	91	25	ND	471	304	0	89394
0511C	50	24	Rain	6537	1140	0	42852
0608B	83	20	NED	2549	531	0	140397
0604	189	14	Rain	4743	1642	4.53	1084810
0608F	53	14	NED	1344	377	0	69709
0608C	55	13	Rain	1161	390	0	127517
0603A	33	9	Rain	1266	294	0	38617
0607C	34	5	ND	307	263	0	60223
0502A	11	4	NED	-	904	0	74347

Cluster with Low Bacterial Loading from All Sources

The fourth group of watersheds consists of 14 watersheds that have high percentage of watershed area under pastureland. The bacterial loadings from livestock and wildlife are low compared to other group of watersheds. The densities of households on public sewer and on OSSF are also low. Out of all the groups of watersheds this group has low contribution of bacteria from all the different sources. Four watersheds showed high means during dry periods, seven watersheds during wet periods and for two watersheds there was no significant difference between the means of two periods (Table 4.6). Few of the stream segments have reasonable number of

WWTPs within their watersheds along with high mean dry period concentration. The HSPF model can be used for modeling the watersheds.

Table 4.6 Water quality characteristics of stream segments belonging to cluster with low loading from all sources.

Segment ID	Total No. of observations	Observations with >400 (cfu/100ml)	Period of larger means	Wet period mean (cfu/100ml)	Dry period mean (cfu/100ml)	WWTP capacity (mgd)	Watershed area (ac)
1242	419	183	Rain	1888	568	1440.3	1087540
1901	543	165	Base	1548	4362	3.88	487408
2202	179	122	Base	966	2636	39.59	198136
1245	286	119	Base	1296	3633	9.48	167572
1903	462	102	Rain	1435	597	55.95	260613
0611	250	74	Rain	872	390	599.24	502576
0507A	78	38	ND	3614	2042	0	90052
0605A	64	19	Rain	1076	834	0	183995
0611C	135	17	Base	53	1561	0	280526
1226C	224	13	Rain	180	154	0	119715
1226D	251	12	Rain	296	86	0	87565
0507B	20	8	Rain	1520	222	0	7764
0505D	33	7	Rain	5499	659	0	94455
0612B	14	4	ND	274	449	0	19368

Livestock and Wildlife Dominated Cluster

The final group of watersheds consists of 6 watersheds with a high percentage of watershed area under pastureland and cropland and with very high bacterial loadings from both livestock and wildlife compared to other groups of watersheds. The densities of households on public sewer and on OSSF are low. Table 4.7 shows some important water quality characteristics of the stream segments within this cluster. Four watersheds showed high means during wet periods, and for one watershed there was no significant difference between the means of two periods. Except for one, all the stream segments showed relatively low mean concentrations during the dry period. The stream segment that showed high mean concentrations during the dry period (Segment 1255) also showed very high mean concentrations during the wet period. Hence, it can be assumed that this stream segment has high contribution from a nonpoint source and either high background concentrations, a constant point source, or presence of animals within accessible distance of the waterbody. The stream segments 1217A and 1222A had very few in-stream bacterial quality observations. Out of the few observations only one observation exceeded the water quality standard of 400 #/100ml. Detailed modeling of these watersheds is not possible because of the limited observations. Hence, either more data should be collected for these stream segments before a detailed modeling or preliminary modeling should be done using simple a model like PLOAD. The other watersheds can be modeled using HSPF, QUAL2E, or MIKE-BASIN. Since MIKE-BASIN is a commercial software and QUAL2E is only capable of addressing steady state condition, HSPF is the most appropriate model for this group of stream segments.

Table 4.7 Water quality characteristics of stream segments belonging to livestock and wildlife dominated cluster.

Segment ID	Total No. of observations	Observations with >400 (cfu/100ml)	Period of larger means	Wet period mean (cfu/100ml)	Dry period mean (cfu/100ml)	WWTP capacity (mgd)	Watershed area (ac)
1226	2037	483	Rain	1154	572	1.83	505641
1255	567	333	Rain	10250	2189	0.71	85146
1221	418	119	Rain	487	249	2.45	765811
1226A	135	18	Rain	1148	97	0	57937
1217A	46	1	ND	248	166	0	122866
1222A	3	1	NED	911	-	0	19262

Validation of model selection

The validation of the process to select a model for a particular cluster will be carried out by calibrating the selected model for one of the stream segments belonging to the cluster and then validating the calibrated model on a second stream segment in the cluster. The model parameters that most affect the bacteria model of HSPF are land use dependent. Hence, the model parameters that are land use dependent will be calibrated to the first stream segment and then applied to the second stream segment in the same cluster. It is highly unlikely that the two stream segments will have the same calibrated parameter values. However, similar watershed characteristics in a particular cluster will likely result in similar model parameters for the two stream segments. Two stream segments from two clusters (cluster 4 and 5) were selected for the analysis (Table 4.8). The HSPF model water quality predictions are driven by the model hydrology and precipitation patterns are different for each watershed in the same cluster. Therefore, the HSPF model hydrology will first be calibrated individually for stream segments 1903, 0611C, 1255 and 1226. Next, stream segments 1903 and 1255 will be calibrated for bacteria. Finally, the HSPF in-stream bacterial predictions for stream segment 0611C using the water quality parameters from the calibrated model for stream segment 1903 will be compared to the observed in-stream bacterial concentrations. Similarly, the HSPF in-stream bacterial predictions for stream segment 1226 using the water quality parameters from the calibrated model for stream segment 1255 will be compared to the observed in-stream bacterial concentration. The closer the predictions are to the observed values, the better the applicability of the group-based modeling approach.

Table 4.8 Stream segments and models selected for validation of the model selection approach.

Cluster	Model	Stream Segment IDs
Cluster with Low loading from all sources (Cluster 4)	HSPF	1903 and 0611C
Livestock and wildlife dominated cluster (Cluster 5)	HSPF	1255 and 1226

CHAPTER V

VALIDATION OF GROUP-WISE TMDL APPROACH FOR BACTERIALLY IMPAIRED WATERBODIES

Synopsis

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) listed 110 stream segments in Texas with pathogenic bacteria impairment in the year 2000. In order to test the hypothesis of grouping the stream segments based on watershed characteristics and conducting a modeling study using a group-wise approach, the impaired stream segments within five river basins Brazos, Neches, Nueces-Rio Grande Coastal, Sabine and San Antonio were selected based on their association with expected point and nonpoint sources. Using different multivariate statistical techniques such as, factor analysis/principal component analysis, cluster analysis, and discriminant analysis the waterbodies were clustered into five groups having similar watershed characteristics. The main characteristics considered for the classification of waterbodies were designated use of the waterbody, land use distribution, density of stream network, average distance of a land of a particular use to the closest stream, household population, density of on-site sewage facilities (OSSF), bacterial loading due to the presence of different types of farm animals and wildlife, and average climatic conditions. Five clusters of watersheds were formed as a result of the statistical analysis. In order to test the possibility of applying the same model for a group of watersheds, two watersheds each from two clusters formed during the multivariate statistical analysis were selected. Hydrological Simulation Program-FORTRAN (HSPF) model was calibrated for one watershed within each group and validated for the other watershed in the same group to study the similarity in the optimal parameter sets due to the similarity in watershed

characteristics. The study showed that the watersheds within a given cluster yielded similar model results for same model input parameters.

Introduction

Under the Clean Water Act (CWA) program the Texas Commission on Environmental Quality (TCEQ) listed 110 stream segments in the year 2000 with pathogenic bacteria impairment (TNRCC, 2000). The next logical step is to verify impairment and if sufficient evidence is present then to develop Total Maximum Daily Loads (TMDLs) for each of these impaired waterbodies. The development of such a large number of TMDLs will require a colossal resource of both money and man power. It is possible that many of the waterbodies considered for TMDL development listed under the current CWA §303(d) for Texas may be grouped based on their watershed characteristics and the potential sources of pollution. Such a grouping scheme would be helpful in reducing the cost of restoration of water quality by restricting the development of TMDLs for only one or two representative waterbodies under a single group and applying the knowledge to other waterbodies in the same group. In order to test this hypothesis the impaired stream segments within five river basins Brazos, Neches, Nueces-Rio Grande Coastal, Sabine and San Antonio were selected based on expected point and nonpoint sources. Using different multivariate statistical techniques such as, factor analysis/principal component analysis, cluster analysis, and discriminant analysis the waterbodies were clustered into five groups having similar watershed characteristics. The first group of stream segments showed a high density of public sewer systems and a relatively high percentage of urban land use compared to the other clusters of stream segments. The second group had a high density of On-site Sewage Facilities (OSSF), and a relatively high contribution of bacterial loading from livestock. The third group of stream segments was located in areas dominated by forestland together with a relatively high contribution of bacterial loading from livestock, relatively low bacterial loading from wildlife and low densities of OSSF. The fourth group of stream segments had a high percentage of pastureland but low bacterial loading from animals and low densities of OSSF. The fifth group of stream segments was located in areas with high percentages of pastureland and cropland and had high bacterial loading from both livestock and wild animals. The primary aim of the current study is to test the validity of the selection of an appropriate water quality model for a given cluster of stream segments and test whether the same model input parameters for watersheds in the same cluster will produce similar model results. Two stream segments were selected each from the fourth and fifth clusters of stream segments. Since HSPF is considered to be the most comprehensive and flexible model for watershed hydrology and water quality available (Zoppou, 2001), it was selected for modeling these watersheds. The HSPF model is also widely used for bacterial TMDL studies across the United States (Donigian and Huber, 1991). The model parameters that most affect the bacteria model of HSPF are land use dependent. Hence, the land use dependent model parameters are calibrated for the first stream segment in each cluster and then the model is validated for the second stream segment in the same cluster. It is highly unlikely that the two stream segments yield the same calibrated parameter values. However, similar watershed characteristics in a particular cluster will likely result in similar model parameters for the two stream segments. If the selected model is appropriate for a given cluster, the comparisons between the model-predicted in-stream bacterial concentrations and the observed in-stream bacterial concentrations for the same model input parameters would yield similar results for all stream segments in the same cluster.

Methodology

Overview

The stream segments selected are Medina River and Mud Creek in cluster four and Upper North Bosque River and North Bosque River in cluster five. Stream segments in cluster four have a high percentage of pastureland and low bacterial loading from animals and low densities of OSSF. Upper North Bosque River and North Bosque River were grouped under cluster five which contained stream segments with high percentages of pastureland and cropland and had high bacterial loading from both livestock and wild animals. The HSPF model was selected for modeling in-stream bacterial concentrations for all of these stream segments.

The methodology of the current study was as follows:

1. Calibrate the HSPF model hydrology for each of the four watersheds.
2. Estimate the initial values of the water quality parameters for the four watersheds based on the number of farm animals and wildlife.
3. Modify the initial values of water quality parameters for the Medina River watershed to calibrate HSPF for in-stream bacterial concentration at the watershed outlet using observed in-stream bacterial concentration at the same location.
4. Run the HSPF model for the Mud Creek watershed with the calibrated model parameters from the Medina River watershed and compare the goodness-of-fit measures for the two watersheds.
5. Repeat steps 3 and 4 for the Upper North Bosque River and the North Bosque River watersheds, respectively.

Description of Watersheds and Input Data

The Upper North Bosque River (segment number 1255) is one of the tributaries of the Brazos River. The watershed area of the Upper North Bosque River lies between latitude 32.127 N and 32.285 N and longitude 98.151 W and 98.250 W (Figure 5.1). The total length of the stream segment is 20.8 km. The stream originates in the north central region of the Brazos River Basin and flows towards the southeast where it joins the North Bosque River. The watershed covers an area of 344.6 km^2.

Figure 5.2 shows the Medina River (segment number 1903) and its watershed. The watershed area of the creek lies between latitude 29.227 N and 29.511 N and longitude 98.401 W and 98.934 W. The Medina River is one of the major tributaries of the San Antonio River with a total length of approximately 128 km. The total area of the delineated watershed is 1054.7 km^2.

The North Bosque River (segment number 1226) is a major tributary of the Brazos River with a total length of 164.8 km (Figure 5.3). The watershed covers an area of 2046.3 km^2 and lies between latitude 31.593 N and 32.127 N and longitude 97.287 W and 98.164 W. The stream originates in the north central region of the Brazos River Basin and flows towards the southeast where it joins the Brazos River.

Figure 5.4 shows Mud Creek (segment number 0611C) and its watershed. The watershed area of the Mud Creek lies between latitude 31.792 N and 32.213 N and longitude 94.963 W and 95.210 W. Mud Creek is one of the tributaries of the Neches River with a total length of approximately 81.6 km. The creek originates from below Lake Tyler. The total area of the delineated watershed is 1135.3 km^2.

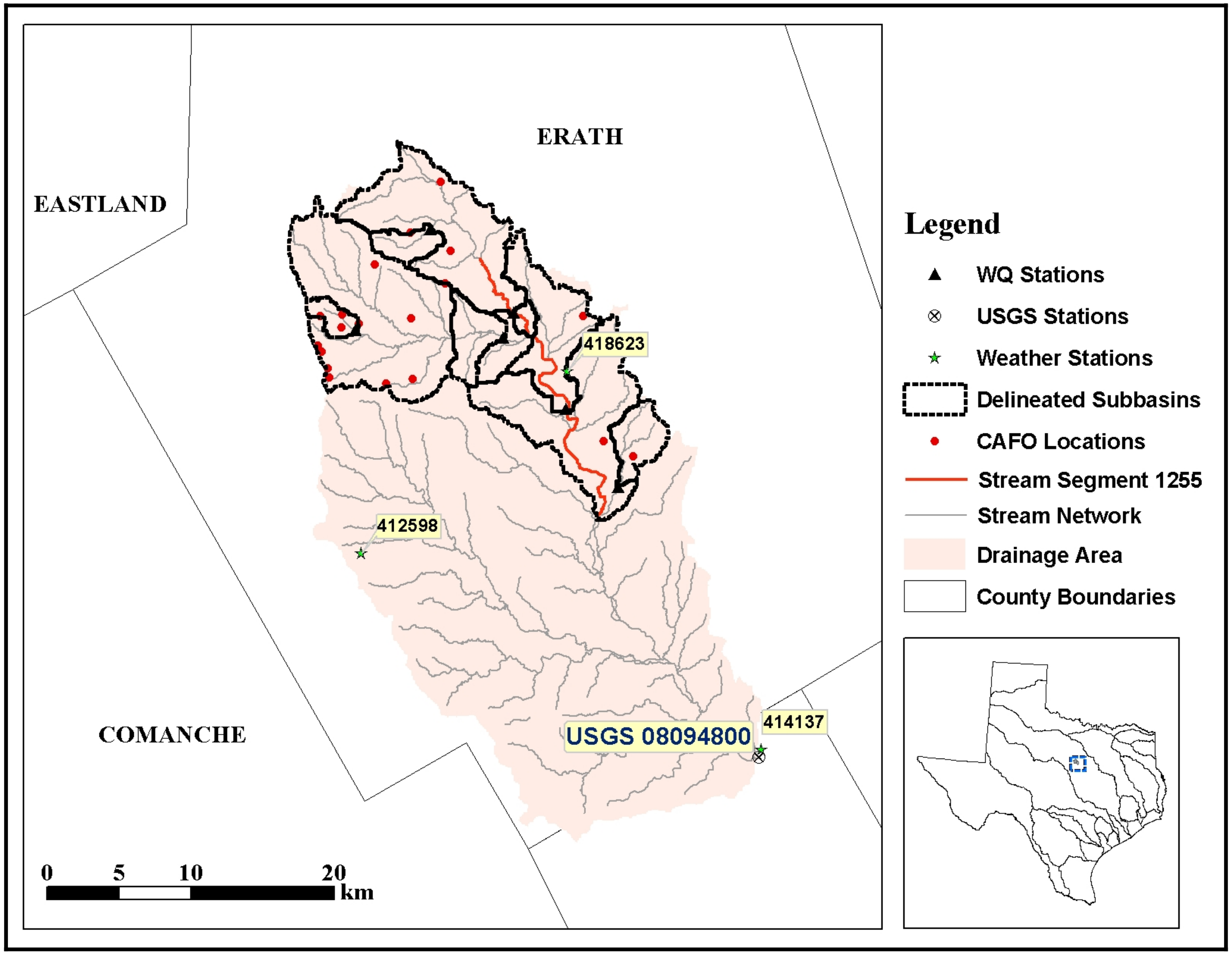

Figure 5.1 Location of weather stations, USGS gauging stations and water quality observation stations within the Upper North Bosque River Watershed.

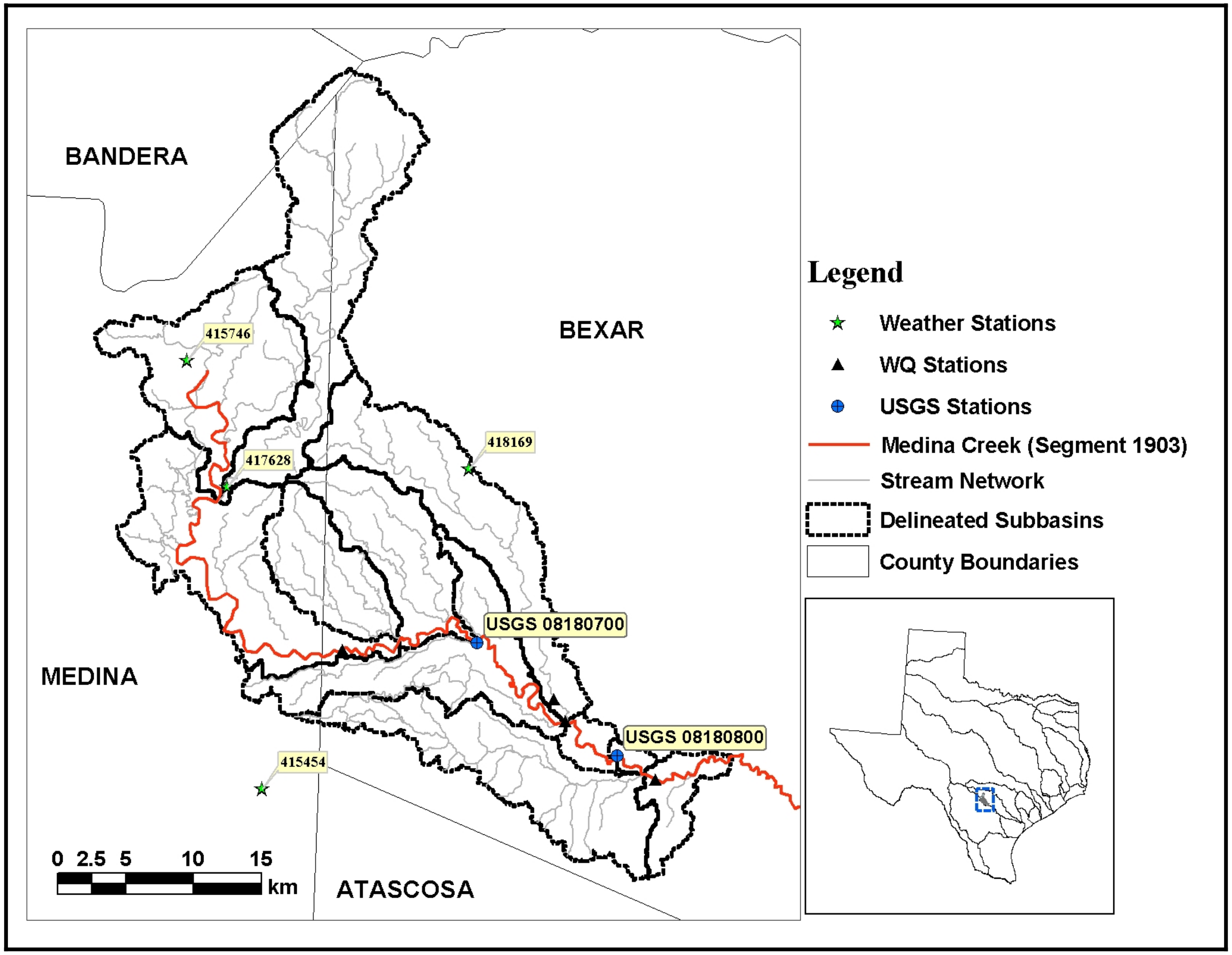

Figure 5.2 Location of weather stations, USGS gauging stations and water quality observation stations within the Medina River Watershed.

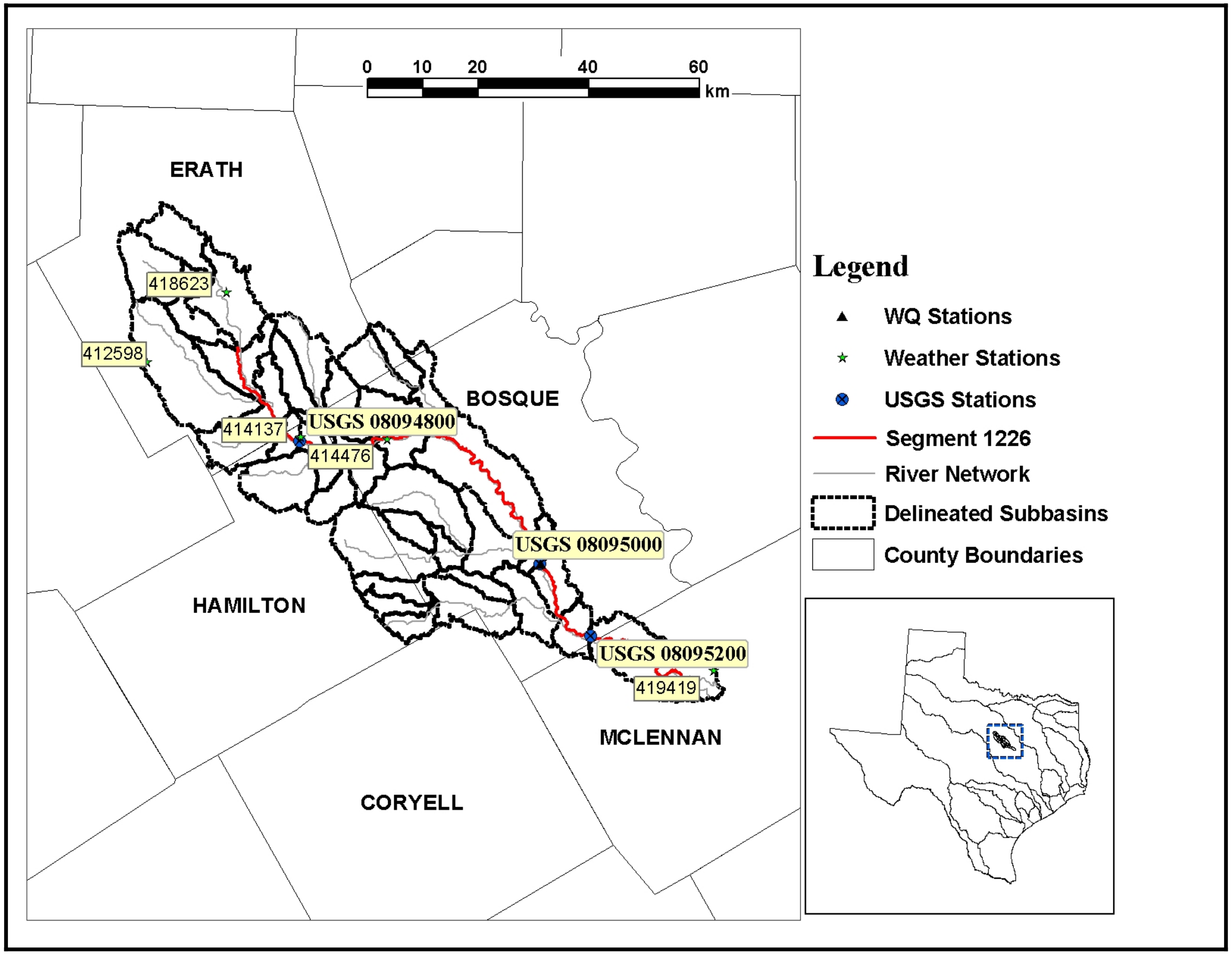

Figure 5.3 Location of weather stations, USGS gauging stations and water quality observation stations within the North Bosque River Watershed.

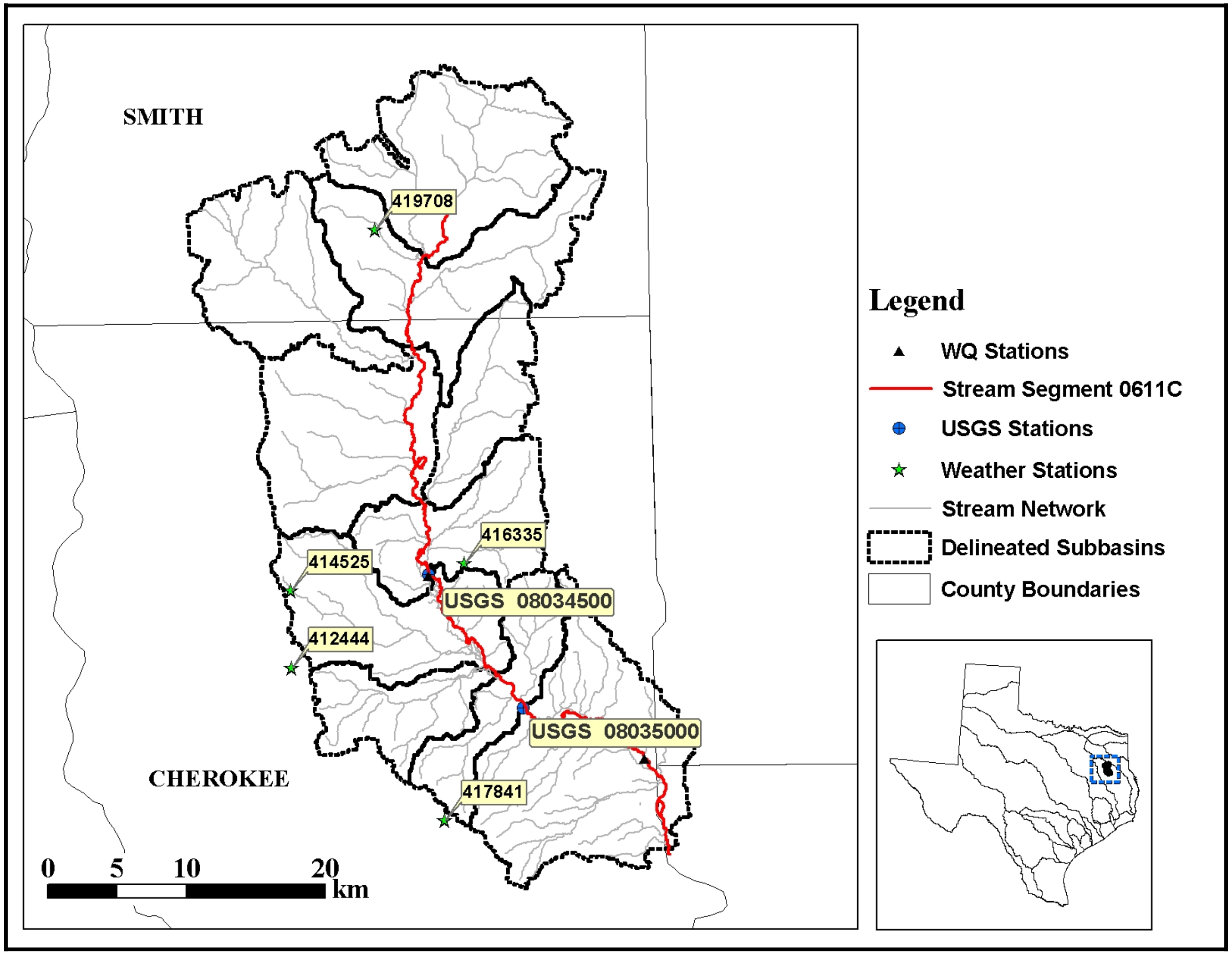

Figure 5.4 Location of weather stations, USGS gauging stations and water quality observation stations within the Mud Creek Watershed.

The land use distribution was determined using the National Land Cover Data (NLCD) obtained from the USGS. The NLCD data were developed from the Landsat satellite Thematic Mapper data acquired by the Multi-resolution Land Characterization (MRLC) Consortium with a resolution of 30m. Table 5.1 shows the total area and percentages of principal land uses of the selected watersheds. The sub-basin level land use distributions are given in Appendix C.

Table 5.1 Principal land uses of watersheds selected for modeling study.

Watershed Name	Area (km^2)	Forestland %	Cropland %	Urban %	Pasture/ Range land %
Mud Creek	1135.3	44	1	2	43
North Bosque River	2046.3	17	7	1	38
Upper North Bosque River	344.6	11	18	5	47
Medina River	1054.7	36	17	4	30

A Digital Elevation Model (DEM) at 1:24,000 scale with a spatial resolution of 85m was obtained from USGS. The detailed stream network was obtained from the National Hydrography Dataset (NHD). NHD is based upon the content of the USGS Digital Line Graph (DLG) hydrography data integrated with reach-related information from the EPA Reach File Version 3 (RF3). It is a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells. NHD is an improvement of DLG and RF3 without replacing them. It is based on 1:100,000-scale data. The watersheds corresponding to the selected stream segments were delineated using the DEM and NHD stream network.

Soil data was taken from the State Soil Geographic database (STATSGO) available at EPA's BASINS data web site. The STATSGO data has a resolution of 250m at 1:250,000 scale.

Climatic data such as daily precipitation, daily maximum and minimum temperatures, and lake evaporation were obtained from the National Climatic Data Center (NCDC). Figures 5.1 through 5.4 show the locations of NCDC weather stations within each watershed. Since HSPF requires an hourly timeseries of precipitation, the available daily values were disaggregated into hourly values using WDMUtil tool available with BASINS (USEPA, 1996a).

