CHAPTER I

INTRODUCTION

STUDY RATIONALE

The microbiological quality of U.S. waters is a topic of great concern and intense research. In 2000, the latest report of the National Water Quality Inventory reported that approximately 40% of streams, 45% of lakes, and 51% of estuaries were not clean enough to support recreational uses such as fishing and swimming. The leading causes of impairments to the rivers and streams are pathogens, sedimentation, and habitat alteration (Environmental Protection Agency (EPA), 2002a). The microbiological maintenance of waters that have primary or secondary contact with humans is imperative, as contamination can create serious health risks. Waters polluted with human feces are generally regarded as a greater risk to human health as they would be more likely to contain human specific pathogens.

Establishing the source of fecal contamination is crucial for the evaluation of the health risks as well as for the direction of remediation efforts. Direct monitoring of human pathogens has proven difficult because many pathogens are not readily detectable in the environment and of even greater concern are the pathogens that pose danger at low infectious doses (Scott et al., 2002). This has led to the use of indicator organisms that circumvent the need to test for every potential pathogen. Fecal coliforms, specifically *E.coli,* and *Enterococci* are commonly used indicators. Fecal coliforms are bacteria that can be found in the intestines of warm-blooded animals. They are a subset of the total coliforms and include *Klebsiella, Citrobacter, Enterobacter,* and *Escherichia*. These organisms are gram-negative aerobic or facultatively anaerobic rods and are distinguished by the ability to produce gas upon lactose fermentation. *Enterococci* are gram-positive facultatively aerobic

cocci. *Enterococci* are taxonomically separate from the genus streptococci, and include at least eighteen species (Holt, 1994). Enterococci are good indicators in marine environments due to the ability to survive in 6.5% salt (Holt, 1994).

The EPA has the responsibility for recommending water quality criteria to states (Table 1). Recent changes have been made to bacterial standard recommendations that now advise no more than 126 *E.coli* per100mL or 33 enterococcus per 100mL in freshwater based on a geometric mean of at least five samples taken within a 30-day period. Marine water recommendations call for no more than 35 enterococcus per100mL as a geometric mean over a 30-day period (EPA, 2003a). States are free to adapt the standards. The new Virginia recreational water quality standards, for example, require no more than 235 *E.coli* per 100mL of freshwater or 104 enterococcus per 100mL of saltwater for a single sampling event and no more than a geometric mean of 126 fecal coliforms per 100mL of freshwater or 35 enterococcus per 100mL of saltwater for multiple samples taken over a thirty-day period. (Virginia DEQ, 2003). The District of Columbia chose not to create new bacterial standards, so freshwater requirements call for no more than 200 fecal coliforms per 100mL as a geometric mean for multiple samples taken over a thirty-day period, and no more than 1000 fecal coliforms per100mL at any time (EPA, 2003a).

Table 1: EPA Recommended Bacterial Standards

Bacterial Standards	*5 sample geometric mean*	*Maximum Allowance*
	(CFU/100 mL)	(CFU/100 mL)
Fecal Coliforms (old)	200	1000
E.coli (freshwater new)	126	
Enterococci (freshwater new)	33	
Enterococci (marine new)	35	

Over 29% of impaired rivers and streams in Washington D.C. are polluted by fecal coliform levels that exceed water quality standards (EPA, 2003b). The Clean Water Act requires each state or territory to submit a list of waters that do not meet water quality standards. In addition for each impaired water, a Total Maximum Daily Load (TMDL) plan must be developed to estimate all significant sources of a pollutant, the specific amount of pollutant coming from each source, and the reduction needed to bring the impaired waterbody back into compliance with water quality standards (Kern et al., 2002). A majority of the TMDLs completed or to be completed in Washington D.C. involve determining sources and creating loading allocations for fecal pollution. Upon completion of the TMDL plans, the state or territory must implement Best Management Practices (BMP) that will meet the goals set forth.

The sources for fecal contamination are ubiquitous. Any means by which fecal contamination can reach receiving waters is a potential source. Agricultural runoff from livestock is often a source in rural watersheds. Origins may include runoff from manure deposited on grazed pastures, runoff from the application of fecal slurries, and runoff or discharge from concentrated animal feedlot operations (Simpson et al., 2002). Human fecal pollution can enter waterways through combined sewer overflows, separate sanitary sewer overflows that can result from leaky or undersized sanitary sewer pipes, and stormwater runoff that includes overland flow and flow conveyed through storm sewer pipes. Other contributions include wildlife and domestic animals. In the Washington D.C. area, the major concern is human fecal contamination through the overflow of the combined sewer/stormwater system.

