Advancing Quantitative Understanding of Escherichia Coli Concentrations in a Contemporary Mixed Land-Use Watershed, in West Virginia, USA

Fritz Petersen
West Virginia University, fp0008@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Part of the Water Resource Management Commons

Recommended Citation
https://researchrepository.wvu.edu/etd/7781

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
Advancing Quantitative Understanding of Escherichia Coli Concentrations in a Contemporary Mixed Land-Use Watershed, in West Virginia, USA

Fritz Petersen

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Part of the Water Resource Management Commons
Advancing Quantitative Understanding of *Escherichia Coli* Concentrations in a Contemporary Mixed Land-Use Watershed, in West Virginia, USA

Fritz Petersen

Dissertation submitted to the
Faculty of the Davis College of Agriculture, Natural Resources and Design
at West Virginia University

In partial fulfilment for the degree of

Doctor of Philosophy in:

Plant and Soil Sciences

Jason A. Hubbart, PhD, Chair
Zachary B. Freedman, Ph.D.
Michael Strager, Ph.D.
Eugenia Pena-Yewtukhiw, Ph.D.

Division of Plant and Soil Sciences

Morgantown, West Virginia
2020

Keywords: *E. coli*; Surface Water; Mixed-Land-Use; Water Quality; Suspended Particulate Matter

Copyright 2020 Fritz Petersen
A scarcity of field-based research investigating the relationships between \textit{E. coli} (fecal matter) concentration, suspended particulate matter (SPM; including size distribution), physicochemical parameters (water temperature, pH, SPC, DO, chloride) and land use practices and the importance of these relationships, particularly for human health, policy makers, and water resource managers, provided the impetus for this research. A 22-site, nested-scale, experimental watershed study design was implemented to investigate \textit{E. coli} concentrations and the aforementioned relationships in a contemporary mixed land use watershed in the Appalachian region of the eastern United States. Results from a reduced (n=4 sites) study showed SPM in the <5 μm interval to be the most important for \textit{E. coli} with more than 90% of \textit{E. coli} data found in this smallest interval. An annual study including all 22 sites revealed elevated \textit{E. coli} concentrations in agricultural sub-catchments (avg. 560 CFU per 100 mL) relative to mixed development (avg. 330 CFU per 100 mL) and forested (avg. 206 CFU per 100 mL) sub-catchments. Additionally, annual \textit{E. coli} and SPM concentration data displayed a statically significant relationship (p < 0.01) in agricultural areas. Results reflect the impact of historic land use practices (e.g. AMD caused by mining) as Spearman’s correlation coefficient (SCC) results showed significant correlation (p < 0.05) between water pH and \textit{E. coli} concentrations at 77% of sample sites. Furthermore, a pH tipping point (threshold) in the range of 7.68–7.76 was identified in the current investigation, with pH values below this range including significant negative correlations (p < 0.05) with \textit{E. coli} concentrations. A land cover tipping point of 25–30% was also identified for mixed development land use practices and significant (p < 0.05) negative correlations between \textit{E. coli} and chloride concentrations. This research provides validation of the use of the experimental watershed study design to advance understanding of fecal contamination regimes in contemporary mixed land use watersheds. Results from the work can inform policy makers and land use managers about factors impacting fecal microbes in mixed land use watersheds, thereby aiding in effective water quality management.
The undersigned, approve the dissertation entitled

Advancing Quantitative Understanding of *Escherichia Coli* Concentrations in a Contemporary Mixed Land-Use Watershed, in West Virginia, USA

Presented by Fritz Petersen,

as partial fulfillment to the requirement of the degree of PhD Plants and Soil Sciences

and hereby certify that, in their opinion, it is worthy of acceptance.

________________________________________
Dr. Jason A. Hubbard (Advisor)

________________________________________
Dr. Eugenia Pena-Yewtukhiw

________________________________________
Dr. Zachary B. Freedman

________________________________________
Dr. Michael Strager
ACKNOWLEDGEMENTS

I would like to thank Dr. Jason Hubbart for his support and guidance. My completion of this program would not have been possible without his patience, support, and instruction.

I would like to thank my mother and fiancé for their unwavering support and continuous motivation during my studies. I also wish to dedicate this work to my mother Marie-Lou Petersen as without her many sacrifices I would never have made it this far.

I would also like to thank committee members Dr. Zachary B. Freedman, Dr. Eugenia Pena-Yewtukhiw and Dr. Michael Strager for sharing their expertise and guidance during this program, as well as taking the time to assist my growth as a scientist.

Special thanks for funding provided by the National Science Foundation under Award Number OIA-1458952, the USDA National Institute of Food and Agriculture, Hatch project accession number 1011536, and the West Virginia Agricultural and Forestry Experiment Station. Results presented may not reflect the views of the sponsors and no official endorsement should be inferred. Special thanks are due to many scientists of the Interdisciplinary Hydrology Laboratory (https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart).
# TABLE OF CONTENTS

Table of Contents

ABSTRACT .......................................................................................................................... ii
ACKNOWLEDGEMENTS .................................................................................................... iv
TABLE OF CONTENTS ....................................................................................................... v
LIST OF FIGURES ............................................................................................................. viii
LIST OF TABLES ................................................................................................................ xiii

CHAPTER 1: INTRODUCTION .............................................................................................. 1

1. Background ..................................................................................................................... 1
   1.1 Fecal Contamination and Monitoring ...................................................................... 1
   1.2 General description of *Escherichia coli* ................................................................. 2
   1.3 Virulence of *E. coli* .............................................................................................. 2
   1.4 Lifecycle of *E. coli* .............................................................................................. 3
   1.5 *E. coli* and Land Use practices ............................................................................ 5
   1.6 *E. coli* and Hydroclimate ................................................................................... 6
   1.7 *E. coli* and Suspended Particulate Matter ............................................................ 7

2. Statement of need ............................................................................................................ 8

3. Objectives ......................................................................................................................... 9

4. Methods .......................................................................................................................... 10
   4.1 Study Watershed ...................................................................................................... 10
   4.2 Study Sites and Design ......................................................................................... 11
   4.3 GIS methodology .................................................................................................... 14

5. Methods: Chapter 2 .......................................................................................................... 15
   5.1 Water Sampling ...................................................................................................... 16
   5.2 Quantifying *E. coli* ............................................................................................. 17
   5.3 Gravimetric Analysis ............................................................................................. 18
   5.4 Sample Filtration ................................................................................................... 18

6. Methods: Chapter 3 .......................................................................................................... 19
   6.1 Measuring Stream Stage ....................................................................................... 19

7. Methods: Chapter 4 .......................................................................................................... 19
   7.1 Measuring Water Quality Variables ..................................................................... 19

8. Dissertation Organization ................................................................................................. 9

References ............................................................................................................................ 23
## CHAPTER 4: ADVANCING UNDERSTANDING OF LAND USE AND PHYSICOCHEMICAL IMPACTS ON FECAL CONTAMINATION IN MIXED-LAND-USE WATERSHEDS

1. Introduction .......................................................................................................................... 107
2. Materials and Methods ........................................................................................................ 110
   2.1. Study Site Description ................................................................................................. 110
   2.2. Data Collection ........................................................................................................... 113
   2.3. Data Analysis ............................................................................................................... 115
3. Results and Discussion ......................................................................................................... 116
   3.1. Climate During Study .................................................................................................. 116
   3.2. *E. Coli* Concentrations ............................................................................................ 117
   3.3. Physicochemical Parameters ...................................................................................... 119
   3.4. Annual Non-Parametric Statistical Results ............................................................... 126
   3.5. Quarterly Non-Parametric Statistical Results ............................................................ 133
   3.6. Study Implications and Future Directions .................................................................. 136
4. Conclusions .......................................................................................................................... 137

Appendix A .................................................................................................................................. 140

References .................................................................................................................................... 149

## CHAPTER 5: REVIEW OF FACTORS IMPACTING THE SURVIVAL AND OCCURRENCE OF ESCHERICHIA COLI IN SECONDARY MICROPHYSICAL HABITATS AND DISSERTATION CONCLUSION

1. Introduction .......................................................................................................................... 158
2. *E. coli* and the Environment ............................................................................................... 161
   2.1. Temperature .................................................................................................................. 161
   2.2. Solar Insolation ............................................................................................................. 163
   2.3. Suspended and Settled Solids ...................................................................................... 164
   2.4. Hydrologic conditions ................................................................................................. 165
   2.5. Water chemistry ........................................................................................................... 167
   2.6. Nutrients and nutrient availability .............................................................................. 168
   2.7. Land-use practices ....................................................................................................... 169
3. Mitigation Strategies ............................................................................................................. 171
4. Future Directions ................................................................................................................. 172
5. Dissertation Conclusion ...................................................................................................... 173

References .................................................................................................................................... 176
LIST OF FIGURES

Figure 1. West Run Watershed experimental watershed study design comprising 22 monitoring sites, in West Run Watershed, West Virginia, USA..................................................14

Figure 2. Filtration work study design, comprising 4 monitoring sites, in West Run Watershed, West Virginia, USA.................................................................16

Figure 3. Monitoring/sampling locations for the current investigation, with land use/land cover, in West Run Watershed, Morgantown, West Virginia, USA.........................41

Figure 4. Thirty-minute time series of climate variables (recorded at the climate station) during study period (20 July 2018–27 October 2018) in West Run Watershed, West Virginia, USA. *Note: Stream stage was monitored in the primary stream of WRW, West Run Creek, within approximately 100m of the confluence of Site #1 and West Run Creek (39°40’3.20” N 79°55’48.99” W).................................................................46

Figure 5. Box and whisker plot of suspended particulate matter (SPM) (mg L−1) and E. coli concentration (CFU per 100 mL) at each sampling location (n = 4) during the study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.................................................................48

Figure 6. Top: cumulative total SPM (mg L−1). Bottom: cumulative total E. coli concentration (CFU per 100 mL) at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA. *Only predominant land use percentage indicated, for full land use information refer to Table 3. .................................................................................................................................52

Figure 7. Fractions of E. coli concentration (CFU) in three filtered size intervals, at four monitoring sites during the study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA..........................................................53

Figure 8. Suspended particulate matter (SPM) and E. coli concentration percent difference, separated into filtration intervals (<5 μm; 5 μm < interval < 60 μm; and >60); and total, at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA............................................................55

Figure 9. Ln of E. coli concentration (CFU per 100 mL) relative to SPM in three filtration size intervals (<5 μm; 5 μm < interval < 60 μm; >60 μm), at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, West Virginia,
USA. Note all *E. coli* concentration values are +1 to avoid ln values being undefined where *E. coli* concentrations were zero. Additionally, interval size was dictated by the filter sizes (5 μm and 60 μm) that we used during the filtration of the extracted water samples.

Figure 10. Results of principal components analysis, including biplots A) showing the data in the smallest interval (*E. coli* and SPM) and B) the total data set (defined as all he SPM and *E. coli* data >0.7 μm), for extracted principal components of *E. coli* concentration and SPM concentration at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, West Virginia, USA.

Figure 11. Monitoring/sampling locations for the current investigation, with land use/land cover, in West Run Watershed, Morgantown, West Virginia, USA.

Figure 12. Thirty-minute time series of climate variables during the study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: stream stage was measured in the primary stream of WRW, West Run Creek, within approximately 30 m of Site #13 and West Run Creek.

Figure 13. Box and whisker plots of suspended particulate matter (mg L$^{-1}$; log$^{10}$ scale) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA. Boxes delineate 25th and 75th percentiles; lines denote medians; squares show means; whiskers describe 10th and 90th percentiles; x shows maxima and minima when above and below, respectively. Note: different box colors represent data from different sites.

Figure 14. Box and whisker plot of *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA. Boxes delineate 25th and 75th percentiles; lines denote medians; squares show means; whiskers describe 10th and 90th percentiles; x shows maxima and minima when above and below, respectively. Note: different box colors represent data from different sites.

Figure 15. Average suspended particulate matter (mg L$^{-1}$) and *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA.

Figure 16. Land use percentage relative to cumulative annual *E. coli* concentration (CFU per 100 mL) and SPM concentration (mg/L) at West Run Creek monitoring sites (n = 9) during the study period (2 January 2018–1 January 2019) in West Run Watershed,
Morgantown, West Virginia, USA. Note: West Run Creek included the following site numbers: #3, #4, #6, #10, #13, #18, #19, #21 and #22.

Figure 17. Quarterly average *E. coli* concentration (CFU per 100 mL) and SPM concentration (mg/L) at each sampling location (n = 22) during the study period (1/2/18–1/1/19) in West Run Watershed, Morgantown, West Virginia, USA. Note: (A) represents Quarter 1 (winter: 2 January 2018–27 March 2018); (B) represents Quarter 2 (spring: 3 April 2018–26 June 2018); (C) represents Quarter 3 (summer: 3 July 2018–25 September 2018); (D) represents Quarter 4 (fall: 2 October 2018–1 January 2019).

Figure 18. Results of principal component analysis, including biplots, for extracted principal components of annual *E. coli* concentration (CFU per 100 mL) and annual SPM concentration (mg/L) at 22 monitoring sites (indicated by the different colors) during the study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA.

Figure 19. Results of principal component analysis, including biplots, for extracted principal components of quarterly *E. coli* concentration (CFU per 100 mL) and quarterly SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: (A) represents Quarter 1 (winter: 2 January 2018–27 March 2018); (B) represents Quarter 2 (spring: 3 April 2018–26 June 2018); (C) represents Quarter 3 (summer: 3 July 2018–25 September 2018); (D) represents Quarter 4 (fall: 2 October 2018–1 January 2019).

Figure 20. Land use/land cover of West Run Watershed, Morgantown, WV, USA, including monitoring/sampling locations for the current investigation.

Figure 21. Thirty-minute time series of climate variables during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: Stream stage was monitored in the primary stream of WRW, West Run Creek, at site #13 (Figure 20).

Figure 22. Box and whisker plot of *E. coli* concentration (CFU per 100 mL) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Figure 23. Box and whisker plot of water temperature (°C) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line
Figure 24. Box and whisker plot of pH descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Figure 25. Box and whisker plot of specific conductance (μS/cm) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Figure 26. Box and whisker plot of dissolved oxygen (%) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Figure 27. Box and whisker plot of chloride ion (mg/L) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Figure 28. Results of principal components analysis, including biplots, for extracted principal components of annual E. coli concentration (CFU per 100 mL), water temperature (°C), pH, specific conductance (μS/cm), dissolved oxygen (DO; %), and chloride ion (mg/L) at 22 monitoring sites (indicated by the different colors) during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA.

Figure 29. Results of principal components analysis, including biplots, for extracted principal components of quarterly E. coli concentration (CFU per 100mL), water temperature (°C), pH, specific conductance (μS/cm), dissolved oxygen (%) and chloride ion (mg/L) at 22 monitoring sites (indicated by the different colors) during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: (A) represents quarter one (winter: 2 January 2018–27 March 2018); (B) represents quarter two (spring: 3 April 2018–26 June 2018); (C) represents quarter three (summer: 3 July
2018–25 September 2018); (D) represents quarter four (fall: 2 October 2018–1 January 2019).
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at twenty-two monitoring sites comprising the experimental watershed study in West Run Watershed (WRW), West Virginia, USA. Note: Data extracted from NLCD 2011 dataset with ERSI Arc GIS 10.6 Software.</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at four monitoring sites comprising the filtration study in West Run Watershed (WRW), West Virginia, USA. Note: Data extracted from NLCD 2011 dataset.</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at four monitoring sites in West Run Watershed (WRW), West Virginia, USA. Note: land use percentages may not sum to 100%, as not every category is included (i.e., wetland, open water, etc.) and some categories are combinations of others (e.g., developed = urban + residential), or independent (e.g., impervious). Final row indicates total values for the entire watershed.</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>Results of principal component analysis comprising 5 components (E. coli concentration, SPM concentration, percentage of agricultural land use, percentage of forested land use and percentage of developed land use) displaying eigenvalues, percentage of variance and cumulative variance during the study period (20 July 2018–27 October 2018) for the smallest interval (&lt;5 µm) and the total data set across the four monitoring sites in West Run Watershed, West Virginia, USA. Note: bold numbers indicate eigenvalues exceeding 1 (representing importance).</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at 22 monitoring sites in West Run Watershed (WRW), West Virginia, USA. Note: land use percentages may not sum to 100%, as not every category is included (i.e., wetland, open water, etc.) and some categories are combinations of others (e.g., mixed development = urban + residential). Final row (Site #22) indicates the total values for the entire watershed.</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Descriptive statistics of suspended particulate matter (mg L$^{-1}$) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: all average values presented in the current work constitute arithmetic means.</td>
<td>81</td>
</tr>
</tbody>
</table>
Table 7. Descriptive statistics of E. coli concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA

Table 8. Results of Spearman’s correlation test, including annual E. coli concentration (CFU per 100 mL) and annual SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: bold values indicate significant correlations (p < 0.05)

Table 9. Results of Spearman’s correlation test, including quarterly E. coli concentration (CFU per 100 mL) and quarterly SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: Quarter 1 represents 2 January 2018–27 March 2018; Quarter 2 represents 3 April 2018–26 June 2018; Quarter 3 represents 3 July 2018–25 September 2018; Quarter 4 represents 2 October 2018–1 January 2019. Bold values indicate significant correlations (p < 0.05)

Table 10. Land use/land cover characteristics (% cover) of 22 monitoring sites in West Run Watershed (WRW), West Virginia, USA, including total drainage area (km²)

Table 11. Results of Spearman’s Correlation test, including annual E. coli concentration (colony forming units (CFU) per 100mL) water temperature (°C), pH, specific conductance (SPC; µS/cm), dissolved oxygen (DO; %) and chloride ion (CL⁻; mg/L) at each sampling location (n = 22) during study period (2 January 2018—1 January 2019) in West Run Watershed, WV, USA

Table 12. Results of principal component analysis comprising 9 variables (E. coli concentration, water temperature, pH, SPC, DO, chloride, percentage of agricultural land use, percentage of forested land use, and percentage of developed land use) used to define 9 principal components, displaying eigenvalues, percentage of variance, and cumulative variance during the study period (2 January 2018–1 January 2019) across the 22 monitoring sites in West Run Watershed, West Virginia, USA

Table 13. Results of principal component analysis comprising 9 variables (E. coli concentration, water temperature, pH, SPC, DO, chloride, percentage of agricultural land use, percentage of forested land use and percentage of developed land use) used to define 9 principal components, displaying eigenvalues, percentage of variance and cumulative variance during quarter one (winter: 2 January 2018–27 March 2018); quarter two (spring: 3 April 2018–26 June 2018); quarter three (summer: 3 July, 2018–25 September, 2018); and quarter four (fall: 2 October 2018–1 January 2019).
2018); and quarter four (fall: 2 October 2018–1 January 2019) across the 22 monitoring sites in West Run Watershed, West Virginia, USA.

Table 14. Descriptive statistics of *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 15. Descriptive statistics of water temperature (°C) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 16. Descriptive statistics of pH at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 17. Descriptive statistics of water specific conductance (µS/cm) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 18. Descriptive statistics of dissolved oxygen (%) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 19. Descriptive statistics of chloride ion (mg/L) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

Table 20. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA.

Table 21. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA.
use) used to define 9 principal components, during quarter one (winter: 2 January 2018–27 March 2018) in West Run Watershed, WV, USA. ................................................................. 147

Table 22. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during quarter two (spring: 3 April 2018–26 June 2018) in West Run Watershed, WV, USA. ................................................................. 147

Table 23. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components during quarter three (summer: 3 July 2018–25 September 2018) in West Run Watershed, WV, USA. ................................................................. 148

Table 24. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components during quarter four (fall: 2 October 2018–1 January 2019) in West Run Watershed, WV, USA. ................................................................. 148
CHAPTER 1: INTRODUCTION

1. Background

1.1 Fecal Contamination and Monitoring

Fecal matter can contain potentially harmful microbes such as *Escherichia coli* (*E. coli*) and Enterococci (Price and Wildeboer 2017). Consequently, if fecal matter enters fresh water supplies it can be a key source of waterborne pathogens (causative agent of disease (Meriam-Webster 2018)) (Cabral 2010; Pandey et al 2014). Globally, waterborne pathogens cause substantial mortality and morbidity among human populations (Mahmoud et al. 2016, Price and Wildeboer 2017). For example, 2 million deaths (mainly in children under 5 years of age) are attributed to waterborne diarrheal diseases annually (WHO 2018). Waterborne pathogens can also affect social and economic sectors, by increasing medical expenditures, hampering productivity due to sick leave, and decreasing tourism (Ishii and Sadowsky 2008). The health, social and economic impacts of waterborne pathogens, necessitate the monitoring of waterbodies for fecal (pathogenic) contamination (Ishii and Sadowsky 2008).

Direct monitoring pathogenic fecal bacteria is complicated due to challenges in terms of sampling and detection (Cabral 2010). Sampling can be difficult as concentrations of pathogenic bacteria in waterbodies can be sporadic and remain low under certain conditions (Cabral 2010). Furthermore, detection can be difficult due to the complex isolation and culturing procedures required to quantify pathogenic fecal bacteria (Cabral 2010). In general, working with pathogenic microbes also includes risk of infection and inability to maintain the microbe in a laboratory environment (Cabral 2010). Due to these complexities, pathogenic fecal microbes are typically not directly monitored to analyze water for pathogenic fecal contamination. Fecal indicator organisms (e.g., *E. coli*) are typically used for this purpose, as water contaminated with fecal matter will contain non-disease-causing intestinal microbes in addition to pathogenic microbes (Stewart et al. 2007). Therefore, fecal indicator bacteria are useful proxies for monitoring fecal contamination and potential pathogens (Stewart et al. 2007). From this group, *E. coli* has been shown to have the highest correlation to the outbreak of diseases in both marine and freshwater systems and is therefore the most commonly used indicator organism (Price and...
The widespread use of *E. coli* as a pathogenic indicator organism is also facilitated by the generally quicker and less-expensive methodologies associated with detection, relative to other indicator organisms (Allen 2010).

1.2 General description of *Escherichia coli*

*E. coli* are a rod shaped, facultative anaerobic, gram-negative, coliform, gammaproteobacteria (Ishii and Sadowsky 2008). *E. coli* are typically found in the intestines of endotherms, animals capable of internal heat generation independent from the environment (Meriam-Webster 2018), including humans, before they enter the environment as part of expelled fecal matter (Price and Wildeboer 2017). Most *E. coli* strains have a commensal relationship with their host organisms. However, some are known to cause disease in humans and other endotherms (Ishii and Sadowsky 2008). Even non-pathogenic strains of *E. coli* can lead to infection, should the immune system be hampered in its function or when the gastrointestinal barriers are disrupted (Nataro and Kaper 1998). *E. coli* has been linked to cases of gastrointestinal, urinary, and central nervous system diseases in hosts of varying immune and physical health (Nataro and Kaper 1998). *E. coli*’s ability to cause disease in various endotherms (cross-species transmission) means outbreaks can affect large proportions of populations, eventually leading to epidemics (Nataro and Kaper 1998).

1.3 Virulence of *E. coli*

Virulence is defined as the severity or harmfulness of a disease or poison (Meriam-Webster 2018). The virulence of *E. coli* is severe enough that even minimal contact with contaminated waterbodies can result in outbreaks of gastroenteritis (Madoux-Humery et al. 2016). However, the amount of contaminated water that is ingested is important when determining the risk posed by *E. coli* (Dorevitch et al. 2011). Six pathotypes of *E. coli* (termed diarrheagenic *E. coli*) are associated with diarrhea. (CDC 2018). Diarrheagenic *E. coli* includes the following strains: Shiga toxin-producing *E. coli* (STEC), also known as enterohemoraghic *E. coli* (EHEC), is responsible for fatal cases of bloody diarrhea and diseases such as hemolytic uremic syndrome (HUS) or hemorrhagic colitis (HC); enterotoxigenic *E. coli* (ETEC) is responsible for traveler’s diarrhea; enteropathogenic *E. coli* (EPEC) is responsible for watery diarrhea in infants; enteroaggregative *E. coli* (EAEC) is responsible for persistent diarrhea; enteroinvasive *E. coli* (EIEC) is responsible for profuse diarrhea and fever, and diffusely
adherent *E. coli* (DAEC) is responsible for acute diarrhea in young children (CDC 2018; Scallstsky et al. 2002; Nataro and Kaper 1998). The six-diarrheagenic strains of *E. coli* are traditionally the causative agents behind most pediatric diarrhea cases globally (Nataro and Kaper 1998). The EHEC serotype *E. coli* O157:H7 has been identified as particularly harmful, given numerous cases where it caused large-scale outbreaks of both food- and water-borne diseases (Ishii and Sadowsky 2008). Apart from the six-diarrheagenic *E. coli*, extraintestinal pathogenic *E. coli* (ExPEC) (such as uropathogenic and avian pathogenic strains) can also be considered among pathogenic *E. coli* strains (Welch et al. 2002). While ExPEC is harmless in the intestinal tract of its original host, it can lead to urinary tract infections, cause neonatal meningitis or sepsis if it infects different organisms than its original host (Welch et al. 2002).

*E. coli* infection of organisms, such as endotherms, is termed colibacillosis. Colibacillosis can lead to septicemia and diarrhea in calves, acute mastitis in dairy cows, perihepatitis, pericarditis, septicemic lungs and peritonitis in poultry, and Alabama rot in canines (Gordon 2013). Fish can also be negatively impacted by *E. coli* (Ibrahim et al. 2014). Thus, efficient control of pathogens is vital for successful aquaculture (Mahmoud et al. 2016). The proportion of the public sector that consumes market fish that originated from aquaculture farms are also at risk to contract pathogens when ingesting the fish (Ibrahim et al. 2014). Different endotherms pose varying degrees of risk as transmitters of *E. coli* to humans, determined by the number of pathogenic strains that they can potentially carry or host (Ishii et al. 2007). An endotherm that can carry a great (diverse) number of pathogenic *E. coli* strains will pose a greater hazard than an endotherm that can carry a smaller, less diverse number of pathogenic *E. coli* strains.

1.4 Lifecycle of *E. coli*

*E. coli* utilizes two reproduction processes, namely asexual reproduction and conjugation. Asexual reproduction, which takes place during binary fission, is the more prevalent reproduction process implemented by *E. coli* (Wang and Levin 2009). Asexual reproduction is initiated by DNA replication (Wang and Levin 2009). Following replication, the genetic material attaches to the membranes of the cell. Simultaneously the cell grows until it reaches approximately double its original size and all components of the cell multiply until each daughter cell receives a complete chromosome, sufficient macromolecules, monomers, and inorganic ions to survive as independent cells (Wang and Levin 2009). Once the required size has been reached,
the cell membrane starts pinching inward forming a partition known as a septum (Burdett and Murray 1974; Wang and Levin 2009). The septum is extended through the continued inward growth of the cytoplasmic membrane and the cell wall from opposing directions until the two daughter cells are pinched off (Burdett and Murray 1974). The full development of the cell walls marks the complete split of the bacteria into two daughter cells and the conclusion of the growth cycle (Wang and Levin 2009). Under ideal conditions, the growth cycle of *E. coli* takes approximately 20 minutes (Fossum et al. 2007) but is dependent on both nutritional and genetic factors (Wang and Levin 2009).

Conjugation, the second reproduction process that *E. coli* utilizes, involves the transfer of genetic material by the sex pili between two bacteria (Volk and Wheeler 1964, Pérez-Mendoza and de la Cruz 2009). Conjugation is not classified as sexual reproduction as it does not involve the combination of gametes and is classified as a form of horizontal gene transfer (Berg and Trevors 1990). Conjugation is initiated by the secretion of enzymes from an *E. coli* cell that contains an extrachromosomal F plasmid (a mobile intracellular genetic structure capable of self-replication, typically a small circular DNA strand in the cytoplasm of a bacterium) signalling it is ready to “mate” (transfer its genetic material) (Pérez-Mendoza and de la Cruz 2009). Plasmid-containing strains of *E. coli* are termed F+ and function as donors, while strains lacking the F plasmid are F− and behave as recipients (Baron 1996). Once the F+ cell finds a suitable recipient, the donor attaches itself to the sex pilus of the receiving cell forming an intercellular cytoplasmic bridge (Baron 1996). Following attachment, a nick (discontinuity in the double stranded DNA molecule) is created in one of the strands of the plasmid. The nicked strand is then unwound from the unbroken strand and transferred to the recipient in the 5′ to 3′ direction (Baron 1996). The remaining unbroken single strand in the donor is then replicated creating a new complete circular double-stranded plasmid (Baron 1996). The genetic material transferred to the recipients is used to create a complimentary DNA strand and is either incorporated into the recipient’s genome or, more commonly, forms a double-stranded plasmid in the recipient (Baron 1996). Thus, at the conclusion of the conjugation process both the donor and recipient possess completely formed plasmids, thus both are now F+ (Baron 1996). Resistance genes located on the F-plasmids can be transferred during conjugation (Baron 1996). Thus, conjugation can allow for the transfer of resistance genes, which can increase the persistence of *E. coli* under environmental conditions.
1.5 E. coli and Land Use practices

The relationship between land use practices and fecal indicator organism (e.g. *E. coli*) concentration has been the subject of previous work (Petersen et al. 2018; Gotkowska-Plachta et al. 2016). Previous investigations showed that agricultural areas have statistically significant increased (*p* < 0.001) microbial concentrations relative to other land use types (Gotkowska-Plachta et al. 2016). In a study in a mixed land use watershed, Petersen et al. (2018) showed that percent agricultural land exhibited a significantly (*p* < 0.04) positive correlation with study average *E. coli* concentrations. Elevated levels of fecal contamination in agricultural areas have typically been attributed to the rearing of livestock (Rwego et al. 2008), with livestock population density being correlated to fecal indicator organism concentration (Causse et al. 2015). Manure application in agricultural areas has also been linked to increased concentrations of fecal microbes in receiving waters (Jamieson et al. 2002) as well as increased pathogen transport in the vadose (unsaturated) zone (Unc and Goss 2004).

Urban land use practices can also negatively impact microbial water quality (Wu et al. 2016). Based on logistic regression models and negative binomial regression models, Wu et al. (2016) concluded that there was a significant (*p* = 0.001) positive relationship (CI = 95%) between the percentage of urban land use and the presence of *E. coli* in waterbodies of a region in Bangladesh. Wu et al. (2016) reported similar results for the presence of both *E. coli* and total coliforms and population density (*p* = 0.001). Due to higher population density (and its link with elevated levels of fecal contamination as well as increased risk of exposure) in urban areas, waste water infrastructure and management is required to prevent the spread of potentially harmful fecal bacteria (e.g. *E. coli*) (USGS 2018). However, improper waste management practices can amplify the effect that both human and animal populations have on microbial water quality (Causse et al. 2015). For example, where wastewater is discharged directly into the waterbodies (Lyna River in northern Poland), water samples typically featured the highest number of bacteria (Gotkowska-Plachta et al. 2016). Similarly, damage to waste water infrastructure can lead to decreases in groundwater quality and contamination of surface waters (Wu et al. 2016). Sewage leakage or overflow sewage systems (especially during heavy rainfall events) are common contributors to the contamination of inland freshwater sources by fecal bacteria (Fewtrell and Kay 2015). The effects of damaged infrastructure can be exacerbated by impervious surfaces,
which are commonplace in urban areas. Impervious surfaces expedite the transport of potentially harmful microbes, such as *E. coli*, from the soil surface to receiving waters, thus spreading the contamination (Rochelle-Newall et al. 2016; Causse et al. 2015; Wilson and Weng, 2010).

Despite the progress of previous works, few studies have been published based on investigations of the relationship between land use practices and fecal contamination outside of storm events (Kandler et al. 2017) in mixed land use settings. Therefore, sites that feature numerous land use practices, such as West Run Watershed where the proposed study will take place, are well suited for future investigations. Furthermore, previous work typically utilized shorter (weeks or months) sampling periods (Petersen et al. 2018), which fail to account for seasonal variability in land use practices. Thus, studies that account for seasonal variability of land use practices, by utilizing a longer (annual) time-series of *E. coli* concentration are needed.

1.6 *E. coli* and Hydroclimate

Weather and climate have been reported to enhance the effect of land use practices on fecal concentrations in receiving waters (Shi et al. 2017). For example, agricultural non-point source fecal contamination is often greatly increased during times of heavy precipitation and runoff events. The effect of precipitation events on fecal contamination can be impacted by numerous factors such as the intensity and duration of the event, presence and amount of fecal matter (e.g. manure) in the area, fecal absorption to soil particles and the age and type of fecal matter (Rochelle-Newall et al. 2015). In a previous study Wu et al. (2016) reported that precipitation amounts (especially times of heavy rain storms) shared a significant (p < 0.05) positive association (risk ratio of 1.38 (95% CI = 1.16, 1.65) with the presence of *E. coli* in surface water. The study further showed that the effects of precipitation could remain detectable for relatively long periods (hours and even days) depending on the severity and duration of the precipitation event (Wu et al. 2016). The effects of air temperature of *E. coli* concentration in this study were more complex. Average air temperature increases over longer periods (7 – 15 days) were reported to increase the likelihood of finding *E. coli* in fresh water (Wu et al. 2016). Conversely, average air temperature increases for shorter time periods (3 days) had no effect on numbers of *E. coli*. Wu et al. (2016) concluded that changes in air temperature required more time (numerous days) to affect *E. coli* in waterbodies. These results indicate that temperature and precipitation can influence *E. coli* in different ways and at various temporal scales. However, the
The exact effect of climate variability on fecal contamination has not been well described in the literature (Wu et al. 2016).

1.7 E. coli and Suspended Particulate Matter

Suspended particulate matter in aquatic systems, defined as heterogeneous aggregates of organic matter, mineral fragments, and microbiological fractions (NAS 1977), can facilitate the growth and persistence (survivability) of fecal and pathogenic bacteria (Jamieson et al. 2005; Amalfitano et al. 2017). For example, storm-water-suspended particles were shown to extend the survival of fecal indicator organisms for several days before returning to background levels (Jeng et al. 2005). Similarly, re-suspended sediment has been hypothesized to increase the risk of contracting an E. coli related infection by approximately tenfold, due to increased persistence of particle-associated bacteria (Abia et al. 2016). Increased persistence is caused by increased nutrient and organic matter availability (comparable to biofilms), and optimal light exposure (Grossart, 2010; Drummond et al. 2015). In addition, the proximity of particle associated microbes to each other allows for the horizontal transfer and consequent proliferation of resistance genes (Allen et al. 2010, Corno et al. 2014), thus allowing associated microbes to become resistant to stresses such as chemical disinfectants, excessive photosynthetically active radiation (PAR) radiation, ultraviolet (UV) radiation, and predation (Mamane, 2008; Tang et al. 2011; Callieri et al. 2016). Thus, suspended particulate matter can cause longer periods of microbial contamination, increasing risk of exposure (Drummond et al. 2015; Garcia-Armisen and Servais 2009; Smith et al. 2008).