Historical daily mean stream flow data corresponding to the USGS gauging stations (Figures 5.1 through 5.4) were obtained from USGS for the simulation period. Historical water quality data for fecal coliform at various water quality stations (Figures 5.1 through 5.4) in the watershed were obtained from the Texas Commission on Environmental Quality [TCEQ] (J. Allen, personal communication, 2002).

Bacterial Source Assessment

There are no known facilities within the watersheds selected for the current study which are permitted to discharge bacteria into the receiving waterbodies. Hence, the main sources of fecal bacteria considered for the current study were limited to nonpoint sources. Nonpoint sources of fecal coliform include on-site sewage facilities (OSSF), livestock, wildlife, and pets. There is only limited information regarding the pet population of these watersheds, therefore this source was not considered. The description of data related to other sources is given in the following sections.

On-site Sewage Facilities

The number of OSSF within a watershed was calculated based on US Census data and data obtained from TCEQ. The GIS layer that shows the boundaries of census blocks and the population data for each census block were obtained from the US Census Bureau (US Census Bureau, 2000). The population and number of households within a watershed were calculated by summing data from census blocks intersecting the watershed under consideration. The number of OSSF installed before 1990 was obtained from the 1990 US Census. The number of OSSF installed after 1990 was obtained from the TCEQ (K. Neimann, personal communication, 2003). It was assumed that the OSSF were generally present only outside major cities. An assumption was made that households located within major cities or Metropolitan Statistical Areas (MSA) are generally connected to the public sewer systems. The GIS data layer corresponding to these MSA was obtained from the Spatial Science Laboratory at Texas A & M University. The number of OSSF within the watershed, NSS_{ws}, was calculated as:

$$NSS_{ws} = \sum_{c=1}^{C} \frac{HH_{ws} \times NSS_c}{HH_c} \tag{5.1}$$

where, HH_{ws} is the number of households within the watershed, HH_c is the number of households within the county, NSS_c is the number of OSSF in the county, and C is the number of counties intersecting the watershed. Table 5.2 shows the estimated numbers of OSSF within each watershed.

Table 5.2 Estimated numbers of OSSF within each watershed.

Watershed Name	Number of OSSF
Mud Creek	8594
North Bosque River	6366
Upper North Bosque River	3906
Medina River	3233

Wildlife Population

The largest relevant group of wildlife in Texas is the white-tailed deer. Availability of information regarding wildlife was limited to the number of white-tailed deer by county in Texas. The estimated number of deer present in all Texas counties was obtained from the Texas Parks and Wildlife Department for the last 5 years. Deer were assumed to only be present in land use areas of forestland, barren land and pastureland. The number of wildlife in a watershed (ND_{ws}) was estimated using the number of white-tailed deer reported for the county (ND_c) and the land use distribution within the county and the watershed using the following equation:

$$ND_{ws} = \sum_{c=1}^{C} ND_c \times \frac{Forest_{ws} + Barren_{ws} + Pasture_{ws}}{Forest_c + Barren_c + Pasture_c} \qquad (5.2)$$

where, *Forest* is the area of forestland, *Barren* is the area of barren land, *Pasture* is the area of pastureland, and *ws* and *c* represent watershed and county, respectively. The areas corresponding to the different land uses were calculated using the NLCD GIS layers. Table 5.3 shows the estimated number of deer within each watershed selected for the current study.

Table 5.3 Estimated wildlife populations within each watershed.

Watershed Name	Number of deer
Mud Creek	3557
North Bosque River	23658
Upper North Bosque River	2198
Medina River	8095

Livestock

The major source of fecal coliform bacteria from agricultural land and pastureland is manure application or direct contribution from livestock and poultry. The numbers of different types of livestock and poultry within the watershed was estimated based on data from the US Department of Agriculture (USDA) Agricultural Statistics Database (USDA, 2002) and data from Concentrated Animal Feeding Operations (CAFO). The livestock data collected included the number of cattle, swine, sheep, and goats. Cattle data was available for each county. Goat, swine and sheep data were available partially on the county level and partially on the agricultural district level. The location and numbers of CAFOs were obtained from TCEQ. The calculation of the total number of animals in each watershed, NA_{ws}, was based on the USDA database and the CAFO data layer. Livestock and poultry were assumed to be present only in pastureland and cropland. The number of animals present in the CAFO layer within a county or an agricultural district was subtracted from the total number of animals reported in the USDA database. The result is the total number of animals that are not in CAFO. Then the number of animals within the county or the agricultural district was multiplied by the ratio of pastureland area within the watershed to the pastureland area within the given county or agricultural district, to obtain (non-CAFO) animals within the watershed. The number of animals within a watershed that were part of a CAFO was obtained directly from the CAFO data layer. Thus, the total number of animals within a watershed would be the sum of animals present in the CAFO and the number of animals that were not part of any CAFO. The calculations are shown as below;

$$NA_{ws} = \sum_{c=1}^{C} (NA_c - CA_c) \times \frac{Pasture_{ws}}{Pasture_c} + CA_{ws} \tag{5.3}$$

or

$$NA_{ws} = \sum_{c=1}^{D} (NA_d - CA_d) \times \frac{Pasture_{ws}}{Pasture_d} + CA_{ws} \tag{5.4}$$

where *NA* is the number of given animal, *CA* is the number of animals reported in the CAFO data layer, *Pasture* is the area under pastureland, and *ws, c* and *d* represent watershed, county, and district, respectively. Table 5.4 shows the estimated numbers of different livestock within each watershed.

Table 5.4 Estimated livestock populations within each watershed.

Watershed Name	Cattle	Goat	Swine	Sheep
Mud Creek	39238	86	251	16
North Bosque River	77065	0	1	2
Upper North Bosque River	29751	990	84	376
Medina River	23225	1715	101	4

HSPF Model Description

HSPF is a continuous hydrologic modeling software that can be used to simulate a comprehensive range of hydrologic and water quality processes (Bicknell et al., 1997). In HSPF, modules are divided into pervious land, impervious land, and reaches, *PERLND, IMPLND,* and *RCHRES,* respectively. Each land segment is considered as a lumped catchment and simulations are based on an hourly time step.

The hydrologic simulation in HSPF is based on a mass balance approach. On pervious land surfaces the amount of precipitation is divided into the following components; direct runoff, direct evaporation from the land surface, surface storage followed by evaporation, surface storage followed by interflow, and infiltration to the subsurface zone. The subsurface zone is considered to be divided further into upper zone, lower zone and deep groundwater zone. Water reaching the subsurface layer either remains in storage, percolates to the zone below the current zone or evaporates. Water that percolates deep below the groundwater zone is considered to be lost from the system. The amount of precipitation over an impervious land surface contributes to overland flow, evaporates directly, or remains in surface detention storage. The surface runoff calculation is based on Chezy-Manning's equation and an empirical relationship between the outflow depth and detention storage. The evapotranspiration (ET) is calculated based on the user supplied potential ET and moisture available in surface and subsurface zones to meet the potential ET demand.

The water quality simulation in HSPF assumes constituents are associated with one or more of the following; sediment, overland flow, interflow, and groundwater flow. Most frequently, fecal coliform is assumed to be flow associated. Flow associated constituents are assumed to be accumulated on the land surface until the occurrence of a rainfall event. The contribution of fecal coliform from the land surface to the stream is calculated using the following equation:

$$SOQO = SQO\left[1.0 - e^{(SURO \times WSFAC)}\right] \tag{5.5}$$

where, SOQO is the washoff of fecal coliform from the land surface (colony forming unit [cfu] acre^{-1} day^{-1}) , SQO is the storage of fecal coliform on the surface (cfu acre^{-1}), SURO is the surface outflow of water (in. acre^{-1}), and WSFAC is the susceptibility of the quality constituent to washoff (in.$^{-1}$). The amount of storage on the land surface depends on the accumulation rate of bacteria and the maximum limit of storage depending upon the land use type. The contribution of fecal coliform through interflow or groundwater flow is calculated using constant concentrations in these flows. The in-stream fecal coliform dynamics is calculated based on Chick's law (Crane and Moore, 1986).

Hydrology Calibration

The HSPF model hydrology calibration was carried out using a model-independent parameter estimator program, PEST (Doherty, 2001a; Doherty, 2002) together with a model-independent time-series processor, TSPROC (Doherty, 2001b). PEST achieves model calibration in fewer model runs than many other automatic calibration programs by minimizing the objective function using a robust variant of the Gauss-Marquardt Levenberg method (Doherty and Johnston, 2003). The objective function is usually the sum of squares of the differences between observed and modeled flows, volumes and/or exceedence times for various thresholds of the flows. The HSPF model was calibrated for the watersheds by adjusting the parameters listed in Table 5.5 to obtain a good agreement between the observed and modeled daily flows and baseflows. These parameters were selected based on the fact that they are the most sensitive parameters that control the different phases of the water balance. All of the selected parameters Except RETSC all parameters belong to PERLND module of HSPF. RETSC belongs to IMPLND module.

Table 5.5 HSPF parameters considered during model calibration and their descriptions.

Parameter	Description	Unit
LZSN	Lower zone nominal storage	in.
INFILT	Soil infiltration capacity related parameter	in. hr^{-1}
AGWRC	Basic groundwater recession rate if KVARY is zero	d^{-1}
DEEPFR	Fraction of groundwater inflow that will enter deep groundwater and be lost	Fraction
BASETP	Fraction of remaining PET which can be satisfied from baseflow	Fraction
AGWETP	Fraction of remaining PET which can be satisfied from active groundwater storage	Fraction
MON-UZSN	Upper zone nominal storage, monthly varied	in.
INTFW	Interflow inflow parameter	None
IRC	Interflow recession parameter	d^{-1}
MON-LZETPARM	Lower zone ET parameter, an index to the density of deep-rooted vegetation, monthly varied	None
RETSC	Retention storage capacity of the surface	in.

In the current study, in each models runs during hydrology calibration PEST modified each input parameters (Table 5.5) based on the current value of that parameter and factor. The relationship of parameters IRC and AGWRC with the HSPF hydrology outputs is highly nonlinear (Doherty, 2001b). In order to overcome this nonlinearity problem, new parameters were derived from these parameters with the following transformations and used during the parameter estimation process. The relationships between the derived parameters and the original parameters are:

$$IRCTRANS = IRC/(1 - IRC) \tag{5.6}$$

and

$$AGWRCTRANS = AGWRC/(1 - AGWRC) \tag{5.7}$$

Also for better numerical stability and to reduce any nonlinearity problem, all the parameters were log-transformed during the parameter estimation process.

The objective function for the calibration comprised of the sum of squares of the differences between the observed and simulated daily flows, daily baseflows calculated from observed and simulated daily flows, monthly volumes calculated from observed and simulated daily flows, and the exceedence probabilities for various flow thresholds corresponding to observed and simulated daily flows. The objective of PEST runs was to minimize the objective function value. The baseflow, monthly volumes, and exceedence probabilities were all calculated using TSPROC program. Different weights were assigned to the four groups for obtaining an approximately equal contribution from all the groups. The low flows were given more weight to reduce the effect of the extremely high flows on the objective function. The initial values for the parameters were calculated in accordance to the BASINS Technical Note 6 (USEPA, 2000b). The bounds for each of the parameters were selected based on the literature values (USEPA, 2000b) and data layers corresponding to the soil types, and land use distribution.

Statistics Used to Evaluate Model Results

The following statistics were used to determine when adequate calibration of simulated daily and monthly flow to measured data from USGS gages was achieved.

1. Nash-Sutcliffe model efficiency, *E* (Nash and Sutcliffe, 1970):

$$E = 1 - \frac{\sum_{i}^{n} (O_i - S_i)^2}{\sum_{i}^{n} (O_i - O')^2} \tag{5.8}$$

where O_i and S_i are the observed and simulated daily flows for the ith day respectively, O' is the mean of the observed daily flow and n is the total number of days. A positive value of *E* indicates an acceptable fit between the observed and simulated flows, therefore, the model can be considered to be a better predictor of the system than the mean of the observed data. The closer *E* is to 1, the better the agreement between the observed and simulated flow.

2. Standard error, *SE*:

$$SE = \sqrt{\frac{\sum_{i}^{n} (O_i - S_i)^2}{n-1}} \tag{5.9}$$

where the parameters are defined as above. The closer SE is to zero, the better the simulation results.

3. Modified index of agreement, $d_1^{'}$ (Legates and McCabe, 1999):

$$d_1^{'} = 1.0 - \frac{\sum_{i}^{n} \left|O_i - S_i\right|}{\sum_{i}^{n} \left(\left|S_i - O^{'}\right| + \left|O_i - O^{'}\right|\right)} \tag{5.10}$$

The modified index of agreement varies from 0.0 to 1.0. The closer the value is to 1.0, the better the agreement between the simulated and measured flows.

4. Modified coefficient of efficiency, $E_1^{'}$ (Legates and McCabe, 1999):

$$E_1^{'} = 1 - \frac{\sum_{i}^{n} |O_i - S_i|}{\sum_{i}^{n} |O_i - O'|} \tag{5.11}$$

The interpretation of modified coefficient of efficiency is similar to that of the Nash-Sutcliffe model efficiency, except that errors and differences are not inflated by their squared values.

Though the interpretation of model efficiency and modified coefficient of efficiency are similar to each other, they are different from correlation based statistics such as the coefficient of determination, R^2. An R^2 value of 0.7 indicates that the model is capable of explaining 70% of the total variance in the observed data. A value 0.7 for model efficiency indicates that mean of the squared differences between the observed and model predictions is equal to 30% of the variance in the observed data (Legates and McCabe, 1999). A positive value for E and $E_1^{'}$ indicates that the model can be used as an adequate prediction tool. In general, a value equal to or greater than 0.5 for E and $E_1^{'}$ is considered to be very good (Santhi et al., 2001).

Water Quality Calibration

After the model hydrology was calibrated for each watershed, the water quality component of HSPF was calibrated for the Medina River and Upper North Bosque watersheds. The parameters that were adjusted to calibrate the water quality component of HSPF were the monthly varying maximum accumulation of fecal coliform bacteria on land (MON-SQOLIM), the monthly varying rate of accumulation of fecal coliform bacteria on land (MON-ACCUM), the rate of surface runoff per hour that will remove 90% of stored fecal coliform (WSQOP), the in-stream first-order decay rate of bacteria (FSTDEC), and the monthly varying concentration of fecal coliform in interflow (MON-IFLW-CONC). The objective was to calibrate the model for these watersheds and to validate the model for another watershed in the same cluster. The model was validated for the Mud Creek and North Bosque River watersheds using the calibration parameters obtained from the Medina River and Upper North Bosque River watersheds, respectively.

Calculation of Initial Fecal Coliform Accumulation Rates

The calculation of these parameters followed the procedures outlined in "Protocol for Developing Pathogen TMDLs" (USEPA, 2001c) and in "EPA's Bacterial Indicator Tool User's Guide" (USEPA, 2000a).

Cropland

Sources of fecal coliform on cropland are primarily wildlife, hog manure, and cattle manure. Bacterial loading from wildlife on cropland is calculated as:

$$Wild_c = \frac{\left(N_{deer} \times LR_{deer}\right)}{Area_{crop}} \tag{5.12}$$

where, $Wild_c$ is the fecal coliform contribution from wildlife on cropland, (cfu.d^{-1}.ac.$^{-1}$), N_{deer} is the total number of deer on cropland, $Area_{crop}$ is the area under cropland (ac.), LR_{deer} is the fecal coliform bacteria production rate of deer (cfu. d^{-1}).

Bacterial loading on cropland from hog manure is assumed to vary by month and is calculated as:

$$Hog_{c,i} = \frac{365 * N_{hog} \times LR_{hog} \times MF_i \times FR_{hog,i}}{ND_i \times Area_{crop}} \tag{5.13}$$

where, $Hog_{c,i}$ is the fecal coliform contribution from hog manure on cropland in the ith month (cfu.d^{-1}.ac.$^{-1}$), N_{hog} is the total number of hogs/swine on cropland, LR_{hog} is the fecal coliform bacteria production rate of hog (cfu.d^{-1}), MF_i is the fraction of annual manure applied in ith month, $FR_{hog,i}$ is the fraction of hog manure available for runoff in ith month, and ND_i is the number of days in ith month

Monthly bacterial loading from cattle on cropland is calculated as:

$$dairy_{c,i} = \frac{365 * N_{dairy} \times LR_{dairy} \times MF_i \times FR_{dairy,i}}{ND_i \times (Area_{crop} + Area_{pasture})}$$

$$beef_{c,i} = \frac{(365 - ND_{grazing}) \times N_{beef} \times LR_{beef} \times MF_i \times FR_{beef,i}}{ND_i \times (Area_{crop} + Area_{pasture})}$$

$$Cattle_{c,i} = dairy_{c,i} + beef_{c,i} \tag{5.14}$$

where, $Cattle_{c,i}$ is the fecal coliform contribution from cattle manure on cropland in ith month (cfu.d^{-1}.ac.$^{-1}$), N_{dairy} is the total number of dairy cows on cropland, N_{beef} is the total number of beef cows on cropland, $Area_{pasture}$ is the area under pastureland (ac.), $FR_{dairy,i}$ is the fraction of dairy manure available for runoff in ith month, $FR_{beef,i}$ is the fraction of beef manure available for runoff in ith month, LR_{dairy} is the fecal coliform bacteria production rate of dairy cows (cfu.d^{-1}), LR_{beef} is the fecal coliform bacteria production rate of beef cows (cfu.d^{-1}), and $ND_{grazing}$ is the number of grazing days for beef cows.

Pastureland

Sources of fecal coliform on pastureland are mainly wildlife, cattle manure (both dairy and beef cows) and from grazing beef cattle. The bacterial contribution from wildlife is calculated as:

$$Wild_p = \frac{\left(N_{deer} \times LR_{deer}\right)}{Area_{pasture}} \tag{5.15}$$

The contribution from cattle is calculated as:

$$dairy_{p,i} = \frac{365 * N_{dairy} \times LR_{dairy} \times MF_i \times FR_{dairy,i}}{ND_i \times (Area_{crop} + Area_{pasture})}$$

$$beef_{p,i} = \frac{(365 - ND_{grazing}) \times N_{beef} \times LR_{beef} \times MF_i \times FR_{beef,i}}{ND_i \times (Area_{crop} + Area_{pasture})} + \frac{N_{beef} \times LR_{beef} \times (1 - FT_{con})}{Area_{pasture}}$$

$$Cattle_{p,i} = dairy_{p,i} + beef_{p,i} \tag{5.16}$$

where, $Cattle_{p,i}$ is the fecal coliform contribution from cattle manure over pastureland in i^{th} month ($cfu.d^{-1}.ac.^{-1}$) and FT_{con} is the fraction of time beef cows are in confinement.

Forestland

The major source of fecal coliform on forestland is wildlife. The contribution of bacteria from wildlife on forestland is calculated as:

$$Wild_f = \frac{\left(N_{deer,forest} \times LR_{deer}\right)}{Area_{forest}} \tag{5.17}$$

where, $Wild_f$ is the fecal coliform contribution from wildlife on forestland ($cfu.d^{-1}.ac.^{-1}$), $N_{deer,forest}$ is the total number of deer over forest land, $Area_{forest}$ is the area under the forestland (ac.).

Urban/Built-up Land

Accumulation rate in urban/built-up land was calculated as the weighted average of background concentrations of fecal coliform for four different types of urban landscapes, commercial and services, mixed urban, residential, and transportation, communications and utilities. The values of background concentrations were obtained from the literature (USEPA, 2000a).

The initial values of MON-ACCUM for all the watersheds are given in Appendix C (Tables C.5 through C.7). The initial values of MON-SQOLIM are assumed to be 1.5 times that of MON-ACCUM value for the same month. The initial values of WSQOP, MON-IFLW-CONC, and FSTDEC were kept at their default values and adjusted during the calibration process. In the current study, in most cases, the parameter values for a given land use were kept the same throughout the entire watershed. This was done because the study was intended to evaluate the transferability of model parameters to a different watershed in the same cluster, thus it was desirable to keep generality in the parameter values. However, when the sources of fecal coliform within a sub-watershed were found to be very different from the other sub-watersheds, the parameter values were adjusted for that particular sub-watershed.

Water Quality Calibration and Validation

In order to gain insight into the nature of in-stream water quality in the selected stream segments, data from in-stream fecal coliform samples collected between January 1985 and December 2002 were analyzed. Based on a statistical analysis it was found that in-stream bacterial concentrations during runoff periods for the Upper North Bosque and North Bosque rivers were significantly higher (at α=0.05) than the bacterial concentrations during baseflow periods. This indicates the presence of nonpoint sources in these watersheds. However, the statistical analysis rejected any significant difference between the means of bacterial concentration during stormflow and baseflow periods for the Medina River and Mud Creek watersheds. The details of the statistical analysis are given in Chapter II. To get some idea of the distribution of bacterial concentration observations, boxplots were made for all four stream segments and are given in Figures 5.5 through 5.8. Table 5.6 shows the summary of the bacterial observations from the water quality stations within the watershed boundaries of the selected stream segments.

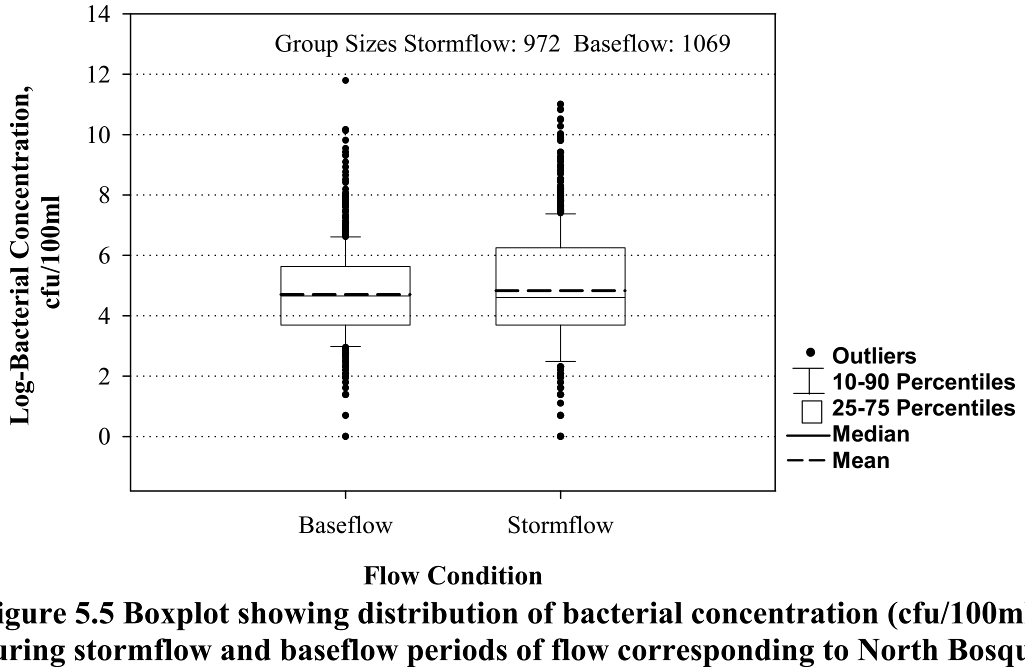

Figure 5.5 Boxplot showing distribution of bacterial concentration (cfu/100ml) during stormflow and baseflow periods of flow corresponding to North Bosque River.

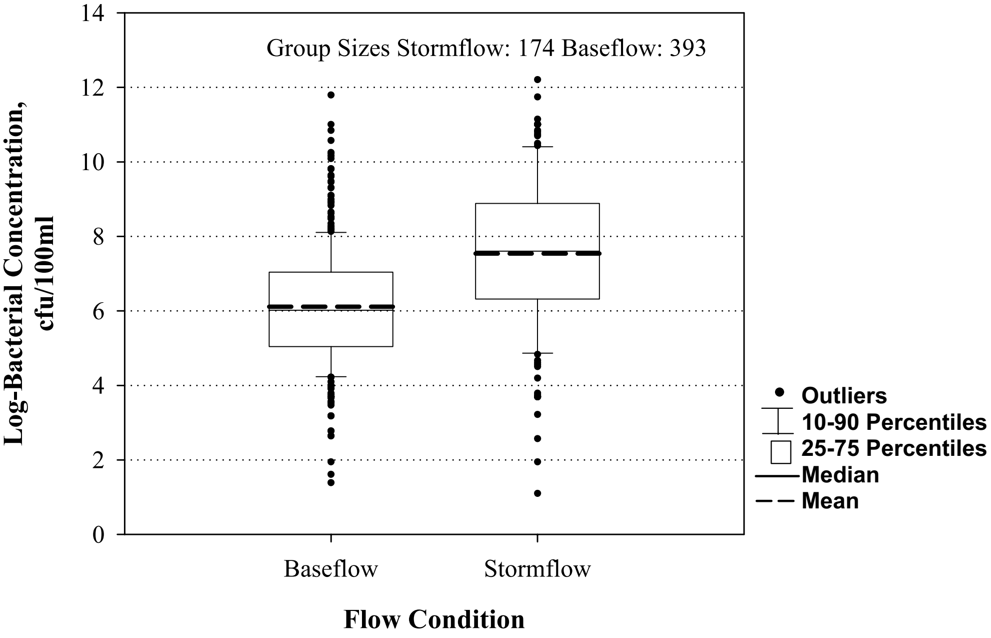

Figure 5.6 Boxplot showing distribution of bacterial concentration (cfu/100ml) during stormflow and baseflow periods of flow corresponding to Upper North Bosque River.

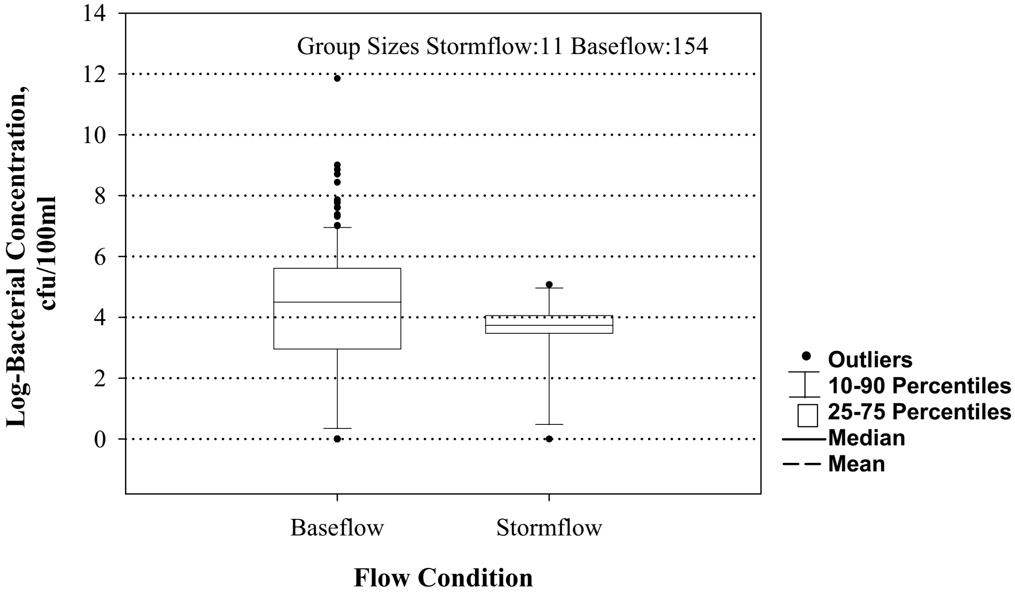

Figure 5.7 Boxplot showing distribution of bacterial concentration (cfu/100ml) during the two periods of flow corresponding to Mud Creek.

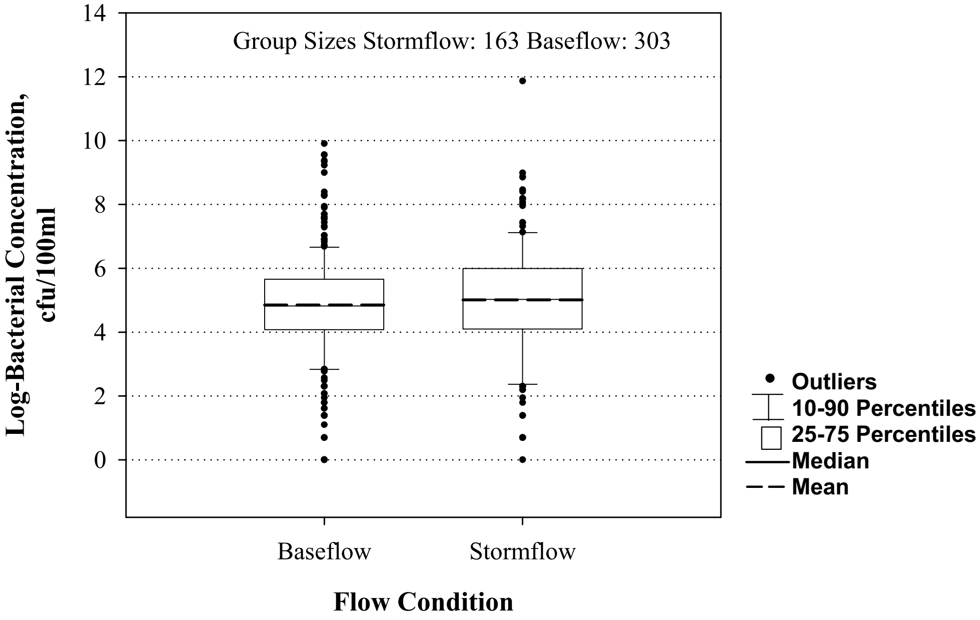

Figure 5.8 Boxplot showing distribution of bacterial concentration (cfu/100ml) during the two periods of flow corresponding to Medina River.

Table 5.6 Summary of bacterial quality observations within the selected watersheds.

