Analysis of Water Quality Data Using Multivariate Statistics to Identify Patterns Associating Land Use to Fecal Pollution Sources

K.K. Hall, B.G. Evanshen, K.J. Maier and P.R. Scheuerman
Department of Environmental Health, East Tennessee State University
Johnson City, Tennessee

ABSTRACT

The fecal contamination of surface water bodies and their inclusion on impaired waters lists (303d) has resulted in the need to find methods that effectively and universally identify sources of fecal pollution. The Total Maximum Daily Load (TMDL) process is used to address these listings, but it is based on the limited 30-day geometric mean and does not take into consideration microbial and chemical processes that influence fate and transport of fecal indicators from their sources to receiving streams. To better understand current water quality, assess the extent of fecal pollution, and predict changes in water quality as development occurs, a water quality monitoring program was initiated using a targeted sampling approach. Four creeks within the Watauga River watershed in Northeast Tennessee were monitored based on the existence of TMDLs or the expectation that they will require a pathogen TMDL, while a TMDL for the Watauga River watershed has been approved. The data demonstrate that the creeks vary in their extent of fecal pollution (309 – 1175 CFU/100ml), and that land use patterns influence the microbial and chemical water quality parameters that are correlated with fecal indicator concentrations. This suggests that not only may TMDL development require multi-year data using a targeted sampling approach instead of a 30-day geometric mean, but the development of TMDLs for entire watersheds may be limited in their ability to effectively identify sources of fecal pollution throughout the entire watershed. The objective of this research is to update results of previous reports and apply multivariate statistical techniques to identify common patterns that influence the fate and transport of fecal indicators from various pollution sources. We will discuss the conclusion and usefulness of these data to accomplish this objective.

Keywords: water quality, TMDL, multivariate statistical analysis

INTRODUCTION

The rapid growth and urbanization that is occurring in many locations is increasing the already significant deterioration of surface water quality. Increased listing of surface water bodies on impaired waters (303d) lists for pathogen impairment and the need to address these through the Total Maximum Daily Load (TMDL) process as required under the Clean Water Act of 1972 has resulted in increased research to find methods that effectively and universally identify sources of fecal pollution. A fundamental requirement to identify such methods is understanding land use patterns, and microbial and chemical processes that influence the fate and transport of fecal indicators and pathogens from various sources to receiving streams. To better understand the chemical and biological factors influencing water quality and to predict water quality changes as development occurs, we initiated a water quality monitoring program.

Four creeks within the Watauga River watershed in Northeast Tennessee were routinely monitored using a targeted sampling approach to identify pollution sources and monitor remediation efforts. Boones, Buffalo, Cash Hollow and Sinking Creeks are listed on the State of Tennessee's 303d list due to pathogen loading (1), and a TMDL for *E. coli* was approved by the U.S. Environmental Protection Agency (USEAP) for the Watauga River watershed in 2006 to address these listings (2). TMDL development is currently based on a limited 30-day geometric mean, but this method does not take into consideration seasonal effects, variability in land use patterns, or the influence of runoff events on water quality. In addition, the interactions between chemical and microbial processes in the water add to the complexity of understanding pathogen loading, fate and transport within the watershed.

To account for these sources of variability, alternative methods of water quality monitoring and data analysis may be necessary to remove impaired waters from 303d lists. Examining these complex relationships using multivariate statistical tools such as canonical correlation and canonical discriminant analyses can help quantify nonpoint sources of pollution and improve our understanding of the influences of chemical and microbial processes on water quality to help effectively identify sources of fecal pollution. An extension of multiple regression analysis, canonical correlation analysis examines the linear combinations of variables in two or more data sets to determine the largest correlation between the data sets, providing a measure of the strength of association between data sets to help explain how chemical parameters influence fecal coliform loading, fate and transport in the watershed (3). Canonical discriminant analysis further identifies common patterns associating fecal pollution to sources and land use patterns by identifying the canonical variables that discriminate between groups based on the strength of their linear associations. Using this statistical technique, it is possible to identify common pollution sources based on the key discriminatory variables and associate them with specific land use patterns within the watershed. In addition to a targeted sampling approach, the identification of pollution sources using such data analysis tools can prove to be a cost-effective method for water quality monitoring and assessment to aid in effective TMDL development and the implementation of best management practices (BMPs).

OBJECTIVES

1. Assess the overall chemical and microbial factors that influence the water quality in the Watauga River watershed using a targeted sampling approach,
2. Apply multivariate statistical methodology to the collected data to better understand how chemical and biological factors are influencing water quality, and
3. Using this watershed as a model, determine the usefulness of this approach to identify common patterns associating these monitored water quality parameters to sources of fecal pollution.