The health consequences of bacterial attachment (association), due to weak adsorption (van der Waals forces) or strong absorption (cellular appendages or extracellular polymers) (Palmateer et al. 1993; Jamieson et al. 2005), to suspended particulate matter, creates a need to describe the factors that drive these processes. These factors include, microbial motility (McCaulou et al. 1995), microbe type (Characklis et al. 2005), Coulomb forces (Marshall, 1975), microbe hydrophobicity (Strenstrom, 1989), extracellular polysaccharides (Lundsford et al. 2000), the ionicity of the environment (Gannon et al. 1991), and suspended particulate matter size (Muirhead et al. 2006; Amalfitano et al. 2017). For example, fecal bacteria (e.g. E. coli) associated with large suspended particles settle to the stream bed more readily than smaller particles (Oliver et al. 2006). Conversely, fecal bacteria that are unattached or associated with
small particles (≤2 μm) remain buoyant under lower flow conditions and can decrease water quality further downstream (Oliver et al. 2006). This latter point is of importance given previous work reported that microbes attach predominantly to small particles (e.g. fine clay particles (<2 μm)) (Davies and Bavor 2001; Muirhead et al. 2006).

Despite progress of previous studies, knowledge gaps, regarding *E. coli* (fecal matter) concentration relative to suspended particulate matter size distribution, remain. Previous investigations were either limited to laboratory simulations (Oliver et al. 2006), few sampling locations (Jeng et al. 2005), or similar type of land use (Characklis et al. 2005; Abia et al. 2016). The latter is important given land use practices can alter particle size characteristics of suspended particulate matter in aquatic systems (Hubbart, 2012; Kellner and Hubbart 2017a). For example, in a mixed land use watershed (Hinkson Creek, central Missouri, USA), Hubbart (2012), and Kellner and Hubbart (2019) reported that suspended sediment displayed a decreasing trend in particle size from agricultural headwaters, to urban areas and a subsequent increase to suburban areas in the lower watershed. Similarly, Kellner et al. (2013) reported a disproportionate contribution of fine sediment from urban areas, relative to receiving waters in Flat Branch Creek and Hinkson Creek. Thus, the relationship between *E. coli* (fecal matter) concentration and suspended particulate matter size can potentially be influenced by particle size distribution stemming from land use practices. The impacts of land use practices on the relationship between *E. coli* (fecal matter) concentration and suspended particulate matter size has been noted as an area requiring further research (Hubbart, 2012; Chen and Chang 2014; Kellner and Hubbart, 2019).

2. Statement of need

This dissertation research is warranted given the following arguments:

- There is a need to investigate the effect of land use practices on *E. coli* concentrations in receiving waters that occur irrespective of storm events, feature numerous sampling locations, and longer sampling periods, as historic studies focus primarily on episodic events and fail to account for seasonal variability.
• *E. coli*-based fecal contamination studies that implement the experimental watershed study are needed, as it has seldom been done and could provide new insight into *E. coli* regimes in receiving waters.
• The limited field-based publications characterizing the influence of land use practices on *E. coli* (fecal matter) concentration relative to suspended particulate matter size distribution creates a need for investigation, particularly field-based studies featuring different land use practices, on this relationship.
• Given coupled *E. coli* and physicochemical and hydroclimate metrics in previous work, there is a need for studies encompassing relationships of physicochemical parameters and hydroclimate metrics on *E. coli* concentration in receiving waters.

3. Objectives

The overarching objective of the following dissertation research was to quantify yearlong (weekly sampling interval) *E. coli* concentrations at sub-watershed scales, with varying dominant land-use types, using an experimental watershed approach. Sub-objectives were be to a) (chapter 2) characterize *E. coli* concentration relative to suspended particulate matter size distribution (filtration study), including quantifying the effect of contrasting land use practices (n = 4 monitoring sites), b) (chapter 3) analyze the influence of suspended solids, and varying land use practices on *E. coli* concentrations in receiving waters (n = 22 monitoring sites, and c) (chapter 4) quantitatively describe relationships between *E. coli* concentrations, physicochemical parameters, hydroclimate metrics and varying land use practices (n = 22 monitoring sites).

4. Dissertation Organization

This dissertation is formatted as follows. Chapter 1(current chapter) is an introductory chapter comprising background information, study objectives and applied methodologies of the dissertation. Methods (except gravimetric analysis and filtration) apply to the entire work as the same study watershed and design is relevant to the entire work. Published manuscripts are presented in chapters 2, 3, and 4 (i.e. primary literature publications), and were published in peer reviewed journals prior to dissertation defence (summer 2020). The formatting of chapters 2 to 5 were kept consisted with their published formats (or intended publication in the case of chapter
5), hence referencing styles may vary. Chapter 2 comprised characterization of *E. coli* (fecal matter) concentrations relative to suspended particulate matter size distribution, and an investigation into the influence of land use practices on this relationship. Chapter 3 includes a quantitative description of the relationship between *E. coli* concentration, land-use practices and suspended particulate matter. Chapter 4 consists of a quantitative analysis of *E. coli* concentration in relation to land-use practices, stream physicochemical parameters, and hydroclimate metrics. Finally, chapter 5 includes a review of environmental factors impacting the survival of *E. coli* in the secondary habitat (natural environment) and a section concluding the dissertation. Ultimately, the flow of the work proceeds from a smaller scale (both spatially and temporally) study focussed on investigating SPM size distribution and *E. coli* concentration relationships (Chapter 2) to a larger scale investigation of SPM, land use and *E. coli* concentration relationships across all 22 sampling sites (Chapter 3). Thereafter, the work focusses on investigating physicochemical impacts of land use practices (both historic and contemporary) on *E. coli* concentrations. The work concludes with a review of the factors known to impact environmental *E. coli* populations, the identification of current knowledge gaps and proposing areas of required future research.

The methods section below is separated into four distinct portions. The first portion of the methods section contains a description of the methods applied during the entire investigation, including descriptions of WRW, the study sites and experimental design and applied GIS methodology. The following three portions includes descriptions of the methods used during the work comprising chapter 2, chapter 3 and chapter 4, respectively. Instances where similar methods were utilised in multiple manuscripts (e.g. *E. coli* quantification; used in chapter 2, 3 and 4) have been noted, to avoid unnecessary repetition of the same descriptions.

5. Methods

5.1 Study Watershed

The research was conducted in West Run Watershed (WRW) located in Morgantown, West Virginia, USA. WRW is 23 km² in area and its primary stream, West Run Creek, is a tributary of the Monongahela River (WVWRI 2008). West Run Creek is typically narrow with small floodplains and is considered a moderately entrenched stream (Petersen et al. 2018;
Kellner et al. 2018). There are no dams or reservoirs in the watershed and its narrow, shallow streams render boats unusable. Elevation ranges from 420 m above mean sea level, at its highest point, to 240 m above mean sea level at the confluence of the Monongahela River (WVWRI, 2008). The oldest recorded geological formation in the watershed is the Upper Kittanning coal, while the most recent formation is the Monongahela series (located in the headwaters) (WVWRI, 2008). The Pittsburgh coal seam is considered the most valuable mineral resource in the county (WVWRI 2008) and is at least 2 m thick and stretches across a wide proportion of WRW (WVGES 1913). Historic mining of the seam in the upper area of the watershed, created water quality problems in parts of WRW).

Coal mining (especially in the areas surrounding the headwaters) took place in WRW with peak production between 1930 and 1950 (WVWRI 2008). Underground mining had completely ceased by the time the Surface Mine Control and Reclamation Act of 1977 (SMCRA) was enacted (WVWRI 2008). However, the effects of mining activities were still evident after mining had ceased. For example, 24 Problem Area Descriptions (all directly related with mine water) were identified by the Abandoned Mine Land (AML) Reclamation Program in 1984. The water quality of approximately 11 km of the primary stream of WRW and about 9.5 km of its tributaries have been impaired by historic mining activities (WVWRI 2008).

The WRW was rapidly developing; however, some areas were still undeveloped, lacking infrastructure such as buildings or paved surfaces (Petersen et al. 2018). Much of the undeveloped land in WRW was used primarily for agricultural purposes (WVWRI 2008). At the time of the investigation West Virginia University (WVU) was the largest individual contributor to agriculture in WRW, owning the organic, animal husbandry, and agronomy farms (Petersen et al. 2018; WVWRI 2008). Apart from agriculture, WRW was characterized by numerous additional land use activities (Table 1.1, Figure 1.1) ranging from urban and business areas to forested regions (WVWRI 2008). The mixed-land-uses in WRW allowed for an assessment of how land use activities influence fecal contamination and suspended sediment in the watershed.

5.2 Study Sites and Design

A scale-nested and paired experimental watershed study design (Hubbart 2007, Hubbart et al. 2011, Zeiger and Hubbart 2016, Hubbart et al. 2017, Wei et al. 2017) comprising twenty-two study sites (i.e. gauge sites) was implemented in 2017. Gauging sites were characterized by
contrasting land use/land cover (Table 1). Study sites and associated sub-catchments were identified and selected via a combination of GIS (see GIS methodology section) and field surveys. Sites #3, #4, #6, #10, #13, #18, #19, #21 and #22 (numbered in downstream order) were located on West Run Creek, while sites #1, #2, #5, #7, #8, #9, #11, #12, #14, #15, #16, #17, and #20 were located on tributaries of the creek (Figure 1). Site #17 (0.7 km²) which had the largest percentage of forested land use (82%) was considered a reference (control) site. In 2011 forested land-use comprised 50.2% of the land-use practices (NLCD 2011) in the watershed (land use delineations for WRW are shown in Figure 1) and was the dominant land use type in all sub-catchments except sites #1, #15, and #20. The primary land-use practices at sites #1, #15, and #20 was developed land, comprising 56.4%, 40.6%, and 54% of the land-use at these sites respectively (Table 1). In general, developed land accounted for 27.2% of the land-use in WRW. Agricultural land use accounted for 22.6% of the total land-use in WRW, however 16 of the 22 sites contained more than 20% agricultural land-use in their sub-catchments. For example, site #16 (0.2 km²) had 42.1% agricultural land-use and drained dairy cattle grazing pastures, holding pens, and manure stacks located on the WVU Experimental Agricultural Farm. Ultimately, agricultural land-use was widely distributed throughout WRW. The sampling period for the study (apart from the filtration work) was one calendar year to account for seasonal variability in the E. coli concentration, physicochemical and hydroclimate data. The sampling duration was longer than typical studies on fecal contamination (Dusek et al. 2018; Wu et al 2011; Desai and Rifai 2010), allowing for comprehensive quantification of E. coli regimes at sub-watershed mixed-land-use scales. Given related costs (labor and fiscal) of weekly analyses (n = 53) and the number of sampling sites (n = 22), a study beyond one year was infeasible. The high-resolution study design resulted in a total of 1166 spatio-temporally delineated E. coli concentration values.
Table 1. Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at twenty-two monitoring sites comprising the experimental watershed study in West Run Watershed (WRW), West Virginia, USA. Note: Data extracted from NLCD 2011 dataset with ERSI ArcGIS 10.6 Software.

<table>
<thead>
<tr>
<th>Site</th>
<th>Forest (%)</th>
<th>Ag (%)</th>
<th>Developed (%)</th>
<th>Drainage Area (km$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.3</td>
<td>30.2</td>
<td>56.4</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>78.3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>61.5</td>
<td>18.4</td>
<td>20.1</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>59.3</td>
<td>17.8</td>
<td>22.9</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>54.0</td>
<td>25.4</td>
<td>20.6</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>57.9</td>
<td>21.7</td>
<td>20.4</td>
<td>3.7</td>
</tr>
<tr>
<td>7</td>
<td>64.7</td>
<td>26.6</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>61.0</td>
<td>19.6</td>
<td>19.3</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>58.4</td>
<td>23.4</td>
<td>18.1</td>
<td>2.3</td>
</tr>
<tr>
<td>10</td>
<td>58.2</td>
<td>22.6</td>
<td>19.2</td>
<td>6.2</td>
</tr>
<tr>
<td>11</td>
<td>49.9</td>
<td>31.3</td>
<td>18.8</td>
<td>1.7</td>
</tr>
<tr>
<td>12</td>
<td>43.7</td>
<td>28.2</td>
<td>27.9</td>
<td>1.2</td>
</tr>
<tr>
<td>13</td>
<td>51.8</td>
<td>26.0</td>
<td>15.1</td>
<td>10.5</td>
</tr>
<tr>
<td>14</td>
<td>54.0</td>
<td>31.8</td>
<td>7.6</td>
<td>3.3</td>
</tr>
<tr>
<td>15</td>
<td>27.6</td>
<td>20.8</td>
<td>40.6</td>
<td>1.0</td>
</tr>
<tr>
<td>16</td>
<td>49.1</td>
<td>42.1</td>
<td>6.6</td>
<td>0.2</td>
</tr>
<tr>
<td>17</td>
<td>82.4</td>
<td>13.3</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>18</td>
<td>52.0</td>
<td>26.8</td>
<td>14.2</td>
<td>16.4</td>
</tr>
<tr>
<td>19</td>
<td>52.0</td>
<td>24.9</td>
<td>22.9</td>
<td>18.8</td>
</tr>
<tr>
<td>20</td>
<td>34.7</td>
<td>11.2</td>
<td>54.0</td>
<td>3.4</td>
</tr>
<tr>
<td>21</td>
<td>49.8</td>
<td>22.7</td>
<td>27.4</td>
<td>22.9</td>
</tr>
<tr>
<td>22</td>
<td>50.2</td>
<td>22.6</td>
<td>27.2</td>
<td>23.2</td>
</tr>
</tbody>
</table>
Figure 1. West Run Watershed experimental watershed study design comprising 22 monitoring sites, in West Run Watershed, West Virginia, USA

5.3 GIS methodology

The 2011 land use and land cover (LULC) raster dataset, provided by National Land Cover Database (NLCD), was used to extract geographical data for WRW. Additionally, a 1-meter digital elevation model (DEM), was obtained from the WVU GIS Technical Center (located in the Department of Geology and Geography at WVU). This DEM was compiled from multiple lidar-based Statewide Addressing and Mapping Board (SAMB) and United States Geological Survey (USGS) DEM datasets. Data were processed using the ERSI ArcGIS 10.6 and ArcGIS Pro Hydrology toolsets to acquire drainage areas for the 22 sampling locations. Processing included using the Flow Direction tool to calculate the direction of surface runoff, using a 3x3 grid around every cell (dimensions 3m x 3m). The cell with lowest elevation was assigned as the direction of flow. Subsequently flow accumulation was calculated, by
determining the number of upstream cells that flow through each cell, and outputting accumulation as raster grid. The raster grid was then used to derive pour points, which were located on the junctions of the stream network in the watershed. Finally, the sub-catchments were delineated with the derived datasets using the watershed tool (Wei et al. 2017; Jenson and Domingue 1988; Watson 2016). Various land use classes within each sub-catchment were grouped and reclassified into four broad categories, namely forest, agriculture, developed and open water. Finally, the raster calculator was used to calculate percentage cover of each broad land use category, for each sub-catchment, using the number of cells within each sub-catchment and area (square kilometers) per land cover (Table 1). The GIS methodology provided here was repeated and updated with the National Agriculture Imagery Program (NAIP) 2018 dataset for the work completed in chapters 3 and 4 as the data were released too late to be used for chapter 2.

6. Methods: Chapter 2

Due to the labor and supply requirements for this portion of the work, a year-long sampling regime was infeasible. As such, a sub-set of four sampling sites (#14, #15, #16, and #17) (Table 2), all first and second order tributaries of West Run Creek, was selected (Figure 2) and sampling took place from 7/20/2018 to 10/27/2018. During the sampling period 32 samples were collected from each site, thus a total of 128 samples were collected. The study sites featured contrasting land use practices, making them ideal to assess the impacts of varying land use practices on E. coli concentration relative to suspended particulate matter size distribution. Flow conditions at the sampling locations determined the timing of sample extraction, as high, low and normal flows were sampled. Sampling during various flow conditions meant results would not be constrained by flow conditions.

Table 2. Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at four monitoring sites comprising the filtration study in West Run Watershed (WRW), West Virginia, USA. Note: Data extracted from NLCD 2011 dataset.

<table>
<thead>
<tr>
<th>Site</th>
<th>Forest (%)</th>
<th>Ag (%)</th>
<th>Developed (%)</th>
<th>Drainage Area (km$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54.0</td>
<td>31.8</td>
<td>7.6</td>
<td>3.3</td>
</tr>
<tr>
<td>2</td>
<td>27.6</td>
<td>20.8</td>
<td>40.6</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>49.1</td>
<td>42.1</td>
<td>6.6</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>82.4</td>
<td>13.3</td>
<td>0.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>
WRW streams are typically shallow (stream order $\leq 3$), therefore the grab sample method was used to collect water samples as per USGS (2006) and as per Petersen et al. (2018), Hubbart et al. (2017), Kellner and Hubbart (2017b), Zeiger and Hubbart (2017), and Zeiger and Hubbart (2016). Samples were collected in a sterile 1000 ml polypropylene bottle, after rinsing the bottle three times using water from the site. Rinsing the bottle decreased the chance of sample contamination by removing impurities. During sampling, the open mouth of the bottle was plunged vertically downward (i.e. bottle mouth down) to approximately 60% stream-depth, at which point the bottle was turned so that the mouth points upstream to collect the sample. Care was taken to avoid disruption of the streambed, minimizing the risk of sample contamination,
and ensuring that the sample was representative. If it was necessary to wade into the stream for sample collection, samples were extracted upstream from point of entry and any bed disturbance. Grab sample collection was completed prior to any other sampling to reduce the risk of grab sample contamination. These sampling procedures were based on the method proposed and endorsed by the United States Geological Survey (USGS 2006).

6.2 Quantifying E. coli

Once the samples were transported to the laboratory (The Interdisciplinary Hydrology Laboratory; https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart), the process of quantifying E. coli concentrations using the IDEXX™ method was initiated. The quantification method was based on the EPA-approved Colilert Quanti-Tray Enumeration Procedure, described in the Colilert Procedure Manual (IDEXX 2017). The first step of the method was the addition of one packet of Colilert’s Defined Substrate Technology (DST) to 100 mL of sampled water and thoroughly mixing the sample until the DST completely dissolved. The DST allows for simultaneous detection of both total coliforms and E. coli (IDEXX 2107). DST contains two nutrient indicators: ortho-Nitrophenyl-β-galactoside (ONPG) and 4-methylumbelliferyl-beta-D-glucuronide (MUG) (IDEXX 2107). These indicators serve as the primary carbon sources in the Colilert test and can be metabolized by the coliform enzyme β-galactosidase and the E. coli enzyme β-glucuronidase, respectively (IDEXX 2107). When ONPG is metabolized it changes from colorless to yellow, while the metabolism of MUG results in a change from colorless to fluorescent (IDEXX 2107). As the study focused on E. coli as an indicator organism, only the enzyme β-glucuronidase (fluorescence) was measured. Following the addition of the DST, the sample was poured into a Quanti-Tray and sealed using an IDEXX™ Quanti-Tray Sealer PLUS and incubated at 35 °C for 24 hours.

Following incubation, the Quanti-Trays, consisting of 96 wells, divided into 48 large wells and 48 small wells, were exposed to UV light (a 6-watt, 365-nm wavelength UV light) within a box (UV lamp black-box). The UV light caused E. coli containing wells to fluoresce which were then enumerated. The number of fluorescing wells were then compared to the Quanti-Tray Most Probable Number (MPN) table, originating from the Standard Methods Approach, yielding a concentration value in colony forming units (CFU) per 100 mL. The Quanti-Tray eliminated the need for sample dilution and allowed for a counting range of 1–
1,011.2 CFU per 100 mL. The counting range allowed for an assessment of potential risk and the identification of harmful sites, as *E. coli* concentrations exceeding 126 CFU per 100 mL can be considered harmful during fresh water recreational activities (USEPA, 1986).

The accuracy and reliability of the IDEXX™ method stems from the fact that most non-coliform bacteria lack the enzymes to metabolize the provided carbon source (MUG in this study). Furthermore, the few non-coliform bacteria that can metabolize MUG (interfere), are eliminated by a selectively suppressing formulated matrix, created for this purpose by IDEXX™. Therefore, the chances of reporting false positives during testing was extremely low. However, it remains a possibility that up to 10% of *E. coli* populations can consist of non-gas producing strains (IDEXX 2017). These strains were not detectable with the applied method.

### 6.3 Gravimetric Analysis

Water samples were refrigerated (at 3.3°C) and gravimetric analyses (vacuum filtration) were conducted as per the American Society for Testing and Materials, test number D 3977 – 97, (ASTM; 2007) within a few days of collection. Gravimetric analysis was used to determine the mass and size distribution of suspended particulate matter. The procedure included drying (at 105°C for 1 hour) and weighing filters prior to filtration. Water samples were then passed through the filters, (and incubated to test for *E. coli* during the filtration study, see Sample Filtration section below) following which the filters were again dried (at 105°C for 1 hour) and weighed. The change in mass of the filter, and the amount of water that was filtered, was used to determine the concentration of suspended particulate matter (mg/L) in the stream, at or above the size of the aperture of the filter. This process was repeated for every filter aperture size (60 µm, 5 µm, 0.7 µm, and 0.2 µm), yielding a particle size distribution of the suspended particulate matter.

### 6.4 Sample Filtration

Currently, no widely accepted method exists for investigating *E. coli* concentration relative to suspended particulate matter size distribution (Amalfitano et al. 2017). However, numerous previous studies implemented one-size filtration to separate particle-attached from free-living microbes (Crump et al. 1999; Riemann and Winding 2001; Ortega-Retuerta et al. 2013; Rieck et al. 2015). For the current work, samples from each site were subdivided into three water sub-samples. One sub-sample was incubated without filtration (i.e. normal sample
processing), while the other two subsamples were filtered using different filter pore sizes (60 µm and 5 µm) and then incubated. Filter pore sizes that were used were selected by soil particle size classifications (i.e. larger than 60 µm = sand; smaller than 60 µm and larger than 4 µm = silt, and smaller than 4 µm = fine silt and clay) (Wentworth 1922) and by accounting for the dimensions of E. coli (2 µm X 1 µm) (NRC 1999). Thus, the 5 µm filter was the closest available filter to the 4 µm size classification of silt, which would allow E. coli to pass through.

7. Methods: Chapter 3

Chapter 3 comprises an investigation into the relationship between E. coli concentration, land-use practices and suspended particulate matter. Chapter 3 data collection included Water Sampling, Gravimetric Analysis (using a 0.7 µm filter), and Quantifying E. coli methods discussed in the Methods: Chapter 2 section. Additional Chapter 3 methods were as follows:

7.1 Measuring Stream Stage

Stream stage (at 30-min intervals) was monitored at the twenty-two study sites shown in Figure 1, during the study period (01/02/2018 – 01/01/2019), with Solinst Levelogger Gold pressure transducers, installed in two-inch poly vinyl chloride (PVC) stilling wells.

8. Methods: Chapter 4

Chapter 4 consisted of studying the relationship between E. coli concentration, land-use practices, stream physicochemical parameters, and hydroclimate metrics. Data collection for Chapter 4 included Water sampling and Quantifying E. coli methods discussed in the Methods: Chapter 2 section and Measuring Stream Stage method discussed in the Methods: Chapter 3 section. Additional Chapter 4 methods were as follows:

8.1 Measuring Water Quality Variables

Concurrent with the collection of water grab sampling (i.e. weekly from twenty-two study sites, and additionally from the Chapter 2 study), five physicochemical variables were monitored. The variables included water temperature, dissolved oxygen (DO), specific conductance (SPC), pH, and Chloride ion. Data were collected using a handheld multi-parameter water quality sonde (YSI Inc. / Xylem Inc.) fitted with an Ion Selective Electrode (ISE) multi-probe (YSI 2017). The ISE probe sensed water temperatures ranging from -5 °C to 70 °C with an
accuracy of ±0.2 °C, SPC ranging from 0 to 200 mS cm⁻¹ with an accuracy of ±0.5% of reading or ±0.001 mS cm⁻¹, whichever is greater, (for readings 0 - 100 mS cm⁻¹) or ±1.0% of reading (for readings 100 - 200 mS cm⁻¹), DO ranging from 0 to 50 mg L⁻¹, with an accuracy of ±1% (for readings 0-20 mg L⁻¹) or ±8% (for readings 20-50 mg L⁻¹), pH ranging from 0 to 14 units, with an accuracy of ±0.2 units, and Chloride ranging from 0 to 1000 mg/L (at water temperatures from 0 to 40°C) with an accuracy ±15% of reading or 5 mg/L (whichever is greater) (YSI 2017).

The method of sensing of water quality variables by the Yellow Springs Instrument (YSI) differed among the respective variables. Water temperature was measured by a thermistor (combination sensor with conductivity) (YSI 2019a). The YSI utilized a high-precision thermistor, 2252 ohms at 25°C (±1%), for temperature measurement. Resistance changes with temperature and the 6-Series sondes converted resistance into °C, °F, or K automatically (YSI 2019a). The YSI used the thermistor readings to temperature-correct other parameter measurements. The specific conductance reading was based on a calculation (involving conductivity and temperature readings) made by the YSI. The YSI measured conductivity with a four-nickel electrode cell (YSI 2019a). Two of the electrodes were current driven, and two were used to measure the reduction in voltage. The measured voltage reduction was then converted into a conductance value in milli-Siemens (millimhos). This value was then converted into a specific conductance value (milli-Siemens per cm) (YSI 2019a).

PH was measured by a glass bulb combination electrode consisting of a proton selective glass reservoir, filled with buffer at approximately pH 7, and a Ag/AgCl reference electrode utilizing a gelled electrolyte (YSI 2019b). The buffer reservoir contained a silver wire coated with AgCl. This created a potential gradient across the glass membrane as protons (H⁺ ions), on both sides of the glass (media and buffer reservoir), selectively interact with the glass. Since the hydrogen ion concentration in the internal buffer solution was invariant, the potential difference, determined relative to the Ag/AgCl reference electrode, was proportional to the pH of the media (water) (YSI 2019b).

DO was measured by optical luminescence (lifetime method) (YSI 2011), which utilizes a light source (LED) to create a beam of excitation light (YSI 2019c). The excitation light travels to the end of the probe and strikes an oxygen-permeable, luminophore-embedded sensing layer (or substrate or foil) (YSI 2019c). The excitation light is absorbed and excites the luminophore...
molecules. While excited, the quality of the absorbed energy and its modulation are affected by the environment surrounding the lumiphores. The excited luminophore molecules release the excitation energy by emitting the energy back as light. The amount of time between absorption and emittance is the luminescence lifetime. The luminescence lifetime is a precisely known quantity and is predictably affected by several physical factors, including the amount of oxygen present in the medium YSI 2019c). Thus, the environmental oxygen concentration quantitatively causes an effect on the luminescence lifetime. The altered light that is emitted back to the probe is measured by an electronic light sensor (photodiode). The signal from the emitted light is then used to calculate a value for the quantity of dissolved oxygen (YSI 2019c).

Chloride was measured by an ISE, which is a potentiometric procedure, meaning the measured signal is in the form of potential differences, or voltages in mV (Higgins 2013). The ISE comprises two components: an ion sensitive measuring system reacting on a special ion type, and a reference electrode. Once immersed in the sample (water) the ion sensitive measuring system provides an electrochemical potential that is influenced by the concentration of the ion type to be measured. Simultaneously, the reference electrode, which is uninfluenced by the sample, also builds up an electrochemical potential. The difference of these potentials, the voltage displayed by the meter, can be described by the Nernst equation (Equation 3) (Higgins 2013):

\[ U_{\text{ion}} = U_{\text{ion}}^0 \pm S \log (a_{\text{ion}}) \]  

(1)

Where \( U_{\text{ion}}^0 \) is a fixed value given by the measuring system. The slope S (Nernst slope is approximately equal to 59.16 mV at 25°C) defines how much the measurement signal is increased or reduced when the concentration changes. The arithmetic sign depends on the charge of the measured ions (i.e. a plus is for positively charged cations, e.g. K\(^+\), a minus for anions such as Cl\(^-\)) (Higgins 2013). The activity a is the active concentration of an ion type that considers the influence of other ions in the sample.

The probes on the YSI were calibrated weekly, as is standard protocol (YSI 2017). Calibration of pH, conductivity, and chloride probes consisted of placing the probes into the relevant calibration solutions, waiting for the YSI to calibrate accordingly, and rinsing the probes with distilled water (DI) between calibrations (YSI 2017). The calibration process was initiated by removing the chloride probe and placing it in a beaker filled with DI (YSI 2017). Following
the removal of the chloride probe, the pH sensing probe was calibrated using three pH solutions in the sequence of pH 7, pH 4, and pH 10 (YSI 2017). Conductivity calibration, requiring two solutions 447.1 (used first) and 1413, commenced after pH calibration was completed (YSI 2017). Following the conductivity calibration, the chloride probe was reattached to be calibrated. This calibration required a 10mg/L (used first) and a 1000mg/L solution (YSI 2017).

Data collection with the YSI consisted of placing the sonde into the stream, upstream of the operator, at 60% depth to sense a representative well-mixed sample (USGS 1995). Data collection commenced at 10:00 (site #1) and continued through the numerical order of sites from 01/02/18 to 01/01/2019. Sites #9 and #10 were exceptions, as they were sampled before sites #7 and #8, due to their location relative to other sites (Figure 1). The proximity of sites #9 and #10 to site #6 meant that overall sampling time was reduced by sampling them after site #6. This reduced the delay between sample extraction and sample processing, increasing the representativeness of the extracted samples.
References


Baron, S. 1996. Medical Microbiology. 4th edition. Chapter 5: Genetics, University of Texas Medical Branch at Galveston. Available at: https://www.ncbi.nlm.nih.gov/books/NBK7627/


Goyal, S.M., Gerba, C.P., Melnick, J.L. 1977. Occurrence and Distribution of Bacterial Indicators and Pathogens in Canal Communities along the Texas Coast. Applied and Environmental Microbiology, p. 139-149


Kandler, M., Blechinger, K., Seidler, C., Pavlu, V., Sanda, M., Dostal, T., Krasa, J., Vitvar, T., Stich, M. 2017. Impact of land use on water quality in the upper Nisa catchment in the
Czech Republic and in Germany, Science of the Total Environment, Volume 586: 1316–1325


Mann, H. B., & Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50-60.


Microtrac. 2014. Microtrac S3500 Particle Size Analyzer Operation and Maintenance Manual for Model Numbers Beginning with ‘M355x–…’


http://dep.wv.gov/WWE/Programs/nonptsource/WBP/Documents/ WP/WestRun_WBP.pdf


ISBN 0-397-54313-1


Yellow Springs Instruments (YSI). 2019a. YSI Model 85 Operations Manual, YSI [available online]:

Yellow Springs Instruments (YSI). 2019b. YSI Model 60 Operations Manual, YSI [available online]:

Yellow Springs Instruments (YSI). 2019c. Luminescent-Based Detection Methods for Dissolved Oxygen Proposed by U.S. Environmental Protection Agency, YSI [available online]:


CHAPTER 2: QUANTIFYING E. COLI AND SUSPENDED PARTICULATE MATTER CONCENTRATIONS IN A MIXED-LAND USE APPALACHIAN WATERSHED

Citation: Petersen, F.; Hubbart, J.A. Quantifying Escherichia coli and Suspended Particulate Matter Concentrations in a Mixed-Land Use Appalachian Watershed. Water 2020, 12, 532. (https://doi.org/10.3390/w12020532)

Abstract: The relationships between Escherichia (E) coli concentration, suspended particulate matter (SPM) particle size class, and land use practices are important in reducing the bacterium’s persistence and health risks. However, surprisingly few studies have been performed that quantify these relationships. Conceivably, such information would advance mitigation strategies for practices that address specific SPM size classes and, by proxy, E. coli concentration. To advance this needed area of research, stream water was sampled from varying dominant land use practices in West Run Watershed, a representative mixed-land use Appalachian watershed of West Virginia in the eastern USA. Water samples were filtered into three SPM intervals (<5 µm; 5 µm to 60 µm; and >60 µm) and the E. coli concentration (colony forming units, CFU) and SPM of each interval was quantified. Statistically significant relationships were identified between E. coli concentrations and size intervals (α < 0.0001), and SPM (α = 0.05). The results show a predominance (90% of total) of E. coli CFUs in the <5 µm SPM interval. The results show that land use practices impact the relationships between SPM and E. coli concentrations. Future work should include additional combined factors that influence bacterial CFUs and SPM, including hydrology, climate, geochemistry and nutrients.

Keywords: Escherichia coli; Suspended particulate matter; Water quality; Land use practices; Watershed management

1. Introduction

Globally, there is a need to investigate factors that influence risks facilitated by pathogenic microbes in water sources [1]. The need for research is stimulated by the common and widespread occurrence of fecal pollution and pathogenic water contamination in many parts of the globe [2–5]. The World Health Organization provided an estimate of 2.2 million deaths that occur annually
due to waterborne diseases, making it the leading cause of deaths in the developing world [6]. Between 2013 and 2014, pathogen-contaminated drinking water resulted in 1006 cases, 42 outbreaks of disease, 124 hospitalizations, and 13 deaths in the United States of America alone [7]. Conceivably, advancing the understanding of variables that influence the persistence and risk of exposure to potentially harmful fecal microbes (e.g., *E. coli, enterococci* or fecal coliform), including the association of bacteria with suspended particulate matter (SPM), will better inform policy makers and contribute to decreased morbidity and mortality caused by pathogenic microbes.

SPM, defined as heterogeneous aggregates of organic matter, mineral fragments and microbiological fractions [8] in aquatic ecosystems, can facilitate the growth and persistence (survivability) of fecal and pathogenic bacteria [9,10]. For example, Jeng et al. [11] reported that suspended particles extended the survival of fecal indicator organisms for several days through the physical and chemical protection of microbes from biotic and abiotic stresses [12]. Other studies showed that SPM-associated bacteria occur in conjunction with, or adsorbed to, suspended organic or inorganic matter via physical, electrostatic or chemical binding [8,13], and benefit from the increased nutrient and organic matter availability (comparable to biofilms), and optimal light exposure [14,15]. Increased nutrients, organic matter and light that SPM-associated bacteria are exposed to can increase the growth rate of microbes by up to 50% relative to free-floating microbes. SPM-associated bacteria are typically larger, more abundant and feature increased cellular uptake of sugars and amino acids relative to free-floating microbes [16,17]. In addition, the proximity of SPM-associated microbes to each other facilitates horizontal transfer and potential proliferation of resistance genes [18,19]. Advancing the understanding of variables that influence bacterial association with SPM is important to better manage the growth and persistence of microbes in the environment.