Identifying the sources of contamination is not only a necessary tool for TMDL development, but is also an important tool for assessing potential health risks to humans. Waters contaminated with human feces are more likely to carry human pathogens such as *Salmonella enterica* serovar Typhi, *Shigella spp*, hepatitis A, and Norwalk-group viruses. There are pathogens associated with animal feces that pose risks to human health as well as various types of *E. coli*, and *Cryptosporidium spp* (Scott et al., 2002).

For water quality specialists and stormwater managers, locating and identifying sources of fecal contamination is high priority but doing so is difficult. In early attempts to differentiate human feces from all other sources, calculation of a fecal coliform to fecal streptococci ratio was performed on environmental samples. The procedure was based on the premise that human feces would produce a ratio of 4.0 or greater while animals would produce a ratio of below 0.7. This technique was strongly criticized, however, because variations were found in fecal *Enterococci* densities among individuals subjected to different diets and environmental conditions (American Public Health Association, 1998).

In the early 1990's technologies known as Bacterial Source Tracking (BST) were introduced which enabled researchers to pinpoint sources of fecal pollution and even calculate loads from the sources. The application of this technology to TMDL studies for fecally contaminated waters has proven to be the most affective tool to date. There are several method variations currently available for BST related projects. These include molecular (genotypic) methods and biochemcial (phenotypic) methods. Molecular methods determine DNA fingerprints of bacteria or viruses, and rely on the unique genetic makeup of each microbial pathogen or indicator organism (Hager, 2001). Molecular methodologies currently under investigation and showing good potential for use in BST projects are

Ribotyping, Repetative PCR, and Pulse Field Gel Electrophoresis (EPA, 2002b). Biochemical methods measure the type and quantity of biochemical substances produced by some effect of an organism's genes (Hager, 2001). Biochemcial methods currently under investigation include Antibiotic Resistance Analysis (ARA), cell wall analysis of fatty acid methyl ester (FAME), F-specific coliphage typing, and Carbon Source Utilization Profiling (BIOLOG) (EPA, 2002b).

At this point, no single BST method has risen head and shoulders above the rest (Hager, 2001), however, there are some advantages to the use of ARA, which include simple techniques, quick sample turnaround times, cost effectiveness, and similar rates of correct classifications in comparison to many molecular methods. ARA is the most widely used BST method to date, Simpson et al. (2002) state that ARA appears to be the most practical approach for source tracking in small watersheds. ARA was the method selected for the study of fecal pollution sources in Washington D.C. waterways.

OBJECTIVES

The primary objective of this project was to determine the sources of fecal contamination in Rock Creek, the Anacostia River, and the Potomac River. The goal was to provide the Metropolitan Washington D.C. Council of Governments with the knowledge needed to make sound decisions regarding the restoration of water quality on these three waterways. Unknown stream isolate profiles from each of these waterways were collected periodically and compared against a known source library using logistic regression, which provided the sources of contamination and their relative contributions to the fecal pollution.

There were several secondary objectives addressed during the course of this study, which attempted to expand the limits of past ARA projects. The first objective to apply

ARA to a highly urbanized region. Traditional ARA studies in the past have been completed on small rural watersheds, where the main source of fecal contamination was livestock. The second objective was to compare results obtained with statistical analysis of the known source library and unknown stream isolates with the use of discriminate analysis and logistic regression. Discriminate analysis was widely used in the medical field prior to its adaptation as an analytical tool for source tracking, first applied by Bruce Wiggins (1996). DA is a more conservative statistical tool than is logistic regression (LR), but it harbors the assumption that the subjects under scrutiny are uniform and this is not entirely true when the subjects are microorganisms.

The third objective of this study was the analysis of probabilities to determine if the unknown isolates were correctly identified. The statistical model (DA or LR) forces each isolate into a source category regardless of the level of certainty that the isolate matches any one source. An unknown category was created to accommodate isolates that do not meet a specific level of certainty for source identification. Some molecular techniques have reported this analysis but up until now, no attempt had been made to compare the mismatch probabilities that ARA produces to the mismatch probabilities that molecular techniques produce.

The final objective was to perform some very stringent tests for representativeness of the known source libraries. These tests included a pulled sample cross validation analysis and a removal of all duplicate patterns. Pulled sample cross validation involves removal of isolates taken from the same sample and comparing those groupings against the known source patterns in the library one at a time. Proponents of this research claim that this analysis removes the bias created when the DA or LR models calculate the typical rates of

correct classification (through resubstitution analysis) by eliminating the possibility that the isolates can be compared against their own pattern. In a similar manner, if all duplicate patterns are removed from the library leaving only unique profiles, proponents claim that resubstitution analysis would then provide an accurate depiction of library representativeness through the elimination of all bias associated with identical patterns from the same sample and within the same source. These tests were performed on the libraries and the results were weighed against the implications the tests have for ARA.