Watershed	Count	Minimum (cfu/100ml)	Maximum (cfu/100ml)	Mean (cfu/100ml)	Median (cfu/100ml)
Upper North Bosque River	567	3	200000	4661	590
Medina River	462	1	142000	889	130
Mud Creek	165	1	140000	1241	79
North Bosque River	2037	0	132000	849	102

The objective of this study was to first calibrate the HSPF model for the Upper North Bosque River and Medina River watersheds. The model was then validated using the North Bosque River and Mud Creek watersheds and the calibrated parameters from the Upper North Bosque River and Medina River watersheds, respectively. Since many of the HSPF model parameters vary on monthly basis, monthly in-stream water quality was evaluated by plotting the mean monthly bacterial concentrations for the Upper North Bosque River and Medina River watersheds against time (Figure 5.9). The mean monthly bacterial concentrations of the Upper North Bosque River watershed showed large variation between months, while the mean monthly concentrations of Medina River watershed did not vary much over the period.

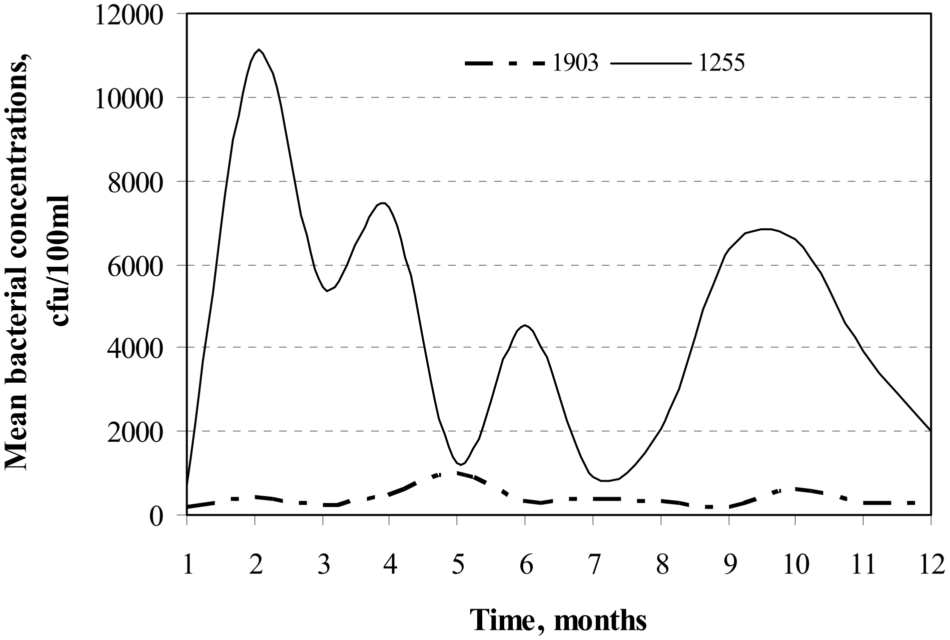

Figure 5.9 Mean monthly bacterial concentrations during 1985-2002 for Medina River (1903) and the Upper North Bosque River (1255).

Results and discussion

Hydrology Calibration

HSPF hydrology was calibrated for all four watersheds in the study using the model independent parameter estimator program, PEST. The selection of simulation periods for each watershed was based on the availability of weather data. The calibration period for the Medina River watershed was between January 1, 1998 and December 31, 1999 and the validation period was January 1, 2000 to December 31, 2000. The calibration and validation periods for both the Upper North Bosque River and North Bosque River watersheds were from January 1, 1980 to December 31, 1993 and January 1, 1994 to December 31, 1995, respectively. For Mud Creek watershed, the calibration and validation periods were from January 1, 1992 to December 31, 1996 and January 1, 1998 to December 31, 2000, respectively. Initial calibration attempts for the North Bosque River watershed showed that a single set of parameter values for the entire watershed would not result in a proper hydrology calibration. Hence, it was decided to use three sets of parameter values. The first set of parameter values were calibrated for the watershed area contributing flow to the USGS gauging station located on the North Bosque River at Hico, TX (08094800), which is the same set for the Upper North Bosque River watershed. The second set of parameter values were calibrated to the watershed area located below the USGS gauging station on the North Bosque River at Hico, TX and contributing flow to the USGS gauging station located on the North Bosque River at Clifton, TX (08095000). The third set of parameter values were calibrated to the watershed area not covered by the first two sets. Similarly, for Mud Creek watershed two sets of parameter values were developed, one for the watershed area above USGS gauging station on the Mud Creek near Jacksonville, TX (08034500) and one for the watershed area below the same station. The calibrated parameter values for the watersheds are listed in Tables 5.7 through 5.9.

Table 5.7 Calibrated values for the constant parameters.

Parameter	Unit	Medina River	Upper North Bosque River	North Bosque 2[+]	North Bosque 3[++]	Mud Creek 1[*]	Mud Creek 2[**]
LZSN	in.	1.010	6.199	4.723	10.930	1.00	0.7
INFILT	in. hr^{-1}	0.050	0.066	0.041	0.059	0.132	0.042
AGWRC	d^{-1}	0.984	0.997	0.994	0.988	0.959	0.94
DEEPFR	Fraction	0.207	0.660	0.010	0.055	0.01	0.01
BASETP	Fraction	0.041	0.013	0.010	0.095	0.01	0.01
AGWETP	Fraction	0.001	0.061	0.081	0.022	0.001	0.001
INTFW	None	10.00	10.00	8.30	8.53	9.91	9.91
IRC	d^{-1}	0.995	0.785	0.843	0.330	0.951	0.951
RETSC	in.	4.27	7.78	9.12	10.00	0.10	0.10

[+] North Bosque 2 is the watershed area below the USGS gauging station on the North Bosque River at Hico, TX and above the USGS gauging station on the North Bosque River at Clifton, TX; [++] North Bosque 3 is the watershed area below the USGS gauging station on the North Bosque River at Clifton, TX and above the USGS gauging station on the North Bosque River at Valley Mills, TX; [*] Mud Creek 1 is the watershed area above the USGS gauging station on Mud Creek near Jacksonville, TX; [**] Mud Creek 2 is the watershed area above the USGS gauging station on Mud Creek at Ponta, TX

Table 5.8 Calibrated values for monthly-varied UZSN (in.) parameter.

Parameter	Medina River	Upper North Bosque River	North Bosque 2	North Bosque 3	Mud Creek 1	Mud Creek 2
Jan	3.68	1.46	3.83	1.46	0.53	4.24
Feb	3.67	1.46	3.52	1.46	0.49	3.9
Mar	3.66	1.45	2.85	1.45	0.39	3.15
Apr	3.64	1.44	1.92	1.44	0.26	2.11
May	3.62	1.44	1.04	1.44	0.14	1.14
Jun	3.61	1.43	0.41	1.43	0.05	0.433
Jul	3.6	1.43	0.22	1.43	0.03	0.227
Aug	3.61	1.43	0.53	1.43	0.07	0.564
Sep	3.62	1.44	1.25	1.44	0.17	1.37
Oct	3.64	1.45	2.15	1.45	0.3	2.37
Nov	3.66	1.45	3.05	1.45	0.42	3.38
Dec	3.67	1.46	3.65	1.46	0.51	4.05

Table 5.9 Calibrated values for monthly-varied LZETP (dimensionless) parameter.

Parameter	Medina River	Upper North Bosque River	North Bosque 2	North Bosque 3	Mud Creek 1	Mud Creek 2
Jan	0.178	1.05	0.99	1.05	0.73	1.09
Feb	0.175	1.03	0.97	1.03	0.70	1.05
Mar	0.169	0.975	0.94	0.97	0.64	0.97
Apr	0.161	0.901	0.89	0.90	0.57	0.85
May	0.153	0.831	0.85	0.83	0.49	0.74
Jun	0.147	0.781	0.82	0.78	0.44	0.66
Jul	0.146	0.766	0.81	0.77	0.42	0.64
Aug	0.148	0.79	0.82	0.79	0.45	0.67
Sep	0.155	0.847	0.86	0.85	0.51	0.76
Oct	0.163	0.919	0.90	0.92	0.59	0.88
Nov	0.171	0.991	0.95	0.99	0.66	0.99
Dec	0.176	1.04	0.98	1.04	0.71	1.07

The goodness-of-fit statistics for the calibration and validation periods for each watershed are given in Tables 5.10 and 5.11, respectively. The observed and simulated daily flows, baseflows, and monthly volumes were in very good agreement for the Medina River and Upper North Bosque River watersheds. Though the Mud Creek watershed at station 08034500 showed good agreement between observed and simulated flows during the calibration period, Mud Creek watershed at station 08035000 showed only acceptable values for the same period. Moreover, for both sites during the validation period the values were found to be acceptable ($E > 0.0$) but poor. The hydrology calibration results for all gauging stations in the North Bosque River watershed were good. For visual comparison, the daily observed and simulated flows during both calibration and validation periods for the main watershed outlets were plotted against time and are given in Figures 5.10 through 5.17.

Table 5.10 Hydrology calibration results.

Watershed	USGS gauging station Location	Calibration period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
Medina River below Medina diversion lake (Segment 1903)	Medina River near Somerset, TX (08180800)	01/01/1998 to 12/31/1999	E	0.54	0.72	0.90
			SE	8.20	2.00	--
			E_1'	0.58	0.64	0.76
			d_1'	0.77	0.80	0.87
Upper North Bosque River (Segment 1255)	North Bosque River at Hico, TX (08094800)	01/01/1980 to 12/31/1993	E	0.53	0.81	0.87
			SE	6.89	1.51	--
			E_1'	0.58	0.65	0.70
			d_1'	0.78	0.82	0.85

Table 5.10 Continued.

Watershed	USGS gauging station Location	Calibration period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
Mud Creek (Segment 0611C)	Mud Creek near Jacksonville, TX (08034500)	01/01/1992 to 12/31/1996	E	0.36	0.52	0.50
			SE	-0.19	5.97	--
			$E_1^{'}$	0.44	0.44	0.48
			$d_1^{'}$	0.64	0.65	0.67
	Mud Creek at Ponta, TX (08035000)	01/01/1992 to 12/31/1996	E	0.03	0.06	0.01
			SE	48.95	29.77	--
			$E_1^{'}$	0.30	0.32	0.29
			$d_1^{'}$	0.58	0.58	0.58

Table 5.10 Continued.

Watershed	USGS gauging station Location	Calibration period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
North Bosque River	North Bosque River at Hico, TX (08094800)	01/01/1980 to 12/31/1993	E	0.53	0.81	0.87
			SE	6.89	1.51	--
			$E_1^{'}$	0.58	0.65	0.70
			$d_1^{'}$	0.78	0.82	0.85
	North Bosque River at Clifton, TX (08095000)	01/01/1980 to 12/31/1993	E	0.06	0.25	0.14
			SE	56.30	9.91	--
			$E_1^{'}$	0.45	0.51	0.47
			$d_1^{'}$	0.68	0.72	0.69
	North Bosque River at Valley Mills, TX (08095200)	01/01/1980 to 12/31/1993	E	0.11	0.32	0.71
			SE	58.90	10.89	--
			$E_1^{'}$	0.47	0.51	0.56
			$d_1^{'}$	0.70	0.73	0.76

[1] E is Nash-Sutcliff Efficiency; SE is Standard Error; $E_1^{'}$ is the modified coefficient of efficiency; and $d_1^{'}$ is the modified index of agreement

Table 5.11 Hydrology validation results.

Watershed	USGS gauging station Location	Validation period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
Medina River below Medina diversion lake (Segment 1903)	Medina River near Somerset, TX (08180800)	01/01/2000 to 12/31/2000	E	0.38	0.52	0.58
			SE	5.50	1.30	
			$E_1^{'}$	0.28	0.16	0.30
			$d_1^{'}$	0.66	0.65	0.67
Upper North Bosque River (Segment 1255)	North Bosque River at Hico, TX (08094800)	01/01/1994 to 12/31/1995	E	0.81	0.82	0.95
			SE	4.63	1.38	--
			$E_1^{'}$	0.60	0.57	0.76
			$d_1^{'}$	0.81	0.81	0.89

Table 5.11 Continued.

Watershed	USGS gauging station Location	Validation period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
Mud Creek (Segment 0611C)	Mud Creek near Jacksonville, TX (08034500)	01/01/1998 to 12/31/2000	E	0.00	0.11	-0.02
			SE	14.20	7.09	--
			E_1'	0.26	0.26	0.22
			d_1'	0.59	0.61	0.59
	Mud Creek at Ponta, TX (08035000)	01/01/1998 to 12/31/2000	E	-0.10	-0.19	-0.27
			SE	53.16	29.33	--
			E_1'	0.19	0.14	0.13
			d_1'	0.55	0.53	0.53

Table 5.11 Continued.

Watershed	USGS gauging station Location	Validation period	Statistics[1]	Variable		
				Daily flow	Daily baseflow	Monthly flow volume
North Bosque River	North Bosque River at Hico, TX (08094800)	01/01/1994 to 12/31/1995	E	0.81	0.82	0.95
			SE	4.63	1.38	--
			E_1'	0.60	0.57	0.76
			d_1'	0.81	0.81	0.89
	North Bosque River at Clifton, TX (08095000)	01/01/1994 to 12/31/1995	E	0.32	0.63	0.19
			SE	35.90	5.83	--
			E_1'	0.44	0.51	0.48
			d_1'	0.69	0.74	0.70
	North Bosque River at Valley Mills, TX (08095200)	01/01/1994 to 12/31/1995	E	0.42	0.68	0.73
			SE	41.56	7.47	--
			E_1'	0.50	0.55	0.57
			d_1'	0.72	0.76	0.77

[1] E is Nash-Sutcliff Efficiency; SE is Standard Error; E_1' is the modified coefficient of efficiency; and d_1' is the modified index of agreement

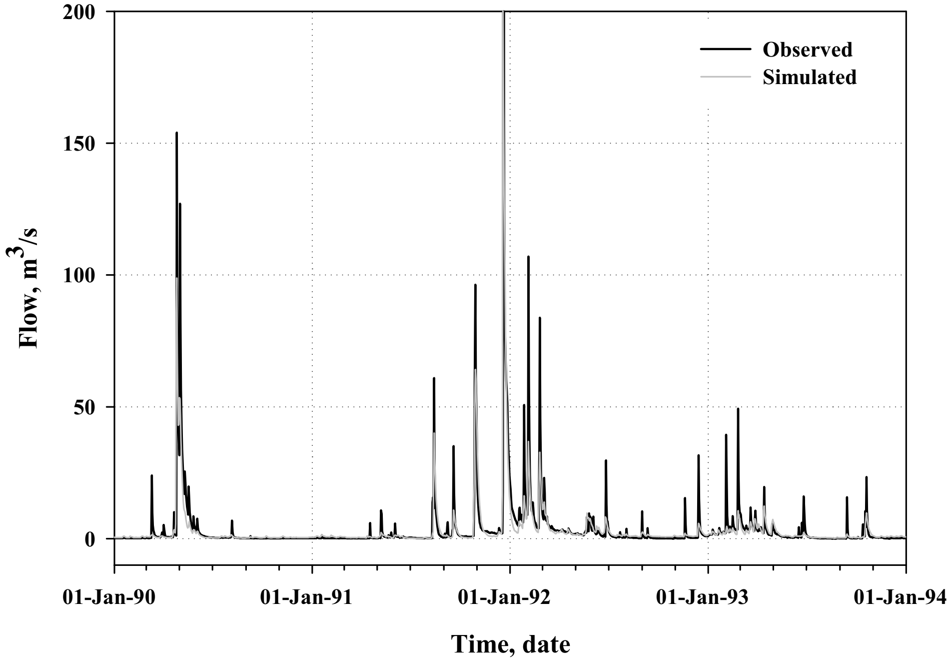

Figure 5.10 Observed and simulated flows corresponding to North Bosque River at Hico, TX during the calibration period (1 January 1980 to 31 December 1993).

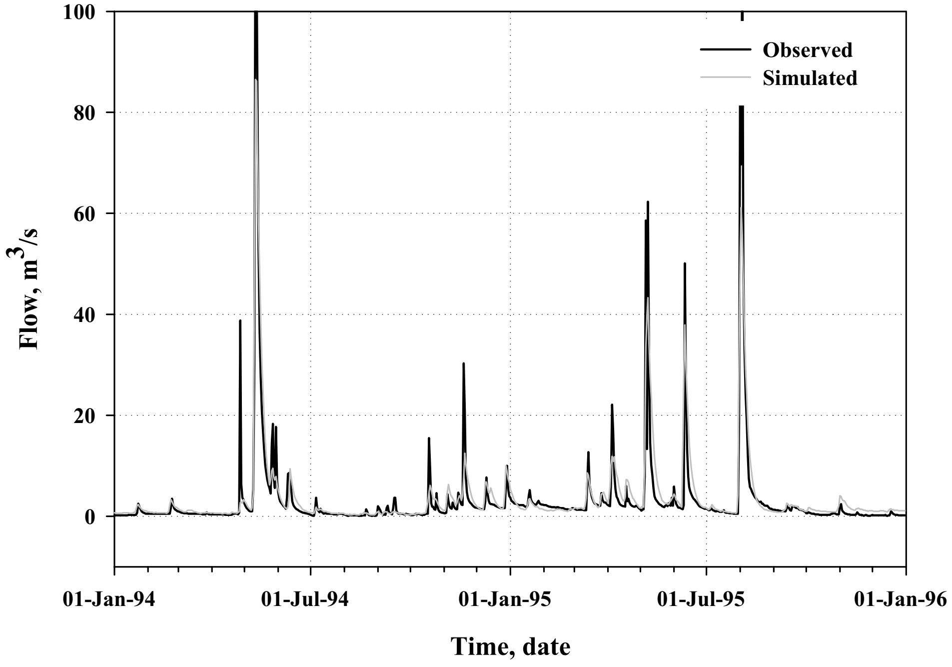

Figure 5.11 Observed and simulated flows corresponding to North Bosque River at Hico, TX during the validation period (1 January 1994 to 31 December 1995).

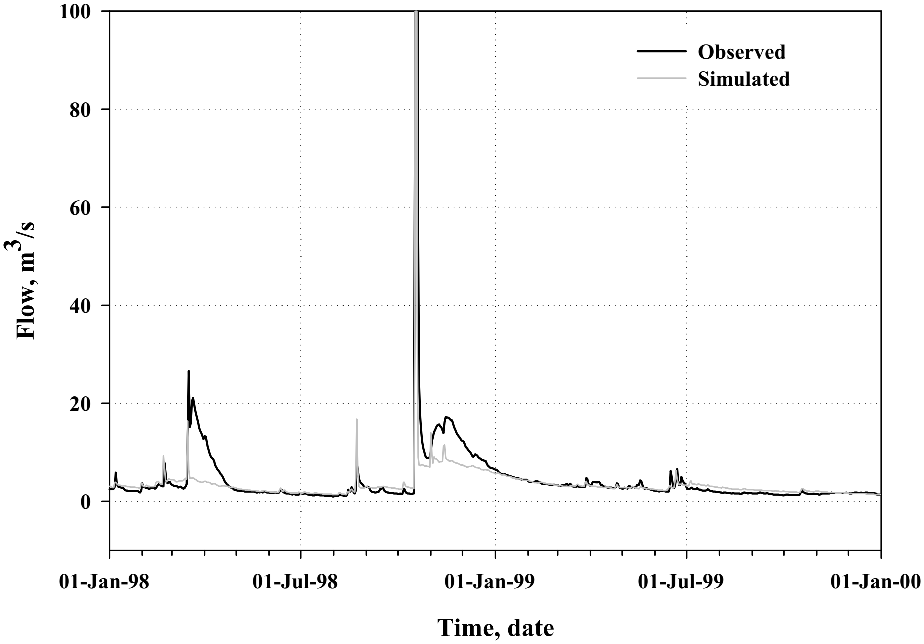

Figure 5.12 Observed and simulated flows corresponding to Medina River near Somerset, TX during the calibration period (1 January 1998 to 31 December 1999).

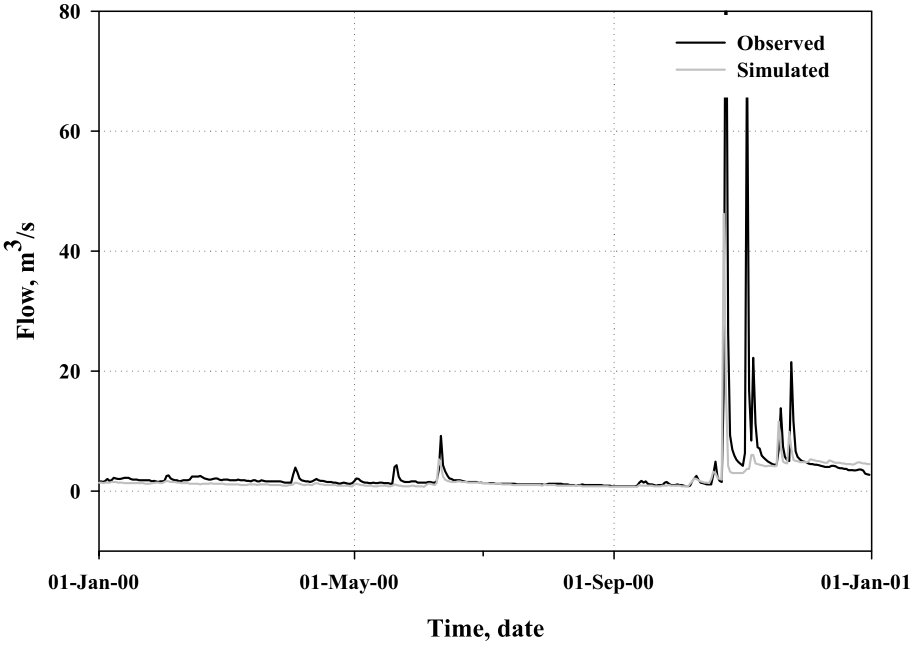

Figure 5.13 Observed and simulated flows corresponding to Medina River near Somerset, TX during the validation period (1 January 2000 to 31 December 2000).

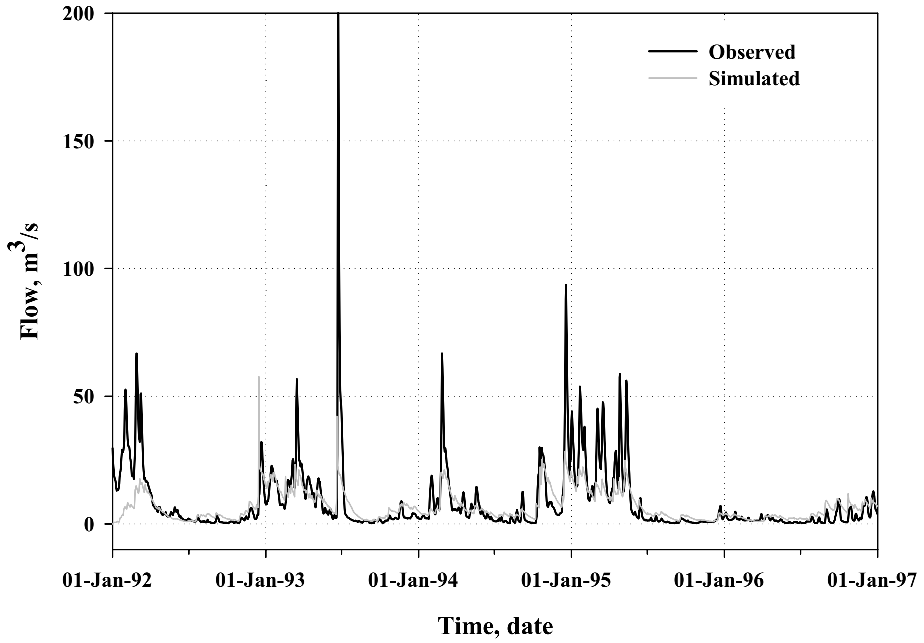

Figure 5.14 Observed and simulated flows corresponding to Mud Creek near Jacksonville, TX during the calibration period (1 January 1992 to 31 December 1996).

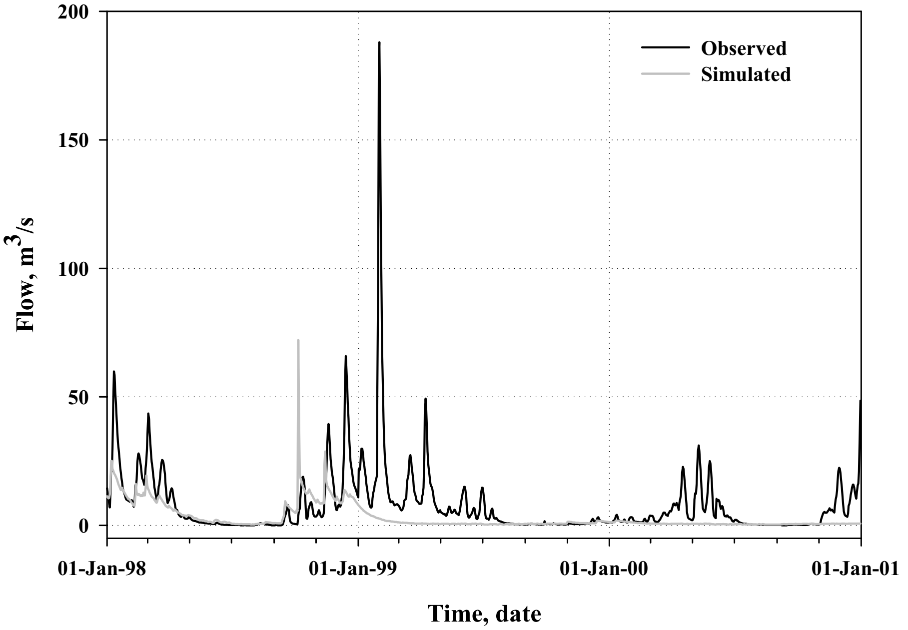

Figure 5.15 Observed and simulated flows corresponding to Mud Creek near Jacksonville, TX during the validation period (1 January 1998 to 31 December 2000).

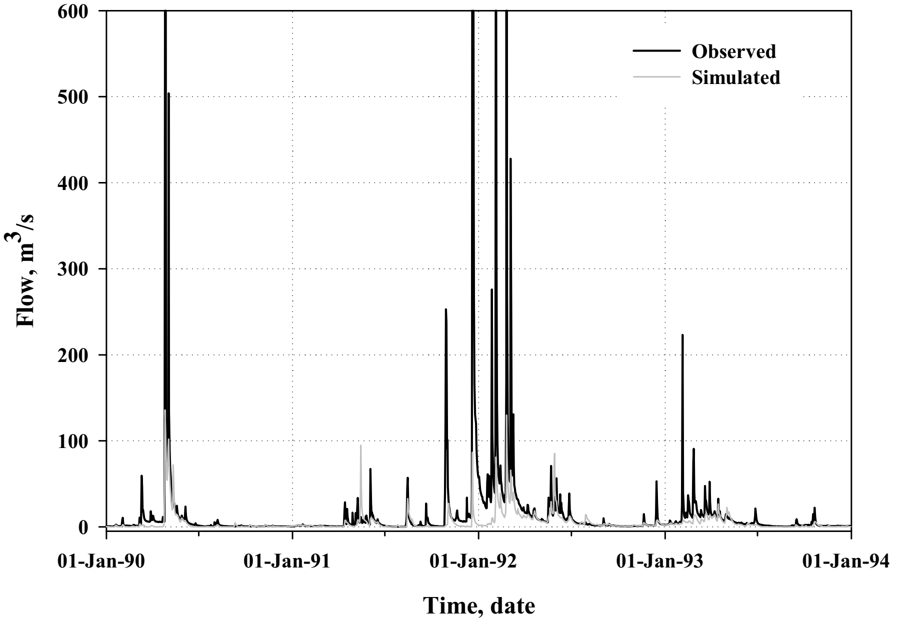

Figure 5.16 Observed and simulated flows corresponding to North Bosque River at Valley Mills, TX during the calibration period (1 January 1980 to 31 December 1993).

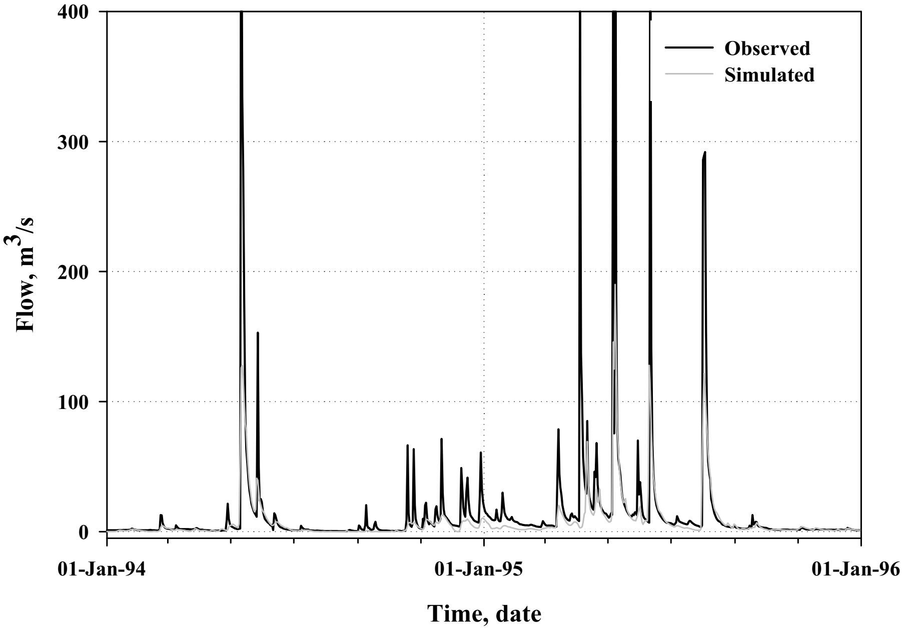

Figure 5.17 Observed and simulated flows corresponding to North Bosque River at Valley Mills, TX during the validation period (1 January 1994 to 31 December 1995).