Previous work showed that microbes attach preferentially to small particles (e.g., fine clay particles; <2 µm) due to the increased surface area to volume and electrostatic charge [1,12]. Conversely, fecal bacteria (e.g., *E. coli*) associated with large suspended particles will settle to the stream bed more readily than smaller particles [1], potentially decreasing microbial water quality impairment once re-suspension occurs. The potential risk of exposure to *E. coli* through its association with SPM is important given that *E. coli* can be pathogenic [20], while simultaneously indicating the presence of other harmful fecal microbes (e.g., *E. coli* serves as an indicator organism) [2]. This is important because association can increase the risk posed by *E. coli* or by
the other fecal microbes that *E. coli* is an indicator for. Thus, an improved understanding of the relationships between *E. coli* concentration and SPM could aid in determining the influence of SPM size on other potentially harmful fecal bacteria, commonly occurring with *E. coli*, such as *Enterococcus* [21]. Pathogenic strains of *E. coli* can cause gastroenteritis, diarrhea, urinary tract infections hemolytic uremic syndrome, and meningitis [22]. Therefore, the increased risk of exposure, through association with SPM, can potentially increase instances of disease outbreaks and subsequent morbidity and mortality [2,3]. In short, advanced understanding of the correlation between *E. coli* concentrations and SPM size class will reduce human health risks by advancing the understanding of exposure and informing effective remediation.

Many knowledge gaps exist regarding *E. coli* (fecal matter) concentration relative to SPM size distribution. For example, previous investigations have been limited by laboratory simulations [1], few sampling locations [11], and/or low to no land use variability [23,24]. The latter is important given land use and land use change has been shown to alter particle size characteristics of SPM in aquatic systems [25,26]. For example, in a mixed-land use watershed (Hinkson Creek, Columbia, Missouri, USA), Hubbart [25] and Kellner and Hubbart [27] reported that suspended sediment displayed a decreasing trend in particle size from agricultural headwaters to urban areas and a subsequent increase in particle size to suburban areas in the lower watershed. Kellner et al. [28] reported a disproportionate contribution of fine sediment from urban areas, relative to receiving waters comprising different land uses (e.g., rural and agriculture) of Flat Branch Creek (a tributary to Hinkson) and Hinkson Creek. The increased fine sediment from urban areas was attributed to preexisting compaction in these areas, increased impervious surface cover relative to other land use types, and in-stream weathering of sediment in conjunction with preferential deposition [25,28]. Ultimately, the results of previous studies have stimulated questions regarding the influence of land use practices, particle size class distribution and *E. coli* (fecal matter) concentration [29].

The Appalachian region of the United States of America (USA) is well-suited for investigations that will resolve existing knowledge gaps regarding the relationship between *E. coli* concentration and SPM size. This region of the USA comprises diverse physiography and widespread, frequent, and problematic fecal pollution [30]. On the basis of physiography alone, Appalachia is comparable to other regions globally. For example, the Central Appalachian region, encompassing a temperate climate, distinct winter and summer periods, and a precipitation regime
that is nearly evenly distributed throughout the year is comparable to similarly temperate locations comprising year round precipitation (e.g., Northern Honshu in Japan [31,32]). In rural areas of Appalachia, thousands of residents are exposed to water security issues, in particular microbial contamination [33]. The vulnerabilities of rural Appalachia to decreased water quality due to microbial contamination is exacerbated by some of the greatest levels of poverty, isolation, rough geographical terrain, and inadequate septic treatment systems in the USA [33]. Therefore, water quality is a primary concern and studies investigating fecal pollutants (e.g., *E. coli*) and factors increasing the risk posed by fecal pollution (e.g., by increasing risk of exposure) in this region are greatly needed.

The overarching objective of the current investigation was to quantitatively characterize *E. coli* concentration relative to SPM size distribution from multiple sites in a mixed-land use watershed of Appalachia. A sub-objective was to evaluate the influence of varying land use practices on the relationship between *E. coli* concentration and SPM particle size distribution. This work was also intended to serve as a valuable springboard for future investigations on the alteration of exposure and subsequent health risks of *E. coli* facilitated by SPM of various sizes.

2. Methods

2.1. Study Site Description

This investigation took place in West Run Watershed (WRW), a 3rd order tributary of the Monongahela River, located in Morgantown, West Virginia, USA. The WRW is 23 km² in area and is a mixed-land use urbanizing watershed comprising many land use practices, including agriculture, urban and forested areas. At the time of this investigation, forested land use accounted for 50.1%, and agricultural and developed (urban and commercial areas) land use practices accounted for 22.6% and 19%, respectively, of the land use of WRW. The primary stream of WRW, West Run Creek, is typically narrow with small floodplains and is considered to be a moderately entrenched stream [34,35]. Elevation ranges from 420 m above mean sea level, at the headwaters, to 240 m above mean sea level at the confluence of the Monongahela River [35]. The physiography of the watershed comprises relatively rugged terrain, featuring numerous rock outcroppings dating to the Paleozoic era [35]. The oldest recorded geological formation in the watershed is the Upper Kittanning coal, while the most recent formation is the Monongahela series
The water quality in WRW, specifically in the headwaters, has been negatively impacted by historic mining of the Pittsburg coal seam [36].

The climate of West Virginia ranges from temperate and humid with hot summers to cold and humid with warm summers [37]. The climate of Morgantown, WV, located in Monongalia County (and including the WRW), is characterized by warm to hot summers (mean monthly temperature >22 °C), cold winters (mean monthly temperature <0 °C), and no dry season [37]. The average annual precipitation in Morgantown between 1981 and 2010 was approximately 1060 mm. During this time period, July (typically warmest and wettest month) had an average daily temperature of approximately 23 °C and average monthly precipitation of 117 mm. Conversely, the coldest (January) and driest (February) months have an average daily temperature of −0.4 °C and average monthly precipitation of 66 mm, respectively [38].

For the current investigation, a study design including four monitoring sites (gauged sampling locations) was implemented. The sites included varying land use practices (Table 3) and (numbered in downstream order) consisted of 1st and 2nd order confluence tributaries of the West Run Creek (Figure 3). A combination of geographic information system (GIS) data and field surveys were implemented to identify study sites and associated sub-catchments. At the time of this investigation, Site #1 comprised developed and forested lands in the upper sub-catchment, and actively grazed pasture in the lower sub-catchment. The primarily urban site (Site #2) drained a commercial area located on the southern side of the Watershed. Site #3 drained a local farm which includes dairy cattle grazing pastures, holding pens, and livestock manure stacks. Site #4 comprised predominantly (82.4%) forested land use and served as a reference sub-catchment (control) for the current work.
Table 3. Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at four monitoring sites in West Run Watershed (WRW), West Virginia, USA. Note: land use percentages may not sum to 100%, as not every category is included (i.e., wetland, open water, etc.) and some categories are combinations of others (e.g., developed = urban + residential), or independent (e.g., impervious). Final row indicates total values for the entire watershed.

<table>
<thead>
<tr>
<th>Site Number and General Description</th>
<th>Forest (%)</th>
<th>Agriculture (%)</th>
<th>Developed (%)</th>
<th>Drainage Area (km$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 (54.0% Forested)</td>
<td>54.0</td>
<td>31.8</td>
<td>7.6</td>
<td>3.3</td>
</tr>
<tr>
<td>#2 (40.6% Developed)</td>
<td>27.6</td>
<td>20.8</td>
<td>40.6</td>
<td>1.0</td>
</tr>
<tr>
<td>#3 (49.1% Forested)</td>
<td>49.1</td>
<td>42.1</td>
<td>6.6</td>
<td>0.2</td>
</tr>
<tr>
<td>#4 (82.4% Forested)</td>
<td>82.4</td>
<td>13.3</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>WRW (Total)</td>
<td>50.1</td>
<td>22.6</td>
<td>19.0</td>
<td>23.3</td>
</tr>
</tbody>
</table>
2.2. Data Collection

During the study period (20 July 2018–27 October 2018), climate data were recorded using research-grade climate instrumentation located within approximately 100 m of Site #1 (Figure 3). The climate variables (recorded at a height of 3 m) included precipitation (TE525 Tipping Bucket Rain Gauge), average air temperature and relative humidity (Campbell Scientific HC2S3 Temperature and Relative Humidity Probe), and average wind speed (Met One 034B Wind Set instrument).

Stream water grab-samples were collected following the USGS methods described in the National Field Manual for the Collection of Water-Quality Data [39] and as per Petersen et al. [35], Hubbart et al. [40], Kellner and Hubbart [41], and Zeiger and Hubbart [42,43] from each
monitoring site (stream order \(\leq 3\)). The sampling regime was determined by stream stage (based on the streamflow descriptions by Zeiger and Hubbart [42]) at the sites to ensure sample extraction occurred during low (approximately 25% bankfull), medium (approximately 40%–60% bankfull) and high(er) (approximately >60% bankfull) stages. This sampling regime facilitated a distributed SPM concentration data set, providing for robust investigation of the general relationship between SPM and *E. coli* concentration. The stream stage-based sampling regime resulted in irregular sample collection. For example, multiple samples were collected on some days during runoff events (leading and receding limbs of hydrograph, etc.) and other consequent samples extracted multiple days apart. During the course of the investigation, 32 samples were collected from each of the sampling locations. Following collection, the samples were transported to the Interdisciplinary Hydrology Laboratory (https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart), located in the Davis College of Agriculture, Natural Resources and Design at West Virginia University, for analyses. Previous investigations have implemented one filter size (size varies between studies) to separate particle-attached from free-living microbes [13, 44-46], although, as noted above, no widely established method exists. During the current investigation, the extracted water grab-samples were subdivided into three water subsamples. One subsample was processed as per the standard Colilert test (see below) procedures and incubated without filtration (i.e., normal sample processing), while the other two subsamples were filtered using different filter matrices (60 µm and 5 µm), processed (as below) and then incubated. Hydrophilic, nylon net, Merck Millipore filters were used for filtration. The filtration of the samples resulted in water samples containing SPM and *E. coli* of sizes <60 µm and <5 µm, respectively. Therefore, after incubating the samples (process described below), the *E. coli* concentrations of the total sample and the <60 µm and <5 µm sizes were known. Subsequently, subtracting the *E. coli* concentration in the <60 µm filtered sample from the *E. coli* concentration of the total sample thus yielded the *E. coli* concentration in the respective SPM size class interval. Additionally, the *E. coli* concentrations in the <5 µm size were subtracted from the <60 µm size, resulting in *E. coli* concentration data for the intermediate (5 µm < interval < 60 µm) interval. The selection of the filter aperture was determined by approximate soil particle size classifications (i.e., larger than 60 µm = sand; smaller than 60 µm and larger than 4µm = silt, and smaller than 4µm = fine silt and clay) [47].
The U.S. Environmental Protection Agency (EPA)-approved Colilert test [29], developed by IDEXX Laboratories Inc., was used to quantify *E. coli* colony forming units (CFU) in the filtered and unfiltered samples. The test, included in Standard Methods for Examination of Water and Wastewater, was developed to estimate fecal concentration in water samples without requiring sample dilution [48,49]. The chances of obtaining inaccurate results with the test is low (chance of reporting false positives ±10%) due to a combination of Colilert’s Defined Substrate Technology (DST) nutrient-indicator (ONPG), and a selectively suppressing formulated matrix. ONPG was formulated to render the majority of non-target organisms unable to grow or interfere as they lack the enzyme to metabolize the provided carbon source [48]. The few non-target organisms that can metabolize ONPG were suppressed by the selectively suppressing formulated matrix [48].

The Colilert (ONPG) substrate was added to 100 mL of sampled water and sealed in the Quanti-Tray, prior to the 24-hour incubation at 35 °C [29]. Following incubation, the number of fluorescing (positive for *E. coli*) wells was converted (with a 95% confidence interval) into a concentration of *E. coli* CFU per the 100 mL of filtered and unfiltered sample water using a Most Probable Number (MPN) table. The MPN table facilitated an estimation of the concentration value, as it yielded a number of CFU per 100 ml. The Quanti-Tray/MPN table method allowed for the estimation of *E. coli* concentration in the range <1 to 1011.2 CFU. The upper limit of the estimation range from the Quanti-Tray/MPN table method presented a challenge for the current investigation, as *E. coli* concentrations in excess of 1011 CFU per 100 mL could not be accurately estimated, as the test becomes effectively saturated. Due to the fiscal cost associated with the dilution and subsequent incubation of extracted water samples, serial dilutions were not feasible for the current work, as it would have significantly increased the number of samples to be analyzed. Regardless, the method as applied allowed for the detection of small changes in *E. coli* concentration, which are important for small SPM sizes (a focus of the current work), which the literature has shown to be more closely associated with pathogen persistence.

2.3. Data Analysis

Descriptive statistics were generated for *E. coli* and SPM concentrations aggregated to the study period level. The total *E. coli* concentrations (i.e., colony forming unit; CFU) from each filtration interval (<5 µm; 5 µm to 60 µm; and >60) were divided by the total unfiltered *E. coli*
concentration. This facilitated the estimation of the average percentage *E. coli* concentration resultant from each respective SPM interval. This process was completed for all four sites for the duration of the study period (20 July 2018–27 October 2018). The average SPM and average *E. coli* concentration from each interval (n = 32) (and totals) were compared to each other to determine the average percentage difference between the sampling locations. Statistical analyses were conducted using Origin Pro 2018 (OriginLab Corporation, Massachusetts, USA). Normality testing was completed using the Anderson Darling Test, which tests whether a sample of data is drawn from a given probability distribution (normal distribution for the current work) [50]. Ordinal logistic regression was used to explore the effect of SPM size concentrations on *E. coli* concentrations in the same size intervals, while also accounting for the effect of varying land use practices at the four sampling locations. This regression is used to predict an ordinal dependent variable given one or more independent variables and requires that the continuous *E. coli* concentration and SPM concentration data be converted to ordinal data. Therefore, the *E. coli* concentration and SPM concentration data were divided into tertiles in which data in the lowest tertile were assigned a value of 1, data in the second tertile were assigned a value of 2, and data in the upper tertiles were assigned a value of 3. Following the conversion to ordinal data, the regression was calculated using JMP software (JMP®, Version Pro 12.2, SAS Institute Inc., Cary, USA, Copyright ©2015), following the method presented by Stokes et al. [51]. The significance threshold for all statistical tests was set at \( \alpha = 0.05 \). Principal component analysis (PCA) was used to investigate the relationships between *E. coli* concentrations, SPM and land use practices for both the smallest interval (<5 µm) and the total values (presented in biplots) across all four sampling locations.

3. Results

3.1. Climate during Study

During the study period (20 July 2018–27 October 2018), the total precipitation was 449 mm, the average air temperature was 19 °C, the relative humidity was 82.7%, and the average wind speed was 0.89 m/s. Historic records since 2007 [52] indicated that the study period comprised normal temperatures (average 19 °C), but received 99 mm more precipitation than average (350 mm). A monthly analysis of climate data showed that the month of July included normal average monthly temperatures (difference between recorded values and average was less than 1 °C) for the
time period (22 °C). However, there was approximately 20 mm less precipitation than average [52]. Conversely, August had average temperatures that were 2 °C cooler than average and the month was approximately 39 mm wetter than average [52]. Precipitation during September (186mm) was more than double the long-term average (80mm), but the average temperature was consistent with the long-term average (18 °C). October also exhibited long-term average temperatures (12 °C); however, it was drier than average, receiving approximately 25mm less precipitation.

The climate during the period of study (20 July 2018–27 October 2018) was predictably variable and consistent with historic trends (Figure 4). The weather was characteristically humid and warm at the start of the investigation, with temperatures decreasing with transition to winter. As previously noted, stream stage was used to determine the timing of sampling events given that streamflow (depth and velocity) is directly related to SPM transport processes and has been reported to influence *E. coli* concentrations with elevated concentrations typically occurring during high flows [53]. This approach facilitated the opportunity to investigate the relationship between *E. coli* and SPM size under a variety of hydrologic conditions. Thus, climatic conditions that could directly influence stream stage, such as precipitation, were particularly important for this work. During the investigation large precipitation events (e.g., 7/25 and 9/9) resulted in maximum stream stages, while drier periods (e.g., 10/12) resulted in low stream stages, as is common in late summer in the region (Figure 4) [35].
Figure 4. Thirty-minute time series of climate variables (recorded at the climate station) during study period (20 July 2018–27 October 2018) in West Run Watershed, West Virginia, USA. *Note: Stream stage was monitored in the primary stream of WRW, West Run Creek, within approximately 100m of the confluence of Site #1 and West Run Creek (39°40'3.20" N 79°55'48.99" W).

3.2. Suspended Particulate Matter Concentrations and Escherichia coli

The average total SPM concentration (38.52 mg L\(^{-1}\)) was the highest at Site #3, which included 49.1% forested area, the highest percentage of agricultural land-use (42.1%) and the highest maximum SPM (402 mg L\(^{-1}\)). Site #3 also had the largest standard deviation of SPM (83.91 mg L\(^{-1}\)), almost double the standard deviation of the second highest standard deviation (Site #4: 43.28 mg L\(^{-1}\)). Site #1 had the highest median (10.83 mg L\(^{-1}\)) and minimum (0.67 mg L\(^{-1}\)) SPM values. Conversely, Site #2 had the lowest average SPM (12.25 mg L\(^{-1}\)), maximum (70 mg L\(^{-1}\)), median SPM (2.67 mg L\(^{-1}\)), and minimum (0 mg L\(^{-1}\)) among the sampled sites.

Site #2 (land use: 40.6% developed, Table 1) had the lowest average E. coli concentration (596 CFU per 100 mL) and the lowest median (629 CFU per 100 mL) among the sites. The lowest
minimum *E. coli* concentration (38 CFU per 100 mL) was recorded at Site #4 (82.4% forested). Conversely, the highest mean value for *E. coli* concentration (708 CFU per 100 mL), the highest median value for *E. coli* concentration (961 CFU per 100 mL), and the lowest standard deviation (351 CFU per 100 mL) were recorded at Site #3 (42.1% Agricultural, and 49.1% Forested). Site #1, which had the second highest percentage of agricultural land use (31.8%) among the subcatchments, had the second highest mean value for *E. coli* concentration (676 CFU per 100 mL), the second highest median value for *E. coli* concentration (813 CFU per 100 mL), and second lowest standard deviation (355 CFU per 100 mL).

Box and whisker plots reflecting descriptive statistics of SPM (mg L\(^{-1}\)) and *E. coli* concentration (CFU per 100 mL) are provided in Figure 5. In the >60 µm interval, Site #2 (40.6% developed) had both the smallest SPM concentration range (0–2.5 mg L\(^{-1}\)) and the largest *E. coli* concentration range (0–262 CFU per 100 mL). However, in the 5 to 60 µm particle size class, Site #4 (82.4% forested) had the largest SPM concentration range (0–171.5 mg L\(^{-1}\)), whereas Site #3 (49.1% forested and 42.1% agriculture) had the largest *E. coli* concentration range (0–276 CFU per 100 mL). In the <5 µm interval, Site #3 had the highest SPM (0–181 mg L\(^{-1}\)) and *E. coli* concentration values (0–1011 CFU per 100 mL), while Site #1 (54% forested) had the lowest SPM concentration range (0–30.5 mg L\(^{-1}\)) and Site #4 (82.4% forested) generally had the lowest *E. coli* concentration values. The raw (unfiltered) samples had a similar distribution to the <5 µm interval, with Site #3 again featuring the highest SPM (0–402 mg L\(^{-1}\)) and *E. coli* concentration values (0–1011 CFU per 100 mL). Similarly, Site #1 (54% forested) comprised the lowest SPM concentration range (0–216 mg L\(^{-1}\)); however, Site #2 generally had the lowest *E. coli* concentration values (Figure 5).
Figure 5. Box and whisker plot of suspended particulate matter (SPM) (mg L\(^{-1}\)) and \(E.\ coli\) concentration (CFU per 100 mL) at each sampling location (n = 4) during the study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.
3.3. Non-Parametric Statistical Analysis

Normality test results indicated that *E. coli* concentration data were non-normally distributed, thus a non-parametric statistical method was used for further analysis. Due to the detection range (including an upper limit) of the Colilert method, ordinal logistic regression was deemed the most suitable form of regression to analyze the data [51]. *E. coli* concentration data for all sites were used as the dependent variable with the corresponding size interval, SPM and land use percentages (forest %, agriculture %, and developed %) data from all sites used as independent (explanatory) variables. Across the sites, the results indicate that size interval (α < 0.0001), and SPM (α = 0.05) showed statistically significant relationships with *E. coli* concentration. Percentage of land use practices across all sites did not display any statistically significant correlations with *E. coli* concentrations, with the percentage of forested land, percentage of agricultural land and percentage of developed land having significance levels (α) of 0.8425, 0.8478, and 0.9031, respectively.

The results from the initial PCA identify two principal components with eigenvalues exceeding 1, an accepted threshold of importance [54] for the smallest interval (<5 µm) (Table 4). *E. coli* concentrations (Eigenvalue = 1.89) and SPM concentrations (Eigenvalue = 1.35) explained approximately 65% of the cumulative variance of the smallest interval data set. The percentage of agricultural land comprised an Eigenvalue of 0.91 and its inclusion accounted for 83% of the cumulative variance of the smallest interval data set. For the total data set (defined as all he SPM and *E. coli* data >0.7 µm) PCA, three principal components were identified with eigenvalues exceeding 1; *E. coli* concentration (1.90), SPM concentration (1.42) and the percentage of agricultural land (1.01) and accounted for 87% of the cumulative variance of the total data set (Table 4). For both PCAs, the percentage of forested land and percentage of developed land accounted for relatively small percentages of variance in the data. Specifically, in the smallest interval, 17% and 0.1% of the variance of the data were accounted for by the percentage of forested land and the percentage of developed land, respectively. Similarly, for the total data set, forested land use accounted for 13% of the data variance and the percentage of developed land accounted for 0.04%.
Table 4. Results of principal component analysis comprising 5 components (E. coli concentration, SPM concentration, percentage of agricultural land use, percentage of forested land use and percentage of developed land use) displaying eigenvalues, percentage of variance and cumulative variance during the study period (20 July 2018–27 October 2018) for the smallest interval (<5 µm) and the total data set across the four monitoring sites in West Run Watershed, West Virginia, USA. Note: bold numbers indicate eigenvalues exceeding 1 (representing importance).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Eigenvalue</th>
<th>Percentage of Variance</th>
<th>Cumulative Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Smallest Interval (&lt;5 µm)</td>
</tr>
<tr>
<td>E. coli concentration</td>
<td>1.89</td>
<td>38%</td>
<td>38%</td>
</tr>
<tr>
<td>SPM concentration</td>
<td>1.35</td>
<td>27%</td>
<td>65%</td>
</tr>
<tr>
<td>% Agriculture</td>
<td>0.91</td>
<td>18%</td>
<td>83%</td>
</tr>
<tr>
<td>% Forested</td>
<td>0.84</td>
<td>17%</td>
<td>100%</td>
</tr>
<tr>
<td>% Developed</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Total Data Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli concentration</td>
<td>1.90</td>
<td>38%</td>
<td>38%</td>
</tr>
<tr>
<td>SPM concentration</td>
<td>1.41</td>
<td>28%</td>
<td>66%</td>
</tr>
<tr>
<td>% Agriculture</td>
<td>1.01</td>
<td>20%</td>
<td>86%</td>
</tr>
<tr>
<td>% Forested</td>
<td>0.67</td>
<td>14%</td>
<td>100%</td>
</tr>
<tr>
<td>% Developed</td>
<td>0</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

4. Discussion

4.1. E. coli and SPM Concentrations

At the time of this study, Site #3 (Table 3) had the highest percentage of agricultural land use among the study sites (42.1%) and comprised the greatest cumulative SPM (1232 mg L\(^{-1}\)) (Figure 6). The cumulative SPM recorded at Site #3 was more than three times the cumulative SPM recorded at Site #2 (40.6% developed), which had the lowest cumulative SPM (391 mg L\(^{-1}\)). Previous studies investigating land use practices and SPM also recorded elevated levels of SPM in agricultural land use areas, thereby supporting the results from Site #3 [55]. The flattened cumulative curves (Figure 6) from 28 September 2018 to 19 October 2018 reflect a lack of sample collection during a period of negligible rainfall (and thus runoff; see Figure 2). Cumulative E. coli was the highest at Site #1 (21632 CFU per 100 mL) and Site #3 (22641 CFU per 100 mL) during the sampling period (20 July 2018–27 October 2018) (Figure 6). These two sites drained the highest (42.1%; Site 3) and second highest (31.8%; Site #1) area of agricultural land use practices among the sampled sites (Table 3). The results are supported by previous work that showed
increased fecal matter content with agricultural land use practices [53,56]. Conversely, Site #2 and Site #4 had the lowest cumulative *E. coli* CFU (19082 CFU per 100 mL and 20080 CFU per 100 mL, respectively) during the study period. Site #4’s low cumulative *E. coli* CFU was anticipated, as this site consisted primarily of forested land use (82.4%) (Table 3) and previous work showed negative correlations between fecal concentration and forest land cover [56]. Furthermore, Site #4 lacked artificial sources of *E. coli* (e.g., livestock manure stacks present at Site #3), which could have increased the *E. coli* concentration in its receiving waters. The low cumulative *E. coli* CFU recorded at Site #2 is contrary to results from previous work linking urban land use practices to increased *E. coli* concentrations in receiving waters [57–59]. Given the differing study design (sampling regime) relative to other studies, these differences are not confounding. It is, however, worth noting that in cases where the current study agreed with previous studies, those agreements are important, given the different sampling regime and yet agreement(s) in results.
Figure 6. Top: cumulative total SPM (mg L\(^{-1}\)). Bottom: cumulative total *E. coli* concentration (CFU per 100 mL) at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA. *Only predominant land use percentage indicated, for full land use information refer to Table 3.*

The average percentage *E. coli* concentration in the ≤5 µm interval exceeded 90 % of the total *E. coli* at all four sites (Figure 7). This finding was supported by ordinal logistic regression results, which showed that the <5 µm size interval had the strongest relationship (\(\alpha < 0.0001\)) with *E. coli* concentration. The results for Site #2 (40.6% developed land use) differ from the other sites as it had a higher *E. coli* concentration in the >60 µm interval than in the intermediate (5 µm< interval < 60 µm) interval (Figure 8). This result differs from the negative correlations (\(p < 0.05\)) between particle size and bacterial association typically reported in the literature [60,61]. In the current work, *E. coli* concentrations generally increased with decreasing SPM interval size, with lowest *E. coli* concentration associated with the >60 µm interval, while the highest concentration of *E. coli* were coincident with the <5 µm interval (Figure 7). This result is supported by previous work that reported negative correlations (\(p < 0.05\)) between particle size and bacterial association.
Of importance, and as noted earlier, greater concentrations in the smallest interval increase the likelihood that the bacteria will remain buoyant for longer time periods, thereby increasing the downstream extent of decreased microbial water quality [1,12].

Figure 7. Fractions of *E. coli* concentration (CFU) in three filtered size intervals, at four monitoring sites during the study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA.

Site #1 (54% forested) had the lowest SPM in the <5 µm interval among the sites (Figure 8). This is consistent with previous studies that showed that developed land use practices typically produce greater quantities of finer (<5 µm) particles relative to other land uses [27], and forested areas originate fewer suspended materials [62–64]. Site #3 (49.1% forested) generally had increased SPM and *E. coli* concentrations in the intermediate (5 µm < interval < 60 µm), <5 µm, and total intervals (Figure 9). Although this site was predominantly forested, it included agricultural land use (42.1%) (including livestock manure stacks), which may account for the elevated SPM and *E. coli* levels [36,65] given the percentage of agricultural land use has been
reported to be significantly correlated (p < 0.04) with *E. coli* concentrations in receiving waters [35]. Site #4 (82.4% forested; control site) had the lowest average *E. coli* concentration in the < 5 µm interval and the second lowest average total *E. coli* concentration. The low *E. coli* concentrations at Site #4 were expected, given that previous studies reported negative correlations (p < 0.01) between *E. coli* concentration and forested land use practices [56]. Decreased *E. coli* in receiving waters from forested locations is usually attributed to decreased endotherm population density relative to agricultural areas [58], lack of artificial sources of *E. coli* (e.g., water infrastructure in urban areas) [65], and decreased run-off compared to urban areas [65]. The decreased *E. coli* and SPM (Figure 6 and Figure 8) from forested areas indicates that the association of *E. coli* with SPM might not be as concerning in forested areas relative to other land use types.

At Site #2, in the >60 µm interval, there was an increase in *E. coli* concentration and a simultaneous decrease in SPM concentration relative to the other sites (Figure 8). This contradicts the negative correlation between *E. coli* concentration and SPM size reported by previous work [60,61]. However, these results support Figure 7, which shows the apparent preferential association of *E. coli* to SPM in the >60 µm interval relative to the 5 µm to 60 µm interval in developed areas. It is conceivable that there may be an unknown variable that altered the SPM size interval that the *E. coli* at Site #2 preferentially associated with in the current work. This result must be interpreted with caution, however, given differences in sampling regime between this and other studies (as noted earlier). Additionally, given the lack of previous studies, further investigation is needed to validate (or refute) these findings [1].
Figure 8. Suspended particulate matter (SPM) and *E. coli* concentration percent difference, separated into filtration intervals (<5 μm; 5 μm < interval < 60 μm; and >60 μm); and total, at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, Morgantown, West Virginia, USA.

4.2. Non-Parametric Statistical Analysis

The significant relationship between *E. coli* and SPM discovered using ordinal logistic regression corresponds well with the results from previous work [10,11], which also reported strong relationships between *E. coli* concentrations and SPM. The relationship between *E. coli* and SPM has been shown to be a function of physical processes including runoff events that can influence respective concentrations or it can indicate a potential preferential association of *E. coli* to SPM. However, differentiating between these two possibilities was beyond the scope of the current investigation. The results from ordinal logistic regression indicate no statistically significant (95% confidence interval) relationship between land use practices and *E. coli* concentrations. Therefore, land use practices alone could not be used to predict *E. coli* concentrations in receiving waters in the current investigation. The results from the current work
emphasize that additional factors that influence *E. coli* concentrations (i.e., water temperature, pH, and geochemistry) should be addressed in future work.

Increasing SPM concentrations in the intermediate (5 µm < interval < 60 µm) or large (>60 µm) intervals did not result in similar increases to the corresponding *E. coli* concentration (Figure 9). However, increased SPM in the smallest interval (<5 µm) had corresponding increases in the *E. coli* concentration in this interval. This result was attributed to bacteria becoming predominantly attached to, and subsequently transported with, SPM in the smallest interval (<5 µm), a relationship shown preliminarily in previous investigations [66–68]. This relationship may also be a function of the similar transport physics of *E. coli* and SPM particles <5 µm in size, due to similar sizes and buoyancy [69]. In general, this relationship supports the likelihood of *E. coli* remaining in suspension for longer time periods, thereby increasing the stream areas affected by fecal contamination.
Principle component analysis (PCA) showed that the components that account for the maximal variance within a given data set can be identified through the computation of multiple principal components and their respective eigenvalues [70]. Eigenvalues represent the variance of the data in a given direction, therefore components with the highest eigenvalues are principal components [70]. However, most data cannot be well described by a single principal component.
Therefore, multiple principal components are typically computed and ranked based on their eigenvalues and displayed visually with biplots [70]. In the current work, principle component biplots showed a distinct spatial distribution of study sites along principal components 1 and 2 for both the smallest interval (<5 µm) and the total data set (defined as all he SPM and E. coli data >0.7 µm) (Figure 10). Both biplots highlight the grouping of each of the sites within the idealized vector space defined by principal components 1 and 2. Given the similarity between the sites in terms of geology, topography, and climate, and their close proximity to each other, it can be concluded that the varying land use practices are the principal factors influencing the grouping of the data observed in both biplots [34]. The most obvious patterns are the strong correlation between E. coli concentration and SPM concentration in both the smallest interval and total data set (Figure 10). Ultimately, the percentage of agriculture land use was most closely associated with both E. coli and SPM concentrations, particularly in the smallest interval, attributable to reasons presented earlier [53,55,56]. This result further emphasizes that agricultural land use practices are strongly correlated with not only the occurrence of fecal contamination [35], but potentially the persistence of fecal microbes in receiving water. Conversely, forested and developed land uses were not as closely correlated with E. coli or SPM concentrations (Figure 10). Ultimately, PCA analysis (Table 4) and biplot results (Figure 10) effectively illustrate spatially the relationships between E. coli concentration, SPM and land use practices.
Figure 10. Results of principal components analysis, including biplots A) showing the data in the smallest interval (E. coli and SPM) and B) the total data set (defined as all he SPM and E. coli data >0.7 µm), for extracted principal components of E. coli concentration and SPM concentration at four monitoring sites during study period (20 July 2018–27 October 2018) in West Run Watershed, West Virginia, USA.

4.3. Study Considerations and Future Directions

It is acknowledged that other variables likely affect the lifecycle of E. coli (e.g., hydrology and climate) or the association of E. coli to SPM (e.g., aquatic geochemistry) [71,72]. Therefore, future work should expand on the results of this investigation by identifying and accounting for these variables. Regression analyses could be a useful tool in establishing the effect of the identified additional variables (e.g., hydrology, climate and aquatic geochemistry) [73]. Future work should also include the use of serial dilutions to avoid sample saturation during the quantification of E. coli concentrations. This would improve the accuracy of the results, specifically in areas comprising frequent elevated fecal pollution.

The current work used a study design and sampling regime that was focused on SPM size class and E. coli, which was dissimilar to previous work. The study design created challenges in comparing the effects of varying land use practices on E. coli or SPM, respectively. Therefore, combining the filtration methodology of the current work with a more traditional sampling regime during future work (i.e., regular temporal and spatial sampling) may facilitate more robust analysis
of the influence of different land use practices on the relationship between *E. coli* and SPM of various size intervals. This is particularly important given the strong relationships between *E. coli* concentrations and land use practices identified in previous work [35] and the influence of land use practices on both *E. coli* and SPM concentrations identified in the current investigation. Additionally, the study design applied in the current work can also be used to investigate the association of other constituents in receiving waters with SPM. For example, investigating the association of various microplastics to SPM, and perhaps *E. coli*, in receiving waters, may enhance the understanding of this relatively novel freshwater pollutant. Given the current general lack of process understanding of *E. coli* and emergent co- or in-dependent human population-induced pollutants, such as microplastics, in freshwater sources [74–76], and the potentially harmful effects of microplastics [74,77,78], improving scientific understanding is critical from a water quality perspective.

5. Conclusions

The paucity of field-based research investigating the relationship between *E. coli* (fecal matter) concentration and SPM size distribution and the importance of this relationship, particularly for human health, policy makers, and water resource managers [23,79], provided the impetus for the current work. Similarly, the need to characterize the relationship between SPM size distribution, *E. coli* (fecal matter) and land use practices provided further motivation. A four-site study design (comprising 1st and 2nd order streams) was implemented to advance this understanding in a contemporary representative, mixed-use, urbanizing watershed in the Appalachian region of the eastern United States. The key results show a statistically significant correlation of SPM (α = 0.05 with *E. coli* concentration. Moreover, the importance of SPM in the <5 µm interval was also highlighted as 1) more than 90% of *E. coli* data were found in this smallest interval at all four sampling locations, 2) this interval also featured the strongest correlation with *E. coli* concentration data (α < 0.0001), and 3) increasing SPM concentrations in the ≤5 µm interval showed corresponding increases in relative *E. coli* concentration. These results generally imply that *E. coli* principally remains free floating or attaches to particles <5 µm in size in receiving waters. Principle component analysis results highlighted the influence of agricultural land use practices on both *E. coli* and SPM concentrations, thereby providing evidence for the potential influence of land use practices on bacterial association with SPM. The work elucidates the effects
of SPM and land use practices on *E. coli* concentrations in receiving waters and provides a valuable steppingstone for future research into microbial water quality and fecal pollution. The results from this work better inform policy makers and water resource managers concerned with microbial and fecal pollution in receiving waters, thereby aiding in decision making and the effective management of freshwater resources.

