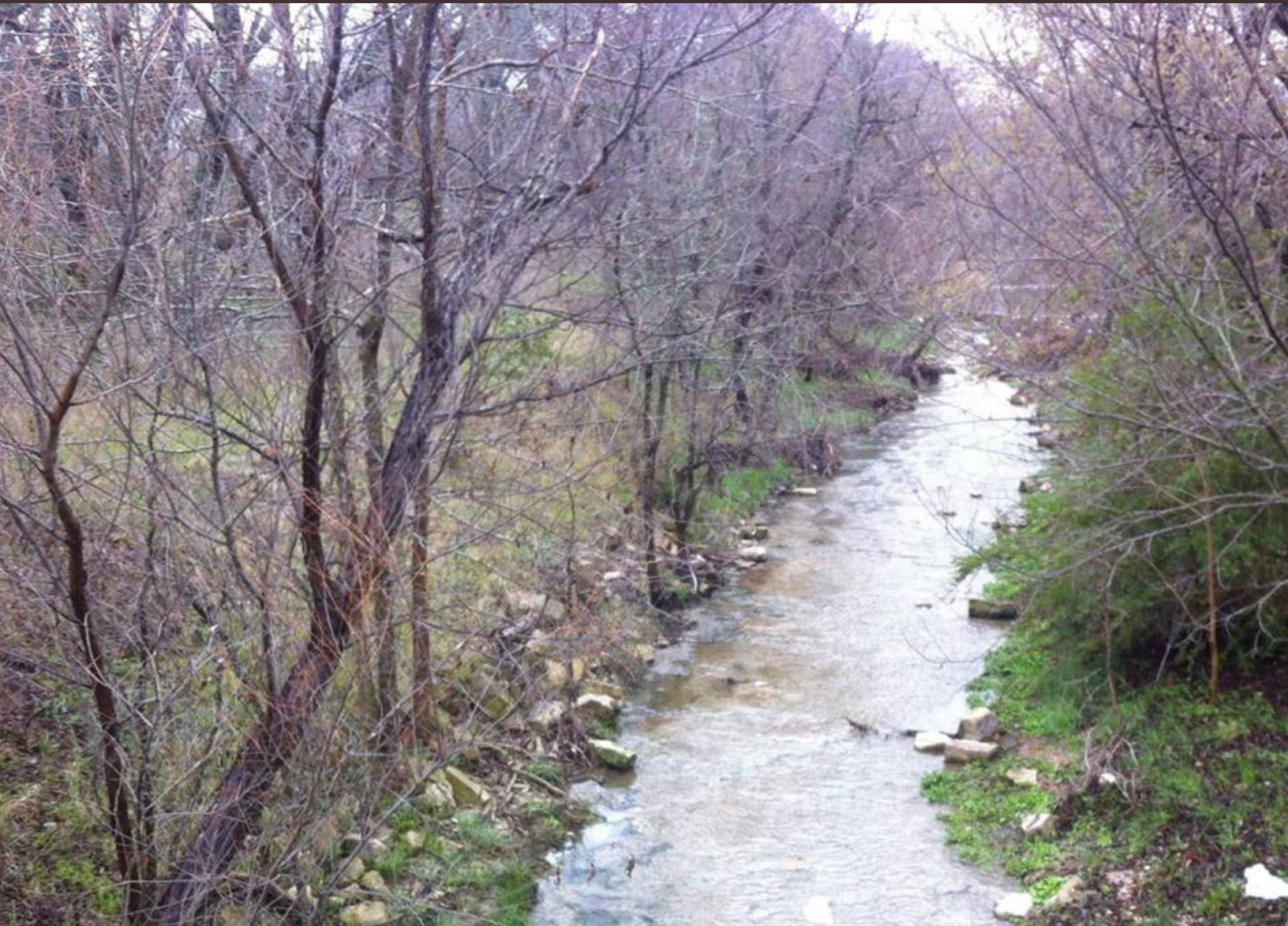


# Distribution of E. coli levels and recreation use as factor of stream order in the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains Ecoregions

Texas Water Resources Institute TR-491  
September 2015



# Distribution of *E. coli* levels and recreation use as factor of stream order in the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains Ecoregions

---

**STATE NONPOINT SOURCE GRANT PROGRAM**

**TSSWCB PROJECT 14-56**

*Prepared for:*

**TEXAS STATE SOIL AND WATER  
CONSERVATION BOARD**

*Prepared by:*

**KHURRAM RAFI  
KEVIN L. WAGNER  
TEXAS WATER RESOURCES INSTITUTE**

-----

**TERRY J. GENTRY  
TEXAS A&M UNIVERSITY  
DEPARTMENT OF SOIL AND CROP SCIENCES**

-----

**R. KARTHIKEYAN  
TEXAS A&M UNIVERSITY  
DEPARTMENT OF BIOLOGICAL AND AGRICULTURAL ENGINEERING**

**SEPTEMBER 2015**

**TEXAS WATER RESOURCES INSTITUTE TR-491**

## Table of Contents

Executive Summary .....	ii
Introduction .....	1
Relationship between Stream Order and <i>E. coli</i> Concentrations .....	3
Relationship between Stream Order and Recreational Use .....	5
Recommendations Regarding Water Quality Standards .....	5
Leon River WPP Revision to Address EPA Comments .....	5
Literature Cited .....	6
Appendix A .....	7
Appendix B .....	68
Appendix C .....	97
Appendix D .....	108
Appendix E .....	111

## Executive Summary

This project examined the relationship of *E. coli* concentrations and recreational use to both stream order and watershed size. To determine possible ecoregion effects, the *E. coli* data used in this study were obtained from monitoring stations located on freshwater streams located in three ecoregions in Texas and Oklahoma – the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains (Ecoregions 27, 29, and 35, respectively). Median *E. coli* concentrations from the monitoring stations were analyzed for correlation with respect to the stream order of each monitoring site as well as the watershed size of each monitoring station. Geospatial analysis was used to determine stream order and watershed size and to identify unimpacted/least impacted streams in each ecoregion. Stream order was classified based on the traditional stream order classification method by Horton (1945) and Strahler (1957) and stream link magnitude analysis method by Shreve (1966).

The analysis of two stream order systems and watershed size for each monitoring site with respect to median *E. coli* showed no significant relationship between *E. coli* and stream orders/watershed size of unimpacted watersheds. However, the watersheds with wastewater outfalls and urban areas exceeding 10% of land use showed a statistically significant, yet weak negative relationship between *E. coli* and stream order/watershed size i.e. *E. coli* decreased with increase in stream order/watershed size.

Generally, a lack of completed RUAA studies prevented analysis of correlation between *E. coli* recreation standards based on the stream orders. However, it was noted that primary contact recreation has been associated with the range of stream orders. This, in addition to the finding that there was no correlation between *E. coli* concentrations and stream order at unimpacted sites, lead to the conclusion that stream order alone cannot be used to guide recreational use designation (i.e. Primary Contact Recreation, Secondary Contact Recreation 1, etc.) as hypothesized. Thus, the hypothesis that stream order could be used to guide recreational use and associated standards designation was not supported by project findings. As such, it is recommended that site specific studies continue to be conducted.

Finally, in order to finalize the Leon River Watershed Protection Plan so that implementation could proceed, through this project Parsons Engineering, under subcontract to TWRI, led efforts to address EPA's comments on the WPP received in November 2012. A detailed response was developed by Parsons addressing each of EPA's comments. Concurrently, revisions were made to the WPP in track changes to address EPA's comments. The response and revisions were reviewed with the Leon River Steering Committee and subsequently provided to the TSSWCB for submission to EPA. As a result of these efforts, on May 8, 2015, EPA informed TSSWCB that the Leon River WPP had been accepted.



## Introduction

Recent findings indicate that *E. coli* concentrations are naturally highest at the edge-of-field and small watershed scale and decline as watershed scale increases. This has tremendous potential impact on Water Quality Standards development and implementation as the likelihood of recreation also increases as waterbody scale (i.e. stream order) increases. Thus, it is expected that findings of stream order correlations with both *E. coli* concentration and level of recreational use could be used to prescribe specific standards based on stream order and help guide use designations prescribed through the RUAA process currently used by the state. This could help the State save money by allowing blanket designations for certain stream orders where recreational use is likely (highest stream orders) and unlikely (lowest stream orders). Further, this information can be used to establish reasonable, science-based standards which take into account natural background concentrations. It is recognized however that the watershed size associated with a particular stream order may vary due to differing geologic and climatic variables. As such, it is anticipated that ecoregion will be an important factor in this assessment.

While standards have been developed to protect waters for recreational use, their applicability has recently been questioned specifically related to scale. A case has been made for regime based standards such as temperature, sediment and stream order (Poole et al., 2004). In central Texas, Harmel et al. (2010) found decreasing *E. coli* concentrations when aggregating sampled segments from upstream to downstream. This trend occurred with both “impacted” (land use of mixed rural with diaries, WWTPs, small communities) and “unimpacted” (land use of mixed rural) streams of the same region. Additionally, Harmel et al. (2010) established that the “unimpacted” stream would be in violation of the Texas Water Quality Standards even though the anthropogenic impact was thought to be minor.

A Canadian study (Lyautey et al., 2010) established that Shreve (1966) stream link was inversely correlated to *E. coli* concentrations with strongest correlations found in the summer. Another Canadian study by Edge et al. (2012) used Horton (1945) – Strahler (1957) stream order and found that small streams (order  $\leq 3$ ) had higher *E. coli* concentrations than large streams (order  $>3$ ). A final Canadian study (Lyautey et al., 2011) explained, “the persistence factor is less important than dilution (i.e. stream order) in describing *E. coli* densities, followed by factors that influence the loading of *E. coli* into watersheds.” Although the reason for decreased *E. coli* densities related to stream order was chiefly cited as dilution (Lyautey et al., 2010; Lyautey et al., 2011), it was also hypothesized that turbidity/sediment processes, which may also be dependent on stream order, influenced *E. coli* concentrations (Bai and Lung, 2005; Brettar and Hofle, 1992; Craig et al., 2004; Oliver et al., 2007). To further evaluate this and its potential implications on Texas Water Quality Standards and watersheds, this project aims to:

- Determine the overall relationship between stream order and *E. coli* concentrations, as well as the relationship by ecoregion
- Determine the overall relationship between stream order and recreational use, as well as the relationship by ecoregion
- Make recommendations regarding the application of findings to water quality standards and recreational use classification.
- Revise the Leon WPP

## Relationship between Stream Order and *E. coli* Concentrations

(EXCERPT FROM K. RAFI THESIS – SEE APPENDIX A FOR FULL THESIS)

The relationship between stream order and *E. coli* concentrations was evaluated by K. Rafi as presented in Appendix A (Thesis), Appendix B (draft journal manuscript), Appendix C (presentation on findings) and Appendix D (project fact sheet). Conclusions from his thesis are presented below.

The data analysis of impacted monitoring sites with wastewater outfalls showed a weak but significant negative correlation of *E. coli* concentrations with Shreve and Strahler orders and watershed area. Analysis of the data showed that at larger watershed sizes and stream orders, *E. coli* concentrations were significantly lower. At smaller stream orders and watersheds, the *E. coli* concentrations ranged from very low to very high. It can be argued that the wider range of *E. coli* concentrations at smaller stream orders and watershed sizes was a driving factor in making the correlation coefficients weak. The *E. coli* concentrations did become lower as the stream orders and watersheds became larger which points to the dilution factor as suggested by Lyautey et al. (2010, 2011). But site specific studies may be a better way to understand *E. coli* relationships at smaller watershed and stream order scales where *E. coli* concentration range from very low to very high.

The data analysis of unimpacted sites showed that there was no correlation between *E. coli* concentrations and Shreve and Strahler orders as well as watershed area. Another significant observation for the unimpacted dataset was that in the Strahler and Modified Shreve stream order studies the median *E. coli* stayed statistically the same through all the stream orders which means that median *E. coli* concentration did not change and stayed within TAC limits as stream length and watershed area increased. This leads to the conclusion that in the absence of anthropogenic factors median *E. coli* will stay the same as watershed area and stream order increase and dilution does not impact the concentrations. The limitations for this conclusion include lack of monitoring stations available at higher stream orders. Many of the monitoring sites which had higher stream orders, e.g. sites at orders 6 and 7 for Strahler order, did not qualify as unimpacted watersheds due the presence of wastewater outfalls. With availability of more data such as discharge volume or hydrologic distance of the wastewater outfalls to the monitoring site, the criteria for exclusion of wastewater outfalls may be modified to include more monitoring sites. This would allow for a better definition of unimpacted watersheds and a decreasing *E. coli* trend may be seen in the unimpacted sites as well at higher stream orders similar to impacted watersheds.

The wastewater outfall impact on the streams was another significant finding of this study. The datasets that included wastewater outfalls, showed a weak but statistically significant negative correlation with *E. coli*, but when the outfalls were removed, there was no relationship. This suggests that wastewater outfalls are a big influence on stream *E. coli* and further research on just the sites that are impacted by these outfalls may

provide more valuable information on *E. coli* concentrations in the streams. Further studies with the types and volume of discharges made by wastewater outfalls may also help to assess the impact of wastewater outfalls on the water bodies.

The comparison of the three ecoregions also revealed that the median *E. coli* concentrations and their correlation with both stream order methods and watershed area decreased moving eastward. Ecoregion 27, which is the western most ecoregion in this study and has a more arid to semi-arid environment compared to the other ecoregions showed the highest *E. coli* concentration as well as correlations. Ecoregion 35, the eastern most region and with wettest climate of the three regions, has the lowest *E. coli* concentrations and correlations. This finding can be a good starting point for a future study that can include additional environmental factors such as precipitation, temperatures, types of streams (seasonal or perennial) etc., which make up each region and can contribute to the hydrology of these areas and its impact on the *E. coli* transport.

The study utilized Strahler and Shreve stream orders and watershed area as the variables to which the *E. coli* was correlated. Shreve stream order and watershed area provide similar information with respect to basin size and results based on them are very similar, it is recommended for future studies only one of the two variables should be used. The Strahler stream order method produces a smaller number of stream orders that helps in conducting the statistical analysis when comparing each stream order. This method is a good tool for comparing individual stream order numbers at ecoregion or river basin scales i.e. compare stream order 1 in ecoregion 27 with stream order 1 in ecoregion 29. Out of the three methods watershed area provides the most relevant information with respect to the scale of the basin but Strahler stream order can be the preferred method when comparing streams with similar characteristics.

For future research, inclusion of other variables such as flow volume at the time of the observation or a classification of the flow, such as low, normal, or high, which could identify if the sample was collected after a storm or in drought or normal conditions, may be beneficial as flow can influence *E. coli* concentrations. Other variables such as temperature at the time of sample collection, and time of collection in terms of seasonality, such as summer, spring etc., may be valuable in understanding *E. coli* relationships. A larger number of *E. coli* observations at monitoring sites at higher stream orders may also help provide a better and clearer understanding of the dilution affect. Further, a study on the major contributors, fate, and transport of *E. coli* in each ecoregion may also help understand the decreasing *E. coli* and correlation trends when moving eastward.



## **Relationship between Stream Order and Recreational Use**

The relationship between stream order and recreational use was evaluated by K. Rafi as presented in Appendix A (Thesis) and Appendix D (Fact Sheet). Generally, a lack of completed RUAA studies did not allow for correlation tests between *E. coli* recreation standards based on the stream orders. However, it was noted that primary contact recreation has been associated with the range of stream orders.

## **Recommendations Regarding Water Quality Standards**

The hypothesis that stream order could be used to guide recreational use and associated standards designation was not supported by research findings. Primary contact recreation has been associated with the range of stream orders. This, in addition to the finding that there was no correlation between *E. coli* concentrations and stream order at unimpacted sites, lead to the conclusion that stream order alone cannot be used to guide recreational use designation (i.e. Primary Contact Recreation, Secondary Contact Recreation 1, etc.) or criteria development as hypothesized. Thus, it is recommended that site specific studies continue to be conducted. In the future, there may be cause for reassessing this information when there are more RUAA studies completed.

## **Leon River WPP Revision to Address EPA Comments**

In order to finalize the Leon River Watershed Protection Plan so that implementation could proceed, Parsons Engineering, under subcontract to TWRI, led efforts to address EPA's comments on the WPP received in November 2012. A detailed response was developed by Parsons addressing each of EPA's comments. Concurrently, revisions were made to the WPP in track changes to address EPA's comments. The response and revisions were reviewed with the Leon River Steering Committee on December 16, 2014 (Appendix E). Following the Steering Committee, Parsons incorporated the modifications recommended by Watershed Steering Committee and provided the track changes and clean electronic files of the WPP and response to EPA comments to TSSWCB on January 26, 2015. TSSWCB finalized the response to comment letter and then submitted the Leon River WPP and response to comment letter to EPA on February 10, 2015. On May 8, 2015, EPA informed TSSWCB that the Leon River WPP had been accepted. The final, accepted WPP can be found on the Leon River Watershed Protection Program website (<http://leonriver.tamu.edu/media/1110/final-leon-wpp.pdf>).

## Literature Cited

- Bai, S. and W. Lung. 2005. "Modeling Sediment Impact on the Transport of Fecal Bacteria." *Water Research* 39(20):5232-5240.
- Brettar, I. and M. Hofle. 1992. "Influence of Ecosystematic Factors on Survival of Escherichia Coli after Large- Scale Release into Lake Water Mesocosms." *Applied and Environmental Microbiology* 58(7):2201-2210.
- Craig, D., H. Fallowfield and N. Cromar. 2004. "Use of Microcosms to Determine Persistence of Escherichia Coli in Recreational Coastal Water and Sediment and Validation with in Situ Measurements." *Journal of Applied Microbiology* 96(5):922-930.
- Edge, T., A. El-Shaarawi, V. Khan, C. Gannon, R. Jokinen, I. Kent, W. Phillips, D. Koning, J. Lapen, N. Wilkes, R. Miller, W. Scott, H. Topp, A. van Bochove, I. Neumann, E. Shtepani, G. Robertson and E. Schreier. 2012. "Investigation of an Escherichia Coli Environmental Benchmark for Waterborne Pathogens in Agricultural Watersheds in Canada." *Journal of Environmental Quality* 41(1):21-30.
- Harmel, R., T. Gentry, R. Karthikeyan, and R. Srinivasan. 2010. "Effects of Agricultural Management, Land Use, and Watershed Scale on *E. Coli* Concentrations in Runoff and Streamflow." *Transactions of the ASABE* 53(6):1833-1841.
- Horton, R. 1945. "Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology." *Bulletin of the Geological Society of America* 56(3):275-370.
- Lyautey, E., G. Wilkes, J. Miller, E. Van Bochove, H. Schreier, W. Koning, T. Edge, D. Lapen and E. Topp. 2011. "Variation of an Indicator of *Escherichia Coli* Persistence from Surface Waters of Mixed- Use Watersheds, and Relationship with Environmental Factors." *Annales de Limnologie - International Journal of Limnology* 47(1):11-19.
- Lyautey, E., Z. Lu, A. Scott, T. Berkens, E. Topp, D.R. Lapen, G. Wilkes and T. Edge. 2010. "Distribution and Diversity of *Escherichia Coli* Populations in the South Nation River Drainage Basin, Eastern Ontario Canada." *Applied and Environmental Microbiology* 76(5):1486-1496.
- Oliver, D., C. Clegg, A. Heathwaite and P. Haygarth. 2007. "Preferential Attachment of *Escherichia Coli* to Different Particle Size Fractions of an Agricultural Grassland Soil." *Water, Air, & Soil Pollution* 185(1-4):369-75.
- Poole, G., J. Dunham, D. Keenan, M. Deppman, S. Sauter, D. McCullough, C. Mebane, J. Lockwood, D. Essig, M. Hicks, D. Sturdevant, E. Materna, S. Spalding and J. Risley. 2004. "The Case for Regime-Based Water Quality Standards." *BioScience* 54(2):155-161.
- Shreve, R. 1966. "Statistical Law of Stream Numbers." *Journal of Geology* 74(1):17-37.
- Strahler, A.N. 1957. "Quantitative Analysis of Watershed Geomorphology." *American Geophysical Union Transactions* 38(6):913-920.

## Appendix A

# DISTRIBUTION OF *E. COLI* LEVELS AND RECREATION USE AS FACTOR OF STREAM ORDER IN THE CENTRAL GREAT PLAINS, CENTRAL OKLAHOMA/TEXAS PLAINS, AND SOUTH CENTRAL PLAINS ECOREGIONS

A Thesis

by

KHURRAM RAFI

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

MASTER OF SCIENCE

Chairs of Committee,	Ronald Kaiser
Co-Chair of Committee,	Kevin Wagner
Committee Member,	Terry Gentry
Committee Member,	Raghupathy Karthikeyan
Intercollegiate Faculty Chair,	Ronald Kaiser

May 2015

Major Subject: Water Management and Hydrological Science

Copyright 2015 Khurram Rafi

## ABSTRACT

This study examines the relationship of *E. coli* concentrations and recreational use to both stream order and watershed size. To determine possible ecoregion effects, the *E. coli* data used in this study were obtained from monitoring stations located on freshwater streams located in three ecoregions in Texas and Oklahoma – the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains (Ecoregions 27, 29, and 35, respectively). Median *E. coli* concentrations from the monitoring stations were analyzed for correlation with respect to the stream order of each monitoring site as well as the watershed size of each monitoring station. Geospatial analysis was used to determine stream order and watershed size and to identify un-impacted/least impacted streams in each ecoregion. Stream order was classified based on the traditional stream order classification method by Horton (1945) and Strahler (1957) and stream link magnitude analysis method by Shreve (1966). The analysis of two stream order systems and watershed size for each monitoring site with respect to median *E. coli* showed no significant relationship between *E. coli* and stream orders/watershed size of unimpacted watersheds. The watersheds with wastewater outfalls and urban areas exceeding 10% of land use showed a statistically significant, yet a weak negative relationship between *E. coli* and stream order/watershed size i.e. *E. coli* decreased with increase in stream order/watershed size.

## ACKNOWLEDGEMENTS

I would like to thank my committee chairs, Drs. Wagner and Kaiser, and my committee members, Drs. Gentry and Karthikeyan for their guidance and support throughout the course of this research. I would also like to thank Amanda Dube and Brian Pierce at IRNR for their valuable input.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.



## NOMENCLATURE

EPA	United States Environmental Protection Agency
IRNR	Institute of Natural and Renewable Resources – Texas A&M University
OCC	Oklahoma Conservation Commission
RUAA	Recreational Use Attainability Analysis
TCEQ	Texas Commission on Environmental Quality
USDA	United States Department of Agriculture

## TABLE OF CONTENTS

	Page
<a href="#"><u>ABSTRACT</u></a> .....	II
<a href="#"><u>ACKNOWLEDGEMENTS</u></a> .....	III
<a href="#"><u>NOMENCLATURE</u></a> .....	IV
<a href="#"><u>TABLE OF CONTENTS</u></a> .....	V
<a href="#"><u>LIST OF FIGURES</u></a> .....	VI
<a href="#"><u>LIST OF TABLES</u></a> .....	VII
 <a href="#"><u>Need for Study</u></a> .....	 1
<a href="#"><u>Stream Order Theory</u></a> .....	4
<a href="#"><u>Microbial Transport in Surface Waters</u></a> .....	7
<a href="#"><u>Objectives</u></a> .....	8
<a href="#"><u>Introduction</u></a> .....	10
<a href="#"><u>Description of Study Area</u></a> .....	11
<a href="#"><u>Data Sources</u></a> .....	14
<a href="#"><u>GIS Methods</u></a> .....	16
<a href="#"><u>Stream delineation and stream order calculation</u></a> .....	16
<a href="#"><u>Watershed and urban area calculation</u></a> .....	17
<a href="#"><u>Unimpacted watersheds</u></a> .....	18
<a href="#"><u>Materials and Methods</u></a> .....	20
<a href="#"><u>Results and Discussion</u></a> .....	25
<a href="#"><u>Analysis of median <i>E. coli</i> by Strahler stream order</u></a> .....	25
<a href="#"><u>Analysis of median <i>E. coli</i> by modified Shreve stream order</u></a> .....	36
<a href="#"><u>Analysis of median <i>E. coli</i> by Shreve stream order and watershed area</u></a> .....	46
<a href="#"><u>Discussion of results</u></a> .....	53
 <a href="#"><u>REFERENCES</u></a> .....	 64

## LIST OF FIGURES

	Page
<a href="#"><b>Figure 1.</b> Strahler stream ordering (a) and Shreve stream ordering (b).</a> .....	6
<a href="#"><b>Figure 2.</b> Study area comprising of ecoregions 27, 29, and 35 in Texas and Oklahoma.</a> .....	12
<a href="#">Figure 3. Location of monitoring sites and wastewater outfalls in the study area.</a> .....	16
<a href="#"><b>Figure 4.</b> Goodness of fit normality test for <i>E. coli</i> distribution.</a> .....	23
<a href="#">Figure 5. Comparison of median <i>E. coli</i> concentrations by Strahler stream order and ecoregions.</a> .....	29
<a href="#"><b>Figure 6.</b> Wilcoxon/Kruskall-Wallis test of <i>E. coli</i> by Strahler order at a) all monitoring sites, b) sites with up to 10% urban area, and c) unimpacted monitoring sites.</a> .....	32
<a href="#"><b>Figure 7.</b> Comparison of median <i>E. coli</i> by modified Shreve order at ecoregion scale.</a> .....	40
<a href="#"><b>Figure 8.</b> Wilcoxon/Kruskall-Wallis test of median <i>E. coli</i> by modified Shreve order.</a> .....	42
<a href="#">Figure 9. Nonlinear model fits for <i>E. coli</i> by Shreve stream order and watershed area.</a> .....	49
<a href="#">Figure 10. RUAA status of <i>E. coli</i> by Strahler stream order.</a> .....	58
<a href="#">Figure 11. RUAA status of <i>E. coli</i> by Shreve stream order.</a> .....	59
<a href="#">Figure 12. RUAA status of <i>E. coli</i> by watershed area.</a> .....	59

## LIST OF TABLES

	Page
<a href="#"><u>Table 1. Summary of <i>E. coli</i> by Strahler stream order for a) all monitoring sites, b) sites with less than 10% urban area, and c) unimpacted sites.</u></a>	26
<a href="#"><u>Table 2. Median <i>E. coli</i> concentrations by Strahler stream order and ecoregions for all monitoring sites.</u></a>	27
<a href="#"><u>Table 3. Median <i>E. coli</i> concentrations by Strahler stream order and ecoregions for sites with &lt;10% urban area.</u></a>	27
<a href="#"><u>Table 4. Median <i>E. coli</i> concentrations by Strahler stream order and ecoregions for unimpacted watershed sites.</u></a>	27
<a href="#"><u>Table 5. Strahler stream order pairs with equal score means using Wilcoxon each pair test.</u></a>	33
<a href="#"><u>Table 6. Wilcoxon/Kruskall-Wallis test probabilities by ecoregion.</u></a>	34
<a href="#"><u>Table 7. Correlation values between <i>E. coli</i> and Strahler stream order.</u></a>	35
<a href="#"><u>Table 8. Correlation values between median <i>E. coli</i> and Strahler stream order at ecoregion level.</u></a>	36
<a href="#"><u>Table 9. Modified Shreve stream orders for Wilcoxon/Kruskall-Wallis test (all monitoring sites).</u></a>	37
<a href="#"><u>Table 10. Summary of <i>E. coli</i> by modified Shreve stream order for all monitoring sites.</u></a>	38
<a href="#"><u>Table 11. Median <i>E. coli</i> concentrations by Shreve stream order and ecoregion for all monitoring sites.</u></a>	38
<a href="#"><u>Table 12. Median <i>E. coli</i> concentrations by Shreve stream order and ecoregion for &lt;10% urban area.</u></a>	39
<a href="#"><u>Table 13. Median <i>E. coli</i> concentrations by Shreve stream order and ecoregion for unimpacted watershed sites.</u></a>	39
<a href="#"><u>Table 14: Modified Shreve stream order pairs with equal score means using Wilcoxon each pair test.</u></a>	43
<a href="#"><u>Table 15. Wilcoxon/Kruskall-Wallis test probabilities by ecoregion.</u></a>	44
<a href="#"><u>Table 16. Correlation values between <i>E. coli</i> and modified Shreve stream order.</u></a>	45
<a href="#"><u>Table 17. Correlation values between median <i>E. coli</i> and modified Shreve stream order at ecoregion level.</u></a>	46

<a href="#"><u>Table 18. Correlation values between <i>E. coli</i> and Shreve stream order.</u></a>	51
<a href="#"><u>Table 19. Correlation values between <i>E. coli</i> and watershed area.</u></a>	51
<a href="#"><u>Table 20. Correlation values between <i>E. coli</i> and Shreve stream order at ecoregion level.</u></a>	52
<a href="#"><u>Table 21. Correlation values between <i>E. coli</i> and watershed area at ecoregion level.</u></a>	52
<a href="#"><u>Table 22. RUAA status of <i>E. coli</i> by Strahler stream order.</u></a>	58



# CHAPTER I

## INTRODUCTION

### **Need for Study**

Pathogen contamination in freshwater streams is a significant water quality concern in the United States. Pathogen contamination is often assessed by determining the amount of an indicator bacteria such as *Escherichia coli* (*E. coli*) in freshwater systems whereas enterococcus is used as an indicator in marine waters (Halliday and Gast 2011).