Water Quality Calibration and Validation

Medina River and Mud Creek Watersheds

The statistics, model efficiency (*E*), modified coefficient of efficiency (E'_1), and modified index of agreement (d'_1) used to measure the goodness-of-fit between the model-predicted and the observed in-stream bacterial concentrations for the Medina River and Mud Creek watersheds are given in Table 5.12. Model run 1 corresponds to the results obtained for the input parameter set calculated based on information regarding the population of livestock and wildlife within the watershed. These parameters were modified to obtain better agreement between the model-predicted and observed in-stream bacterial concentrations corresponding to the Medina River and the subsequent model runs correspond to the results obtained during these simulations. Though, the values of the statistics were different for each watershed, the improvement in fit for each successive model run for the Medina River watershed was mirrored in the Mud Creek watershed. Figure 5.18 is a plot of the index of agreement for successive model simulations of the Mud Creek watershed against the same statistic for the Medina River. The correlation between the indices of agreement for the two watersheds was found to be very good. This suggests that the input parameter sets that yield good fit for the Medina River watershed will produce a good fit in the Mud Creek watershed as well.

Variability in the density of fecal coliform bacteria in feces, the variability in location and timing of fecal deposition, and environmental impacts on regrowth and die-off make the calibration of the fecal coliform bacteria model very difficult (MapTech, 2000). For the same reasons it will be highly unlikely to get good agreement between the model-predicted and observed in-stream bacterial concentrations when the comparisons are made at the exact time of observation. However, comparing the observed in-stream bacterial concentration with model predicted minimum and maximum values within a small window of time around the time of observation will provide a better comparison of modeling results. Figure 5.19 is a plot of model predicted minimum and maximum values in a 2-day window around the time of observation, and the observed in-stream bacterial concentration for the Medina River watershed. The model calibration can be considered reasonable if the observed values of fecal coliform fall between the minimum and maximum values in a 2-day window around the time of observation. In the current study because of the limited bacterial observations the results were not very good. The available observations were for a period between January 1985 and December 2002 and the model input parameters were based on the current watershed conditions. Also the model input parameters are constant for the entire simulation period. Regardless of these limitations, the model was able to predict fecal coliform with close agreement to the observed in-stream bacterial concentrations. Figure 5.20 shows the plot of model predicted minimum and maximum values in a 2-day window around the bacterial observation time, and the observed in-stream bacterial concentration for Mud Creek watershed using the same parameter sets calibrated to the Medina River watershed.

Table 5.12 Comparison of simulation statistics for different HSPF bacterial model runs with Medina River and Mud Creek watersheds.

Model Run	Medina River			Mud Creek		
	E	E'	d'	E	E'	d'
1	-509.94	-7.84	0.09	-48.05	-2.63	0.29
2	-5.18	-0.59	0.37	-0.61	-0.15	0.49
3	-2.95	-0.85	0.41	0.15	0.32	0.63
4	-1.44	-0.65	0.39	0.01	0.28	0.59
5	-0.94	-0.40	0.39	0.05	0.28	0.61

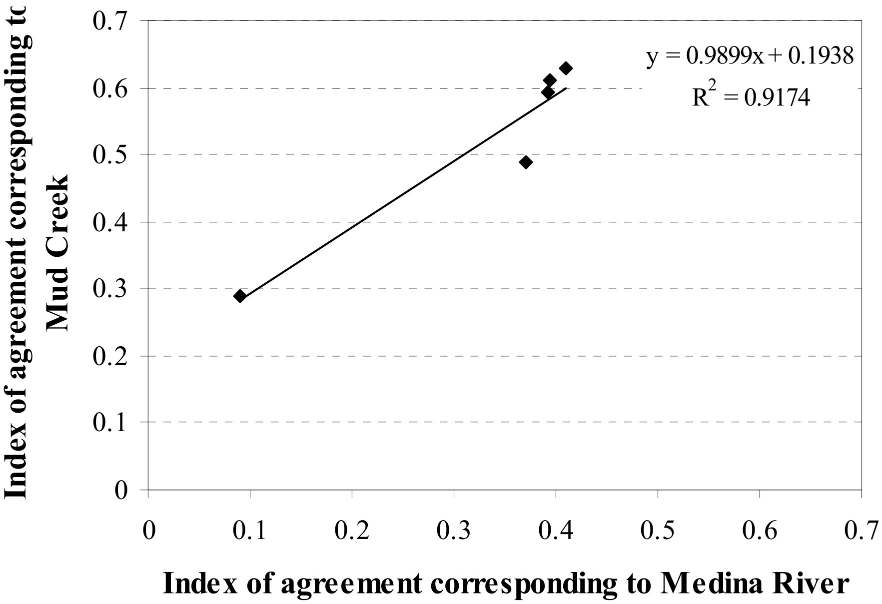

Figure 5.18 Comparison of indices of agreement for HSPF bacterial model runs with Medina River and Mud Creek watersheds using the same model input parameters.

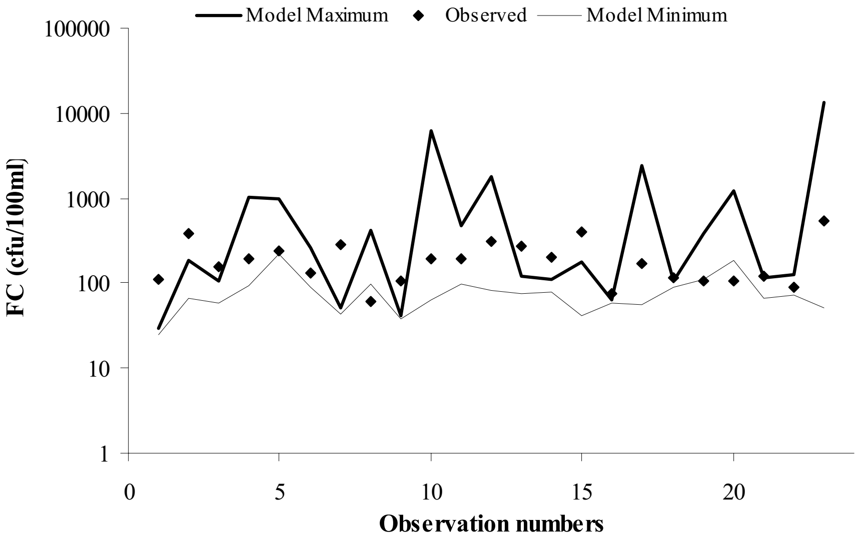

Figure 5.19 Model predicted minimum and maximum values in a 2-day window around the observation time and the observed in-stream bacterial concentration for Medina River.

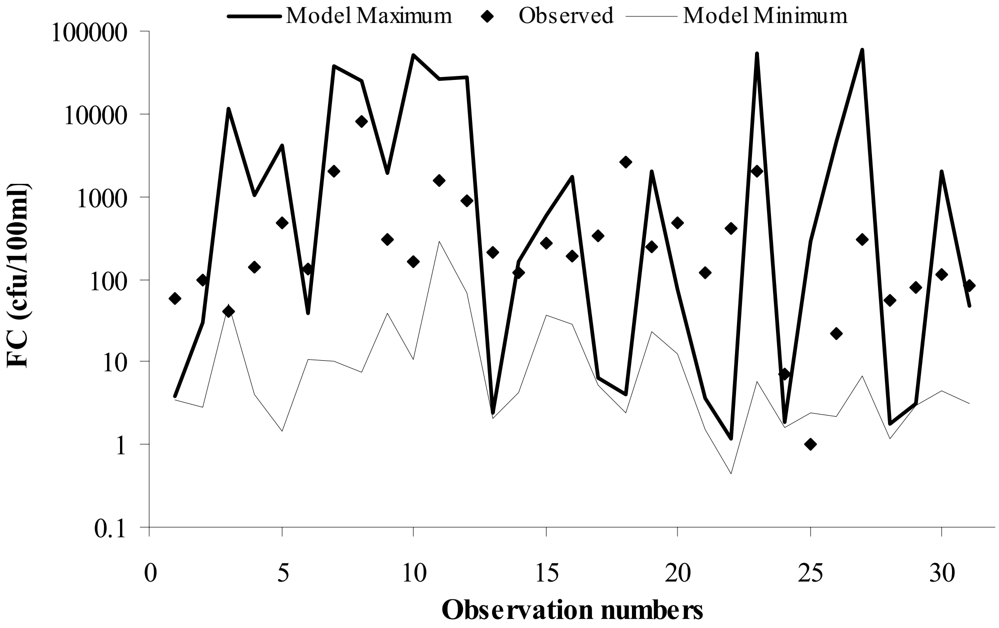

Figure 5.20 Model predicted minimum and maximum values in a 2-day window around the observation time and the observed in-stream bacterial concentration for Mud Creek.

Upper North Bosque and North Bosque Watersheds

Model simulated fecal coliform in the Upper North Bosque River and North Bosque River watersheds were compared to observed values using two different sets of model input values. First, the North Bosque River watershed was simulated using the calibrated data set from the Upper North Bosque River. Table 5.12 shows model efficiency (E), modified coefficient of efficiency (E'_1), and modified index of agreement (d'_1). Similar changes in goodness-of-fit statistics were seen in both watersheds for successive model runs. However, the goodness-of-fit between observed and simulated fecal coliform in the North Bosque River watershed was not very good. This was primarily due to the difference in the densities of livestock and wildlife between the two watersheds. The ratio of fecal coliform bacteria loading rate on the land for the North Bosque River watershed to the fecal coliform bacteria loading rate on the land for the Upper North Bosque River watershed was found to be approximately 0.6. In order to account for this difference in accumulation rate, the input parameter values corresponding to the daily accumulation and maximum accumulation over the land surface were multiplied by 0.6 and the simulations for the North Bosque River watershed were repeated. The goodness-of-fit statistics corresponding to these new parameter values are shown in Table 5.13. The new parameter set resulted in better goodness-of-fit between simulated and observed fecal coliform in the North Bosque River watershed.

Table 5.13 Comparison of simulation statistics for different HSPF bacterial model runs with Upper North Bosque River and North Bosque River watersheds.

Model Run	Upper North Bosque River			North Bosque River					
				Parameter set 1^{+}			Parameter set 2^{++}		
	E	E'_{1}	d'_{1}	E	E'_{1}	d'_{1}	E	E'_{1}	d'_{1}
1	-2.81	-0.15	0.54	-760.69	-12.58	0.10	-276.73	-7.12	0.16
2	-0.21	0.22	0.61	-7.06	-0.76	0.35	-2.85	-0.35	0.40
3	-0.10	0.25	0.62	-0.19	0.18	0.54	-0.15	0.20	0.54
4	-0.09	0.25	0.62	-0.11	0.23	0.55	-0.11	0.23	0.55

$^{+}$ Input parameter set was the same as that used with Upper North Bosque River

$^{++}$ Input parameter set was modified to incorporate the difference in the animal densities in

Based on the observed in-stream fecal coliform concentration it was reasonable to believe that the main sources of pollution within the Upper North Bosque River and North Bosque River watersheds are nonpoint sources. Hence, the primary interest of the current study was to obtain a reasonable agreement between the model-predicted and observed in-stream bacterial concentrations during the stormflow periods. The total rainfall depth over the watershed for a three day window before the time of bacterial observation was calculated and goodness-of-fit statistics were determined for several ranges of rainfall depths. The model efficiency (E), modified coefficient of efficiency (E'_1), and modified index of agreement (d'_1) for one of the calibration simulations for the Upper North Bosque River and North Bosque River watersheds using both the same parameter sets and the parameter sets modified to incorporate the change in bacterial accumulation rates are given in Table 5.14. The goodness-of-fit between model-predicted and observed in-stream bacterial concentrations was the closest when the three day rainfall values were between 0 - 0.5 inches. The goodness-of-fit in the Upper North Bosque River watershed was the worst when the three day rainfall value was 1.0 inches or more. This may be the result of calibrating the hydrology of the model for better agreement between the daily observed and model-predicted baseflows. This might have resulted in under prediction of hydrograph peaks. Since in-stream bacterial concentration predictions of the model are closely related to flow volume, bacterial concentration prediction results during very high rainfall periods tend to be of poor quality.

Table 5.14 Comparison statistics for different ranges of rainfall values corresponding to Upper North Bosque River and North Bosque River watersheds.

Rainfall Condition	Upper North Bosque River			North Bosque River					
				Parameter set 1^{+}			Parameter set 2^{++}		
	E	E'_1	d'_1	E	E'_1	d'_1	E	E'_1	d'_1
0-0.5 in.	0.18	0.37	0.67	-0.04	0.37	0.61	-0.05	0.36	0.61
0.5 - 1.0 in.	-0.04	0.07	0.48	-48.9	-3.22	0.16	-30.5	-2.49	0.19
>1.0 in.	-1074	-19	0.07	-	-	-	-	-	-
Rain	-0.26	0.19	0.59	-0.35	0.11	0.51	-0.24	0.16	0.53
No rain	-1.8	-0.41	0.52	-0.25	0	0.45	-0.24	0.03	0.48
Over all	-0.1	0.25	0.62	-0.32	0.12	0.52	-0.21	0.17	0.53

$^{+}$ Input parameter set was the same as that used with Upper North Bosque River

$^{++}$ Input parameter set was modified to incorporate the difference in the animal densities in Upper North Bosque and North Bosque River watersheds

A careful observation of Table 5.13 shows that the agreement between the simulated and observed in-stream bacterial concentrations is better during stormflow days than during baseflow days. Figures 5.21 and 5.22 show the model predicted minimum and maximum values in a 2-day window around the time of bacterial observation, and the observed in-stream bacterial concentrations for the Upper North

Bosque River and North Bosque River watersheds, respectively, when simulated for the same input parameters. It was found that about 35% of the observations were not within the range of simulated two day minimum and maximum values. The model input values could be further adjusted to obtain a better calibration for both watersheds.

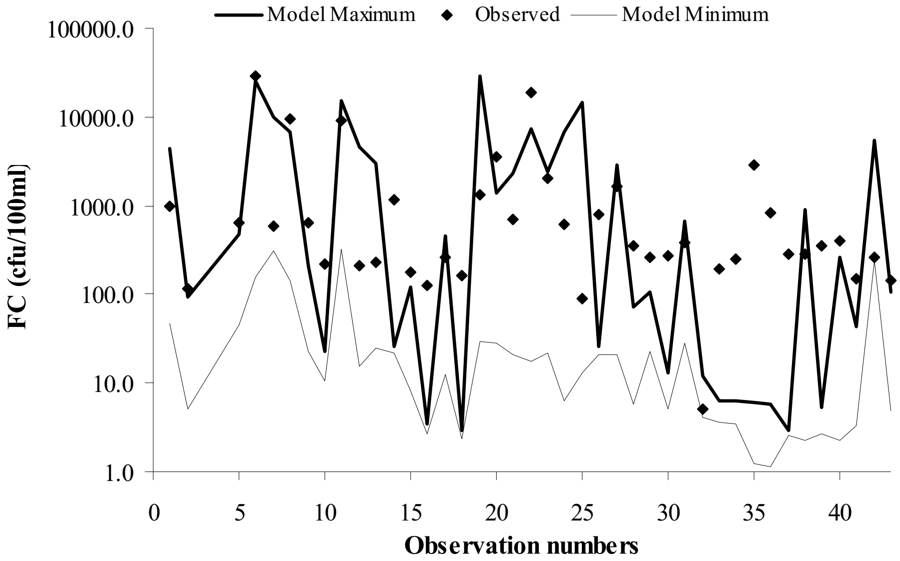

Figure 5.21 Model predicted minimum and maximum values in a 2-day window around the observation time and the observed in-stream bacterial concentration for Upper North Bosque River.

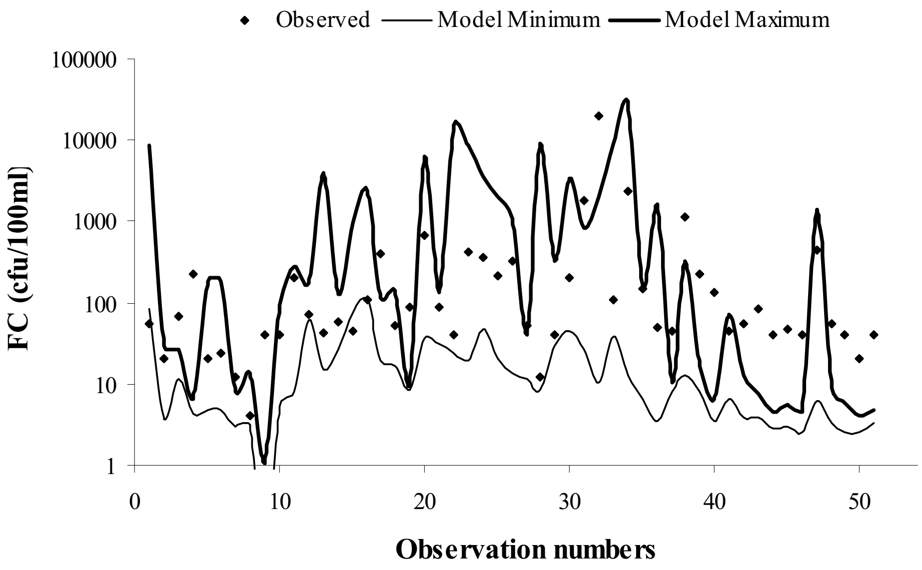

Figure 5.22 Model predicted minimum and maximum values in a 2-day window around the observation time and the observed in-stream bacterial concentration for North Bosque River.

Conclusions

The HSPF model was calibrated for model hydrology for two watersheds each in two different clusters. The HSPF model was able to predict the in-stream bacterial concentrations fairly well, though the model parameters need to be finely adjusted to get a better agreement between the observed values and the model-predicted values. The specific conclusions derived from the current study are as follows:

1. The model input parameters for a well calibrated watershed in a cluster provides very good initial input parameter values for the other watersheds in the same cluster.
2. For the same change in input parameters goodness-of-fit statistics showed the same trend for the watersheds with the same characteristics.
3. Adjusting the input parameters based on the expected density of livestock and wildlife within the watershed improved the simulation results for the validation watershed.

4. For proper calibration of the watersheds more information on specific sources, especially point sources, is required.
5. The comparison of model predictions with observed in-stream bacterial concentrations would be improved by collecting more grab samples at uniform time intervals. The calibration results can be improved considerably by intensive sample collections, thereby limiting the simulation period to a very short time period. This would greatly reduce the effect of variability of watershed characteristics on the calibration process.
6. Urban areas in the Medina River watershed tend to have high in-stream bacterial concentrations compared to other areas in the watershed. This may be due to the high presence of domestic pets and hence the high accumulation of fecal coliform bacteria in those land areas followed by higher washoff during rainfall events. Also the steady rate of high concentration of in-stream bacterial concentration may be due to the fact that the domestic pets have constant access to the waterbody.

CHAPTER VI

EFFECT OF PARAMETER UNCERTAINTY ON HSPF IN-STREAM FECAL COLIFORM CONCENTRATION PREDICTIONS

Synopsis

Salado Creek in Bexar County, Texas is one of 110 streams listed in the Clean Water Act section 303(d) list, as an impaired waterbody for its high concentration of fecal coliform bacteria. The Hydrological Simulation Program-FORTRAN (HSPF) in Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) was applied to the Salado Creek watershed to study its applicability as a prediction tool for in-stream fecal coliform bacterial concentration from both point and nonpoint sources associated with different landuse types in the watershed. The sensitivity of simulated in-stream fecal coliform concentrations to changes in model parameters associated with the bacterial simulation was evaluated using weather data obtained from the Next Generation Weather Radar (NEXRAD) system and the National Climatic Data Center (NCDC). First Order Approximation and Monte Carlo Simulation (MCS) were conducted to determine the effect of variability in the most sensitive parameters identified in the sensitivity analysis on the variability in the model predicted maximum monthly geometric mean and maximum daily mean in-stream bacterial concentrations. MCS results showed that the major contribution of the variance in maximum monthly geometric mean predictions of fecal coliform concentrations was contributed by the parameter that controls the evaporation from the lower soil zone (38.18%), parameter related to soil infiltration capacity (35.41%), and interflow inflow parameter of the pervious land (23.33%) when the NEXRAD rainfall data was used. The parameter related to soil infiltration capacity (26.09%), parameter that controls the evaporation from the active ground water zone (19.15%), interception storage capacity (11.82%),

and interflow inflow parameter of the pervious land (11.23%) contributed most to the variance in maximum daily mean predictions of fecal coliform concentrations when NEXRAD rainfall data was used. The major contribution of the variance in maximum monthly geometric mean predictions of fecal coliform concentrations was contributed by lower zone nominal storage (41.76%), parameter that controls the evaporation from the lower soil zone (32.81%) and parameter related to soil infiltration capacity (12.48%) when the NCDC rainfall data was used. Manning's n for the flow on impervious land (14.09%), parameter related to soil infiltration capacity (16.91%), the parameter that controls the evaporation from the lower soil zone (11.92%), average in-stream water temperature (11.18%), and parameter that controls the evaporation from baseflow (10.88%) contributed most to the variance in maximum daily mean predictions of fecal coliform concentrations when the rainfall data was from the NCDC weather station.

Introduction

Salado Creek in Bexar County, Texas is one of 110 streams listed as impaired water bodies in the 2000 Clean Water Act section 303(d) list for its high concentration of fecal coliform bacteria (TNRCC, 2000). Fecal coliforms are a group of bacteria that primarily live in the lower intestines of warm-blooded animals, including humans. The US Environmental Protection Agency (EPA), under its Clean Water Action Plan of 1996, has emphasized the need for State, local and tribal authorities to carry out a watershed level study and management approach in order to address the issues of nonpoint source runoff and pollution and restore the health of impaired waters. The restoration of water quality of the impaired streams starts with acquiring knowledge about the system, including the amount and sources of pollutant loading and the sources that are to be controlled.

There are a number of water quality models available for modeling fecal coliform. Some of the models used are Agricultural Runoff Management II: Animal Waste Version (ARM II) model (Overcash et al., 1983); the Utah State (UTAH) model (Springer et al., 1983); the MWASTE model (Moore et al., 1988); the COLI model (Walker et al., 1990); and Hydrologic Simulation Program- FORTRAN (HSPF) model (Bicknell et al., 1997). All of these models calculate the bacterial die-off using the first order exponential decay expressed as Chick's Law directly or with some modifications (Moore et al., 1988). According to Chick's law (Crane and Moore, 1986), the die-off of fecal coliform bacteria follows a first order decay rate given by the equation,

$$\frac{N_t}{N_0} = 10^{-kt} \tag{6.1}$$

where N_0 and N_t are number of coliform bacteria at time 0 and at time t days (colony forming unit [cfu]/100ml), respectively, and k is the first order decay rate, (d^{-1}).

The MWASTE model modifies Chick's Law by adjusting the die-off rate constant based on temperature, manure application method and soil pH. The COLI model combines the Modified Universal Soil Loss Equation (MUSLE), Chick's Law, a cell density factor and a temperature adjustment equation (Walker et al., 1990, DeGuise et al., 1999) to determine the amount of bacteria lost from land-applied waste. The HSPF model allows both accumulated and fresh manure to contribute to bacterial losses from land-applied manure.

Although the usefulness of water quality models as an aid in stream water quality restoration is unquestionable, in recent years EPA started emphasizing the importance of incorporating variability and uncertainty in the modeling process (USEPA, 1997b). They pointed out that probability analysis techniques like Monte Carlo analysis are useful tools in adequately quantifying these variabilities and uncertainties (Chang, 1999).

In most watershed-level assessment and management activities the only thing we are sure of is that we are “in doubt” (Hession et al., 1996a, 1996b, 1996c). There are many uncertainties inherent in such activities including monitoring/measurement error, model error, model input parameter errors, spatial variability, errors in spatial data layers within a Geographic Information System (GIS), the effects of aggregation of spatial data when modeling watersheds, and temporal variability. These different errors or uncertainties may or may not be additive.

Many types of uncertainties have been identified in the literature utilizing various taxonomic breakdowns (Morgan and Henrion, 1998). Haan (1989), in discussing uncertainty in hydrologic models, classified uncertainty into three categories: the inherent variability in natural processes, model uncertainty, and parameter uncertainty. Similarly, Suter et al. (1987) and Suter (1999) proposed a taxonomy of uncertainty identifying three sources of analytical uncertainty: 1) errors resulting from our conceptualizations of the world (model error), 2) stochasticity in the natural world, and 3) uncertainties in measuring model parameters (parameter error).

Model errors are errors resulting from faulty conceptualizations of the world (Suter et al., 1987). Model errors result from (1) using a fewer number of variables to represent a larger number of complex phenomena (2) choosing incorrect functional forms for interactions among variables, and (3) setting inappropriate boundaries for the components of the world to be included in the model.

Stochastic variability is the unexplained random variability of the natural environment (Suter et al., 1987; Haan, 1989). This inherent variability in the natural processes can be either variability in space (spatial variability) and/or variability in time (temporal variability). Spatial and temporal variability can be generally observed with environmental factors such as rainfall, temperature, and streamflow.

The current study was aimed at analyzing the HSPF model to determine the effect of uncertainty in sensitive model parameters on the model predicted maximum monthly geometric mean and maximum daily mean in-stream bacterial concentrations, using weather data from both a single weather station and distributed weather data available from Next Generation Weather Radar (NEXRAD) system of National Weather Service.

Methodology

Overview

The study was conducted using the tools available in the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) system framework. The methodology was as follows:

- Calibrate the HSPF hydrology model.
- Perform a sensitivity analysis and select the parameters that are most sensitive to in-stream fecal coliform concentrations.
- Estimate the probability density function (pdf) for each sensitive parameter.
- Perform uncertainty analysis using the First Order Approximation (FOA) technique to determine the fraction of the variance of model-predicted in-stream fecal coliform concentrations attributed to each of the most sensitive model parameters.
- Perform uncertainty analysis using Monte Carlo Simulation (MCS) and identify the probability distributions of the objective function.

The recreational water quality standard of Texas states that the monthly geometric mean based on a minimum of five observations should not exceed 200 colony forming unit (cfu) /100ml for primary contact recreation and 2000 cfu/100ml for secondary contact recreation. However, no samples may exceed 400 cfu/100ml for primary contact recreation and 4000 cfu/100ml for secondary contact recreation (USEPA, 2003). Therefore, the objective functions evaluated were the maximum of mean daily in-stream fecal coliform concentrations and the maximum of monthly geometric mean of the in-stream fecal coliform concentrations during the simulation period.

Model Description

BASINS was developed by the EPA's Office of Water to support environmental and ecological studies in a watershed context (USEPA, 2001a). BASINS works with a Geographic Information System (GIS) framework and is composed of: (1) national databases (2) assessment tools (3) watershed delineation tool (4) classification utilities (5) characterization reports (6) an in-stream water quality model, *QUAL2E* (7) watershed loading and transport models, *HSPF* and Soil and Water Assessment Tool, (*SWAT*); and (9) a simplified GIS based model, *PLOAD* that estimates annual average nonpoint loads.

HSPF is a continuous hydrologic modeling software that can be used to simulate a comprehensive range of hydrologic and water quality processes. In HSPF, modules are divided into pervious land, impervious land, and reaches, *PERLND, IMPLND,* and *RCHRES*, respectively. Each land segment is considered as a lumped catchment and simulations are based on an hourly time step.

The hydrologic simulation in HSPF is based on a mass balance approach. On pervious land surfaces the amount of precipitation is divided into the following components; direct runoff, direct evaporation from the land surface, surface storage followed by evaporation, surface storage followed by interflow, and infiltration to the subsurface zone. The subsurface zone is considered to be divided further into upper zone, lower zone and deep groundwater zone. Water reaching the subsurface layer either remains in storage, percolates to the zone below the current zone or evaporates.

Water that percolates deep below the groundwater zone is considered to be lost from the system. The amount of precipitation over an impervious land surface contributes to overland flow, evaporates directly, or remains in surface detention storage. The surface runoff calculation is based on Chezy-Manning's equation and an empirical relationship between the outflow depth and detention storage. The evapotranspiration (ET) is calculated based on the user supplied potential ET and moisture available in surface and subsurface zones to meet the potential ET demand.

The water quality simulation in HSPF assumes constituents are associated with one or more of the following; sediment, overland flow, interflow, and groundwater flow. Most frequently, fecal coliform is assumed to be flow associated. Flow associated constituents are assumed to be accumulated on the land surface until the occurrence of a rainfall event. The contribution of fecal coliform from the land surface to the stream is calculated using the following equation:

$$SOQO = SQO\left[1.0 - e^{\left(SURO \times WSFAC\right)}\right] \tag{6.2}$$

where, SOQO is the washoff of fecal coliform from the land surface (cfu. ac^{-1} d^{-1}), SQO is the storage of fecal coliform on the surface (cfu. ac^{-1}), SURO is the surface outflow of water (in. ac^{-1}), and WSFAC is the susceptibility of the quality constituent to washoff (in.$^{-1}$). The amount of storage on the land surface depends on the accumulation rate of bacteria and the maximum limit of storage depending upon the landuse type. The contribution of fecal coliform through interflow or groundwater flow is calculated using constant concentrations in these flows. The in-stream fecal coliform dynamics is calculated based on equation 6.1.

Study Area

The study area lies between latitude 29.735 N and 29.147 N and longitude 98.649W and 98.221 W (Figure 6.1). Salado Creek is one of the major tributaries of the San Antonio River with a total length of approximately 51.5 km. The creek originates in the north central region of the San Antonio River Basin, and flows toward the eastern region of the San Antonio metropolitan area and joins the San Antonio River at the south end. The total area of the delineated watershed is 498.39 km^2.