**Author Contributions:** For the current work author contributions were as follows: conceptualization J.A.H.; methodology J.A.H.; formal analysis F.P. and J.A.H.; investigation J.A.H. and F.P.; resources, J.A.H.; data curation J.A.H.; writing—original draft preparation J.A.H. and F.P.; writing—review and editing J.A.H. and F.P.; visualization, J.A.H. and Petersen; supervision, J.A.H.; project administration, J.A.H.; funding acquisition, J.A.H. The authors declare no conflict of interest for the current work.

**Funding:** This work was supported by the National Science Foundation under Award Number OIA-1458952, the USDA National Institute of Food and Agriculture, Hatch project accession number 1011536, and the West Virginia Agricultural and Forestry Experiment Station. Results presented may not reflect the views of the sponsors and no official endorsement should be inferred. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgments:** Special thanks are due to many scientists of the Interdisciplinary Hydrology Laboratory (https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart), and the Institute of Water Security and Science (https://iwss.wvu.edu/). The authors extend great appreciation to Ida Holásková of the Davis College of Agriculture, Natural Resources and Design office of statistics for her expertise and guidance. The authors also appreciate the feedback of anonymous reviewers whose constructive comments improved the article.

**Conflicts of Interest:** The authors declare no conflict of interest for the current work.

**References**


25. Hubbart, J.A. Using sediment particle size class analysis to better understand urban land-use effects. *Int. J. Appl.* 2012, 2, 12-27.


69. Southard, J. *Introduction to Fluid Motions, Sediment Transport, and Current-Generated Sedimentary Structures Course Textbook*; MIT OpenCourseWare; Massachusetts Institute of Technology: Cambridge, MA, USA, 2006.


79. Pandey, P.K.; Soupir, M.L. Assessing linkages between E. coli levels in streambed sediment and overlying water in an agricultural watershed in Iowa during the first heavy rain event of the season. Trans. ASABE 2014, 57, 1571.

© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
CHAPTER 3: SPATIAL AND TEMPORAL CHARACTERIZATION OF *ESCHERICHIA COLI*, SUSPENDED PARTICULATE MATTER AND LAND USE PRACTICE RELATIONSHIPS IN A MIXED-LAND USE CONTEMPORARY WATERSHED


**Abstract:** Understanding land use practice induced increases in *Escherichia (E.) coli* and suspended particulate matter (SPM) concentrations is necessary to improve water quality. Weekly stream water samples were collected from 22 stream gauging sites with varying land use practices in a representative contemporary mixed-land use watershed of the eastern USA. Over the period of one annual year, *Escherichia (E.) coli* colony forming units (CFU per 100 mL) were compared to suspended particulate matter (SPM) concentrations (mg/L) and land use practices. Agricultural land use sub-catchments comprised elevated *E. coli* concentrations (avg. 560 CFU per 100 mL) compared to proximate mixed development (avg. 330 CFU per 100 mL) and forested (avg. 206 CFU per 100 mL) sub-catchments. Additionally, agricultural land use showed statistically significant relationships (*p* < 0.01) between annual *E. coli* and SPM concentration data. Quarterly PCA biplots displayed temporal variability in land use impacts on *E. coli* and SPM concentrations, with agricultural land use being closely correlated with both pollutants during Quarters 2 and 3 but not Quarters 1 and 4. The data collected during this investigation advance the understanding of land use impacts on fecal contamination in receiving waters, thereby informing land use managers on the best management practices to reduce exposure risks.

**Keywords:** *Escherichia coli*; suspended particulate matter; water quality; land use practices; experimental watershed

1. **Introduction**

Fecal pollution is the greatest contributor to water borne disease human morbidity and mortality rates globally [1]. Freshwater fecal pollution and subsequent increases in pathogenic bacteria (e.g., *Escherichia (E) coli*), cause disease outbreaks, including diarrhea, urinary tract
infections, respiratory illness and pneumonia [2,3]. The World Health Organization reported that 2.2 million deaths are caused by diarrhea annually, due to the consumption of fecal contaminated water [1]. An improved understanding of factors leading to increased fecal contamination in receiving waters will be useful in reducing outbreaks of waterborne disease and improving water quality. Furthermore, understanding the factors impacting the health and exposure risks of fecal pollution can be used to decrease the threat posed by fecal organisms. For example, the environmental persistence of fecal microbes can be extended when occurring with suspended particulate matter (SPM) [4,5]. Therefore, monitoring SPM in conjunction with fecal pollution can provide greater insight into water quality, through more accurate assessments of the persistence of fecal microbes.

Suspended particulate matter (SPM), defined as heterogeneous aggregates of mineral fragments, organic matter and microbial fractions, comprises the greatest water pollutant by volume globally [6]. Excess SPM in freshwater can impact water quality by decreasing the amount of transmitted light, thereby restricting or eliminating the photosynthesis of aquatic plants and dramatically influencing the aquatic food chain [7,8]. Therefore, understanding the factors influencing fluxes in SPM (e.g., land use practices) is important from an ecosystem management perspective. Additionally, increases in SPM can clog the gills of fish, thus lowering resistance to disease and decreasing developmental growth rates [7] while also elevating water temperatures, thereby disrupting the metabolic processes of various aquatic biota [8]. Thus, changes in SPM concentrations in receiving waters can entail serious consequences for various aquatic organisms. SPM can also act as a conveyance system for other pollutants including heavy metals, chemicals and pathogens, including fecal microbes (as discussed above) [8–12]. Insofar as excess sediment can be harmful to aquatic ecosystems, too little sediment can also be harmful, leading to the scouring of river channels, erosion and reduced nutrient inputs [8,13]. Consequently, understanding the factors leading to increases or decreases of SPM in receiving waters is important from a water quality perspective [14].

Previous work investigating fecal contamination and SPM reported statically significant, Pearson’s product moment, correlations \( r > 0.9 \) between the two pollutants [15]. Moreover, the strength of the relationships between fecal pollution and SPM reported in previous work [15,16] has led to speculation that SPM concentrations (and the turbidity caused by SPM) can potentially serve as a proxy for fecal contamination [16]; however, this is yet to be verified. The relationship
between fecal contamination and SPM has been attributed to similar transport processes influencing both pollutants during run-off events [17], including similar in-stream transport physics [18], and the sorption of fecal microbes to SPM [4,5]. Additionally, certain land use practices can simultaneously elevate both SPM and fecal microbe concentrations in receiving waters [19,20]. For example, agricultural land use practices are commonly associated with increased fecal (e.g., *E. coli*) and SPM concentrations relative to other land use types [19–22]. This is often related to the presence of livestock [23], with livestock population density being correlated to fecal indicator organism concentration [24]. Manure application in agricultural areas has also been linked to increased concentrations of fecal microbes in receiving waters [4]. Conversely, agricultural practices such as soil tillage and soil exposure yield increased SPM concentrations in the receiving waters of agricultural areas [19,21]. Differing land use practices in a given area (watershed) can therefore account for differing *E. coli* and SPM concentrations in receiving waters and should thus be accounted for when monitoring these pollutants.

Despite the progress of previous research, knowledge gaps regarding the relationship between fecal contamination and SPM remain. For example, few studies investigated the relationships between SPM and fecal contamination in mixed land use settings. Furthermore, the majority of studies included limited sampling locations [20,25] and tended to occur in areas of similar land use types [16,26], or were controlled laboratory simulations [27]. Additionally, previous work on fecal contamination typically focused on storm events and therefore report disproportionately elevated fecal concentrations in receiving waters [28]. Clearly, knowledge regarding fecal concentrations outside of storm events is lacking, creating challenges for proper management practices. Similarly, previous work investigating fecal concentrations in receiving waters typically utilized shorter (weeks or months) sampling periods [20], which fail to account for the seasonal variability in land use practices. For example, seasonal variability in agricultural land use practices can lead to changes in fecal microbe concentrations in receiving water, particularly in areas where manure is applied [29]. Manure is typically only applied in specific seasons, thereby leading to corresponding increases in the fecal contamination of associated receiving waters during these seasons [29]. Therefore, a study design capable of distinguishing the effects of different land uses and seasonality would be useful for investigating fecal contamination and SPM in receiving waters and for better informing water resource managers.
Previous studies have used many different study design methods, including different sampling regimes, to advance the understanding of *E. coli* regimes. Study designs have included laboratory based designs comprising simulations [30] and field based designs comprising event based sampling [31], periodic sampling [32], stochastic sampling [33] and nested-scale experimental watersheds [20]. The nested-scale and paired experimental watershed study design is a method that has been successfully used to quantify the effects of land use practices on receiving waters in mixed land use settings [34–40]. Nested watershed study designs divide a larger watershed into a series of sub-catchments to investigate the influence of land use practices on the environmental variables of interest [35,37,41–43]. Sub-catchment delineation isolates different land use practices and hydrologic characteristics [41]. Paired watersheds comprise at least two physiographically similar watersheds (control and treatment) from which data are collected [41]. The study design enables the identification of the influence and cumulative effect of various land use practices on the response variable of interest through the quantification of the influencing processes observed at the sub-catchment scale [44]. Therefore, the approach allows for the effective disentanglement of factors (e.g., land use practices and SPM) that influence a given response variable of interest (e.g., fecal microbe concentration), thus providing quantitative information regarding hydrologic and water quality regimes related to specific land-uses [41]. Given its proven application in over a century of studies, the nested-scale and paired experimental watershed study design is an accepted optimal study design for investigating current knowledge gaps regarding fecal contamination, SPM and land use practices.

The Appalachian region of the USA is well-suited for researching knowledge gaps concerning fecal contamination, SPM and land use practices. The region is representative of many locations globally given that it suffers from widespread, frequent, and problematic fecal pollution [45]. Additionally, Appalachia is physiographically diverse, encompassing distinct Northern, Central and Southern regions, consisting of dissimilar geographic, climatological, and ecological characteristics [46]. For example, the temperate climate and well-distributed year-round rainfall characteristics of Central Appalachia [47] are similar to those of areas such as Uruguay or Southern Brazil [48], and many other locations. Conceivably, other temperate areas comprising year-round precipitation (e.g., Uruguay) will benefit from research conducted in the Central Appalachian region as the results will be comparable and transferable. Furthermore, water quality is a primary concern in rural Appalachia as thousands of residents are exposed to water quality problems,
specifically regarding microbial contamination [49]. Water quality problems in rural areas are exacerbated by inadequate wastewater treatment infrastructure, isolation due to geographically inaccessible terrain, and poverty [49]. Consequently, water quality is a primary concern, and insight into both SPM and fecal pollution (e.g., \textit{E. coli} concentrations) is necessary to effectively inform policy makers and water resource managers regarding water quality to make the best management practice decisions in Appalachia and physiographically similar locations globally.

The overarching objective of the current investigation was to quantify fecal contamination (\textit{E. coli} concentration) and SPM concentrations in receiving waters relative to differing land use practices from numerous sites in a mixed-land use contemporary watershed of Appalachia. Sub-objectives included 1) investigating the relationship between fecal microbe concentration (\textit{E. coli} colony forming units) and SPM concentrations in receiving waters, and 2) determining the influence of quarterly (seasonal) changes on the relationship between \textit{E. coli} and SPM. The study outcomes were to improve the understanding of the influence of land use practices on both fecal contamination and SPM pollution, providing land use managers with insight into factors influencing water quality in receiving waters.

2. Methods

2.1. Study Site Description

This research took place in West Run Watershed (WRW) a 23 km\(^2\) mixed-land use urbanizing watershed located in Morgantown, West Virginia, USA. West Virginia’s climate varies between cold and humid with warm summers, to temperate and humid with hot summers [50]. In Morgantown, WV, located in Monongalia County (and including the WRW), the climate is characterized by the lack of a dry season, cold winters (mean monthly temperature <0 °C) and warm-to-hot summers (mean monthly temperature >22 °C) [50]. Historically (1981–2010), Morgantown received approximately 1060 mm of average annual precipitation, with the coldest (January) and driest (February) months having an average daily temperature of −0.4 °C and an average monthly precipitation of 66 mm, respectively [51]. Conversely, the warmest and wettest month (July) comprised an average daily temperature of approximately 23 °C and an average monthly precipitation of 117 mm [51].

West Run Creek, the primary drainage of WRW, is a third order tributary of the Monongahela River, and includes many land use practices including agriculture, urban and forested areas [22].
Based on the 2016 National Agriculture Imagery Program (NAIP) land use and land cover data, WRW includes 42.7% forested land use, 37.7% mixed development (urban and commercial areas) land use and 19.4% agricultural land use practices. West Run Creek is a narrow, moderately entrenched stream with multiple small floodplains [20,52]. The elevation of the headwaters of WRW is 420 m above mean sea level [22]. Conversely, the elevation of the confluence of WRW with the Monongahela River is 240 m above mean sea level [20]. The watershed includes relatively rugged terrain, featuring numerous Paleozoic era rock outcroppings [20]. The headwaters of WRW contain the most recent geological formation (Monongahela series) [20]. Two coal formations are also present in the watershed, namely the Upper Kittanning coal and the Pittsburg coal seam [20]. Historic mining of the Pittsburg coal seam negatively impacted water quality in WRW, particularly in the headwaters [53].

A nested-scale and paired experimental watershed study design [35,44,54–56] comprising twenty-two study sites (i.e., gauge sites) was implemented in 2017. Sampling sites (numbered in downstream order) were located in West Run Creek (#3, #4, #6, #10, #13, #18, #19, #21 and #22) and its first and second order confluence tributaries (#1, #2, #5, #7, #8, #9, #11, #12, #14, #15, #16, #17 and #20) and included varying land use practices (Table 5; Figure 10). Both field surveys and GIS were used to identify the study sites and related sub-catchments. At the time of this investigation, forested land use was the predominant land use in WRW, accounting for 42.7% of the total land use practices in the watershed. Additionally, forested land use was the predominant land use type in all sub-catchments except #1, #11, #15, #16 and #20. Sub-catchments #1, #15 and #20 were primarily mixed development, whereas sub-catchments #11 and #16 where primarily agricultural (Table 5). Conversely, 85.84% of sub-catchment #17 was forested land use, the highest among the sub-catchments. This sub-catchment, which served as a reference sub-catchment (control) for the current work, also comprised 9.4% agricultural and 4.8% mixed development land use practices. Sub-catchment #17 is, therefore, considerably different to sub-catchment #12 (34.5% forested, 33.7% agriculture and 31.7% agriculture) despite both comprising predominantly forested land use practices. In general, at the time of the investigation, mixed development comprised the second largest percentage of land use practices (37.7%) and agricultural land use practices accounted for the lowest percentage of land use practices (19.4%) in WRW.
Table 5. Land use/land cover characteristics (% cover) and total drainage area (km$^2$) at 22 monitoring sites in West Run Watershed (WRW), West Virginia, USA. Note: land use percentages may not sum to 100%, as not every category is included (i.e., wetland, open water, etc.) and some categories are combinations of others (e.g., mixed development = urban + residential). Final row (Site #22) indicates the total values for the entire watershed.

<table>
<thead>
<tr>
<th>Site</th>
<th>Mixed Development (%)</th>
<th>Agriculture (%)</th>
<th>Forested (%)</th>
<th>Drainage Area (km$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>53.23%</td>
<td>38.70%</td>
<td>8.07%</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>13.58%</td>
<td>12.20%</td>
<td>74.21%</td>
<td>0.29</td>
</tr>
<tr>
<td>3</td>
<td>22.35%</td>
<td>16.17%</td>
<td>61.32%</td>
<td>1.87</td>
</tr>
<tr>
<td>4</td>
<td>25.88%</td>
<td>14.91%</td>
<td>59.00%</td>
<td>2.48</td>
</tr>
<tr>
<td>5</td>
<td>23.35%</td>
<td>25.51%</td>
<td>51.14%</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>23.91%</td>
<td>17.25%</td>
<td>58.70%</td>
<td>3.72</td>
</tr>
<tr>
<td>7</td>
<td>16.33%</td>
<td>28.60%</td>
<td>54.91%</td>
<td>0.78</td>
</tr>
<tr>
<td>8</td>
<td>30.78%</td>
<td>16.47%</td>
<td>52.35%</td>
<td>1.55</td>
</tr>
<tr>
<td>9</td>
<td>27.57%</td>
<td>19.33%</td>
<td>52.84%</td>
<td>2.29</td>
</tr>
<tr>
<td>10</td>
<td>24.92%</td>
<td>18.40%</td>
<td>56.49%</td>
<td>6.18</td>
</tr>
<tr>
<td>11</td>
<td>18.15%</td>
<td>41.87%</td>
<td>39.16%</td>
<td>1.75</td>
</tr>
<tr>
<td>12</td>
<td>31.77%</td>
<td>33.72%</td>
<td>34.51%</td>
<td>1.75</td>
</tr>
<tr>
<td>13</td>
<td>26.83%</td>
<td>25.77%</td>
<td>47.15%</td>
<td>10.53</td>
</tr>
<tr>
<td>14</td>
<td>16.19%</td>
<td>26.43%</td>
<td>56.92%</td>
<td>3.36</td>
</tr>
<tr>
<td>15</td>
<td>70.28%</td>
<td>10.31%</td>
<td>19.42%</td>
<td>0.98</td>
</tr>
<tr>
<td>16</td>
<td>5.38%</td>
<td>58.72%</td>
<td>35.16%</td>
<td>0.25</td>
</tr>
<tr>
<td>17</td>
<td>4.78%</td>
<td>9.38%</td>
<td>85.84%</td>
<td>0.75</td>
</tr>
<tr>
<td>18</td>
<td>25.98%</td>
<td>24.88%</td>
<td>48.86%</td>
<td>16.41</td>
</tr>
<tr>
<td>19</td>
<td>29.45%</td>
<td>22.45%</td>
<td>47.85%</td>
<td>18.88</td>
</tr>
<tr>
<td>20</td>
<td>89.16%</td>
<td>4.19%</td>
<td>6.61%</td>
<td>3.42</td>
</tr>
<tr>
<td>21</td>
<td>38.10%</td>
<td>19.46%</td>
<td>42.23%</td>
<td>22.93</td>
</tr>
<tr>
<td>22</td>
<td>37.71%</td>
<td>19.38%</td>
<td>42.66%</td>
<td>23.24</td>
</tr>
</tbody>
</table>
Figure 11. Monitoring/sampling locations for the current investigation, with land use/land cover, in West Run Watershed, Morgantown, West Virginia, USA.

2.2. Data Collection

Climate data collected for the current work included precipitation (Campbell Scientific TE525 Tipping Bucket Rain Gage), average air temperature, relative humidity (Campbell Scientific HC2S3 Temperature and Relative Humidity Probe), and average wind speed (Campbell Scientific Met One 034B Wind Set instrument). Data were recorded at a 3 m height during the study period (2 January 2018–1 January 2019) by a climate station located within approximately 30 m of Site #13 (Figure 10).

For the current work, weekly water grab-samples were collected as per Petersen et al. [20], Hubbart et al. [57], Kellner and Hubbart [43], and Zeiger and Hubbart [42,58] from each monitoring site (stream order ≤ 3). Water sample collection was initiated at 09:00 at Site #1 and continued in numerical order of sites. Sites #9 and #10 were exceptions, as they were sampled before Sites #7 and #8, due to their location relative to other sites (Figure 10). The proximity of Sites #9 and #10 to Site #6 meant that overall sampling time was reduced by sampling them after Site #6, increasing the comparability of the samples during sample processing. The sampling
period for the study (2 January 2018–1 January 2019, thus 53 weeks) was one calendar year to account for seasonal variability in the *E. coli* concentration and SPM data. Notably, the sampling period was longer than in typical studies on fecal contamination [59,60], allowing for a comprehensive quantification of fecal contamination (*E. coli*) regimes at sub-catchment mixed-land-use scales. The high-resolution study design resulted in a total of 1166 spatio-temporally delineated fecal contamination (*E. coli*) concentration and SPM concentration values.

Following collection, the samples were transported to the Interdisciplinary Hydrology Laboratory, located in the Davis College of Agriculture, Natural Resources and Design at West Virginia University, for analyses. In the laboratory, water samples were refrigerated (at 3.3 °C), and gravimetric analyses (vacuum filtration) were conducted as per the American Society for Testing and Materials, test number D 3977-97, [61] within a few days of collection to determine the mass of suspended particulate matter (SPM). Additionally, fecal contamination was quantified immediately upon arrival at the laboratory using *Escherichia (E) coli* as an indicator organism, as per previous work [20,62]. *E. coli* coliform forming units (CFU) were enumerated using the U.S. Environmental Protection Agency (EPA) approved Colilert test [63], developed by IDEXX Laboratories Inc. The applied method used an MPN approach to estimate the *E. coli* CFU concentration; therefore, *E. coli* concentration data were referred to as CFU, not MPN, during the investigation. The test, included in Standard Methods for Examination of Water and Wastewater was developed to estimate fecal concentrations in water samples without requiring sample dilution [63,64]. A combination of Colilert’s Defined Substrate Technology nutrient-indicator (ONPG), and a selectively suppressing formulated matrix created low chances of recording inaccurate results (chance of reporting false positives ±10%). With this test, most non-target organisms are unable to grow given that they lack the enzyme to metabolize the provided carbon source (ONPG) [63]. The formulated matrix selectively suppresses the few non-target organisms that can metabolize ONPG [63]. The number of *E. coli* colony-forming units (CFU) per 100 mL of sampled water was estimated using the Quanti-Tray system, comprising 96 total wells: 48 large wells (49, including the overflow well) and 48 small wells [63]. The Colilert (ONPG) substrate was added to 100 mL of sampled water, sealed in the Quanti-Tray, and incubated at 35 °C for 24 hours, as per Colilert’s instructions [29]. Following incubation, fluorescing (positive for *E. coli*) wells were quantified using a UV light and converted, with a 95% confidence interval, into a concentration of *E. coli* (CFU per 100 mL) using the Quanti-
Tray Most Probable Number (MPN) table. The \textit{E. coli} concentration range resultant from the Quanti-Tray/MPN table method was \(<1\) to \(10^{11.2}\) CFU. Therefore, \textit{E. coli} concentrations in excess of \(10^{11}\) CFU per 100 mL could not be accurately estimated. This limitation was an allowable shortcoming of the current work given the focus on the weekly detection of small \textit{E. coli} concentrations occurring between storm events.

2.3. \textit{Data Analysis}

Descriptive statistics were generated for \textit{E. coli} and SPM concentrations and aggregated for the study period. Average percentage differences between sites were determined by comparing the average SPMs and average \textit{E. coli} concentrations between sites. Statistical analyses were conducted using Origin Academic 2018 (OriginLab Corporation). Normality testing was completed using the Anderson Darling Test \cite{65}. Land use practices were reclassified (lumped) into three major categories prior to analysis, namely mixed development, agriculture, and forested \cite{20}. Mixed development constituted roads, impervious surfaces, mixed developments and barren areas. Agriculture included low vegetation, hay pasture and cultivated crops. Forested land use included mine grass, forest, mixed mesophytic forest, dry mesic oak forest, dry oak (pine) forest and small stream riparian habitats. Annual data were also analyzed in four quarter data subsets, comprising all weekly samples collected in three month blocks starting on January 1st, 2018, to analyze seasonal variation. Thus, Quarter 1 included 2 January 2018–27 March 2018 (winter), Quarter 2 included 3 April 2018–26 June 2018 (Spring), Quarter 3 included 3 July 2018–25 September 2018 (summer), and Quarter 4 included 2 October 2018–1 January 2019 (fall). Spearman correlation tests, with a significance threshold of \(\alpha = 0.05\) \cite{66}, were used to analyze the relationship between \textit{E. coli} concentration, suspended sediment, and land use practices at all twenty-two sites, as per Petersen et al. \cite{20} for the complete annual data set and the four quarterly data subsets. Finally, principal component analysis (PCA) was used to investigate the relationships between \textit{E. coli} concentrations, SPM and land use practices (presented in biplots) across all 22 sampling locations for the annual data set and the four quarterly data subsets.
3. Results and Discussion

3.1. Climate during Study

Total precipitation was 1378 mm in 2018 in WRW. This was approximately 20% more precipitation than the historic annual average (1096 mm) dating back to 2007 [67]. September (186 mm) and October (47 mm) were the wettest and driest months, respectively, during 2018 (Figure 12). Approximately 14% of the annual precipitation was received in September. This was more than double the historic average precipitation (80 mm) for that month [67]. The average air temperature, during the study period was approximately 12 °C, which is close to the historic average of 11 °C [67]. July (22 °C) and January (−4 °C) comprised the warmest and coldest average monthly temperatures, respectively, in WRW during 2018. Relative humidity was characteristically high during 2018 (Figure 12), comprising a yearly average of 76%. Generally, climate during the period of study (2 January 2018–1 January 2019) was predictably variable and consistent with historic trends (Figure 12), including humid and warm weather during the summer months, with temperatures decreasing over the transition to winter (Figure 12). As is typical of the region, there was no dry season; however, large precipitation events during Quarters 2 (spring; e.g., May 6th: 24 mm) and 3 (summer; e.g., September 9th: 60 mm) resulted in greater quarterly (seasonal) variation in precipitation (Figure 2) [67]. Quarters 2 and 3 (spring and summer; 850 mm) therefore received 67% more precipitation than Quarters 1 and 4 (winter and fall; 510 mm).
Figure 12. Thirty-minute time series of climate variables during the study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: stream stage was measured in the primary stream of WRW, West Run Creek, within approximately 30 m of Site #13 and West Run Creek.

3.2. Annual Suspended Particulate Matter, E. coli Concentrations and Land Use Practices

The results showed that forested sub-catchments had the highest average (Site #7; 55% forested; 78.4 mg/L), maximum (Site #9; 53% forested; 1140 mg/L) and minimum (Site #8; 52% forested; 12.7 mg/L) SPM concentrations (Table 6; Figure 13). Notably, these sub-catchments constituted one of the paired watersheds of the paired study design and were in close proximity to each other (Figure 10). Consequently, these sub-catchments were subject to similar land use activities and processes leading to elevated SPM in this region of the watershed. For example, the agricultural land use practices in the headwaters of sub-catchments #7 and #8 (Figure 10) could have elevated the SPM in the entire paired catchment (Sites #7, #8 and #9) as previous work has reported increased SPM in agricultural areas [19–22]. SPM concentrations were also elevated in West Run Creek (combined average of sites in West Run Creek, 39 mg/L) relative to sites located in the first and second order confluence tributaries (combined average, 35 mg/L) (Table 6; Figure 13). The increased SPM in West Run Creek was attributable to (1) the greater volumetric
streamflow in West Run Creek relative to in its tributaries, as increased streamflow can increase the SPM concentration [68]; and (2) increased SPM sources due to an increased drainage area relative to its tributaries (23 km² and 15 km² respectively) (Table 5; Figure 10). Conversely, SPM concentrations were decreased in mixed development sub-catchments (Sites #15: 70% mixed development and #20: 89% mixed development) comprising the lowest average (6.5 mg/L), lowest median (1 mg/L), and lowest minimum (0 mg/L) (Table 6; Figure 13). Site #15 also had low SPM concentrations during previous work conducted in the WRW, thereby supporting the results from the current investigation [20]. Mixed development areas can comprise decreased exposed soil surfaces and subsequent reductions in the SPM sources relative to other land use types, which can account for the decreased SPM concentrations [69].

Table 6. Descriptive statistics of suspended particulate matter (mg L⁻¹) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: all average values presented in the current work constitute arithmetic means.
Figure 13. Box and whisker plots of suspended particulate matter (mg L$^{-1}$; log$^{10}$ scale) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA. Boxes delineate 25th and 75th percentiles; lines denote medians; squares show means; whiskers describe 10th and 90th percentiles; x shows maxima and minima when above and below, respectively. Note: different box colors represent data from different sites.

The study results showed that *E. coli* concentrations were the highest at sub-catchments comprising the greatest percentage agricultural land use area (Site #16: 59% agricultural). These results are similar to those of previous investigations in WRW reporting increased *E. coli* concentrations in agricultural land use sub-catchments [20]. This predominantly agricultural sub-catchment comprised the highest average (560 CFU per 100 mL) and median (575 CFU per 100 mL) (Table 7; Figure 14) *E. coli* concentrations over the period of investigation. Previous investigations in the USA (California and Ohio) reported increased fecal contamination with agricultural land use practices [31,70], and a significant correlation (p < 0.04) between agricultural land use and *E. coli* concentrations [20], thereby supporting the results recorded during the current
work. The lowest \textit{E. coli} concentrations were recorded at two forested sites (Site #2: 74% forested and Site #5: 51% forested) comprising the lowest median (3 CFU per 100 mL) and average (34 CFU per 100 mL) amongst the sites respectively (Table 7; Figure 14). These two sites, and the forested sub-catchments comprising one of the paired watersheds (Sites #7, #8 and #9) located in the headwaters of WRW (Figure 10), were heavily impacted by acid mine drainage (AMD) from historic mining activities [20,22,53], which likely, at least in part, explains the low \textit{E. coli} concentrations observed at these sites. This is an important finding given that previous studies showed that AMD lowers the pH of receiving waters [71] and that the current results indicate that AMD may also lower \textit{E. coli} concentrations. Forested sites generally had lower \textit{E. coli} concentrations (e.g., Site #17: 86% forested; average \textit{E. coli} concentration: 206 CFU per 100 mL) during the study period than sites comprising other land use practices (e.g., Site #20: 89% mixed development; average \textit{E. coli} concentration: 415 CFU per 100 mL) (Figure 14). These results align well with previous studies reporting decreased fecal contamination in forested areas [70] and are attributable to the increased quality of receiving waters in forested areas [72]. Consequently, both forested land use practices and AMD lowered \textit{E. coli} concentrations in WRW during the investigation. Notably, the low average \textit{E. coli} concentrations recorded during the study period (2 January 2018–1 January 2019), specifically in the headwaters, affirms the study objective of analyzing samples collected between storm events that comprise lower \textit{E. coli} concentrations. Additionally, in no other study has there been such high spatial and temporal resolution sampling over a full annual year. This allowed for a more comprehensive analysis of \textit{E. coli} concentration regimes and relationships with SPM and land use, including accounting for seasonality, than is available in the literature surrounding contemporary mixed land use watersheds. The current study therefore lends greatly needed confirmation through high spatial and temporal resolution of previous studies.
Table 7. Descriptive statistics of *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>170</td>
<td>66</td>
<td>0</td>
<td>1011</td>
<td>251</td>
</tr>
<tr>
<td>#2</td>
<td>38</td>
<td>3</td>
<td>0</td>
<td>961</td>
<td>139</td>
</tr>
<tr>
<td>#3</td>
<td>397</td>
<td>260</td>
<td>15</td>
<td>1011</td>
<td>315</td>
</tr>
<tr>
<td>#4</td>
<td>429</td>
<td>361</td>
<td>107</td>
<td>1011</td>
<td>249</td>
</tr>
<tr>
<td>#5</td>
<td>34</td>
<td>4</td>
<td>0</td>
<td>914</td>
<td>129</td>
</tr>
<tr>
<td>#6</td>
<td>269</td>
<td>194</td>
<td>2</td>
<td>1011</td>
<td>276</td>
</tr>
<tr>
<td>#7</td>
<td>84</td>
<td>20</td>
<td>0</td>
<td>1011</td>
<td>179</td>
</tr>
<tr>
<td>#8</td>
<td>89</td>
<td>32</td>
<td>0</td>
<td>1011</td>
<td>180</td>
</tr>
<tr>
<td>#9</td>
<td>127</td>
<td>25</td>
<td>0</td>
<td>1011</td>
<td>241</td>
</tr>
<tr>
<td>#10</td>
<td>210</td>
<td>93</td>
<td>3</td>
<td>870</td>
<td>273</td>
</tr>
<tr>
<td>#11</td>
<td>98</td>
<td>16</td>
<td>0</td>
<td>3</td>
<td>202</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#12</td>
<td>234</td>
<td>88</td>
<td>0</td>
<td>1011</td>
<td>305</td>
</tr>
<tr>
<td>#13</td>
<td>215</td>
<td>91</td>
<td>0</td>
<td>1011</td>
<td>266</td>
</tr>
<tr>
<td>#14</td>
<td>457</td>
<td>299</td>
<td>0</td>
<td>1011</td>
<td>406</td>
</tr>
<tr>
<td>#15</td>
<td>330</td>
<td>211</td>
<td>5</td>
<td>1011</td>
<td>293</td>
</tr>
<tr>
<td>#16</td>
<td>560</td>
<td>575</td>
<td>22</td>
<td>1011</td>
<td>373</td>
</tr>
<tr>
<td>#17</td>
<td>206</td>
<td>93</td>
<td>3</td>
<td>1011</td>
<td>288</td>
</tr>
<tr>
<td>#18</td>
<td>324</td>
<td>218</td>
<td>0</td>
<td>1011</td>
<td>342</td>
</tr>
<tr>
<td>#19</td>
<td>466</td>
<td>436</td>
<td>1</td>
<td>1011</td>
<td>339</td>
</tr>
<tr>
<td>#20</td>
<td>415</td>
<td>299</td>
<td>0</td>
<td>1011</td>
<td>340</td>
</tr>
<tr>
<td>#21</td>
<td>471</td>
<td>397</td>
<td>2</td>
<td>1011</td>
<td>342</td>
</tr>
<tr>
<td>#22</td>
<td>452</td>
<td>397</td>
<td>3</td>
<td>1011</td>
<td>345</td>
</tr>
</tbody>
</table>
Figure 14. Box and whisker plot of \textit{E. coli} concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA. Boxes delineate 25th and 75th percentiles; lines denote medians; squares show means; whiskers describe 10th and 90th percentiles; x shows maxima and minima when above and below, respectively. Note: different box colors represent data from different sites.

\textit{E. coli} concentrations showed a general increase from the headwaters of WRW to the confluence of the Monongahela River, with larger average concentrations typically being observed in the lower portions of the watershed (Figure 15). In the current work, AMD may account for the lower \textit{E. coli} concentrations in the upper watershed (as discussed above). However, in the lower elevations of WRW, land use practices may be the predominant factor influencing \textit{E. coli} concentrations. For example, in West Run Creek (Sites #13–#21), there was a notable increase in cumulative \textit{E. coli} concentrations and a simultaneous increase in agricultural and mixed development land use practices (Figure 16). Previous work reported on the increased fecal contamination associated with increased agricultural and urban areas [31,70,73,74], commonly attributed to increased sources (livestock and manure) [23] and increased (concentrated flow) run-
off during precipitation events, respectively [73], and urban stream syndrome [75], thereby supporting the results from the current investigation. The inter-site relationship between *E. coli* concentrations and SPM was not clearly discernable based on average values (Figure 15) or cumulative values in West Run Creek (Figure 16), as increases in SPM were not always accompanied by similar increases or decreases in *E. coli* between the different sampling locations. A potential explanation for these results may be that SPM and *E. coli* concentrations are affected by different factors at different sites (e.g., geochemistry, land use and antecedent soil water conditions). Thus, in WRW, there may exist a spatial disconnect regarding the factors influencing *E. coli* concentrations and subsequently impacting the relationship between *E. coli* and SPM in the watershed. To account for the spatial disconnect, site specific analysis of *E. coli* and SPM correlations, including separate analysis for different time periods in the year (quarters), was required to improve the current understanding regarding this relationship and to assess the use of SPM as a proxy for fecal contamination.