MATERIALS AND METHODS

Site Description – The Watauga River watershed (*HUC 06010103*) is located in Northeast Tennessee in a region that is experiencing rapid growth in the form of new residential developments (Figure 1). Boones Creek contains 19.3 miles of impaired water due to pasture grazing and land development and Buffalo Creek contains 6 miles of impaired water due to pasture grazing. Sinking Creek and Cash Hollow Creek contain 10 and 3.4 miles of impaired water, respectively. Both creeks are impacted by a combination of pasture grazing and discharges from MS4 areas (1). Land use classification is based on human population density. Low human populations and the presence of agricultural activity characterize agricultural land use. High human population densities characterize urban land use, while low human populations and the lack of agricultural activity are characteristic of forested land use.

Sample Collection – . All creeks were sampled monthly during the first year and quarterly for every year after. Ten stations on Boones Creek were sampled from 2005 – 2008. Twelve stations on Buffalo Creek were sampled 2004 – 2008. Nine stations on Cash Hollow Creek and 14 stations on Sinking Creek were sampled from 2002 – 2008. Water samples for chemical analyses were collected in triplicate in 2L Nalgene™ bottles, and water and sediment samples for microbial analyses were collected in 100ml sterile Whirl-pak bags (Nasco, Fort Atkinson, WI). Water samples for Colilert® Quanti-Tray method were collected in sterile 100ml plastic bottles. All samples were transported to the laboratory on ice and analyzed within 6 hours of arrival.

Microbial Analyses – Analyses for the determination of total/fecal coliform concentrations, standard plate count (SPC) concentrations, and the Colilert® Quanti-Tray method were performed according to Standard Methods for Examination of Water and Wastewater (4). Microbial Enzyme Activity (MEAs) assays for acid phosphatase, alkaline phosphatase, dehydrogenase, galactosidase and glucosidase activities were also performed (5, 6).

Chemcial Analyses - Nitrate, phosphate, alkalinity and hardness analyses were performed in triplicate using colorimetric HACH™ methods as described by the manufacturer (7). All HACH™ analyses were performed in triplicate. Biochemical oxygen demand (BOD_5) analyses were conduced in triplicate using the YSI Model 5000 dissolved oxygen meter (YSI Inc., Yellow Springs, OH) according to Standard Methods for Examination of Water and Wastewater (4).

Statistical Analysis – Statistical analyses were performed using SAS/STAT software (SAS Institute, Cary, NC). Fecal coliform data were analyzed by ANOVA to compare concentrations by site and season. Canonical correlations were performed using the CANCORR procedure and discriminant analysis was performed using the CANDISC procedure. Canonical correlations and canonical discriminant analyses were performed at the season and site levels.

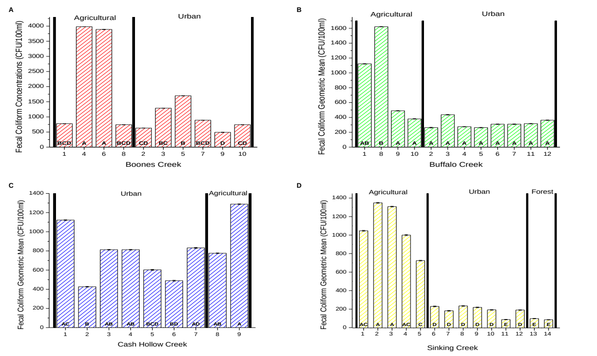

Figure 2. Fecal coliform concentrations in Boones (A), Buffalo (B), Cash Hollow (C) and Sinking (D) Creeks by site number with land use classifications

RESULTS

- Significant differences in fecal coliform concentrations are observed for Boones, Buffalo, Cash Hollow and Sinking Creeks within the Watauga River watershed (Fig. 3a). Seasonal variation of fecal coliform concentrations was also observed with the highest concentrations observed in the summer and fall months (Fig. 3b).
- Fecal coliform concentrations vary within and between creeks. Typically, the highest concentrations are observed at agricultural land use sites (Fig 4a-d).
- Yearly fecal coliform concentrations also display spatial and seasonal trends. All creeks exceed regulatory limits for fecal coliform concentrations (Fig. 5a) in each year with the highest fecal coliform concentrations observed during the summer and fall months (Fig. 5b).
- Canonical correlation analysis of water quality data indicates sources of fecal pollution vary within the watershed. Soil erosion primarily impact Boones and Buffalo Creeks . Cash Hollow Creek is impacted by urban runoff and Sinking Creek is impacted by a combination of soil erosion and urban runoff.
- Canonical discriminant analysis demonstrates a grouping effect of class means by site and land use pattern. A plot of the first two canonical variables displays the degree of discrimination between each group (Fig 6a-b). The canonical variables show strong separation between sites and land uses, demonstrating the strong influence of spatial specific conditions on water quality within and between creeks in the Watauga River watershed (Table 1).