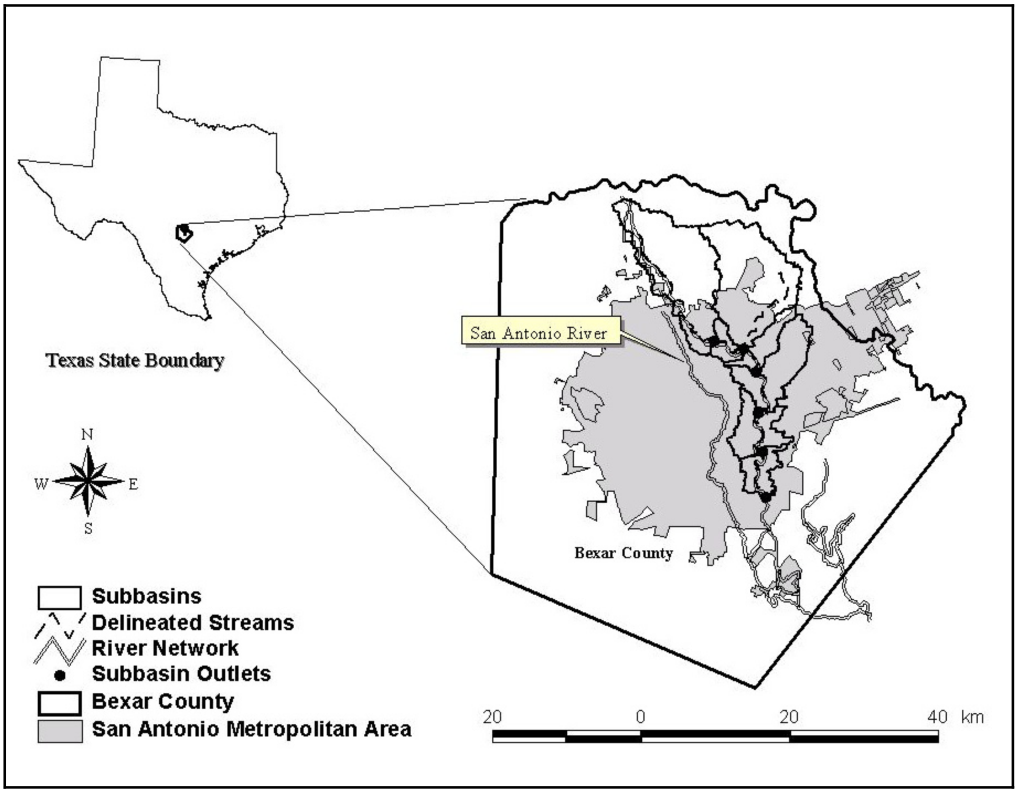

Figure 6.1 Salado Creek Watershed.

Description of Data

The Salado Creek Watershed is located within the Hydrologic Cataloging Boundary (HUC) 12100301. The land use distribution was determined using the National Land Cover Data (NLCD) obtained from the US Geological Survey (USGS).

The NLCD data were developed from the Landsat satellite Thematic Mapper data acquired by the Multi-resolution Land Characterization (MRLC) Consortium with a resolution of 30m. Soil data was the State Soil Geographic (STATSGO) available with the EPA's BASINS dataset. The STATSGO data has a resolution of 250m at 1:250,000 scale.

A Digital Elevation Model (DEM) at 1:24,000 scale with a spatial resolution of 30m was obtained from Texas Natural Resources Information System (TNRIS). These are grids covering a full 7.5-minute USGS quadrangle. The DEMs for the entire Bexar County were obtained from TNRIS and mosaiced together for the current study.

The detailed stream network for the HUC was obtained from National Hydrography Dataset (NHD) using the NHD download tool available with BASINS 3.0. NHD is based upon the content of USGS Digital Line Graph (DLG) hydrography data integrated with reach-related information from the EPA Reach File Version 3 (RF3). It is a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells. NHD is an improvement of DLG and RF3 without replacing them. It is based on 1:100,000-scale data.

The Salado Creek main watershed was sub-divided into hydrologically connected sub-watersheds using the DEMs and the Automatic Watershed Delineation tool available with BASINS 3.0. The NHD stream network for HUC 12100301 was used to obtain the correct path of the streams.

In general, the rainfall data required for model was obtained from the network of raingages available at the National Climatic Data Center (NCDC). However, there was a single weather station close to the study area. The spatial variability of rainfall cannot be accounted for by a single weather station for the entire study area. Another source of rainfall data with a higher spatial resolution was the data from the weather radars operated by the National Weather Service (NWS). The Next Generation Weather Radar (NEXRAD) maintained by NWS provides a better spatial resolution compared to the use of single weather station for the entire watershed. NEXRAD is a Doppler radar known as the Weather Surveillance Radar-1988 Doppler (WSR-88D). NEXRAD provides precipitation estimates for 4 km x 4 km grids defined in the Hydrologic Rainfall Analysis Project [HRAP] (Jayakrishnan, 2001). Figure 6.2 shows the locations of HRAP grids (identified by 6 digit number). For each subbasin the area weighted daily rainfall was calculated from the NEXRAD data and used as the input to the HSPF model.

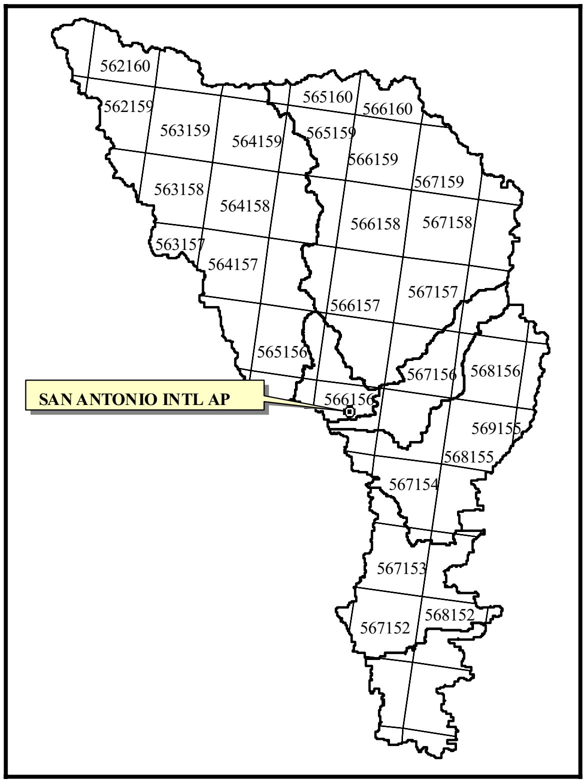

Figure 6.2 Location of NEXRAD grids and the San Antonio International Airport weather station.

The NCDC weather station within the Salado Creek watershed is located at the San Antonio International Airport (Figure 6.2). Other climatic data required for HSPF were lake evaporation and potential evapotranspiration. These data were obtained from the above NCDC weather station.

Historical daily mean stream flow data for USGS gauging station located at Salado Creek at Loop 13 at San Antonio, Texas (station number:08178800) was obtained from the USGS web site (USGS, 2001) for the simulation period. USGS gauging station 08178800 is located at the outlet of the main watershed. Hydrology calibration was carried out by statistically comparing the daily mean measured streamflow data at 08178800 with the model predicted streamflow at the same location. Figure 6.3 shows the location of the USGS gage station, the locations of the water quality stations, and the landuse distribution within the study area.

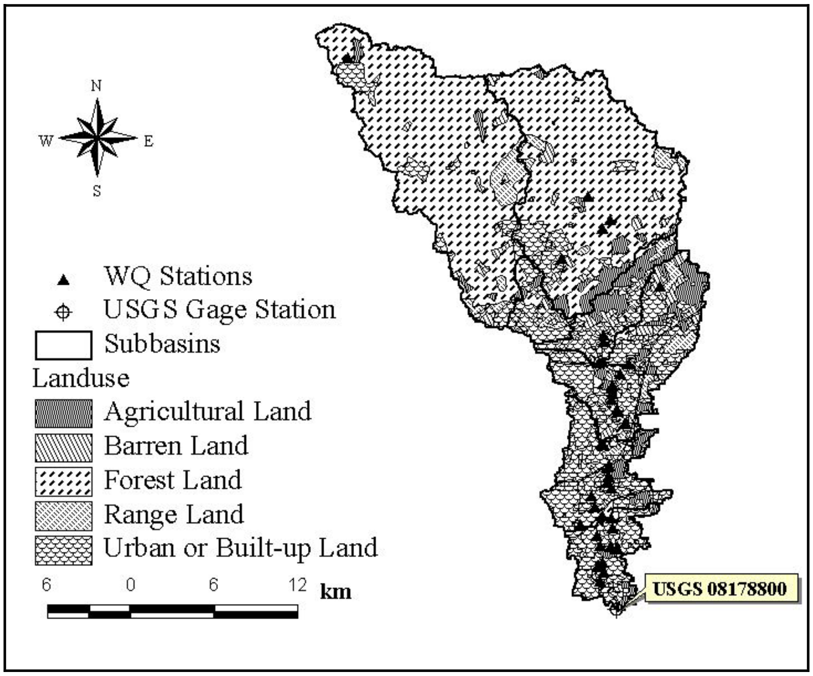

Figure 6.3 Landuse classification of the study area and locations of USGS gauging station and water quality observation stations.

Model Calibration for Hydrology

The HSPF model hydrology calibration was carried out using a model-independent parameter estimator program, PEST (Doherty, 2001a; Doherty, 2002) together with a model-independent time-series processor, TSPROC (Doherty, 2001b). PEST achieves model calibration in fewer model runs by minimizing the objective function using a robust variant of the Gauss-Marquardt Levenberg method (Doherty and Johnston, 2003). The objective function is usually the sum of squares of the differences between observed and modeled flows, volumes and/or exceedence times for various flow thresholds of the flows. The HSPF model was calibrated for the watersheds by adjusting the parameters listed in Table 6.1 to obtain a good agreement between the observed and modeled daily flows and baseflows.

Table 6.1 HSPF parameters considered during model calibration, their descriptions and default values.

Parameter	Description	Unit	Default
LZSN_PERLND	Lower zone nominal storage	in.	14.1
INFILT_PERLND	Soil infiltration capacity related parameter	in. hr^{-1}	0.16
AGWRC_PERLND	Basic groundwater recession rate if KVARY is zero	d^{-1}	0.98
DEEPFR_PERLND	Fraction of groundwater inflow that will enter deep groundwater and be lost	Fraction	0.1
BASETP_PERLND	Fraction of remaining PET which can be satisfied from baseflow	Fraction	0.02
AGWETP_PERLND	Fraction of remaining PET which can be satisfied from active groundwater storage	Fraction	0
MON-UZSN_PERLND	Upper zone nominal storage, monthly varied	in.	1.128
INTFW_PERLND	Interflow inflow parameter	None	0.75
IRC_PERLND	Interflow recession parameter	d^{-1}	0.5
MON-LZETPARM_PERLND	Lower zone ET parameter, an index to the density of deep-rooted vegetation, monthly varied	None	0.1
RETSC_IMPLND	Retention storage capacity of the surface	in.	0.065

The suffix (*PERLND, IMPLND,* and RCHRES) of the parameters in the following sections indicate the HSPF module to which the parameters belong to. These parameters were selected based on the fact that they are most sensitive parameters that control the different phases of water balance (USEPA, 2000b). All of the selected parameters except RETSC belong to PERLND module of HSPF, while RETSC belongs to IMPLND module.

The relationship of parameters IRC_PERLND and AGWRC_PERLND with the HSPF hydrology outputs is highly nonlinear (Doherty, 2001b). In order to overcome this nonlinearity problem, new parameters were derived from these parameters with the following transformations and used during the parameter estimation process. The relationships between the derived parameters and the original parameters are:

$$IRCTRANS = IRC\_PERLND/(1 - IRC\_PERLND) \tag{6.3}$$

and

$$AGWRCTRANS = AGWRC\_PERLND/(1 - AGWRC\_PERLND) \tag{6.4}$$

where IRCTRANS and AGWRCTRANS are the transformations for IRC_PERLND and AGWRC_PERLND, respectively. Also for better numerical stability and to reduce any nonlinearity problem, all the parameters were log-transformed during the parameter estimation process.

The objective function for the calibration is comprised of the following:

1. The sum of the squares of the differences between observed and simulated daily flows.
2. The sum of the squares of the differences between daily baseflows calculated from observed and simulated daily flows using a digital filter within TSPROC (Doherty, 2001b).

3. The sum of the squares of the differences between monthly volumes calculated from observed and simulated daily flows.
4. The sum of the squares of the differences between exceedence probabilities for various flow thresholds corresponding to observed and simulated daily flows.

The objective of PEST run was to minimize the objective function value. The baseflow, monthly volumes, and exceedence probabilities were all calculated using TSPROC (Doherty, 2001b). Different weights were assigned to the four groups for obtaining approximately equal contributions from all the groups. Low flows were given more weight to reduce the effect of the extremely high flows on the objective function. This will guarantee a better matching during the regular flow regimes. However, the extremely high flows may be under predicted by the model. The initial values for the parameters were calculated in accordance to the BASINS Technical Note 6 (USEPA, 2000b). The bounds for each of the parameters were selected based on the literature values (USEPA, 2000b) and data layers corresponding to the soil types, and land use distribution.

The statistics used to compare the observed daily mean flow at the USGS gage station and the simulated daily mean flow at the same location were the Nash-Sutcliffe model efficiency, modified index of agreement, and the mean absolute error.

The model efficiency, E (Nash and Sutcliffe, 1970) is given as:

$$E = 1 - \frac{\sum_{i}^{n} (O_i - S_i)^2}{\sum_{i}^{n} (O_i - O')^2} \tag{6.5}$$

where O_i and S_i are the observed and simulated daily flows for the ith day respectively, O' is the mean of the observed daily flow and n is the total number of days. A positive value of E indicates an acceptable fit between the observed and simulated flows and thus the model can be considered to be a better predictor of the system than the mean of the observed data. The closer *E* is to 1, the better the agreement between the observed and simulated flow.

The modified index of agreement, *d′* (Legates and McCabe, 1999):

$$d' = 1.0 - \frac{\sum_i^n \left| O_i - S_i \right|}{\sum_i^n \left(\left| S_i - O' \right| + \left| O_i - O' \right| \right)} \tag{6.6}$$

The modified index of agreement varies from 0.0 to 1.0. The closer the value to 1.0, the better is the agreement between the model and the observations.

The mean absolute error, MAE (Weglarczyk, 1998) is given by:

$$MAE = \frac{\sum_i^n \left| (O_i - S_i) \right|}{n} \tag{6.7}$$

where the parameters are the same as defined above. The closer MAE is to zero, the better the simulation results.

Sensitivity Analysis

Sensitivity analysis determines the sensitivity of model outputs to changes in model parameters and thus helps to identify the parameters to which the model output is most sensitive. Two types of sensitivity coefficients, absolute sensitivity and relative sensitivity can be calculated (Haan, 2002). The absolute sensitivity, S is calculated as a partial derivative:

$$S = \frac{\partial O}{\partial P} \tag{6.8}$$

where, O is the output value corresponding to the parameter value P. The absolute sensitivity is the absolute change in the output for a unit change in the input value. The relative sensitivity, S_r is calculated as:

$$S_r = \frac{\partial O}{\partial P} \times \frac{P}{O} \tag{6.9}$$

Numerical methods were used to find out the sensitivities of the parameters. The derivatives were calculated as:

$$\frac{\partial O}{\partial P} = \frac{\left(O_{P+\Delta P} - O_{P-\Delta P}\right)}{2\Delta P} \tag{6.10}$$

where ΔP is the change in the input parameter values.

The relative sensitivity is the percent change in the output for a unit percent change in the input. Since the relative sensitivity is dimensionless, it can be used to compare across parameters to select those parameters that when changed cause the greatest change in model outputs. The sensitivity of simulated peak in-stream fecal coliform concentrations at the main outlet of the Salado Creek to changes in model water quality parameters was evaluated.

Uncertainty Analysis

There are various techniques available for propagating parameter uncertainty through a model. Some of the popular techniques are Monte Carlo Simulation, Latin Hypercube Sampling and First Order Approximation (Haan, 2002). These techniques vary in their conceptual approach and the effort required for computation. In Latin Hypercube Sampling (LHS), a stratified sampling approach is used (Haan, 2002). The probability distributions of each uncertain parameter are divided into non-overlapping, equi-probable intervals. Random values of the parameters are simulated such that each range is sampled just once and the model is run. Since the outputs obtained from LHS are not completely random, the precision of the results obtained from LHS may be inaccurate (Morgan and Henrion, 1998).

First-Order Approximation

In First Order Approximation (FOA) an estimate of the variance of the output, Var(O) is given by (Haan, 2002):

$$Var(O) = \sum_{i=1}^{N} S_i^2 Var(P_i) \tag{6.11}$$

where S_i is the absolute sensitivity of the model output with respect to parameter P_i.

The fraction of the total variance of the model output that can be attributed to a particular input parameter is given by:

$$F_i = \frac{S_i^2 Var(P_i)}{\sum_{i=1}^{N} S_i^2 Var(P_i)} \tag{6.12}$$

where N is the number of sensitive parameters.

Because FOA is computationally efficient and easy to apply, the method is widely used for uncertainty analysis (Tyagi and Haan, 2001). The knowledge of input parameter distributions is not required for FOA. FOA is an approximate method that can be used for many modeling applications (Haan, 2002). Tyagi and Haan (2001) listed some of the assumptions made in FOA as (1) the near linearity in functional relations, (2) small coefficients of variation of the most sensitive uncertain variables, and (3) near normal distributions for the uncertain variables. Despite the limitations, since the FOA contributes significant understanding of the processes being modeled, it has been applied in both watershed modeling and water quality studies (Prabhu, 1995; Zhang and Haan, 1996; Haan and Skaggs, 2003a, 2003b). The FOA was used for propagation of parameter uncertainty in this study because with little information on uncertain model parameters, a general understanding of which parameters will have the most effect on the variance in the model outputs can be made.

Monte Carlo Simulation

In Monte Carlo Simulation (MCS), the uncertain parameters are assigned probability distributions and based on these distributions random values are generated for each parameter (Haan, 2002). The model is run many times using the randomly generated inputs to obtain statistics and an empirical probability distribution of the model output. MCS is computationally demanding as the number of simulations required for obtaining convergence of the model output distribution can be on the order of thousands (Haan, 2002).

The contribution of a parameter to the total variance of the objective function using a Monte Carlo simulation can be calculated using the relationship (Haan, 2002):

$$F_i = \frac{r_i^2}{\sum r_i^2} \tag{6.13}$$

where r_i is the correlation between the ith parameter and the objective function.

Results and discussion

Hydrology Calibration

HSPF hydrology was calibrated using both rainfall data from a single NCDC weather station and distributed rainfall data from NEXRAD dataset. The NCDC dataset was available for a longer time period and therefore the model simulations were done for a time period of January 1, 1970 to December 31, 2000. The model was calibrated for a period from January 1, 1980 to December 31, 1993 for the above USGS gauging station 08178800 located at the main outlet of the Salado Creek watershed. The model efficiency (*E*), modified index of agreement (*d′*), and mean absolute error (*MAE*) corresponding to daily mean stream flows during the calibration period were 0.48, 0.72 and 0.83 m^3s^{-1}, respectively. The mean of the observed daily stream flows during the same time period was 1.42 m^3s^{-1}. The comparison of simulated baseflow with the observed baseflow showed a higher agreement. The values of *E* and *d′* corresponding to daily baseflow were 0.62 and 0.7, respectively. The *MAE* was not evaluated for baseflow. Figure 6.4 shows a comparison of monthly mean observed and simulated stream flows. The respective values of *E*, *d′* and *MAE* for the monthly mean values were 0.84, 0.79 and 0.5 m^3s^{-1}.

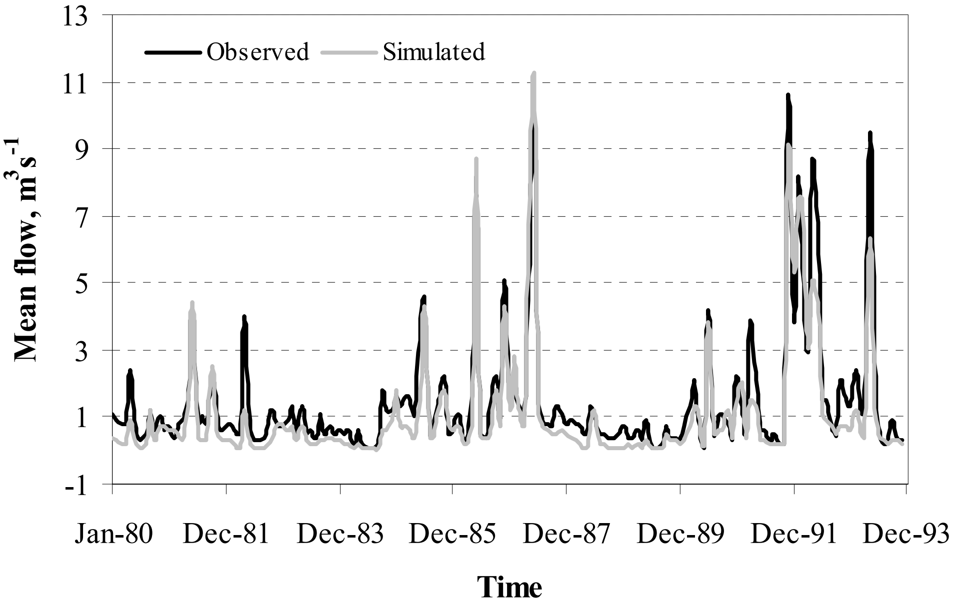

Figure 6.4 Comparison of monthly streamflows at the outlet of Salado Creek during calibration period corresponding to NCDC rainfall data.

The NEXRAD data is available starting in 1995. Hence, the model simulations using the NEXRAD data was limited for a time period of January 1, 1995 to December 31, 2000. The model hydrology was calibrated for a time period of January 1, 1996 to December 31, 2000. The model efficiency (*E*), modified index of agreement (*d′*), and mean absolute error (*MAE*) corresponding to daily mean stream flows during the calibration period were 0.51, 0.69 and 1.17 m^{3}s^{-1}, repectively. The mean of the observed daily stream flows during the same time period was 1.32 m^{3}s^{-1}. The values of *E* and *d′* corresponding to baseflow during the same period were 0.51 and 0.65, respectively. Figure 6.5 shows a comparison of monthly mean observed stream flow and simulated stream flow using NEXRAD rainfall data. The values of *E*, *d′* and *MAE* for the monthly mean values were fount to be 0.94, 0.82, and 0.55 m^{3}s^{-1}, respectively.

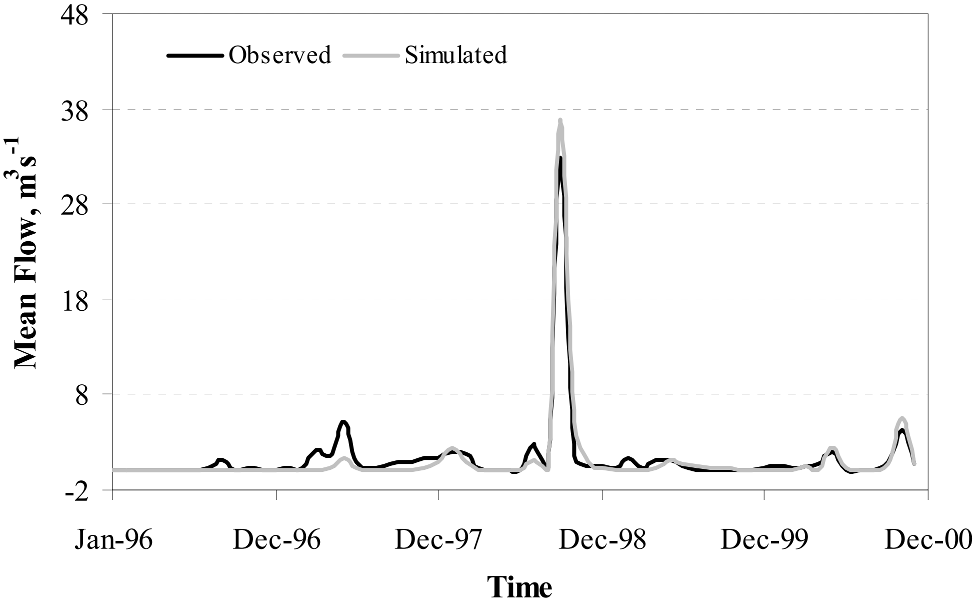

Figure 6.5 Comparison of monthly streamflows at the outlet of Salado Creek during calibration period corresponding to NEXRAD rainfall data.

Sensitivity Analysis

A sensitivity analysis of the HSPF model was done to determine which parameters, when changed, caused the greatest change in model predicted maximum monthly geometric mean concentrations and maximum daily mean concentrations of fecal coliform at the outlet of the Salado Creek. Haan (2002) describes that in general the FOA is carried out using either a 10% change or 15% change. If the assumption of a linear relationship between the input parameter and the model output is true, then the sensitivity analysis results can be assumed to be identical for a different percentage change in the input parameter. In the current study, the input parameters were varied by ±10%, ±15%, and ±20% and the average sensitivities were calculated for each input parameters. All the parameters considered during model hydrology calibration were considered for sensitivity analysis (Table 6.1). Additional parameters considered for sensitivity analysis are listed in Table 6.2.

Table 6.2 Description of model parameters considered for uncertainty analysis other than those selected for hydrology calibration.

Parameter	Description	Units
CEPSC_PERLND	Interception storage capacity	in.
ACQOP	Rate of accumulation of bacteria	cfu $ac^{-1}d^{-1}$
SQOLIM	Maximum storage of bacteria over land surface	cfu $ac^{-1}d^{-1}$
WSQOP	Rate of surface runoff which will remove 90 percent of stored bacteria	in. hr^{-1}
NSUR	Manning's n for the overland flow plane	
SQO	Initial storage of fecal coliform bacteria on the surface	cfu ac^{-1}
IOQC_PERLND	Concentration of the constituent in interflow outflow	cfu ft^{-3}
FSTDEC	First-order decay rate for bacteria	d^{-1}
AOQC_PERLND	Concentration of the constituent in active groundwater outflow	cfu ft^{-3}
THFST_RCHRES	Temperature correction coefficient for first-order decay of bacteria	None
TWAT_RCHRES	Water temperature	F

The parameters NSUR, ACQOP, SQOLIM, WSQOP, and SQO are present in both pervious land (*PERLND*) and impervious land (*IMPLND*) modules of HSPF.

Table 6.3 shows the average relative sensitivities of the HSPF parameters. The parameters ACQOP_PERLND, SQOLIM_PERLND, WSQOP_PERLND, SQO_PERLND, AOQC_PERLND, IOQC_PERLND, NSUR_PERLND, ACQOP_PERLND, and SQO_IMPLND showed negligible values of sensitivities. Hence, these parameters were not considered for the uncertainty analysis.

Table 6.3 Average relative sensitivity (S_r) values of HSPF objective functions corresponding to NCDC and NEXRAD rainfall datasets.

Parameter	Base Value	NEXRAD^{+}		NCDC^{++}	
		MaxGM*	MaxDC**	MaxGM	MaxDC
AGWETP_PERLND	2.00E-03	0.035	**3.489**	0.000	0.014
AGWRC_PERLND	0.75	**-0.942**	**-1.267**	**-3.090**	**-30.464**
BASETP_PERLND	2.00E-02	**0.167**	**2.273**	0.018	-0.084
CEPSC_PERLND	0.1	**-0.335**	**-5.527**	-0.001	**1.307**
DEEPFR_PERLND	0.6	**0.675**	**1.027**	**0.602**	**-1.870**
FSTDEC_RCHRES	0.55	-0.012	0.000	**-0.102**	-0.067
INFILT_PERLND	0.3	**-0.101**	**-5.210**	**-0.271**	**0.502**
INTFW_PERLND	8	0.000	0.042	0.000	**0.112**
IRC_PERLND	0.75	0.000	**-17.613**	**-0.826**	**-1.985**
LZETP_PERLND	0.1	**0.574**	**16.398**	-0.040	**-47.179**
LZSN_PERLND	9	**0.465**	**6.471**	0.007	**0.673**
NSUR_IMPLND	0.05	**0.239**	**25.149**	**0.566**	**-0.950**
RETSC_IMPLND	0.4	**-12.017**	**0.237**	**0.457**	**0.551**
SQOLIM_IMPLND	9.79E+12	**0.623**	**0.889**	**0.948**	**0.768**
THFST_RCHRES	1.3	-0.020	0.000	**-0.161**	**-0.110**
TWAT_RCHRES	70.99	**-0.163**	0.000	**-1.201**	**-0.905**
WSQOP_IMPLND	1.25	**-0.599**	**-0.954**	**-0.745**	-0.069
WSQOP_PERLND	1.25	0.000	0.000	0.000	-0.007
SQOLIM_PERLND	9.79E+12	0.000	0.000	0.000	0.007
ACQOP_PERLND	5.84E+12	0.000	0.000	0.000	0.000
ACQOP_IMPLND	5.84E+12	0.002	0.050	0.002	0.090
SQO_PERLND	9.79E+12	0.000	0.000	0.000	0.000
SQO_IMPLND	9.79E+12	0.000	0.000	0.000	0.000
NSUR_PERLND	0.3	0.000	0.000	0.000	0.000
AOQC_PERLND	10000	0.000	0.000	0.000	0.000
IOQC_PERLND	10000	0.000	0.000	0.000	0.000

Note: Parameters changed by ±10%, ±15%, and ±20% and the average sensitivities were calculated. Bold values equal to or greater than |0.10|

*MaxGM is the maximum of monthly geometric mean concentrations

**MaxDC is the maximum of daily mean concentrations

$^{+}$ Base values: MaxGM = 8.21E+5 cfu/100 ml, MaxDC = 1.13E+9 cfu/100 ml

$^{++}$ Base values: MaxGM = 2.33E+7 cfu/100 ml, MaxDC = 4.22E+9 cfu/100 ml

The maximum monthly geometric mean (MaxGM) predictions of HSPF showed high sensitivities to parameters such as AGWRC_PERLND, DEEPFR_ PERLND, LZETP_PERLND, NSUR_IMPLND, RETSC_IMPLND, SQOLIM_ IMPLND, TWAT_RCHRES, and IRC_PERLND. In addition to the above parameters, LZSN_PERLND, BASETP_PERLND, AGWETP_PERLND, INFILT_PERLND, and

CEPSC_PERLND had high effect on the maximum daily in-stream bacterial concentration (MaxDC) predictions. The absolute sensitivities are listed later along with the FOA results.