Figure 15. Average suspended particulate matter (mg L\(^{-1}\)) and *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, West Virginia, USA.
3.3. Quarterly Suspended Particulate Matter, *E. coli* Concentrations and Land Use Practices

Average *E. coli* and SPM concentration data showed notable temporal variation during 2018, based on quarterly analysis (Figure 17). *E. coli* concentrations were elevated during Quarter 2 (spring; 3 April 2018–26 June 2018) and Quarter 3 (summer; 3 July 2018–25 September 2018) of 2018, coinciding with the warmer spring and summer months (average maximum daily temperatures: Quarters 2 and 3 = 26 °C; Quarters 1 and 4 = 9 °C) of the year and higher stream flows induced by larger and more frequent precipitation events (Figure 12). Previous work noted correlations between elevated air temperatures and *E. coli* concentrations in freshwater [74], thereby supporting the results from the current work reporting greater *E. coli* concentrations during warmer months. Furthermore, the second and third quarters included some of the largest precipitation events of 2018 (Figure 12). These precipitation events could, at least in part, account...
for the elevated *E. coli* concentrations recorded during this time period, as previous work linked precipitation events to elevated *E. coli* concentrations in receiving waters [32,74]. Conversely, the highest average SPM concentrations were recorded during Quarters 1 (winter) and 2 (spring), with average concentrations decreasing during the second half of the year (Figure 17). The high SPM recorded during this time period may be attributable to decreased vegetation cover throughout WRW, leading to increased exposed soil surfaces, owing to the seasonal changes in vegetation (i.e., many plant species senesce during the cold winter months) [76,77]. Once vegetation throughout WRW increases during Quarter 3 (summer), exposed soil surfaces and sources of SPM become more limited. Quarter 4 (fall) comprised low *E. coli* and SPM concentrations. During this time period, WRW received less precipitation than in the preceding quarters (Figure 12), which could have limited the transport of pollutants (i.e., *E. coli* and SPM) to receiving waters [78]. Additionally, *E. coli* concentrations in the receiving water could have been suppressed by the colder temperatures [79] and decreased nutrient availability owing to the drier antecedent conditions and greater infiltration [77,80].
Figure 17. Quarterly average *E. coli* concentration (CFU per 100 mL) and SPM concentration (mg/L) at each sampling location (n = 22) during the study period (1/2/18–1/1/19) in West Run Watershed, Morgantown, West Virginia, USA. Note: (A) represents Quarter 1 (winter: 2 January 2018–27 March 2018); (B) represents Quarter 2 (spring: 3 April 2018–26 June 2018); (C) represents Quarter 3 (summer: 3 July 2018–25 September 2018); (D) represents Quarter 4 (fall: 2 October 2018–1 January 2019).

3.4. Non-Parametric Statistical Results

Normality test results indicated that the *E. coli* concentration data were non-normally distributed, thus Spearman correlation coefficients (nonparametric version of the Pearson product moment correlation) were used to quantify the relationships between *E. coli* concentration, SPM concentration, and land use at each site. *E. coli* concentrations and SPM concentrations were not significantly correlated at all sites; however, nine of the 22 sites (Sites #4, #7, #8, #9, #11, #15, #16, #17 and #20) did have significant correlations (Table 8). Notably, Sites #7, #8 and #9, which had the highest SPM during the investigation (Table 6; Figure 13), displayed statistically significant positive correlations ($p < 0.01$ for all three sites) between *E. coli* concentrations and
SPM concentrations. Based on these relationships, SPM may serve as a relatively accurate proxy for *E. coli* concentrations in similarly physiographic catchments. Mixed development and forested sub-catchments did not display consistent significant correlations (*p* < 0.05) between *E. coli* and SPM concentrations. For example, Sites #15 (70.3% mixed development) and #17 (85.8% forested) both comprised statistically significant relationships between SPM and *E. coli* concentrations (*p* < 0.01 and *p* < 0.05, respectively) despite comprising different predominant land use practices. However, Site #1 (53.2% mixed development) and Site #5 (51.1 % forested) displayed statistically insignificant correlations, despite including approximately similar dominant land use practices to Sites #15 and #17. As discussed in the preceding sections, AMD in the headwaters of WRW could be affecting *E. coli* concentrations, thereby creating inconsistency in the correlations between the *E. coli* and SPM concentrations. Therefore, the *E. coli* and SPM correlation results imply a spatial disconnect in terms of the influence of land use practices—in particular, mixed development and forested areas—on the use of SPM as a proxy for fecal contamination. Notably, it seems likely, based on these results, that legacy effects (AMD) of historic land use practices (coal mining) may be impacting *E. coli* concentrations and affecting the observed relationships between *E. coli* and SPM concentrations. Conversely, both of the predominantly agricultural sites (Site #11 and #16) displayed significant correlations between SPM concentrations and *E. coli* concentrations, supporting previous work that reported elevated *E. coli* and SPM concentrations in the receiving waters of agricultural areas [19,20]. Therefore, in agricultural areas, SPM could be implemented as a proxy for *E. coli* with greater accuracy and less difficulty than in other land use areas.
Table 8. Results of Spearman’s correlation test, including annual *E. coli* concentration (CFU per 100 mL) and annual SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: bold values indicate significant correlations ($p < 0.05$).

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCC</td>
<td>0.19</td>
<td>0.18</td>
<td>0.04</td>
<td>0.30</td>
<td>−0.11</td>
<td>0.04</td>
<td>0.46</td>
<td>0.52</td>
<td>0.73</td>
<td>0.24</td>
<td>0.56</td>
</tr>
<tr>
<td><em>p</em>-value</td>
<td>0.17</td>
<td>0.19</td>
<td>0.77</td>
<td>0.03</td>
<td>0.42</td>
<td>0.78</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.09</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCC</td>
<td>0.00</td>
<td>0.13</td>
<td>−0.25</td>
<td>0.42</td>
<td>0.64</td>
<td>0.27</td>
<td>0.02</td>
<td>−0.14</td>
<td>0.55</td>
<td>0.01</td>
<td>0.17</td>
</tr>
<tr>
<td><em>p</em>-value</td>
<td>0.99</td>
<td>0.34</td>
<td>0.07</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.05</td>
<td>0.88</td>
<td>0.33</td>
<td>&lt;0.01</td>
<td>0.94</td>
<td>0.24</td>
</tr>
</tbody>
</table>

SCC = Spearman correlation coefficient.

Explanatory variables that account for the maximal variance in a data set can be identified via principle component analysis (PCA), through the computation of multiple principal components and their respective Eigenvalues [81]. Components comprising the highest Eigenvalues are assumed principal components, given that Eigenvalues represent the variance of the data in that direction [81]. A principle component is a linear function of the variables in an original data set that successively maximize variance and that are uncorrelated with each other [82]. Multiple principal components are typically calculated and ranked based on their Eigenvalues as most data cannot be well-described by a single principal component [81]. For the current work, the results showed three principal components with Eigenvalues exceeding 1 (an accepted threshold of importance [22,83]), including Principal Component 1 (Eigenvalue = 1.83), Principal Component 2 (Eigenvalue = 1.22) and Principal Component 3 (Eigenvalue = 1.17). These three principle components explained approximately 85% of the cumulative variance of the data set. Conversely, Principal Components 4 and 5 accounted for approximately 16% and 0% of the variance of the data set. For the current work, the principle component biplots showed distinct spatial distributions for study sites along Principal Components 1 and 2 (Figure 18). The idealized biplot vector space defined by Principal Components 1 and 2 is characterized by the grouping of the sites. Given these results, it can be concluded that land use practices are the primary factors influencing the grouping of the data in the biplot, given the similarity of the sites in terms of geology, topography and climate, and their close proximity to each other [52]. The strongest correlation illustrated by the biplot is between SPM concentration and agricultural land use practices (Figure 18). However, *E. coli* concentration is also closely related to both, attributable to agricultural land use practices (i.e.,
the rearing of livestock, manure application, soil tillage and increased exposed soil surfaces), as discussed in the preceding sections and reported by previous investigations [19–23]. Ultimately, the PCA results were analogous to the Spearman correlation coefficient results, indicating that SPM could potentially serve as proxy for *E. coli* in agricultural areas, especially during periods with lower levels of fecal contamination between storm events.

![Figure 18](image)

Figure 18. Results of principal component analysis, including biplots, for extracted principal components of annual *E. coli* concentration (CFU per 100 mL) and annual SPM concentration (mg/L) at 22 monitoring sites (indicated by the different colors) during the study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA.

The Spearman correlation test results based on the quarterly analysis of data indicated predictable temporal variation in the correlation between *E. coli* and SPM concentrations, with only Site #9 showing significant correlations (*p* < 0.05) throughout all four quarters (*p* < 0.01 Quarters 1 to 3; *p* = 0.02 Quarter 4) (Table 5). Quarter 1 displayed the most significant correlations (10; Site #2, #7, #8, #9, #10, #11, #12, #13, #15, and #18), whereas Quarter 4 had the fewest
significant correlations (two; Site #9 and #16) (Table 9). The temporal variation can be explained by the different impacts seasonal variation have on *E. coli* and SPM concentrations. For example, as discussed above, SPM concentrations will be influenced by changes in vegetation, with elevated concentrations typically occurring when vegetation cover decreases and decreasing as vegetation cover increases [76,77]. Therefore, elevated SPM concentrations can be expected during and immediately after the cold winter months, with decreased concentrations during the warmer summer months. Conversely, previous work linked elevated *E. coli* concentrations with warmer water (and air) temperatures [84–86]. Consequently *E. coli* concentrations can be expected to be elevated during the warmer summer months and decrease during the colder winter months.

Ultimately, there is a temporal (seasonal) difference between periods of elevated SPM and *E. coli* concentrations. For example, Quarter 3 included the highest average (424 CFU per 100 mL) *E. coli* concentrations across all 22 sampling locations, while Quarter 1 comprised the lowest average (187 CFU per 100 mL) *E. coli* concentration. Thus, between Quarter 1 and 3 there was more than a 100% increase in average *E. coli* concentrations across the 22 sampling sites. Conversely, Quarter 1 had the highest average (55.7 mg/L) SPM across all 22 sampling locations, while Quarter 4 comprised the lowest average (18.2 mg/L) SPM. Consequently, there was more than a 65% decrease in average SPM, across the 22 sampling locations, between Quarter 1 and Quarter 4. These temporal differences, driven by changes in precipitation [78], antecedent conditions [87], seasonal land cover [76] or land use practices [29] may account for the variable *E. coli* and SPM concentrations correlations identified in Table 9.

The varying correlations displayed in Table 9 constitute important results regarding the use of SPM as a proxy for *E. coli*. Temporal changes in correlation significance indicate that SPM cannot be a consistently accurate proxy for fecal contamination throughout all the quarters (seasons) of a year. Even agricultural sub-catchments displayed insignificant correlations during certain quarters (Quarter 3 and 4 for Site #11 and Quarter 1 for Site #16), despite being significantly correlated for the annual time period. These results differ from previous investigations that showed strong correlations between *E. coli* and SPM concentrations [15,16], particularly small SPM particles [22]. However, these studies did not comprise high frequency or sufficiently long sampling regimes that would allow for quarterly (seasonal) analysis. Therefore, the current work is among the first to include seasonal analysis of the relationship (correlation) between *E. coli* and SPM concentrations in receiving waters.

93
Table 9. Results of Spearman’s correlation test, including quarterly *E. coli* concentration (CFU per 100 mL) and quarterly SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: Quarter 1 represents 2 January 2018–27 March 2018; Quarter 2 represents 3 April 2018–26 June 2018; Quarter 3 represents 3 July 2018–25 September 2018; Quarter 4 represents 2 October 2018–1 January 2019. Bold values indicate significant correlations (p < 0.05).

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quarter 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>0.40</td>
<td>0.60</td>
<td>0.26</td>
<td>0.48</td>
<td>0.52</td>
<td>0.35</td>
<td>0.76</td>
<td>0.70</td>
<td>0.73</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>p-value</td>
<td>0.18</td>
<td>0.03</td>
<td>0.39</td>
<td>0.09</td>
<td>0.09</td>
<td>0.23</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>&lt;0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Quarter 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>0.19</td>
<td>0.18</td>
<td>0.03</td>
<td>0.26</td>
<td>-0.04</td>
<td>0.07</td>
<td>0.78</td>
<td>0.88</td>
<td>0.78</td>
<td>0.51</td>
<td>0.93</td>
</tr>
<tr>
<td>p-value</td>
<td>0.53</td>
<td>0.56</td>
<td>0.91</td>
<td>0.38</td>
<td>0.89</td>
<td>0.81</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.08</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><strong>Quarter 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>-0.05</td>
<td>-0.40</td>
<td>0.32</td>
<td>0.50</td>
<td>0.76</td>
<td>0.69</td>
<td>-0.27</td>
<td>0.58</td>
<td>0.79</td>
<td>0.25</td>
<td>0.54</td>
</tr>
<tr>
<td>p-value</td>
<td>0.88</td>
<td>0.17</td>
<td>0.29</td>
<td>0.08</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.37</td>
<td>0.04</td>
<td>&lt;0.01</td>
<td>0.42</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Quarter 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>-0.30</td>
<td>-0.32</td>
<td>-0.39</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.09</td>
<td>0.62</td>
<td>-0.22</td>
<td>0.51</td>
</tr>
<tr>
<td>p-value</td>
<td>0.31</td>
<td>0.27</td>
<td>0.17</td>
<td>0.89</td>
<td>0.95</td>
<td>0.98</td>
<td>0.78</td>
<td>0.75</td>
<td>0.02</td>
<td>0.45</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Similarly to in the annual results, Quarter 1 (winter) had three principal components with Eigenvalues exceeding 1 (Eigenvalues = 1.86, 1.35 and 1.20, respectively), which accounted for 88% of the cumulative variance in the data. Quarters 2 and 3 also included three principal components with Eigenvalues exceeding 1 (1.85, 1.29 and 1.15; and 1.88, 1.22 and 1.16, respectively). The principal components explained 86% of the variance in Quarter 2 and 85% of that in Quarter 3. Two principal components, accounting for 63% of the data variance, were identified for Quarter 4, comprising Eigenvalues of 1.95 and 1.21. The quarterly PCA results illustrated the correlation between agricultural land use and *E. coli* and SPM during Quarters 2 (spring) and 3 (summer) (Figure 19). The agricultural land use impacts on *E. coli* and SPM concentrations during these quarters were subjected to the largest precipitation events during the study period (Figure 12). Precipitation has been reported to exacerbate the impact of land use practices on receiving waters [19,74,88]; therefore, land use impacts (particularly agriculture) were
elevated during Quarters 2 (spring) and 3 (summer). The elevated impacts of land use practices during precipitation events are attributable to the increased transport of both *E. coli* and SPM during runoff events and subsequent increased *E. coli* and SPM concentrations in the receiving waters [88]. Conversely, during Quarters 1 (winter) and 4 (fall), none of the land use classes were closely correlated with either SPM or *E. coli* concentrations (Figure 19). Small and fewer precipitation events during Quarters 1 (winter) and 4 (fall) (Figure 12) may account for these results as fewer runoff events would lead to decreased concentrations of *E. coli* and SPM in the associated receiving waters [88]. The reduced precipitation during Quarters 1 and 4 would also have led to drier antecedent soil water conditions [89], leading to greater infiltration during subsequent precipitation events [90], further reducing the transport of both *E. coli* and SPM to the receiving waters. These results highlight the varied seasonality of land use impacts on *E. coli* and SPM concentrations and thus advance the science-based understanding of temporal fluctuations in *E. coli* concentration regimes.
Figure 19. Results of principal component analysis, including biplots, for extracted principal components of quarterly *E. coli* concentration (CFU per 100 mL) and quarterly SPM concentration (mg/L) at each sampling location (n = 22) during the study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Note: (A) represents Quarter 1 (winter: 2 January 2018–27 March 2018); (B) represents Quarter 2 (spring: 3 April 2018–26 June 2018); (C) represents Quarter 3 (summer: 3 July 2018–25 September 2018); (D) represents Quarter 4 (fall: 2 October 2018–1 January 2019).

### 3.5. Study Implications and Future Work

The scale-nested experimental watershed study design and the high spatial and temporal sampling period implemented during the current work allowed for the collection of a unique data set. PCA biplots illustrated the close correlation between agricultural land use practices and both *E. coli* and SPM concentrations, relative to mixed development and forested land use practices. Additionally, spatial and temporal variability in the significant correlations between *E. coli* and SPM concentrations indicated that SPM would not be a suitable proxy for fecal contamination.
The recorded lack of consistent *E. coli* and SPM relationships constitutes an important result for the development of accurate predictive fecal pollution models. The investigation emphasized the efficacy of the nested-scale experimental watershed study design to elucidate land use influences on fecal pollution in receiving waters. Future work should expand on the results from the current investigation by attempting to determine the precise tipping points associated with different land use practices’ influence on *E. coli* and SPM concentrations. Implementing a similar study design, as in the current work, in mixed land use watersheds not impacted by legacy land use impacts (e.g., AMD) could provide useful information regarding the precise tipping points for various land use practices. Furthermore, due to the results of the current investigation indicating that *E. coli* concentrations are not solely influenced by land use practices and given the previously reported influence of physicochemical parameters (e.g., pH and water temperature) on *E. coli* [84–86], future work should focus on the identification of additional variables (e.g., physicochemical and geochemical) influencing *E. coli* concentrations in receiving waters. The incorporation of a multi-year study period would also allow future work to expand on the results of this study to account for annual variations in climate.

4. Conclusions

A 22-site, nested-scale, experimental watershed study design was implemented to investigate *E. coli* concentrations in a mixed land use watershed in the Appalachian region of the eastern United States. Specific focus was given to the relationship between *E. coli* concentrations, SPM concentrations and land use practices, including an evaluation of the potential use of SPM concentrations as a proxy for *E. coli* concentrations. Agricultural land use sub-catchments comprised elevated *E. coli* concentrations (avg. 560 CFU per 100 mL) compared to adjacent mixed development (avg. 330 CFU per 100 mL) and forested (avg. 206 CFU per 100 mL) sub-catchments. Annual *E. coli* and SPM concentration data displayed a statically significant relationship (p < 0.01) in agricultural areas. However, quarterly SCC analysis highlighted fluctuations between significance (p < 0.05) and insignificance (p > 0.05) in the correlations between *E. coli* and SPM concentrations across all land use classes. Therefore, SPM lacked the consistent significant correlations with *E. coli* concentrations required to be a suitable proxy for fecal contamination. The annual PCA results illustrated the influence of agricultural land use practices on both *E. coli* and SPM concentrations, serving as validation for previous investigations,
which typically included less temporally and spatially robust sampling regimes. The quarterly PCA results highlighted the seasonal variability of land use impacts on both E. coli and SPM concentrations, with Quarters 2 and 3’s biplots displaying greater correlations between agricultural land use practices, E. coli and SPM concentrations than Quarters 1 and 2. Combined, Quarters 2 and 3 received 67% more precipitation (850 mm) than Quarters 1 and 2 (510 mm), accounting for the temporal variation in land use impacts depicted by the quarterly biplots. Ultimately, the current investigation advances the understanding of the influence of land use practices on E. coli and SPM concentrations, thereby contributing to the current understanding of fecal contamination regimes in contemporary mixed land use watersheds. The results better inform model builders, policy makers and land use managers regarding the factors influencing freshwater fecal contamination, thereby aiding in effective decision making and effective water quality management.

Author Contributions: For the current work author contributions were as follows: conceptualization, J.A.H.; methodology, J.A.H.; formal analysis, F.P. and J.A.H.; investigation, F.P. and J.A.H.; resources, J.A.H.; data curation, J.A.H.; writing—original draft preparation, F.P. and J.A.H.; writing—review and editing, F.P. and J.A.H.; visualization, F.P. and J.A.H.; supervision, J.A.H.; project administration, J.A.H.; funding acquisition, J.A.H.

Funding: This work was supported by the National Science Foundation under Award Number OIA-1458952, the USDA National Institute of Food and Agriculture, Hatch project accession number 1011536, and the West Virginia Agricultural and Forestry Experiment Station. Results presented may not reflect the views of the sponsors and no official endorsement should be inferred. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments: Special thanks are due to many scientists of the Interdisciplinary Hydrology Laboratory (https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart). The authors also appreciate the feedback of anonymous reviewers whose constructive comments improved the article.

Conflicts of Interest: The authors declare no conflict of interest for the current work.
References


CHAPTER 4: ADVANCING UNDERSTANDING OF LAND USE AND PHYSICOCHEMICAL IMPACTS ON FECAL CONTAMINATION IN MIXED-LAND-USE WATERSHEDS


**Abstract:** Understanding mixed-land-use practices and physicochemical influences on *Escherichia (E.) coli* concentrations is necessary to improve water quality management and human health. Weekly stream water samples and physicochemical data were collected from 22 stream gauging sites representing varying land use practices in a contemporary Appalachian watershed of the eastern USA. Over the period of one annual year, *Escherichia (E.) coli* colony forming units (CFU) per 100 mL were compared to physicochemical parameters and land use practices. Annual average *E. coli* concentration increased by approximately 112% from acid mine drainage (AMD) impacted headwaters to the lower reaches of the watershed (approximate averages of 177 CFU per 100 mL vs. 376 CFU per 100 mL, respectively). Significant Spearman’s correlations (p < 0.05) were identified from analyses of pH and *E. coli* concentration data representing 77% of sample sites; thus highlighting legacy effects of historic mining (AMD) on microbial water quality. A tipping point of 25–30% mixed development was identified as leading to significant (p < 0.05) negative correlations between chloride and *E. coli* concentrations. Study results advance understanding of land use and physicochemical impacts on fecal contamination in mixed-land-use watersheds, aiding in the implementation of effective water quality management practices and policies.

**Keywords:** *Escherichia coli*; physicochemistry; water quality; land use practices; experimental watershed

1. **Introduction**

Fecal microbes (e.g., *Escherichia (E.) coli*) are sources of waterborne pathogens and water contamination, causing substantial mortality and morbidity among human populations globally.
Outbreaks of diarrhea, urinary tract infections, respiratory illness, and pneumonia have been traced to increased fecal microbes (e.g., *E. coli*) in freshwater systems [3,4]. In 2018, the World Health Organization (WHO) reported that waterborne diarrheal diseases are the leading cause of mortality in the developing world, causing 2.2 million human deaths annually [5]. Obviously, understanding factors conducive to elevated fecal microbe (e.g., *E. coli*) concentrations in receiving waters is important from water quality and human health perspective. Improved understanding of factors favorable to fecal microbes can be used to inform land use managers in terms of how to effectively reduce fecal contamination, thereby improving water quality, freshwater security, and human health [6].

Fecal microbes in receiving waters can be variously impacted by physicochemical parameters. Previous investigations reported negative correlations between temperature, salinity, oxygen content, pH, and fecal microbe concentrations [7]. However, information on fecal microbes and physicochemical parameters in freshwater systems is limited as most previous investigations occurred in saline environments [7,8]. The available freshwater investigations comprised notable shortcomings including: short sampling periods and/or few collected samples [9], limited sampling locations (in similar land use types) [9], long periods of time (weeks or months) between sample collection [10], sample collection solely during baseflow conditions [10], or failure to account for constituents such as chloride (Cl$^-$), which has been shown to influence microbe concentrations in laboratory settings [10,11]. Studies from freshwater systems affirmed relationships between temperature, pH, and fecal microbe concentrations [9,10,12]. However, investigations, including many physicochemical (e.g., chloride) relationships with fecal contamination, are greatly lacking. There is, therefore, need for investigations that include high spatio-temporal sampling regimes, occurring at different stage (streamflow and hydro-climate) conditions, in areas comprising different land use practices, over longer time periods. Information gained from these investigations could aid in the implementation of effective strategies to reduce fecal microbe concentrations in receiving waters.

Land use practice impacts on fecal contamination is relatively well-documented [6,13–18]. Both agricultural and urban land use practices have been shown to generally increase fecal concentrations [6,17], attributed to livestock husbandry [14], manure application [16], or poorly maintained wastewater infrastructure [17], and increased impervious surfaces [19,20]. Previous work reported negative correlations between forested areas and fecal concentrations in receiving
waters [21]. Shortcomings of previous work include that many investigations occurred in areas comprising similar land use practices (i.e., lacking variability) [22,23] or included few sampling locations [24]. Therefore, information regarding the influence of land use practices on fecal microbe concentrations in mixed-land-use watersheds (comprising the majority of watersheds globally) are quite limited. Furthermore, the predominant focus on storm events [25], limits information regarding the influence of land use practices on E. coli concentrations during other flow and transport conditions. Simultaneous combined analysis of fecal pollution (E. coli concentration), land use practices, and physicochemical parameters relationships is also lacking in previous investigations. Such advanced integrated understanding will provide land use managers with more detailed information regarding land use impacts on fecal pollution; thereby, aiding in freshwater quality management decisions.

Multiple study designs and sampling regimes have been implemented to investigate fecal pollution in receiving waters, including laboratory and field based designs. Laboratory studies usually included simulations [26], whereas field based designs comprise event based sampling [27], periodic sampling [19], stochastic sampling [28] and, in limited number, scale-nested experimental watersheds [6]. The scale-nested experimental watershed study design has been particularly effective for quantifying factors (e.g., physicochemical parameters, land use practices) influencing response variables of interest (e.g., E. coli concentration) in receiving waters, particularly in mixed-land-use watersheds [29–36]. To achieve this, nested watershed study designs divide larger watersheds into a series of sub-catchments, each with a monitoring (gauging) site at its drainage terminus [30,32,36–38]. Hydrologic characteristics and land use practices can then be isolated though sub-catchment delineation [36]. Quantification of the influencing processes recorded at the sub-catchment scale allows for the identification of the influence and cumulative effect of various land use practices on the response variable of interest [39]. Given its use in numerous peer reviewed publications over multiple decades, the scale-nested and paired [36]) experimental watershed study design is the optimal study design for investigating current knowledge gaps regarding fecal contamination, physicochemical parameters, and land use practices.

The widespread, frequent, and persistent fecal pollution characteristics of the Appalachian region of the USA are similar (representative) to numerous locations globally. The region is, therefore, well suited for (transferrable) research into factors affecting fecal contamination in
receiving waters [40]. Water quality and security is a primary concern in rural Appalachia, as fecal contamination poses substantial risk to residents [41]. The risk is elevated by inadequate wastewater treatment infrastructure, geographical isolation, inaccessible terrain, and poverty [41]. Fecal contamination is, therefore, a primary concern, and improved understanding of factors increasing fecal contamination can inform better management practices and improve water quality in the region. Furthermore, the Appalachian region is also physiographically diverse, consisting of distinct geographic, climatological, and ecological areas, typically divided into distinct Northern, Central and Southern regions [42]. The diverse physiography makes results from investigations in these distinct regions comparable and transferable to similar areas globally. For example, the central Appalachian region is comparable to areas such as Hokkaido or Northern Honshu in Japan as these areas comprise temperate climates and well-distributed year-round rainfall [43,44].

The overarching objective of the current work was to quantitatively compare relationships between fecal concentration (E. coli colony forming units), physicochemical parameters, and land use practices. Sub-objectives included (1) identifying the dominant factors that influence fecal microbe concentrations in receiving waters, and (2) investigating the effects of seasonal variation on factors (e.g., physicochemical parameters, land use practices) influencing fecal microbes (E. coli) in freshwater streams. Study outcomes provide new quantitative insight to these issues thereby providing land use managers with advanced science-based information to improve management practices and policies in freshwater systems.

2. Materials and Methods

2.1. Study Site Description

The location for the current work was a 3rd order tributary of the Monongahela River, the 23 km² mixed–land use urbanizing West Run Watershed (WRW). The WRW is located in Morgantown, WV, USA, and contains many different land use practices ranging from various mixed development (e.g. urban and commercial), agriculture, and forested practices. Based on 2016 National Agriculture Imagery Program (NAIP) land use and land cover data, at the time of this investigation WRW consisted of 42.7% forested, 37.7% mixed development and 19.4% agricultural land use (Table 1, site #22). West Run Creek, the primary drainage of WRW includes small floodplains and is a narrow, moderately entrenched stream [6,45]. The elevation in WRW ranges between 420 to 240 m above mean sea level from the headwaters to the confluence of the
Monongahela River [6]. The geology of WRW comprises numerous Paleozoic era rock outcroppings and the Monongahela series located in the headwaters [6]. Historic mining of the two coal seams in WRW, the Upper Kittanning, and more specifically, the Pittsburg coal seam resulted in pervasive water quality problems in the watershed (particularly in the headwaters) [6,46].

The climate regime in West Virginia and the city of Morgantown, residing in part, in the WRW and in Monongalia County WV, has a climate lacking a dry season and warm summers (average monthly temperature > 22 °C) and cold winters (average monthly temperature < 0 °C) [47]. Between 1981 and 2010 Morgantown received approximately 1060 mm of average annual precipitation, with the wettest and warmest month (July) comprising an average daily temperature and monthly precipitation of 23 °C and 117 mm, respectively [48]. Conversely, the coldest (January) month included an average daily temperature of −0.4 °C and the driest (February) month an average monthly precipitation of 66 mm [48].

A twenty-two study site (i.e., n = 22 stream gauging sites) scale-nested and paired experimental watershed study design [30,39,49–51] was implemented in WRW in 2016. Field surveys and GIS were used to identify study site locations and associated sub-catchments. The sampling sites (numbered in downstream order were located on the 1st and 2nd order confluence tributaries of WRW (#1, #2, #5, #7, #8, #9, #11, #12, #14, #15, #16, #17, and #20) and along West Run Creek (#3, #4, #6, #10, #13, #18, #19, #21, and #22) and included many land use practices (Table 10; Figure 20). Forested land use was the predominant land use practice in WRW during the time of the investigation, accounting for 42.7% of the total land use practices in the watershed (Table 10, site #22). All the sub-catchments except #1, #11, #15, #16, and #20 were by majority forested. Sub-catchments #11 and #16 were primarily agricultural, whereas sub-catchments #1, #15 and #20 were primarily mixed development (Table 10).
Table 10. Land use/land cover characteristics (% cover) of 22 monitoring sites in West Run Watershed (WRW), West Virginia, USA, including total drainage area (km²).

<table>
<thead>
<tr>
<th>Site</th>
<th>Mixed Development (%)</th>
<th>Agriculture (%)</th>
<th>Forested (%)</th>
<th>Drainage Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>53.23%</td>
<td>38.70%</td>
<td>8.07%</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>13.58%</td>
<td>12.20%</td>
<td>74.21%</td>
<td>0.29</td>
</tr>
<tr>
<td>3</td>
<td>22.35%</td>
<td>16.17%</td>
<td>61.32%</td>
<td>1.87</td>
</tr>
<tr>
<td>4</td>
<td>25.88%</td>
<td>14.91%</td>
<td>59.00%</td>
<td>2.48</td>
</tr>
<tr>
<td>5</td>
<td>23.35%</td>
<td>25.51%</td>
<td>51.14%</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>23.91%</td>
<td>17.25%</td>
<td>58.70%</td>
<td>3.72</td>
</tr>
<tr>
<td>7</td>
<td>16.33%</td>
<td>28.60%</td>
<td>54.91%</td>
<td>0.78</td>
</tr>
<tr>
<td>8</td>
<td>30.78%</td>
<td>16.47%</td>
<td>52.35%</td>
<td>1.55</td>
</tr>
<tr>
<td>9</td>
<td>27.57%</td>
<td>19.33%</td>
<td>52.84%</td>
<td>2.29</td>
</tr>
<tr>
<td>10</td>
<td>24.92%</td>
<td>18.40%</td>
<td>56.49%</td>
<td>6.18</td>
</tr>
<tr>
<td>11</td>
<td>18.15%</td>
<td>41.87%</td>
<td>39.16%</td>
<td>1.75</td>
</tr>
<tr>
<td>12</td>
<td>31.77%</td>
<td>33.72%</td>
<td>34.51%</td>
<td>1.75</td>
</tr>
<tr>
<td>13</td>
<td>26.83%</td>
<td>25.77%</td>
<td>47.15%</td>
<td>10.53</td>
</tr>
<tr>
<td>14</td>
<td>16.19%</td>
<td>26.43%</td>
<td>56.92%</td>
<td>3.36</td>
</tr>
<tr>
<td>15</td>
<td>70.28%</td>
<td>10.31%</td>
<td>19.42%</td>
<td>0.98</td>
</tr>
<tr>
<td>16</td>
<td>5.38%</td>
<td>58.72%</td>
<td>35.16%</td>
<td>0.25</td>
</tr>
<tr>
<td>17</td>
<td>4.78%</td>
<td>9.38%</td>
<td>85.84%</td>
<td>0.75</td>
</tr>
<tr>
<td>18</td>
<td>25.98%</td>
<td>24.88%</td>
<td>48.86%</td>
<td>16.41</td>
</tr>
<tr>
<td>19</td>
<td>29.45%</td>
<td>22.45%</td>
<td>47.85%</td>
<td>18.88</td>
</tr>
<tr>
<td>20</td>
<td>89.16%</td>
<td>4.19%</td>
<td>6.61%</td>
<td>3.42</td>
</tr>
<tr>
<td>21</td>
<td>38.10%</td>
<td>19.46%</td>
<td>42.23%</td>
<td>22.93</td>
</tr>
<tr>
<td>22</td>
<td>37.71%</td>
<td>19.38%</td>
<td>42.66%</td>
<td>23.24</td>
</tr>
</tbody>
</table>

Note: due to the omission of certain categories (i.e., wetland, open water, etc.) and certain categories comprising combinations other others (e.g., mixed development = urban and residential) land use percentages may not sum to 100%. Final row (site #22) indicates total values for the entire watershed.

The predominance of forested cover among the sites should not be taken to imply that catchments were equivalent in terms of land use. For example, the reference sub-catchment (control) for the current work, sub-catchment #17 comprised 85.84% forested land use 9.4% agricultural and 4.8% mixed development land use practices. This sub-catchment comprised the largest percentage of forested land use practices among the sub-catchments (hence, selection as reference sub-catchment). Conversely, sub-catchment #12 comprised 34.5% forested, 33.7% agriculture, and 31.7% agriculture land use practices and had the lowest percentage forested land use among the primarily forested sites (Table 10). Sub-catchment #17 is therefore distinct in terms of land use practices relative to sub-catchment #12, despite both being predominantly forested.
general, at the time of the investigation, mixed development comprised the second largest percentage of land use practices (37.7%) and agricultural land use practices accounted for the lowest percentage of land use practices (19.4%) in WRW.