Pathogen contamination is by far the largest cause for water quality impairment with approximately 11,000 instances reported to date. Over 80 percent of the pathogen impairment is caused by excessive fecal coliform bacteria which include *E. coli* (USEPA 2014a). Diarrhea, gastrointestinal illness, and diseases such as cholera can be attributed to waterborne pathogens along with approximately 900 deaths and 900,000 illnesses each year (Arnone and Walling, 2007). This makes setting appropriate recreational use standards for water bodies very important. In Texas, the recreational use standard for a water body is assumed primary body contact unless determined otherwise after a Recreational Use Attainability Analysis (RUAA) has been conducted. The RUAA is a site-specific study to determine what uses a stream can and does support based on its current and historic use and its physical and flow characteristics (TCEQ 2014). An appropriate recreational standard for a water body can then be set for the water body based on the findings of RUAA and approval by EPA.

The Texas Commission on Environmental Quality (TCEQ) is responsible for development and implementation of the Texas Water Quality Standards including the recreational use standards as specified in the Texas Administrative Code (TAC). The contact recreation standards with respect to *E. coli* concentrations as listed in the TAC regulations are as follows (TCEQ 2010):

I. Primary contact recreation – activities involving significant risk of water ingestion (e.g. swimming, diving, surfing, etc.). The geometric mean criterion for *E. coli* is 126 per 100 ml. In addition, the single sample criterion for *E. coli* is 399 per 100 ml.

II. Secondary contact recreation 1 – commonly occurring activities with limited body contact incidental to shoreline activity (e.g. fishing, kayaking, rafting, etc.). The geometric mean criterion for *E. coli* is 630 per 100 ml.

III. Secondary contact recreation 2 – activities with limited body contact incidental to shoreline activities with less significant risk of water ingestion than secondary contact recreation 1. The geometric mean criterion for *E. coli* is 1,030 per 100 ml.

IV. Noncontact recreation – activities with no significant risk of ingestion. The geometric mean criterion for *E. coli* is 2,060 per 100 ml.

As previously stated, in Texas, all fresh water bodies are classified as primary contact recreation unless a RUAA study determines a change in the classification is required (TCEQ 2010). The applicability of these standards has been questioned especially related to scale. Regime based standards, which account for temperature,

sediment, and stream order of the water bodies have been proposed as possible alternatives (Poole et al., 2004). A central Texas study conducted by Harmel et. al. (2010) established a trend of decreasing *E. coli* concentrations when aggregating sampled segments from upstream to downstream. The trend was visible in both the “impacted” (land use of mixed rural with known point sources of pollution such as dairies, waste water treatment plants (WWTP), small communities, etc.) and “unimpacted” (land use of mixed rural excluding point sources of pollution) streams of the study area. The study also found that with increasing watershed scale the median *E. coli* concentrations decreased. Furthermore, the study found “unimpacted streams to be in violation of TAC regulations even though the anthropogenic impact was insignificant.

A study conducted by Lyautey et. al. (2010) in the South Nation River Drainage basin of Canada compared *E. coli* concentrations using Shreve (1966) stream link. The study found Shreve stream link to be inversely correlated to *E. coli* concentrations with strongest correlations seen during summer. Another Canadian study by Edge et al. (2012) used Horton (1945) and Strahler (1957) stream order and found that streams with order 3 and lower had higher *E. coli* concentrations compared to the streams of order greater than 3. Another study (Lyautey et. al. 2011) explained, “the persistence factor is less important than dilution (i.e. stream order) in describing *E. coli* densities, followed by factors that influence the loading of *E. coli* into watersheds.”

Conversely, some studies dispute the inverse correlation between stream order and *E. coli* concentrations and suggest other factors (i.e. nonpoint source inputs) as more vital. A study conducted by the Virginia Department of Environmental Quality (VDEQ)

found no correlation to stream order or ecoregions (level III) when investigating fecal coliform bacteria. The report found highest fecal coliform concentrations at the stream orders 3 and 4. The report suggested use of *E. coli* in future study as a better indicator for pathogen contamination (VDEQ 2003). Another study by Byappanahalli et. al. (2003) found *E. coli* increasing steadily downstream in the Dunes Creek watershed in Indiana. The study found a relation between excessive ditching and consequent increase in non-point source input responsible for increase in *E. coli*.

Lyautey et. al. (2010, 2011) attribute the inverse relationship of *E. coli* and stream order to the dilution in the water body as smaller stream orders combine to make larger streams. However, other studies hypothesize an influence of physical stream processes such as turbidity/sediment process and flow rates, on the *E. coli* concentrations, (Bai and Lung 2005; Brettar and Hofle 1992; Craig et. al. 2004; Oliver et. al. 2007) which may be processes dependent on stream order as well. This study investigates the relationships between stream order and *E. coli* concentrations of unimpacted and impacted watersheds in the Texas and Oklahoma ecoregions 27 (Central Great Plains), 29 (Central Oklahoma/Texas Plains), and 35 (South Central Plains).

### **Stream Order Theory**

The relative position of a stream within a stream system can provide valuable information regarding the stream segment and help to explain similarities and differences within a stream network. Stream order, a measure of the relative size of streams, can be easily associated to the drainage area, stream size, and other physical properties of the streams. Classifying a stream according to stream order provides an

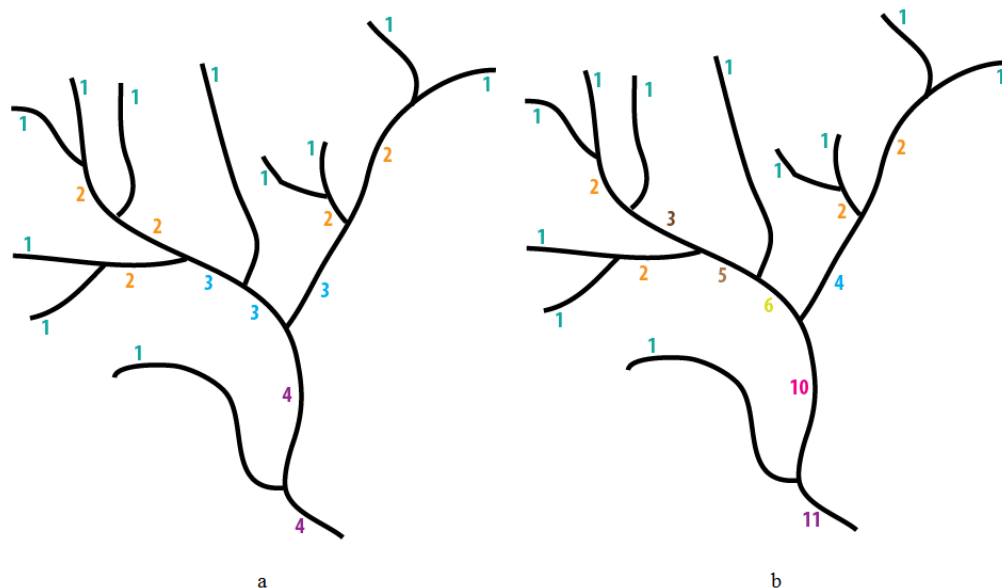
important measure for the size and characteristics of specific segments within the stream network, which is essential for water management. Classifying stream order can also help in studying the amount of sediment in an area and facilitate a more effective use of waterways as natural resources. Ichim (1987) noted a strong relationship of decreasing sediment due to runoff with increasing stream order in easily erodible rocks and soils whereas a more moderate relationship was seen when the sediment source was less erodible. In lower order streams, the erosion of soils is usually more dominant than the deposition, therefore making stream order a factor in the sampling of the sediment (Otteson & Theobald 1994).

Gravelius (1914) as quoted by Horton (1945) was one of the first scientists to propose stream ordering. According to Gravelius, the source to mouth segment of the river (i.e. the segment that does not become a tributary of another stream and drain into a larger water body) has the first order since it is the greatest collector of the water. The direct tributaries of the first order stream have the order of two; the direct tributaries of the second order have the stream order of three; and so forth. The Gravelius method did not account for the length, catchment area, or flow of the streams and was subjective in nature.

Horton (1945) modified Gravelius subjective and somewhat difficult method and presented a more objective and simplified method of stream ordering in a dimensionless way. Horton reversed Gravelius ordering method and assigned the stream order of one to the stream with the initial concentrated flow. Meeting of two first order streams leads to a second order stream and the order of the stream will stay two as long as all of its



tributaries have an order of one. Similarly meeting of the two-second order streams results in a 3<sup>rd</sup> order stream and it will stay a 3<sup>rd</sup> order stream as long as all of its tributaries have an order of two or less. This ordering of streams will continue until the mouth of the stream is reached.



**Figure 1.** Strahler stream ordering (a) and Shreve stream ordering (b).

In 1952, Strahler created his own system based on Horton's theory. The Strahler classification system (Strahler 1952) assigned a stream order of one to the basic water course and increased the order as theorized by Horton (1945) where the streams order does not change if a stream receives a lower order tributary (Figure 1). Zăvoianu (1985) noted the similarity in catchment area, average length of the water network, average slope, average flow rates, and other characteristics of streams when they had the same stream orders.

The Shreve classification system (Shreve 1966) provides a more representative depiction of the basin size by taking into account all tributaries of a stream system. In this method, all of the streams are accounted for as links in a network. Similar to the Strahler method, all of the basic watercourses considered exterior links are assigned an order of one. For order of all the other interior links, the orders of the links intersecting to make that interior are added together. This means if an interior link consists of the intersection of two first order streams links than it will have an order of two. Similarly, if a stream link of order 2 intersects with a stream link of order 5 the resulting link will have an order of 7. Due to the additive nature of the Shreve classification method, the stream numbers are sometimes referred to as magnitude instead of orders (Tarboton et. al. 1991). Therefore, the magnitude of a stream link in the Shreve ordering system is the number of all the links upstream to it.

### **Microbial Transport in Surface Waters**

There are many factors that affect microbial transport in surface waters. Microbes may be found in the water column or within the sediments associated with the surface waters. These microbes may be unattached, attached, or resuspended within the water body. The unattached microbes do not get absorbed in the sediment and flow with velocities similar to the water since they are osmotically similar. Due to their electro-potential attractions, these unattached microbes bond easily with other particles that decrease their buoyancy while increasing their settling rates. On the other hand, attached microbes may be surficial in nature with a weak bond or may become completely absorbed into another particle. Detaching microbes from the absorbed particles is very

difficult and often requires sediment removal for complete decontamination (Berry 1991).

The unattached microbes are less likely to deposit in the streambed and become part of the sediment whereas the attached microbes will deposit at the same rate as their host particle. The attached microbes deposited in the sediment bed may resuspend due to natural processes such as high flows due to flooding. This resuspension may cause the microbes to dislodge from host particles and flow unattached or attach to another particle and resettle once the resuspension event is over (Pachepsky and Shelton 2011). The attachment rates for different microbes vary significantly. A study done by Krometis et. al. (2007) found attachment rates of approximately 40% for the fecal coliforms, *E. coli*, and Enterococci. Krometis also concluded that these attachment rates were the same through the storm events and the resuspension rates of attached particles were higher during the storm events.

### **Objectives**

The overall goal of this research is to better understand and predict the relationships between stream orders, *E. coli* concentrations, and recreational use and determine whether current recreational use standards are suitable.

The specific objectives are to:

1. Assess stream order and watershed size of all sites used in the study using GIS. Various stream order classification systems will be evaluated for suitability including traditional stream order classification (Horton 1945; Strahler 1957) and stream link magnitude analysis (Shreve 1966).

2. Evaluate correlations between concentrations and stream order (plus watershed size) as impacted by ecoregion
3. Evaluate correlation between recreation standard and stream order/watershed size.

## CHAPTER II

### STUDY AREA, DATA COLLECTION, AND GIS METHODS

#### **Introduction**

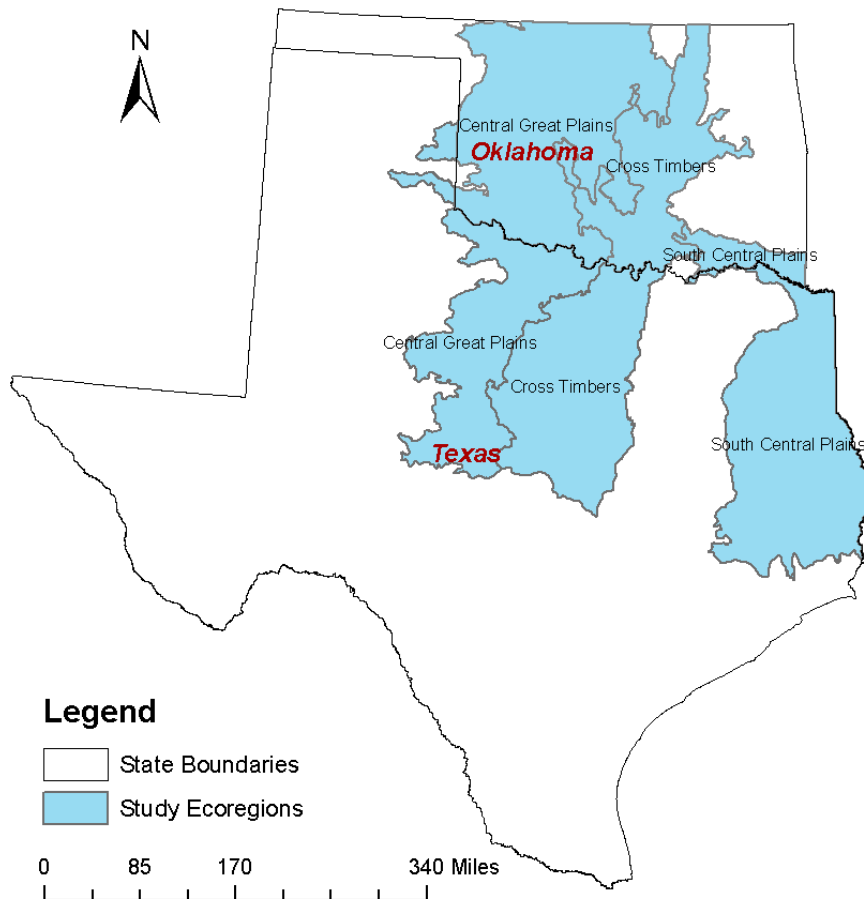
The purpose of this chapter is to provide a brief description of the study area as well as to detail the data sources used in this study. GIS steps undertaken such as delineation of streams, calculation of stream orders and watershed area, and determination of unimpacted watersheds are also described. Study area chosen for this study includes parts of ecoregions 27 (Central Great Plains), 29 (Central Oklahoma/Texas Plains also known as Cross Timbers), and 35 (South Central Plains) located within the states of Texas and Oklahoma. With the exception of some large population urban areas such as Dallas and Fort Worth in Texas and Oklahoma City in Oklahoma most of the study area is of rural to semi-rural in nature. The large rural area allowed for a greater selection of unimpacted streams and watersheds within the study area. Selecting the study area at ecoregional scale also allowed for a larger number of monitoring stations for the study. This was important since a significant number of the rural monitoring sites do not collect *E. coli* data regularly and would not be included in the analysis due to lack of *E. coli* data. A smaller number of monitoring sites would have therefore curtailed the *E. coli* data required for analysis in this study. Collecting data at ecoregion scale also provides an opportunity to compare the *E. coli* relationships in these ecoregions as well. Many state resource management agencies including Texas agencies have used ecoregions when setting up water quality standards and non-point pollution management goals (Omernik and Bailey 1997). The ecoregional comparison will also

help in understanding *E. coli* relationships in the selected ecoregions and their comparison to the state standards.

### **Description of Study Area**

The three ecoregions selected for this study are shown in Figure 2. Ecoregion 27 is the most westward of the three ecoregions. The average temperatures in the region range from a low of minus eight and high of 13 °C during January and a low of 19 and high of 37 °C during July. The temperature increases moving southwards in the ecoregion. The vegetation in the ecoregions was once a mostly mixed-grass prairie but has evolved into mostly cropland today. Range and grassland are mostly located in more rugged areas (Griffith et. al. 2004). The terrain of the region mainly consists of irregular plains and broad alluvial valleys. The hydrology of the area includes mostly intermittent streams with some perennial streams and a few natural lakes. There are some large rivers such as Red and Brazos River, which pass through the region as well. The larger rivers have braided sandy channels with mostly turbid flows. After heavy rains, the streams have stronger flows with large amounts of suspended sediments. The streams draining from rangeland contain less sediment load than the streams downstream of croplands. Many streams in the region have been channelized and/or impounded leading to unnatural flow regimes, higher erosion, and loss of the riparian forests. The precipitation in the area ranges from 560 to 965 mean annual millimeters (mm) and increases from west to east and the region is mostly semi-arid in nature. The wildlife in region mostly consists of white-tailed deer, mule deer, pronghorn, coyote, and jackrabbit. The land use in the region is dominated by dryland and irrigated cropland with some pasture and

rangeland also occurring. Grazing for cattle, sheep and goat along with oil and gas production is also a common land use. (Griffith et. al. 2004; Woods et. al. 2005).



**Figure 2.** Study area comprising of ecoregions 27, 29, and 35 in Texas and Oklahoma.

Ecoregion 29 occurs in the north-central Texas, central Oklahoma, and southeastern Kansas and is located between ecoregions 27 and 35 the other two ecoregions included in this study. The region is transitional area between winter wheat growing areas in the ecoregion 27 and the forested low mountains of eastern Oklahoma.

Vegetation is mainly comprised of forest, woodland, savannah, and prairie. Cropland is not very common and is restricted to valleys near channelized streams. Both intermittent and perennial streams can be found in the region, which have a low to moderate gradient due to the rolling plains type terrain found in the region. Several large rivers cross the ecoregions along with some large reservoirs and lakes to serve the urban areas located in the region. The temperature in the region ranges from a low of minus six and high of 14 °C during January and a low of 21 and high of 36 °C during July. The precipitation in the region ranges from 780 to 1170 mean annual mm with higher precipitation occurring on the eastern side of the region. Wildlife in the region mainly consists of white-tailed deer, gray fox, bobcat, black-tailed jackrabbit, and prairie chicken etc. Main landuse in the area is pastureland and rangeland. Areas of woodland are also dominant, along with some cropland. Oil and gas production is a major landuse and the region also includes some major urban centers such as Dallas, Fort Worth, and Arlington in Texas. (Griffith et. al. 2004; Woods et. al. 2005).

Ecoregion 35 is the eastern region in the study area. The region comprises mainly a temperate coniferous forest with several species of pine along with hardwoods such as hickory and oak. Approximately one sixth of the region consists of cropland mainly within the Red River floodplain. Perennial streams are most common in the region but the flow can become limited during the summer months. The streams located in the forested regions normally have lower concentrations of suspended solids whereas the Red River the largest river flowing through the region is mostly turbid. The region lacks natural lakes but some reservoirs have been built within the region to account for



inconsistent flow. The average temperatures range from a low of minus two and high of 16 °C during January and a low of 21 and high of 34 °C during July. The precipitation ranges between 1066 to 1422 mean annual mm with higher rainfall towards the east. The wildlife mainly consists of white-tailed deer, coyote, beaver, raccoon, muskrats, and rabbits. Major land uses include commercial pine plantations, timber production, livestock grazing, and oil and gas production. (Griffith et. al. 2004; Woods et. al. 2005).

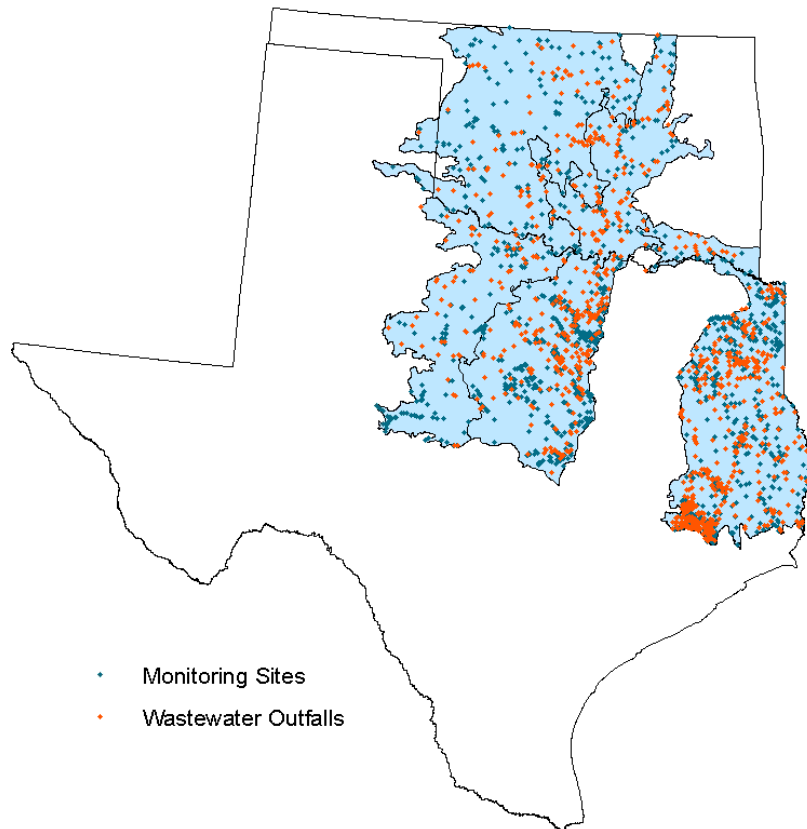
The three ecoregions have very different characteristics as detailed above. Major differences include the amounts of precipitations and temperatures as well as the hydrology of each region. The type of vegetation changes as we move eastward from dryland and irrigated cropland in ecoregion 27 to a predominantly coniferous forest in ecoregion 35. The analysis of *E. coli* relationship with stream orders and watershed area by ecoregion will be an important part of this study to assess if these differences in characteristics have any bearing on the *E. coli* relationships.

### **Data Sources**

Where available the data for the study ecoregions was collected for each monitoring station location in the study area. All of the *E. coli* and monitoring sites data for this study was collected from TCEQ (TCEQ 2014, updated daily) for the Texas portion of the ecoregions and the OCC for the Oklahoma portion of the ecoregions. The *E. coli* reported by TCEQ was in the units of MPN/100ml (most probable number/100 milliliter). TCEQ data was collected from the years of 2006 to 2014. Similarly, the OCC data from 2006 to 2013 was used. The most recent 7 years data was used in the study to ensure that the data is representative of the current hydrologic and environmental

condition. Hydrologic regimes, precipitation patterns, and land use practices all change over time and can have a big impact on both the *E. coli* concentrations and its transport through the streams. Use of recent data allows for investigation of current conditions and ensures that historic conditions do not affect the results. The *E. coli* observations for each monitoring site ranged from five to 112. On average, there were 28 observations per monitoring site and median number of observations was 22. Out of the final 742 monitoring sites selected for this study, 74 sites had less than 10 *E. coli* observations available while 245 monitoring sites had more than 28 observations.

TCEQ provided the ID's of monitoring sites along with the *E. coli* data. These site ID's were matched with the TCEQ GIS layer for monitoring sites (TCEQ 2014a) to conduct the GIS steps of this study. OCC provided the longitude and latitudes for each of their monitoring sites, which were used to determine the geographic location of the sites for the GIS steps. Figure 3 shows the geographic location of the monitoring sites and wastewater outfalls in the study area. The ecoregion boundaries were obtained from the EPA website (USEPA 2014b). The flow direction and flow accumulation raster's based on 30-m digital elevation models (DEMs), which are provided in the NHDPlus Version 2, were obtained from Horizons Systems Corporation (Horizon 2014). The land cover/land use data for the watersheds was obtained from the USDA National Agricultural Statistics Service (NASS) 2013 Cropland Data Layer from USDA CropScape (NASS 2014). Wastewater outfall locations were obtained from TCEQ (TCEQ 2014a) for Texas regions and from EPA (USEPA 2014c) for Oklahoma regions.



**Figure 3.** Location of monitoring sites and wastewater outfalls in the study area.

## GIS Methods

### *Stream delineation and stream order calculation*

Streams were delineated using the ArcHydro add-in for ArcMap. The flow direction and flow accumulation raster's obtained from Horizon were used as inputs in the ArcHydro stream definition tool. The stream threshold was selected at 5000 30 x 30 m cells (an area of 4.5 km<sup>2</sup>) for catchment definition, which is based on the USGS Elevation Derivatives for National Applications (EDNA 2014). The threshold determines the density of the stream network. For example, a smaller threshold will

result in a denser stream network and a larger threshold will result in a less dense stream network. The stream raster is then used in the stream segmentation and drainage line processing tools to get the stream vectors. The monitoring stations locations are added from TCEQ for Texas regions and from the latitude and longitude provided by OCC in ArcMap. The Monitoring stations are snapped using the Snap Pour Point tool to the stream segments if they are within 200 meter of the stream segment. This slightly moved the monitoring location site so that stream order could be extracted. Not all monitoring sites were snapped to streams using a 200-meter distance and those sites were excluded from statistical analysis.

#### *Watershed and urban area calculation*

Using the watershed tool in the Spatial Analyst toolset in ArcMap, the watershed raster's are delineated for each monitoring station point. The flow direction raster and the snap pour points are used as inputs in the watershed tool to get the watershed area. Once the watershed raster's are calculated, the Raster to Polygon conversion tool in ArcMap is used to get vector polygons for each watershed.

Urban area polygons are determined from the cropland GIS layer (NASS 2014) and are delineated for each watershed using the clip tool in ArcMap. The percentage of each watershed's urban area is then determined by dividing the area of the urban area within each watershed by the total watershed area and added to the attributes of the monitoring station layer. The wastewater outfall layers obtained from TCEQ and EPA were added to ArcMap at this point. The count of wastewater outfalls in each watershed was added to the attributes of the monitoring station shapefile. The attribute table of the

monitoring station layer, which includes urban area percentages, watershed area in square meters (m), wastewater outfall counts, and the Shreve and Strahler stream orders, is exported to MS Excel for further analysis.

The stream order is obtained using the spatial analyst toolset in the ArcMap. Stream order tool in the spatial analyst menu is applied to determine the stream order of each flow line segment in the stream raster using flow accumulation and flow direction rasters. The tool allows calculation of a stream order raster based on both Strahler and Shreve stream order systems. Once the stream order raster is calculated, the Raster to Poly line tool is used to convert the stream order raster into a stream vector shapefile. This shapefile contains the stream order for each segment in its attributes. Stream vector is then joined with the monitoring stations to get the stream order of each monitoring site.

### *Unimpacted watersheds*

The criterion for unimpacted watersheds was based on the percentage of urban areas and the number of wastewater outfall. Since the volume discharge of each wastewater outfall was not readily available from EPA, only the watersheds that did not have a wastewater outfall were selected as unimpacted. This lead to 278 monitoring sites with no wastewater outfall in their watersheds. To identify the rural watersheds, only the watersheds that had a maximum of 10 percent urban area were selected in the unimpacted classification. This classification was made after considering other thresholds of 50, 25, and 5 percent urban areas and conducting the statistical tests discussed in Chapter III. The statistical tests showed that urban area percentage did not

have a major impact on the significance of the results therefore the 10 percent threshold was chosen as representative of below which an area would be considered rural. This classification lead to a final data with 252 monitoring sites which also was not a significant reduction from the 278 sites that did not have a wastewater outfall in the their watersheds.

## CHAPTER III

### STATISTICAL ANALYSIS

#### **Materials and Methods**

This chapter will discuss the statistical methods undertaken to evaluate relationships between *E. coli* and the two stream order methods and *E. coli* and watershed areas, which were obtained from the GIS methods in Chapter II. *E. coli* data was collected by TCEQ and OCC over the 7-year study period at irregular intervals and at different flow condition for many of the monitoring sites. This can affect the *E. coli* concentrations measured in the stream since very low and very high concentrations can be result of specific conditions such as floods or droughts in the area or landuse practices that are time sensitive. Median *E. coli* would be representative of the normal conditions of the stream eliminating the too low or too high concentrations, which may occur due to other than normal conditions and was therefore used as the metric in this study. Average *E. coli* measurements were also considered but not used as many of the sites in the study had less than 10 observations and a small number of high or low observations could easily skew the average, which would not be representative of the normal stream conditions. Another measure that would be appropriate for similar analysis but was not considered for this study is the geometric mean, which is the  $n$ th root of the product of  $n$  numbers. Geometric mean also reduces the effect of very low and very high numbers and could be an appropriate measure for a future study of similar nature.