First Order Approximation Results

The FOA results corresponding to rainfall data from a single NCDC weather station are given in Tables 6.4. Most of the contribution to the variance in MaxGM corresponding to NCDC weather dataset comes from the variances in AGWRC_PERLND (13.68%), and NSUR_IMPLND (83.76%). The parameters that contributed the most variance to MaxDC were LZETP_PERLND (48.08%), NSUR_IMPLND (7.8%), and AGWRC_PERLND (44.0%) corresponding to NCDC. A small portion of the variance in both MaxGM and MaxDC were attributed to other parameters. The objective functions were found to have high absolute sensitivities to the parameters that contributed the most to the variance.

The FOA results corresponding and spatially distributed rainfall data from NEXRAD are given in Table 6.5. RETSC_IMPLND (82.1%), NSUR_IMPLND (6.74%), and LZETP_PERLND (9.71%) contributed the major portion of variance in MaxGM corresponding to NEXRAD weather dataset.

Table 6.4 FOA results corresponding to NCDC.

Parameters	Base value	Variance	MaxGM			MaxDC		
			S_i	$S_i^2Var(P_i)$	%	S_i	$S_i^2Var(P_i)$	%
NSUR_IMPLND	0.05	0.0324	2.64E+08	2.26E+15	**83.76**	-8.02E+10	2.08E+20	**7.80**
AGWRC_PERLND	0.75	0.04	-9.60E+07	3.69E+14	**13.68**	-1.71E+11	1.17E+21	**44.00**
WSQOP_IMPLND	1.25	0.1849	-1.39E+07	3.56E+13	1.32	-2.34E+08	1.01E+16	0.00
TWAT_RCHRES	70.99	74.13	-3.94E+05	1.15E+13	0.43	-5.38E+07	2.14E+17	0.01
IRC_PERLND	0.75	0.0121	-2.57E+07	7.96E+12	0.30	-1.12E+10	1.51E+18	0.06
RETSC_IMPLND	0.4	0.01	2.66E+07	7.08E+12	0.26	5.81E+09	3.38E+17	0.01
LZETP_PERLND	0.1	0.0324	-9.30E+06	2.80E+12	0.10	-1.99E+11	1.28E+21	**48.08**
DEEPFR_PERLND	0.6	0.0036	2.34E+07	1.97E+12	0.07	-1.32E+10	6.23E+17	0.02
FSTDEC_RCHRES	0.55	0.0676	-4.17E+06	1.18E+12	0.04	-5.16E+08	1.80E+16	0.00
INFILT_PERLND	0.3	0.0016	-2.11E+07	7.10E+11	0.03	7.06E+09	7.97E+16	0.00
BASETP_PERLND	0.02	0.0001	2.06E+07	4.25E+10	0.00	-1.78E+10	3.15E+16	0.00
THFST_RCHRES	1.3	0.0016	-2.89E+06	1.33E+10	0.00	-3.57E+08	2.04E+14	0.00
LZSN_PERLND	9	3.96	1.79E+04	1.27E+09	0.00	3.15E+08	3.94E+17	0.01
AGWETP_PERLND	0.002	0.0001	1.68E+06	2.81E+08	0.00	2.92E+10	8.54E+16	0.00
SQOLIM_IMPLND	9.79E+12	2.34E+17	2.26E-06	1.19E+06	0.00	3.31E-04	2.56E+10	0.00
CEPSC_PERLND	0.1	5.76E-06	-3.08E+05	5.46E+05	0.00	5.52E+10	1.75E+16	0.00
INTFW_PERLND	8	0.25	-3.64E+02	3.32E+04	0.00	5.89E+07	8.67E+14	0.00
Total Variance			2.70E+15			2.67E+21		
Total Standard Deviation			5.19E+07			5.17E+10		

Table 6.5 FOA results corresponding to NEXRAD.

Parameters	Base value	Variance	MaxGM			MaxDC		
			S_i	$S_i^2Var(P_i)$	%	S_i	$S_i^2Var(P_i)$	%
RETSC-IMPLND	0.4	0.01	-2.47E+07	6.09E+12	**82.10**	6.68E+08	4.46E+15	0.00
LZETP-PERLND	0.1	0.0324	4.71E+06	7.20E+11	**9.71**	1.85E+11	1.11E+21	**9.28**
NSUR-IMPLND	0.05	0.0324	3.93E+06	5.00E+11	**6.74**	5.67E+11	1.04E+22	**87.36**
AGWRC-PERLND	0.75	0.04	-1.03E+06	4.26E+10	0.57	-1.90E+09	1.45E+17	0.00
WSQOP-IMPLND	1.25	0.1849	-3.94E+05	2.86E+10	0.39	-8.60E+08	1.37E+17	0.00
AGWETP-PERLND	0.002	0.0001	1.44E+07	2.06E+10	0.28	1.97E+12	3.86E+20	3.24
LZSN-PERLND	9	3.96	4.24E+04	7.13E+09	0.10	8.10E+08	2.60E+18	0.02
BASETP-PERLND	0.02	0.0001	6.86E+06	4.71E+09	0.06	1.28E+11	1.64E+18	0.01
DEEPFR-PERLND	0.6	0.0036	9.24E+05	3.07E+09	0.04	1.93E+09	1.34E+16	0.00
TWAT-RCHRES	70.99	74.13	-1.89E+03	2.64E+08	0.00	0.0	0.0	0.00
INFILT-PERLND	0.3	0.0016	-2.77E+05	1.23E+08	0.00	-1.96E+10	6.13E+17	0.01
CEPSC-PERLND	0.1	5.76E-06	-2.75E+06	4.36E+07	0.00	-6.23E+10	2.23E+16	0.00
FSTDEC-RCHRES	0.55	0.0676	-1.82E+04	2.25E+07	0.00	0.0	0.0	0.00
THFST-RCHRES	1.3	0.0016	-1.26E+04	2.56E+05	0.00	0.0	0.0	0.00
SQOLIM-IMPLND	9.79E+12	2.34E+17	5.22E-08	6.39E+02	0.00	1.02E-04	2.45E+09	0.00
INTFW-PERLND	8	0.25	7.40E+00	1.37E+01	0.00	5.93E+06	8.78E+12	0.00
IRC-PERLND	0.75	0.0121	-2.98E-01	1.07E-03	0.00	-2.65E+10	8.47E+18	0.07
Total Variance				7.41E+12			1.19E+22	
Total Standard Deviation				2.72E+06			1.09E+11	

NSUR_IMPLND (87.36%) and LZETP_PERLND (9.28%) contributed the most to the variance in MaxDC when the rainfall data was from NEXRAD. Similar to NCDC dataset, the high absolute sensitivities of the objective functions to the parameter resulted in the high contribution of variance from these parameters.

The difference in the rainfall pattern of NCDC dataset and NEXRAD dataset can be attributed to the difference in sensitivities of the objective functions to different input parameters and hence the contribution from the input parameters to the variances of the objective functions. With the NCDC data, the rainfall pattern over the watershed area was assumed to be uniform, while the NEXRAD data was distributed at the sub-watershed level.

Number of MCS Runs

To calculate the total number of simulations required to make sure that the MCS produces representative results, thirty initial Monte Carlo simulations were run and sample means and standard deviations of the objective functions were calculated. The number of required simulation runs was calculated based on the equation (Morgan and Henrion, 1998):

$$m > \left(\frac{2cs}{w} \right)^2 \tag{6.14}$$

where m is the number of simulations required, c is the deviation for the unit normal enclosing probability α, s is the sample standard deviation, and w is the not-to-exceed width of confidence interval. The value of *w* was set based on the sample means and standard deviation. The mean and standard deviation of maximum monthly geometric mean in-stream bacterial concentration for 30 simulations were 9343 and 7412 cfu/100ml, respectively. Setting *w*=1000, showed that the minimum number of simulations required for 95% confidence interval is 879. The mean and standard deviation of the maximum daily mean in-stream bacterial concentration for 30 simulations were 2.26E+9 and 1.1E+10, respectively. Setting *w*=1.0E+9 (plus or minus a twentieth of the sample standard deviation), showed that the minimum number of simulations required for 95% confidence interval is 1934. However the numbers corresponding to 99% confidence interval for the same values of *w* were 1463 and 3218, respectively for maximum monthly geometric mean in-stream bacterial concentration and maximum daily mean in-stream bacterial concentration. Based on these results, the number of model runs was selected as 4000 for the current study.

Probability Distribution Functions of Input Parameters

The historical values of HSPF parameters listed in Table 6.3 with the exception of SQOLIM_IMPLND, THFST_RCHRES, WSQOP_IMPLND, FSTDEC_RCHRES, and TWAT_RCHRES, were obtained from HSPFParm (Donigian et al., 1999). These values were used in various watersheds across the United States for watershed hydrology calibrations. The HSPFParm database was developed with funding from EPA to provide starting parameter values for a new watershed model development. The database contains parameter values for model applications in over 40 watersheds in 14 states in the US, characterizing a broad variety of physical settings, land use practices and water quality constituents. The probability distribution functions of these parameters were determined based on the sample values obtained from the database.

The availability of historical values for SQOLIM_IMPLND was very limited. In the current study it was assumed that the value of SQOLIM_IMPLND is 1.6 times the value of daily accumulation rate, ACOQP_IMPLND, of bacteria on impervious land surface (USEPA, 2000a). The value of ACOQP_IMPLND depends on the landuse distribution and the number of animals present for each landuse. The parameter ACOQP_IMPLND was calculated for a 26-year period from 1975 to 2000, using the landuse distribution for Bexar County (Kreuter et al., 2001) and yearly population of livestock (USDA, 2002) and wildlife (TPWD, 2002). Based on the empirical frequency distribution corresponding to these calculated values, the probability distribution function of ACOQP was assumed to follow a lognormal distribution. The probability distribution function of TWAT_RCHRES was determined using historical values available from EPA's STORET database. The values corresponding to THFST_RCHRES, WSQOP_ IMPLND and FSTDEC_RCHRES were obtained from USEPA (1985, 2000c). The probability distribution functions of THFST_RCHRES, WSQOP_IMPLND and FSTDEC_RCHRES were assumed based on a study of 41 samples.

MCS Results

Table 6.6 shows the selected probability distributions and the statistics of the randomly generated parameter values. The generated parameter values were analyzed graphically to make sure they follow the selected probability distributions. The correlations between the parameters corresponding to randomly generated values were analyzed and it was found that the input parameters showed negligible correlation between each other. Thus the generated input parameters were found to be appropriate for using with MCS. The model was run 4000 times for a period between January 1, 1995 and December 31, 2000 using the randomly generated input parameters. Parameters that were found to have low sensitivities were kept at their base values during MCS runs.

Table 6.6 Selected distributions and parameter statistics.

Parameter	Assumed Distribution				Simulated Data	
	Distribution	Mean (μ)	Standard Deviation (σ)	CV	Mean ($\bar{x}$)	Standard Deviation (s_x)
IRC_PERLND	Beta	0.73	0.11	0.16	0.72	0.12
NSUR_IMPLND	Lognormal	0.18	0.18	1.01	0.18	0.18
INTFW_PERLND	Lognormal	0.73	0.50	0.70	0.74	0.51
INFILT_PERLND	Lognormal	0.05	0.04	0.78	0.05	0.04
LZETP_PERLND	Normal	0.29	0.18	0.63	0.29	0.18
LZSN_PERLND	Normal	5.79	1.99	0.34	5.82	1.97
CEPSC_PERLND	Exponential	0.0024	0.0024	1.01	0.0023	0.0023
AGWETP_PERLND	Exponential	0.01	0.01	1.01	0.01	0.01
DEEPFR_PERLND	Exponential	0.06	0.06	0.99	0.06	0.06
BASETP_PERLND	Exponential	0.014	0.014	0.99	0.014	0.014
TWAT_RCHRES	Normal	71.15	8.61	0.12	71.25	8.72
RETSC_IMPLND	Uniform	0.18	0.10	0.57	0.18	0.10
AGWRC_PERLND	Uniform	0.65	0.20	0.31	0.65	0.20
WSQOP_IMPLND	Uniform	1.25	0.43	0.35	1.25	0.43
THFST_RCHRES	Uniform	1.12	0.04	0.04	1.12	0.04
FSTDEC_RCHRES	Uniform	0.56	0.26	0.47	0.55	0.26
ACOQP_IMPLND	Lognormal	5.08E8	2.88E8	0.57	5.06E8	2.89E8

Table 6.7 shows the statistics describing the objective functions corresponding to NCDC and NEXRAD datasets. The probability distributions of the MCS generated objective functions were examined visually, using relative frequency histograms. The probability distributions of the MCS generated objective functions were statistically analyzed to test the goodness of fit of the normal, lognormal and exponential distributions to the simulated data using Kolmogorov-Smirnov (K-S) test (Haan, 2002). The null hypothesis (H_0) of K-S test is that the set of the empirical observations come from a known cumulative distribution. H_0 is rejected if the magnitude of the maximum deviation between the theoretical cumulative distribution function, $P_x(x)$, and the empirical cumulative density function, $S_n(x)$, is more than the critical tabulated value of the K-S statistic. $S_n(x)$ is calculated by dividing the number of observations less than or equal to x by the total number of observations. The Kolmogorov-Smirnov test was done in SAS (SAS, 1999). The results of the Kolmogorov-Smirnov test are given in

Table 6.8. It was found that none of the objective functions statistically fit the assumed normal or exponential distributions. Figures 6.6 through 6.9 show the relative frequencies and the expected lognormal probability distribution functions.

Table 6.7 MCS statistics of the simulated objective functions.

	NEXRAD		NCDC	
	MaxGM	MaxDC	MaxGM	MaxDC
Mean	1.66E+06	6.60E+09	7.51E+05	3.88E+09
Standard Error	3.95E+04	6.71E+08	8.40E+04	5.31E+08
Median	7.58E+05	6.74E+08	2.14E+05	7.03E+08
Mode	1.03E+07	1.13E+08	2.12E+05	1.79E+09
Standard Deviation	2.53E+06	4.28E+10	5.38E+06	3.40E+10
Sample Variance	6.39E+12	1.83E+21	2.89E+13	1.16E+21
Kurtosis	19.3	462.0	700.3	1585.5
Skewness	3.7	19.4	24.3	35.8
Range	2.59E+07	1.28E+12	1.92E+08	1.68E+12
Minimum	1.00E+04	3.75E+07	8.34E+03	3.85E+07
Maximum	2.60E+07	1.28E+12	1.92E+08	1.68E+12

Table 6.8 Kolmogorov-Smirnov goodness of fit test results for the simulated objective functions corresponding to NCDC and NEXRAD rainfall datasets.

Objective function	Hypothesis test for	Calculated D	Critical D $\alpha = 0.05$	Reject or do not reject H_0	Critical D $\alpha = 0.01$	Reject or do not reject H_0
NEXRAD						
MaxGM	lognormality	0.012	0.019	Do not reject	0.025	Do not reject
MaxDC	lognormality	0.04	0.019	Reject	0.025	Reject
NCDC						
MaxGM	lognormality	0.06	0.019	Reject	0.025	Reject
MaxDC	lognormality	0.05	0.019	Reject	0.025	Reject

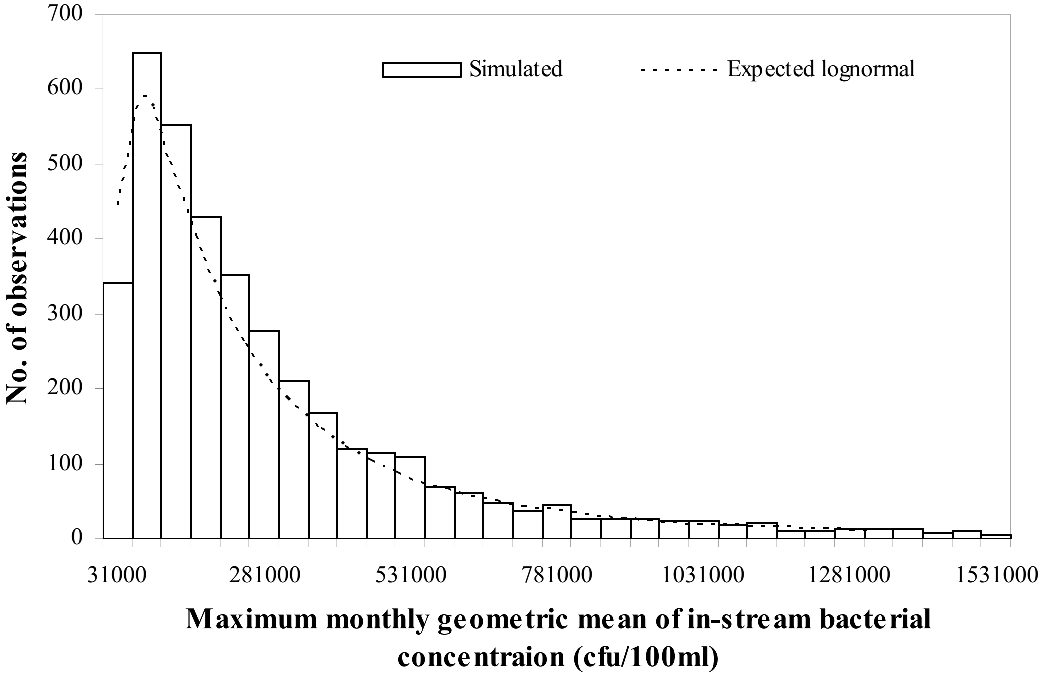

Figure 6.6 Expected lognormal distribution and simulated relative frequency histogram of maximum monthly geometric mean bacterial concentration corresponding to NCDC weather data.

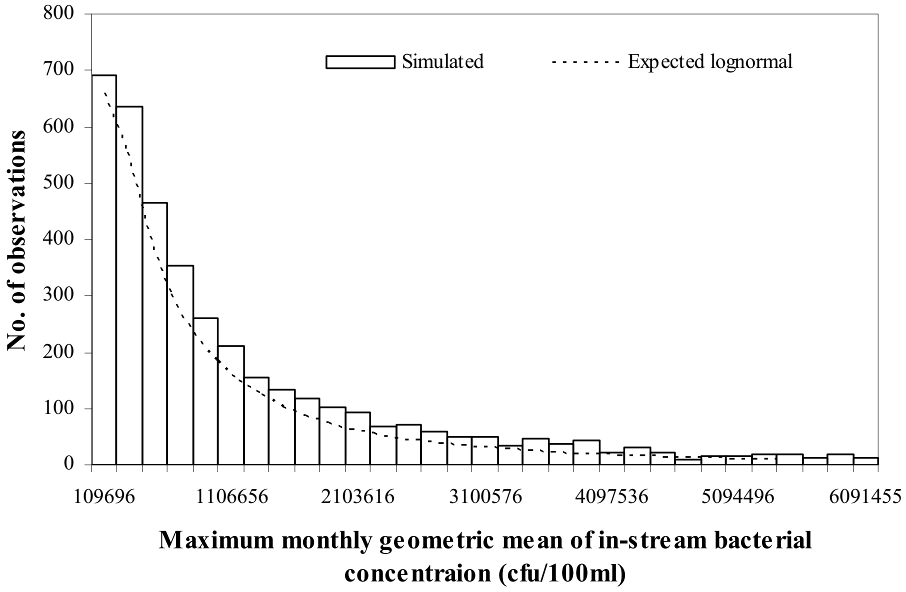

Figure 6.7 Expected lognormal distribution and simulated relative frequency histogram of maximum monthly geometric mean bacterial concentration corresponding to NEXRAD weather data.

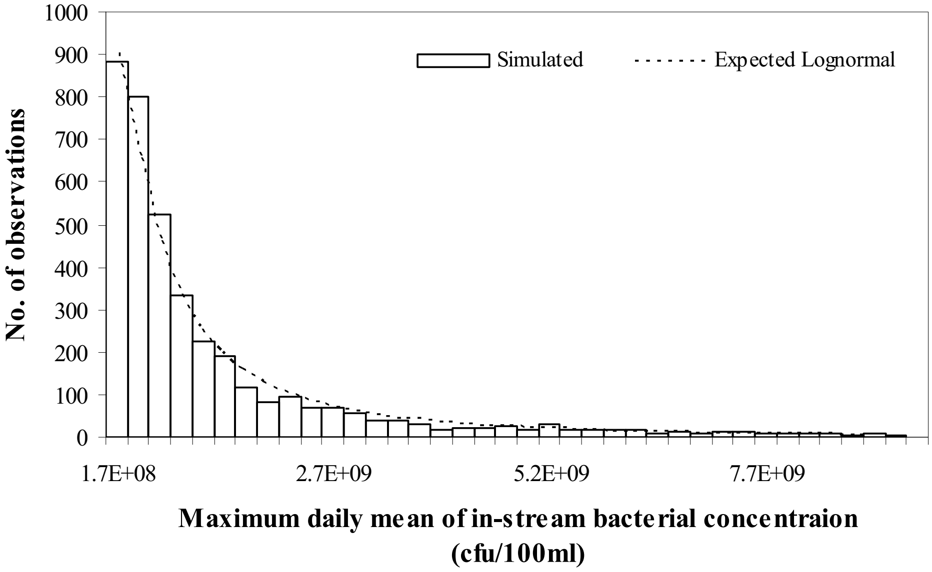

Figure 6.8 Expected lognormal distribution and simulated relative frequency histogram of maximum daily mean bacterial concentration corresponding to NCDC weather data.

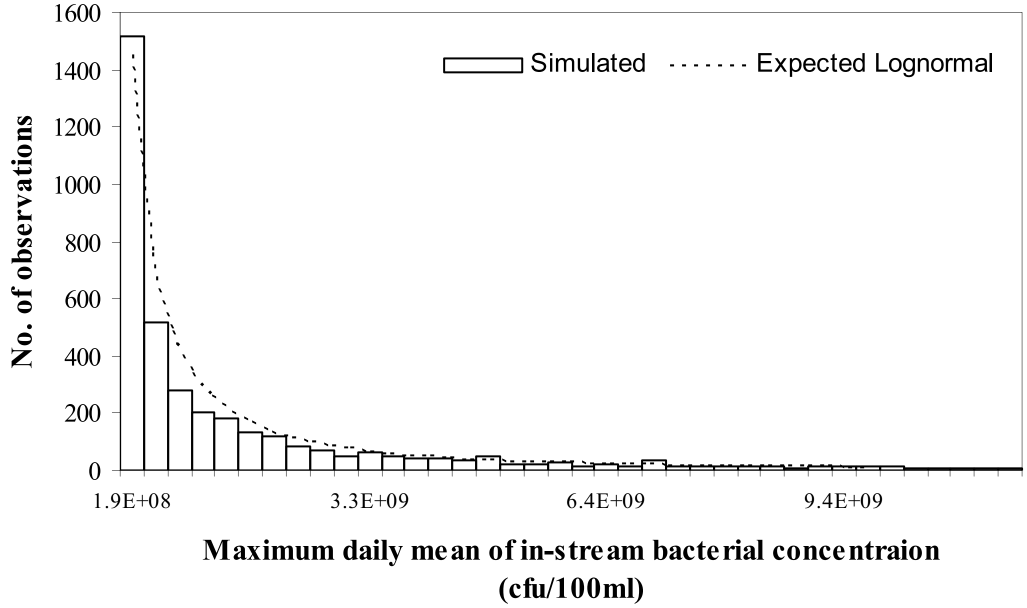

Figure 6.9 Expected lognormal distribution and simulated relative frequency histogram of maximum daily mean corresponding to NEXRAD weather data.

Table 6.9 shows the relative contributions of each input parameter to the variance in the objective functions, MaxGM and MaxDC using NEXRAD data. The parameters LZETP_PERLND (38.18%), INFILT_PERLND (35.41%), and INTFW_PERLND (23.33%), contributed more to the variance of MaxGM using NEXRAD weather data. The parameters that contributed the most to the variance of MaxDC are INFILT_PERLND (26.09%), AGWETP_PERLND (19.15%), CEPSC_PERLND (11.82%), BASETP_PERLND (5.91%), AGWRC_PERLND (6.01%), INTFW_PERLND (11.23%), and IRC_PERLND (6.22%) using NEXRAD weather data.

Table 6.9 Relative contribution of each parameter to the variance in objective functions for NEXRAD.

Parameter (i)	MaxGM			MaxDC		
	$r_{o,i}$	$r_{o,i}^2$	%	$r_{o,i}$	$r_{o,i}^2$	%
LZETP_PERLND	-0.39	0.152	**38.18**	-0.01	0.000	0.19
INFILT_PERLND	-0.38	0.141	**35.41**	-0.07	0.006	**26.09**
INTFW_PERLND	-0.31	0.093	**23.33**	-0.05	0.002	**11.23**
CEPSC_PERLND	-0.09	0.007	1.86	0.05	0.003	**11.82**
RETSC_IMPLND	0.06	0.003	0.76	0.03	0.001	4.58
NSUR_IMPLND	0.03	0.001	0.18	-0.02	0.000	1.82
TWAT_RCHRES	-0.02	0.001	0.15	0.00	0.000	0.09
AGWRC_PERLND	0.01	0.000	0.03	0.04	0.001	**6.01**
BASETP_PERLND	0.01	0.000	0.03	0.04	0.001	**5.91**
LZSN_PERLND	0.01	0.000	0.02	-0.01	0.000	0.87
FSTDEC_RCHRES	-0.01	0.000	0.02	-0.03	0.001	3.38
AGWETP_PERLND	-0.01	0.000	0.01	0.06	0.004	**19.15**
DEEPFR_PERLND	0.01	0.000	0.01	0.00	0.000	0.10
WSQOP_IMPLND	0.00	0.000	0.00	0.02	0.000	1.79
IRC_PERLND	0.00	0.000	0.00	-0.04	0.001	**6.22**
THFST_RCHRES	0.00	0.000	0.00	0.01	0.000	0.19
SQOLIM_IMPLND	0.00	0.000	0.00	-0.01	0.000	0.58
Sums		0.399	100		0.022	100

Table 6.10 shows the relative contributions of each input parameter to the variance in the objective functions, MaxGM and MaxDC for the NCDC dataset. The parameters that contributed more than 5% to the variance of MaxGM using NCDC weather data were INTFW_PERLND (8.92%), INFILT_PERLND (12.48%), LZETP_PERLND (32.81%), and LZSN_PERLND (41.76%). NSUR_IMPLND (14.09%), INFILT_PERLND (16.91%), LZETP_PERLND (11.92%), BASETP_PERLND (10.88%), TWAT_RCHRES (11.18%), RETSC_IMPLND (7.03%), THFST_RCHRES (11.3%), and SQOLIM_IMPLND (8.11%).

A careful observation of Table 6.9 and 6.10 shows that the use of a single NCDC weather station rainfall data resulted in smaller correlations between the input parameters and the model objective functions. With the NCDC data, the rainfall pattern over the watershed area was assumed to be uniform, while the NEXRAD data was distributed at the sub-watershed level. This difference in the rainfall pattern can be attributed to the difference in the contributions from the input parameters to the variances of the objective functions.

Table 6.10 Relative contribution of each parameter to the variance in objective functions for NCDC.

Parameter	MaxGM			MaxDC		
(i)	$r_{o,i}$	$r_{o,i}^2$	%	$r_{o,i}$	$r_{o,i}^2$	%
LZSN_PERLND	-0.13	0.016	**41.76**	0.00	0.000	0.00
LZETP_PERLND	-0.11	0.012	**32.81**	-0.03	0.001	**11.92**
INFILT_PERLND	-0.07	0.005	**12.48**	-0.04	0.001	**16.91**
INTFW_PERLND	-0.06	0.003	**8.92**	-0.02	0.000	3.56
IRC_PERLND	0.02	0.000	0.82	-0.01	0.000	0.99
BASETP_PERLND	-0.02	0.000	0.77	0.03	0.001	**10.88**
CEPSC_PERLND	-0.02	0.000	0.65	0.01	0.000	0.68
WSQOP_IMPLND	-0.01	0.000	0.54	-0.01	0.000	1.76
SQOLIM_IMPLND	-0.01	0.000	0.32	-0.03	0.001	**8.11**
DEEPFR_PERLND	-0.01	0.000	0.30	0.00	0.000	0.09
RETSC_IMPLND	0.01	0.000	0.27	0.02	0.001	**7.03**
FSTDEC_RCHRES	-0.01	0.000	0.18	0.00	0.000	0.09
NSUR_IMPLND	0.01	0.000	0.12	-0.03	0.001	**14.09**
TWAT_RCHRES	0.00	0.000	0.04	-0.03	0.001	**11.18**
THFST_RCHRES	0.00	0.000	0.01	-0.03	0.001	**11.30**
AGWETP_PERLND	0.00	0.000	0.01	0.01	0.000	0.35
AGWRC_PERLND	0.00	0.000	0.00	0.00	0.000	0.08
Sums		0.038	100		0.009	100

The confidence intervals (CI) on the objective functions corresponding to the NEXRAD and NCDC datasets were calculated. The calculation of CI was done using the empirical probabilities calculated from the simulation results. Table 6.11 shows the CI at 80, 90 and 95% for both NEXRAD and NCDC dataset. The CI values corresponding to both NEXRAD and NCDC datasets showed similar values. A careful observation of Table 6.11 shows that NEXRAD data tends to give higher MaxGM values for all the probability levels. The confidence intervals of MaxDC corresponding to NEXRAD data were wider compared to confidence intervals of the same corresponding to NCDC weather station data. The difference between the values corresponding to NCDC and NEXRAD may be due to the effect of better spatial resolution in NEXRAD dataset compared to the single NCDC weather station data.