REFERENCES

American Public Health Association, American Water Works Association, and Water Environment Federation. 1998. M. A. Franson, Editor. Standard Methods for the Examination of Water and Wastewater, 20th Edition. American Public Health Association, Washington D.C. pp.9-74-9-78.

Environmental Protection Agency. 2003a. Bacterial Water Quality Standards for Recreational Waters, Status Report. Office of Water. EPA-823-R-03-008.

Environmental Protection Agency. 2003b. 1998 Section 303(d) List Fact Sheet for the District of Columbia. http://oaspub.epa.gov/waters/state_rept.control?p_state=DC#TPOL [Online] Last Accessed November 9, 2003.

Environmental Protection Agency. 2002a. Water Quality Conditions in the United States, A Profile from the 2000 National Water Quality Inventory. http://www.epa.gov/305b/2000report/factsheet.pdf. [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2002b. Wastewater Technology Fact Sheet, Bacterial Source Tracking. Office of Water. EPA 832-F-02-010.

Holt, J.G., N.R. Krieg, P. Sneath, J.T. Staley, S.T. Williams. 1994. Bergey's Manual of Determinative Bacterialogy, 9th Edition. Williams & Wilkins, Baltimore. pp. 538-539.

Kern, J., B. Petrauskas, P. McCelellan, V. Shanholtz, C. Hagedorn. 2002. Bacterial Source Tracking: A Tool for Total Maximum Daily Load Development. Advances in Water Monitoring Research. 125-142.

Hager, M.C. 2001. Detecting Bacteria in Coastal Waters 1 and 2. Stormwater Magazine. 2 (3): 16-25. http://www.forester.net/sw_0105_detecting.html and http://www.forester.net/sw_0106_detecting.html [Online] Last accessed November 9, 2003.

Scott, T.M., J.B. Rose, T.M. Jenkins, S.R. Farrah, J. Lukasik. 2002. Microbial Source Tracking: Current Methodology and Future Directions. *Applied and Environmental Microbiology.* 68 (12): 5796-5803.

Simpson, J.M, J.W. Santo Domingo, D.J. Reasoner. 2002. Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36 (24): 5279-5288.

Virginia Department of Environmental Quality. 2003. Memorandum Commonwealth of Virginia Department of Environmental Quality, Water Division. http://www.deq.state.va.us/waterguidance/pdf/032007.pdf. [Online] Last Accessed November 9, 2003.

Wiggins, B. A. 1996. Discriminate Analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animals sources of fecal pollution in natural waters. Applied Environmental Microbiology. 62: 3997-4002.

CHAPTER II

LITERATURE REVIEW

Total Maximum Daily Load Studies

A Total Maximum Daily Load (TMDL) is formally defined as a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and an allocation of that amount to the pollutant's sources (EPA 2003a). Section 303 of the Clean Water Act establishes TMDL programs and requires each state to create their own water quality standards. Through this legislation, states also identify and list impaired waters every two years (EPA, 2003b), and prioritize and perform TMDL studies for those waters listed. Loadings are based on evaluations of point source estimates, non point source estimates, natural background loading, a margin of safety, and seasonal variation (EPA, 2003c).

Though TMDLs have been law since the inception of the Clean Water Act from 1972, they have only recently been put into action and the ideals of managing waterbodies at the watershed level have taken shape. Approximately 45 legal actions in 37 states and Washington D.C. have forced the EPA under court order/consent decree to ensure that TMDLs are established (EPA, 2003d). Pressure has been mounting to apply the implementation strategies and targets as indicated in the TMDL studies. The EPA passed a ruling in 1992 that sought to establish a national consensus for the development and implementation of TMDLs. It set the guidelines for TMDL development and public interactions and pushed states to develop TMDLs within 8-13 years of the listing of a waterbody. In 2000, the EPA attempted to pass another rule which provided for a more comprehensive list: state impairment listing to occur every four years, impaired waterways

to remain on the 303 (d) list until successfully removed, and (most importantly) put a time limit of ten years (if possible) on the attainment of state water quality standards (EPA, 2003e). However this rule was not implemented and the TMDL program still follows the guidelines established under the rule of 1992.

The most recent national water quality inventory showed a list of 41,973 impaired waters. Of the impaired waters, 6,045 or 12.5% are attributed to pathogens (EPA, 2003f). The EPA classifies all waters impaired by high levels of any fecal-indicator organism as pathogenic (Kern et al., 2002). Since January 1996, 1,825 or 19.2% of the total TMDLs that have received approval have been pathogenic impairments (EPA, 2003f). Pathogens are a main cause of impairment to our nation's waterways and pose the greatest health risk, thus have been the focus of a majority of the TMDLs.