For calculation of median *E. coli*, monitoring stations data was first sorted based on the *E. coli*'s date of observation. Only the monitoring stations that had minimum of

five *E. coli* observations during the study period of January 2006 to May 2014 were included in the analysis. Cursory analysis of sites that had less than five observations showed that those observations were collected either too sporadically or in a short frame of time. These sites were also impacted by high and low concentrations more severely, which made the use of median inappropriate. Therefore, a cutoff was made at five *E. coli* observations per site to calculate median *E. coli*, which would be better representative of the normal stream conditions. Using the median function in MS Excel, the median for each monitoring station was obtained. The data obtained from GIS methods, which included stream orders (both Strahler and Shreve), watershed area, urban area percentages, ecoregion number, and the wastewater outfall count for each monitoring site, was joined with *E. coli* data. The monitoring stations, which did not snap to 200 meters of a stream segment or lacked sufficient *E. coli* data, were eliminated from analysis. This resulted in a final database of 742 monitoring stations for the study regions.

The data included a maximum median *E. coli* concentration of 2400 MPN/100ml with next largest value of 1000 MPN/100ml. The maximum *E. coli* concentration occurred at monitoring station number 10786 in ecoregion 29. The monitoring station had six *E. coli* observation all made between 11/22/2011 to 05/22/2012 with four observations that had concentrations of 2400 MPN/100ml or more. Since all of the *E. coli* observations were made within 7 months and had high concentration values it was decided to eliminate this monitoring site from analysis as the period when observations



were collected was too small and concentrations measured may not have been representative of the true conditions at the site.

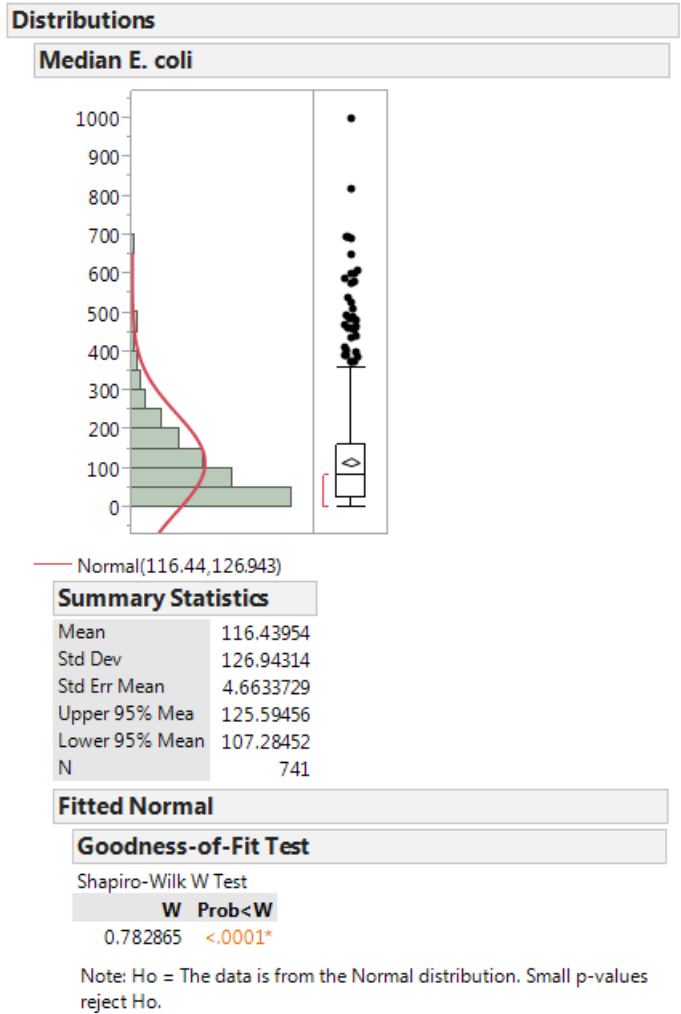
From this database, two subsets were extracted. For the first subset only those monitoring sites were selected where the urban area percentage was 10 percent or less. In the second subset, monitoring sites with wastewater outfalls in their watersheds were also removed in addition to limiting the urban area at less than 10 percent, which is the unimpacted watersheds definition for this study. The following three datasets were then analyzed using statistical techniques:

- a) All sites with at least five observations;
- b) Sites with 10 percent or less urban area in their watersheds;
- c) Unimpacted watershed sites with 10 percent or less urban area and no wastewater outfalls in their watersheds.

The first two datasets are referred to as impacted with wastewater outfall (WWTF) datasets in this study. All three data sets were analyzed in the JMP statistical software (JMP<sup>®</sup>, Version 11) according to procedures described by Helsel and Hirsch (2002) and Haan (2002).

The *E. coli* data was first analyzed in its entirety without the consideration of impacted or unimpacted monitoring sites to determine the suitability of the tests. The analysis of *E. coli* showed that it was not normally distributed as shown in Figure 4 below. The p-value for the Shapiro Wilk test for the goodness of a normal fit was less than 0.0001, which signifies that the data is not normally distributed. The box plot analysis in Figure 4 shows most of the *E. coli* observations over 400 MPN/100ml are

plotted as outliers. Similarly lognormal and exponential distribution goodness of fit tests showed p-values of 0.01, which signified a lack of fit for these distributions as well. Based on these results the non-parametric tests are most suitable for this dataset.



**Figure 4.** Goodness of fit normality test for *E. coli* distribution.

The non-parametric analysis of the data was conducted using Wilcoxon/Kruskall-Wallis tests for Strahler and modified Shreve stream orders with a null hypothesis of

mean of each monitoring sites median *E. coli* being equal at each stream order. The alpha threshold of 5 percent was used to determine the significance of the probability values obtained throughout the analysis. The same analysis for each stream order method was conducted at ecoregion scale as well. Further, each stream order method was tested for correlation between stream order and *E. coli* concentration using Kendall  $\tau$  and Spearman's  $\rho$ .

Continuous nonlinear tests were used to evaluate if a nonlinear distribution would fit the *E. coli* data with respect to Shreve order and watershed size since they represent continuous variables. Data distributions modeled using nonlinear fit allow testing of the correlation using the parametric Pearson's  $r$  coefficient. Non-parametric Kendall  $\tau$  and Spearman's  $\rho$  tests were used to determine correlation between *E. coli* and watershed size/Shreve stream order when nonlinear models were not applicable. Kendall  $\tau$  and Spearman's  $\rho$  are monotonic correlation tests and apply well to the data similar to this study. These tests are also resistant to the effect of outliers. The Kendall's  $\tau$  approaches a normal distribution more rapidly than  $\rho$ , as  $N$ , the sample size, increases.  $\tau$  is also a rank based test and is therefore more resistant to small number of unusual values (Helsel and Hirsch 2002).

The correlation coefficient values range from -1 to +1. The values closer to  $\pm 1$  indicate that the variables are highly correlated and the values closer to zero indicate they are not correlated. Negative correlation signifies that the variables are inversely correlated such that one variable decreases when the other increases. Conversely, a positive coefficient means the variables are moving in the same direction. The tests also

provide a probability value for the correlation co-efficient, which is compared at the 5% alpha threshold. If the p-value is less than 0.05 then the test is significant and there is correlation present between the x and y variables.

## **Results and Discussion**

### *Analysis of median *E. coli* by Strahler stream order*

#### **Data distribution**

The Strahler stream order for the study areas ranged from one to seven for all data. However, when only unimpacted watersheds are considered, the range is reduced to orders one to five due to reduction in the sampled data. Table 1 summarizes the number of monitoring site at each stream order for the three datasets analyzed in this study and the median *E. coli* concentration for those sites. The table shows a consistent *E. coli* distribution for all three datasets with increasing *E. coli* for the first three orders and a drop in the concentrations for later orders. Study conducted by Edge et. al. (2012) had found similar relationships between stream order and *E. coli* concentrations where the concentrations increased for the first three orders and then decreased. Another study by VDEQ (2003) had found the highest fecal coliform concentration at orders 3 and 4, which can also be seen in the table below for datasets with limited urban areas. For unimpacted watersheds, it is significant to note that the median *E. coli* concentration increases initially and then remains constant. The median *E. coli* for each of the three datasets stays below the geometric mean standard of 126 cfu/100ml required for primary contact recreation in Texas but approaches the standard limit at stream order 3.

**Table 1.** Summary of *E. coli* by Strahler stream order for a) all monitoring sites, b) sites with less than 10% urban area, and c) unimpacted sites.

Strahler stream order	a) All monitoring sites		b) <10% Urban area		c) Unimpacted watersheds	
	Number of monitoring sites	Median <i>E. coli</i>	Number of monitoring sites	Median <i>E. coli</i>	Number of monitoring sites	Median <i>E. coli</i>
1	34	91.75	22	65.25	22	65.25
2	144	95.00	113	80.00	82	80.00
3	214	106.50	194	105.00	99	100.00
4	175	93.00	171	90.00	39	100.00
5	106	62.00	106	62.00	10	102.25
6	30	28.00	27	26.00		
7	38	21.25	37	20.50		
Total	741	81.00	670	78.00	252	93.75

Analysis of *E. coli* at ecoregion scales reveals differences in the *E. coli* concentration in each ecoregion. Table 2, 3, and 4 provide a breakdown of the distribution for each ecoregion. Ecoregion 27 shows consistently higher median *E. coli* concentrations than the other two ecoregions. The differences in median *E. coli* between ecoregions 29 and 35 are not as significant but at higher stream orders ecoregion 35 has lower *E. coli* than the other two ecoregions. Ecoregion 27 is also the only ecoregion that shows a consistently decreasing trend in *E. coli* concentrations as stream order increases for all three datasets. This trend is similar to what has been found by Harmel et. al. (2010) and Lyautey et. al. (2010) in their studies. The median *E. coli* concentrations in ecoregion 27 are also higher than the geometric mean standard of 126 cfu/100ml required for primary contact recreation for the first four stream orders in all three datasets.

Ecoregion 29 has a decreasing trend for median *E. coli* concentration for all data similar to ecoregion 27. However, for less than 10 percent urban area and unimpacted

datasets there is increasing *E. coli* concentrations for first few orders and then decreasing concentrations. Ecoregion 35 shows an increasing trend at first and then a decreasing one for all datasets. Both ecoregion 29 and 35 have median *E. coli* values much lower than the 126 cfu/100ml required for primary contact recreation.

**Table 2.** Median *E. coli* concentrations by Strahler stream order and ecoregions for all monitoring sites.

Strahler stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	2	22	10	293.75	92.50	66.50
2	16	65	63	192.50	79.50	104.00
3	59	66	89	155.00	80.75	110.00
4	54	59	62	135.00	61.00	89.00
5	31	45	30	70.00	62.00	34.50
6	19	1	10	30.50	130.00	25.00
7	5	27	6	37.00	20.00	21.25

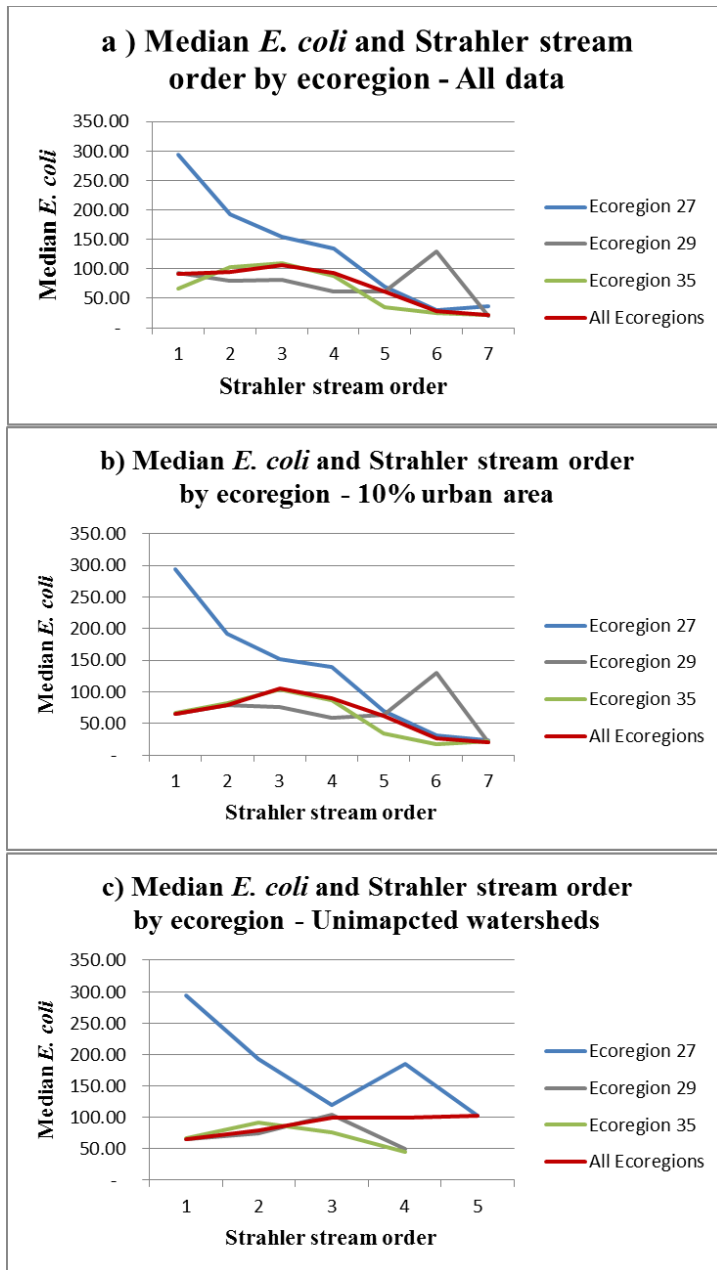
**Table 3.** Median *E. coli* concentrations by Strahler stream order and ecoregions for sites with <10% urban area.

Strahler stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	2	12	8	293.75	65.25	66.50
2	16	53	44	192.50	79.00	82.75
3	58	56	80	152.50	76.00	104.00
4	55	58	58	140.00	58.50	87.00
5	30	46	30	69.50	64.50	34.50
6	20	1	6	30.75	130.00	17.50
7	4	27	6	23.50	20.00	21.25

**Table 4.** Median *E. coli* concentrations by Strahler stream order and ecoregions for unimpacted watershed sites.

Strahler stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	2	12	8	293.75	65.25	66.50
2	14	39	29	192.50	75.00	92.50
3	40	25	34	120.00	104.00	76.00
4	23	9	7	185.00	50.00	45.00
5	10			102.25		

Figure 5 shows a graphic comparison of *E. coli* at different stream orders for all datasets by ecoregions. The figure shows an increasing *E. coli* trend for ecoregions 29 and 35 in the first three stream orders and then a decreasing trend for remaining orders. Ecoregion 27 shows a decreasing trend only from lower to higher stream orders. For ecoregion 27, there is an increase in *E. coli* at order 6, which can be attributed to a lack of observations, as there is only one monitoring station available with *E. coli* data.



**Figure 5.** Comparison of median *E. coli* concentrations by Strahler stream order and ecoregions.



### Wilcoxon/Kruskall-Wallis test

Strahler stream order follows discrete variable characteristics and along with non-normality of the *E. coli* data, the Wilcoxon/Kruskall-Wallis test is the most suitable nonparametric test for this method.

Figure 6 shows the results of the tests for the three datasets below. The tests were run using JMP software (JMP®, Version 11) using JMP software procedures outlined by Schlotzhauer (2007).

The Wilcoxon/Kruskall-Wallis method tests the null hypothesis that the mean ranks of observations at each stream order are equal i.e.:

- Null hypothesis ( $H_0$ ):  $\mu(\text{stream order } 1) = \mu(\text{stream order } 2) = \dots = \mu(\text{stream order } 7)$
- Alternate hypothesis ( $H_a$ ): at least two means are different.

In this study, the observations are the median *E. coli* calculated for each monitoring site to which Wilcoxon/Kruskall-Wallis method will be applied. For example in

Figure 6a the level column represents the stream order, the count column represent the number of monitoring sites with median *E. coli*, score sum and expected score columns are computed based on the test formula using the median *E. coli*, and score mean is score sum divided by the count. This score mean is the statistics that is compared by the test to calculate the significance probabilities. The test results are deemed significant, i.e. at least two score means are different, if the Chi Square p-value is less than  $\alpha = 0.05$ . The p-value for the WWTF impacted datasets are less than  $<0.0001$

from which it can be concluded that there are at least two stream orders with different means. Whereas the p-value of 0.7848 for unimpacted watersheds is not significant and it can be concluded that, the score means for all stream orders are equal. The plots in

Figure 6 show the median *E. coli* plotted on the Y-axis and Strahler stream order on X-axis. The median *E. coli* for each monitoring site in the analysis is plotted as dots in the figure. The straight grey line horizontal to the x-axis is the grand mean of each dataset and the blue line is connecting the mean of each stream order's median *E. coli*.

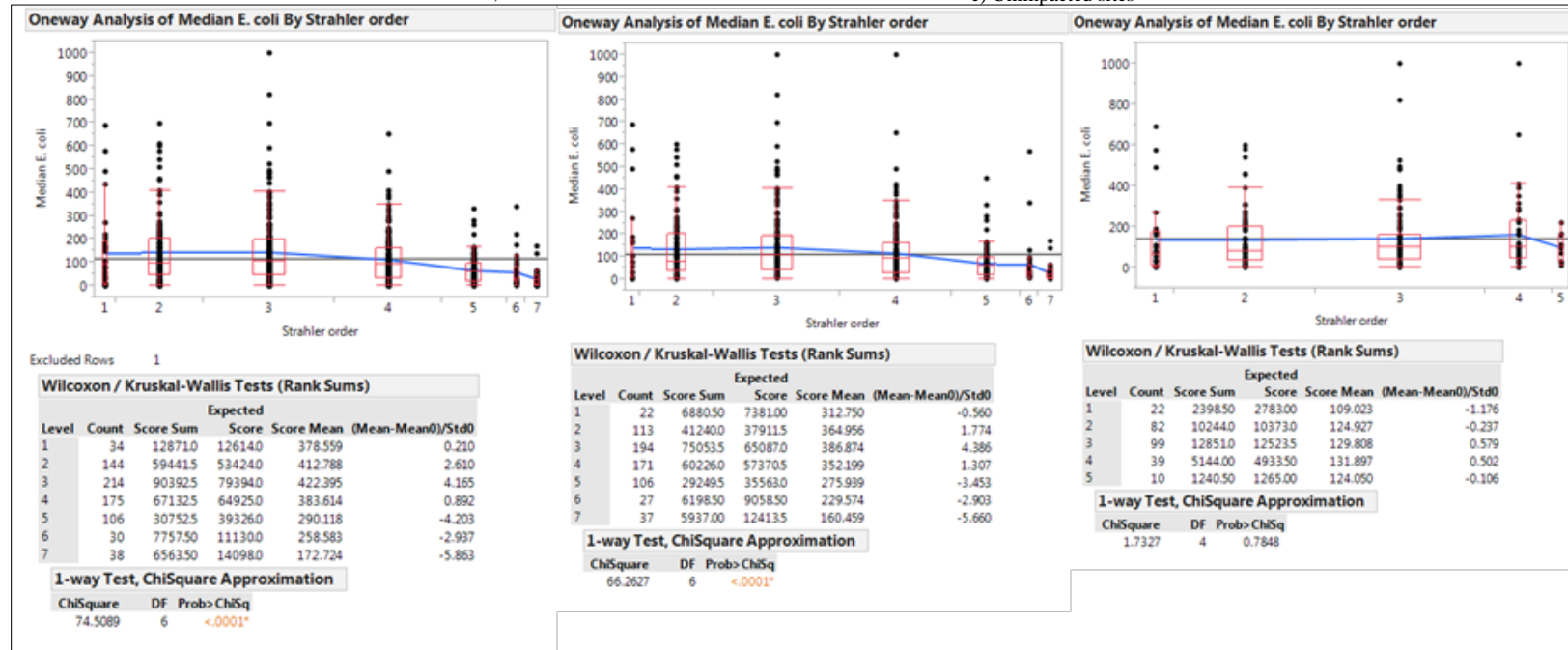
Further analysis for each pair of stream orders was conducted using nonparametric comparisons for each pair using the Wilcoxon method. The comparison of each pair showed that most of the pairs of stream orders for impacted datasets had different score means at 5% alpha as seen in Table 5. Whereas all pairs for unimpacted sites showed score means were statistically equal using the 5% alpha threshold.



a) All monitoring sites

b) Sites with less than 10% urban area

c) Unimpacted sites



**Figure 6.** Wilcoxon/Kruskal-Wallis test of *E. coli* by Strahler order at a) all monitoring sites, b) sites with up to 10% urban area, and c) unimpacted monitoring sites.

**Table 5.** Strahler stream order pairs with equal score means using Wilcoxon each pair test.

a) All monitoring sites			b) <10% Urban sites			c) Unimpacted sites		
Stream order	Stream order	p-Value	Stream order	Stream order	p-Value	Stream order	Stream order	p-Value
1	2	0.4750	1	2	0.3211	1	2	0.3177
1	3	0.3921	1	3	0.1625	1	3	0.2476
1	4	0.9913	1	4	0.4421	1	4	0.2998
1	5	0.0669	1	5	0.6698	1	5	0.5150
1	6	0.1048	1	6	0.4269			
1	7	0.0011	1	7	0.0322			
2	3	0.6830	2	3	0.3660	2	3	0.6240
2	4	0.2095	2	4	0.5420	2	4	0.6314
2	5	<.0001	2	5	0.0007	2	5	0.9600
2	6	0.0004	2	6	0.0016			
2	7	<.0001	2	7	<.0001			
3	4	0.0613	3	4	0.0757	3	4	0.8629
3	5	<.0001	3	5	<.0001	3	5	0.8419
3	6	0.0001	3	6	0.0001			
3	7	<.0001	3	7	<.0001			
4	5	0.0002	4	5	0.0007	4	5	0.6822
4	6	0.0026	4	6	0.0020			
4	7	<.0001	4	7	<.0001			
5	6	0.2263	5	6	0.1000			
5	7	<.0001	5	7	<.0001			
6	7	0.0391	6	7	0.0694			

The Wilcoxon/Kruskal-Wallis test for each of the ecoregions showed similar results to the ones for entire study area. The main challenge arising for the test at the ecoregion level was the lack of observations at certain stream orders. As seen in Table 2, 3, and 4, stream order 6 in ecoregion 29 has only one monitoring site and a few others have less than 10 sites. This causes problems comparing score means of these low observation orders with better populated orders. The score means of the lower count stream orders will be skewed and the comparisons may provide us with incorrect

significance values for alpha comparison. However, even considering this issue, the Wilcoxon test results were very similar, significance wise, for the three datasets. The impacted datasets had significance probabilities of less than 0.05, which meant that at least two of the stream order score means were different in those datasets. The unimpacted sites data set had probabilities of greater than 0.05 for all three ecoregions signifying that the score means of the stream orders were equal for all stream orders. Table 6 summarizes the Chi square probabilities for each test.

**Table 6.** Wilcoxon/Kruskal-Wallis test probabilities by ecoregion.

Ecoregion	Probability > Chi Square		
	All monitoring sites	<10% Urban sites	Unimpacted sites
27	<0.0001	<0.0001	0.2446
29	<0.0001	0.0002	0.4076
35	0.0085	0.0144	0.5716

### Correlation tests

Nonparametric Kendall  $\tau$  and Spearman's  $\rho$  correlation tests were applied to the three datasets with the results shown in Table 7. The WWTF impacted datasets showed significant correlation probabilities but the correlation co-efficient for both were closer to 0 than -1. This meant that the test did show there was a statistically significant negative correlation between Strahler order and median *E. coli* but it was a weak correlation. The correlation test for unimpacted sites resulted in a non-significant probability. The correlation co-efficient was also close to zero which means there was no correlation between stream order and median *E. coli* of unimpacted monitoring sites.

**Table 7.** Correlation values between *E. coli* and Strahler stream order.

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2491	<0.0001	-0.2346	<0.0001	0.0622	0.3254
Kendall's $\tau$	-0.1882	<0.0001	-0.1786	<0.0001	0.0483	0.3128

The correlation tests at the ecoregion level revealed similar results to the analysis of all data.

Table 8 shows the correlation co-efficient and the probability associated with them. Ecoregion 27 showed much stronger correlation between *E. coli* and stream order for the impacted datasets compared to the other two ecoregions but it was still lower than -0.50 for both Kendall  $\tau$  and Spearman's  $\rho$  which would be the criterion for a strong correlation. Ecoregion 29 and 35 showed very similar correlation coefficients and stayed with the trend where the *E. coli* concentrations in the unimpacted monitoring sites were not correlated with the Strahler stream order.

Another observation was the reduction in correlation coefficient values as the data moved eastward i.e. ecoregion 27 had the highest coefficients followed by 29 and lastly ecoregion 35. This signified that *E. coli* became less correlated to stream orders moving eastward.

**Table 8.** Correlation values between median *E. coli* and Strahler stream order at ecoregion level.

<b>Ecoregion 27</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4795	<0.0001	-0.4476	<0.0001	-0.1444	0.1768
Kendall's $\tau$	-0.3696	<0.0001	-0.3447	<0.0001	-0.116	0.153
<b>Ecoregion 29</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2785	<0.0001	-0.2431	<0.0001	0.0673	0.5405
Kendall's $\tau$	-0.2051	<0.0001	-0.1778	<0.0001	0.0592	0.4814
<b>Ecoregion 35</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.1787	0.0032	-0.1627	0.0131	-0.154	0.1784
Kendall's $\tau$	-0.1424	0.0017	-0.1359	0.0055	-0.1177	0.1846

#### *Analysis of median E. coli by modified Shreve stream order*

#### **Data distribution**

Since Shreve orders are distributed from one to 7,447 for the 742 monitoring sites in the data, the stream orders were grouped to apply the non-parametric Wilcoxon/Kruskal-Wallis test. The GIS calculated stream order were grouped into 9



categories with the first three orders being kept the same as the original order, and remaining observations distributed fairly equally in groups 4 to 9 as seen in Table 9. This was an important step since large numbers of orders can invalidate the Wilcoxon/Kruskall-Wallis test. In addition, small numbers of observations i.e. 10 or less for each stream order may also produce less reliable statistics.

**Table 9.** Modified Shreve stream orders for Wilcoxon/Kruskall-Wallis test (all monitoring sites).

Shreve modified	Shreve order	No. of monitoring sites
1	1	34
2	2	36
3	3	31
4	4-8	109
5	9-16	108
6	17-33	108
7	34-97	108
8	98-374	108
9	375-7,447	100

Similar to the Strahler order the modified Shreve order showed that the median *E. coli* for each stream order increased initially and decreased for higher stream orders. The analysis of the *E. coli* at the ecoregion scale also had a similar pattern to the Strahler order where Ecoregion 27 had consistently higher median *E. coli* compared to ecoregions 29 and 35. Lack of observations at ecoregion scale was another issue that was common with a few orders having less than 10 monitoring sites.