Figure 20. Land use/land cover of West Run Watershed, Morgantown, WV, USA, including monitoring/sampling locations for the current investigation.

2.2. Data Collection

Climate data, including precipitation (Campbell Scientific TE525 Tipping Bucket Rain Gauge), average air temperature, relative humidity (Campbell Scientific HC2S3 Temperature and Relative Humidity Probe), and average wind speed (Campbell Scientific Met One 034B Wind Set instrument), were recorded at a three-meter height within approximately 30 m of site #13 (Figure 20), for the duration of the study period (2 January 2018–1 January 2019).

For the current work, weekly water grab-samples were collected as per Petersen et al. [6], Hubbart et al. [52], Kellner and Hubbart [38], and Zeiger and Hubbart [37,53] from each monitoring site (stream order ≤ 3). Water sample collection proceeded through numerical order of
sites starting at 09:00 at site #1. To reduce overall sampling time (increasing sample representativeness during processing), sites #9 and #10 were sampled before sites #7 and #8, due to their proximity relative to site #6 (Figure 20). The calendar year duration of the sampling period (1 February 2018–1 January 2019) was longer than previous work on fecal contamination, allowing for assessment of seasonal variability of *E. coli* concentration, physicochemical, and hydro-climate data [54,55]. This resulted in a more comprehensive quantification of fecal contamination (*E. coli*) regimes at sub-catchment mixed-land-use scales than offered through most, if not all previous studies in the published literature. A total of 1166 spatio-temporally delineated fecal contamination (*E. coli*) concentration values were obtained during this high-resolution study.

Once collected, water samples were transported to the Interdisciplinary Hydrology Laboratory, located in the Davis College of Agriculture, Natural Resources and Design at West Virginia University, for analyses. *Escherichia (E) coli* was used as an indicator organism to quantify fecal contamination as per previous work investigating fecal contamination [2,6]. The Colilert test, developed by IDEXX Laboratories Inc. and approved by the U.S. Environmental Protection Agency (EPA) [56], was used to quantify *E. coli* colony forming units (CFU). The test was designed to eliminate the need for sample dilution when evaluating fecal concentration in water samples and is included in the Standard Methods for Examination of Water and Wastewater [56,57]. The likelihood of reporting inaccurate results (i.e., false positives ± 10%) when using the test is low due to Colilert’s Defined Substrate Technology nutrient-indicator (ONPG), and a selectively suppressing formulated matrix. The ONPG is a carbon source that most non-target organisms lack the enzyme to utilize, rendering them unable to grow or interfere [56]. The few non-target organism that can use ONPG as a carbon source is suppressed by a selective matrix [56]. As per Colilert instructions, the Colilert (ONPG) substrate was added to 100 mL of sampled water, sealed in the Quanti-Tray, and incubated at 35 °C for 24 h. The Quanti-Tray system comprises 96 total wells: 48 large wells (49, including an overflow well) and 48 small wells. Results are reported in CFU per 100 mL [56]. Following incubation, fluorescing (positive for *E. coli*) wells were enumerated using a UV light (6 watt, 365 nm wavelength) and compared to the Quanti-Tray Most Probable Number (MPN) table. The MPN table is used to convert the number of positive wells to an *E. coli* concentration value (CFU per 100mL), with a 95% confidence interval. The applied method thus used a MPN approach to estimate *E. coli* CFU concentration;
therefore, \textit{E. coli} concentration data was referred to as CFU not MPN during the investigation. The \textit{E. coli} concentration range that could be detected using this method was <1 to 1011.2 CFU. \textit{E. coli} concentrations exceeding 1011 CFU per 100 mL; therefore, could not be estimated accurately. However, this shortcoming was deemed acceptable due to the accuracy at detecting low \textit{E. coli} concentrations provided by the method, which is important when sampling for \textit{E. coli} outside of storm events or in land use areas less prone to fecal contamination.

Concurrent with the collection of water grab samples, five physicochemical variables were collected using a handheld multi-parameter water quality sonde (YSI Inc./Xylem Inc.) fitted with an Ion Selective Electrode (ISE) multi-probe [58]. Variables included water temperature (°C), dissolved oxygen (DO), specific conductance (SPC), pH, and Chloride ion. The ISE probe sensed water temperatures ranging from −5 °C to 70 °C with an accuracy of ±0.2 °C, SPC ranging from 0 to 200 mS cm$^{-1}$ with an accuracy of ±0.5% of reading or ±0.001 mS cm$^{-1}$, whichever is greater, (for readings 0–100 mS cm$^{-1}$) or ±1.0% of reading (for readings 100–200 mS cm$^{-1}$), DO ranging from 0 to 50 mg L$^{-1}$, with an accuracy of ±1% (for readings 0–20 mg L$^{-1}$) or ±8% (for readings 20–50 mg L$^{-1}$), pH ranging from 0 to 14 units, with an accuracy of ±0.2 units, and Chloride ranging from 0 to 1000 mg/L (at water temperatures from 0 to 40˚C) with an accuracy ±15% of reading or 5 mg/L (whichever is greater) [58].

2.3. Data Analysis

Descriptive statistics for \textit{E. coli}, physicochemical variables and hydro-climate data aggregated to the study period were calculated. Statistical analyses were conducted using OriginPro 2019 (OriginLab Corporation). The Anderson Darling Test was used for normality testing [59]. Land use practices were reclassified into three major categories prior to analysis including mixed development, agriculture, and forested land use [6]. Mixed development included roads, impervious surfaces, mixed development, and barren areas. Agriculture comprised low vegetation, hay pasture, and cultivated crops. Forested land use constituted mine grass, forest, mixed mesophytic forest, dry mesic oak forest, dry oak (pine) forest, and small stream riparian habitats. Seasonal variation was analyzed by dividing annual data into four quarter data subsets, comprising all weekly samples collected in three-month blocks starting 1 January 2018. Consequently, quarter one included 2 January 2018–27 March 2018 (winter); quarter two included 3 April 2018–26 June 2018 (Spring); quarter three included 3 July 2018–25 September 2018 (summer); and quarter four
included 2 October, 2018–1 January, 2019 (fall). Spearman correlation tests for both annual and quarterly datasets, with a significance threshold of $\alpha = 0.05$ [60], were used to analyze the relationship between $E. \text{coli}$ concentration, physicochemical parameters, and land use practices at all twenty-two sites. Finally, the annual data and quarterly data subsets comprising $E. \text{coli}$ concentrations, physicochemical parameters, and land use practices were analyzed using principal component analysis (PCA) (presented in biplots) across all 22 sampling locations.

3. Results and Discussion

Given the scope of this research and large subsequent dataset, the authors have included the most salient tables and figures in text to facilitate presentation of results and discussion. For the reader wishing to learn more, comprehensive descriptive statistics tables are provided in Appendix A (referenced throughout).

3.1. Climate During Study

West Run Watershed received approximately 20% more total precipitation in 2018 (1378 mm) than annual averages dating back to 2007 (1096 mm) [61]. October was the driest (47 mm) and September was the wettest (186 mm) months during the study period (Figure 21), with September receiving more than double the historic average precipitation (80 mm) [61]. Consequently, September included approximately 14% of the precipitation received during 2018. During 2018, average annual air temperature ($12 \, ^\circ\text{C}$) was close to the historic annual average ($11 \, ^\circ\text{C}$) [61]. The coldest average monthly temperature in WRW was recorded in January ($-4 \, ^\circ\text{C}$), whereas July had the warmest average monthly temperature ($22 \, ^\circ\text{C}$) [61]. Average annual relative humidity was 76% during 2018 and was, therefore, characteristically high (Figure 21), as is common in the region [61]. Ultimately, the climate in WRW during 2018 was predictably variable and consistent with historic trends. As is characteristic of WRW, and the region, 2018 did not include a dry season; however, quarters two and three (spring and summer) received more precipitation than quarters one and four (winter and fall). This was the result of large precipitation events during quarters two/spring (e.g., 6 May; 24 mm) and three/summer (e.g., 9 September; 60 mm) (Figure 21) [61].
Figure 21. Thirty-minute time series of climate variables during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: Stream stage was monitored in the primary stream of WRW, West Run Creek, at site #13 (Figure 20).

3.2. E. Coli Concentrations

Sub-catchment #16, comprising predominantly agricultural land use practices (59%) had the highest E. coli concentration (560 CFU per 100 mL) during the current investigation, similar to previous inventions in WRW [6,18]. This sub-catchment also had the highest median concentration (575 CFU per 100 mL) (Figure 22; Appendix A Table 14). These results are supported by previous work that reported significant correlations (p < 0.04) between agricultural land use and E. coli concentrations [6] and increased fecal contamination in agricultural areas [21,27]. Conversely, forested land use areas (site #2; 74% Forested and site #5; 51% Forested) comprised the lowest E. coli concentrations, including the lowest median (3 CFU per 100 mL), and average (34 CFU per 100 mL), respectively (Figure 22; Appendix A Table 14). The low E. coli concentrations recorded at these two sites, and others including sites #7, #8, and #9 located in the headwaters of WRW (Figure 20) were impacted by acid mine drainage (AMD), which lowers
the pH of receiving waters [62], potentially killing fecal bacteria [63]. For a more comprehensive discussion of pH, the reader is referred to the physicochemical parameters Section 3.3. Under approximately average pH conditions, forested land use areas had lower *E. coli* concentrations (e.g., site # 17: 86% forested average (avg.) 206 CFU per 100 mL) relative to areas comprising either agriculture (e.g., site # 16: 59% agriculture, avg. 560 CFU per 100 mL) or mixed development (e.g., site # 20: 89% mixed development, avg. 415 CFU per 100 mL) (Figure 22). In previous investigations, decreased fecal contamination was recorded in forested areas [21], and attributed to increased receiving water quality [64], thereby supporting the results of the current investigation. The motivation for sampling between storm events, is strengthened by the low average *E. coli* concentrations (specifically in the headwaters) recorded during the study period (2 January 2018—1 January 2019). Results support previous investigations that showed increased *E. coli* concentrations in agricultural areas (e.g., site # 16) and decreased *E. coli* concentrations in forested areas (e.g., site # 17). However, no published investigations included sampling over a full annual year using such a high spatial and temporal resolution. Results here are therefore unique, comprehensive, and may increase confidence in previous study outcomes.
Figure 22. Box and whisker plot of E. coli concentration (CFU per 100 mL) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

3.3. Physicochemical Parameters

Mixed development land use (e.g., site #20: 89% mixed development) had elevated annual average water temperatures (avg. 12.88 °C) relative to agriculture (e.g., site #16: 59% agriculture, avg. 12.00 °C) and forested land use practices (e.g., site #17: 86% forested, avg. 11.67 °C) (Figure 23; Appendix A Table 15). Previous investigations linking elevated receiving water temperature to mixed development areas support the results of the current study [65–68]. Elevated water temperatures in mixed development areas are typically attributed to decreased vegetation (reduced stream shading) and warmer impervious surfaces (e.g., road and building surfaces) and increased surface runoff (volume and temperature) during storm events [65–68]. During the current investigation water temperatures were lower in forested areas (e.g., site #17: 86% forested),
including a lower average (11.67 °C), median (10.40 °C), and maximum (21.70 °C) relative to other land uses (Figure 23; Appendix A Table 15), thereby supporting the results from previous work [69]. The study design and high spatial and temporal resolution-sampling regime of the current investigation resulted in an extensive water temperature and *E. coli* concentration dataset. This is important, given temperature has been identified as the primary factor influencing *E. coli* survival in the environment, accounting for up to 61% of the variance (based on inactivation rates) of *E. coli* populations [70]. Therefore, this analysis (see non-parametric results Sections 3.4 and 3.5) expands current understanding of land use practice and water temperature impacts on *E. coli* concentrations.

Figure 23. Box and whisker plot of water temperature (°C) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.
Previous work reported decreased pH in WRW, particularly in the headwaters, attributed to historic mining activities and subsequent AMD [6,18,46]. In the current work, approximately 55% of monitoring sites located in the headwaters had average pH values below six (Figure 24; Appendix A Table 16). The lowest pH was recorded in at site #8 with an annual average pH of 4.37 and median of 4.23. Monitoring sites #7, #8, and #9, (comprising one of the paired watersheds) had the lowest pH values among the sites (annual averages 5.03; 4.37; 5.08 and medians 5.04; 4.23; 4.93, respectively). Sampling locations in lower WRW displayed increased pH values, indicating a dilution of the AMD that was prevalent in the headwaters (no historic mining in lower reaches). For example, when comparing the annual average pH of two sites located on West Run Creek (site #13: 6.10 and #18: 7.18) there was a notable increase in pH (1.08) from acidic to neutral (Figure 24; Appendix A Table 16). Ultimately, the pH data recorded during the current investigation provides high sampling density evidence for legacy effects of historic mining practices. Notably, while mining activities in WRW ceased by 1977 [46]. Long-term effects of those practices were still impacting receiving waters at the time of the current study.
Figure 24. Box and whisker plot of pH descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

Acid mine drainage (AMD) impacted sites, including the paired watershed comprising sites #7, #8 and #9, generally included the highest specific conductance (SPC) data recorded during the study period (2 January 2018–1 January 2019). For example, site #8, which was the most heavily impacted by AMD (see preceding section) comprised the highest average (1661.13 µS/cm) and median (1480.00 µS/cm) SPC values (Figure 25; Appendix A Table 17). These findings are supported by previous work that reported elevated SPC in AMD impacted locations [71], including previous work investigating the influence of coal mining on conductivity of waters in Appalachia [72]. Increased SPC in AMD impacted areas can be attributed to increased iron, sulfate, copper, cadmium, arsenic, and/or other constituents in the water increasing ion availability [73]. Mixed development sub-catchments (sites #15: 70% mixed development and #20: 89% mixed development) also displayed elevated SPC data (averages of 1392.64 µS/cm and 1463.58 µS/cm,
respectively). Moreover, these mixed development areas included the highest recorded maximum SPC (6631.00 µS/cm and 6106.00 µS/cm, respectively) (Figure 25; Appendix A Table 17). These results are supported by previous investigations reporting increased SPC in mixed development locations [74,75], attributed to increased ions in receiving water originating from diverse sources, including transportation, sewage treatment, and infrastructure development [75]. Moreover, the low SPC recorded at forested sites (e.g., site #17: 86% forested; average SPC 249.11 µS/cm), in the lower portion of WRW, during the current work is also supported by previous work that reported negative associations between SPC and forested land cover [76].

![Box and whisker plot of specific conductance (µS/cm) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA.](image)

Dissolved oxygen (DO) was lowest at site #1, comprising an annual average of 85.93% and median of 85.10%, and the highest at site #20, with an average of 104.63% and median of 102.40%
(Figure 26; Appendix A Table 18). Statistical analysis did not reveal a significant (CI = 0.05) relationship between DO and land use practices, attributable to unmeasured DO influencing factors. Variables beyond the scope of the current investigation included ground water depth and antecedent soil water conditions) [77], aquatic macrophytes [78], aquatic plant photosynthesis [79], and aquatic chemical, physical, and biochemical activities [80]. These factors can influence DO independent from changes to land use practices [77–80], thereby, obscuring the influence of land use changes on DO in receiving waters. However, increased DO variability was recorded at mixed development sub-catchments (site #15: 70% mixed development and site #20: 89% mixed development) in the lower portion of WRW (Figure 26). DO in streams can be impacted by urbanization through increased primary production or decomposition of organic matter [81]. Therefore, increased DO variability in mixed development areas can alter microbial community structures [82] of associated receiving waters, potentially indirectly affecting facultative anaerobic bacteria (e.g., *E. coli*) through inter specific competition [83].
Mixed development sub-catchments showed increased chloride ion (Cl\(^{-}\)) concentrations relative to other land use practices. For example, mixed development at site #1 (53%), site #15 (70%), and site #20 (89%) accounted for the highest chloride concentrations among the sampled locations (average concentrations 272.11 mg/L, 220.17 mg/L and 282.87 mg/L, respectively) (Figure 27; Appendix A Table 19). Previous work, using a similar study design, also reported increased chloride relative to increased mixed development land use practices [52]. The application of road salts in mixed development areas has been presented as a contributor to elevated chloride levels in these land use areas [52,84]. Forested land use areas (site #17: 86% forested) comprised the lowest chloride concentrations in WRW, including the lowest average (13.34 mg/L), median (11.79 mg/L), minimum (6.85 mg/L), maximum (39.82 mg/L), and lowest
standard deviation (6.13 mg/L) (Figure 27; Appendix A Table 19). Previous work reported similar low(er) chloride concentrations in forested areas relative to other land use types, thereby validating results from the current investigation [52,84,85]. Given the study design and sampling regime, this work shows convincingly that the impact of chloride on microbe concentrations, including fecal microbes, will be increased in mixed development areas.

Figure 27. Box and whisker plot of chloride ion (mg/L) descriptive statistics at 22 sampling locations during study period (2 January 2018–1 January 2019) in West Run Watershed, Morgantown, WV, USA. Box delineates 25th and 75th percentiles; line denotes median; square shows mean; whisker describes 10th and 90th percentiles; x shows maximum and minimum when above and below, respectively.

3.4. Annual Non-Parametric Statistical Results

Normality testing showed that annual E. coli and physicochemical data were non-normally distributed. Therefore, Spearman’s Correlations tests, the non-parametric equivalent of the Pearson’s correlations tests [86], were used to investigate annual relationships between E. coli concentrations and physicochemical parameters at all 22 sampling locations (Table 11). Results
showed that water temperature was significantly (p < 0.05) positively correlated with \textit{E. coli} concentrations at 14 (64\%) of the 22 sampling locations (Table 11). Six of eight West Run Creek sampling locations (75\%) included significant correlations between \textit{E. coli} concentration and water temperature. Despite previous work also reporting spatial and site specific variation regarding water temperature and \textit{E. coli} concentrations [87], water temperature is historically regarded as the primary environmental variable influencing \textit{E. coli} survival in the environment [70]. However, in WRW, pH was significantly negatively correlated (p < 0.05) to \textit{E. coli} concentrations at 77\% (17 of 22) of sampling locations. Therefore, given the presence of AMD in WRW [6,18,46], which can lower the pH of receiving waters [62] killing (or inactivating) fecal bacteria [63], pH exceeded the influence of water temperatures on \textit{E. coli} concentrations. This is an important finding as it challenges traditional beliefs that temperature is the primary factor influencing the environmental survival of \textit{E. coli}. Moreover, pH values displayed a tipping point (threshold) of between 7.68–7.76, with pH values below this range including significant correlations (p < 0.05) with decreased \textit{E. coli} concentrations. Two West Run Creek sites (site #21 and #22) that had insignificant relationships between pH and \textit{E. coli} concentrations were located near the terminus (confluence with the Monongahela River) of the watershed. At these sites, AMD was diluted to levels not influencing \textit{E. coli} survival. Therefore, results also provide evidence for the dilution of AMD impacted waters and subsequent decreased impact on fecal microbe viability. Subsequently, \textit{E. coli} concentrations could potentially be used to assess the freshwater health for aquatic organisms’ sensitive to decreased pH and AMD, essentially serving as a bioindicator. Notably, SPC, which is known to be impacted by pH and AMD, displayed significant correlations (p < 0.05) with \textit{E. coli} at three sites (#7, #8, and #9) which were particularly heavily impacted by AMD. Generally, SPC did not show consistent correlations with \textit{E. coli} concentrations across the sampling locations, as only 55\% (12 of the 22) sites displayed significant relationships. Therefore, based on the Spearman’s Correlations tests from the current investigation, SPC was poorly related to \textit{E. coli} CFU’s in the current study.

Dissolved oxygen (DO) lacked consistent correlations with \textit{E. coli} concentrations, as only 55\% (12 out of 22) of sites comprised significant correlations (Table 11). Furthermore, no significant relationship between land use practices and \textit{E. coli} concentrations and DO were found. Previous investigations showed that facultative anaerobic characteristics of \textit{E. coli} decrease its dependence on oxygen for survival [88], which may account for these results [89]. Spearman’s
correlation results between *E. coli* and chloride concentrations showed consistent (with the exception of site #1) significant negative correlations in sub-catchments comprising mixed development land uses in excess of 25% to 30% (Figure 20; Table 11). Analyses also identified that if mixed development land use is less than 25% to 30%, chloride is less likely to influence *E. coli* concentrations in receiving waters. Notably, this tipping point should not be interpreted to imply that lower concentrations are ecologically benign. The insignificant correlation at site #1 may be a function of the relatively small drainage of this sub-catchment (0.30 km²) and shorter stream distance, relative to other larger catchments with mixed development land use practices in excess of 25–30%. Previous work investigating the influence of mixed development, specifically urban, land use on chloride concentrations reported tipping points approaching 25%; thus, supporting the results from the current investigation [84]. Of importance, the distinct data set of the current work facilitated advanced understanding of the influence of land use practices on *E. coli* concentrations by Cl⁻ (possible attribute of winter road salting) that may suppress (inactivate or kill) fecal bacteria in receiving waters.
Table 11. Results of Spearman’s Correlation test, including annual *E. coli* concentration (colony forming units (CFU) per 100mL) water temperature (°C), pH, specific conductance (SPC; µS/cm), dissolved oxygen (DO; %) and chloride ion (Cl⁻; mg/L) at each sampling location (n = 22) during study period (2 January 2018—1 January 2019) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Temp.</td>
<td>SCC</td>
<td>0.35</td>
<td>0.59</td>
<td>0.20</td>
<td>0.19</td>
<td>0.75</td>
<td>0.37</td>
<td>0.12</td>
<td>0.26</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.01</td>
<td>0.00</td>
<td>0.18</td>
<td>0.19</td>
<td>0.00</td>
<td>0.01</td>
<td>0.43</td>
<td>0.08</td>
<td>0.18</td>
<td>0.00</td>
</tr>
<tr>
<td>pH</td>
<td>SCC</td>
<td>0.36</td>
<td>0.47</td>
<td>0.54</td>
<td>0.31</td>
<td>0.69</td>
<td>0.44</td>
<td>0.55</td>
<td>0.46</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SPC</td>
<td>SCC</td>
<td>-0.48</td>
<td>-0.23</td>
<td>-0.17</td>
<td>-0.18</td>
<td>-0.16</td>
<td>-0.16</td>
<td>-0.80</td>
<td>-0.51</td>
<td>-0.55</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.11</td>
<td>0.24</td>
<td>0.21</td>
<td>0.27</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>DO</td>
<td>SCC</td>
<td>0.31</td>
<td>-0.14</td>
<td>-0.18</td>
<td>-0.24</td>
<td>-0.01</td>
<td>-0.39</td>
<td>-0.16</td>
<td>-0.15</td>
<td>-0.37</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.03</td>
<td>0.35</td>
<td>0.21</td>
<td>0.10</td>
<td>0.93</td>
<td>0.01</td>
<td>0.27</td>
<td>0.32</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>SCC</td>
<td>-0.14</td>
<td>-0.20</td>
<td>-0.24</td>
<td>-0.26</td>
<td>-0.16</td>
<td>-0.21</td>
<td>-0.35</td>
<td>-0.45</td>
<td>-0.28</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.34</td>
<td>0.18</td>
<td>0.09</td>
<td>0.07</td>
<td>0.28</td>
<td>0.16</td>
<td>0.01</td>
<td>0.00</td>
<td>0.05</td>
<td>0.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Temp.</td>
<td>SCC</td>
<td>0.71</td>
<td>0.66</td>
<td>0.71</td>
<td>0.37</td>
<td>0.13</td>
<td>0.69</td>
<td>0.56</td>
<td>0.46</td>
<td>0.05</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.37</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.73</td>
<td>0.02</td>
</tr>
<tr>
<td>pH</td>
<td>SCC</td>
<td>0.62</td>
<td>0.74</td>
<td>0.69</td>
<td>-0.35</td>
<td>-0.14</td>
<td>0.16</td>
<td>0.56</td>
<td>0.57</td>
<td>-0.08</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.36</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.61</td>
<td>0.14</td>
</tr>
<tr>
<td>SPC</td>
<td>SCC</td>
<td>-0.16</td>
<td>-0.38</td>
<td>-0.06</td>
<td>-0.70</td>
<td>0.21</td>
<td>0.30</td>
<td>-0.30</td>
<td>-0.12</td>
<td>-0.50</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.28</td>
<td>0.01</td>
<td>0.67</td>
<td>0.00</td>
<td>0.16</td>
<td>0.04</td>
<td>0.04</td>
<td>0.43</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>DO</td>
<td>SCC</td>
<td>-0.41</td>
<td>-0.56</td>
<td>0.00</td>
<td>-0.73</td>
<td>-0.47</td>
<td>-0.58</td>
<td>-0.22</td>
<td>-0.20</td>
<td>-0.25</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.97</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
<td>0.17</td>
<td>0.09</td>
<td>0.68</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>SCC</td>
<td>-0.57</td>
<td>-0.41</td>
<td>-0.23</td>
<td>-0.52</td>
<td>0.16</td>
<td>-0.15</td>
<td>-0.44</td>
<td>-0.27</td>
<td>-0.37</td>
<td>-0.41</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.00</td>
<td>0.27</td>
<td>0.33</td>
<td>0.00</td>
<td>0.07</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: bold values indicate significant correlations (p < 0.05).

Principal component analysis (PCA) can be implemented to determine which explanatory variables account for the maximal variance in a data set, through the computation of multiple principal components and their respective Eigenvalues [90]. A principal component is defined as a linear function of original data set variables, which maximize variance and is uncorrelated with
other principal components [91]. Eigenvalues are used to identify principal components based on the assumption that components comprising the highest Eigenvalues will constitute principal components as Eigenvalues symbolize the variance of the data in that direction [90]. Given most data cannot be accurately described by a single principal component, numerous principal components are typically calculated and ranked based on their Eigenvalues [90]. For the current work, annual PCA results displayed three principal components with Eigenvalues exceeding 1 (an accepted threshold of importance [18,92]). The three principal components comprised Eigenvalues of 2.64, 1.99, and 1.33, respectively, and combined accounted for 66.27% of the variance on the annual data (Table 12). Consequently, the remaining six principal components accounted for only 33.73% of the variance of the data, of which principal component four accounted for 10.91% of the variance. Notably, principal component 4 comprised an Eigenvalue of 0.98, very close to the threshold of importance. For more comprehensive list of the coefficients of the variables comprising the three identified important principal components of the annual PCA please see Appendix A (Appendix A Table 20).
Table 12. Results of principal component analysis comprising 9 variables (E. coli concentration, water temperature, pH, SPC, DO, chloride, percentage of agricultural land use, percentage of forested land use, and percentage of developed land use) used to define 9 principal components, displaying eigenvalues, percentage of variance, and cumulative variance during the study period (2 January 2018–1 January 2019) across the 22 monitoring sites in West Run Watershed, West Virginia, USA

<table>
<thead>
<tr>
<th>Principal Component</th>
<th>Eigenvalue</th>
<th>Percentage of Variance</th>
<th>Cumulative Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.64</td>
<td>29.34%</td>
<td>29.34%</td>
</tr>
<tr>
<td>2</td>
<td>1.99</td>
<td>22.14%</td>
<td>51.48%</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>14.79%</td>
<td>66.27%</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td>10.91%</td>
<td>77.18%</td>
</tr>
<tr>
<td>5</td>
<td>0.79</td>
<td>8.80%</td>
<td>85.98%</td>
</tr>
<tr>
<td>6</td>
<td>0.61</td>
<td>6.73%</td>
<td>92.71%</td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>4.94%</td>
<td>97.65%</td>
</tr>
<tr>
<td>8</td>
<td>0.21</td>
<td>2.35%</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Note: bold numbers indicate eigenvalues exceeding 1 (representing importance).

The Annual PCA biplot (Figure 28) compliments the Spearman’ correlation results for pH and E. coli concentration. Water temperature was also closely correlated to E. coli concentrations based on annual PCA results. Both water temperature and pH has been reported to be closely correlated with fecal bacteria concentrations in previous work [9,10,12], thereby supporting results of the current investigation. Historic land use in WRW, specifically mining, influences the pH in the watershed, which is closely correlated with E. coli concentrations (Table 11; Figure 28). Moreover, water temperature, which is influenced by land use practices [65–68] is also closely correlated with E. coli concentrations. Therefore, annual PCA biplot results emphasize the influence of both historic and contemporary land use practices on fecal bacteria in receiving waters. The biplots of annual land use practices, physicochemical parameters, and E. coli concentration relationships facilitates visual assessment of land use impacts on physicochemical parameters, which influences fecal microbes. For example, forested land use showed a negative correlation with water temperature and E. coli concentrations (Figure 28) by reduced solar
radiation reaching the stream [93]. The decreased water temperatures may suppress *E. coli* concentrations, as the microbe is sensitive to temperature changes [94]. The influence of other physicochemical parameters (e.g., chloride) on *E. coli* concentrations were overshadowed by temperature and pH (Figure 28). Therefore, study results (based on annual average values) highlight temperature and pH as priority factors influencing *E. coli* in the receiving waters of WRW. Notably in watersheds with more neutral pH values and decreased legacy land use impacts (e.g., mining), the influence of physicochemical parameters on *E. coli* concentrations may be different [95]. In the current work, biplot results did not display a strong negative correlation between chloride and *E. coli* (Figure 28). Ultimately, results from the current study highlight legacy land use impacts (mining and AMD) as important considerations regarding microbial water quality management.
Figure 28. Results of principal components analysis, including biplots, for extracted principal components of annual *E. coli* concentration (CFU per 100 mL), water temperature (°C), pH, specific conductance (µS/cm), dissolved oxygen (DO; %), and chloride ion (mg/L) at 22 monitoring sites (indicated by the different colors) during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA.

3.5. Quarterly Non-Parametric Statistical Results

Quarterly PCA results displayed varying Eigenvalues between quarters. Quarter one comprised three principal components, which accounted for 70% of the cumulative data variance (Table 13). Conversely, quarters two, three, and four included four principal components that accounted for 81%, 84%, and 80% for the cumulative variance, respectively (Table 13). The remaining six principal components of quarter one and five, principal components of quarter two to four, did not comprise eigenvalues denoting importance, accounted for only 30%, 19%, 16%,
and 20% of the data variance in their respective quarters (Table 13). Appendix A (Tables 21–24) includes a more thorough presentation of the coefficients comprising the principal components of the PCA for all four quarters.

Table 13. Results of principal component analysis comprising 9 variables (E. coli concentration, water temperature, pH, SPC, DO, chloride, percentage of agricultural land use, percentage of forested land use and percentage of developed land use) used to define 9 principal components, displaying eigenvalues, percentage of variance and cumulative variance during quarter one (winter: 2 January 2018–27 March 2018); quarter two (spring: 3 April 2018–26 June 2018); quarter three (summer: 3 July, 2018–25 September, 2018); and quarter four (fall: 2 October 2018–1 January 2019) across the 22 monitoring sites in West Run Watershed, West Virginia, USA.

<table>
<thead>
<tr>
<th>Principal Component</th>
<th>Eigenvalue</th>
<th>Percentage of Variance</th>
<th>Cumulative Variance</th>
<th>Eigenvalue</th>
<th>Percentage of Variance</th>
<th>Cumulative Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.36</td>
<td>37.31%</td>
<td>37.31%</td>
<td>2.81</td>
<td>31.22%</td>
<td>31.22%</td>
</tr>
<tr>
<td>2</td>
<td>1.53</td>
<td>16.95%</td>
<td>54.26%</td>
<td>1.98</td>
<td>21.97%</td>
<td>53.19%</td>
</tr>
<tr>
<td>3</td>
<td>1.44</td>
<td>16.00%</td>
<td>70.25%</td>
<td>1.41</td>
<td>15.63%</td>
<td>68.82%</td>
</tr>
<tr>
<td>4</td>
<td>0.92</td>
<td>10.22%</td>
<td>80.48%</td>
<td>1.06</td>
<td>11.74%</td>
<td>80.56%</td>
</tr>
<tr>
<td>5</td>
<td>0.73</td>
<td>8.13%</td>
<td>88.60%</td>
<td>0.85</td>
<td>9.39%</td>
<td>89.95%</td>
</tr>
<tr>
<td>6</td>
<td>0.53</td>
<td>5.94%</td>
<td>94.54%</td>
<td>0.48</td>
<td>5.36%</td>
<td>95.32%</td>
</tr>
<tr>
<td>7</td>
<td>0.26</td>
<td>2.94%</td>
<td>97.48%</td>
<td>0.30</td>
<td>3.38%</td>
<td>98.70%</td>
</tr>
<tr>
<td>8</td>
<td>0.23</td>
<td>2.52%</td>
<td>100.00%</td>
<td>0.12</td>
<td>1.30%</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0.00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Quarter 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.89</td>
<td>32.07%</td>
<td>32.07%</td>
<td>2.74</td>
<td>30.39%</td>
<td>30.39%</td>
</tr>
<tr>
<td>2</td>
<td>2.14</td>
<td>23.80%</td>
<td>55.88%</td>
<td>1.87</td>
<td>20.73%</td>
<td>51.13%</td>
</tr>
<tr>
<td>3</td>
<td>1.34</td>
<td>14.91%</td>
<td>70.79%</td>
<td>1.47</td>
<td>16.30%</td>
<td>67.43%</td>
</tr>
<tr>
<td>4</td>
<td>1.16</td>
<td>12.92%</td>
<td>83.71%</td>
<td>1.09</td>
<td>12.14%</td>
<td>79.57%</td>
</tr>
<tr>
<td>5</td>
<td>0.58</td>
<td>6.44%</td>
<td>90.15%</td>
<td>0.79</td>
<td>8.83%</td>
<td>88.40%</td>
</tr>
<tr>
<td>6</td>
<td>0.39</td>
<td>4.29%</td>
<td>94.44%</td>
<td>0.57</td>
<td>6.38%</td>
<td>94.78%</td>
</tr>
<tr>
<td>7</td>
<td>0.31</td>
<td>3.49%</td>
<td>97.93%</td>
<td>0.29</td>
<td>3.22%</td>
<td>97.99%</td>
</tr>
<tr>
<td>8</td>
<td>0.19</td>
<td>2.07%</td>
<td>100.00%</td>
<td>0.18</td>
<td>2.01%</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0.00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Quarter 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: bold numbers indicate eigenvalues exceeding 1 (representing importance).