A total of 966 impairments were listed by Virginia under the last listing cycle in 1998, with pathogens as the leading cause at 44.2% of the total impairments (EPA, 2003g). In Virginia, waters are listed for pathogenic impairments if they exceed the standard set for *E.coli* and enterococcus, two commonly used indicators of pathogenic microbes. The Virginia recreational water quality standards for a single sampling event require no more than 235 *E.coli* per 100mL for freshwater or 104 enterococcus per 100mL of saltwater and no more than a geometric mean of 126 *E.coli* per 100mL for freshwater or 35 enterococcus per 100mL of saltwater for multiple samples taken over a thirty-day period. (Virginia DEQ, 2003). Since January 1996 322 TMDLs in Virginia have met approval by the EPA and 44 of these (14%) have been for pathogenic impairments (EPA, 2003g).

Washington D.C. has reported 36 impaired waterways to the EPA. These 36 waterways have 122 total impairments, with a majority (29.5%) caused by pathogens and

priority organics. Five TMDLs from the District have been approved by the EPA, but none of these are for pathogenic impairments (EPA, 2003h).

Due to the dominance of pathogenic impairments and the human health risks associated with them, in Virginia, Washington D.C. and nationwide, there was and still is an intense need to develop technologies that can more accurately address fecal source identification and load allocation issues as required by TMDLs. Bacterial source tracking, which encompasses a variety of techniques (molecular and non molecular) has been the leading source of progress being made to develop more meaningful and useful TMDL information for remediation efforts and for addressing human health risks.

Bacterial Source Tracking

Regulatory agencies in conjunction with scientists across the country have been in search of methods to determine sources of fecal contamination since the inception of the Clean Water Act, though most of the progress has been made since the mid 1990's. Bacterial source tracking technologies can be subdivided into three basic groups: Molecular, Biochemical, and Chemical. Molecular (genotypic) methods involve DNA fingerprinting and are based on the unique genetic makeup of different strains or subspecies of fecal bacteria. Biochemical (phenotypic) methods are based on the reaction of an organism's genes that actively produce certain biochemical substances that are detected and measured. Chemical methods are based on the detection of chemical compounds that are associated with human wastewaters (Hagedorn, 2003). Generally the molecular and biochemical methods require a known source library for a comparison point for unknown stream isolates, but there has been some recent work on methods to circumvent this step in the source tracking process.

A major issue the source tracking community faces in addition to method selection is target organism selection. There are many pathogenic microbes (human and animal) that may be present in fecally polluted waters. Detecting them all is not only impossible, but impractical. There are several organisms whose presence serves as indicators of pathogens. Targets currently under investigation are bacterial, viral, or protozoan. Most source tracking studies have involved bacterial indicators, but there is significant research being completed on other indicators as well.

BST Targets

BST bacterial targets include total/fecal coliforms, *Escherichia coli*, *Enterococcus sp*, *Bifidobacterium sp.*, and *Clostridium perfringens.* Species belonging to these bacterial groups are normally present in the feces of higher mammals and birds. Total and fecal coliforms have been used extensively as indicators of water quality. The pathogenic standards for Washington D.C. are based on fecal coliform levels, but recently scientists have discovered that their ecology, prevalence, and resistance to stress differs from many of the pathogenic microbes found in water (Scott et al., 2002). This has led researchers to search for other more appropriate bacterial indicators. The use of *E. coli* for BST has dominated the scientific literature. *E.coli* is generally non-pathogenic to humans, and is present at much higher concentrations than the pathogens it indicates, but there is evidence that suggests *E.coli* may not be a reliable indicator in tropical and subtropical environments due to its ability to replicate in contaminated soils (Desmarais et al., 2002). *Enteroccocus* is gaining popularity as an indicator organism because for the most part, it is the most frequently found organism in human feces. *Enteroccocus* has also proven reliable as an indicator in marine environments and recreational waters due to the organism's ability to

grow in 6.5% NaCl (Scott et al., 2002). Like *E. coli*, enteroccocus possesses qualities of a good indicator organism as it is non-pathogenic and present in concentrations much higher than the pathogens it indicates. Enterococcus is found in environmental reservoirs where regrowth is highly possible, so its detection may not necessarily indicate recent contamination (Desmarais et al., 2002). Interest in the use of *Bifidobacterium* stems from its inhabitance as a major component of the human intestine, and its rarity in animals. *Bifidobacterium's* unique ability to ferment sorbitol can be used to further differentiate the organism as human derived (Scott et al., 2002). As indicators, *Bifidobacterium* are not the first choice because of variable survival in the environment and weakly defined culture methods. *C. perfringens* is another bacterial target that has potential use in source tracking because it is a good indicator of viruses and remote fecal pollution, but as an indicator *C. perfringens* may not represent recent contamination due to long term persistence in the environment.

