

Chapter V

SUMMARY

The Appomattox Total Maximum Daily Load (TMDL) Development Project had three objectives: to verify that the sampling sites were impaired, to compare the Discriminate Analysis (DA) and Logistic Regression (LR) Statistical models to determine the most appropriate one to use with the Appomattox river and tributaries within the Lower Appomattox River Watershed, and to determine the sources of fecal pollution for the forty sampling sites. Although not part of this thesis project, the source results and fecal coliform data were used to develop TMDLs for sites throughout the Appomattox watershed.

The first objective, to verify that the sampling sites were impaired, was met by testing water from each site for fecal coliforms and *E.coli* once a month for one year. If a site exceeded the state standard 10% of the time during that assessment period, it was considered impaired. The fecal coliform standard was exceeded by thirty-eight of the forty sampling sites over 10% of the time and the *E.coli* standard was violated by thirty-three of the forty sampling sites over 10% of the time. The first objective was met by verifying that thirty-eight of the sampling sites were still impaired based on the 11-14 months of samples analyzed for this study.

In order to address objective two (to determine which statistical model, DA or LR, was the most appropriate) several variations of each regular (containing 1280 isolates divided into four categories: Human, Livestock, Pet, and Wildlife) library model (DA/ LR) were used to analyze ten water samples. First, each regular library was used to classify 10 randomly selected samples each with at least 272 isolates. Next, each library was altered by deleting any sources whose probabilities of classification were below 80% (called the DA/ LR 80% Delete library). As recommended by Simmons and Herbein (1995) an 80% cut was chosen because it appeared that many of the sources that were being misclassified had below an 80% probability of classification. After running the 10 randomly selected samples against each 80% Delete library, the source classification values were compared to the regular libraries. The final variation was carried out by placing the sources whose probabilities were below 80% into a fifth category, called the Unknown category. Samples were classified using the Unknown DA/LR libraries and compared to each

other, as well as the regular and 80% Delete libraries. The Logistic Regression library with Unknown category was the most appropriate choice because it provided results consistent with the land use patterns within the watershed.

Although the Logistic Regression library with Unknown category was the best choice for the Appomattox Total Maximum Daily Load Development Project, each library needs to be analyzed and considered separately. Because the watershed is heavily forested and rural, high Pet and Human contamination is not reasonable, although some Human contamination can be attributed to failing septic systems. Knowing that the land use around these sampling sites was predominately for agriculture and that the sites were heavily forested and rural allowed the most appropriate library to stand out. This illustrates an often-illusory step in Bacterial Source Tracking, using your knowledge of an area to cross validate the results. The second objective was met by the comparison of Discriminate Analysis and Logistic Regression libraries resulting in the formation of an appropriate library. Using the Unknown Logistic Regression library produced Rates of Correct Classification (RCC) for each category of 100% for Human, 99.51% for Livestock, 96.61% for Pet, 91.32% for Unknown, and 97.72% for Wildlife. The Average Rates of Correct Classification (ARCC) value was 97.02%. These values are higher than those typically reported in the literature. For example, Wiggins *et al* (2003) reported ARCC values ranging from 49%-81% from his multi-watershed study. However, Graves *et al* (2003) did obtain values close to those obtained for the Appomattox, 92.02% a 1,174 isolate library. The high ARCC values observed for the Appomattox River TMDL development project are possibly due to including too many isolates from a given source sample, thereby reducing the diversity of the library.

Once the appropriate library was determined, to satisfy objective three 486 water samples were analyzed from forty sampling sites and source classifications made for each of the 9,907 isolates. Although the Unknown category was dominant for twenty-two of the forty sites, the dominant identifiable source of fecal contamination was Livestock for twenty-six of the forty sites. When the Unknown category was the dominant source, it indicates that the bacterial strains isolated from the water samples were not adequately represented in the library. Fourteen sites resulted in Wildlife as the dominant identifiable source while Human and Pet were consistently

secondary contributors of the fecal pollution. These results were reasonable considering that land in the James River Watershed, the larger watershed of which the Appomattox is a part of, is 65% forested, 19% is used for agricultural purposes, and 12% is urbanized (James River Association 2002). Of the sites sampled, only eight could be considered to be located in urban surroundings. Objective three was met because a dominant source contributing to the fecal pollution at each station was identified and it was consistent with the land use patterns around the sites.

The decision to use the standard four categories (Human, Livestock, Pets and Wildlife) for the entire state of Virginia TMDL Development Project, as opposed to tailoring the categories to fit each specific watershed, was not a good idea. The Appomattox Watershed, as were many other watersheds in the Virginia TMDL Development Project, is rural and should not have a significant amount of pollution from Pets. Creating an unnecessary category leads to misclassifications, because there will inevitably be some similarities between strains of different species. Extra known sources for the existing three categories should have been collected to make the Appomattox library more representative. This may have reduced the number of isolates classified as Unknown. Pet resistance patterns need to be represented on a smaller scale, perhaps by placing a few source samples into one of the other categories.

The most important information that was gained from this study was the benefit of using an Unknown category for source classification. Because library size necessary for representativeness has not been standardized, it is important to have a safe guard to prevent misclassifications. In practice, standardization of library size may not be appropriate, as every watershed is unique. Using the Unknown category insures that only strong classification probabilities are used, and therefore more confidence in source classifications is obtained.

References

Graves, A. K., C. Hagedorn, A. Teetor, M. Mahal, A. M. Booth, and R. B. Reneau, Jr. 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. *Journal of Environmental Quality*. 31:1300-1308.

James River Association. 2002. *State of the James*. 52 pages. James River Association, Mechanicsville, Va.

Simmons, G. M., Jr., S. A. Herbein, and C. M. James. 1995. Managing nonpoint fecal coliform sources to tidal inlets. *Water Resources Update* 100: 64-74.

Wiggins, B. A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Hahn, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, G. G. McCarthy, J. A. McDonough, S. A. Merces, M. J. Noto, H. Park, M. S. Phillips, S. M. Perner, B. M. Smith, E. N. Stevens, and A. K. Varner. 2003. Use of Antibiotic Resistance Analysis for representativeness testing of multiwatershed libraries. *Applied Environmental Microbiology*. 69: 3399-3405.