**Table 10.** Summary of *E. coli* by modified Shreve stream order for all monitoring sites.

All monitoring sites

Modified Shreve stream order	Number of monitoring sites	Median <i>E. coli</i>
1	34	91.75
2	36	88.50
3	31	95.00
4	109	96.00
5	107	120.00
6	108	112.50
7	108	100.00
8	108	56.00
9	100	26.00
Total	741	81.00

**Table 11.** Median *E. coli* concentrations by Shreve stream order and ecoregion for all monitoring sites.

Modified Shreve stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	2	22	10	293.75	92.50	66.50
2	2	21	13	355.00	75.00	104.00
3	2	14	15	58.75	87.00	130.00
4	18	38	53	295.00	80.00	97.00
5	30	34	43	172.50	100.50	110.00
6	36	35	37	150.50	93.00	103.00
7	36	35	37	111.25	73.50	120.00
8	27	48	33	90.00	42.00	55.50
9	33	38	29	37.00	23.75	25.00

**Table 12.** Median *E. coli* concentrations by Shreve stream order and ecoregion for <10% urban area.

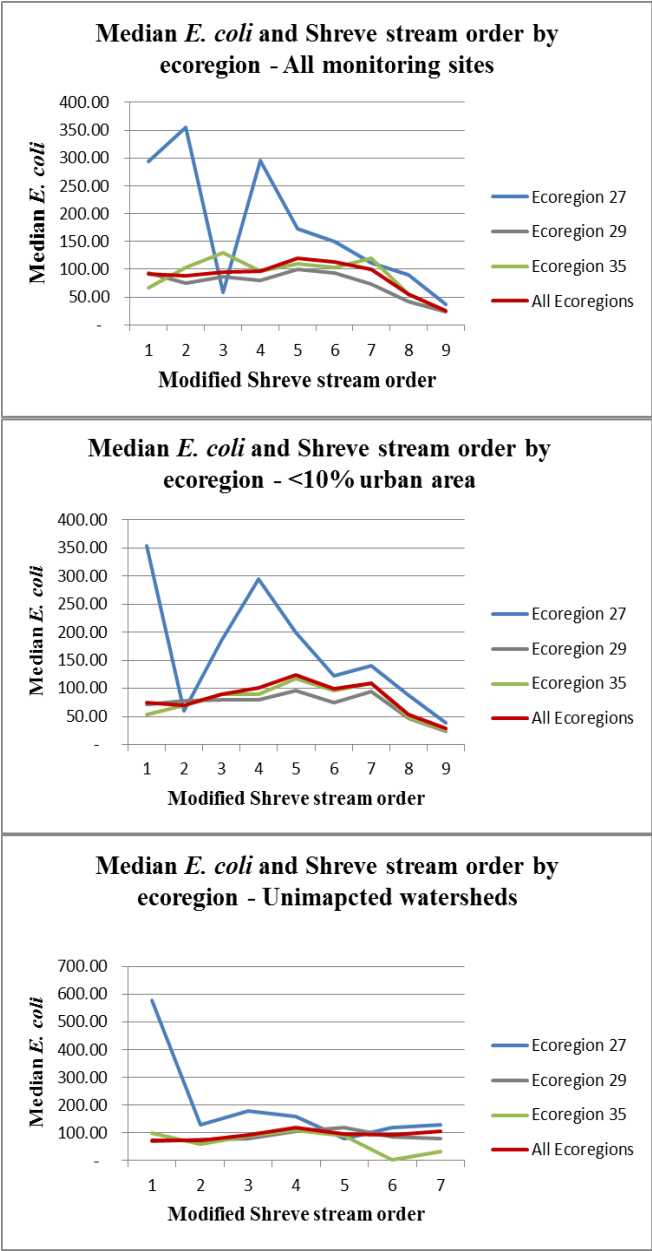
Modified Shreve stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	4	28	16	353.75	72.00	54.25
2	4	18	19	60.00	78.00	70.00
3	9	16	19	185.00	79.50	90.00
4	8	13	17	295.00	80.00	90.00
5	23	15	29	200.00	96.00	117.50
6	36	37	37	122.50	75.00	97.00
7	39	35	35	140.00	95.00	110.00
8	28	52	29	88.75	47.50	49.00
9	34	39	31	39.00	24.00	25.00

**Table 13.** Median *E. coli* concentrations by Shreve stream order and ecoregion for unimpacted watershed sites.

Modified Shreve stream order	Number of monitoring sites			Median <i>E. coli</i>		
	Ecoregion 27	Ecoregion 29	Ecoregion 35	Ecoregion 27	Ecoregion 29	Ecoregion 35
1	3	26	13	577.50	69.00	100.00
2	8	19	18	127.50	77.00	59.75
3	10	11	13	180.00	80.00	90.00
4	13	6	12	160.00	105.00	108.75
5	9	11	11	80.00	120.00	88.00
6	20	4	7	117.50	84.00	4.00
7	26	8	4	128.75	77.50	34.00

Figure 7 shows a graphic comparison of median *E. coli* of each stream order and the trend of decreasing *E. coli* at higher stream orders can be clearly seen for the

impacted datasets. On the other hand, the unimpacted watershed dataset has a minimal difference at higher stream orders and median *E. coli* stays flat throughout.



**Figure 7.** Comparison of median *E. coli* by modified Shreve order at ecoregion scale.

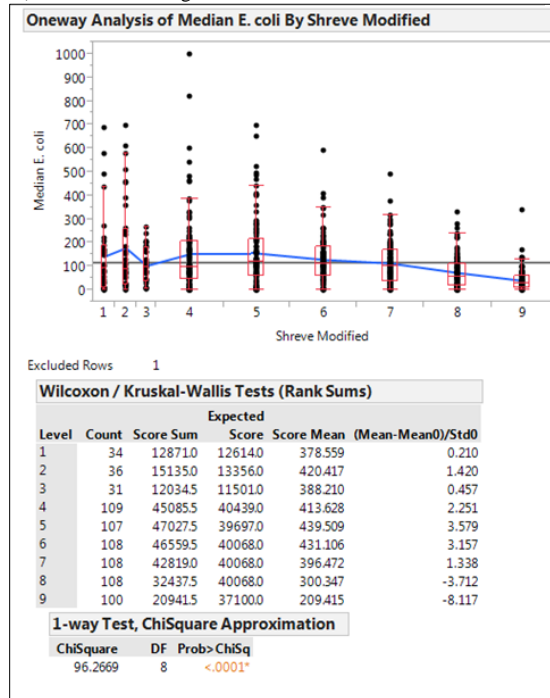
**Wilcoxon/Kruskal-Wallis test**

The modified Shreve stream order also has discrete characteristics therefore the Wilcoxon/Kruskal-Wallis test was applied to the data. The Chi Square p-value of datasets with wastewater outfalls (i.e. impacted sites) returned significant probabilities of less than 0.0001. This meant there are at least two stream orders that have different score means for median *E. coli*. On the other hand, similar to Strahler stream order results for the Chi Square probability for unimpacted watershed sites was not significant as the p-value of 0.4937 was higher than the alpha limit of 0.05 as seen in Figure 8

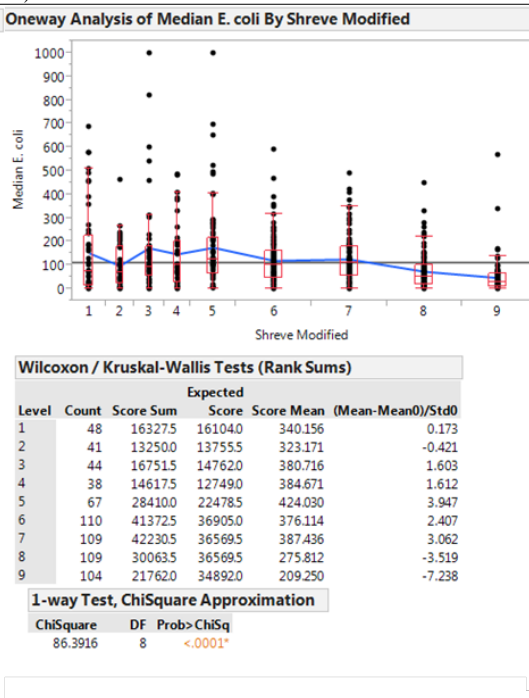
.

The analysis of each stream order pair showed that there were pairs of streams orders that had similar medians for datasets with WWTF outfalls, but the majority of the pairs differed as seen in Table 14. For the unimpacted dataset, all of the stream order pairs showed a probability greater than 5% significance threshold.

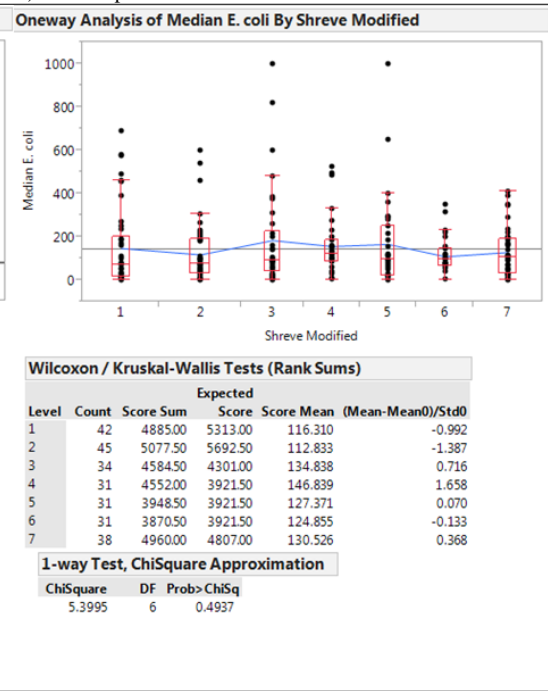
a) All monitoring sites



b) Sites with less than 10% urban area



c) Unimpacted sites



**Figure 8.** Wilcoxon/Kruskal-Wallis test of median *E. coli* by modified Shreve order.

**Table 14:** Modified Shreve stream order pairs with equal score means using Wilcoxon each pair test.

a) All monitoring sites			b) <10% Urban sites			c) Unimpacted sites		
Stream order	Stream order	p-Value	Stream order	Stream order	p-Value	Stream order	Stream order	p-Value
1	2	0.4072	1	2	0.6775	1	2	0.9661
1	3	0.9215	1	3	0.3169	1	3	0.2315
1	4	0.4782	1	4	0.4623	1	4	0.1234
1	5	0.2483	1	5	0.0686	1	5	0.6153
1	6	0.3790	1	6	0.4206	1	6	0.5282
1	7	0.8018	1	7	0.2973	1	7	0.4580
1	8	0.1065	1	8	0.1159			
1	9	0.0014	1	9	0.0019			
2	3	0.4812	2	3	0.1769	2	3	0.1626
2	4	0.7853	2	4	0.1679	2	4	0.0429
2	5	0.7819	2	5	0.0065	2	5	0.4527
2	6	0.9430	2	6	0.1396	2	6	0.3255
2	7	0.5079	2	7	0.0808	2	7	0.3124
2	8	0.0070	2	8	0.2035			
2	9	<.0001	2	9	0.0012			
3	4	0.4713	3	4	0.8561	3	4	0.4422
3	5	0.1731	3	5	0.2232	3	5	0.6504
3	6	0.3090	3	6	0.9061	3	6	0.6316
3	7	0.8238	3	7	0.7917	3	7	0.8790
3	8	0.0319	3	8	0.0021			
3	9	<.0001	3	9	<.0001			
4	5	0.3831	4	5	0.3591	4	5	0.4140
4	6	0.6971	4	6	0.6152	4	6	0.1038
4	7	0.4569	4	7	0.8595	4	7	0.4258
4	8	0.0002	4	8	0.0041			
4	9	<.0001	4	9	<.0001			
5	6	0.5234	5	6	0.0548	5	6	0.8769
5	7	0.0978	5	7	0.1551	5	7	0.9279
5	8	<.0001	5	8	<.0001			
5	9	<.0001	5	9	<.0001			
6	7	0.2222	6	7	0.5813	6	7	0.6596
6	8	<.0001	6	8	<.0001			
6	9	<.0001	6	9	<.0001			
7	8	0.0004	7	8	<.0001			
7	9	<.0001	7	9	<.0001			
8	9	0.0007	8	9	0.0062			



Similar to the Strahler order, the WWTF impacted datasets had significance probabilities of less than 0.05 which means that at least two of the stream order score means were different in those sets. The unimpacted sites data set had probabilities of greater than 0.05 for all three ecoregions signifying that the score means of the stream orders were equal for all stream orders. Table 15 summarizes the Chi square probabilities for each dataset.

**Table 15.** Wilcoxon/Kruskall-Wallis test probabilities by ecoregion.

Ecoregion	Probability > Chi Square		
	All monitoring sites	<10% Urban sites	Unimpacted sites
27	<0.0001	<0.0001	0.647
29	<0.0001	0.0015	0.8269
35	0.0004	0.0025	0.459

### Correlation tests

The Kendall  $\tau$  and Spearman's  $\rho$  correlation tests were significant for the impacted datasets with included the wastewater outfalls but the correlation coefficients as shown in

Table 16 were not very high which meant that the correlation between the modified Shreve order and *E. coli* for the monitoring sites was not very strong. Similar

to the Strahler analysis, the unimpacted watershed sites did not show any correlation between the modified Shreve order and *E. coli*.

**Table 16.** Correlation values between *E. coli* and modified Shreve stream order.

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2716	<0.0001	-0.2484	<0.0001	0.0735	0.2449
Kendall's $\tau$	-0.2	<0.0001	-0.1826	<0.0001	0.0551	0.2244

The analysis of correlations at the ecoregion level had similar results to the Strahler analysis with ecoregion 27 showing the highest correlation coefficients for the datasets with wastewater outfalls. The only difference from Strahler analysis here was that the unimpacted sites stream orders also showed a significant but weak negative correlation with *E. coli* in ecoregion 27. The correlation coefficients also decreased when going eastward, i.e. ecoregion 27 had the highest correlation followed by ecoregion 29 and ecoregion 35 similar to Strahler method.

**Table 17.** Correlation values between median *E. coli* and modified Shreve stream order at ecoregion level.

<b>Ecoregion 27</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4855	<0.0001	-0.4596	<0.0001	-0.2176	0.0405
Kendall's $\tau$	-0.3701	<0.0001	-0.3472	<0.0001	-0.166	0.0319
<b>Ecoregion 29</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.303	<0.0001	-0.2642	<0.0001	0.1114	0.3102
Kendall's $\tau$	-0.2189	<0.0001	-0.1877	<0.0001	0.0969	0.2098
<b>Ecoregion 35</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2082	0.0006	-0.1678	0.0105	-0.1321	0.249
Kendall's $\tau$	-0.158	0.0003	-0.1329	0.0046	-0.0894	0.2664

#### *Analysis of median E. coli by Shreve stream order and watershed area*

#### **Application of nonlinear models**

The *E. coli* data for both the Shreve order and the watershed area plots very similarly and in a nonlinear fashion. At low orders and smaller watershed areas, i.e.

orders less than 15 and watershed areas less than 10,000 square kilometers (km), there is higher density of monitoring stations along with higher *E. coli* concentrations. As the stream order and watershed area increase, *E. coli* concentrations start to get smaller and plots closer to the x-axis. This behavior of the data follows a decreasing exponential curve and application of nonlinear exponential models will show if those are a fit for this data.

Figure 9 shows the application of the three polynomial distribution fits and the two and three parameter exponential distribution fits using the procedures described by Walsh (2013). The impacted datasets with wastewater outfalls had the best fit with the two and three parameter exponential distributions for both Shreve order and watershed size. Based on the lowest AICc value, the three parameter (3P) exponential distribution was the best fit. Akaike information criterion (AIC) is an estimator of the quality of statistical model. AIC computes an estimate of the information lost in certain statistical model as compared to other models and assigns a value based on that. The lower AIC values mean a better model fit (Helsel and Hirsch 2002). Even though the 3P exponential distribution showed a very good fit based on AIC, there were monitoring stations which plotted far from the curve, which means if this fit was adopted those stations would not be represented.

Figure 9 also shows how similar the Shreve order and watershed area *E. coli* distributions are. Both datasets plot very similarly and the nonlinear fits for them are alike as well. From this, it can be concluded that Shreve order and watershed area depict

the same characteristic for the *E. coli* distributions as noted in the stream order theory chapter earlier. Another observation that can be gleaned from

Figure 9 is that most of the higher *E. coli* concentrations plot at lower stream order and watershed sizes. At higher orders and watershed sizes, *E. coli* concentrations stay lower than 150 MPN/100ml. Therefore, it is recommended that future studies should be focused on the lower Shreve stream orders and watershed sizes. These results also compare well with the study conducted by Harmel et. al. (2010) where it was found that *E. coli* concentrations decrease with increase watershed scales. Correlation for these two data would be tested by applying the non-parametric test since all the models showed existence of outliers. The parametric Pearson's  $r$  test result after applying the 3P exponential model will also be presented for comparison purposes only. For the unimpacted watersheds, none of the nonlinear distributions provided a good fit; therefore, only non-parametric correlation tests are applicable.

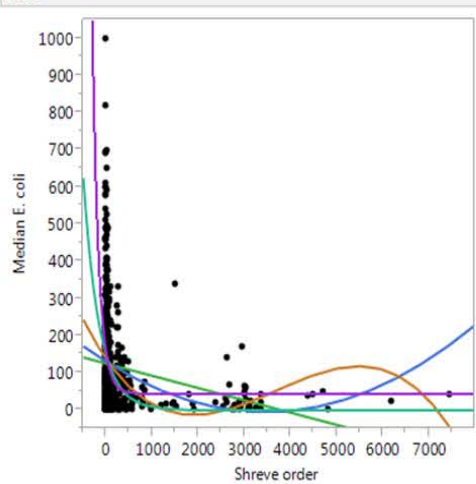
Shreve stream order

All monitoring sites

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	9215.738	0.8996668					9234.1156	10811786	14650.116	121.03766	0.0933
Exponential 2P	9220.1283	0.1001724					9233.9197	10905749	14757.441	121.48021	0.0854
Cubic	9233.0026	0.0001604					9255.961	11036413	14974.78	122.37148	0.0744
Quadratic	9244.826	4.3422e-7					9263.2036	11244643	15236.644	123.4368	0.0570
Linear	9254.9979	2.6847e-9					9268.7893	11431213	15468.488	124.37238	0.0413

Plot



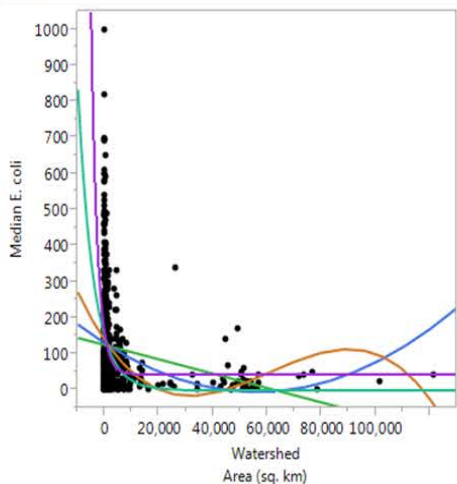
Watershed area (sq. km)

All monitoring sites

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	9216.7114	0.8759269					9235.0891	10825999	14669.375	121.11719	0.0921424
Exponential 2P	9220.6241	0.1238334					9234.4156	10913049	14767.32	121.52086	0.0848424
Cubic	9233.1238	0.0002391					9256.0822	11038217	14977.228	122.38149	0.0743459
Quadratic	9244.9104	6.5939e-7					9263.2881	11245924	15238.38	123.44383	0.0569278
Linear	9254.9872	4.2755e-9					9268.7787	11431048	15468.265	124.37148	0.0414035

Plot

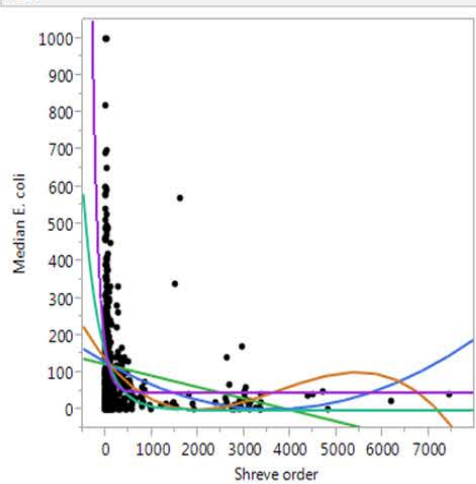


<10% urban sites

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	8398.316	0.9459679					8416.285	10768875	16145.24	127.06392	0.08069
Exponential 2P	8404.0477	0.0538591					8417.5335	10894258	16308.769	127.70579	0.06998
Cubic	8415.6023	0.0001668					8438.0483	11016898	16541.889	128.61528	0.05951
Quadratic	8422.2837	5.907e-6					8440.2526	11161079	16733.252	129.35707	0.04721
Linear	8428.43	2.7335e-7					8441.9158	11298018	16913.2	130.05076	0.03552

Plot

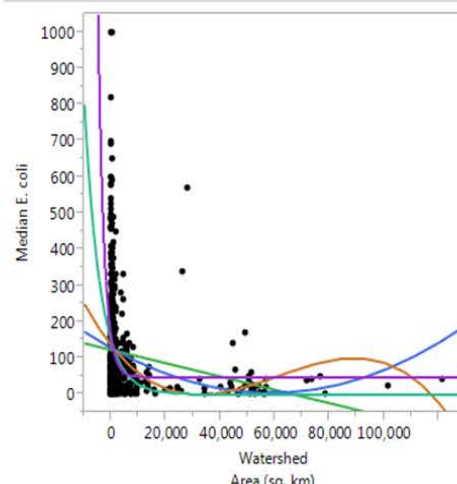


<10% urban area

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	8398.2038	0.9451667					8416.1728	10767072	16142.537	127.05328	0.0808465
Exponential 2P	8403.9041	0.0546655					8417.3899	10891923	16305.274	127.69211	0.0701883
Cubic	8415.5414	0.0001624					8437.9874	11015897	16540.385	128.60943	0.0596051
Quadratic	8422.4754	5.0701e-6					8440.4443	11164273	16738.041	129.37558	0.0469386
Linear	8428.3717	2.6586e-7					8441.8575	11297035	16911.729	130.0451	0.0356051

Plot

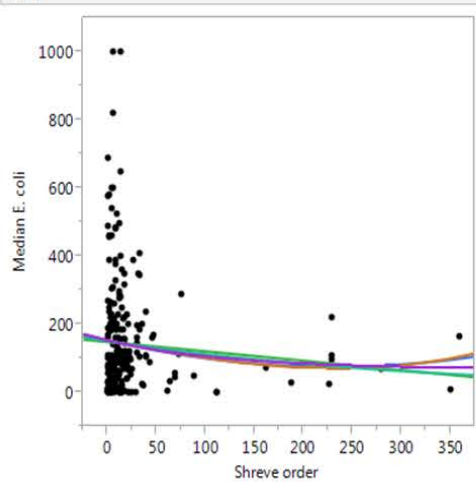


Unimpacted sites

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 2P	3292.8329	0.3526848					3303.3244	6811723.8	27246.895	165.06634	0.00728
Linear	3293.0148	0.3220222					3303.5063	6816642.7	27266.571	165.12592	0.00656
Quadratic	3294.658	0.1416048					3308.6138	6805236.9	27330.268	165.31869	0.00822
Exponential 3P	3294.7768	0.133441					3308.7325	6808444.7	27343.152	165.35765	0.00771
Cubic	3296.7302	0.0502472					3314.1334	6804972.5	27439.405	165.64844	0.00826

Plot



Unimpacted sites

Model Comparison

Model	AICc	AICc Weight	.2	.4	.6	.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 2P	3293.3622	0.2886162					3303.8537	6826044.9	27304.18	165.23976	0.0051941
Exponential 3P	3293.3746	0.2868239					3307.3304	6770668	27191.438	164.89826	0.0132645
Linear	3293.4365	0.2780873					3303.928	6828058.5	27312.234	165.26413	0.0049006
Quadratic	3295.3295	0.1079235					3309.2853	6823395.7	27403.196	165.53911	0.0055801
Cubic	3297.3885	0.038549					3314.7918	6822773.7	27511.184	165.86496	0.0056708

Plot

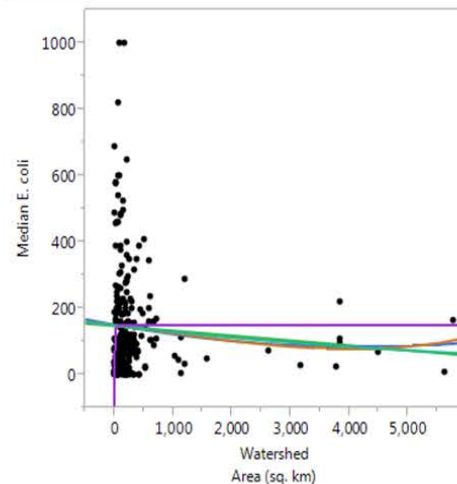


Figure 9. Nonlinear model fits for *E. coli* by Shreve stream order and watershed area.

## Correlation tests

The correlation values for the *E. coli* by Shreve stream order and watershed area are very similar to earlier tests conducted on Strahler and modified Shreve stream order methods. The impacted datasets show weak yet significant negative correlations between *E. coli* and Shreve stream order (

Table 18) and between *E. coli* and watershed size (Table 19). On the other hand, the unimpacted datasets do not show any correlation between Shreve order and *E. coli*. Spearman's  $\rho$  test does show some correlation between *E. coli* and watershed size for unimpacted dataset but the Kendall's  $\tau$  does not. The correlation co-efficient of 0.1292 in this case is small enough to ignore the Spearman's  $\rho$  significance probability of 0.405. The Pearson's  $r$  test, done on the modeled distributions for the data with wastewater outfalls, gives correlation co-efficient of -0.60 for both the Shreve order and the watershed size. This is a strong correlation with *E. coli* but at the same time, the model ignores the effect of outliers. The exponential model calculates a fitted curve based on the median *E. coli* of each monitoring site. As seen in

Figure 9 the fitted curve (purple line) for 3P exponential model goes through many of actual data points but there are some outlier points with high *E. coli* concentrations and higher stream order/watershed sizes that are not well represented by the curve. In addition, a large number of data points below the curve and near the origin point of zero are also not close to the curve but that may be less of a concern since the curve at that point is below the TAC limit of 126 cfu/100ml.

**Table 18.** Correlation values between *E. coli* and Shreve stream order.

Test	All monitoring sites		<10% Urban sites		Unimpacted sites <sup>29</sup>	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2714	<0.0001	-0.2571	<0.0001	0.0733	0.2463
Kendall's $\tau$	-0.1914	<0.0001	-0.1824	<0.0001	0.0509	0.24
Pearson's $r^*$	-0.5963		-0.5939			

\* Using the three parameter exponential model

**Table 19.** Correlation values between *E. coli* and watershed area.

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2572	<0.0001	-0.2441	<0.0001	0.1292	0.0405
Kendall's $\tau$	-0.183	<0.0001	-0.1754	<0.0001	0.0819	0.0539
Pearson's $r^*$	-0.6014		-0.5967			

\* Using the three parameter exponential model

**The correlation analysis at the ecoregion scale also had similar results to the Strahler method. The impacted datasets showed a weak but significant negative correlation for both Shreve stream order (Table 20) and watershed size (**

**Table 21) with *E. coli* for each ecoregion. The unimpacted datasets for Shreve order showed a weak negative correlation with *E. coli* in ecoregion 27 but no correlation in the other two ecoregions. The unimpacted dataset for watershed size at ecoregion 29 showed a weak positive relationship between *E.coli* and watershed size i.e. *E. coli* increased as the watershed size increased. The other two ecoregions showed no correlation between watershed size and *E. coli* for unimpacted data.**



Similar to findings from the Strahler method, the correlation coefficients for both Shreve order and watershed size were highest for ecoregion 27 and lowest for ecoregion 35 showing a decreasing trend going eastward.

**Table 20.** Correlation values between *E. coli* and Shreve stream order at ecoregion level.

<b>Ecoregion 27</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4874	<0.0001	-0.4626	<0.0001	-0.2108	0.0474
Kendall's $\tau$	-0.3519	<0.0001	-0.3340	<0.0001	-0.1508	0.3900
Pearson's $r^*$	-0.6218		-0.6297			
<b>Ecoregion 29</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.3036	<0.0001	-0.2692	<0.0001	0.1074	0.3278
Kendall's $\tau$	-0.2089	<0.0001	-0.1832	<0.0001	0.0935	0.2231
Pearson's $r^*$	-0.6569		-0.6476			
<b>Ecoregion 35</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2025	0.0008	-0.1869	0.0043	-0.1405	0.2198
Kendall's $\tau$	-0.1490	0.0003	-0.1415	0.0015	-0.0966	0.2239
Pearson's $r^*$	-0.5162		-0.5200			

\* Using the three parameter exponential model

**Table 21.** Correlation values between *E. coli* and watershed area at ecoregion level.

<b>Ecoregion 27</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4708	<0.0001	-0.4586	<0.0001	-0.1907	0.0735
Kendall's $\tau$	-0.341	<0.0001	-0.3318	<0.0001	-0.1379	0.0565
Pearson's $r^*$	-0.622		-0.626			
<b>Ecoregion 29</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2671	<0.0001	-0.2414	<0.0001	0.2376	0.0285
Kendall's $\tau$	-0.1859	<0.0001	-0.1668	<0.0001	0.1667	0.0248
Pearson's $r^*$	-0.6669		-0.6516			
<b>Ecoregion 35</b>						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2011	0.0009	-0.168	0.0104	-0.0712	0.5355
Kendall's $\tau$	-0.1498	0.0003	-0.1325	0.0028	-0.0564	0.4682
Pearson's $r^*$	-0.5237		-0.5233			

\* Using the three parameter exponential model

### Discussion of results

Statistical analysis of the data collected was done by dividing it into three datasets – 1) all monitoring sites; 2) sites which had 10 percent or less urban area in their watersheds; and 3) unimpacted watersheds sites (i.e. those without WWTFs and <10% urban area). The datasets were analyzed for *E. coli* relationships by using nonparametric Wilcoxon/Kruskal-Wallis test on Strahler stream order and modified Shreve stream order. Since Shreve order and watershed area were continuously distributed, nonlinear exponential distributions were applied to find a best fit model to test the *E. coli* relationship with them parametrically.

The Strahler and modified Shreve stream order Wilcoxon tests showed very similar results for the two impacted datasets. Both stream orders showed a Chi square probability of less than 0.0001 which means there are at least two stream orders that have different score means for median *E. coli* concentrations of datasets impacted with wastewater outfalls. The unimpacted dataset show that the median *E. coli* for all the stream orders is statistically same due to the insignificant Chi square probabilities of 0.7848 and 0.4937 for Strahler and Shreve stream orders respectively. These results differ somewhat from the Harmel et. al (2010) where a decreasing trend in both impacted and unimpacted datasets was seen. The notable differences between Harmel study and this one include the smaller study area and inclusion of site specific impacted identifiers such as dairy operations etc. in addition to the WWTF in the Harmel study. This study though found similarities with the Edge et. al. (2012) where it was found that streams with order less than 3 had higher *E. coli* concentrations when compared to orders greater than 3. This study found that order 4 and less have higher *E. coli* concentration when compared to higher stream orders. Again the major difference in the two studies is the study area scale with Edge et. al. study area being much smaller than used in this study.