Viral indicators are thought to be good indicators of viral pathogens. Viral targets include *Bacteroides fragillis* bacteriophages, F-specific RNA coliphages, and human enteric viruses. Like bacterial targets each organism or group of organisms has pros and cons associated with its usefulness as an indicator of pathogenic microbes. *Bacteroides fragillis* bacteriophages is a virus that infects *Bacteroides fragillis*, a bacterium that is abundant in the human gut. It makes a good indicator because of its presence in human feces, its inability to replicate in the environment after deposition, and a high correlation with human enteric viruses (Scott et al., 2002). Several studies however have indicated that due to low numbers, the viruses are not always detectable in waters directly impacted by human fecal contamination (Tartera et al. 1989, Puig et al. 1999). Coliphages are viruses that infect

E.coli. The F-specific RNA coliphage has four main subgroups (I-IV). Groups II and III have been shown to be highly associated with human fecal contamination and/or domestic sewage, group IV are associated with animal and livestock waste, and group I has been found in all types of wastes (Hager, 2001). It makes a good target for differentiating between human and animal fecal contamination. As a virus, its numbers are much lower in the environment that its bacterial counterpart *E.coli,* so detection methods are very sensitive and efforts to isolate the F+ RNA coliphage has revealed that only a small percentage of human fecal samples contain these phages (Havelaar et al., 1990).

The most ideal way to complete a risk assessment in fecally polluted waters is to directly test for human enteric viruses. By monitoring for these viruses directly, the uncertainty associated with the use of fecal indicators can be avoided. Unfortunately there are over 120 of these viruses that exist in the human intestinal tract alone, and many of these viruses are not easily cultivated from environmental samples (Scott et al., 2002). These targets are recommended for use as indicators only in conjunction with other methods for prediction of fecal pollution and enteric pathogens.

The most common protozoan targets are *Cryptosporidium* and *Giardia* species. *Giardia* is regarded as the most common flagellate in the human digesticve tract and is highly contagious in the cyst stage of its life cycle. It is pathogenic for humans and animals and is transmitted primarily through contaminated water. Cysts germinate in the gastrointestinal tract and the symptoms of giardiasis, a diarrheal disease, ensue (Webster, 2000). *Cryptosporidium* is another common protozoan found in the intestinal tracts of vertebrates. Cryptosporidiosis is a significant cause of waterborne outbreaks of diarrheal disease and molecular typing has shown it to be caused by two genotypes of

Cryptosporidium parvum (Xiao et al., 2000). Both of these groups of organisms have a high tolerance to harsh environments and can withstand traditional means of wastewater treatment such as chlorination. Another issue related to the detection of these protozoa is that they have low infectious doses, so it is often necessary for health risk assessments to detect organisms that are present in much higher quantities, but are still associated with the presence of these pathogenic organisms.

Chemical Methods

Chemical targets studied for their applicability to source tracking projects are limited at this time. These targets can only be indicative of the presence or absence of contamination sources, with most of the research focusing on chemicals associated with human wastes. Chemical targets include caffeine, fragrance agents, fluorescent whitening agents, and fecal sterols. These compounds are associated with human wastewaters so their presence would indicate human pollution. Caffeine is a compound ingested by humans, which passes through the digestive systems, however, its presence is not necessarily indicative of human contamination, as some plants have the ability to produce caffeine. Another disadvantage to caffeine detection is the difficulty in detecting it because soil microbes readily degrade it. Fragrance agents are another option, but these compounds as well as caffeine are detected chemically, which requires expensive gas chromatography equipment. Fecal sterols and stanols are constituents of fatty acids in cell walls and membranes. Recent research has suggested that types and quantities of theses substances may show differences between human and animal feces (Hagedorn, 2003). Corpostanol is a fecal stanol formed during the breakdown of cholesterol by bacteria present in the gut of humans and other animals. It is the primary stanol detected in domestic wastewater

(MacDonald et al., 1983). This method is currently under development and there are no published reports of its use in fecal source tracking (Hagedorn, 2003). Flourescent whitening agents, sodium tripolyphosphate, and linear alkyl benzenes are other promising chemical targets (Scott et al., 2002). These include optical brighteners and other chemicals associated with laundry wastewater. Their presence may indicate human waste plumes.