Quarterly PCA biplots (Figure 29) show the predominant influence of pH (presumably AMD driven) on fecal bacteria concentrations in WRW, as pH was closely correlated to E. coli concentrations during all four quarters of the study period (2 January 2018–1 January 2019). Conversely, the relationships between E. coli concentration and other physicochemical parameters varied between quarters. For example, water temperature was closely correlated with E. coli concentrations during quarters two and three, but not during quarter’s one and two (Figure 29).
The changing relationship between *E. coli* and water temperature is attributable to seasonal changes in air temperature and water temperatures [96]. Colder temperatures are known to suppress *E. coli* concentrations in receiving waters [94]. Thus, temporal and seasonal fluctuations of physicochemical parameters constitute important considerations regarding fecal concentration regimes (and mitigation practices) in receiving waters. The relationships between *E. coli* concentrations and land use practices also lacked consistency (Figure 29). This may have been due to the consistent close correlation between *E. coli* concentrations and pH. Legacy land use practices could be further confounding the influence of contemporary land use practices. Additionally, seasonal variation in land use practices (e.g., application of road salts during winter months) may contribute to the lack of consistent relationships. The high spatial temporal sampling regime and experimental watershed study design in conjunction with quarterly PCA biplots provided new (high-resolution) insight, emphasizing the complexity of temporal changes in *E. coli* concentration, physicochemical parameters, and land use practices. Results can be used by land use managers to inform water quality management strategies during different quarters (seasons), thereby improving efficacy. For example, focusing management strategies, including limiting livestock stream crossings through temporary fencing [97] and irrigation management [98] in quarter one in agricultural areas may be more effective in reducing *E. coli* concentrations than focusing on quarter four in some geographic locations.
Figure 29. Results of principal components analysis, including biplots, for extracted principal components of quarterly *E. coli* concentration (CFU per 100mL), water temperature (°C), pH, specific conductance (µS/cm), dissolved oxygen (%) and chloride ion (mg/L) at 22 monitoring sites (indicated by the different colors) during study period (2 January 2018–1 January 2019) in West Run Watershed, West Virginia, USA. Note: (A) represents quarter one (winter: 2 January 2018–27 March 2018); (B) represents quarter two (spring: 3 April 2018–26 June 2018); (C) represents quarter three (summer: 3 July 2018–25 September 2018); (D) represents quarter four (fall: 2 October 2018–1 January 2019).

3.6. Study Implications and Future Directions

The scale nested experimental watershed study design, calendar-year sampling period, and high spatial and temporal sampling regime used in this work allowed for the identification of legacy land use impacts on *E. coli* concentrations in receiving waters. Legacy land use impacts, specifically mining and subsequent AMD, were identified major influencers of *E. coli* concentrations (and by extension microbial water quality) in receiving waters, even exceeding the
influence of water temperature, commonly regarded as the primary factor influencing the environmental survival of *E. coli*. Additionally, the 7.68–7.76 pH tipping point identified regarding the significant correlations between pH and *E. coli* concentrations may indicate the use of *E. coli* as a potential bioindicator species for assessing the freshwater health, specifically in AMD impacted streams. Study results clearly identify legacy land use impacts of mining activity as a major influencer of microbial water quality. A threshold (tipping point) of 25–30% was identified regarding mixed development land use practices and significant (p < 0.05) negative correlations between *E. coli* and chloride concentrations. Increased chloride in the receiving waters of mixed development areas has been attributed to road salting [52,84]. This work shows that road salting in mixed development areas exceeding 25–30% total area may impact microbial water quality, through the suppression of *E. coli* concentrations. Future work should include the implementation of similar study design in physiographically dissimilar areas, including simultaneous implementation across different watersheds. This would allow for the comparison of data from climatically distinct regions, further improving understanding regarding physicochemical and land use impacts on *E. coli* regimes. Replication of the experimental watershed study design in areas not effected by AMD could potentially result in the identification of other tipping points regarding land use practices and physicochemical and *E. coli* relationships. For example, the influence of pH in the current work may have obscured the influence of other physicochemical parameters. Future investigations should be undertaken that includes multi-year sampling to better account for annual climate variability.

4. Conclusions

Fecal bacteria concentrations were investigated in a mixed-land-use watershed in the Appalachian region of the eastern United States, using a 22-site nested scale experimental watershed study design. Specific focus was given to the relationships between *E. coli* concentrations, physicochemical parameters (water temperature, pH, SPC, DO, chloride) and land use practices. In the study watershed, there was an approximate 112% increase in *E. coli* concentrations from the AMD impacted headwaters (avg. 177 CFU per 100mL) to the lower portion of the watershed (avg. 376 CFU per 100mL), an approximate 7 Km stream distance. Study results highlight the legacy impacts of historic mining (acid mine drainage) on *E. coli* concentrations, as Spearman correlation test results showed significant correlation (p < 0.05)
between pH and *E. coli* concentrations at 77% of sample sites. Moreover, a pH tipping point (threshold) in the range of 7.68–7.76 was identified in the current investigation, with pH values below this range including significant correlations (p < 0.05) with *E. coli* concentrations. Consequently, pH values in receiving waters below the 7.68–7.76 tipping point will start significantly impacting (decreasing) active *E. coli* concentrations. Furthermore, a land cover tipping point of 25–30% was identified for mixed development land use practices and significant (p < 0.05) negative correlations between *E. coli* and chloride concentrations. Therefore, study results indicate that in areas comprising mixed development in excess of 25% to 30%, the application of road salts may suppress fecal bacteria in receiving waters. The importance of seasonal variability on fecal concentrations in receiving waters was illustrated by temporal variability in quarterly PCA biplots of *E. coli* concentrations, physicochemical parameters, and land use practices. The current work advances understanding of land use practice (both historic and current) and physicochemical parameter influences on *E. coli* concentrations in contemporary mixed-land-use watersheds. Results will aid policy makers and land use managers in effective water quality management, in watersheds with fecal contamination challenges.

**Author Contributions:** For the current work author contributions were as follows: conceptualization, J.A.H.; methodology, J.A.H.; formal analysis, F.P. and J.A.H.; investigation, F.P. and J.A.H.; resources, J.A.H.; data curation, J.A.H.; writing—original draft preparation, F.P. and J.A.H.; writing—review and editing, J.A.H. and F.P.; visualization, F.P. and J.A.H.; supervision, J.A.H.; project administration, J.A.H.; funding acquisition, J.A.H.

**Funding:** This work was supported by the National Science Foundation under Award Number OIA-1458952, the USDA National Institute of Food and Agriculture, Hatch project accession number 1011536, and the West Virginia Agricultural and Forestry Experiment Station. Results presented may not reflect the views of the sponsors and no official endorsement should be inferred. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgments:** Special thanks are due to many scientists of the Interdisciplinary Hydrology Laboratory (https://www.researchgate.net/lab/The-Interdisciplinary-Hydrology-Laboratory-Jason-A-Hubbart). The authors also appreciate the feedback of anonymous reviewers whose constructive comments improved the article.
Conflicts of Interest: The authors declare no conflict of interest for the current work.
Appendix A

Table 14. Descriptive statistics of *E. coli* concentration (CFU per 100 mL) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>170</td>
<td>38</td>
<td>397</td>
<td>429</td>
<td>34</td>
<td>269</td>
<td>84</td>
<td>89</td>
<td>127</td>
<td>210</td>
<td>98</td>
</tr>
<tr>
<td>Med.</td>
<td>66</td>
<td>3</td>
<td>260</td>
<td>361</td>
<td>4</td>
<td>194</td>
<td>20</td>
<td>32</td>
<td>25</td>
<td>93</td>
<td>16</td>
</tr>
<tr>
<td>Min.</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>107</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Max.</td>
<td>1011</td>
<td>961</td>
<td>1011</td>
<td>1011</td>
<td>914</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>870</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>251</td>
<td>139</td>
<td>315</td>
<td>249</td>
<td>129</td>
<td>276</td>
<td>179</td>
<td>180</td>
<td>241</td>
<td>273</td>
<td>202</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>234</td>
<td>215</td>
<td>457</td>
<td>330</td>
<td>560</td>
<td>206</td>
<td>324</td>
<td>466</td>
<td>415</td>
<td>471</td>
<td>452</td>
</tr>
<tr>
<td>Med.</td>
<td>88</td>
<td>91</td>
<td>299</td>
<td>211</td>
<td>575</td>
<td>93</td>
<td>218</td>
<td>436</td>
<td>299</td>
<td>397</td>
<td>397</td>
</tr>
<tr>
<td>Min.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Max.</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
<td>1011</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>305</td>
<td>266</td>
<td>406</td>
<td>293</td>
<td>373</td>
<td>288</td>
<td>342</td>
<td>339</td>
<td>340</td>
<td>342</td>
<td>345</td>
</tr>
</tbody>
</table>
Table 15. Descriptive statistics of water temperature (°C) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>11.77</td>
<td>11.09</td>
<td>11.27</td>
<td>11.48</td>
<td>12.28</td>
<td>11.44</td>
<td>12.36</td>
<td>12.32</td>
<td>11.88</td>
<td>11.72</td>
<td>12.84</td>
</tr>
<tr>
<td>Med.</td>
<td>11.00</td>
<td>9.90</td>
<td>10.20</td>
<td>9.80</td>
<td>10.85</td>
<td>9.70</td>
<td>11.40</td>
<td>11.30</td>
<td>10.60</td>
<td>10.10</td>
<td>12.80</td>
</tr>
<tr>
<td>Min.</td>
<td>5.00</td>
<td>2.60</td>
<td>1.10</td>
<td>-0.10</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.10</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.20</td>
</tr>
<tr>
<td>Max.</td>
<td>20.30</td>
<td>19.10</td>
<td>19.50</td>
<td>21.90</td>
<td>23.10</td>
<td>22.10</td>
<td>21.50</td>
<td>21.50</td>
<td>21.80</td>
<td>22.20</td>
<td>22.10</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>4.20</td>
<td>5.51</td>
<td>6.15</td>
<td>6.95</td>
<td>7.73</td>
<td>7.21</td>
<td>6.65</td>
<td>6.28</td>
<td>6.72</td>
<td>7.02</td>
<td>6.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>12.10</td>
<td>12.09</td>
<td>12.34</td>
<td>12.86</td>
<td>12.00</td>
<td>11.67</td>
<td>12.53</td>
<td>12.64</td>
<td>12.88</td>
<td>12.87</td>
<td>13.74</td>
</tr>
<tr>
<td>Med.</td>
<td>10.70</td>
<td>10.70</td>
<td>10.60</td>
<td>11.40</td>
<td>11.10</td>
<td>10.40</td>
<td>10.90</td>
<td>11.10</td>
<td>11.80</td>
<td>11.60</td>
<td>13.05</td>
</tr>
<tr>
<td>Min.</td>
<td>-0.10</td>
<td>0.10</td>
<td>0.30</td>
<td>-0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>-0.10</td>
<td>-0.20</td>
<td>-0.30</td>
<td>-0.20</td>
<td>-0.10</td>
</tr>
<tr>
<td>Max.</td>
<td>23.50</td>
<td>22.00</td>
<td>24.00</td>
<td>23.00</td>
<td>24.00</td>
<td>21.70</td>
<td>24.00</td>
<td>24.10</td>
<td>23.50</td>
<td>24.60</td>
<td>27.70</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>7.87</td>
<td>6.91</td>
<td>7.66</td>
<td>6.89</td>
<td>6.99</td>
<td>6.99</td>
<td>7.44</td>
<td>7.72</td>
<td>7.76</td>
<td>7.88</td>
<td>8.61</td>
</tr>
</tbody>
</table>
Table 16. Descriptive statistics of pH at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>7.33</td>
<td>5.94</td>
<td>6.68</td>
<td>7.17</td>
<td>6.19</td>
<td>6.79</td>
<td>5.03</td>
<td>4.37</td>
<td>5.08</td>
<td>5.60</td>
<td>5.56</td>
</tr>
<tr>
<td>Med.</td>
<td>7.30</td>
<td>5.62</td>
<td>6.70</td>
<td>7.24</td>
<td>7.17</td>
<td>7.19</td>
<td>5.04</td>
<td>4.23</td>
<td>4.93</td>
<td>5.62</td>
<td>5.70</td>
</tr>
<tr>
<td>Min.</td>
<td>6.57</td>
<td>4.74</td>
<td>5.71</td>
<td>6.29</td>
<td>3.05</td>
<td>3.92</td>
<td>2.89</td>
<td>3.05</td>
<td>3.13</td>
<td>3.43</td>
<td>3.08</td>
</tr>
<tr>
<td>Max.</td>
<td>8.69</td>
<td>7.93</td>
<td>7.92</td>
<td>7.85</td>
<td>7.74</td>
<td>8.06</td>
<td>7.78</td>
<td>7.58</td>
<td>7.83</td>
<td>7.81</td>
<td>7.66</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.44</td>
<td>0.93</td>
<td>0.51</td>
<td>0.42</td>
<td>1.63</td>
<td>0.90</td>
<td>1.43</td>
<td>1.04</td>
<td>1.24</td>
<td>1.08</td>
<td>1.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>6.13</td>
<td>6.10</td>
<td>6.54</td>
<td>7.68</td>
<td>7.80</td>
<td>7.86</td>
<td>7.18</td>
<td>7.38</td>
<td>7.97</td>
<td>7.86</td>
<td>7.76</td>
</tr>
<tr>
<td>Med.</td>
<td>6.38</td>
<td>6.38</td>
<td>6.75</td>
<td>7.97</td>
<td>7.96</td>
<td>7.96</td>
<td>7.51</td>
<td>7.72</td>
<td>8.11</td>
<td>8.06</td>
<td>7.96</td>
</tr>
<tr>
<td>Min.</td>
<td>0.25</td>
<td>3.61</td>
<td>4.01</td>
<td>4.30</td>
<td>4.23</td>
<td>4.00</td>
<td>3.99</td>
<td>4.14</td>
<td>4.27</td>
<td>4.32</td>
<td>4.27</td>
</tr>
<tr>
<td>Max.</td>
<td>7.89</td>
<td>7.85</td>
<td>7.76</td>
<td>8.47</td>
<td>8.36</td>
<td>8.45</td>
<td>7.95</td>
<td>8.23</td>
<td>8.65</td>
<td>8.73</td>
<td>8.67</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.13</td>
<td>1.07</td>
<td>0.77</td>
<td>0.76</td>
<td>0.58</td>
<td>0.59</td>
<td>0.91</td>
<td>0.92</td>
<td>0.66</td>
<td>0.70</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Table 17. Descriptive statistics of water specific conductance (µS/cm) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>1639.28</td>
<td>958.51</td>
<td>778.70</td>
<td>810.22</td>
<td>1320.75</td>
<td>827.24</td>
<td>1053.28</td>
<td>1661.13</td>
<td>1497.19</td>
<td>1143.77</td>
<td>812.42</td>
</tr>
<tr>
<td>Med</td>
<td>1643.00</td>
<td>918.00</td>
<td>718.00</td>
<td>727.00</td>
<td>1353.00</td>
<td>769.00</td>
<td>999.00</td>
<td>1562.00</td>
<td>1480.00</td>
<td>1129.00</td>
<td>695.00</td>
</tr>
<tr>
<td>Min</td>
<td>331.60</td>
<td>358.10</td>
<td>286.20</td>
<td>301.80</td>
<td>633.00</td>
<td>354.30</td>
<td>503.00</td>
<td>813.00</td>
<td>688.00</td>
<td>476.60</td>
<td>530.00</td>
</tr>
<tr>
<td>Max</td>
<td>3778.00</td>
<td>2024.00</td>
<td>1620.00</td>
<td>1709.00</td>
<td>2654.00</td>
<td>1745.00</td>
<td>3378.00</td>
<td>4692.00</td>
<td>4302.00</td>
<td>2572.00</td>
<td>1955.00</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>484.90</td>
<td>301.93</td>
<td>305.86</td>
<td>303.74</td>
<td>307.58</td>
<td>263.52</td>
<td>404.69</td>
<td>555.53</td>
<td>496.47</td>
<td>349.67</td>
<td>284.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>885.06</td>
<td>961.58</td>
<td>503.58</td>
<td>1392.64</td>
<td>413.31</td>
<td>249.11</td>
<td>868.99</td>
<td>877.08</td>
<td>1463.58</td>
<td>945.33</td>
<td>738.12</td>
</tr>
<tr>
<td>Med</td>
<td>861.00</td>
<td>958.00</td>
<td>480.40</td>
<td>1116.00</td>
<td>385.00</td>
<td>226.40</td>
<td>845.00</td>
<td>831.00</td>
<td>1045.00</td>
<td>850.00</td>
<td>651.00</td>
</tr>
<tr>
<td>Min</td>
<td>516.00</td>
<td>449.50</td>
<td>321.50</td>
<td>510.00</td>
<td>231.60</td>
<td>144.40</td>
<td>451.60</td>
<td>494.00</td>
<td>408.90</td>
<td>486.50</td>
<td>297.60</td>
</tr>
<tr>
<td>Max</td>
<td>2072.00</td>
<td>2315.00</td>
<td>1052.00</td>
<td>6631.00</td>
<td>890.00</td>
<td>449.90</td>
<td>2184.00</td>
<td>2690.00</td>
<td>6106.00</td>
<td>3345.00</td>
<td>2185.00</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>279.25</td>
<td>292.86</td>
<td>111.90</td>
<td>1041.80</td>
<td>120.58</td>
<td>68.43</td>
<td>306.89</td>
<td>355.26</td>
<td>1193.23</td>
<td>458.84</td>
<td>359.11</td>
</tr>
</tbody>
</table>
Table 18. Descriptive statistics of dissolved oxygen (%) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>85.93</td>
<td>89.95</td>
<td>98.47</td>
<td>100.78</td>
<td>102.52</td>
<td>100.62</td>
<td>93.20</td>
<td>100.38</td>
<td>101.42</td>
<td>101.16</td>
<td>98.50</td>
</tr>
<tr>
<td>Med.</td>
<td>85.10</td>
<td>94.90</td>
<td>98.50</td>
<td>100.50</td>
<td>102.25</td>
<td>100.60</td>
<td>95.00</td>
<td>100.10</td>
<td>101.70</td>
<td>101.30</td>
<td>98.90</td>
</tr>
<tr>
<td>Min.</td>
<td>76.90</td>
<td>61.30</td>
<td>93.20</td>
<td>90.40</td>
<td>97.10</td>
<td>95.20</td>
<td>75.70</td>
<td>94.60</td>
<td>94.80</td>
<td>95.30</td>
<td>92.70</td>
</tr>
<tr>
<td>Max.</td>
<td>101.10</td>
<td>100.70</td>
<td>102.00</td>
<td>112.10</td>
<td>107.40</td>
<td>105.00</td>
<td>103.20</td>
<td>120.50</td>
<td>106.00</td>
<td>108.10</td>
<td>103.60</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>6.02</td>
<td>10.04</td>
<td>1.97</td>
<td>3.48</td>
<td>1.92</td>
<td>1.68</td>
<td>7.64</td>
<td>3.14</td>
<td>1.60</td>
<td>1.73</td>
<td>1.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>100.76</td>
<td>100.31</td>
<td>100.39</td>
<td>103.78</td>
<td>97.40</td>
<td>96.68</td>
<td>102.30</td>
<td>103.09</td>
<td>104.63</td>
<td>104.22</td>
<td>103.68</td>
</tr>
<tr>
<td>Med.</td>
<td>100.90</td>
<td>100.50</td>
<td>100.50</td>
<td>102.70</td>
<td>97.70</td>
<td>98.50</td>
<td>102.00</td>
<td>103.10</td>
<td>102.40</td>
<td>103.80</td>
<td>103.50</td>
</tr>
<tr>
<td>Min.</td>
<td>94.00</td>
<td>92.90</td>
<td>93.40</td>
<td>89.10</td>
<td>86.80</td>
<td>60.70</td>
<td>93.80</td>
<td>96.90</td>
<td>93.40</td>
<td>96.70</td>
<td>94.80</td>
</tr>
<tr>
<td>Max.</td>
<td>107.20</td>
<td>104.60</td>
<td>109.50</td>
<td>117.80</td>
<td>107.40</td>
<td>103.70</td>
<td>112.90</td>
<td>110.00</td>
<td>129.30</td>
<td>112.00</td>
<td>115.60</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.08</td>
<td>2.09</td>
<td>3.05</td>
<td>5.56</td>
<td>3.97</td>
<td>6.60</td>
<td>3.13</td>
<td>1.78</td>
<td>9.06</td>
<td>2.74</td>
<td>3.80</td>
</tr>
</tbody>
</table>
Table 19. Descriptive statistics of chloride ion (mg/L) at each sampling location (n = 22) during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA. Avg. = average, Med = median, Min. = minimum, Max. = maximum and Std. Dev. = standard deviation.

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>272.11</td>
<td>128.83</td>
<td>101.63</td>
<td>95.87</td>
<td>23.41</td>
<td>62.49</td>
<td>22.53</td>
<td>119.36</td>
<td>109.73</td>
<td>84.04</td>
<td>76.27</td>
</tr>
<tr>
<td>Med.</td>
<td>263.40</td>
<td>118.25</td>
<td>86.62</td>
<td>83.73</td>
<td>12.89</td>
<td>57.47</td>
<td>18.25</td>
<td>105.28</td>
<td>102.46</td>
<td>78.75</td>
<td>52.46</td>
</tr>
<tr>
<td>Min.</td>
<td>27.08</td>
<td>19.78</td>
<td>13.76</td>
<td>14.31</td>
<td>4.97</td>
<td>12.14</td>
<td>6.38</td>
<td>32.65</td>
<td>24.29</td>
<td>16.35</td>
<td>3.39</td>
</tr>
<tr>
<td>Max.</td>
<td>821.26</td>
<td>325.24</td>
<td>388.96</td>
<td>444.44</td>
<td>282.40</td>
<td>289.87</td>
<td>124.07</td>
<td>340.28</td>
<td>263.38</td>
<td>249.22</td>
<td>520.25</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>146.30</td>
<td>55.88</td>
<td>68.56</td>
<td>68.34</td>
<td>45.50</td>
<td>41.00</td>
<td>18.04</td>
<td>55.72</td>
<td>47.16</td>
<td>40.79</td>
<td>84.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Number</th>
<th>#12</th>
<th>#13</th>
<th>#14</th>
<th>#15</th>
<th>#16</th>
<th>#17</th>
<th>#18</th>
<th>#19</th>
<th>#20</th>
<th>#21</th>
<th>#22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>73.42</td>
<td>74.23</td>
<td>27.94</td>
<td>220.17</td>
<td>32.16</td>
<td>13.34</td>
<td>67.52</td>
<td>80.97</td>
<td>282.87</td>
<td>111.21</td>
<td>87.82</td>
</tr>
<tr>
<td>Med.</td>
<td>51.01</td>
<td>63.76</td>
<td>22.94</td>
<td>164.83</td>
<td>21.56</td>
<td>11.79</td>
<td>57.96</td>
<td>66.06</td>
<td>168.78</td>
<td>84.08</td>
<td>56.97</td>
</tr>
<tr>
<td>Min.</td>
<td>16.48</td>
<td>17.88</td>
<td>15.26</td>
<td>29.23</td>
<td>14.13</td>
<td>6.85</td>
<td>17.69</td>
<td>21.87</td>
<td>33.96</td>
<td>27.24</td>
<td>7.71</td>
</tr>
<tr>
<td>Max.</td>
<td>455.64</td>
<td>248.34</td>
<td>101.48</td>
<td>905.31</td>
<td>297.32</td>
<td>39.82</td>
<td>225.82</td>
<td>307.39</td>
<td>1242.68</td>
<td>491.53</td>
<td>472.78</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>74.83</td>
<td>45.41</td>
<td>17.91</td>
<td>176.53</td>
<td>44.33</td>
<td>6.13</td>
<td>40.02</td>
<td>54.18</td>
<td>260.03</td>
<td>95.67</td>
<td>99.41</td>
</tr>
</tbody>
</table>
Table 20. Coefficients of annual principal components comprising 9 variables (*E. coli* concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during study period (2 January 2018–1 January 2019) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients of PC1</th>
<th>Coefficients of PC2</th>
<th>Coefficients of PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>−0.03</td>
<td>0.55</td>
<td>0.00</td>
</tr>
<tr>
<td>Water Temp</td>
<td>−0.09</td>
<td>0.37</td>
<td>−0.14</td>
</tr>
<tr>
<td>DO</td>
<td>0.12</td>
<td>0.11</td>
<td>0.53</td>
</tr>
<tr>
<td>SPC</td>
<td>0.41</td>
<td>−0.37</td>
<td>−0.08</td>
</tr>
<tr>
<td>pH</td>
<td>0.09</td>
<td>0.57</td>
<td>0.06</td>
</tr>
<tr>
<td>Cl−</td>
<td>0.50</td>
<td>−0.12</td>
<td>−0.03</td>
</tr>
<tr>
<td>Mixed Development</td>
<td>0.55</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td>Agriculture</td>
<td>−0.15</td>
<td>0.02</td>
<td>−0.73</td>
</tr>
<tr>
<td>Forested</td>
<td>−0.47</td>
<td>−0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Table 21. Coefficients of annual principal components comprising 9 variables (E. coli concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during quarter one (winter: 2 January 2018–27 March 2018) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients of PC1</th>
<th>Coefficients of PC2</th>
<th>Coefficients of PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>−0.03</td>
<td>0.55</td>
<td>0.00</td>
</tr>
<tr>
<td>Water Temp</td>
<td>−0.09</td>
<td>0.37</td>
<td>−0.14</td>
</tr>
<tr>
<td>DO</td>
<td>0.12</td>
<td>0.11</td>
<td>0.53</td>
</tr>
<tr>
<td>SPC</td>
<td>0.41</td>
<td>−0.37</td>
<td>−0.08</td>
</tr>
<tr>
<td>pH</td>
<td>0.09</td>
<td>0.57</td>
<td>0.06</td>
</tr>
<tr>
<td>Cl−</td>
<td>0.50</td>
<td>−0.12</td>
<td>−0.03</td>
</tr>
<tr>
<td>Mixed Development</td>
<td>0.55</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td>Agriculture</td>
<td>−0.15</td>
<td>0.02</td>
<td>−0.73</td>
</tr>
<tr>
<td>Forested</td>
<td>−0.47</td>
<td>−0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Table 22. Coefficients of annual principal components comprising 9 variables (E. coli concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components, during quarter two (spring: 3 April 2018–26 June 2018) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients of PC1</th>
<th>Coefficients of PC2</th>
<th>Coefficients of PC3</th>
<th>Coefficients of PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>0.03</td>
<td>0.58</td>
<td>−0.09</td>
<td>−0.09</td>
</tr>
<tr>
<td>Water Temp</td>
<td>0.05</td>
<td>0.12</td>
<td>−0.12</td>
<td>0.91</td>
</tr>
<tr>
<td>DO</td>
<td>0.04</td>
<td>0.09</td>
<td>0.59</td>
<td>−0.18</td>
</tr>
<tr>
<td>SPC</td>
<td>0.36</td>
<td>−0.49</td>
<td>−0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>pH</td>
<td>0.17</td>
<td>0.60</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Cl−</td>
<td>0.51</td>
<td>−0.12</td>
<td>−0.01</td>
<td>−0.11</td>
</tr>
<tr>
<td>Mixed Development</td>
<td>0.56</td>
<td>0.09</td>
<td>0.15</td>
<td>−0.01</td>
</tr>
<tr>
<td>Agriculture</td>
<td>−0.12</td>
<td>0.03</td>
<td>−0.71</td>
<td>−0.23</td>
</tr>
<tr>
<td>Forested</td>
<td>−0.50</td>
<td>−0.12</td>
<td>0.31</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Table 23. Coefficients of annual principal components comprising 9 variables (E. coli concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components during quarter three (summer: 3 July 2018–25 September 2018) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients of PC1</th>
<th>Coefficients of PC2</th>
<th>Coefficients of PC3</th>
<th>Coefficients of PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>0.43</td>
<td>0.14</td>
<td>−0.29</td>
<td>−0.20</td>
</tr>
<tr>
<td>Water Temp</td>
<td>0.21</td>
<td>0.18</td>
<td>0.58</td>
<td>0.33</td>
</tr>
<tr>
<td>DO</td>
<td>0.30</td>
<td>0.22</td>
<td>0.48</td>
<td>0.16</td>
</tr>
<tr>
<td>SPC</td>
<td>−0.50</td>
<td>0.03</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td>pH</td>
<td>0.39</td>
<td>0.32</td>
<td>−0.20</td>
<td>−0.18</td>
</tr>
<tr>
<td>Cl(^-)</td>
<td>−0.48</td>
<td>0.21</td>
<td>−0.13</td>
<td>−0.12</td>
</tr>
<tr>
<td>Mixed Development</td>
<td>−0.16</td>
<td>0.63</td>
<td>0.01</td>
<td>−0.15</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.09</td>
<td>−0.13</td>
<td>−0.41</td>
<td>0.78</td>
</tr>
<tr>
<td>Forested</td>
<td>0.11</td>
<td>−0.57</td>
<td>0.26</td>
<td>−0.36</td>
</tr>
</tbody>
</table>

Table 24. Coefficients of annual principal components comprising 9 variables (E. coli concentration, water temperature, DO, SPC, pH, chloride, percentage of mixed developed land use, percentage of agricultural land use and percentage of forested land use) used to define 9 principal components during quarter four (fall: 2 October 2018–1 January 2019) in West Run Watershed, WV, USA.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients of PC1</th>
<th>Coefficients of PC2</th>
<th>Coefficients of PC3</th>
<th>Coefficients of PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>−0.03</td>
<td>0.55</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>Water Temp</td>
<td>−0.09</td>
<td>0.37</td>
<td>−0.14</td>
<td>0.79</td>
</tr>
<tr>
<td>DO</td>
<td>0.12</td>
<td>0.11</td>
<td>0.53</td>
<td>−0.01</td>
</tr>
<tr>
<td>SPC</td>
<td>0.41</td>
<td>−0.37</td>
<td>−0.08</td>
<td>0.19</td>
</tr>
<tr>
<td>pH</td>
<td>0.09</td>
<td>0.57</td>
<td>0.06</td>
<td>−0.14</td>
</tr>
<tr>
<td>Cl(^-)</td>
<td>0.50</td>
<td>−0.12</td>
<td>−0.03</td>
<td>−0.17</td>
</tr>
<tr>
<td>Mixed Development</td>
<td>0.55</td>
<td>0.18</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>Agriculture</td>
<td>−0.15</td>
<td>0.02</td>
<td>−0.73</td>
<td>−0.44</td>
</tr>
<tr>
<td>Forested</td>
<td>−0.47</td>
<td>−0.20</td>
<td>0.40</td>
<td>0.20</td>
</tr>
</tbody>
</table>
References


© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
CHAPTER 5: REVIEW OF FACTORS IMPACTING THE SURVIVAL AND OCCURRENCE OF ESCHERICHIA COLI IN SECONDARY MICROPHYSICAL HABITATS AND DISSERTATION CONCLUSION

Citation: Petersen, F.; Hubbart, J.A. In Preparation. Factors impacting the survival and occurrence of Escherichia coli in secondary microphysical habitats Water 2020, 12

Abstract: Escherichia (E.) coli is a fecal microbe that inhabits the intestines of endotherms (primary habitat) and the natural environment (secondary habitats). Due to prevailing thinking regarding the limited capacity of E. coli to survive in the environment relatively few published investigations exist regarding environmental factors influencing E. coli’s survival. To help guide future research in this area, an overview of factors known to impact the survival of E. coli in the environment is provided. Notably, the lack of historic field-based research holds two important implications: 1) large knowledge gaps regarding environmental factors influencing E. coli’s survival in the environment exists; and 2) the efficacy of implemented management strategies have rarely been assessed on larger field scales, thus leaving their actual impact(s) largely unknown. Moreover, the persistence of E. coli in the environment calls into question its widespread and frequent use as a fecal indicator organism. To address these shortcomings future work should include more field-based studies, occurring in diverse physiographical regions and over larger spatial extents. This information provides scientists and land use managers with new understanding regarding factors influencing E. coli concentrations in the secondary habitat, thereby providing insight into effectively address problematic fecal contamination.

Keywords: Escherichia coli; microphysics; water quality; land use practices; environmental persistence

1. Introduction

Escherichia (E.) coli is a fecal indicator microbe with a life history that cycles between two principal habitats, intestines of endotherms (primary habitat) and environmental water, sediment, and soils (secondary habitats). These habitats differ markedly with respect to physical conditions and nutrient availability [1]. For example, temperature remains relatively constant (approximately
37 °C) in the primary habitat but can vary greatly in the secondary habitat where annual average temperatures can range from below freezing (0°C) to approximately 18°C, or higher [1]. Additionally, the primary habitat is an anaerobic environment [2], whereas the secondary environment varies between aerobic and anaerobic (e.g. deep soil, sediment and water resources) [3]. Nutrients in the secondary environment are also typically less abundant, especially in soil and sediment [4] which, with respect to bacterial growth, are in a state of constant nutrient deficiency [5]. In water nutrients can vary from being abundant (e.g. receiving waters in agricultural areas) to scarce (e.g. open ocean) [6,7]. This contrasts primary habitat (i.e. colon) nutrient conditions, comprising consistently high nutrient levels that support rapid bacterial growth [1]. Consequently, the secondary habitat will place greater strain on the growth and survival of *E. coli*, as it was not the habitat the microbe predominantly populated and evolved in over the past few mega-annum. The conditions of the secondary habitat (environment) will therefore inhibit the growth and survival of *E. coli* when the microbe’s tolerance thresholds are exceeded. The tolerance thresholds of *E. coli* can be used to predict changes in *E. coli* concentrations based on changes in the conditions of the secondary environment. This can be used by scientists and land use managers concerned with microbial pollution in the environment.

When present in the environment “foreign” fecal microbes can pose considerable health risks to endotherms (human and animals) particularly if ingested [8]. Exposure to these microbes contributes significantly to morbidity and mortality in the global human population (2.2 million annual deaths) [9]. Therefore, managing the abundance of fecal microbes in the environment is important from a human health perspective. Managing fecal microbe concentrations (e.g. *E. coli*) and by extension the water quality of receiving waters, requires understanding of factors influencing the lifecycles and concentrations of these organisms in their secondary habitat. Therefore, understanding temperature and solar insolation influences, hydrologic requirements, chemistry and nutrient availability, and land use impacts on *E. coli* populations are critical for the proper implementation of effective management strategies.

