

Identifying Sources of Fecal Pollution in the Appomattox River Watershed

Julie Michelle McKinney
Dr. Charles Hagedorn, Chair
Crop Soil and Environmental Sciences

ABSTRACT

Sources of *E. coli* were determined from impaired waterways in the Appomattox River watershed (in the lower Piedmont and South-Central Virginia) for the development of bacterial Total Maximum Daily Loads (TMDLs). The Appomattox River watershed is primarily undeveloped with 70.8% of the land forested, 17.0% used for agriculture (mainly livestock production), and 7.7% classified as water, wetland or barren land. The remaining 4.5% is developed for residential, commercial, and industrial land uses (mainly within the city of Petersburg).

Using Antibiotic Resistance Analysis, a known source library of 1,280 *E. coli* isolates (320 isolates per source) was constructed. Water samples were collected monthly for between eleven and fourteen months (11/02-12/03) from 40 locations throughout the Appomattox watershed and analyzed for fecal coliforms, *E. coli*, and resistance to 7 antibiotics of varying concentrations. A total of 486 water samples (9,907 isolates) were analyzed during the study. The objectives of this study were verify that each sampling site exceeded state bacterial count standards (using fecal coliform data), to compare the Discriminate Analysis and Logistic Regression statistical models for use in the classification of isolates, and finally to determine the source of contamination at each site.

The fecal coliform and *E.coli* data was used to determine if each site exceeded state standards during the assessment period. Thirty-eight of the sites exceeded the fecal coliform standard at least 10% of the time, and thirty-three exceeded the *E.coli* standard at least 10% of the time.

Discriminate Analysis (DA) is typically used to classify isolates, but the results obtained from the DA model were unrealistic based on the watershed land uses. By statistically analyzing the original 1,280 *E.coli* isolates six different ways, a more appropriate classification of isolates was determined. The six analyzing methods were Regular DA and Logistic Regression (LR); DA

and LR where each isolate whose probability fell below 80% was deleted; DA and LR where each isolate whose probability fell below 80% was used to create an Unknown category. The Logistic Regression model with an Unknown category proved to be the most appropriate. By using the Logistic Regression model, with Unknown category, to classify isolates, twenty five of the forty sites were discovered to be contaminated predominately with Livestock and fourteen of the sites predominately by Wildlife. One site was equally divided between these two categories. Human and Pet contamination were not dominant at any of the forty sites.

This comparison of the DA and LR statistical methods could change the analysis standard for Bacterial Source Tracking and suggests that the model required to classify isolates depends on the watershed characteristics.