Molecular Methods

Molecular (genotypic) methods can be used to differentiate bacterial species found within different animal hosts, but under two very important assumptions. The first is that within a species of bacteria, there are subgroups that have become more adapted to a particular host or environment, and the second is that after adaptation, the subsequent replications of the organism will be genetically identical, which overtime creates a group within a particular host or environment that posses a unique genetic fingerprint (Simpson et al., 2002). Genotypic methods used to date in source tracking related projects include Pulse Field Gel Electrophoresis (PFGE), Ribotyping, Repetitive Polymerase Chain Reaction (Rep-PCR), Length Heterogeneity PCR (LH-PCR), Terminal Restriction Fragment Length Polymorphism (T-RFLP), Denaturing Gradient Gel Electrophoresis (DGGE), Amplified Fragment Length Polymorphism (AFLP), Toxin Biomarkers, and Reverse Transcriptase PCR. The most commonly researched and applied molecular methods to source tracking projects include PFGE, Ribotyping, and Rep-PCR. The other methods will be discussed briefly but not much information is available regarding results and potential as source tracking tools. These methods are being adapted from medical microbial identification applications and have not been thoroughly tested with environmental samples.

LH-PCR and T-RFLP have recently been proposed for source tracking and are based on the detection of flourescently labeled 16S rDNA PCR products. These methods are used to analyze length differences of gene fragments and the abundance of the different fragment sizes is determined. In general, each population of the community contributes a terminal fragment of a distinct size. These methods are non library dependent which saves time and money and have been used to successfully detect rDNA sequences specific to anaerobic fecal bacteria which are found in a higher proportion of the gastrointestinal tract (Simpson et al., 2002). These methods can also be used to detect recent contamination. Bernhard and Field used LH-PCR and T-RFLP to target *Bifidobacterium* and *Bacteroides* species (Bernhard et al., 2000).

DGGE is capable of discriminating between PCR products of similar size based on changes in electrophoretic mobility. These properties are influenced by the melting properties of the DNA fragments (Muyzer et al., 1996). Farnleitner et al. (2000) has demonstrated that DGGE could be used to detect and differentiate *E.coli* populations, but no work has been done to differentiate sources.

AFLP has the ability to inspect an entire genome for polymorphisms. It has been used successfully in taxonomical, epidemiological, and ecological studies to identify and characterize microbial parasites, fungi, and bacteria (Savelkoul et al., 1999). It is a library dependent method requiring the cultivation of target organisms (Simpson et al., 2002).

Toxin biomarkers involve the detection of bacterial contamination sources through the identification of genes that code for toxins in *E. coli* populations. There are currently three biomarkers for source identification; one for human, one for cattle, and one for pig. As an advantage, the approach is not location specific, and is not library dependent, though it

only provides information on the basis of presence/absence. A more serious implication for this method is that the toxin genes are not highly prevalent in *E.coli* populations (Hager, 2001).

Reverse transcriptase PCR (RT-PCR) can be used to detect the RNA of any organism whose genome has been previously sequenced. Dr. Rachel Nobel has applied the method extensively to the detection of enteroviruses in coastal waters (Hager, 2001). The primers it employs are complementary to viral RNA sequences. The process transcribes the detected RNA back into DNA and is amplified via PCR. This method shows promise for the detection of human contamination because enteroviruses are human specific. The method is also non-library dependent or location specific, but can only be used as a presence/absence test. Unlike many other methods that detect indicator organisms, RT-PCR is a direct measure of pathogens in the system (Hager, 2001).

Rep-PCR uses primers corresponding to DNA elements that are repeated in various locations within the genome to generate a highly specific DNA fingerprints. There are three major repetitive elements in bacteria that have been studied for use in source tracking to generate fingerprints; REP, ERIC, and BOX. BOX primers have shown to be the best suited (Scott et al., 2002). This method is library dependent, requiring a significant number of known source isolates. Dombeck et al. (2000) used rep-PCR to differentiate *E.coli* strains from seven known sources. Dombeck used BOX, REP, and BOX plus REP primers, and the best classification rates for all of the known sources were obtained with the BOX primers. The Average Rate of Correct Classification (ARCC) was 87.5% and analysis was completed using BOX PCR DNA fingerprints and Jackknife analysis. Results are to be interpreted loosely because few isolates were used as known source comparisons. Limitations include

cost and labor intensity. Another major criticism goes to the overall reproducibility of the method (Scott et al., 2002)

Ribotyping was first applied to human disease outbreaks, and has been applied to bacterial source tracking 11 years ago. Mansour Samadpour is a leading researcher of ribotyping applications to BST studies. Ribotyping generates genetic fingerprints of *E.coli* from genes that code for highly conserved ribosomal RNA. Restriction enzymes are used to cut the DNA into pieces that are sorted by size through the process of electrophoresis. Highly sophisticated genetic probes visualize the location of DNA fragments, which appear as bands that correspond to relevant rRNA. This is known as a ribotype (Hager, 2001). The ribotyping procedure requires the creation of a large known source database.