Historically *E. coli* was thought to be poorly adapted to survival in the environment, and believed to comprise an average half-life of two days [1]. Additionally, it was believed that *E. coli* cells could only originate in the intestines of endotherms prior to excretion into the secondary habitat. The transfer from primary to secondary habitats was represented by the following two relationships [1]:

```latex
\text{Primary Habitat:} \text{Temperature} 37 \degree C \quad \text{Secondary Habitat:} \text{Temperature from below freezing (0°C) to approximately 18°C, or higher}
```

```latex
\text{Primary Habitat:} \text{Anaerobic Environment} \quad \text{Secondary Habitat:} \text{Aerobic and Anaerobic (deep soil, sediment and water resources)}
```

```latex
\text{Primary Habitat:} \text{Nutrients typically abundant} \quad \text{Secondary Habitat:} \text{Nutrients typically less abundant}
```

```latex
\text{Primary Habitat:} \text{Nutrient levels support rapid bacterial growth} \quad \text{Secondary Habitat:} \text{Inhibit growth and survival}
```
\[
\frac{dP}{dt} = \gamma P - \beta P
\]
(2)

And

\[
\frac{dS}{dt} = \beta P - \delta S
\]
(3)

Where P represents \textit{E. coli} populations in the primary habitat, S represents \textit{E. coli} populations in the secondary habitat, and \(\gamma\), \(\beta\) and \(\delta\) represent effective growth, bulk transfer, and death rate, respectively. Traditional thought surrounding \textit{E. coli}’s limited survival in the secondary habitat held two important implications: 1) deposition of new fecal matter was required to increase \textit{E. coli} populations in the secondary habitat, and 2) the probability of new host colonization was low [1]. Direct deposition of new fecal matter remains an important factor influencing fecal microbe concentrations in the environment, due to the direct proportionality between endotherm population density and fecal contamination (i.e. more organisms produce more waste) [10–12]. However, recent investigations have reported on the ability of \textit{E. coli} to persist (survive and reproduce) in the environment for extended periods of time, thereby increasing the likelihood of new-host colonization. This shift in understanding of the persistence of \textit{E. coli} in the secondary habitat implies that there may be ample time for the microbe becoming naturalized into the soil microbiome [13]. Once naturalized the microbe may become autochthonous and thus capable of surviving and reproducing in the environment without being directly deposited or replaced by \textit{E. coli} from animal feces or water [13]. Given the extended persistence of \textit{E. coli} and its widespread use as a fecal indicator organism [14,15], understanding the relationships between the environment and this microbe will aid land managers and scientists concerned with mitigation of microbial contamination in soil and water resources.

The objective of this review is to provide an overview of factors currently understood to affect \textit{E. coli} lifecycles and concentrations in the secondary habitat (environment), including how those factors can be used to quantitatively predict \textit{E. coli} population changes. A sub-objective was to supply a summary of contemporary management strategies aimed at reducing \textit{E. coli} and fecal microbe concentrations in the environment and identification of research needs.
2. *E. coli* and the Environment

2.1. Temperature

In the secondary habitat of *E. coli*, average annual temperatures range from approximately 0 to 47°C, and is typically considered the primary factor influencing survival, accounting for up to 61% of *E. coli* population variance (centered on inactivation rates) based on an Arrhenius model [16]. Given its relative importance, temperature must be accounted for when assessing local environmental parameters that influence *E. coli* lifecycles and concentrations. Thus, the growth limit and tolerance range of the microbe, 7 °C and -20 °C to 66 °C, respectively, provide insight as to how temperature can alter the bacteria’s lifecycle [17]. Previous investigations reported bacterial temperature dependencies, based on the first-order inactivation rates $k_c$, defined by the Chick equation [18,19] as,

$$C = C_0 e^{-k_c t}$$

Or alternatively

$$\ln C = \ln C_0 - k_c t$$

where $C$ represents bacterial concentration, $C_0$ represents initial bacterial concentration, $k_c$ represents the inactivation rate and $t$ represents time in days [19]. In the above equations the temperature dependence of bacterial inactivation ($k_c$) rates can be expressed using either the $Q_{10}$ or Mancini equations [19]. The $Q_{10}$ equation is defined as follows:

$$\frac{k}{k_*} = Q_{10}^{(T_c-T_c^*)/10}$$

where $k$ represents first order inactivation at temperature $T_c$, $k_*$ represents reference temperature first order inactivation, $Q_{10}$ represents rate of change in the inactivation rate due to temperature increases in 10 °C increments, $T_c$ represents temperature in °C and $T_c^*$ represents a reference temperature (usually 20°C) [19]. Conversely, the Mancini equation, frequently implemented in water quality models [19], can be defined as follows:

$$\frac{k}{k_*} = \theta^{T_c-T_c^*}$$

where $k$ represents first order inactivation at temperature $T_c$, $k_*$ represents reference temperature first order inactivation, $\theta$ represents temperature sensitivity of the microbe, $T_c$ represents temperature in °C and $T_c^*$ represents a reference temperature (usually 20°C). Consequently, $Q_{10}$ can be converted to $\theta$ as follows:
\theta = Q_{10}^{1/10} \tag{8}

Additionally, the temperature dependence of first-order reaction rates can be found through the Arrhenius equation:

\[ k = A e^{(-\frac{E_a}{RT})} \tag{9} \]

Where \( k \) represents the kinetic rate constant at temperature \( T \), \( A \) represents the constant pre factor (collision frequency factor), \( E_a \) represents activation energy, \( R \) represents the gas constant and \( T \) represents temperature in Kelvin.

Notably, environmental factors such as water purity can influence \( E. coli \) inactivation rates, as pristine water usually comprises higher average \( Q_{10} \) values but lower first order inactivation rates at 20 °C. For example, based on a review of 450 \( E. coli \) survival datasets from 70 peer reviewed investigations by Blaustein et al. [19] pristine water (defined by the authors as water originating from caves or springs and including less impurities) comprised an average \( Q_{10} \) of 2.066 ± 0.190 and an average first order inactivation rate at 20 °C (\( k_{20/\text{day}} \)) of 0.063 ± 0.007. Conversely, groundwater, agricultural waters and waste water comprised the following \( Q_{10} \) and \( k_{20/\text{day}} \) values: 1.783 ± 0.702, 0.504 ± 0.136; 1.548 ± 0.161, 0.388 ± 0.024; and 1.358 ± 0.238, 0.672 ± 0.114, respectively [19]. Consequently, the source of water can be used to approximate inactivation for contained bacteria. However, the impacts of biological and physical survival factor variations can cause variability in site and source specific \( E. coli \) survival rates at the same temperature [19]. Therefore, site specific data (e.g. physicochemical parameters) would greatly improve the accuracy of predictions regarding \( E. coli \) inactivation and survival in water resources.

Based on laboratory investigations, \( E. coli \) was shown to survive for up to 32 days in soil when incubated at 15°C [13]. In situ investigations also support the extended survival of \( E. coli \) in soils, even at colder temperatures. For example, \( E. coli \) in soil samples extracted from soils surrounding Lake Superior, USA in October 2003 and April 2004 had a 92% DNA fingerprint similarity [13]. This high similarity indicates these \( E. coli \) strains became naturalized (defined as the process by which non-native \( E. coli \) becomes integrated into the secondary habitat and reproduces at a sufficient rate to maintain its population[20]), autochthonous members of the soil microbial community, capable of enduring the cold winter months (including numerous freeze-thaw cycles) and growing during the warmer summer months [13]. Additionally, the growth and replication of \( E. coli \) in soils has been verified by laboratory studies. For example the bacteria grew to high cell densities (4.2 × 10^5 colony forming units (CFU)/g soil) when incubated at 30°C to 37°C in
nonsterile soils [13]. Additionally, when $9 \times 10^2$ CFU/g soil E. coli was inoculated and incubated for 16 days at 15°C for prior to a temperature increase (37°C for 8 days), cell density decreased to $1.1 \times 10^2$ CFU/g soil (during the 16 day period), before increasing 10-fold to $1.04 \times 10^3$ CFU/g soil 4 days after the temperature increase [13]. However, there was a subsequent decrease in cell numbers to $7.7 \text{ CFU/g soil}$ 8 days after the temperature increase, attributed to nutrient depletion (see nutrient and nutrient availability section) [13]. Soil E. coli population density is also subject to seasonal variation with the highest cell densities, up to $3 \times 10^3$ CFU/g soil, reported in warmer months (summer to autumn), and the lowest numbers, $\leq 1 \text{ CFU/g soil}$, reported in the colder months (winter to spring) [13]. Consequently, E. coli growth is directly influenced by soil temperature fluctuations, with rapid growth increases occurring as soil temperature rises from to 15°C (no growth) to 37°C [13]. Ultimately, given its reported influence on bacterial survival and growth [13,16,21], temperature constitutes a vital environmental factor influencing E. coli’s survival in the secondary habitat.

2.2. Solar Insolation

Research regarding the impact of solar insolation on fecal bacteria (e.g. E. coli) survival and inactivation, have been predominantly focused on marine waters [22–24]. However, in a study conducted at Lake Michigan, USA day length and exposure to insolation during sunny days resulted in an exponential decrease in E. coli counts [25]. Additionally, diminished E. coli inactivation was reported during cloudy days [25]. Cloud cover therefore played an important role in determining E. coli concentrations. For example, E. coli concentrations frequently exceeded safe swimming criteria (threshold E. coli concentration in water at which the bacteria becomes hazardous to human health, approximately $>235 \text{ CFU 100 ml}^{-1}$) during partly cloudy or completely cloudy conditions, but rarely exceeded this threshold during sunny conditions [25]. Furthermore, the submersion depth of E. coli in Lake Michigan also impacted decay rates. For example sunny condition decay rates at 45cm and 90cm depths were $Y_{45} = 48091e^{-0.4682t}$ and $Y_{90} = 12746e^{-0.4184t}$, respectively where Y represents E. coli concentration (CFU 100 ml$^{-1}$)and t represents time (hours) [25]. The impact of insolation on E. coli inactivation was also subject to initial bacterial concentrations, with higher concentrations having quicker decay rates [25]. Moreover, lake E. coli density could be more accurately predicted by exposure time (dosage) than insolation [25]. Thus, the impact of extended periods of insolation exceeded the effect of
intense insolation over shorter time periods. In the Lake Michigan study insolation was the predominant abiotic factor influencing *E. coli* inactivation, accounting for 40% of the variance as opposed to 7% by temperature and 8% for relative lake level [25]. Therefore, the results from this study challenges the assumption that temperature is the primary factor influencing *E. coli* survival, specifically at the surface (upper 90 cm) of freshwater bodies. Consequently, in shallow streams and headwaters, the inactivation of *E. coli* could be primarily driven by insolation and not temperature.

2.3. Suspended and Settled Solids

The survival of *E. coli* in water can be influenced by suspended solids concentrations in terms of how readily microbes can attach to those particles [26]. Association can increase nutrient and organic matter availability, particularly when the suspended solids include organic material (e.g. fallen leaf litter), while also providing optimal light exposure [27,28]. In addition, the close proximity of suspended particle associated microbes to each other can facilitate the horizontal transfer and proliferation of resistance genes [29,30]. The horizontal transfer of genetic material can be expedited when two microbes come into close contact with each other and remain that way until the transfer of genetic material is completed [29,30]. Thus, if two or more microbes associate with the same suspended particle, the likelihood of horizontal genetic transfer increases relative to free-floating microbes [29,30]. If resistance genes are transferred in this manner, over time the microbial population may display increased resistance to stressors such as chemical disinfectants, excessive photosynthetically active radiation (PAR) radiation, ultraviolet (UV) radiation, and predation [31–33]. However, the effect of suspended solids, including sediment, on the inactivation and survival of *E. coli* in the secondary habitat (the environment) is yet to be quantified as the majority of studies that attempted to quantify this relationship also include temperature fluxes which have a greater impact on *E. coli* variance [16,34]. Consequently, no equations are available that relate changes in suspended solids to associated changes in *E. coli* concentrations. Nevertheless, given current understanding decreased suspended solids in receiving waters will decrease *E. coli* survivability, thereby decreasing concentrations of this microbe.

The survivability of *E. coli* in settled sediments have been quantified [35] using an exponential die-off model based on the Chick [18] equation:

\[ \ln C = \ln C_0 - \mu_t \] (10)
where C represents initial concentration; C₀ represents current concentration; μ represents inactivation rate; and t represents time. Different inactivation rates have been reported by previous investigations. For example, sediment from the lakes of the eastern United States comprised an E. coli inactivation rate of approximately 0.54 d⁻¹ [35], whereas sediments of Southern Ontario creeks included an inactivation rate of 0.15 d⁻¹ [36] and the Hillsborough River in Florida comprised a inactivation rate of 0.07 d⁻¹ [37]. Like other environmental (secondary habitat) variables (see temperature section) temperature influences the die-off of E. coli in sediment, with more rapid die-off occurring at warmer temperatures [35]. Additionally, sediment particle size impacts temperature driven die-off rates, with survival rates being less sensitive to temperature in the finer soil. Garzio-Heardick et al. [38] reported the temperature driven die off rates relative to soil types as follows:

Sand: \( \mu = 0.109 \times 1.133^{T-20} \)

Sandy Loam: \( \mu = 0.051 \times 1.105^{T-20} \)

Silt Loam: \( \mu = 0.046 \times 1.054^{T-20} \)

where \( \mu \) represents the die off rate and T represents the temperature in °C. Ultimately, E. coli’s survival in settled sediment will vary geographically, due to changes in both temperature and physical soil characteristics. Therefore, to understand E. coli survival (including changes in survival due to land use changes or mitigation strategies) in sediment at a specified location, site specific information would be required, to avoid broad assumptions potentially leading to prediction inaccuracies.

2.4. Hydrologic conditions

Intense precipitation and subsequent runoff events can increase pollutant transport, thereby deteriorating surface water quality by increasing turbidity, suspended solid concentrations, organic matter and fecal contamination during stormwater discharge events [11,12]. Similarly increased overland and streamflow, during storm events, have been linked to increased E. coli concentrations relative to baseflow conditions [39,40]. The magnitude of the E. coli concentration increase varies between 15-fold [41] to 1000 fold [12], such that the increase can be represented by the formula below:

\[ C_s \geq C_0 k_s \] (11)
Where, $C_s$ and $C_0$ represent storm and base flow $E. coli$ concentrations, respectively and $k_s$ represents coefficient of increase ranging between 15.8283 - 1000. Factors impacting $E. coli$ concentrations during storm-generated overland flow includes rainfall intensity and duration, upland agricultural manure application, type and age of fecal deposits, and $E. coli$ adsorption to soil particles [12]. The coefficient of increase ($k_s$) is subject to change based on these factors. Moreover, the relationship between streamflow and $E. coli$ concentration is not linear, as increases in discharge during stormflow may dilute $E. coli$ concentrations. For example, in systems where the contribution of groundwater flow to stream flow is high, storm events may result in a decrease in receiving water $E. coli$ concentrations [12]. This is due to groundwater typically comprising low $E. coli$ concentrations [42,43]. However, while groundwater can dilute stream water, high groundwater contributions to streamflow can result in increased bed and bank shear stress, increasing resuspension of streambed sediment and elevating $E. coli$ numbers [44]. This resuspension can account for approximately 11% of total $E. coli$ load during storm events [44].

The concentration of $E. coli$ that can become resuspended can be calculated as follows [45]:

$$R_0 = C_s \times E_0 \left( \frac{\tau_b - \tau_{cn}}{\tau_c - \tau_{cn}} \right)^{n_a} \quad (12)$$

where $R_0$ represents resuspended $E. coli$ (CFU/m^2S); $E_0$ represents erosion rate (cm/s); $\tau_b$ represents bottom shear stress caused by water flow (Pa); $\tau_{cn}$ represent the critical shear stresses (N/m2) of non-cohesive sediments; $\tau_c$ represents the critical shear stresses (N/m2) of cohesive sediments; and $n_a$ represents particle size (diameter < 432 µm, $n_a$=2) [45]. The $\tau_b$ can be calculated using the specific gravity of water, $\gamma$ (N/m3), hydraulic radius, $R$ (m), and water surface slope, $S$, (m/m) ($\tau_b = \gamma . R . S$). Conversely, $\tau_{cn}$ can be calculated using particle size, $d$ (m), ($\tau_{cn} = d . 4.14 \times 10^{-3}$), and $\tau_c$ can be calculated using Lick’s equation as follows [45]:

$$\tau_c = \tau_{cn} \{(1 + a \epsilon b_c p^d) \frac{1}{d^2} \} \quad (13)$$

where $a$ and $b_c$ represent constants of $8.5 \times 10^{-16}$, and of 9.07 cm^3/g respectively, $p$ is water density (M L^{-3}) and $d$ is particle size. Resuspension during high flows can be driven by three resuspension mechanisms [46], 1) a steep-fronted wave (caused by the influx of water entering and flowing through a given stream during a precipitation event leading to a flashing leading hydrograph edge), with a wave height in excess of the preceding water depth, can lift microbes from the bottom sediment, holding them in the turbulent wave front [35,46], 2) a less steep front or falling wave can resuspend microbes without maintain them in the wave overrun [35,46], and 3) high flow...
turbulence (decrease in laminar flow due to increased kinetic energy) can cause steady-flow stochastic erosion of bed and bank sources, thereby maintaining elevated microbe concentrations relative to periods of lower flow [35,46]. Currently, equations that describe general relationships between increased *E. coli* concentrations or survival due to storm flow induced increased suspended solids or increased turbidity are lacking in the literature. This is attributable to the high geographic variation and subsequent site specificity of this relationship. Subsequently, compiling site-specific data from diverse geophysical environments would be useful for the development of general equations relating streamflow changes to *E. coli* concentrations and survival.

2.5. Water chemistry

Few published investigations are available regarding the influence of water chemistry on *E. coli* in the environment, therefore laboratory investigations are most often relied on and extrapolated to determine the growth limits of *E. coli* regarding water chemistry variables. However, the sole impact of water chemistry variables of *E. coli* is obscured by the inclusion of temperature as an independent variable in addition to the chemical aspect being investigated, by the majority of previous investigations [16,47–49]. These studies invariably conclude that ambient temperature greatly influences water chemistry impacts on *E. coli*, as for example *E. coli* can tolerate lower pH at higher temperatures [47]. Additionally, Presser et al. 1997, reported the effects of temperature, pH, water activity and lactic acid and concluded that these factors were synergistic in limiting *E. coli* growth [47]. In this investigation water activity (defined as the partial vapor pressure of water in a substance divided by the standard state partial vapor pressure of water) values of 0.985 and 0.975 and temperatures ≥ 25°C, resulted in a minimum *E. coli* growth pH of approximately 4, however temperature decreases raised the minimum pH slightly [47]. Subsequently the growth rate equations presented below assume a constant temperature as temperature fluctuations could impact water chemistry and *E. coli* relationships and alter growth and survival thresholds.

Previous investigations reported optimal *E. coli* survival between pH 5 and pH 7 with increased acidity or alkalinity resulting in decreased survival [50]. The growth limit of *E. coli* is at approximately a pH of 4, however *E. coli* in stationary phase (period during which the number of viable bacteria cells remains the same) can survive in pH 2 to 3 for several hours [51]. The following equation can be used to estimate the growth rate of *E. coli* based on pH, assuming the
growth rate is proportional to the amount by which the pH is more than the minimum value which prohibits growth [47]:

\[ rate = (c \times 10^{-pH_{\text{min}}})(\frac{10^{pH_{\text{min}}}-10^{-pH}}{10^{pH_{\text{min}}}}) \]  

(14)

where rate represents \( E. \ coli \) growth rate; \( c \) represents a constant of proportionality; \( pH_{\text{min}} \) represents minimum growth tolerable pH; and pH represents ambient pH [47]. Similar equations have been created for other chemistry variables constituting linear relationships with \( E. \ coli \) growth. For example, the growth rate of \( E. \ coli \) as determined by organic acid concentration can be estimated as follows:

\[ rate = c' \times (C_{\text{min}} - C) \]  

(15)

where \( c' \) represents a proportionality constant; \( C_{\text{min}} \) represents the theoretical minimum growth inhibitory concentration of the organic acid; and C is the measured concentration of the organic acid [47]. The equation is predicated on the assumption that the growth rate is proportional to the amount by which the concentration of the organic acid is less than the minimum concentration which prevents growth. Applying similar assumptions analogous equations can be created for inorganic acids, water activity chloride and salinity. The concentrations of inorganic and organic acids capable of preventing the growth of \( E. \ coli \) is dependent on the specific chemical under consideration. However, general minimum growth inhibitory concentrations for water activity, chloride concentration and salinity concentrations are 0.95 [52], 1.5 mg/L [53] and 20% NaCl (complete die-off in 72 hours) in nutrient rich media and 3.5% NaCl (limit for growth) in nutrient depleted media [54], respectively.

2.6. Nutrients and nutrient availability

Environmental nutrient conditions also impact \( E. \ coli \) growth and survival in the secondary habitat. For example, previous investigations reported \( E. \ coli \) populations were three times greater in soils rich in organic matter relative to nutrient depleted sandy soils, suggesting that soil nutrients and organic matter facilitate growth of fecal bacteria (e.g. \( E. \ coli \)) [55]. Additionally, \( E. \ coli \) cell density incubated at 30°C and 37°C in soils decreased following rapid initial cell growth [13]. This indicates that the final population of the soil \( E. \ coli \) was determined predominantly through either the exhaustion of bioavailable nutrients or predation [13]. Nutrient limitation on \( E. \ coli \) growth was also evident in laboratory studies where \( E. \ coli \) growth in M9 (minimal growth) medium
without C and N was limited to a less than one log increase in CFU [21]. In addition to nutrients, soil water potential can also influence the growth of E. coli in soil, due to its impact of nutrient availability and bacterial movement. [21]. For example, regression analysis from previous work indicated that E. coli growth (population doubling time) at 37°C was significantly related to soil water potential ($r^2=0.70$, $p < 0.001$) [21]. Soil water potential can also impact the motility of microbes in soil, as water potential less than −0.1 MPa, results in negligible bacterial movement, and decreased solute diffusion (half the rate observed under saturated conditions) [56]. Therefore, low soil water potential (−1.5 or −0.1 MPa) can cause limited nutrient supply for bacterial growth and survival [21]. Ultimately, nutrient abundance and availability (as determined factors such as soil water potential) constitute important factors impacting the survival and growth of E. coli in the environment.

2.7. Land-use practices

Previous work has linked land use practices including agricultural and urban land uses to increased E. coli concentrations in associated receiving waters [14,26,57]. In agricultural regions increased E. coli concentrations are primarily driven by manure applications [58] and the population density of endotherms (livestock) [10,12,59]. During manure application, the environmental inactivation or die-off of E. coli results in decreasing concentrations with time passed since application. This can be estimated using the equation below:

$$C = C_i - C_i (R_d \times t)$$  \hspace{1cm} (16)

where $C$ represents the current E. coli concentration at the soil surface; $C_i$ represents the initial soil E. coli concentration immediately after manure application; $R_d$ represents the die-off rate (or rate of inactivation); and $t$ represents time since the manure was applied. This equation can be adapted to estimate E. coli concentrations in associated receiving waters if the E. coli transferal rate between the soil and associated receiving waters is known or assuming that 89% of stream E. coli concentrations originates directly from surface run-off, as per Ribolzi et al. [44]. Conversely, in agricultural areas comprised primarily of the rearing of livestock the animal population numbers can be used to approximate the input (addition) of E. coli as follows [12]:

$$I = c \times \frac{(N_p \times W_i)}{A}$$  \hspace{1cm} (17)

where $I$ represents the new input of E. coli in a specified area over a specified time period; $c$ represents a constant (which varies based on the type and size of the animals under consideration,
assuming larger animals will produce more waste, contributing more \(E. coli\) to the area); \(N_p\) represents the population number of the animals (endotherms); \(W_i\) represents the waste per individual animal and \(A\) represents the area which the animals inhabit. This equation only approximates new additions of \(E. coli\) over specified period and does not approximate total \(E. coli\) concentrations, bacterial persistence, bacterial inactivation, or die-off rates. Notably, if the area that an animal population inhabits can be expanded the concentration of new \(E. coli\) inputs will decrease, due to the inverse proportionality between land use area and \(E. coli\) concentration based on animal waste. Thus, the concentrations of \(E. coli\) inputs can be reduced without reducing animal population numbers simply by increasing animal grazing area. The above equation can be adapted to estimate contributions of newly deposited \(E. coli\) to associated receiving waters if the \(E. coli\) transferal rate between the soil and associated receiving waters is known or assuming that 89% of stream \(E. coli\) concentrations originates directly from surface run-off, as per Ribolzi et al. [44].

Predicting \(E. coli\) population increases from urban areas is more complex as \(E. coli\) concentrations are elevated due to two primary reasons: 1) leaking wastewater infrastructure [10,11,57] and 2) increased run-off during precipitation events due to increased impervious surfaces [10,60]. Predicting leaks from wastewater infrastructure is difficult and therefore the quantitative effect of leaks in urban areas are rarely accounted for, despite potentially contributing significantly to \(E. coli\) concentrations in receiving waters [61]. In a simplified form, the contribution of a leak on the \(E. coli\) population in a region \((C)\) will be impacted by the fecal concentrations of the leak \((C_l)\); the specific discharge \((q_l)\); and the removal of bacteria in the soil due to the filtering effect of the soil (calculated by the removal rate divided by the distance flowed) \((f_r)\):

\[
C = C_l \times q_l \times f_r \tag{18}
\]

assuming saturated conditions the specific discharge of the leak can be calculated with Darcy’s Law [62]:

\[
q_l = \frac{Q_l}{A_l} = -K_{ht} \frac{dh}{dl} \tag{19}
\]

where \(q_l\) represents the specific discharge of the leak; \(Q_l\) represents the volume discharge of the leak; \(K_{ht}\) represents the hydraulic conductivity and \(dh/dl\) is the gradient of total hydraulic head [62]. Conversely, spatial data software and models such as the fate transport advection dispersion overland flow modeling approach can be used to predict the effect of land cover changes on the transport of pollutants (including \(E. coli\)) during overland flow from precipitation events [63–65].
These models, also known as mass balance water quality models, utilize annual average export coefficients from land use and land cover data to estimate in stream pollutant (e.g. *E. coli*) loadings [63–65]. An example of an overland flow model is the St. Venant equation [66]:

\[
\frac{\partial h(x,t)}{\partial t} + \frac{5(S_0)^{\frac{1}{2}}}{3n_s} \frac{\delta h(x,t)}{\delta x} = f(t) - i(t)
\]  

(20)

and

\[
q = \frac{(S_0)^{\frac{1}{2}}}{n_s} \frac{5}{h^3}
\]  

(21)

where \(h(x,t)\) represents overland flow depth; \(q(x,t)\) represents overland flow discharge per unit width; \(f(t)\) represents rainfall rate; \(i(t)\) represents infiltration rate; \(n_s\) represents Manning’s roughness coefficient; and \(S_0\) is channel slope [66]. Notably, complex hydrological equations can be used to predict vertical flow in soil and flow in unsaturated groundwaters, which can include the transport of *E. coli*.

3. Mitigation Strategies

Current mitigation strategies surrounding freshwater *E. coli* contamination include minimizing the transport of *E. coli* during overland flow or reducing sources of *E. coli*. Strategies include 1) vegetation management; 2) restricting livestock grazing and movement; 3) altering manure application strategies and 4) wastewater infrastructure maintenance. The maintenance of adequate vegetation or use of vegetative filter strips (strips of vegetation planted for the sole purpose of reducing pollutant transport during run-off events) can reduce the rate and energy of run-off thereby reducing the concentration of pollutants transported to receiving waters [12,67]. Given that 89% of stream *E. coli* concentrations result from overland flow [44], a reduction in the transport of *E. coli* from soil surfaces to associated receiving waters will proportionately decrease the concentration of the microbe in the water. Restricting the movement and grazing of cattle, using temporary fencing or active herding will reduce the amount of fecal matter, including *E. coli*, that is deposited in a specified area over a given period [68]. For example, McDowell et al. [68] reported that restricting the grazing time of dairy cows to three hours decreased *E. coli* concentrations in associated receiving water to below water quality guidelines (126 CFU per 100mL). Limiting the use of manure in the growing of crops can also decrease *E. coli* concentrations, due to a reduction in the sources of the microbe. Warnemuende and Kanwar [69] investigated the effects of swine manure application on bacterial quality of leachate and reported
that “an increase in application rate is more likely to cause greater bacterial contamination”. Therefore, limiting application rate (frequency) of manure can improve microbial water quality and decrease *E. coli* concentrations and population numbers in associated receiving waters. Notably, very few large-scale field-based case studies investigating the effect of varying manure application on *E. coli* or fecal concentrations currently exists in the literature. Thus, the true effectiveness of this form of mitigation remains largely unknown. The same holds true for the precise effect of frequent and proper maintenance of wastewater infrastructure in urban land use areas. Due to the sporadic and unpredictable occurrence of leaks it is hard to quantify their exact effect on *E. coli* concentrations. However, studies have reported that in developing nations, leaking wastewater infrastructure contributed significantly to *E. coli* concentrations in urban receiving waters, specifically during storm events [10,11,61].

4. Future Directions

Currently, there exists a general lack of field based investigations studying the environmental factors impacting *E. coli*’s survival in the secondary habitat [70]. Consequently, current understanding is predominantly based on laboratory studies, including many uncertainties and excluding factors that could influence identified relationships in uncontrolled environmental settings. For example, it is known that suspended solids can increase the persistence of *E. coli* in water resources [36], yet no widely accepted equations are available that relate changes in suspended solids to associated changes in *E. coli* concentrations. Additionally, the laboratory or small field scale methods implemented by previous investigations [69] implies that large scale work is needed, especially given the processes governing *E. coli* survival and concentrations could differ at different spatial scales. For example, animal population density could be a key consideration in a small agricultural area [59] but could become irrelevant if the surrounding area comprises of large cities. Site specific modelling techniques, including geophysical characteristics, need to be developed and applied to a wide variety of areas under varying climatic conditions, to improve understanding of the dynamics of the different water masses and their associated *E. coli* concentrations [12]. This type of work is specifically needed given the contrasting roles of groundwater outflow on stream *E. coli* concentrations during precipitation events: 1) the dilution of stream *E. coli* concentrations; 2) and the resuspension of streambed *E. coli* [12]. Having more fine scale data also constitutes a fundamental requirement for improving predictive *E. coli* models.
and process-based understanding of *E. coli* concentration fluctuation in the secondary habitat. Finally, the identification of naturalized soil *E. coli* communities calls into question the microbes use as an indicator of fecal contamination [13,21]. Subsequently, studies are required investigating the movement of naturalized *E. coli* strains and their relative contribution to stream *E. coli* concentrations. Environmental effects on this relative contribution also warrant further investigation, as changing climatic or microphysical conditions could potentially alter the movement of both naturalized and newly deposited *E. coli* to associated receiving waters.

The impact of other microbes, including bacteria, fungi, nematodes, and viruses on both naturalized and newly deposited *E. coli* concentrations should also be in investigated, especially for pathogenic strains of *E. coli*, such as the O157:H7 serotype (71). Interspecific competition and interaction can impact the survival of *E. coli* populations in the secondary habitat. For example, when *E. coli* is competing with more well-adapted microbes, including competition between different strains of *E. coli* for limited resources (e.g. space or nutrients), resulting in the decline of *E. coli* populations as the microbe is out competed (72, 73). Previous works have reported differential antimicrobial activity in soils, due to soil bacteria, after exposure to *E. coli* (74). The secretion of the antimicrobial substances can decrease the survival of *E. coli* populations in the secondary habitat. Similarly, other microbes (e.g. bacteriophages) predating on *E. coli* can also cause population decline and act as a form of population control (75). Consequently, increasing understanding of interactions between secondary habitat microbial communities and *E. coli* can potentially provide insight into novel *E. coli* management strategies, similar to biological control used for macroorganisms (76).

5. Dissertation Conclusion

The paucity of field-based research investigating the relationships between *E. coli* (fecal matter) concentration, suspended particulate matter (SPM; including size distribution), physicochemical parameters (water temperature, pH, SPC, DO, chloride) and land use practices and the importance of these relationships, particularly for human health, policy makers, and water resource managers, provided the impetus for this dissertation research. A 22-site, nested-scale, experimental watershed study design, including a distinct four-site study, was implemented to investigate *E. coli* concentrations and the aforementioned relationships in a contemporary mixed land use watershed in the Appalachian region of the eastern United States. Results from the four-site study showed a
statistically significant correlation of SPM ($\alpha = 0.05$) with *E. coli* concentration. SPM in the <5 µm interval was also identified as the most important as 1) more than 90% of *E. coli* data from all four sampling locations were found in this smallest interval, 2) this interval was the most highly correlated with *E. coli* concentration data ($\alpha < 0.0001$), and 3) increasing SPM concentrations in the ≤5 µm interval showed corresponding increases in relative *E. coli* concentration. The implication of these results is that *E. coli* principally remains free floating or attaches to particles <5 µm in size in receiving waters. The annual duration of the 22-site work revealed elevated *E. coli* concentrations in agricultural sub-catchments (avg. 560 CFU per 100 mL) compared to mixed development (avg. 330 CFU per 100 mL) and forested (avg. 206 CFU per 100 mL) sub-catchments. Additionally, annual *E. coli* and SPM concentration data displayed a statically significant relationship ($p < 0.01$) in agricultural areas. However, quarterly Spearman’s correlation coefficient (SCC) results indicated temporal fluctuations between significance ($p < 0.05$) and insignificance ($p > 0.05$) in the correlations between *E. coli* and SPM concentrations across all land use classes. Due to these temporal fluctuations resulting in periods of insignificant correlations between SPM and *E. coli* concentrations, SPM could not be used as an accurate proxy for fecal contamination. The work also highlighted the impact of historic land use practices (specifically mining which resulted in AMD) as there was an approximate 112% increase, over roughly 7km stream distance, in *E. coli* concentrations from the AMD impacted headwaters (avg. 177 CFU per 100mL) to the lower portion of the watershed (avg. 376 CFU per 100mL). Moreover, SCC results showed significant correlation ($p < 0.05$) between pH and *E. coli* concentrations at 77% of sample sites. The effect of temperature was therefore exceeded by pH in the watershed, as temperature was only significantly correlated at (64%) of sites. Furthermore, a pH tipping point (threshold) in the range of 7.68–7.76 was identified in the current investigation, with pH values below this range including significant negative correlations ($p < 0.05$) with *E. coli* concentrations. Consequently, receiving water pH values below the 7.68–7.76 tipping point will start significantly impacting (decreasing) active *E. coli* concentrations. A land cover tipping point of 25–30% was also identified for mixed development land use practices and significant ($p < 0.05$) negative correlations between *E. coli* and chloride concentrations. Consequently, study results indicated that the application of road salts in areas comprising mixed development exceeding 25% to 30%, may suppress fecal bacteria in receiving waters. The importance of seasonal variability on fecal concentrations in receiving waters was illustrated by temporal variability in quarterly PCA biplots
of *E. coli* concentrations, SPM concentrations, physicochemical parameters, and land use practices.

The results presented in this dissertation advanced understanding of land use practice (both historic and current), SPM (including size distribution) and physicochemical parameter influences on *E. coli* concentrations in contemporary mixed-land-use watersheds. Results contribute to the current understanding of fecal contamination regimes in contemporary mixed land use watersheds. For example, the work is among the first to provide evidence that *E. coli* remains predominantly free-floating or associates with SPM < 5µm in size in receiving water of contemporary mixed-land-use watersheds (Chapter 2). Results from the work also indicated that SPM could not be used as an accurate proxy for *E. coli* concentrations (fecal contamination), as correlations between SPM and *E. coli* concentrations varied (between significant and insignificant) across different land use practices and different quarters of the year (Chapter 3). Results also highlighted the impact of historic land use practices (AMD from historic mining) on *E. coli* concentrations, with pH exceeding the influence of temperature in WRW. Consequently, the results from this dissertation work can be used to inform policy makers and land use managers about effective water quality management, in watersheds with fecal contamination challenges.
References


64. van der Tak, L.; Edwards, C. An ArcView GIS Tool to Calculate Nonpoint Sources of Pollution in Watershed and Stormwater Projects; USEPA, 2001;