The correlation tests on Strahler and modified Shreve order also had very similar results with a weak but significant negative correlation present between *E. coli* and stream orders for the two WWTF impacted datasets i.e. the *E. coli* decreases as the stream order increases. Both stream orders had p-values of less than 0.0001 for the WWTF impacted datasets. But when looking at the unimpacted dataset there is no correlation between *E. coli* and the two stream order methods. The p-values for both methods were greater than 0.22 which are insignificant at 0.05 alpha level.

The tests on Shreve stream order and watershed area also had very similar results and their distributions with *E. coli* plotted in the same way as well which reinforces the theory of Shreve order being more representative of basin size as stated by Tarboton et. al (1991). A 3P exponential model was able to fit very well for the impacted datasets, but it did leave some monitoring stations as outliers. It may be of value to apply this fit to the WWTF impacted monitoring sites excluding the outliers. The outlier sites can then be individually assessed and studied to identify the sources of higher *E. coli* since they did not fall along the curve. For this study, it was decided to use nonparametric correlation tests on the WWTF impacted datasets as preferred tests with all sites included. None of the nonlinear models was a good fit for the unimpacted datasets therefore they could be only tested for nonparametric correlations. The nonparametric correlations tests revealed very similar results to the Strahler method where the WWTF impacted datasets showing weak but statistically significant negative *E. coli* relationship with Shreve order and watershed size and the unimpacted dataset showing no correlation of *E. coli* with Shreve order and watershed size. The WWTF impacted results compared favorably with the Lyautey et. al (2010) study where it was found that shreve orders are inversely related to the *E. coli* concentrations.

The modeled exponential distributions for the impacted datasets did show a relatively strong negative correlation of *E. coli* with Shreve order and watershed size. This strong correlation means that the 3P exponential curve calculated in the study can be very useful tool in identifying the outlier sites that have high *E. coli* concentrations at higher orders and larger watersheds. Those sites can then be individually studied to understand the reasons behind higher *E. coli* concentrations. At lower order and smaller watershed sizes the curve may not be as suitable since the range of *E. coli* concentrations is too large and higher concentrations are

having a large impact on the curve. The Pearson's  $r$  correlation coefficient for both datasets was approximately -0.60 but as mentioned earlier the modeled distribution did not account for all of the monitoring sites therefore this correlation result was not preferred for comparison with Strahler analysis.

The analysis of the datasets by ecoregion had very similar results for each distribution as well. Ecoregion 27 consistently showed higher *E. coli* concentrations than the other 2 regions. The correlation tests at the ecoregion scale also had very similar outcomes to the entire study region where the impacted datasets showed weak negative correlation between *E. coli* and stream orders/watershed size but the unimpacted data had no correlation. Ecoregion 27 showed a stronger correlation for the impacted data between *E. coli* and stream orders/watershed size compared to the other two ecoregions. The correlation at the ecoregion scale decreased as the data moved eastward similar to the over all *E. coli* values. As noted in the study area description that the vegetation types, precipitation amount and temperatures are different in the three ecoregion. Ecoregion 27 the driest of the three region in terms of precipitation and temperatures, showed higher *E. coli* concentrations. Ecoregion 27 also had more dryland and irrigated cropland compared to the other two ecoregion. Ecoregion 35 which had the most rainfall and where streams had the lower concentrations of suspended solids also showed the lowest amounts of *E. coli* concentrations. These variables were not specifically tested in this study but can be important part of future studies done on this subject.

Another important result seen was that the median *E. coli* for stream orders remained below the the geometric mean standard of 126 cfu/100ml for unimpacted datasets. This differed from the Harmel et. al. (2010) observation where the unimpacted streams violated the TAC regulation. Harmel et. al. looked at six sites in the Leon river watershed which showed this trend.

It is possible when looking at much larger areas and numbers of monitoring sites, the finding by Harmel et. al. and other similar watersheds were smoothed with the over all trend at the ecoregion level. Based on this, a recommendation can be made to conduct more site specific studies to understand the dynamics of smaller watersheds.

## CHAPTER IV

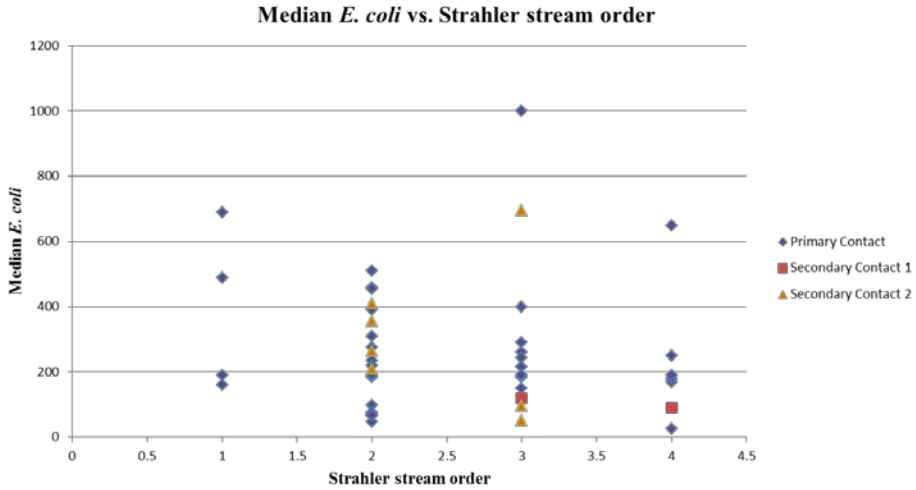
### RUAA ANALYSIS

The RUAA data was collected from TCEQ. The RUAA analysis have been completed on and approved by TCEQ on 94 stream segments. From these 94 segments, 46 monitoring stations were selected for the study due to the *E. coli* data availability from TCEQ and their sites snapping within 200 meters of a stream segment during the GIS steps. Out of those initially selected, one station was deemed an outlier due to a very high *E. coli* measurement of 240,000 MPN/1000ml. Of the remaining selected, 36 stations had been assigned the Primary Contact Recreation (PCR) 1 status, two were assigned Secondary Contact Recreation (SCR) 1 status, and seven stations were assigned SCR2 status. Based on this limited data no significant statistical tests could be conducted.

Table 22 shows the distribution of median *E. coli* by Strahler stream order. The SCR1 and SCR2 statuses are not available for all the monitoring stations and as the table and Figure 10 below show, *E. coli* for available stations does not have any discernable pattern.

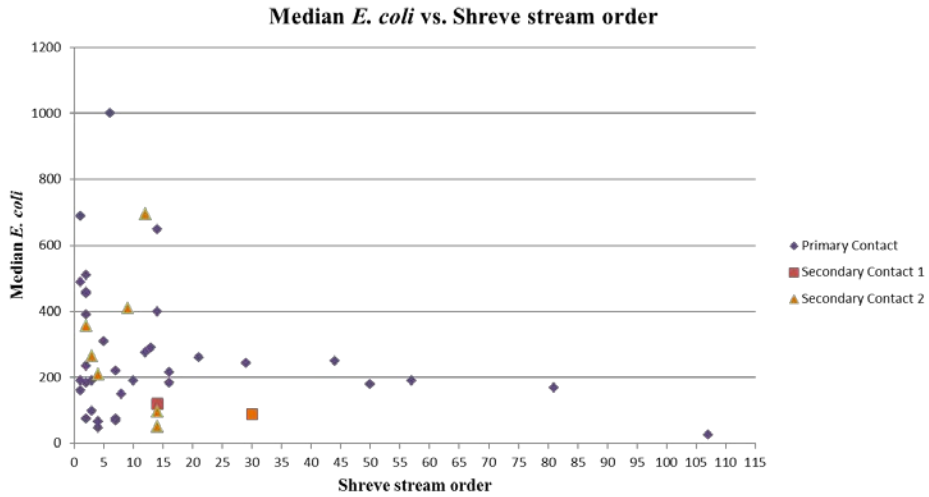
**Table 22.** RUAA status of *E. coli* by Strahler stream order.

Strahler stream order	Number of monitoring stations			Median <i>E. coli</i>		
	PCR1	SCR1	SCR2	PCR1	SCR1	SCR2
1	4			340.00		
2	17		4	190.00		310.75
3	9	1	3	244.00	120.00	96.00
4	6	1		185.00	89.50	

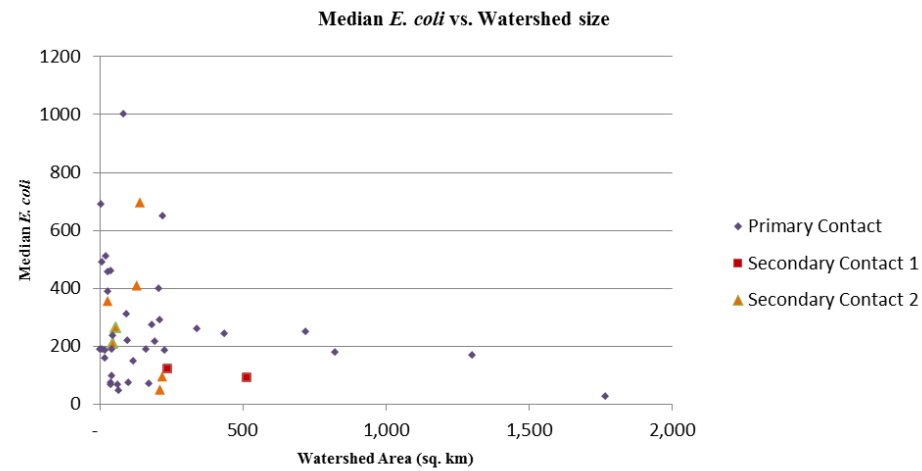


**Figure 10.** RUAA status of *E. coli* by Strahler stream order.

Similarly Shreve stream order and watershed area do not provide any significant information. The *E. coli* concentrations decrease as the Shreve stream order and watershed area grow larger as seen in Figures 11 and 12. With a larger dataset these relationship may be expanded.



**Figure 11.** RUAA status of *E. coli* by Shreve stream order.



**Figure 12.** RUAA status of *E. coli* by watershed area.



189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211

CHAPTER V  
CONCLUSIONS

The overall goal of this research was to understand and predict a relationship between stream order and *E. coli* and evaluate the application of that relationship to the current recreational use standards. The data analysis of impacted monitoring sites with wastewater outfalls showed a weak but significant negative correlation of *E. coli* concentrations with Shreve and Strahler orders and watershed area. Analysis of the data showed that at larger watershed sizes and stream orders, *E. coli* concentrations were significantly lower. At smaller stream orders and watersheds, the *E. coli* concentrations ranged from very low to very high. It can be argued that the wider range of *E. coli* concentrations at smaller stream orders and watershed sizes was a driving factor in making the correlation coefficients weak. The *E. coli* concentrations did become lower as the stream orders and watersheds became larger which points to the dilution factor as suggested by Lyautey et. al. (2010, 2011). But site specific studies may be a better way to understand *E. coli* relationships at smaller watershed and stream order scales where *E. coli* concentration range from very low to very high.

The data analysis of unimpacted sites showed that there was no correlation between *E. coli* concentrations and Shreve and Strahler orders as well as watershed area. Another significant

observation for the unimpacted dataset was that in the Strahler and Modified Shreve stream order studies the median *E. coli* stayed statistically the same through all the stream orders which means that median *E. coli* concentration did not change and stayed within TAC limits as stream length and watershed area increased. This leads to the conclusion that in the absence of anthropogenic factors median *E. coli* will stay the same as watershed area and stream order increase and dilution does not impact the concentrations. The limitations for this conclusion include lack of monitoring stations available at higher stream orders. Many of the monitoring sites which had higher stream orders, e.g. sites at orders 6 and 7 for Strahler order, did not qualify as unimpacted watersheds due the presence of wastewater outfalls. With availability of more data such as discharge volume or hydrologic distance of the wastewater outfalls to the monitoring site, the criteria for exclusion of wastewater outfalls may be modified to include more monitoring sites. This would allow for a better definition of unimpacted watersheds and a decreasing *E. coli* trend may be seen in the unimpacted sites as well at higher stream orders similar to impacted watersheds.

The wastewater outfall impact on the streams was another significant finding of this study. The datasets that included wastewater outfalls, showed a weak but statistically significant negative correlation with *E. coli*, but when the outfalls were removed, there was no relationship. This suggests that wastewater outfalls are a big influence on stream *E. coli* and further research on just the sites that are impacted by these outfalls may provide more valuable information on *E. coli* concentrations in the streams. Further studies with the types and volume of discharges made by wastewater outfalls may also help to assess the impact of wastewater outfalls on the water bodies.

The comparison of the three ecoregions also revealed that the median *E. coli* concentrations and their correlation with both stream order methods and watershed area decreased moving eastward. Ecoregion 27, which is the western most ecoregion in this study and has a more arid to semi-arid environment compared to the other ecoregions showed the highest *E. coli* concentration as well as correlations. Ecoregion 35, the eastern most region and with wettest climate of the three regions, has the lowest *E. coli* concentrations and correlations. This finding can be a good starting point for a future study that can include additional environmental factors such as precipitation, temperatures, types of streams (seasonal or perennial) etc., which make up each region and can contribute to the hydrology of these areas and its impact on the *E. coli* transport.

The study utilized Strahler and Shreve stream orders and watershed area as the variables to which the *E. coli* was correlated. Shreve stream order and watershed area provide similar information with respect to basin size and results based on them are very similar, it is recommended for future studies only one of the two variables should be used. The Strahler stream order method produces a smaller number of stream orders that helps in conducting the statistical analysis when comparing each stream order. This method is a good tool for comparing individual stream order numbers at ecoregion or river basin scales i.e. compare stream order 1 in ecoregion 27 with stream order 1 in ecoregion 29. Out of the three methods watershed area provides the most relevant information with respect to the scale of the basin but Strahler stream order can be the preferred method when comparing streams with similar characteristics.

For future research, inclusion of other variables such as flow volume at the time of the observation or a classification of the flow, such as low, normal, or high, which could identify if the sample was collected after a storm or in drought or normal conditions, may be beneficial as

flow can influence *E. coli* concentrations. Other variables such as temperature at the time of sample collection, and time of collection in terms of seasonality, such as summer, spring etc., may be valuable in understanding *E. coli* relationships. A larger number of *E. coli* observations at monitoring sites at higher stream orders may also help provide a better and clearer understanding of the dilution effect. Further, a study on the major contributors, fate, and transport of *E. coli* in each ecoregion may also help understand the decreasing *E. coli* and correlation trends when moving eastward.

Lack of completed RUAA studies did not allow for correlation tests between *E. coli* recreation standards based on the stream orders. The available data showed a decreasing trend in *E. coli* with increase in stream order and watershed size but due to lack of monitoring sites it is recommended that site specific studies continue to be conducted. In the future, there may be cause for reassessing this information when there are more RUAA studies completed.

- 273 Arnone, R. D. and J. P. Walling. 2007. "Waterborne Pathogens in Urban Watersheds." *Journal of*  
274 *Water Health* 5(1):149-162
- 275 Bai, S. and W. S. Lung. 2005. "Modeling Sediment Impact on the Transport of Fecal Bacteria."  
276 *Water Research* 39(20):5232-5240.
- 277 Berry, D. F. and C. Hagedorn. 1991. "Soil and Groundwater Transport of Microorganisms."  
278 *Assessing Ecological Risks of Biotechnology* 15:57-73.
- 279 Brettar, I. and M. G. Hofle. 1992. "Influence of Ecosystematic Factors on Survival of  
280 *Escherichia Coli* after Large- Scale Release into Lake Water Mesocosms." *Applied and*  
281 *Environmental Microbiology* 58(7):2201-2210.
- 282 Byappanahalli, M., M. Fowler, D. Shively and R. Whitman. 2003. "Ubiquity and Persistence of  
283 *Escherichia Coli* in a Midwestern Coastal Stream." *Applied Environmental Microbiology*  
284 69(8):4549-4555. doi: 10.1128/AEM.69.8.4549-4555.2003.
- 285 Craig, D. L., H. J. Fallowfield and N. J. Cromar. 2004. "Use of Microcosms to Determine  
286 Persistence of *Escherichia Coli* in Recreational Coastal Water and Sediment and Validation with  
287 in Situ Measurements." *Journal of Applied Microbiology* 96(5):922-930. doi: 10.1111/j.1365-  
288 2672.2004.02243.x.
- 289 Edge, T. A., A. El-Shaarawi, V. Khan, C. Gannon, R. Jokinen, I. U. H. Kent, W. Phillips, D.  
290 Koning, J. Lapen, N. Wilkes, R. Miller, W. Scott, H. Topp, A. van Bochove, I. Neumann, E.  
291 Shtepani, G. Robertson and E. Schreier. 2012. "Investigation of an *Escherichia Coli*  
292 Environmental Benchmark for Waterborne Pathogens in Agricultural Watersheds in Canada."  
293 *Journal of Environmental Quality* 41(1):21-30. doi: 10.2134/jeq2010.0253.
- 294 Elevation Derivatives for National Applications (EDNA). 2014. "Elevation Derivatives for  
295 National Applications". Available at <http://edna.usgs.gov/>. Accessed 08/27/2014.
- 296 Griffith, G.E., S.A. Bryce, J.M. Omernik, J.A. Comstock, A.C. Rogers, B. Harrison, S.L. Hatch  
297 and D. Bezanson. 2004. "Ecoregions of Texas (Color Poster with Map, Descriptive Text, and  
298 Photographs)." Reston, Virginia, U.S. Geological Survey (map scale 1:2,500,000). Available at  
299 <http://www.epa.gov/naaujydh/pages/ecoregions/>  
300 [tx\\_eco.htm](http://www.epa.gov/naaujydh/pages/ecoregions/tx_eco.htm). Accessed 06/21/2014.
- 301 Haan, C. T. 2002. Statistical Methods in Hydrology. 2nd ed. Ames, Iowa: Iowa State Press.
- 302 Halliday, E. and R. J. Gast. 2011. "Bacteria in Beach Sands: An Emerging Challenge in  
303 Protecting Coastal Water Quality and Bather Health." *Environmental Science & Technology*  
304 45(2):370-379. doi: 10.1021/es102747s.

305 Harmel, R. D., T. Gentry, R. Karthikeyan, and R. Srinivasan. 2010. "Effects of Agricultural  
306 Management, Land Use, and Watershed Scale on *E. Coli* Concentrations in Runoff and  
307 Streamflow." *Transactions of the ASABE* 53(6):1833-1841.

308 Helsel, D. R. and R. M. Hirsch. 2002. "Chapter A3 Statistical Methods in Water Resources." in  
309 Techniques of Water-Resources Investigations of the United States Geological Survey Book 4:  
310 U. S. Geological Survey.

311 Horizon. 2014. "NHDPlusV2 Data" Available at [http://www.horizon-systems.com/  
312 NHDPlus/NHDPlusV2\\_data.php](http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php). Accessed 05/26/2014.

313 Horton, Robert E. 1945. "Erosional Development of Streams and Their Drainage Basins;  
314 Hydrophysical Approach to Quantitative Morphology." *Bulletin of the Geological Society of  
315 America* 56(3):275-370. doi: 10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2  
316

317 Ichim, Ionita. 1990. "The Relationship between Sediment Delivery Ratio and Stream Order; a  
318 Romanian Case Study." Pp. 79-86, Vol. 189, edited by D. E. Walling, A. Yair and S. Berkowicz.  
319 [Louvain]: [Louvain], International: International Association of Hydrological Sciences.  
320

321 JMP®, Version 11. SAS Institute Inc., Cary, NC, 1989-2007.

322 Krometis, L.A. H., G. W. Characklis, O. D. Simmons, C. A. Likirdopulos, M. D. Sobsey and M.  
323 J. Dils. 2007. "Intra- Storm Variability in Microbial Partitioning and Microbial Loading Rates."  
324 *Water Research* 41(2):506-516. doi: 10.1016/j.watres.2006.09.029.

325 Lyautey, E., G. Wilkes, J. J. Miller, E. Van Bochove, H. Schreier, W. Koning, T. A. Edge, D. R.  
326 Lapen and E. Topp. 2011. "Variation of an Indicator of *Escherichia Coli* Persistence from  
327 Surface Waters of Mixed- Use Watersheds, and Relationship with Environmental Factors."  
328 *Annales de Limnologie - International Journal of Limnology* 47(1):11-19. doi:  
329 10.1051/limn/2010033.

330 Lyautey, E., Z. Lu, A. Scott, T. Berkers, E. Topp, D. R. Lapen, G. Wilkes and T. A. Edge. 2010.  
331 "Distribution and Diversity of *Escherichia Coli* Populations in the South Nation River Drainage  
332 Basin, Eastern Ontario Canada." *Applied and Environmental Microbiology* 76(5):1486-1496.  
333 doi: 10.1128/AEM.02288-09.

334 National Agricultural Statistics Service (NASS). 2014. "CropScape – Cropland Data Layer".  
335 Available at <http://nassgeodata.gmu.edu/CropScape/>. Accessed 05/26/2014.

336 Oliver, D. M., C. D. Clegg, A. L. Heathwaite and P. Haygarth. 2007. "Preferential Attachment of  
337 *Escherichia Coli* to Different Particle Size Fractions of an Agricultural Grassland Soil." *Water,  
338 Air, & Soil Pollution* 185(1-4):369-75. doi: 10.1007/s11270-007-9451-8.

339 Omernik, J. M. and R. G. Bailey. 1997. "Distinguishing between Watersheds and Ecoregions."  
340 *Journal of the American Water Resources Association* 33(5):935-949.

341 Pachepsky, Y. A., D. R. Shelton and D. R. Pachepsky. 2011. "*Escherichia Coli* and Fecal  
 342 Coliforms in Freshwater and Estuarine Sediments." *Critical Reviews in Environmental Science*  
 343 *and Technology* 41(12):1067-1110. doi: 10.1080/10643380903392718.

344 Poole, G. C., J. B. Dunham, D. M. Keenan, M. Deppman, S. T. Sauter, D. A. McCullough, C.  
 345 Mebane, J. C. Lockwood, D. A. Essig, M. R. Hicks, D. J. Sturdevant, E. J. Materna, S. A.  
 346 Spalding and J. Risley. 2004. "The Case for Regime- Based Water Quality Standards."  
 347 *BioScience* 54(2):155-161.

348 Schlotzhauer, S. D. 2007. *Elementary Statistics Using JMP®*. Cary, NC: SAS Institute Inc.

349 Shreve, R. L. 1966. "Statistical Law of Stream Numbers." *The Journal of Geology* 74(1):17-37.  
 350 doi: 10.1086/627137.

351 Strahler, A. N. 1957. "Quantitative Analysis of Watershed Geomorphology." *American*  
 352 *Geophysical Union Transactions* 38(6):913-920. doi: 10.1029/TR038i006p00913.

353 Texas Commission on Environmental Quality (TCEQ). 2010. Texas Administrative Code, Title  
 354 30 Environmental Quality, Part 1 Texas Commission on Environmental Quality, Chapter 307  
 355 Texas Surface Water Quality Standards, Volume 307.7 Site-Specific Uses and Criteria.

356 Texas Commission on Environmental Quality (TCEQ). 2014. "Recreational Use Attainability  
 357 Analyses." Available at [https://www.tceq.texas.gov/waterquality/standards](https://www.tceq.texas.gov/waterquality/standards/ruaas)  
 358 [/ruaas](https://www.tceq.texas.gov/waterquality/standards/ruaas). Accessed 08/28/2014.

359 Texas Commission on Environmental Quality (TCEQ). 2014a. "Permitted Wastewater Outfalls."  
 360 Available at <http://www.tceq.state.tx.us/gis/download-tceq-gis-data>. Accessed 07/12/2014.  
 361

362 Texas Commission on Environmental Quality (TCEQ). 2014, updated daily. Surface Water  
 363 Quality Monitoring Information System, 09/01/1968-05/02/2014 Bacteria-ER-27-29-35-  
 364 050214.xlsx. Compiled by Data Management & Analysis Team. Austin, TX USA. Data Request  
 365 ID 210.  
 366

367 Tarboton, D. G. 1997. "A New Method for the Determination of Flow Directions and Upslope  
 368 Areas in Grid Digital Elevation Models." *Water Resources Research* 33(2):309-319.

369 United States Environmental Protection Agency (USEPA). 2014a. "National Summary of  
 370 Impaired Waters and TMDL Information." Available at [http://iaspub.epa.gov/](http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T)  
 371 [waters10/attains\\_nation\\_cy.control?p\\_report\\_type=T](http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T). Accessed 08/18/2014.

372 United States Environmental Protection Agency (USEPA). 2014b. "Level III and IV Ecoregions  
 373 of the Continental United States." Available at [http://www.epa.gov/wed/](http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm)  
 374 [pages/ecoregions/level\\_iii\\_iv.htm](http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). Accessed 08/18/2014.

375 United States Environmental Protection Agency (USEPA). 2014c. "EPA Clip N Ship  
 376 Application." Available at <https://edg.epa.gov/clipship/>. Accessed 11/16/2014.

377 Virginia Department of Environmental Quality (VDEQ). 2003. The Quality of Virginia Non-  
378 Tidal Streams: First Year Report: Virginia Department of Environmental Quality. Available at  
379 [http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring](http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/ProbabilisticMonitoring/report1.pdf)  
380 [/ProbabilisticMonitoring/report1.pdf](http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/ProbabilisticMonitoring/report1.pdf). Accessed 06/30/2014.

381 Walsh S. 2013. When a Linear Model Just Won't Do: Fitting Nonlinear Models Using JMP®.  
382 JMP Discovery Conference. SAS Institute Inc. Available at [http://www.jmp.com](http://www.jmp.com/about/events/summit2013/resources/Paper_Susan_Walsh.pdf?utm_source=newswire1013&utm_medium=newsletter&utm_campaign=summit2013)  
383 [/about/events/summit2013/resources/Paper\\_Susan\\_Walsh.pdf?utm\\_source=newswire1013&utm](http://www.jmp.com/about/events/summit2013/resources/Paper_Susan_Walsh.pdf?utm_source=newswire1013&utm_medium=newsletter&utm_campaign=summit2013)  
384 [\\_medium=newsletter&utm\\_campaign=summit2013](http://www.jmp.com/about/events/summit2013/resources/Paper_Susan_Walsh.pdf?utm_source=newswire1013&utm_medium=newsletter&utm_campaign=summit2013). Accessed 02/18/2015.

385 Woods, A.J., J.M. Omernik, D.R. Butler, J.G. Ford, J.E. Henley, B.W. Hoagland, D.S. Arndt,  
386 and B.C. Moran. 2005. Ecoregions of Oklahoma (color poster with map, descriptive text,  
387 summary tables, and photographs): Reston, Virginia, U.S. Geological Survey (map scale  
388 1:1,250,000). Available at [http://www.epa.gov/naaujydh/pages](http://www.epa.gov/naaujydh/pages/ecoregions/ok_eco.htm)  
389 [/ecoregions/ok\\_eco.htm](http://www.epa.gov/naaujydh/pages/ecoregions/ok_eco.htm). Accessed 06/21/2014.

390 Zăvoianu I. 1985. Morphometry of drainage basins, Editura Academiei Republicii Socialiste  
391 România, București. Amsterdam. Elsevier Science.

392

393



## Appendix B

### **Distribution of *E. coli* Levels and Recreation Use as Factor of Stream Order in the Central Great Plains, Central Oklahoma/Texas Plains, and South Central Plains Ecoregions**

Khurram Rafi<sup>1</sup>, Kevin L. Wagner<sup>1#</sup>, Terry Gentry<sup>2</sup>, Raghupathy Karthikeyan<sup>3</sup>, Amanda Dube<sup>4</sup>

<sup>1</sup>Texas Water Resources Institute, Texas A&M AgriLife Research, College Station, TX

<sup>2</sup>Department of Soil and Crop Sciences, Texas A&M University, College Station, TX

<sup>3</sup>Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas.

<sup>4</sup>Institute of Natural and Renewable Resources, Texas A&M AgriLife Research, College Station, TX

#### **#Corresponding author:**

Kevin L. Wagner

Texas Water Resources Institute

1500 Research Parkway A110,

2260 TAMU,

College Station, TX 77843-2260.

phone: 979- 845-2649

email: [klwagner@ag.tamu.edu](mailto:klwagner@ag.tamu.edu)

**Running Title: Distribution of *E. coli* Levels and Recreation Use as Factor of Stream Order...**

## **ABSTRACT**

This research examines the relationship of *E. coli* concentrations and recreational use to stream order and watershed size. To determine possible ecoregion effects, *E. coli* data were obtained from monitoring stations in three ecoregions in Texas and Oklahoma – the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains (Ecoregions 27, 29, and 35, respectively). Median and geometric mean *E. coli* concentrations from the monitoring stations were analyzed for correlation with respect to the stream order of each monitoring site as well as the watershed size of each monitoring station. Geospatial analysis was used to determine stream order and watershed size and to identify un-impacted/least impacted streams in each ecoregion. Stream order was classified based on the traditional stream order classification method by Horton (1945) [GSA Bulletin 56(3),275-370] and Strahler (1957) [AGU Trans. 38(6),913-920] and stream link magnitude analysis method by Shreve (1966) [J. Geol. 74(1),17-37]. The analysis of two stream order systems and watershed size for each monitoring site with respect to median and geometric mean of *E. coli* showed no significant relationship in unimpacted watersheds. The watersheds impacted by wastewater outfalls and urban areas showed a weak negative relationship between *E. coli* and stream order/watershed size i.e. *E. coli* decreased with increase in stream order/watershed size. Ecoregion 27 consistently showed higher concentrations of *E. coli* and ecoregion 35 the lowest.