In 1995 Samadpour and Chechowitz performed a groundbreaking ribotyping study of *E.*coli patterns found the Little Soos Creek in Washington. Eight hundred twenty-three various animal isolates and 227 septic tank isolates were collected and used to build a known source database. When unknown stream isolates were compared to the database, 71% of the isolates were identified as belonging to one of the known sources, while 29% were not matched. Another ribotyping study completed in 1999 by Parveen et al. (1999) differentiated between human and non-human sources and obtained a correct classification rate of 67% for humans and 97% for non-humans. Parveen et al. (1999) also found that human isolates showed much less diversity than non-human isolates. A ribotyping study performed by Carson et al. (2001) compared ARCC's of the library broken into only two host classes (human and non-human) to the library broken into eight host classes. The two-way host analysis resulted in a 95% RCC for human and 99.2% RCC for non-human. The ARCC for the eight-way host split was 73.6%. The library was then tested for clustering in

the comparison of three hosts at a time, and there was a distinct separation of patterns from each class. Carson et al. (2001) showed that the accuracy of classification can be greatly increased by limiting the number of classes compared. The clustering analysis showed that in order to distinguish between hosts, the library should be split into no more than three host sources. Carson et al. (2001) also claimed that ribotyping will be most effective as a source tracking tool when there are only a few obvious potential sources of pollution in the environmental samples, however field tests were not completed on ribotyping as a tool for BST. Ribotyping has produced variable results as to the ARCC's and the relative ability to descriminate among host sources. The dissimilarities in the literature are most likely caused by method variations including the use of different restriction enzymes, different Southern blotting procedures, and differences in analysis and interpreation of ribotype patterns. Other limitations to ribotyping include limited geographic stability, labor intensity, and high cost (Scott et al., 2002).

PFGE analysis begins with pure culture bacterial cells placed in agarose plugs where the DNA is digested using rare restriction enzymes. The digested plugs are placed in a special gel and electrophoresed with alternating currents. This allows for superior band separation by molecular size, creating isolate fingerprints. PFGE has been used extensively in clinical microbial studies, but is time consuming, expensive, and the number of isolates processed simultaneously is limited (Hager, 2001). Dr. George Simmons pioneered PFGE as a BST method. The original study completed on Four Mile Run in Virginia found that 30% of the unknown stream isolates could not be classified based on an 80% similarity to DNA banding patterns in the known source database. Several interesting conclusions were reached based in this study such as a high non-human signature, mostly attributed to

waterfowl, only localized human contamination, a significant role in the elevated fecal coliform levels attributed to discharge from storm drains during baseflow, and *E.coli* regrowth through cloning within the storm drains and stream sediments. The regrowth issue has major implications for all source-tracking methods, but has not been thoroughly researched or addressed. The data collected by Simmons et al. (2002) also suggested that certain wildlife species have a greater disproportionate representation in the DNA profile analysis, but it may not feasible to eliminate wildlife. An important implication from the results of this study is that a surge in heterotrophic microorganisms in a watershed represents an ecosystem that is out of balance.

A study completed by Kariuki et al. (1999) compared *E.coli* isolates from rectal swabs obtained from 62 chickens and 42 children living in close proximity to the chickens to compare their genetic relatedness. The PFGE analysis showed 14 distinct clusters (6 from the children, seven from chickens, and only one overlapping cluster with 60% coefficient of similarity). The results indicate that although several different genotypes were identified for each source, the *E.coli* strains from the two sources were distinct. PFGE is a method with the ability to differentitate between sources with greater certainty, as the patterns produced for discrimination purposes are less likely to overlap among sources.

Non-molecular Methods

Non-molecular methods (phenotypic) focus on the traits that bacteria may have acquired from exposure to different hosts and environments. These techniques generally require less training for lab personnel, cost less per isolate, and typically can be performed on hundreds of isolates per week. These methods under investigation for use in BST include

the fecal coliform/fecal streptocci ratio, F+ RNA coliphage analysis, Carbon Utilization Profiling (CUP), and Antibiotic Resistance Analysis (ARA).