CONCLUSIONS

- Fecal coliform concentrations exhibit spatial, seasonal and yearly variation. This variation suggests the influence of climate within and between years, and the effects of development and urbanization within the watershed. This variability within the Watauga River watershed also suggests that TMDL development may require multi-year data using a targeted sampling approach instead of a 30-day geometric mean.
- Fecal coliform concentrations vary between and within creeks. This difference is attributed to variation in land use patterns along the creeks. Agricultural land use sites demonstrate the highest fecal coliform concentrations, particularly in the summer and fall months.
- Canonical correlation analyses demonstrate that sources of impairment are related to land use patterns in the watershed and the sources of impairment differ within and between creeks. Soil erosion/organic matter input and urban runoff are the result of agricultural and/or urban activities including land disturbances and municipal sewage contamination.
- Canonical discriminant analysis further suggests the influence of land use patterns on fecal coliform concentrations within and between sites. The inclusion of some agricultural land use sites in the urban cluster suggests mixed sources of fecal pollution. The first canonical variable is influenced by alkalinity and hardness. The second canonical variable is influenced by total/fecal coliform, nitrate and phosphate concentrations. The monitored sites within the Watauga River watershed are clustered based on these variables, further indicating the influence of land use on water quality.
- The application of multivariate statistical methods to water quality data has helped to identify common patterns associating monitored water quality parameters to various pollution sources. Combined with a targeted water quality monitoring program, this data analysis approach is a useful method to identify sources of impairment and to identify BMPs that can prevent and remediate the effects of rapid urbanization.

BIBLIOGRAPHY

1. TDEC. 2010. *2010 303(d) Report, The Status of Water Quality in Tennessee, August 2010*. State of Tennessee, Department of Environment and Conservation, Division of Water Pollution Control. Nashville, TN.
2. TDEC. 2006. *Total Maximum Daily Load (TMDL) for E. coli in the Watauga River Watershed (HUC 06010103). Carter, Johnson, Sullivan, Unicoi, and Washington Counties, Tennessee*. State of Tennessee, Department of Environment and Conservation, Division of Water Pollution Control. Nashville, TN.
3. Johnson, R.A., D.W. Wichern. 1992. Applied multivariate statistical analysis. Prentice Hall, Englewood Cliffs, New Jersey
4. APHA, AWWA, and WPCF. 1995. Standard Methods for the Examination of Water and Wastewater, 19th ed. American Public Health Association, Washington D.C., USA.
5. Sayler, G.S. M. Puziss, M. Silver. 1979. Alkaline phosphatase assay for freshwater sediments: application to perturbed sediment systems. *Amer. Soc. Microbiol.* 38:922-927
6. Morrison, S.J., F.D. King, R.J. Bobbie, R.E. Beechtold, D.C. White. 1977. Evidence for microfloral succession on allochthonous plant littler in Apalachicola Bay, Florida, USA. *Marine Bio.* 41:229-240
7. Anonymous. 1997. HACH Water Analysis Handbook. 3rd ed. HACH Company Loveland, CO.

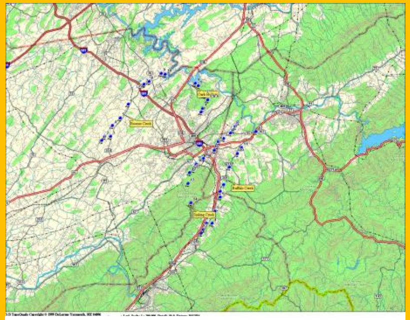

Figure 1. Map of Boones, Buffalo, Cash Hollow and Sinking Creeks with sampling sites and surrounding region

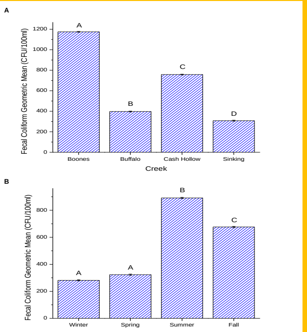

Figure 3. Mean fecal coliform concentrations by creek (A) and season (B)

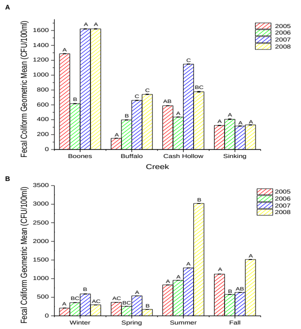

Figure 5. Fecal coliform concentrations by year

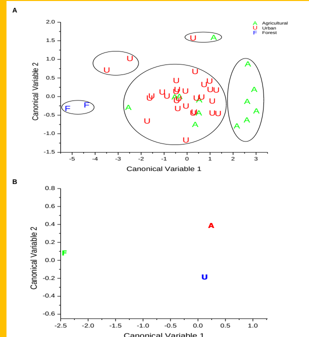

Figure 5. Class means on canonical variables by site (A) and land use (B)

Table 1. Description of Canonical Structure by Land Use

Canonical Variable	Water Quality Variables Describing Canonical Structure
Canonical Variable 1	Alkalinity (0.9453) Hardness (0.8874)
Canonical Variable 2	Total Coliforms (0.7492) NO_3 (0.4298) PO_4 (0.4278) Fecal Coliforms (0.3282) Fecal Coliforms in Sediment (0.3136)