**Keywords:** Stream order; *E. coli*, Ecoregion, Wastewater outfalls.

## **Highlights:**

## **1. INTRODUCTION**

Pathogen contamination in freshwater streams is a significant water quality concern in the United States. Pathogen contamination is often assessed by determining the amount of an indicator bacteria such as *Escherichia coli* (*E. coli*) in freshwater systems whereas enterococcus is used as an indicator in marine waters (Halliday and Gast 2011).

Pathogen contamination is by far the largest cause for water quality impairment with approximately 11,000 instances reported to date. Over 80 percent of the pathogen impairment is caused by excessive fecal coliform bacteria which include *E. coli* (USEPA 2014a). Diarrhea, gastrointestinal illness, and diseases such as cholera can be attributed to waterborne pathogens along with approximately 900 deaths and 900,000 illnesses each year (Arnone and Walling, 2007). This makes setting appropriate recreational use standards for water bodies very important. In Texas, the recreational use standard for a water body is assumed primary body contact unless determined otherwise after a Recreational Use Attainability Analysis (RUAA) has been conducted. The RUAA is a site-specific study to determine what uses a stream can and does support based on its current and historic use and its physical and flow characteristics (TCEQ 2014). An appropriate recreational standard for a water body can then be set for the water body based on the findings of RUAA and approval by EPA.

The Texas Commission on Environmental Quality (TCEQ) is responsible for development and implementation of the Texas Water Quality Standards including the recreational use standards as specified in the Texas Administrative Code (TAC). The contact recreation standards with respect to *E. coli* concentrations as listed in the TAC regulations are as follows (TCEQ 2010):

- I. Primary contact recreation – activities involving significant risk of water ingestion (e.g. swimming, diving, surfing, etc.). The geometric mean criterion for *E. coli* is 126 per 100 ml. In addition, the single sample criterion for *E. coli* is 399 per 100 ml.

- II. Secondary contact recreation 1 – commonly occurring activities with limited body contact incidental to shoreline activity (e.g. fishing, kayaking, rafting, etc.). The geometric mean criterion for *E. coli* is 630 per 100 ml.
- III. Secondary contact recreation 2 – activities with limited body contact incidental to shoreline activities with less significant risk of water ingestion than secondary contact recreation 1. The geometric mean criterion for *E. coli* is 1,030 per 100 ml.
- IV. Noncontact recreation – activities with no significant risk of ingestion. The geometric mean criterion for *E. coli* is 2,060 per 100 ml.

As previously stated, in Texas, all fresh water bodies are classified as primary contact recreation unless a RUAA study determines a change in the classification is required (TCEQ 2010). The applicability of these standards has been questioned especially related to scale. Regime based standards, which account for temperature, sediment, and stream order of the water bodies have been proposed as possible alternatives (Poole et al., 2004). A central Texas study conducted by Harmel et. al. (2010) established a trend of decreasing *E. coli* concentrations when aggregating sampled segments from upstream to downstream. The trend was visible in both the “impacted” (land use of mixed rural with known point sources of pollution such as dairies, waste water treatment plants (WWTP), small communities, etc.) and “unimpacted” (land use of mixed rural excluding point sources of pollution) streams of the study area. The study also found that with increasing watershed scale the median *E. coli* concentrations decreased. Furthermore, the study found “unimpacted streams to be in violation of TAC regulations even though the anthropogenic impact was insignificant.

A study conducted by Lyautey et. al. (2010) in the South Nation River Drainage basin of Canada compared *E. coli* concentrations using Shreve (1966) stream link. The study found Shreve stream

link to be inversely correlated to *E. coli* concentrations with strongest correlations seen during summer. Another Canadian study by Edge et al. (2012) used Horton (1945) and Strahler (1957) stream order and found that streams with order 3 and lower had higher *E. coli* concentrations compared to the streams of order greater than 3. Another study (Lyautey et. al. 2011) explained, “the persistence factor is less important than dilution (i.e. stream order) in describing *E. coli* densities, followed by factors that influence the loading of *E. coli* into watersheds.”

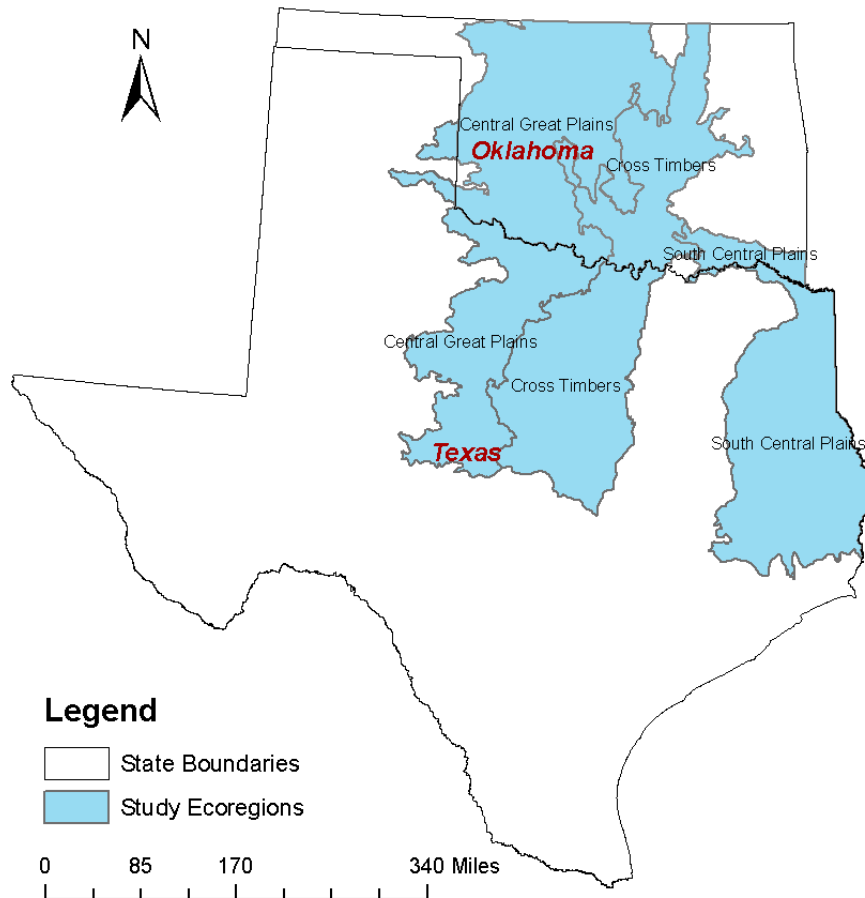
Conversely, some studies dispute the inverse correlation between stream order and *E. coli* concentrations and suggest other factors (i.e. nonpoint source inputs) as more vital. A study conducted by the Virginia Department of Environmental Quality (VDEQ) found no correlation to stream order or ecoregions (level III) when investigating fecal coliform bacteria. The report found highest fecal coliform concentrations at the stream orders 3 and 4. The report suggested use of *E. coli* in future study as a better indicator for pathogen contamination (VDEQ 2003). Another study by Byappanahalli et. al. (2003) found *E. coli* increasing steadily downstream in the Dunes Creek watershed in Indiana. The study found a relation between excessive ditching and consequent increase in non-point source input responsible for increase in *E. coli*.

Lyautey et. al. (2010, 2011) attribute the inverse relationship of *E. coli* and stream order to the dilution in the water body as smaller stream orders combine to make larger streams. However, other studies hypothesize an influence of physical stream processes such as turbidity/sediment process and flow rates, on the *E. coli* concentrations, (Bai and Lung 2005; Brettar and Hofle 1992; Craig et. al. 2004; Oliver et. al. 2007) which may be processes dependent on stream order as well.

## **2. MATERIALS AND METHODS**

## ***2.1 Study Area***

The three ecoregions selected for this study are shown in Figure 1. Ecoregion 27 is the most westward of the three ecoregions. The average temperatures in the region range from a low of minus eight and high of 13 °C during January and a low of 19 and high of 37 °C during July. The temperature increases moving southwards in the ecoregion. The vegetation in the ecoregions was once a mostly mixed-grass prairie but has evolved into mostly cropland today. Range and grassland are mostly located in more rugged areas (Griffith et. al. 2004). The terrain of the region mainly consists of irregular plains and broad alluvial valleys. The hydrology of the area includes mostly intermittent streams with some perennial streams and a few natural lakes. There are some large rivers such as Red and Brazos River, which pass through the region as well. The larger rivers have braided sandy channels with mostly turbid flows. After heavy rains, the streams have stronger flows with large amounts of suspended sediments. The streams draining from rangeland contain less sediment load than the streams downstream of croplands. Many streams in the region have been channelized and/or impounded leading to unnatural flow regimes, higher erosion, and loss of the riparian forests. The precipitation in the area ranges from 560 to 965 mean annual millimeters (mm) and increases from west to east and the region is mostly semi-arid in nature. The wildlife in region mostly consists of white-tailed deer, mule deer, pronghorn, coyote, and jackrabbit. The land use in the region is dominated by dryland and irrigated cropland with some pasture and rangeland also occurring. Grazing for cattle, sheep and goat along with oil and gas production is also a common land use. (Griffith et. al. 2004; Woods et. al. 2005).



**Figure 13.** Study area comprising of ecoregions 27, 29, and 35 in Texas and Oklahoma.

Ecoregion 29 occurs in the north-central Texas, central Oklahoma, and southeastern Kansas and is located between ecoregions 27 and 35 the other two ecoregions included in this study. The region is transitional area between winter wheat growing areas in the ecoregion 27 and the forested low mountains of eastern Oklahoma. Vegetation is mainly comprised of forest, woodland, savannah, and prairie. Cropland is not very common and is restricted to valleys near channelized streams. Both intermittent and perennial streams can be found in the region, which have a low to moderate gradient due to the rolling plains type terrain found in the region. Several large rivers cross the ecoregions along with some large reservoirs and lakes to serve the urban areas located in the region. The temperature in the region ranges from a low of minus six and high of 14 °C during January and a low of 21 and high of 36 °C during July. The precipitation in

the region ranges from 780 to 1170 mean annual mm with higher precipitation occurring on the eastern side of the region. Wildlife in the region mainly consists of white-tailed deer, gray fox, bobcat, black-tailed jackrabbit, and prairie chicken etc. Main landuse in the area is pastureland and rangeland. Areas of woodland are also dominant, along with some cropland. Oil and gas production is a major landuse and the region also includes some major urban centers such as Dallas, Fort Worth, and Arlington in Texas. (Griffith et. al. 2004; Woods et. al. 2005).

Ecoregion 35 is the eastern region in the study area. The region comprises mainly a temperate coniferous forest with several species of pine along with hardwoods such as hickory and oak. Approximately one sixth of the region consists of cropland mainly within the Red River floodplain. Perennial streams are most common in the region but the flow can become limited during the summer months. The streams located in the forested regions normally have lower concentrations of suspended solids whereas the Red River the largest river flowing through the region is mostly turbid. The region lacks natural lakes but some reservoirs have been built within the region to account for inconsistent flow. The average temperatures range from a low of minus two and high of 16 °C during January and a low of 21 and high of 34 °C during July. The precipitation ranges between 1066 to 1422 mean annual mm with higher rainfall towards the east. The wildlife mainly consists of white-tailed deer, coyote, beaver, raccoon, muskrats, and rabbits. Major land uses include commercial pine plantations, timber production, livestock grazing, and oil and gas production. (Griffith et. al. 2004; Woods et. al. 2005).

The three ecoregions have very different characteristics as detailed above. Major differences include the amounts of precipitations and temperatures as well as the hydrology of each region. The type of vegetation changes as we move eastward from dryland and irrigated cropland in ecoregion 27 to a predominantly coniferous forest in ecoregion 35. The analysis of *E. coli*



relationship with stream orders and watershed area by ecoregion will be an important part of this study to assess if these differences in characteristics have any bearing on the *E. coli* relationships.

## ***2.2 GIS Methods***

Streams were delineated using the ArcHydro add-in for ArcMap. The flow direction and flow accumulation raster's obtained from Horizon (2014) were used as inputs in the ArcHydro stream definition tool. The stream threshold was selected at 5000 30 x 30 m cells (an area of 4.5 km<sup>2</sup>) for catchment definition based on the USGS Elevation Derivatives for National Applications (EDNA 2014). The stream raster is then used in the stream segmentation and drainage line processing tools to get the stream vectors. The Monitoring stations obtained from TCEQ and OCC are snapped using the Snap Pour Point tool to the stream segments if they are within 200 meter of the stream segment. The watersheds are delineated for each monitoring station point in ArcMap. The flow direction raster and the snap pour points are used as inputs in the watershed tool to get the watershed area. Urban area polygons are determined from the cropland GIS layer (NASS 2014) and are delineated for each watershed in ArcMap. The percentage of each watershed's urban area is then determined by dividing the area of the urban area within each watershed by the total watershed area. The wastewater outfall layers obtained from TCEQ and Oklahoma Department of Environmental Quality were used to determine the number of wastewater outfalls in the watersheds and the distance of the outfall to the monitoring sites. The stream order is obtained using the spatial analyst toolset in the ArcMap. The tool allows calculation of a stream order raster based on both Strahler and Shreve stream order systems.

## ***2.2 Statistical analysis***

*E. coli* data was collected by Texas Commission on Environmental Quality (TCEQ) and Oklahoma Conservation Commission (OCC) over the 7-year study period at irregular intervals and at different flow condition for many of the monitoring sites. Median and geometric mean of *E. coli* for each monitoring site was used in the study to normalize the effect of the too low or too high *E. coli* concentrations.

Monitoring stations that had minimum of five *E. coli* observations during the study period of January 2006 to May 2014 were included in the analysis. Cursory analysis of sites that had less than five observations showed that those observations were collected either too sporadically or in a short frame of time. These sites were also impacted by high and low concentrations more severely, which made the use of median or geometric mean inappropriate. Therefore, a cutoff was made at five *E. coli* observations per site to calculate median and geometric mean *E. coli*. The data obtained from GIS methods, which included stream orders (both Strahler and Shreve), watershed area, urban area, ecoregion number, and the wastewater outfall count, and the distance to the nearest outfall for each monitoring site, was joined with *E. coli* data. The final database included 743 monitoring stations for the study regions.

From this database, a subsets for unimpacted monitoring stations was extracted. For this purpose, the urban area percentage was limited to 10% or less for each monitoring site. Along with the urban area, the monitoring sites that had in their watersheds less than 10 wastewater outfalls with the nearest outfall at least 2.5 km away were chosen as unimpacted. The following two datasets were then analyzed using statistical techniques:

- d) All sites with at least five observations;
- e) Unimpacted watershed sites with 10 percent or less urban area and less than 10 wastewater outfalls with the nearest outfall at least 2.5 km away.

Other combinations of data were also tested to ensure the best definition of unimpacted sites. Combinations such as minimum distance to the nearest outfall ranging from 1 to 5 km and urban area percentages ranging from 5 to 25 percent were also tested but it was felt that the 10 percent urban area cap and the minimum outfall distance of 2.5 km provided the best unimpacted sites definition. Both of the above datasets were then analyzed in the JMP statistical software (JMP®, Version 11). The non-parametric analysis of the data was conducted using Wilcoxon/Kruskal-Wallis tests for Strahler stream orders with a null hypothesis of median or geometric mean *E. coli* being equal at each stream order. The same analysis was conducted at ecoregion scale as well. Further, the correlation between stream order and *E. coli* concentration was tested using Kendall  $\tau$  and Spearman's  $\rho$ .

Continuous nonlinear tests were used to evaluate if a nonlinear distribution would fit the *E. coli* data with respect to Shreve order and watershed size since they represent continuous variables. Data distributions modeled using nonlinear fit allow testing of the correlation using the parametric Pearson's  $r$  coefficient. Non-parametric Kendall  $\tau$  and Spearman's  $\rho$  tests were used to determine correlation between *E. coli* and watershed size/Shreve stream order when nonlinear models were not applicable. Kendall  $\tau$  and Spearman's  $\rho$  are monotonic correlation tests and apply well to the data similar to this study. These tests are also resistant to the effect of outliers. The Kendall's  $\tau$  approaches a normal distribution more rapidly than  $\rho$ , as  $N$ , the sample size, increases.  $\tau$  is also a rank based test and is therefore more resistant to small number of unusual values (Helsel and Hirsch 2002). The alpha threshold of 5 percent was used to determine the significance of the probability values obtained throughout the analysis.

### **3. RESULTS AND DISCUSSION**

### 3.1 Strahler stream order

The Strahler stream order for the study areas ranged from one to seven for all data. However, when only unimpacted watersheds are considered, the range is reduced to orders one through six due to reduction in the sampled data. Table 1 summarizes the number of monitoring site at each stream order for the three datasets analyzed in this study and the median and geometric mean *E. coli* concentrations for those sites. The table shows a consistent *E. coli* distribution for the datasets with increasing *E. coli* for the first three orders and a drop in the concentrations for later orders. Study conducted by Edge et. al. (2012) had found similar relationships between stream order and *E. coli* concentrations where the concentrations increased for the first three orders and then decreased. Another study by VDEQ (2003) had found the highest fecal coliform concentration at orders 3 and 4, which can also be seen in the table below. The median *E. coli* for each datasets stays below the geometric mean standard of 126 cfu/100ml required for primary contact recreation in Texas as well.

Table 23. Summary of *E. coli* by Strahler stream order for a) all monitoring sites, b) unimpacted sites.

Strahler stream order	a) All monitoring sites			b) Unimpacted sites		
	Number of monitoring sites	Median <i>E. coli</i>	Geometric mean of <i>E. coli</i>	Number of monitoring sites	Median <i>E. coli</i>	Geometric mean of <i>E. coli</i>
1	34	91.75	50.77	23	74.00	40.54
2	145	95.00	80.00	118	80.00	73.73
3	214	106.50	79.10	210	110.00	82.21
4	176	93.00	65.84	160	100.00	72.73
5	106	62.00	41.83	49	68.50	57.71
6	30	28.00	34.85	15	31.00	33.24
7	38	21.25	17.33			
Total	743	81.00	60.80		96.00	71.58

Since Strahler stream order follows discrete variable characteristics and along with non-normality of the *E. coli* data, the Wilcoxon/Kruskall-Wallis test is the most suitable

nonparametric test for this method. The tests were run using JMP software ((JMP<sup>®</sup>, Version 11) using JMP software procedures outlined by Schlotzhauer (2007).

The Wilcoxon/Kruskal-Wallis method tests the null hypothesis that the mean ranks of median or geometric mean *E. coli* at each stream order are equal. The test results are deemed significant, i.e. at least two score means are different, if the Chi Square p-value is less than  $\alpha = 0.05$ . The p-value for all data is less than  $<0.0001$  from which it can be concluded that there are at least two stream orders with different means. Similarly, the p-value of 0.0010 for unimpacted watersheds is also significant. The plots in Figure 2 and Figure 3 show the median and geometric mean *E. coli* plotted on the Y-axis and Strahler stream order on X-axis for all and unimpacted monitoring sites. From the figures we can see that the means stay fairly same for the first 4 stream orders and then drop significantly for order 5 and higher for both median and geometric mean.

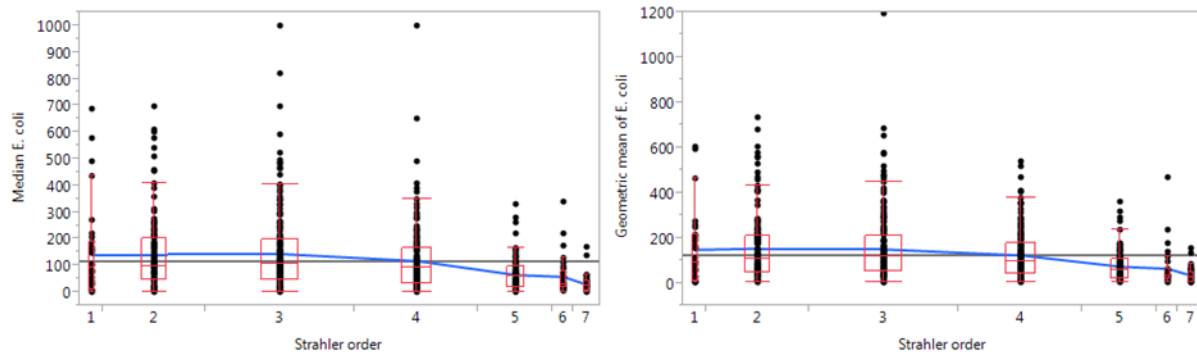


Figure 14: Distribution of median and geometric mean *E. coli* concentrations by Strahler stream order for all monitoring sites (the whiskers represent the maximum and minimum for the typical data quality scenario, black dots represent each monitoring site, blue line represents mean of each order, and grey line is the grand mean of the data).

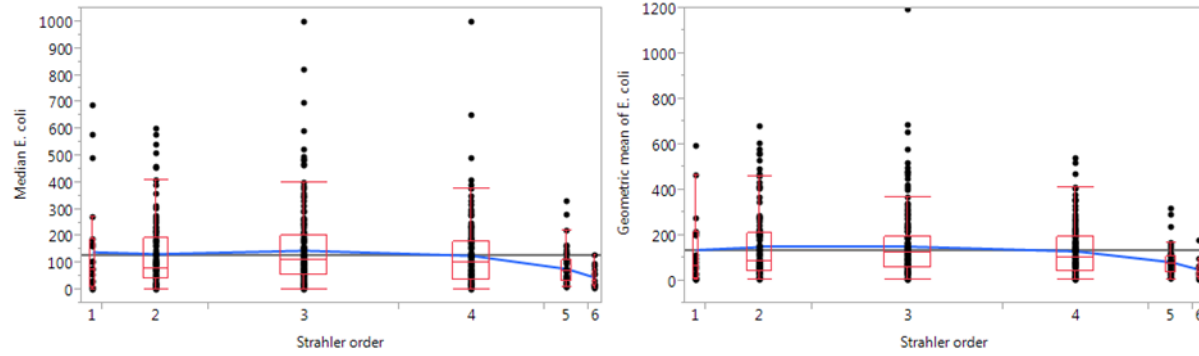


Figure 15: Distribution of median and geometric mean *E. coli* concentrations by Strahler stream order for unimpacted sites (the whiskers represent the maximum and minimum for the typical data quality scenario, black dots represent each monitoring site, blue line represents mean of each order, and grey line is the grand mean of the data).

The Wilcoxon/Kruskal-Wallis test for each of the ecoregions showed similar results to the ones for entire study area. The main challenge arising for the test at the ecoregion level was the lack of observations at certain stream orders. However, even considering this issue, the Wilcoxon test results were very similar, significance wise, for the three datasets. The dataset with all sites had significance probabilities of less than 0.05 (Table 2) for all ecoregions, which meant that at least two of the stream order score means were different in those datasets, whereas only ecoregion 27 showed significant probability for unimpacted sites. Table 2 summarizes the Chi square probabilities for each test.

Table 24: Wilcoxon/Kruskal-Wallis test probabilities by ecoregion for median and geometric mean *E. coli*.

Ecoregion	Median <i>E. coli</i>		Geometric mean <i>E. coli</i>	
	Probability > Chi Square		Probability > Chi Square	
	All monitoring sites	Unimpacted sites	All monitoring sites	Unimpacted sites
27	<0.0001	<0.0001	<0.0001	<0.0001
29	<0.0001	0.4009	<0.0001	0.3960
35	0.0085	0.9496	0.0205	0.9964

Nonparametric Kendall  $\tau$  and Spearman's  $\rho$  correlation tests were applied to the datasets with the results shown in Table 3. The dataset with all monitoring sites showed significant correlation probabilities but the correlation co-efficient for both were closer to 0 than -1. This meant that the test did show there was a statistically significant negative correlation between Strahler order and median or geometric mean *E. coli* but it was a weak correlation. The correlation test for unimpacted sites resulted in a non-significant probability apart from the Kendall's  $\tau$  for geometric mean *E. coli*. Since all correlation co-efficient were very close to zero from which it can be concluded that there is no significant correlation between stream order and median or geometric mean *E. coli* of unimpacted monitoring sites.

Table 25. Correlation values between median and geometric mean *E. coli* and Strahler stream order for the study area.

<b>Study Area</b>				
<b>Median <i>E. coli</i></b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2470	<0.0001	-0.0730	0.0804
Kendall's $\tau$	-0.1863	<0.0001	-0.0567	0.0698
<b>Geometric mean <i>E. coli</i></b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2422	<0.0001	-0.0813	0.0514
Kendall's $\tau$	-0.1823	<0.0001	-0.0633	0.0425

The correlation tests at the ecoregion level revealed similar results to the analysis of entire study area. Table 4 and Table 5 show the correlation co-efficient and the probability associated with

them. Ecoregion 27 showed much stronger correlation between *E. coli* and stream order for both datasets compared to the other two ecoregions but it was still lower than -0.50 for both Kendall  $\tau$  and Spearman's  $\rho$ . Ecoregion 29 and 35 showed lower correlation coefficients in both datasets. The probabilities showed that *E. coli* concentrations in the unimpacted monitoring sites were not correlated with the Strahler stream order. Another observation was the reduction in correlation coefficient values as the data moved eastward i.e. ecoregion 27 had the highest coefficients followed by 29 and lastly ecoregion 35. This signified that *E. coli* became less correlated to stream orders moving eastward.

Table 26. Correlation values between median *E. coli* and Strahler stream order at ecoregion level.

<b>Ecoregion 27</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4738	<0.0001	-0.4323	<0.0001
Kendall's $\tau$	-0.3644	<0.0001	-0.3337	<0.0001
<b>Ecoregion 29</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2786	<0.0001	-0.0615	0.3806
Kendall's $\tau$	-0.205	<0.0001	-0.0443	0.3990
<b>Ecoregion 35</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.1787	0.0032	0.0135	0.8601
Kendall's $\tau$	-0.1424	0.0017	0.0101	0.8630

Table 27. Correlation values between geometric mean *E. coli* and Strahler stream order at ecoregion level.



<b>Ecoregion 27</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.4636	<0.0001	-0.4283	<0.0001
Kendall's $\tau$	-0.3516	<0.0001	-0.3250	<0.0001
<b>Ecoregion 29</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2700	<0.0001	-0.0587	0.4031
Kendall's $\tau$	-0.1974	<0.0001	-0.0416	0.4275
<b>Ecoregion 35</b>				
Test	All monitoring sites		Unimpacted sites	
	Correlation co-efficient	Prob>  $\rho$	Correlation co-efficient	Prob>  $\rho$
Spearman's $\rho$	-0.1600	0.0084	0.0123	0.8725
Kendall's $\tau$	-0.1287	0.0044	0.0446	0.9374

### 3.2 Shreve stream order and watershed area

The *E. coli* data for both the Shreve order and the watershed area plots very similarly and in a nonlinear fashion. At low orders and smaller watershed areas, i.e. orders less than 15 and watershed areas less than 10,000 square kilometers (km), there is higher density of monitoring stations along with higher *E. coli* concentrations. As the stream order and watershed area increase, *E. coli* concentrations start to get smaller and plots closer to the x-axis. This behavior of the data follows a decreasing exponential curve and application of nonlinear exponential models will show if those are a fit for this data. Figure 4 and Figure 5 show the application of the two and three parameter exponential distribution fits using the procedures described by Walsh (2013). The dataset with all monitoring sites had the best fit with the two and three parameter

exponential distributions for both Shreve order and watershed size. Based on the lowest AICc value, the three parameter (3P) exponential distribution was the best fit. Even though the 3P exponential distribution showed a very good fit based on AIC, there were monitoring stations which plotted far from the curve, which means if this fit was adopted those stations would not be represented.