Table 6.11 Confidence intervals on objective functions.

	Confidence Level (%)	Confidence Intervals	
		NEXRAD	NCDC
MaxGM*	80	2.48E+05 - 2.38E+06	9.19E+04 - 5.31E+05
	90	1.31E+05 - 4.15E+06	6.00E+04 - 9.82E+05
	95	8.67E+04 - 6.45E+06	4.34E+04 - 1.81E+06
MaxDC**	80	1.40E+08 - 3.85E+09	2.72E+08 - 2.36E+09
	90	1.00E+08 - 9.45E+09	1.60E+08 - 5.24E+09
	95	7.89E+07 - 1.98E+10	1.14E+08 - 1.02E+10

[+] Mean values: NEXRAD = 1.66E+6 cfu/100 ml, NCDC = 7.51E+5 cfu/100 ml
[++] Mean values: NEXRAD = 6.60E+9 cfu/100 ml, NCDC = 3.88E+9 cfu/100 ml

Comparison of FOA and MCS Results

Table 6.12 compares the mean and standard deviation of each objective functions from both the FOA and MCS methods corresponding to NEXRAD and NCDC weather dataset. FOA tends to produce higher variance values for both objective functions when using either type of rainfall data. Careful observation of Tables 6.4, 6.5, 6.9 and 6.10 shows that for MCS the parameters that contributed the most to variance of a given objective function is the same for both types of rainfall data. FOA showed that for the same objective functions the contributions of variance from the same parameter were different when the rainfall data was changed. This may be because the MCS takes into account the full range of values for each input parameter in the form of probability density functions, while FOA results are highly dependent on the base values selected for each input parameters. From Table 6.6 it is clear that most of the sensitive parameters did not meet the FOA assumption of near normal distributions.

Also the large coefficient of variation values for many of the parameters may have resulted in incorrect FOA estimates of objective function variances. In addition, if the relationships of objective functions with the input parameters are not linear, then the results obtained from FOA will not be reliable.

Table 6.12 Means and standard deviations of FOA and MCS.

Objective Function	Weather Data	FOA		MCS	
		$\bar{x}$	s_x	$\bar{x}$	s_x
MaxGM	NEXRAD	8.21E+05	2.72E+06	1.66E+06	2.53E+06
	NCDC	2.33E+07	5.19E+07	7.51E+05	5.38E+06
MaxDC	NEXRAD	1.13E+09	1.09E+11	6.60E+09	4.28E+10
	NCDC	4.22E+09	5.17E+10	3.88E+09	3.40E+10

Conclusions

The Salado Creek Watershed in Bexar County, Texas was modeled using the HSPF model in BASINS. A sensitivity analysis was performed to determine the parameters that are most sensitive to in-stream fecal coliform concentrations. The effect of uncertainty of the sensitive parameters on the maximum monthly geometric mean of the in-stream fecal coliform concentration and maximum daily mean in-stream fecal coliform concentration were evaluated using both FOA and MCS methods using rainfall data from a single NCDC weather station and spatially distributed rainfall data from NEXRAD. The specific findings from the study include:

1. The predicted in-stream bacterial concentration at the outlet of Salado Creek watershed showed very small sensitivities to the water quality parameters corresponding to the pervious land segment. This is believed to be because of high urbanization of the watershed.
2. MCS results showed that the major contribution of the variance in maximum monthly geometric mean predictions of fecal coliform concentrations was

contributed by the parameter that controls the evaporation from the lower soil zone (38.18%), parameter related to soil infiltration capacity (35.41%), and interflow inflow parameter of the pervious land (23.33%) when the NEXRAD rainfall data was used. The parameter related to soil infiltration capacity (26.09%), parameter that controls the evaporation from the active ground water zone (19.15%), interception storage capacity (11.82%), and interflow inflow parameter of the pervious land (11.23%) contributed most to the variance in maximum daily mean predictions of fecal coliform concentrations when NEXRAD rainfall data was used.

3. The major contribution of the variance in maximum monthly geometric mean predictions of fecal coliform concentrations was contributed by lower zone nominal storage (41.76%), parameter that controls the evaporation from the lower soil zone (32.81%) and parameter related to soil infiltration capacity (12.48%) when the NCDC rainfall data was used. Manning's n for the flow on impervious land (14.09%), parameter related to soil infiltration capacity (16.91%), the parameter that controls the evaporation from the lower soil zone (11.92%), average in-stream water temperature (11.18%), and parameter that controls the evaporation from baseflow (10.88%) contributed most to the variance in maximum daily mean predictions of fecal coliform concentrations when the rainfall data was from the NCDC weather station.
4. From MCS results it was found that the variance of maximum daily mean predictions of the model was found to be affected by more input parameters than variance in the maximum monthly geometric mean predictions.
5. In the current study, the effect of variability in all sensitive parameters on the objective functions of the model was evaluated. However, the information on the parameters that controls the HSPF model hydrology can be acquired with higher accuracy. Discarding these parameters from the uncertainty analysis may provide a clear understanding on the effect of variability in those parameters that control only the water quality predictions of the model.

6. Though NEXRAD data provided better spatial resolution, the temporal resolution of the data was better for NCDC dataset. In the current study, the daily NEXRAD rainfall data was disaggregated into hourly data using WDMUtil tool with the help of a triangular distribution and hourly rainfall pattern at the NCDC station located at San Antonio. This might have resulted in hourly estimates that are incapable of providing the exact temporal distribution of the rainfall.
7. Quantifying the uncertainties of input parameters such as maximum storage of bacteria over impervious land surface, rate of surface runoff which will remove 90% of stored bacteria, first-order decay rate of bacteria, and temperature correction coefficient for first-order decay rate of bacteria was found to be difficult, since the available information was limited. This may have led to some unrealistic conclusions. Thus, the actual variability of these parameters for the study area should be studied in detail to get a better understanding of the system.
8. Though FOA is a good method to get a reasonable understanding of the parameter that contribute to the variance in model predicted maximum monthly geometric mean in-stream bacteria concentration and maximum daily mean in-stream bacteria concentration, the choice of the base value may have a significant effect on the results.
9. The coefficient of variation of most of the input parameters was found to be large. Hence, the use of FOA to quantify the effect of parameter uncertainty on the model objective functions may not be appropriate. If the relationships of objective functions with the input parameters are not linear, then the results obtained from FOA will not be reliable.

These results point out the importance of parameterization in modeling with any complex, process-based watershed model. The current study showed that most of the uncertainty in in-stream water quality predictions was accounted for by the hydrologic parameters. Improving the knowledge of these hydrologic parameters would greatly improve the in-stream bacteria predictions. We also need to have an accurate prediction of hydrology before an attempt to predict the in-stream water quality is made. This would in turn provide a reliable total maximum daily load.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Summary

The Total Maximum Daily Load (TMDL) development process for impaired waterbodies is designed to maintain surface waterbodies for safe use. One of main components of a TMDL is creating a linkage between the source of pollution and some water quality target through the use of a water quality prediction model. However, performing such modeling studies on every individual impaired waterbody would require a tremendous amount of input in terms of money and human labor. The current study was intended to determine the possibility of using a common model on a group of stream segments that share similar watershed characteristics such as designated use of the waterbody, land use distribution, density of stream network, average distance of a land of a particular use to the closest stream, household population, density of on-site sewage facilities (OSSF), bacterial loading due to the presence of different types of farm animals and wildlife, and average climatic conditions. The watershed characteristics for 110 stream segments listed for bacterial water quality impairment in Texas during the year 2000 were collected and analyzed to obtain six clusters of stream segments with similar watershed characteristics. Observed in-stream bacterial concentrations from water quality stations within the watersheds were statistically analyzed to determine whether there was a significant difference in in-stream bacterial concentrations between baseflow and stormflow periods. The primary factors that differentiated the clusters were bacterial contribution from farm animals and wildlife, density of on-site sewage facility, density of households connected to public sewers, and the land use distribution. To test the appropriateness of a group-based TMDL development approach, two watersheds each were selected from two of the watershed clusters. The Hydrological Simulation Program-FORTRAN (HSPF) model was calibrated for one watershed in each cluster and validated with the other watershed in the same cluster. The study showed that watersheds with similar watershed characteristics yielded similar model goodness-of-fits for the same model inputs. The effect of parameter uncertainty on in-stream bacterial concentration predictions by HSPF was evaluated for the Salado Creek watershed, in Bexar County. The parameters that control the HSPF model hydrology contributed the most to the variance in in-stream fecal coliform bacterial concentration predictions corresponding to a simulation period between 1 January 1995 and 31 December 2000.

Conclusions

1. There was not enough data for analysis for 11 stream segments. Based on a pooled T-test, 44 stream segments showed significantly higher bacterial concentrations during the stormflow period, 10 during the baseflow period, and 45 stream segments showed no significant difference between the two periods. It may be concluded that there was a considerable increase in in-stream bacterial concentrations for 44 stream segments due to runoff, while the effect of rainfall on the in-stream bacterial concentrations of 45 stream segments was negligible.
2. The waterbodies with high mean bacterial concentration during the baseflow period and relatively low mean bacterial concentrations during the stormflow period may indicate continuous bacterial loading from point sources such confined animal feeding operations and discharges from waste water treatment plants. The reduction in bacterial concentrations for these stream segments during the stormflow period may be the result of dilution during large rainfall events.
3. Lower concentrations of bacteria during the baseflow period may indicate the absence of continuous point sources. An increase in concentration of bacteria

during stormflow periods may be the result of an increased load of bacteria in streams from nonpoint source runoff.

4. Higher concentrations during both stormflow and baseflow periods may indicate both point and nonpoint sources of bacteria. However, in a few of these waterbodies there were no known point sources and the density of livestock and/or wildlife were high in the contributing watersheds. One reason for higher concentrations of bacteria during baseflow periods for these stream segments may be contributions from animals having ready access to streams.
5. Four stream segments had high concentrations of bacteria during both stormflow and baseflow periods, but there was no evident source within the contributing watersheds.
6. Further research is required to understand the differences in the effect of rainfall.
7. The Texas waterbodies listed for bacterial water quality violation under CWA §303(d) could be clustered into six homogenous clusters based on their watershed characteristics using the multivariate statistical techniques of factor analysis/principal component analysis, cluster analysis, and discriminant analysis.
8. The primary watershed characteristics that differentiate the clusters are bacterial contribution from farm animals and wildlife, density of OSSF, density of households connected to public sewers, and the land use distribution.
9. A few of the watersheds were found to share a border with other states or another country (Mexico). The data collection was done only within the boundaries of the state of Texas. This may have left some of the potential sources of bacteria out of the analysis. The effect of potential sources across the boundaries on the in-stream water quality in these watersheds should be studied.

10. Presence of point and nonpoint sources within the watershed boundaries was apparent for many watersheds regardless of their membership to a certain cluster.
11. The currently available information on domestic pets and migratory birds was insufficient to be incorporated into the multivariate analysis.
12. The use of GIS is found to be very useful in disaggregating the data available at County or State level to the watershed level for the analysis. However, an extensive data collection at the watershed level will greatly improve the results.
13. The model input parameters for one well calibrated watershed in a cluster provides very good initial input parameter values for the other watersheds in the same cluster.
14. Water quality model runs corresponding to the same modification in model input parameters for two watersheds in a cluster resulted in similar improvements in model goodness-of-fit statistics for both watersheds.
15. Adjusting the input parameters based on the expected density of livestock and wildlife within the watershed improved the simulation results for the second watershed.
16. The goodness-of-fit of model predictions to the observed in-stream bacterial concentration would be improved by collecting more grab samples at uniform time intervals.
17. Urban sub-basins in the Medina Creek watershed tend to have high in-stream bacterial concentrations compared to other sub-basins in the watershed. This may be the result of a large number of domestic pets. Therefore, there may be a large accumulation of fecal coliform bacteria on those sub-basins resulting in a higher volume of fecal coliform being washed off during rainfall events.
18. The predicted in-stream bacterial concentrations at the outlet of the Salado Creek watershed showed little sensitivity to the water quality parameters related to the pervious land segment. This is believed to be because of the reduction in pervious land area resulting from high urbanization of the watershed.

19. Monte Carlo Simulation (MCS) results showed that regardless of the spatial resolution of rainfall, the major contribution of the variance in maximum monthly geometric mean predictions of the model was due to uncertainty in the interflow inflow parameter of the pervious land, parameter related to soil infiltration capacity, and the parameters that control the evaporation from the lower soil zone.
20. The number of parameters that contributed to the variance in the objective functions was different for maximum monthly geometric mean and maximum daily mean. The number of uncertain parameters contributing to the variance was greater for the maximum daily mean bacterial concentration
21. In the current study, effect of variability in all parameters to which the objective functions are sensitive was evaluated. However, the information on the parameters that controls the HSPF model hydrology can be acquired with higher accuracy. Discarding these parameters from the uncertainty analysis may provide a clear understanding on the effect of variability in those parameters that control only the water quality predictions of the model.
22. Though NEXRAD rainfall data provided better spatial resolution, the temporal resolution of the rainfall was better for the NCDC dataset for the current study. The daily NEXRAD rainfall data was disaggregated into hourly data using WDMUtil tool with the help of a triangular distribution and hourly rainfall pattern at the NCDC station located at San Antonio. This may have resulted in hourly estimates that are incapable of providing the exact temporal distribution of the rainfall.
23. Quantifying the uncertainties of input parameters such as maximum storage of bacteria over impervious land surface, rate of surface runoff which will remove 90% of stored bacteria, first-order decay rate of bacteria, and temperature correction coefficient for first-order decay rate of bacteria was found to be difficult, since the available information was limited. This might have led to some unrealistic conclusions. Thus, the actual variability of these parameters

for the study area should be studied in detail to get a better understanding of the system.

24. Though FOA is a good method to understand which parameters contribute to the most to the variance in model objective functions, the choice of the mean value may have a significant effect on the results when the functional relationship between the inputs and the model outputs are not linear.
25. The coefficient of variation of most of the input parameters was found to be large. Hence, the use of FOA to quantify the effect of parameter uncertainty on the model objective functions may not be appropriate.

Recommendations for future research

The grouping of the stream segments were based on readily available dataset at the time of analysis. The data regarding the livestock (except the CAFO) and wildlife were available only on County or State level. The watershed level data was obtained by disaggregating the county or state level data based on land use pattern. The use of GIS is found to be very useful in disaggregating the data available at County level or State level into watershed level for analysis. However, disaggregating the dataset to the watershed level introduced errors in the final analysis for many watersheds. An extensive data collection at watershed level will greatly improve the results.

The data regarding the OSSF failure rates were very limited and in most cases were not reliable. Therefore the grouping process was done based on the total number of OSSF present within the watershed boundaries. The assumption of the more the number of OSSF within a watershed, the more is the contribution of bacteria from OSSF, may not be valid when other factors that affect the failure of OSSF are dominant. For example, based on experts opinion (K. Neimann, personal communication, April 2003), the contribution of bacteria from failed OSSF is evident for watershed corresponding to stream segment 2202 (Arroyo Colorado above tidal).

This was not apparent from the cluster analysis results due to lack of reliable data regarding OSSF failure rates. Hence, the clustering analysis presented here can be used as guidelines, however, local knowledge is needed to supplement the findings to make sure the segments are classified correctly in a group.

The incorporation of location specific knowledge on the application of manure on the land surface would increase the accuracy of the results. Another drawback was the lack of distributed data regarding migratory birds and domestic pets. Collection of watershed level data regarding migratory birds and domestic pets would improve the results. The effect of potential sources across state boundaries on the in-stream water quality in these watersheds should also be studied.

More information is needed on the maximum storage of bacteria over impervious land surface, rate of surface runoff which will remove 90% of stored bacteria, first-order decay rate of bacteria, and temperature correction coefficient for first-order decay rate of bacteria and conduct MCS to obtain a more accurate result.

The selection of different water quality models was based on the current knowledge of the sources and land use distribution. Hence, if the watershed characteristics for particular impaired stream segment is found to be different from that given in this report, a different water quality model may be more appropriate for the situation and that model should be selected for the modeling study.

A few of the water quality models reviewed under the current study such as CE-QUAL-RIV1, CE-QUAL-W2, EFDC and WASP were mainly developed to address the receiving water quality. Conversely, other watershed models such as SWAT and HSPF could be used to estimate the bacterial loading from land surface to the receiving waters. Watershed level load estimations may be better addressed by linking watershed and receiving water quality models. This linkage will help reduce the need for additional resources by utilizing existing models and eliminating requirements for adding new routines to receiving water quality models.

For a TMDL modeling study, local data should also be considered from the respective watersheds. "A model can be no better than the data available for parameter estimation" (Haan, 2002). Hence, for proper calibration of the watersheds more information on specific sources, especially point sources, is required.

REFERENCES

Alberto, W. D., M. L. Del Pilar, A. M. Valeria, P. S. Fabiana, H. A. Cecilia, and B. M. De Los Angeles. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba-Argentina). *Water Res.* 35(12): 2881-2894.

Arnold, J. G., P. M. Allen, R. Muttiah, and G. Bernhardt. 1995. Automated baseflow separation and recession analysis techniques. *Groundwater* 33: 1010-1018.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. 1998. Large area hydrologic modeling and assessment, Part I: model development. *J. AWRA*. 34(1):73-89.

Arnold, J. G., and P. M. Allen. 1999. Automated methods for estimating baseflow and ground water recharge from streamflow records. *J. AWRA*. 35(2): 441- 424.

AVMA. 2002. *U.S. Pet Ownership & Demographics Sourcebook*. Schaumburg, Ill.: American Veterinary Medical Association.

Barwick, R. S., D. A. Levy, G. F. Craun, M. J. Beach, and R. L. Calderon. 2000. Surveillance for waterborne-disease outbreaks-United States 1997-1998. *Morbidity and Mortality Weekly Report*. 49(SS04):1-35.

Baumgartner, D. J., W. E. Frick, and P. J. W. Roberts. 1994. *Dilution Models for Effluent Discharges*, 3rd Ed. EPA-600-R-94-086. Newport, Ore.: US Environmental Protection Agency.

Bicknell, B. R, J. C. Imhoff, J. L. Kittle Jr., A. S. Donigian Jr., and R. C. Johanson. 1997. *Hydrological Simulation Program FORTRAN, User's Manual for Version 11*. EPA/600/R-97/080, Research Triangle Park, N.C.: National Exposure Research Laboratory, US Environmental Protection Agency.

Blumberg, A. F., and G. L. Mellor. 1987. A description of a three-dimensional coastal ocean circulation model. In *Three-Dimensional Coastal Ocean Models, Coastal and Estuarine Science*, Heaps, N.S., ed. vol. 4. 1-19. Washington D.C.: American Geophysical Union.

Box, G. E. P., and Cox, D. R. 1964. An analysis of transformations. *J. Royal Stat. Soc.*, 26: 211-253.

Bras, R. L. 1990. *Hydrology: An Introduction to Hydrologic Science*. New York, N.Y.: Addison-Wesley Publishing Co.

Brown, L.C. and T.O. Barnwell Jr. 1987. *The Enhanced Stream Water Quality Models QUAL2E and QUAL2E- UNCAS: Documentation and User Manual.* EPA/600/3-87/007. Athens, Ga.: US Environmental Protection Agency.

Carter, A. M., R. E. Pacha, G. W. Clark, and E. A. Williams. 1987. Seasonal occurrence of *Campylobacter* spp in surface water and their correlation with standard indicator bacteria. *Appl. Environ. Microbiol*. 53: 523-526.

Cattell, R. B. 1966. The scree test for the number of factors. *Multivariate Behavioral Res.*1: 245-276.

Chang, S. S. 1999. Implementing probabilistic risk assessment in USEPA Superfund program. *Human and Ecol. Risk Assessment* 5(4): 737-754.

Choi, S., W. Chu, J. Brown, S.J. Becker, V.J. Harwood, and S.C. Jiang. 2003. Application of enterococci antibiotic resistance patterns for contamination source identification at Huntington Beach, California. *Marine Pollution Bulletin.* In press.

Chow V. T., D. R. Maidment, L. W. Mays. 1988. *Applied Hydrology*. New York, N.Y.: McGraw-Hill.

Clucas, S.R. 1997. Construction as a Curriculum Organizer for Technology Education. PhD diss. Blacksburg, Va.: Virginia Polytechnic Institute and State University, Department of Teaching and Learning.

Cole, T. M., and E. M. Buchak. 1995. CE-QUAL-W2: A two-dimensional, laterally averaged, Hydrodynamic and Water Quality Model, Version 2.0, Instruction Report EL-95-, Vicksburg, Miss.: US Army Engineer Waterways Experiment Station.

Crane, S. R. and J. A. Moore. 1986. Modeling enteric bacterial die-off: A review. *Water Sci. and Soil Pollution.* 27(3/4): 411-439.

Crosetto, M., S. Terantola, and A. Sattelli. 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agri. Eco. Env.* 81:71-79.

Cruz, J. R., P. Caceres, F. Cano, J. Flores, A. Bartlett, and B. Torun. 1990. Adenovirus types 40 and 41 and rotaviruses associated with diarrhea in children from Guatemala. *J. Clin Microbiol*. 28: 1780-1784.

DeGuise, K. A., M. A. Wolfe, and T. A. Dillaha III. 1999. A foundation for modeling bacterial contamination of surface water due to nonpoint sources. ASAE Paper No. 99-2153. St. Joseph, Mich.: ASAE.

DHI. 2003a. *MIKE SWMM: User Guide MIKE SWMM 2002*. Horsholm, Denmark: Danish Hydraulic Institute.

DHI. 2003b. *MIKE BASIN 2002: A Versatile Decision Support Tool for Integrated Water Resources Management and Planning, Guide to Getting Started Tutorial*. Horsholm, Denmark: Danish Hydraulic Institute.

Doherty, J. 2001a. *PEST-ASP User's Manual*. Brisbane, Australia: Watermark Numerical Computing.

Doherty, J. 2001b. *PEST Surface Water Utilities User's Manual*. Brisbane, Australia: Watermark Numerical Computing and Moscow, Idaho: University of Idaho.

Doherty, J. 2002. *Addendum to PEST Manual for Version 6 of PEST*. Brisbane, Australia: Watermark Numerical Computing.

Doherty, J., and J. M. Johnston. 2003. Methodologies for calibration and predictive analysis of a watershed model. *J.AWRA*. 39(2): 251-267.

Donigian Jr., A. S. and W.C. Huber. 1991. Modeling of nonpoint source water quality in urban and non-urban areas. EPA/600/3/91/039. Athens, Ga.: Environmental Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Donigian Jr., A. S., J. C. Imhoff, and J. L. Kittle, Jr. 1999. *HSPFParm: An Interactive Database of HSPF Model Parameters* Version 1.0. EPA/823/R/99/004. Washington D.C.: Office of Water, US Environmental Protection Agency.

Environmental Laboratory. 1995. *CE-QUAL-RIV1: A Dynamic, One-Dimensional (Longitudinal) Water Quality Model for Streams: User's Manual*, Instruction Report EL-95-2, Vicksburg, Miss.: U.S. Army Engineer Waterways Experiment Station.

ESRI. 1999. *Getting to Know ArcView GIS: The Geographic Information System (GIS) for Everyone.* 3rd ed. Redlands, Calif.: Environmental Systems Research Institute.

Fischer, H.B., E.J. List, R.C.Y. Koh, J. Imberger, and N.H. Brooks. 1979. *Mixing in Inland and Coastal Waters*. Orlando, Fla.: Academic Press.

Geiger, W.P., and H.R. Dorsch. 1980. Quantity-Quality Simulation (QQS): A Detailed Continuous Planning Model for Urban Runoff Control, Volume 1, Model Description, Testing and Applications. EPA/600/2-80-011. Cincinnati, Ohio: US Environmental Protection Agency.

Haan, C. T. 1989. Parametric uncertainty in hydrologic modeling. *Trans. ASAE* 32(1):137-146.

Haan, C. T. 2002. *Statistical Methods in Hydrology*. 2nd ed. Ames: Iowa State University Press.

Haan, P. K., and R. W. Skaggs. 2003a. The effect of parameter uncertainty on DRAINMOD predictions: I. Hydrology and yield. *Trans. ASAE*. 46(4): 1061–1067.

Haan, P. K., and R. W. Skaggs. 2003b. The effect of parameter uncertainty on DRAINMOD predictions: II. Nitrogen loss. *Trans. ASAE*. 46(4): 1069–1075.

Hamrick, J. M. 1992. A Three-Dimensional Environmental Fluid Dynamics Computer Code: Theoretical and Computational Aspects. Special Report 317. Williamsburg, Va.: The College of William and Mary, Virginia Institute of Marine Science.

Hamrick, J. M. 1996. User's Manual for the Environmental Fluid Dynamics Computer Code. Special Report 331. Williamsburg, Va.: The College of William and Mary, Virginia Institute of Marine Science.

Hamrick, J. M., and W. B. Mills. 2000. Analysis of water temperatures in Conowingo Pond as influenced by the Peach Bottom atomic power plant thermal discharge. *Environmental Science and Policy*. 3:S197-S209.

Havno, K., M. N. Madsen, and J. Dorge. 1995. MIKE 11 - A generalised river modelling package. In *Computer Models of Watershed Hydrology*. Singh, V.P., ed. 733-782. Littleton, Colo.: Water Resources Publications.

HEC. 1977. *Storage, Treatment, Overflow, Runoff Model, STORM, Generalised Computer Program 723-58-L77520*. Davis, Calif.: Hydrologic Engineering Center, US Corps of Engineers.

HEC. 1978. *WQRRS Water Quality for River-Reservoir Systems: User's Manual*. CPD-8. Davis, Calif.: Hydrologic Engineering Center, US Army Corps of Engineers.

Helena, B., R. Pardo, M. Vega, E. Barrado, J. M. Fernandez, and L. Fernandez. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* 34(3): 807-816.

Hession, W. C., D. E. Storm, and C. T. Haan. 1996a. Two-phase uncertainty analysis: An example using the universal soil loss equation. *Trans. ASAE*. 39(4): 1309-1319.

Hession, W. C., D. E. Storm, C. T. Haan, S. L. Burks, and M. D. Matlock. 1996b. A watershed-level ecological risk assessment methodology. *Water Resour. Bulletin*. 32 (5): 1039-1054.

Hession, W.C., D.E. Storm, C. T. Haan, K.H. Reckhow, M.D. Smolen, and S.L. Burks. 1996c. Risk analysis of TMDLs in an uncertain environment. *Lake and Reservoir Management*. 12(3): 331-347.

Hoffpauir, R. 2002 Calculating Hydrologic Parameters for Estimating Surface Water Flow with GIS. SR 2002-020, College Station, Tex.: Texas Water Resources Institute, Texas A&M University.

Huber, W.C., J.P. Heaney, S.J. Nix, R.E. Dickinson, and D.J. Polmann. 1984. *Storm Water Management Model, User's Manual,* Version III. EPA/600/2-84-109a. Athens, Ga.: US Environmental Protection Agency.

Huber. WC. and R.E. Dickinson. 1988. *Storm Water Management Model, Version 4, User's Manual.* EPA/600/3-88-001a. Athens, Ga.: US Environmental Protection Agency.

Illinois Water Survey. 1990. Simulation of Urban Runoff and Pollutant Loading From the Greater Lake Calumet Area. Champaign, Ill.: Office of Spatial Data Analysis and Information, Illinois Water Survey.

Jayakrishnan, R. 2001. Effect of rainfall variability on hydrologic simulation using WSR-88D (NEXRAD) data. PhD diss. College Station: Texas A&M University, Department of Biological and Agricultural Engineering.

Jenerette, G. D., J. Lee, D. W. Waller, and R.E. Carlson. 2002. Multivariate analysis of the ecoregion delineation for aquatic systems. *Env. Mgmt*. 29(1): 67-75.

Ji, Z.-G., M. R. Morton, and J. M. Hamrick. 2001. Wetting and drying simulation of estuarine processes. *Estuarine, Coastal and Shelf Science*. 53:683-700.

Jirka, G.H. 1992. Use of Mixing Zone Models in Estuarine Waste Load Allocations. Technical Guidance Manual for Performing Water Load Allocations, Book III, Part 3. EPA/823/R/92/004. Washington, D.C.: Office of Water, US Environmental Protection Agency.

Jirka, G. H., R. L. Doneker, and S. W. Hinton. 1996. *User's Manual for CORMIX: A Hydrodynamic Mixing Zone Model and Decision Support System for Pollutant Discharges into Surface Waters.* Washington, D.C.: Office of Science and Technology, US Environmental Protection Agency.

Johnson, B. H., K. W. Kim, R. E. Heath, B. B. Hsieh, and H. L. Butler. 1993. Validation of three-dimensional hydrodynamic model of Chesapeake Bay. *J. Hyd. Engng*. 119: 2-20.

Johnson, R. A. and D. W. Wichern. 2002. *Applied Multivariate Statistical Analysis*. 5th ed., Upper Saddle River, N.J.: Prentice Hall International.

Jørgensen, B. S. 2002. A river rehabilitation study in Malaysia. MSc. thesis. Lyngby, Denmark: Technical University of Denmark, Department of Environment and Resources.

Kaiser, H. F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20: 141-151.

Kalkstein, L. S., G. Tan, and J. A. Skindlov. 1987. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate and Applied Meteorology*. 26: 717-730.

Kanji, G. K. 1999. 100 *Statistical Tests*. Thousand Oaks, Calif.: Sage Publications.

Karson, M. J. 1982. *Multivariate Statistical Methods*. Ames: Iowa State University Press.

Kreuter, U. P., H. G. Harris, M. D. Matlock, and R. E. Lacey. 2001. Change in ecosystem service values in the San Antonio area, Texas. *Ecol. Economics*. 39: 333-346.