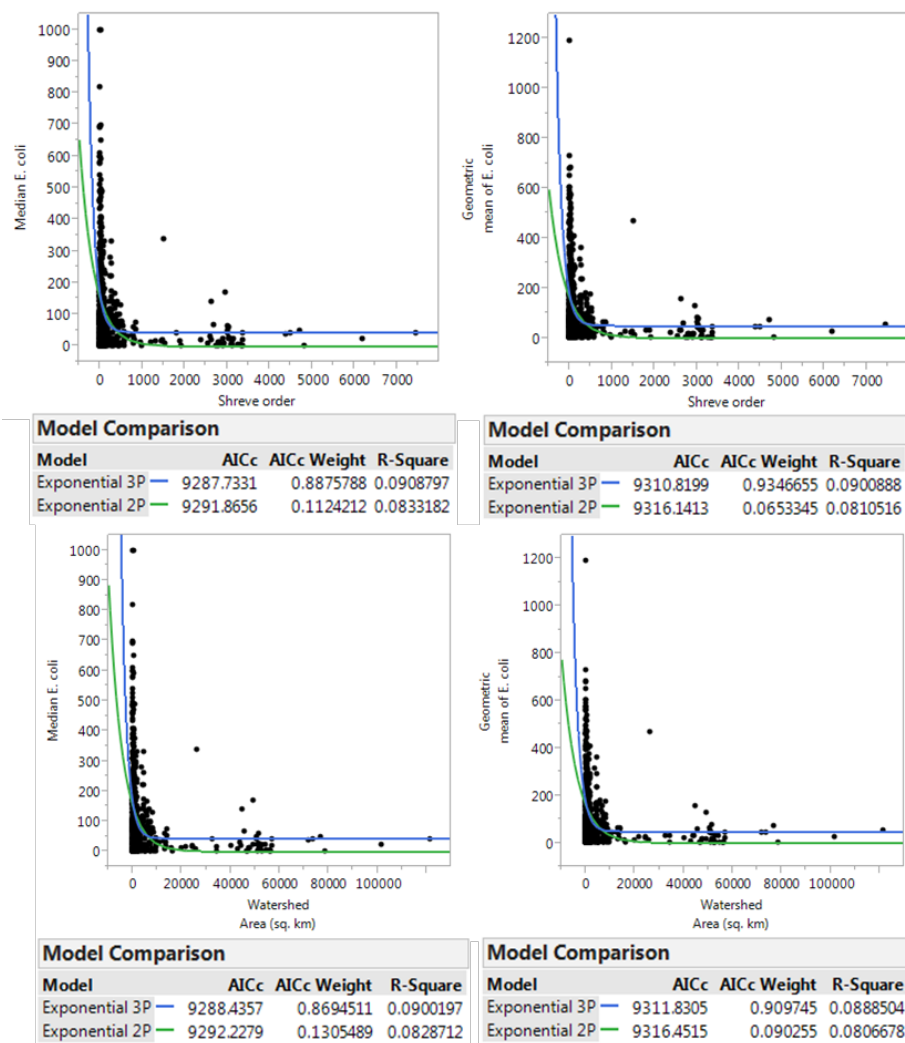


Figure 16. Nonlinear model fits for *E. coli* by Shreve stream order and watershed area for all monitoring sites.

For the unimpacted watersheds, none of the nonlinear distributions provided a good fit as seen in Figure 5; therefore, only non-parametric correlation tests were applicable.

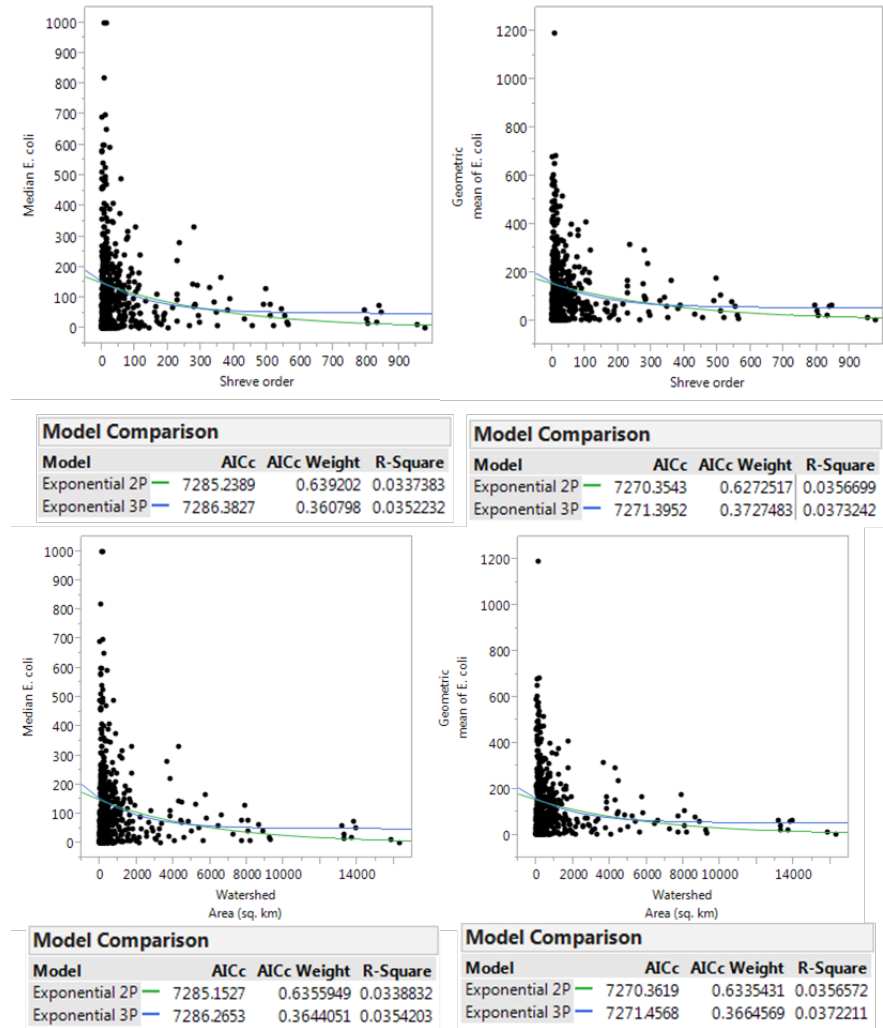


Figure 17. Nonlinear model fits for *E. coli* by Shreve stream order and watershed area for unimpacted monitoring sites.

Figure 4 and Figure 5 also show how similar the Shreve order and watershed area *E. coli* distributions are. Both datasets plot very similarly and the nonlinear fits for them are alike as well. From this, it can be concluded that Shreve order and watershed area depict the same characteristic for the *E. coli* distributions. Another observation that can be gleaned from these

figures is that most of the higher *E. coli* concentrations plot at lower stream order and watershed sizes. At higher orders and watershed sizes, *E. coli* concentrations stay lower than 150 MPN/100ml. These results also compare well with the study conducted by Harmel et. al. (2010) where it was found that *E. coli* concentrations decrease with increase in watershed scales. Non-parametric test are more suitable for these distribution due to the existence of outliers but the 3P exponential distribution may also be a good model to assess distribution of *E. coli*.

The correlation values for the *E. coli* by Shreve stream order and watershed area are very similar to earlier tests conducted on Strahler stream order method. The dataset with all sites show weak yet significant negative correlation between *E. coli* and Shreve stream order (Table 18) and between *E. coli* and watershed size (Table 6 and Table 7). On the other hand, the unimpacted datasets do not show much correlation between Shreve order and *E. coli*. Both Spearman's  $\rho$  and Kendall's  $\tau$  for geometric mean of *E. coli*, and Kendall's  $\tau$  for the median *E. coli* show a significance probability of less than 5 percent (Table 6 and Table 7) but since the correlation coefficient are very close to zero it can be concluded that the correlation is not present. The Pearson's  $r$  test, done on the 3P exponentially modeled distributions for all monitoring sites, gives a correlation coefficient of approximately -0.60 for both the Shreve order and the watershed size. The exponential model calculates a fitted curve based on the median or geometric mean *E. coli* of each monitoring site. As seen in Figure 4 and Figure 5 the fitted curve for 3P exponential model goes through many of actual data points but there are some outlier points with high *E. coli* concentrations and higher stream order/watershed sizes that are not well represented by the curve. In addition, a large number of data points below the curve and near the origin point of zero are also not close to the curve but that may be less of a concern since the curve at that point is below the TAC limit of 126 cfu/100ml.

Test	Median <i>E. coli</i>				Geometric mean <i>E. coli</i>			
	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2711	<0.0001	-0.0894	0.0321	-0.2688	<0.0001	-0.0993	0.0172
Kendall's $\tau$	-0.1910	<0.0001	-0.0639	0.0237	-0.1884	<0.0001	-0.0698	0.0133
Pearson's $r^*$	-0.5949				-0.6002			

\* Using the three parameter exponential model

Table 28. Correlation values between median and geometric mean *E. coli* and Shreve stream order.

Test	Median <i>E. coli</i>				Geometric mean <i>E. coli</i>			
	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2573	<0.0001	-0.0734	0.0788	-0.2571	<0.0001	-0.0822	0.0488
Kendall's $\tau$	-0.1829	<0.0001	-0.0551	0.0491	-0.1812	<0.0001	-0.0605	0.0305
Pearson's $r^*$	-0.5961				-0.6093			

\* Using the three parameter exponential model

Table 29. Correlation values between median and geometric mean *E. coli* and watershed size.

The correlation analysis at the ecoregion scale also had similar results to the Strahler method.

The dataset with all sites showed a weak but significant negative correlation for both Shreve stream order (Table 20) and watershed size (Table 21) with *E. coli* for each ecoregion. The unimpacted datasets for Shreve order and watershed size showed a weak negative correlation with *E. coli* in ecoregion 27 but no correlation in the other two ecoregions.

Similar to findings from the Strahler method, the correlation coefficients for both Shreve order and watershed size were highest for ecoregion 27 and lowest for ecoregion 35 showing a decreasing trend going eastward.

	Median <i>E. coli</i>				Geometric mean <i>E. coli</i>			
Ecoregion 27								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ
Spearman's ρ	-0.4903	<0.000 1	-0.4416	<0.000 1	-0.4665	<0.000 1	-0.4158	<0.000 1
Kendall's τ	-0.3538	<0.000 1	-0.3143	<0.000 1	-0.3278	<0.000 1	-0.2883	<0.000 1
Ecoregion 29								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ
Spearman's ρ	-0.3034	<0.000 1	-0.0994	0.1562	-0.2985	<0.000 1	-0.1028	0.1423
Kendall's τ	-0.2087	<0.000 1	-0.0659	0.1675	-0.2044	<0.000 1	-0.0662	0.1644
Ecoregion 35								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ
Spearman's ρ	-0.2025	0.0008	-0.0104	0.8918	-0.1942	0.0013	-0.0292	0.7026
Kendall's τ	-0.1490	0.0003	-0.0085	0.8705	-0.1439	0.0005	0.0247	0.6361

Table 30. Correlation values between median and geometric mean *E. coli* and Shreve stream order at ecoregion level.

Median <i>E. coli</i>				Geometric mean <i>E. coli</i>				
Ecoregion 27								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ	Correlati on co-efficient	Prob>  ρ

Spearman's $\rho$	-0.4747	<0.000 1	-0.4397	<0.000 1	-0.4607	<0.000 1	-0.4166	<0.000 1
Kendall's $\tau$	-0.3438	<0.000 1	-0.3131	<0.000 1	-0.3249	<0.000 1	-0.2900	<0.000 1
<b>Ecoregion 29</b>								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2672	<0.000 1	-0.0527	0.4532	-0.2661	<0.000 1	-0.0602	0.3914
Kendall's $\tau$	-0.1857	<0.000 1	-0.0371	0.4316	-0.1843	<0.000 1	-0.0415	0.3778
<b>Ecoregion 35</b>								
Test	All monitoring sites		Unimpacted sites		All monitoring sites		Unimpacted sites	
	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$	Correlati on co- efficient	Prob>  $\rho$
Spearman's $\rho$	-0.2011	0.0009	0.0058	0.9392	-0.1896	0.0018	-0.0036	0.9628
Kendall's $\tau$	-0.1498	0.0003	-0.0012	0.9822	-0.1406	0.0006	-0.0084	0.8706

Table 31. Correlation values between median and geometric mean *E. coli* and watershed size at ecoregion level.

#### 4. CONCLUSIONS

The overall goal of this research was to understand and predict a relationship between stream order and *E. coli* and evaluate the application of that relationship to the current recreational use standards. The data analysis of all monitoring sites which included urban area and wastewater outfalls showed a weak but significant negative correlation of *E. coli* concentrations with Shreve and Strahler orders and watershed area. Analysis of the data showed that at larger watershed sizes and stream orders, *E. coli* concentrations were significantly lower. At smaller stream orders and watersheds, the *E. coli* concentrations ranged from very low to very high. It can be argued that the wider range of *E. coli* concentrations at smaller stream orders and watershed sizes was a

driving factor in making the correlation coefficients weak. The *E. coli* concentrations did become lower as the stream orders and watersheds became larger which points to the dilution factor as suggested by Lyautey et. al. (2010, 2011). But site specific studies may be a better way to understand *E. coli* relationships at smaller watershed and stream order scales where *E. coli* concentration range from very low to very high.

The data analysis of unimpacted sites showed that there was no correlation between *E. coli* concentrations and Shreve and Strahler orders as well as watershed area. Many of the monitoring sites which had higher stream orders, e.g. sites at orders 6 and 7 for Strahler order, did not qualify as unimpacted watersheds due the presence of wastewater outfalls. With availability of more data such as discharge volume, the criteria for exclusion of wastewater outfalls may be modified to include more monitoring sites. This would allow for a better definition of unimpacted watersheds.

The wastewater outfall impact on the streams was another significant finding of this study. The datasets that included wastewater outfalls, showed a weak but statistically significant negative correlation with *E. coli*, but when the outfalls were removed, there was no correlation. This suggests that wastewater outfalls are a big influence on stream *E. coli* and further research on just the sites that are impacted by these outfalls may provide more valuable information on *E. coli* concentrations in the streams. Further studies with the types and volume of discharges made by wastewater outfalls may also help to assess the impact of wastewater outfalls on the water bodies.

The comparison of the three ecoregions also revealed that the median and geometric mean *E. coli* concentrations and their correlation with both stream order methods and watershed area decreased moving eastward. Ecoregion 27, which is the western most ecoregion in this study and



has a more arid to semi-arid environment compared to the other ecoregions showed the highest *E. coli* concentration as well as correlations. Ecoregion 35, the eastern most region and with wettest climate of the three regions, has the lowest *E. coli* concentrations and correlations. This finding can be a good starting point for a future study that can include additional environmental factors such as precipitation, temperatures, types of streams (seasonal or perennial) etc., which make up each region and can contribute to the hydrology of these areas and its impact on the *E. coli* transport.

The study utilized Strahler and Shreve stream orders and watershed area as the variables to which the *E. coli* was correlated. Shreve stream order and watershed area provide similar information with respect to basin size and results based on them are very similar, it is recommended for future studies only one of the two variables should be used. The Strahler stream order method produces a smaller number of stream orders that helps in conducting the statistical analysis when comparing each stream order. This method is a good tool for comparing individual stream order numbers at ecoregion or river basin scales i.e. compare stream order 1 in ecoregion 27 with stream order 1 in ecoregion 29. Out of the three methods watershed area provides the most relevant information with respect to the scale of the basin but Strahler and Shreve stream order can be the preferred method when comparing streams with similar characteristics.

For future research, inclusion of other variables such as flow volume at the time of the observation or a classification of the flow, such as low, normal, or high, which could identify if the sample was collected after a storm or in drought or normal conditions, may be beneficial as flow can influence *E. coli* concentrations. Other variables such as temperature at the time of sample collection, and time of collection in terms of seasonality, such as summer, spring etc.,

may be valuable in understanding *E. coli* relationships. A larger number of *E. coli* observations at monitoring sites at higher stream orders may also help provide a better and clearer understanding of the dilution effect. Further, a study on the major contributors, fate, and transport of *E. coli* in each ecoregion may also help understand the decreasing *E. coli* and correlation trends when moving eastward.

## 5. ACKNOWLEDGMENTS

The continued support by the Texas State Soil and Water Conservation Board in related projects to better understand bacteria fate and transport and mitigate its negative impacts is appreciated.

## 6. REFERENCES

- Arnold, R. D. and J. P. Walling. 2007. "Waterborne Pathogens in Urban Watersheds." *Journal of Water Health* 5(1):149-162
- Bai, S. and W. S. Lung. 2005. "Modeling Sediment Impact on the Transport of Fecal Bacteria." *Water Research* 39(20):5232-5240.
- Berry, D. F. and C. Hagedorn. 1991. "Soil and Groundwater Transport of Microorganisms." *Assessing Ecological Risks of Biotechnology* 15:57-73.
- Brettar, I. and M. G. Hofle. 1992. "Influence of Ecosystematic Factors on Survival of *Escherichia Coli* after Large- Scale Release into Lake Water Mesocosms." *Applied and Environmental Microbiology* 58(7):2201-2210.
- Byappanahalli, M., M. Fowler, D. Shively and R. Whitman. 2003. "Ubiquity and Persistence of *Escherichia Coli* in a Midwestern Coastal Stream." *Applied Environmental Microbiology* 69(8):4549-4555. doi: 10.1128/AEM.69.8.4549-4555.2003.
- Craig, D. L., H. J. Fallowfield and N. J. Cromar. 2004. "Use of Microcosms to Determine Persistence of *Escherichia Coli* in Recreational Coastal Water and Sediment and Validation with in Situ Measurements." *Journal of Applied Microbiology* 96(5):922-930. doi: 10.1111/j.1365-2672.2004.02243.x.

Edge, T. A., A. El-Shaarawi, V. Khan, C. Gannon, R. Jokinen, I. U. H. Kent, W. Phillips, D. Koning, J. Lapen, N. Wilkes, R. Miller, W. Scott, H. Topp, A. van Bochove, I. Neumann, E. Shtepani, G. Robertson and E. Schreier. 2012. "Investigation of an *Escherichia Coli* Environmental Benchmark for Waterborne Pathogens in Agricultural Watersheds in Canada." *Journal of Environmental Quality* 41(1):21-30. doi: 10.2134/jeq2010.0253.

Elevation Derivatives for National Applications (EDNA). 2014. "Elevation Derivatives for National Applications". Available at <http://edna.usgs.gov/>. Accessed 08/27/2014.

Griffith, G.E., S.A. Bryce, J.M. Omernik, J.A. Comstock, A.C. Rogers, B. Harrison, S.L. Hatch and D. Bezanson. 2004. "Ecoregions of Texas (Color Poster with Map, Descriptive Text, and Photographs).": Reston, Virginia, U.S. Geological Survey (map scale 1:2,500,000). Available at [http://www.epa.gov/naaujydh/pages/ecoregions/tx\\_eco.htm](http://www.epa.gov/naaujydh/pages/ecoregions/tx_eco.htm). Accessed 06/21/2014.

Haan, C. T. 2002. Statistical Methods in Hydrology. 2nd ed. Ames, Iowa: Iowa State Press.

Halliday, E. and R. J. Gast. 2011. "Bacteria in Beach Sands: An Emerging Challenge in Protecting Coastal Water Quality and Bather Health." *Environmental Science & Technology* 45(2):370-379. doi: 10.1021/es102747s.

Harmel, R. D., T. Gentry, R. Karthikeyan, and R. Srinivasan. 2010. "Effects of Agricultural Management, Land Use, and Watershed Scale on *E. Coli* Concentrations in Runoff and Streamflow." *Transactions of the ASABE* 53(6):1833-1841.

Helsel, D. R. and R. M. Hirsch. 2002. "Chapter A3 Statistical Methods in Water Resources." in Techniques of Water-Resources Investigations of the United States Geological Survey Book 4: U. S. Geological Survey.

Horizon. 2014. "NHDPlusV2 Data" Available at [http://www.horizon-systems.com/NHDPlus/NHDPlusV2\\_data.php](http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php). Accessed 05/26/2014.

Horton, Robert E. 1945. "Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology." *Bulletin of the Geological Society of America* 56(3):275-370. doi: 10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2

Ichim, Ionita. 1990. "The Relationship between Sediment Delivery Ratio and Stream Order; a Romanian Case Study." Pp. 79-86, Vol. 189, edited by D. E. Walling, A. Yair and S. Berkowicz. [Louvain]: [Louvain], International: International Association of Hydrological Sciences.

JMP<sup>®</sup>, Version 11. SAS Institute Inc., Cary, NC, 1989-2007.

Krometis, L.A. H., G. W. Characklis, O. D. Simmons, C. A. Likirdopulos, M. D. Sobsey and M. J. Dilts. 2007. "Intra- Storm Variability in Microbial Partitioning and Microbial Loading Rates." *Water Research* 41(2):506-516. doi: 10.1016/j.watres.2006.09.029.

Lyautey, E., G. Wilkes, J. J. Miller, E. Van Bochove, H. Schreier, W. Koning, T. A. Edge, D. R. Lapen and E. Topp. 2011. "Variation of an Indicator of *Escherichia Coli* Persistence from Surface Waters of Mixed- Use Watersheds, and Relationship with Environmental Factors." *Annales de Limnologie - International Journal of Limnology* 47(1):11-19. doi: 10.1051/limn/2010033.

Lyautey, E., Z. Lu, A. Scott, T. Berkers, E. Topp, D. R. Lapen, G. Wilkes and T. A. Edge. 2010. "Distribution and Diversity of *Escherichia Coli* Populations in the South Nation River Drainage Basin, Eastern Ontario Canada." *Applied and Environmental Microbiology* 76(5):1486-1496. doi: 10.1128/AEM.02288-09.

National Agricultural Statistics Service (NASS). 2014. "CropScape – Cropland Data Layer". Available at <http://nassgeodata.gmu.edu/CropScape/>. Accessed 05/26/2014.

Oliver, D. M., C. D. Clegg, A. L. Heathwaite and P. Haygarth. 2007. "Preferential Attachment of *Escherichia Coli* to Different Particle Size Fractions of an Agricultural Grassland Soil." *Water, Air, & Soil Pollution* 185(1-4):369-75. doi: 10.1007/s11270-007-9451-8.

Omernik, J. M. and R. G. Bailey. 1997. "Distinguishing between Watersheds and Ecoregions." *Journal of the American Water Resources Association* 33(5):935-949.

Pachepsky, Y. A., D. R. Shelton and D. R. Pachepsky. 2011. "*Escherichia Coli* and Fecal Coliforms in Freshwater and Estuarine Sediments." *Critical Reviews in Environmental Science and Technology* 41(12):1067-1110. doi: 10.1080/10643380903392718.

Poole, G. C., J. B. Dunham, D. M. Keenan, M. Deppman, S. T. Sauter, D. A. McCullough, C. Mebane, J. C. Lockwood, D. A. Essig, M. R. Hicks, D. J. Sturdevant, E. J. Materna, S. A. Spalding and J. Risley. 2004. "The Case for Regime- Based Water Quality Standards." *BioScience* 54(2):155-161.

Schlotzhauer, S. D. 2007. *Elementary Statistics Using JMP®*. Cary, NC: SAS Institute Inc.

Shreve, R. L. 1966. "Statistical Law of Stream Numbers." *The Journal of Geology* 74(1):17-37. doi: 10.1086/627137.

Strahler, A. N. 1957. "Quantitative Analysis of Watershed Geomorphology." *American Geophysical Union Transactions* 38(6):913-920. doi: 10.1029/TR038i006p00913.

Texas Commission on Environmental Quality (TCEQ). 2010. Texas Administrative Code, Title 30 Environmental Quality, Part 1 Texas Commission on Environmental Quality, Chapter 307 Texas Surface Water Quality Standards, Volume 307.7 Site-Specific Uses and Criteria.

Texas Commission on Environmental Quality (TCEQ). 2014. "Recreational Use Attainability Analyses". Available at <https://www.tceq.texas.gov/waterquality/standards/ruaas>. Accessed 08/28/2014.

Texas Commission on Environmental Quality (TCEQ). 2014a. "Permitted Wastewater Outfalls". Available at <http://www.tceq.state.tx.us/gis/download-tceq-gis-data>. Accessed 07/12/2014.

Texas Commission on Environmental Quality (TCEQ). 2014, updated daily. Surface Water Quality Monitoring Information System, 09/01/1968-05/02/2014 Bacteria-ER-27-29-35-050214.xlsx. Compiled by Data Management & Analysis Team. Austin, TX USA. Data Request ID 210.

Tarboton, D. G. 1997. "A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models." *Water Resources Research* 33(2):309-319.

United States Environmental Protection Agency (USEPA). 2014a. "National Summary of Impaired Waters and TMDL Information". Available at [http://iaspub.epa.gov/waters10/attains\\_nation\\_cy.control?p\\_report\\_type=T](http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T). Accessed 08/18/2014.

United States Environmental Protection Agency (USEPA). 2014b. "Level III and IV Ecoregions of the Continental United States". Available at [http://www.epa.gov/wed/pages/ecoregions/level\\_iii\\_iv.htm](http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). Accessed 08/18/2014.

United States Environmental Protection Agency (USEPA). 2014c. "EPA Clip N Ship Application". Available at <https://edg.epa.gov/clipship/>. Accessed 11/16/2014.

Virginia Department of Environmental Quality (VDEQ). 2003. The Quality of Virginia Non-Tidal Streams: First Year Report: Virginia Department of Environmental Quality. Available at <http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/ProbabilisticMonitoring/report1.pdf>. Accessed 06/30/2014.

Walsh S. 2013. When a Linear Model Just Won't Do: Fitting Nonlinear Models Using JMP®. JMP Discovery Conference. SAS Institute Inc. Available at [http://www.jmp.com/about/events/summit2013/resources/Paper\\_Susan\\_Walsh.pdf?utm\\_source=newswire1013&utm\\_medium=newsletter&utm\\_campaign=summit2013](http://www.jmp.com/about/events/summit2013/resources/Paper_Susan_Walsh.pdf?utm_source=newswire1013&utm_medium=newsletter&utm_campaign=summit2013). Accessed 02/18/2015.

Woods, A.J., J.M. Omernik, D.R. Butler, J.G. Ford, J.E. Henley, B.W. Hoagland, D.S. Arndt, and B.C. Moran. 2005. Ecoregions of Oklahoma (color poster with map, descriptive text, summary tables, and photographs): Reston, Virginia, U.S. Geological Survey (map scale 1:1,250,000). Available at [http://www.epa.gov/naaujydh/pages/ecoregions/ok\\_eco.htm](http://www.epa.gov/naaujydh/pages/ecoregions/ok_eco.htm). Accessed 06/21/2014.

Zăvoianu I. 1985. Morphometry of drainage basins, Editura Academiei Republicii Socialiste România, București. Amsterdam. Elsevier Science.

## Appendix C

“Distribution of *E. coli* levels and recreation use as  
factor of stream order in the Central Great Plains,  
Central Oklahoma-Texas Plains, and South Central  
Plains Ecoregions”

by

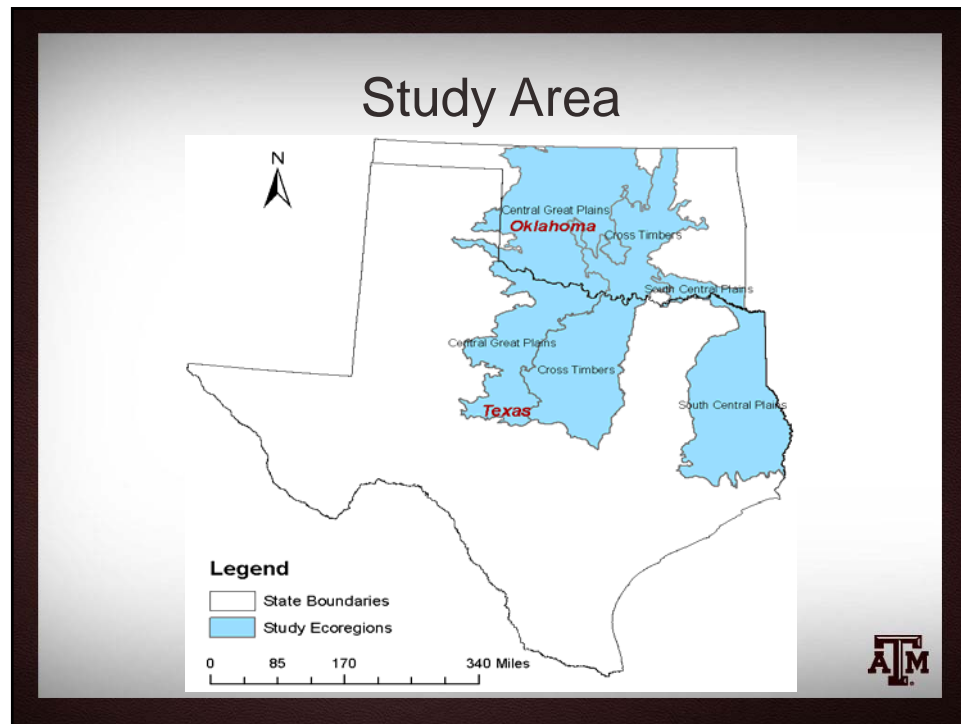
Khurram Rafi, Kevin Wagner, Amanda Dube, Terry  
Gentry, and Raghupathy Karthikeyan



## Objectives

- Assess stream order and watershed size of all sites used in the study using GIS.
- Evaluate correlations between *E. coli* concentrations and stream orders (plus watershed size) as impacted by ecoregion.
- Make recommendations on recreational standards based on the study findings.