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

Leonard, B. P. 1979. A stable and accurate convective modelling procedure based on upstream interpolation. *Computer Methods in Applied Mech. Eng*. 19: 59-98.

Liggett, J. A. and D. A. Woolhiser. 1967. Difference solutions of the shallow water equation. *J. Eng. Mech. Div*. 93(2): 39-71.

MapTech. 2000. Fecal Coliform TMDL (Total Maximum Daily Load) Development for Middle Blackwater River, Virginia. Blacksburg, Va.: MapTech Inc.

Milton, J. S. and J. C. Arnold. 1995. *Introduction of Probability and Statistics. Principles and Application for Engineering and the Computing Sciences*. 3rd Ed. New York, N.Y.: McGraw-Hill, Inc.

Moore, J. A., J. Smyth, S. Baker, and J. R. Miner. 1988. Evaluating coliform concentrations in runoff from various animal waste management systems. Special Report 817, Corvallis, Ore.: Agricultural Experiment Station, Oregon State University.

Morgan, M. G., and M. Henrion. 1998. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, N.Y.: Cambridge University Press.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models. Part 1. A discussion of principles. *J. Hydrology*. 10: 282-290.

Nathan, R. J. and T. A. McMahon. 1990a. Evaluation of automated techniques for baseflow and recession analysis. *Water Resour. Res.* 26(7): 1465-1473.

Nathan, R. J. and T. A. McMahon. 1990b. Identification of homogeneous regions for the purposes of regionalization. *J. Hydrology*, 121 (1990) 217-238.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, J. R. Williams. 2001. *Soil and Water Assessment Tool User's Manual:* Version 2000. Temple, Tex.: Blackland Research Center, Texas Agricultural Experiment Station.

Nix, S. J. 1990. Mathematical modeling of the combined sewer system. In *Control and Treatment of Combined Sewer Overflows*. 23- 78. New York, N.Y.: Van Nostrand Reinhold.

Overcash, M. R., K. R. Reddy, and R. Khaleel. 1983. Chemical processes and transport of animal waste pollutants. In *Agricultural Management and Water Quality*. F. W. Shaller and D. W. Bailey, eds. 109-125. Ames: Iowa State University Press.

Parker, A. and M. Lahlou. 2001. Application of a pathogen loading estimation tool to TMDL development. In *TMDL Science Issues Conference* Alexandria, Va.: Water Environment Federation. Available at: http://www.wef.org/pdffiles/TMDL/Parker_ Andrew.pdf. Accessed 1 March 2002.

Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin. 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. *Appl. Environ. Microbiol.* 65: 3142-3147.

Paul, S., M. D. Matlock, P. K. Haan, S. Mukhtar, and S. Pillai. 2002. Uncertainty Analysis as a First Step of Developing a Risk-Based Approach to Nonprofit Source Modeling of Fecal Coliform Pollution for Total Maximum Daily Load

Estimates. Technical Report No. 192. College Station, Tex.: Texas Water Resources Institute, Texas A&M University.

Pitt, R., and J. Voorhees. 2000. The Source Loading and Management Model (SLAMM) A Water Quality Management Planning Model for Urban Stormwater Runoff. Available at: http://www.eng.ua.edu/~rpitt/SLAMMDETPOND/WinSlamm/MainWINSLAMM_book.html. Accessed 1 March 2003.

Prabhu, S. 1995. Application of stochastic model evaluation protocol on EPIC and AGNPS. MS thesis. Stillwater: Oklahoma State University, Department of Biosystems Engineering.

Roesner, LA., J. A. Aldrich, and R. E. Dickinson. 1988. *Storm Water Management Model, User's Manual,* Version 4*:* Addendum I*, EXTRAN*. EPA/600/3-83-001b. Athens, Ga.: US Environmental Protection Agency.

Sadeghi, A. M. and J. G. Arnold. 2002. A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. In *Proc. Total Maximum Daily Load (TMDL) Environmental Regulations Conference*. St. Joseph, Mich.: ASAE.

Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck. 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. AWRA*. 37(5): 1169-1187.

SAS. 1999. *SAS/STAT User's Guide*. Version 8, Volumes 1-3. Cary, N.C.: SAS Institute, Inc.

Schnoor, J. L. 1996. *Environmental Modeling Fate and Transport of Pollutants in Water, Air and Soil.* New York, N.Y.: John Wiley & Sons, Inc.

Springer, E. P., G. F. Gifford, M. P. Windham, R. Thelin, and M.Kress. 1983. Fecal coliform release studies and development of a preliminary nonpoint source transport model for indicator bacteria. Logan, Utah: Utah Water Research Laboratory, Utah State University.

Srivastava, M. S., and E. M. Carter. 1983. *An Introduction to Applied Multivariate Statistics*. New York, N.Y.: Elsevier Science Publishing Co.

Suter, G. W. II. 1999. Developing conceptual models for complex ecological risk assessments. *Human and Ecol. Risk Assessment* 5(2): 375-396.

Suter, G. W. II, L. W. Barthouse, and R. V. O'Neil. 1987. Treatment of risk in environment impact assessment. *Env. Mgmt*. 11(3): 295-303.

TNRCC. 2000. Texas 2000 Clean Water Act Section 303(d) List. Austin, Tex.: Texas Commission on Environmental Quality. Available at: http://www.tnrcc.state.tx.us/ water/ quality/ 00-303dlist.pdf. Accessed 18 September 2001.

TPWD. 2002. South Texas Wildlife Management-White-tailed Deer Management. Austin, Tex.: Texas Parks and Wildlife Department. Available at: http://www.tpwd.state.tx.us/southtx_plain/deer/index.htm. Accessed 1 February 2002.

TWDB. 1970. *Simulation of Water Quality in Streams and Canals, Program Documentation and User's Manual*. Austin, Tex.: Texas Water Development Board.

Tyagi, A., and C. T. Haan. 2001. Uncertainty analysis using corrected first-order approximation method. *Water Resour. Res*. 37(6): 1847-1858.

US Census Bureau. 2000. *Census 2000 TIGER/Line® Files*. Washington D.C.: US Census Bureau. Available at: http://www.census.gov/geo/www/tiger/tiger2k/ tgr2000.html. Accessed 30 October 2000.

USDA. 2002. *USDA-NASS Published Estimates Data Base: 2000-2001*. National Agricultural Statistics Database. Washington D.C.: USDA National Agricultural Statistics Service. Available at: http://www.nass.usda.gov: 81/ ipedb/. Accessed 25 February 2002.

USEPA. 1985. Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling (Second Edition). EPA/600/3-85/040. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1986. Ambient water quality criteria for bacteria. EPA 440/5/84/002. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1991a. Guidance for Water Quality-based Decisions: The TMDL Process. EPA/440/4-91/001. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1991b. Modeling of nonpoint source water quality in urban and non-urban areas. EPA 600/3/91/039. Washington D.C.: Office of Research and Development, US Environmental Protection Agency.

USEPA. 1995. Watershed protection: A statewide approach. EPA 841/R/95/004., Washington D.C.: Office of Wetlands, Oceans, and Watersheds, US Environmental Protection Agency.

USEPA. 1996a. Better Assessment Science Integrating Point and Nonpoint Sources (BASINS). EPA823/R/96/001. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1996b. Dynamic Toxics wasteload allocation model (DYNTOX), version 2.1, user's manual. Washington D.C.: Office of Science and Technology, US Environmental Protection Agency.

USEPA. 1996c. TMDL Development cost estimated: Case studies of 14 TMDLs. EPA 841/R/96/001. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1997a. Compendium of tools for watershed assessment and TMDL development. EPA 841/B/97/006. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 1997b. Policy for use of probabilistic analysis in risk assessment at the U.S. Environmental Protection Agency. Washington D.C.: US Environmental Protection Agency. Available at: http://www.epa.gov/ncea/mcpolicy.htm. Accessed 18 October 2000.

USEPA. 1998a. Clean Water Action Plan. Washington D.C.: US Environmental Protection Agency. Available at: http://www.epa.gov/history/topics/cwa/03.htm. Accessed 12 December 1999.

USEPA. 1998b. *National Water Quality Inventory: 1998 Report to Congress.* Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2000a. Bacterial Indicator Tool User's Guide. EPA-823-B-01-003. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2000b. BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF. EPA/823/R00/012. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2000c. Fecal coliform TMDL modeling report Cottonwood Creek watershed Idaho County, Idaho. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2001a. *Better Assessment Science Integrating point and Nonpoint Sources, User's Manual for Release 3.0*. EPA/823/B-01/001 Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2001b. PLOAD Version 3.0: An ArcView GIS tool to calculate nonpoint sources of pollution in watershed and stormwater projects. EPA-WDC0101-HSQ, Washington, D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2001c. Protocol for developing pathogen TMDLs (First Edition). EPA/841/R-00/002. Washington D.C.: Office of Water, US Environmental Protection Agency.

USEPA. 2003. Bacterial Water Quality Standards for Recreational Waters (Freshwater and Marine Waters) Status Report. EPA/823/R/03/008. Washington D.C.: US Environmental Protection Agency.

USGS. 2001. Daily Streamflow for the Nation: USGS 08178800Salado Ck at Loop 13 at San Antonio, TX. Reston, Va.: US Geological Survey. Available at:

http://waterdata.usgs.gov/nwis/discharge/?site_no =08178800 Accessed 9 April 2001.

Vega M., R. Pardo, E. Barrado, and L. Debn. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res*. 32(12): 3581-3592.

Walker, S. E., S. Mostaghimi, T. A. Dillaha, and F. E. Woeste. 1990. Modeling animal waste management practices: impacts on bacteria levels in runoff from agricultural lands. *Trans. ASAE* 33(3): 807-817.

Weglarczyk, S. 1998. The interdependence and applicability of some statistical quality measures for hydrological models. *J. Hydrology* 206: 98-103.

Whitlock, J. E., D. T. Jones, and V. J. Harwood. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Res.* 36: 4265-4274.

Wiggins B. A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman. 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Appl. Environ. Microbiol.* 65: 3483-3486.

Wool T. A, R. B. Ambrose, J. L. Martin, and E. A. Comer. 2002. *Water Quality Analysis Simulation Program (WASP)* Version 6.0 Draft: User's Manual. Atlanta, Ga.: US Environmental Protection Agency.

Young, K. D. and E. L. Thackston. 1999. Housing density and bacterial loading in urban streams. *J. Env. Engng.* December: 1177-1180.

Zhang, H. and C. T. Haan. 1996. Evaluation of uncertainty in estimated flow and phosphorous loads by FHANTM. *Applied Engng. Agric.* 12(6): 663-669.

Zoppou, C. 2001. Review of urban storm water models. *Env. Modelling and Software*. 16: 195-231.

APPENDIX A

GENERAL CHARACTERISTICS OF THE BACTERIALLY IMPAIRED WATERBODIES

Table A.1 List of bacterially impaired stream segments in Texas.

Segment ID	Segment Name	Type^{+}	Size	Size Units	Count	Mean (cfu/100ml)
2426	Tabbs Bay	E	4	Sq. miles	411	435.1
2429	Scott Bay	E	2	Sq. miles	1749	16100
0204	Red River Above Lake Texoma	FS	158	Miles	175	977
0205	Red River Below Pease River	FS	66	Miles	171	3100
0306	Upper South Sulphur River	FS	42	Miles	36	771.3
0513	Big Cow Creek	FS	30	Miles	173	498.8
0604	Neches River Below Lake Palestine	FS	231	Miles	189	2922
0607	Pine Island Bayou	FS	81	Miles	319	3338
0611	Angelina River Above Sam Rayburn Reservoir	FS	104	Miles	250	468.7
0804	Trinity River Above Lake Livingston	FS	160	Miles	485	1296
0805	Upper Trinity River	FS	100	Miles	335	46980
0806	West Fork Trinity River Below Lake Worth	FS	33	Miles	223	12910
0810	West Fork Trinity River Below Bridgeport Reservoir	FS	36	Miles	387	368.6
0819	East Fork Trinity River	FS	29	Miles	138	9517
0841	Lower West Fork Trinity River	FS	27	Miles	191	17320
0902	Cedar Bayou Above Tidal	FS	25	Miles	32	896
1008	Spring Creek	FS	69	Miles	495	1219
1009	Cypress Creek	FS	53	Miles	1046	3520
1014	Buffalo Bayou Above Tidal	FS	24	Miles	846	11800
1016	Greens Bayou Above Tidal	FS	24	Miles	737	13620
1017	Whiteoak Bayou Above Tidal	FS	23	Miles	699	17260
1102	Clear Creek Above Tidal	FS	30	Miles	320	2671

Table A.1 Continued.

Segment ID	Segment Name	Type^{+}	Size	Size Units	Count	Mean (cfu/100ml)
1104	Dickinson Bayou Above Tidal	FS	7	Miles	83	3125
1108	Chocolate Bayou Above Tidal	FS	22	Miles	19	1018.8
1110	Oyster Creek Above Tidal	FS	77	Miles	44	1260
1218	Nolan Creek /South Nolan Creek	FS	29	Miles	332	413.2
1221	Leon River Below Proctor Lake	FS	173	Miles	418	326.4
1226	North Bosque River	FS	103	Miles	2037	849.4
1242	Brazos River Above Navasota River	FS	223	Miles	419	389
1245	Upper Oyster Creek	FS	30	Miles	286	3192
1255	Upper North Bosque River	FS	13	Miles	567	4661
1414	Pedernales River	FS	125	Miles	426	391.3
1427	Onion Creek	FS	78	Miles	579	1190
1428	Colorado River Below Town Lake	FS	41	Miles	1875	875.4
1430	Barton Creek	FS	40	Miles	757	379.9
1502	Tres Palacios Creek Above Tidal	FS	34	Miles	300	1635
1901	Lower San Antonio River	FS	153	Miles	543	2849
1903	Medina River Below Medina Diversion Lake	FS	80	Miles	462	416.66
1906	Lower Leon Creek	FS	32	Miles	393	2381
1910	Salado Creek	FS	44	Miles	816	4247
1911	Upper San Antonio River	FS	85	Miles	2601	3480
2107	Atascosa River	FS	103	Miles	75	1773
2110	Lower Sabinal River	FS	27	Miles	84	211.6
2117	Frio River Above Choke Canyon Reservoir	FS	158	Miles	139	556.6
2202	Arroyo Colorado Above Tidal	FS	63	Miles	179	2562
2302	Rio Grande Below Falcon Reservoir	FS	231	Miles	949	1167

Table A.1 Continued.

Segment ID	Segment Name	Type^{+}	Size	Size Units	Count	Mean (cfu/100ml)
2304	Rio Grande Below Amistad Reservoir	FS	226	Miles	1344	54610
2306	Rio Grande Above Amistad Reservoir	FS	313	Miles	246	536.5
0101A	Dixon Creek	FS	19	Miles	47	384.4
0202D	Pine Creek	FS	27	Miles	70	2005
0203A	Big Mineral Creek	FS	18	Miles	26	358.4
0207A	Buck Creek	FS	50	Miles	30	467.47
0404B	Tankersley Creek	FS	8	Miles	114	717.9
0502A	Nichols Creek	FS	32	Miles	12	904
0505B	Grace Creek	FS	14	Miles	30	3394
0505D	Rabbit Creek	FS	29	Miles	33	1832
0507A	Cowleech Fork Sabine River	FS	30	Miles	78	2808
0507B	Long Branch	FS	8	Miles	20	1001
0508A	Adams Bayou Above Tidal	FS	8	Miles	368	5445
0508B	Gum Gully	FS	4	Miles	44	5472
0511A	Cow Bayou Above Tidal	FS	11	Miles	165	1762
0603A	Sandy Creek	FS	23	Miles	33	441.1
0604A	Cedar Creek	FS	23	Miles	131	877.9
0604B	Hurricane Creek	FS	4	Miles	38	1171
0604C	Jack Creek	FS	16	Miles	47	528.5
0605A	Kickapoo Creek	FS	31	Miles	64	909.9
0607C	Willow Creek	FS	15	Miles	34	270.6
0608B	Big Sandy Creek	FS	41	Miles	82	629.35
0608C	Cypress Creek	FS	24	Miles	55	488.1
0608D	Hickory Creek	FS	24	Miles	107	475
0608F	Turkey Creek	FS	29	Miles	53	449.7

Table A.1 Continued.

Segment ID	Segment Name	Type$^+$	Size	Size Units	Count	Mean (cfu/100ml)
0610A	Ayish Bayou	FS	32	Miles	91	357.3
0611B	La Nana Bayou	FS	32	Miles	80	871.4
0611C	Mud Creek	FS	51	Miles	135	1482
0612B	Waffelow Creek	FS	10	Miles	14	373.9
1113A	Armand Bayou Above Tidal	FS	6	Miles	11	4875
1209C	Carters Creek	FS	17	Miles	141	569.5
1217A	Rocky Creek	FS	12	Miles	46	215.9
1222A	Duncan Creek	FS	12	Miles	1	288
1226A	Duffau Creek	FS	21	Miles	135	447.3
1226C	Meridian Creek	FS	25	Miles	224	163.4
1226D	Neils Creek	FS	26	Miles	251	192.4
1304A	Linville Bayou	FS	23	Miles	36	473.3
1403A	Bull Creek	FS	10	Miles	802	454.4
1427A	Slaughter Creek	FS	16	Miles	26	264.8
1427B	Williamson Creek	FS	16	Miles	162	863.4
1427C	Bear Creek	FS	15	Miles	-	-
1428A	Boggy Creek	FS	7	Miles	25	141.6
1428B	Walnut Creek	FS	20	Miles	126	1222
1428C	Gilleland Creek	FS	24	Miles	61	673.9
1429A	Shoal Creek	FS	10	Miles	292	641.8
1429B	Eanes Creek	FS	6	Miles	466	279.9
1803A	Elm Creek	FS	24	Miles	44	1231
1804B	Peach Creek	FS	64	Miles	65	851.1
1811A	Dry Comal Creek	FS	30	Miles	191	684.9

Table A.1 Continued.

Segment ID	Segment Name	Type[+]	Size	Size Units	Count	Mean (cfu/100ml)
0105	Rita Blanca Lake	R	524	Acres	30	98.5
1403	Lake Austin	R	1830	Acres	1873	164.5
2116	Choke Canyon Reservoir	R	26000	Acres	160	798.33
0508	Adams Bayou Tidal	TS	8	Miles	352	5498
0511	Cow Bayou Tidal	TS	20	Miles	595	1075
0901	Cedar Bayou Tidal	TS	19	Miles	34	1101
1001	San Jacinto River Tidal	TS	17	Miles	1346	1312
1013	Buffalo Bayou Tidal	TS	4	Miles	931	13320
1101	Clear Creek Tidal	TS	12	Miles	706	2687
1103	Dickinson Bayou Tidal	TS	15	Miles	83	3125
1109	Oyster Creek Tidal	TS	25	Miles	83	905.7
1113	Armand Bayou Tidal	TS	8	Miles	59	2853
1304	Caney Creek Tidal	TS	32	Miles	99	999.1
0511B	Coon Bayou	TS	5	Miles	416	942.7
0511C	Cole Creek	TS	10	Miles	50	2327

[+] E = Estuary, FS = Freshwater stream, R = Reservoir, TS = Tidal Stream

APPENDIX B

WATERSHED CHARACTERISTICS

Table B.1 Descriptive Statistics of watershed parameters.

Variable	Mean	Std Dev	Minimum	Maximum
Percent Forest (%)	32.91	22.94	0.00	89.57
Percent Cropland (%)	6.66	10.22	0.00	48.75
Percent Water (%)	2.35	6.93	0.03	52.17
Percent Residential (%)	9.77	13.44	0.00	59.02
Percent Commercial (%)	3.80	5.34	0.00	29.01
Percent Wetland (%)	3.45	5.41	0.00	38.76
Percent Pasture (%)	31.36	16.78	2.80	71.93
Percent Barrenland (%)	9.71	15.78	0.02	94.94
Distance Factor Forest (m)	147.11	138.31	1.00	657.51
Distance Factor Cropland (m)	361.49	307.79	1.00	1627.39
Distance Factor Water (m)	274.03	226.89	0.00	1358.27
Distance Factor Residential (m)	245.06	268.98	0.00	1408.55
Distance Factor Commercial (m)	458.25	283.52	1.00	2074.33
Distance Factor Wetland (m)	229.59	330.09	0.00	1671.30
Distance Factor Pasture (m)	168.62	140.29	6.24	875.62
Distance Factor Barren (m)	429.74	456.40	3.02	3654.16
Population Density (number. ac^{-1})	1.13	1.68	0.002	8.92
Density of Households (number. ac^{-1})	0.42	0.68	0.0006	4.35
Density of OSSF (number. ac^{-1})	0.02	0.04	0	0.25
Density of Other Septic systems (number. ac^{-1})	0.001	0.001	0.000	0.01
Average Age of Households (years)	32.21	16.77	21.11	100.00
Density of Public Sewers (number. ac^{-1})	0.29	0.49	0.0003	2.71
Density of Cattle (number. ac^{-1})	0.04	0.03	0.0025	0.19
Density of Swine (number. ac^{-1})	0.0011	0.0031	0.00	0.02
Density of Sheep (number. ac^{-1})	0.0007	0.0021	0.000	0.0110
Density of Deer (number. ac^{-1})	0.02	0.03	0.00	0.16
Stream Density (m. ac^{-1})	3.97	6.75	0.33	69.91
Average Precipitation (mm)	1025.40	266.42	363.53	1497.62
Average Temperature (°C)	18.32	1.47	12.00	22.00

Calculation of bacterial loading rates

1. Loading from deer = Density of Deer *5.0E8
2. Loading from Farm Animals = (Density of Cattle * 5.4E9 + Density of Swine * 8.9E9/2 + Density of Sheep * 1.8E10)/ (Percent Cropland+ Percent Pasture)

Table B.2 Varimax rotated factors considering only the discriminating parameters.

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Percent Wetland	0.89	-0.08	-0.10	-0.09	0.10	-0.07
Average Precipitation	0.78	-0.40	-0.01	0.09	0.22	0.13
Bacterial Loading from Deer	-0.65	-0.24	-0.07	-0.24	0.30	-0.33
Distance Factor -Forestland	-0.06	0.91	-0.13	-0.04	-0.05	-0.01
Percent Forestland	0.21	-0.77	0.21	-0.10	0.33	-0.01
Distance Factor -Pastureland	-0.06	0.01	0.93	-0.01	-0.07	-0.01
Percent Cropland	-0.32	0.32	-0.52	-0.17	-0.41	-0.24
Percent Pasture	0.03	0.28	-0.88	-0.02	-0.06	-0.05
Average Age of Households	-0.03	0.14	0.06	0.94	-0.04	0.00
Bacterial Loading from Farm Animal	0.13	-0.22	0.01	0.02	0.86	-0.12
Density of Other Septic Systems	0.11	-0.19	-0.03	0.81	0.09	0.25
Distance Factor -Residential	0.37	-0.34	0.06	0.00	-0.45	-0.31
Density of Households	0.08	-0.02	0.06	0.15	-0.07	0.95
Distance Factor -Water	0.04	0.03	0.02	0.10	0.12	0.00
Average Temperature	0.08	-0.04	0.16	0.06	-0.02	0.11

APPENDIX C

DETAILS OF WATERSHEDS SELECTED FOR MODELING

Table C.1 Principal Land Uses by Sub watershed: Upper North Bosque watershed.

Sub basin	Urban %	Forestland %	Pasture/Range land %	Cropland %
1	1	9	67	14
2	20	13	30	31
4	7	9	36	14
36	1	17	44	24
Total	5	12	45	18

Table C.2 Principal Land Uses by Sub watershed: Medina Creek Watershed.

Sub basin	Urban %	Forestland %	Pasture/Range land %	Cropland %
1	0	71	5	17
2	1	62	20	12
3	1	37	31	23
4	2	20	28	33
5	17	38	18	10
6	4	10	41	27
7	1	13	65	14
8	1	10	69	10
9	1	4	50	15
10	3	41	33	7
11	0	5	61	19
Total	4	36	29	18

Table C.3 Principal Land Uses by Sub watershed: Mud Creek Watershed (0611C).

Sub basin	Urban %	Forestland %	Pasture/Range land %	Cropland %
1	2	41	37	1
2	2	31	56	1
3	1	51	39	1
4	1	47	37	0
5	1	43	51	1
6	6	38	44	1
7	2	43	44	1
8	3	44	43	1
9	1	42	50	1
10	1	50	40	1
Total	2	43	43	1

Table C.4 Principal Land Uses by Sub watershed: North Bosque Watershed.

Sub basin	Urban %	Forestland %	Pasture/Range land %	Cropland %
1	4	11	41	17
2	1	10	27	15
3	0	13	62	2
4	1	19	54	3
5	0	18	32	9
6	0	19	38	3
7	9	18	22	5
8	0	9	24	3
9	1	22	47	7
10	0	27	47	5
11	0	32	50	1
12	0	19	55	2
13	3	27	33	15
14	0	32	44	4
15	1	20	48	6
16	1	13	24	4
17	7	13	40	14
18	2	33	28	13
Total	1	19	41	8

Table C.5 Initial values of MON-ACCUM (#org. d^{-1}) corresponding to cropland for different watersheds.

Month/Watershed	Mud Creek	North Bosque	Upper North Bosque	Medina Creek
Jan	8.25E+09	8.60E+09	1.40E+10	4.98E+09
Feb	4.81E+08	4.96E+08	1.55E+10	3.15E+08
Mar	4.34E+08	4.48E+08	1.40E+10	2.84E+08
Apr	9.86E+10	1.03E+11	1.45E+10	5.93E+10
May	9.54E+10	9.96E+10	1.40E+10	5.74E+10
Jun	9.86E+10	1.03E+11	1.45E+10	5.93E+10
Jul	3.58E+10	3.73E+10	1.42E+10	2.19E+10
Aug	3.58E+10	3.73E+10	1.42E+10	2.19E+10
Sep	3.70E+10	3.86E+10	1.47E+10	2.26E+10
Oct	3.58E+10	3.73E+10	1.40E+10	2.15E+10
Nov	4.49E+08	4.63E+08	1.45E+10	2.94E+08
Dec	4.34E+08	4.48E+08	1.40E+10	2.84E+08

Table C.6 Initial values of MON-ACCUM (#org. d^{-1}) corresponding to pastureland for different watersheds.

Month/Watershed	Mud Creek	North Bosque	Upper North Bosque	Medina Creek
Jan	8.25E+09	8.64E+09	1.40E+10	5.00E+09
Feb	4.82E+08	5.30E+08	8.60E+08	3.30E+08
Mar	4.36E+08	4.82E+08	7.78E+08	3.00E+08
Apr	9.86E+10	1.03E+11	1.67E+11	5.93E+10
May	9.54E+10	9.96E+10	1.62E+11	5.74E+10
Jun	9.86E+10	1.03E+11	1.67E+11	5.93E+10
Jul	3.58E+10	3.74E+10	6.11E+10	2.19E+10
Aug	3.58E+10	3.74E+10	6.11E+10	2.19E+10
Sep	3.70E+10	3.86E+10	6.31E+10	2.26E+10
Oct	3.58E+10	3.74E+10	6.07E+10	2.16E+10
Nov	4.50E+08	4.97E+08	8.04E+08	3.09E+08
Dec	4.36E+08	4.82E+08	7.78E+08	3.00E+08

Table C.7 Initial values of MON-ACCUM (#org. d^{-1}) corresponding to forestland and urban land for different watersheds.

Watershed	Urban	Forest
Mud Creek	1.18E+07	9.24E+06
North Bosque	1.14E+07	5.98E+07
Upper North Bosque	1.15E+07	3.39E+07
Medina Creek	1.44E+07	2.93E+07

VITA

Sabu Paul
Tetra Tech, Inc.
10306 Eaton Place Suite 340
Fairfax, VA 22030
Tel. 703-385-6000; 703-385-6007 (Fax)
Email: sabupaul28@yahoo.com

Educational background:

B.Tech., Agricultural Engineering, 1996, Kerala Agricultural University, India
M.Tech., Water Resources, 1998, Indian Institute of Technology, India
Ph.D., Biological and Agricultural Engineering, 2003, Texas A&M University, College Station, Texas.

Experience:

Graduate Research/Teaching Assistant, Texas A&M University, College Station, Texas - August 1999 to Present.
Summer Intern, Tetra Tech, Inc., Fairfax, VA - May 2001 to August 2001.
Research Assistant/Junior Project Officer, Indian Institute of Technology, India - June 1997– June1999.

Publications:

Paul, S., P.K. Haan, M. Matlock, and S. Mukhtar. 2002. Analysis of the HSPF Parameter Uncertainty in Predicting Peak In-Stream Fecal Coliform Concentrations. *Transactions of the ASAE* (accepted).
Paul, S., R. Srinivasan, P. Haan, S. Mukhtar, and K. Neimann. 2002. Bacterial Total Maximum Daily Load (TMDL): A new classification scheme for impaired waterbodies of Texas. 2002 National TMDL Science and Policy conference at Phoenix, Arizona, Paper No. 10E. Alexandria, Va.: WEF.
Paul, S., P. Haan, M. D. Matlock, and S. Mukhtar. 2002. Analysis of the HSPF Model for Predicting In-stream Fecal Coliform Concentrations. ASAE Meeting Paper No. 022157. St. Joseph, Mich.: ASAE.
Paul, S., M. Matlock, M., P.K. Haan, S. Mukhtar, and S. Pillai. 2002. Uncertainty Analysis as a First Step of Developing a Risk-Based Approach to Nonprofit Source Modeling of Fecal Coliform Pollution for Total Maximum Daily Load Estimates. Technical Report No. 192, Texas Water Resources Institute. Texas A&M University
Paul, S., S.N. Panda and D.N. Kumar. 2000. Optimal irrigation allocation: A multilevel approach. *Journal of Irrigation and Drainage Engineering*, 126(3): 149-156