## Data Sources and Characteristics

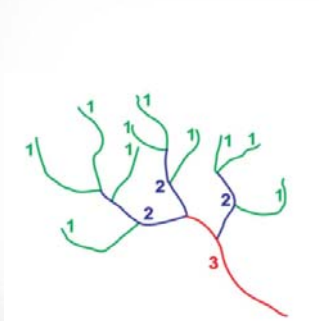
- *E. coli* and Monitoring Station Data Sources:
  - Texas Commission on Environmental Quality (TCEQ)
  - Oklahoma Conservation Commission (OCC).
- Date Range:
  - TCEQ – 1/1/2006 – 02/2014
  - OCC – 1/1/2006 – 12/2013.
- Median *E.coli* for monitoring sites with minimum 5 *E. coli* observations used in the analysis.
- 741 Monitoring stations included in the 3 ecoregions based on the dates and minimum observations.



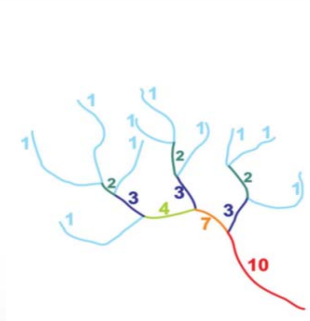


## Stream Order Systems

Strahler stream order



Shreve stream order

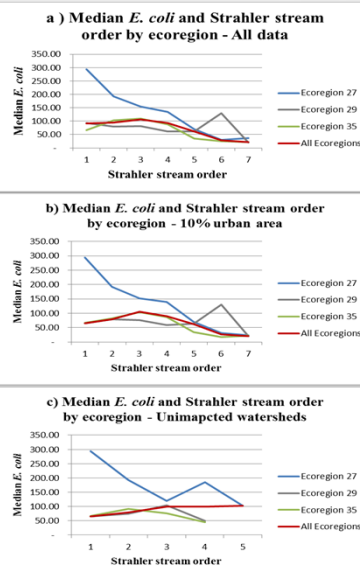


## Methods

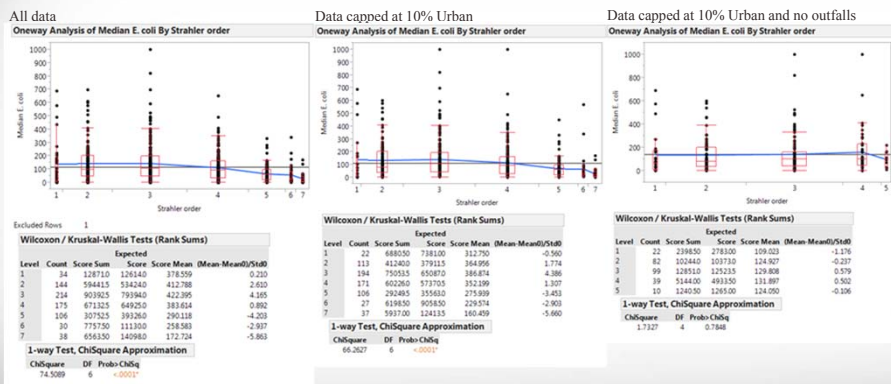
- ArcGIS used to delineate streams, stream orders, and watershed size.
- Three datasets for analysis:
  - All monitoring sites (741 sites);
  - Sites with 10 percent or less urban area in their watersheds including wastewater outfalls (670 sites);
  - Unimpacted watershed sites with 10 percent or less urban area and no wastewater outfalls in their watersheds (252 sites).
- Parametric and Non-Parametric correlation tests applied based on the data characteristics.



## Strahler Stream Order – *E. coli* Comparison

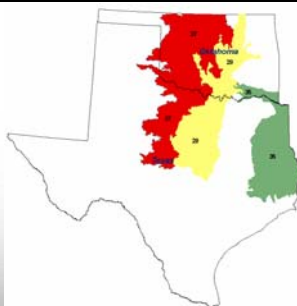


## Strahler Stream Order Wilcoxon/Kruskal-Wallis Test



## Strahler Stream Order - Wilcoxon/Kruskal-Wallis Test at Ecoregion Scale

Ecoregion	Probability > Chi Square		
	All monitoring sites	<10% Urban sites	Unimpacted sites
27	<0.0001	<0.0001	0.2446
29	<0.0001	0.0002	0.4076
35	0.0085	0.0144	0.5716



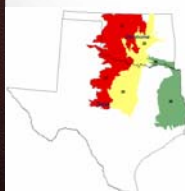
## Strahler Stream Order Correlation Tests

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2491	<0.0001	-0.2346	<0.0001	0.0622	0.3254
Kendall's $\tau$	-0.1882	<0.0001	-0.1786	<0.0001	0.0483	0.3128

Correlation values between *E. coli* and Strahler stream order.



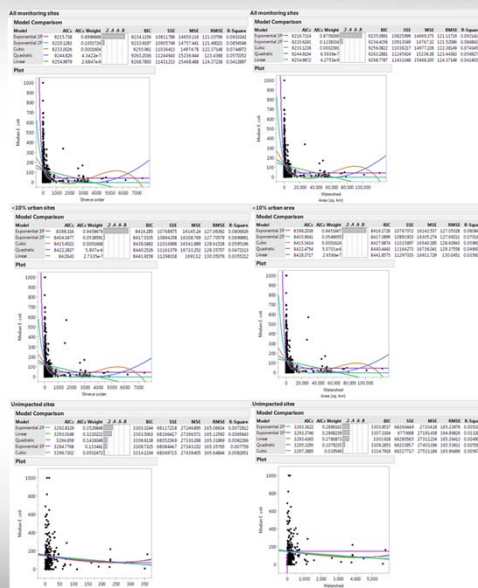
## Strahler Stream Order Correlation Tests at Ecoregion Scale



Ecoregion 27						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.4795	<0.0001	-0.4476	<0.0001	-0.1444	0.1768
Kendall's $\tau$	-0.3696	<0.0001	-0.3447	<0.0001	-0.116	0.153
Ecoregion 29						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2785	<0.0001	-0.2431	<0.0001	0.0673	0.5405
Kendall's $\tau$	-0.2051	<0.0001	-0.1778	<0.0001	0.0592	0.4814
Ecoregion 35						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.1787	0.0032	-0.1627	0.0131	-0.154	0.1784
Kendall's $\tau$	-0.1424	0.0017	-0.1359	0.0055	-0.1177	0.1846



## Nonlinear Tests – *E. coli* by Shreve Order & Watershed Size



## Shreve Stream Order - Correlation Tests

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2714	<0.0001	-0.2571	<0.0001	0.0733	0.2463
Kendall's $\tau$	-0.1914	<0.0001	-0.1824	<0.0001	0.0509	0.24
Pearson's $r^*$	-0.5963		-0.5939			

\* Using the three parameter exponential model

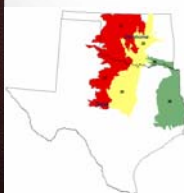
Correlation values between *E. coli* and Shreve stream order.



## Shreve Stream Order - Correlation Tests at Ecoregion Scale

Ecoregion 27						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.4874	<0.0001	-0.4626	<0.0001	-0.2108	0.0474
Kendall's $\tau$	-0.3519	<0.0001	-0.3340	<0.0001	-0.1508	0.3900
Pearson's $r^*$	-0.6218		-0.6297			
Ecoregion 29						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.3036	<0.0001	-0.2692	<0.0001	0.1074	0.3278
Kendall's $\tau$	-0.2089	<0.0001	-0.1832	<0.0001	0.0935	0.2231
Pearson's $r^*$	-0.6569		-0.6476			
Ecoregion 35						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2025	0.0008	-0.1869	0.0043	-0.1405	0.2198
Kendall's $\tau$	-0.1490	0.0003	-0.1415	0.0015	-0.0966	0.2239
Pearson's $r^*$	-0.5162		-0.5200			

\* Using the three parameter exponential model



## Watershed Size - Correlation Tests

Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2572	<0.0001	-0.2441	<0.0001	0.1292	0.0405
Kendall's $\tau$	-0.183	<0.0001	-0.1754	<0.0001	0.0819	0.0539
Pearson's $r^*$	-0.6014		-0.5967			

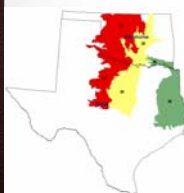
\* Using the three parameter exponential model

Correlation values between *E. coli* and watershed area.



## Watershed Size - Correlation Tests at Ecoregion Scale

Ecoregion 27						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.4708	<0.0001	-0.4586	<0.0001	-0.1907	0.0735
Kendall's $\tau$	-0.341	<0.0001	-0.3318	<0.0001	-0.1379	0.0565
Pearson's $r^*$	-0.622		-0.626			
Ecoregion 29						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2671	<0.0001	-0.2414	<0.0001	0.2376	0.0285
Kendall's $\tau$	-0.1859	<0.0001	-0.1668	<0.0001	0.1667	0.0248
Pearson's $r^*$	-0.6669		-0.6516			
Ecoregion 35						
Test	All monitoring sites		<10% Urban sites		Unimpacted sites	
	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p	Correlation co-efficient	Prob> p
Spearman's $\rho$	-0.2011	0.0009	-0.168	0.0104	-0.0712	0.5355
Kendall's $\tau$	-0.1498	0.0003	-0.1325	0.0028	-0.0564	0.4682
Pearson's $r^*$	-0.5237		-0.5233			



\* Using the three parameter exponential model



## Conclusions

- Wastewater outfall impacted datasets - weak but significant negative correlation between *E. coli* and stream orders and *E. coli* and watershed size.
- Unimpacted dataset - no correlation between *E. coli* and stream orders and *E. coli* and watershed size.
- Median *E. coli* and correlation coefficients showed a decreasing trend for ecoregions moving eastward.
- Dilution effect on *E. coli* concentrations at higher stream orders and watershed sizes.
- Results for each of Strahler order, Shreve order, and watershed size were similar for each dataset. This means the three variables are a good substitute of each other since they provide similar results.
- Based on the results the recreational standards should be based on the site specific study for each segment.



## Future Recommendations

- Inclusion of variables such as stream flow, temperature, and seasonality etc.
- Inclusion of wastewater type and discharge volume for each site as a factor.
- Increasing study area by incorporating more ecoregions.
- Compare more arid to semi arid ecoregions with wetter ecoregions to confirm higher *E. coli* pattern in arid to semi arid regions.
- Analysis of factors impacting each ecoregion *E. coli* concentrations i.e. agricultural practices, major industries etc.





**Thank You!**





## Appendix D

# *E. coli* Concentration and Recreation Use as a Function of Stream Order and Watershed Size

While statewide water quality standards have been developed to protect waters for recreational use, recently a case has been made for regime-based standards such as temperature, sediment and stream order (Poole et al., 2004). In central Texas, Harmel et al. (2010) found decreasing *E. coli* concentrations when aggregating sampled segments from upstream to downstream. This trend occurred with both “impacted” (land use of mixed rural with diaries, WWTPs, small communities) and “unimpacted” (land use of mixed rural) streams of the same region.

Further, Harmel et al. (2010) found that the “unimpacted” stream would be in violation of the Texas Water Quality Standards even though the anthropogenic impact was minor. Similarly, Lyautey et al. (2010) established that Shreve (1966) stream link was inversely correlated to *E. coli* concentrations. Further, Edge et al. (2012) used Horton (1945) – Strahler (1957) stream order and found that small streams (order  $\leq 3$ ) had higher *E. coli* concentrations than large streams (order  $>3$ ). Although the reason for decreased *E. coli* densities related to stream order was chiefly cited as dilution (Lyautey et al., 2010; Lyautey et al., 2011), it was also hypothesized that turbidity/sediment processes, which may also be dependent on stream order, influenced *E. coli* concentrations (Bai and Lung, 2005; Brettar and Hofle, 1992; Craig et al., 2004; Oliver et al., 2007).

## Objectives

To evaluate this and its implications on Texas Water Quality Standards and watersheds, this project:

- Determined the relationship between stream order and *E. coli* concentrations
- Determined the relationship between stream order and recreational use
- Made recommendations regarding the application of findings to water quality standards and recreational use classification

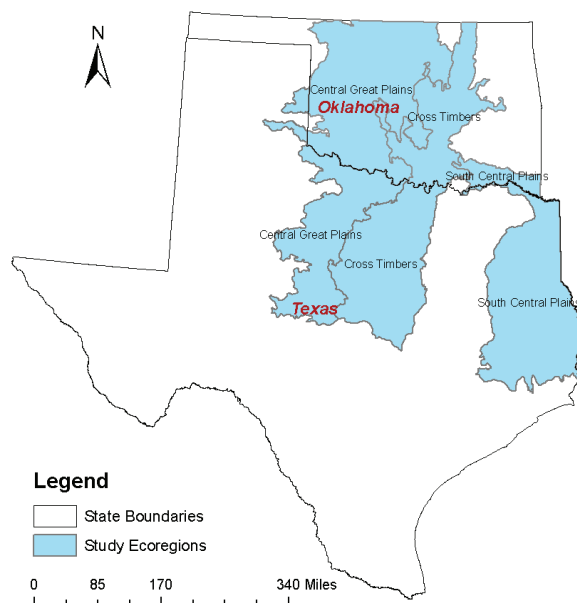
## Methods

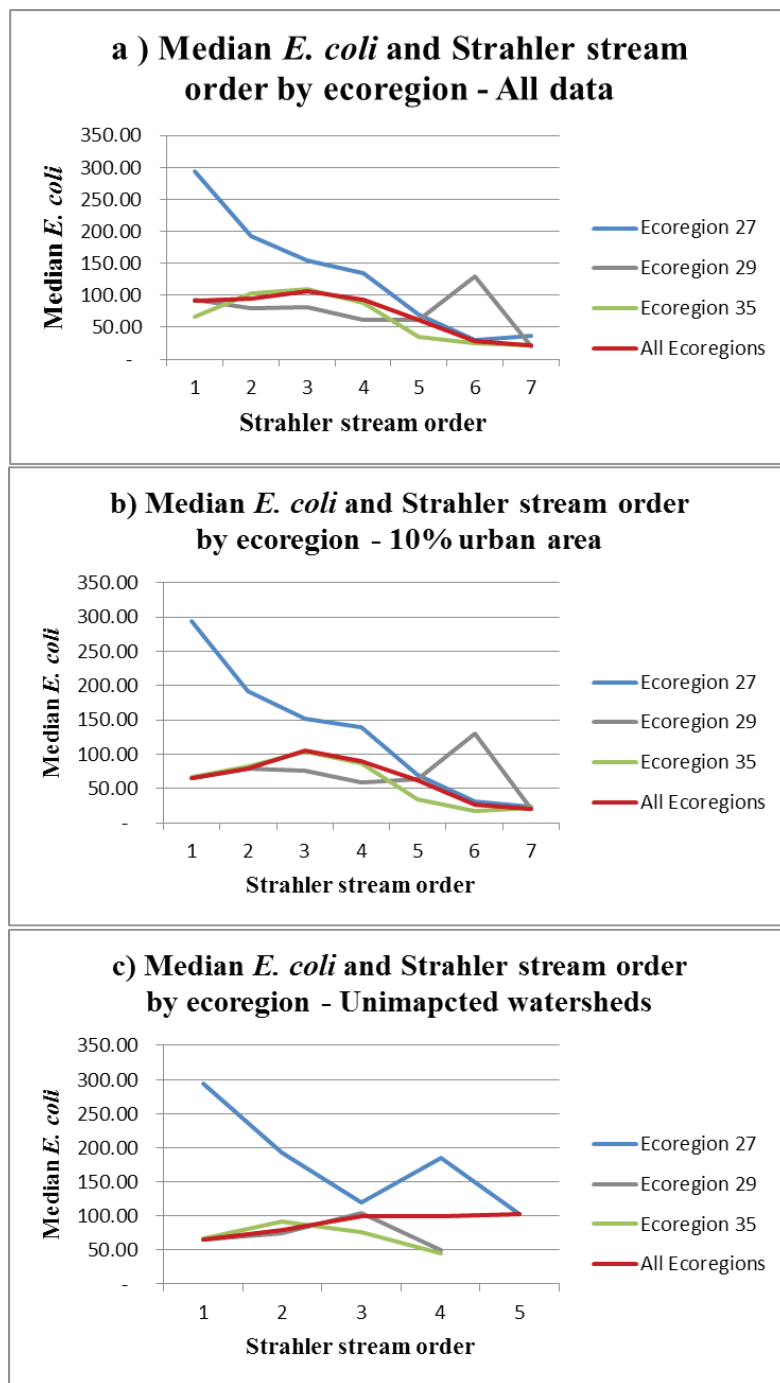
*E. coli* data used in this study were obtained from existing monitoring stations located on freshwater streams in three ecoregions in Texas and Oklahoma – the Central Great Plains, Central Oklahoma/Texas Plains and South Central Plains (Ecoregions 27, 29, and 35, respectively). This allowed determination of not only overall relationships but also ecoregion effects. Median *E. coli* concentrations from the monitoring stations were analyzed for correlation with the stream order of each monitoring site as well as the watershed size above each site. Geospatial analysis was used to determine stream order and watershed size and identify un-impacted/least impacted streams in each ecoregion. Stream order was classified based on Horton (1945) and Strahler (1957) stream order classification and stream link magnitude analysis method (Shreve 1966).

## Findings

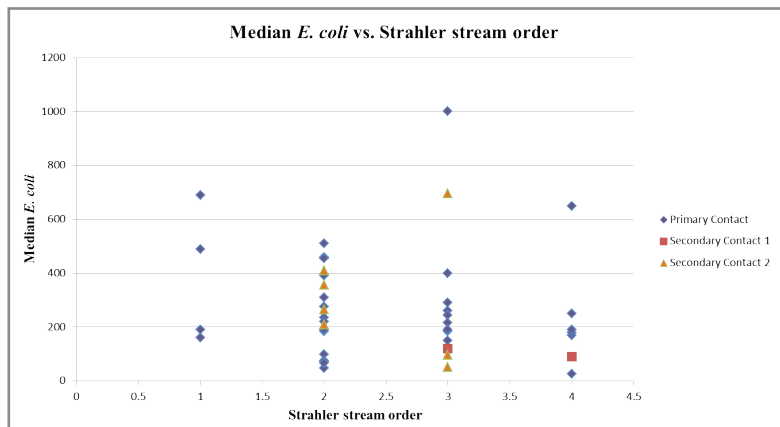
No significant relationships were observed between *E. coli* and stream orders/watershed size of unimpacted watersheds. However, watersheds with wastewater outfalls and urban areas exceeding 10% of land use showed a statistically significant, yet weak negative relationship between *E. coli* and stream order/watershed size, i.e. *E. coli* decreased with increase in stream order/watershed size. Similar to the findings of VDEQ (2003) and Edge et al. (2012), *E. coli* concentrations generally increased for the first three stream orders and then dropped for later stream orders. It is noteworthy that the median *E. coli* for each of the three datasets stays below the geometric mean standard of 126 cfu/100ml required for primary contact recreation in Texas.

Comparison of the three ecoregions revealed that median and geometric mean *E. coli* concentrations and their correlation with both stream order methods and watershed area decreased moving eastward. Ecoregion 27, which is the western most ecoregion in this





**Figure 1.** Comparison of median *E. coli* concentrations by Strahler stream order and ecoregions.



**Figure 2.** RUAA status of *E. coli* by Strahler stream order.

study and has a more arid to semi-arid environment compared to the other ecoregions, showed the highest *E. coli* concentration as well as correlations. Ecoregion 35, the eastern most region and with the wettest climate of the three regions, had the lowest *E. coli* concentrations and correlations.

Finally, the relationship between stream order and recreational use was evaluated; however, a lack of completed RUAA studies did not allow for correlation tests between *E. coli* recreation standards based on the stream orders. It is noteworthy that primary contact recreation has been associated with all stream orders.

## Implications

The hypothesis that stream order could be used to guide recreational use and associated standards designation was not supported by research findings. Primary contact recreation has been associated with the range of stream orders. This, in addition to the finding that there was no correlation between *E. coli* concentrations and stream order at unimpacted sites, lead to the conclusion that stream order alone cannot be used to guide recreational use designation (i.e. Primary Contact Recreation, Secondary Contact Recreation 1, etc.) as hypothesized. Thus, it is recommended that site-specific studies continue to be conducted. In the future, there may be cause for reassessing this information when there are more RUAA studies completed.



## Appendix E

## Leon River Watershed Protection Plan: Revisions to Plan for Resubmission to EPA



December XX, 2014

### Discussion Topics

- Refresher of EPA Concerns
- Key elements revised in WPP
- Next Steps

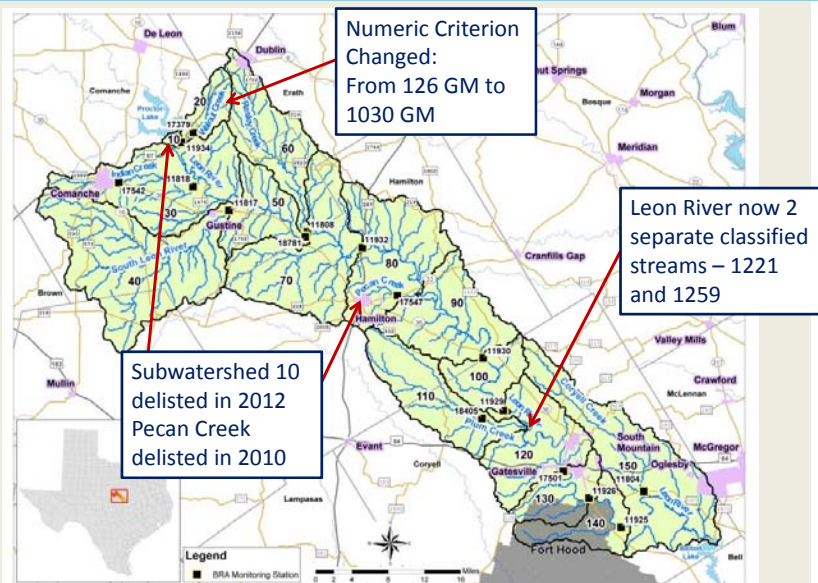
## General Comments

- “Challenges to acceptance of the document under the NPS Program Guidelines focus on
  - a clear commitment to state water quality standards as the restoration goal,
  - BMPs and water quality goals that address nutrients, chlorophyll-a, and dissolved oxygen,
  - size of the watershed, and
  - lack of sufficient documentation of modeling efforts to answer technical questions about sources, load reductions, and the level of implementation needed for desired water quality outcomes.
- These factors make it unclear to what degree water quality restoration would be accomplished through implementation of the WPP, and should be addressed before we can accept this plan.”

## Clarify WQ Goals

Waterbodies in Leon River Watershed	Pre-2014 Use Category	Pre-2014 Geometric Mean <i>E. coli</i> Criterion for Freshwater Streams (cfu/100mL)
Leon River (1221) and all tributaries	Contact Recreation	126
	2014 Approved Use Categories	2014 Approved Geometric Mean <i>E. coli</i> Criteria for Freshwater (cfu/100mL)
<ul style="list-style-type: none"> <li>• Leon River below Proctor Lake (1221) including Pecan Creek</li> <li>• Leon River above Lake Belton (1259) and all tributaries that flow to 1259</li> </ul>	Primary Contact Recreation 1	126
<ul style="list-style-type: none"> <li>• Walnut Creek</li> </ul>	Secondary Contact Recreation 2	1030

## Adoption of WQS Revisions and De-Listings



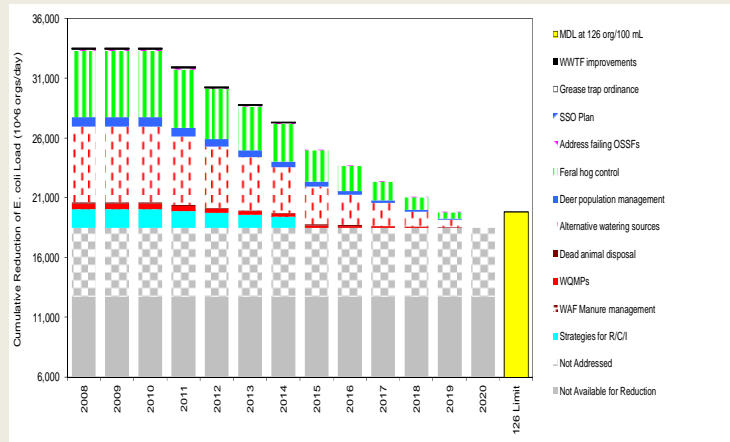
## Implications – Additional Load Reductions Needed

SW	Proposed Evaluation Location	E. coli Geometric Mean (Modeled Data)		
	SWQM Station	2008	2015	2020
10	11934	85	73	64
20	17379	301	266	248
30	11818	130	108	102
40	11817	229	203	188
50	18781	111	95	88
60	11808	253	210	191
70	11932	106	91	85
80	17547	93	79	73
90	11930	61	52	48
100	11929	62	53	49
110	18405	97	90	83
120	17501	93	79	74
130	11926	109	92	87
140	11925	99	83	79
150	11804	75	64	60

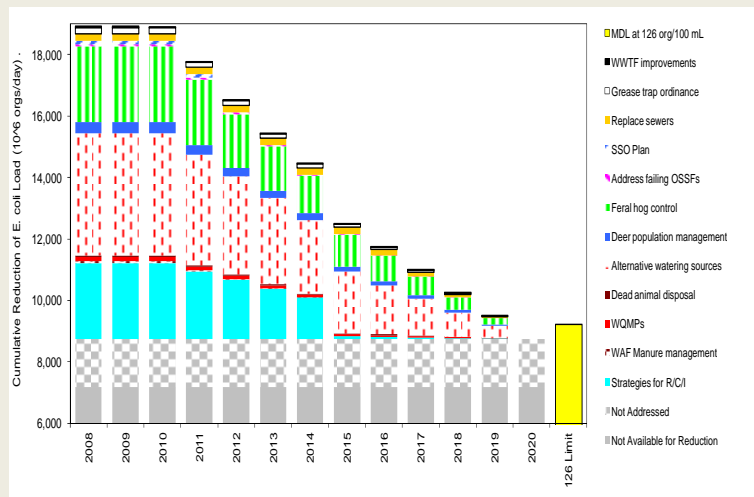
  

SW	Proposed Evaluation Location	E. coli Geometric Mean (Modeled Data)		
	SWQM Station	2008	2015	2020
10	11934	85	73	64
20 <sup>1</sup>	17379	301	262	241
30	11818	130	107	100
40	11817	229	169	126
50	18781	111	93	83
60	11808	253	169	126
70	11932	106	88	79
80	17547	93	78	70
90	11930	61	52	47
100	11929	62	53	48
110	18405	97	89	82
120	17501	93	79	73
130	11926	109	92	85
140	11925	99	83	78
150	11804	75	63	60

## Cumulative Reduction of Loads in Subwatershed 40 (South Leon River)



## Cumulative Reduction of Loads in Subwatershed 60 (Resley Creek)





## Additional Costs

Management Strategy	Original Costs	New Costs
WWTF improvements	\$ 11,729,600	\$ 11,729,600
Grease trap ordinance	\$ 10,000	\$ 10,000
Replace sewers	\$ 1,539,000	\$ 1,539,000
SSO Plan	\$ 375,000	\$ 375,000
Address failing OSSFs	\$ 2,247,500	\$ 2,247,500
Feral hog control	\$ 796,000	\$ 1,057,227
Deer population management	\$ 100,000	\$ 100,000
Alternative watering sources	\$ 2,080,000	\$ 2,740,000
Dead animal disposal	\$ 484,000	\$ 515,389
WQMPs	\$ 9,630,000	\$ 13,027,500
WAF Manure management	\$ 1,665,000	\$ 1,665,000
Strategies for R/C/I	\$ 1,765,000	\$ 2,545,000
<b>Subtotal</b>	<b>\$ 32,276,100</b>	<b>\$ 37,551,216</b>
<b>Total increase</b>		<b>\$ 5,275,116</b>

## Nutrients, Chlorophyll-*a*, DO

- Fact: The watershed is not meeting certain water quality criteria beside bacteria: nutrients and chlorophyll A
- Issue: The WPP states only that it is expected that what will be done for bacteria will also reduce these parameters with no specific projects, quantification of reduction, or other discussion to show a focus on nutrients
- Response: To further characterize nutrients and the associated effects on dissolved oxygen, a greater level of monitoring is necessary to determine baseline conditions and establish an acceptable weight-of-evidence approach between estimated nutrient reductions and long-term water quality conditions in Resley Creek and Indian Creek.

## Next Steps

- Parsons incorporates any modifications recommended by Watershed Steering Committee (January)
- Parsons provides track changes and clean electronic file to TSSWCB for submission to EPA (January)
- TSSWCB finalizes the response to comment letter (January)
- TSSWCB submits Leon River WPP and response to comment letter to EPA (February)