The fecal coliform/fecal streptococci ratio and F+ RNA coliphage analysis have briefly been described above. To reiterate and expound upon the previous information provided, the fecal coliform/fecal streptococci ratio was proposed as a source tracking method as far back as 1969. Prior to more modern BST techniques it was the most widely accepted and used means of differentiating between human and non-human sources. It is based on the premise that human feces have higher levels of fecal coliform counts while animals have higher levels of fecal streptococci. A ratio of greater than 4 would thus indicate human pollution while a ratio of less than 0.7 would indicate non-human pollution (Scott et al.., 2002). The method was however proven unreliable due to variable survival rates of fecal streptococci and differences in fecal *Enterococci* densities found in individuals with different diets (Simpson et al.., 2002). Like the fecal coliform/fecal streptococci ratio, the F+RNA coliphage method only allows for the differentiation of human and non-human pollution. The F-specific RNA coliphage has four main subgroups (I-IV). Groups II and III have been shown to be highly associated with human fecal contamination and/or domestic sewage, group IV are associated with animal and livestock waste, and group I has been found in all types of wastes (Hager, 2001). As a virus, its numbers are much lower in the environment that its bacterial counterpart *E.coli* so detection methods are very sensitive and efforts to isolate the F+ RNA coliphage has revealed that only a small percentage of human fecal samples contain these phages (Havelaar et al.., 1990).

Carbon Utilization Profiling (CUP) is used to differentiate sources of bacteria based on differences among bacteria and their use of a wide range of carbon and nitrogen sources

for energy and growth. This method has worked well in the laboratory environment, but many environmental factors impact bacterial nutrient requirements, which may make this method impractical for field determination. BIOLOG is a commonly used system for performing CUP studies. It rapidly scores and tabulates a 96-carbon source utilization test per isolate, which generates a nutritional profile that can be used to build a known source library (Hagedorn, 2003). A study performed by Hagedorn et al.. (2003) involved the identification of 375 *E.coli* isolates down to species and the creation of nutritional patterns based on 30 of the 96 nutrient wells provided by the BIOLOG system. A two way human and non-human split gave and ARCC of 92.7%, a three way source split gave an ARCC of 85.7%, and a four way source split gave an ARCC of 81.9%. The results were based on a modest *Enterococcus* library and a preliminary field validation test, but did suggest that this method has potential as a phenotypic BST method.

Antibiotic Resistance Analysis is based on patterns of antibiotic resistance found in bacteria from human and animal sources. The premise is that human fecal bacterial will demonstrate greater resistance to certain antibiotics, and animals, specifically livestock, will be resistant to others. *E.coli* and *Enterococci* are the commonly used target organisms with this method. To determine characteristic antibiotic resistance patterns, fecal bacteria isolates are tested on a variety of antibiotics and concentrations. After incubation, the isolates are scored for growth/no growth resulting in a form of an organisms specific resistance pattern (Hager, 2001).

ARA requires a known source database for which to compare unknown isolates. While it appears that molecular methods are more precise, ARA has shown results that are just as precise in identifying sources and is also the most commonly applied BST method

thus far (Hagedorn, 2003). In 1997 Parveen et al. used Multiple Antibiotic Resistance Profiles (now commonly referred to as ARA) of 765 *E.coli* isolates from point and nonpoint sources collected from the Appalachicola National Estuarine Reserve. Nonpoint sources included isolates collected from, estuarine surface waters while point sources included direct sampling of human and animal feces. *E.coli* isolates from point sources showed significantly greater resistance to antibiotics and had higher concentration tolerances than isolates from non point sources. In 1999, Hagedorn et al. published results from an ARA study performed on a rural Virginia watershed. The results were analyzed and formatted in the same fashion as most molecular techniques. A known source database was built containing 7, 058 isolates with and ARCC of 89%. The method was then field tested on a watershed improvement project on Page Brook. The results indicated that cattle were the major source contributions and so cattle were restricted from the stream. Post fencing, fecal coliform levels were reduced at three sites by an average of 94% and less than 45% of the isolates were then classified as being cattle. The results demonstrated that ARA in fecal streptococcus could be used to reliably determine sources of fecal contamination in efforts for remediation.

At the same time, Dr. Bruce Wiggins was also developing ARA as a technique for BST. He was interested in determining the reliability and repeatability of ARA. Four large sets of isolates of fecal streptococci ranging from 2,635 to 5,990 isolates were obtained from several known sources and from pristine waters. ARA patterns were analyzed using discriminate analysis and it was found that when isolates were classified individually, the ARCC's ranged from 67-78%. When the resistance patterns of all isolates from each sample were averaged and classified, the ARCC's jumped to 96-100%. This indicated a measurable

and consistent difference in antibiotic resistance patterns isolated from various sources (Wiggins et al., 1999).

Jody Harwood et al. (2000) used ARA of fecal streptococci and fecal coliforms isolated from domestic wastewater and animal feces to build a known source library. Discriminate analysis was used to establish the relationship between antibiotic resistance patterns and bacterial source. The ARCC obtained from this study was 62.3% for *Enterococci* and 63.9% for *E.coli*. Both databases identified the source of indicator bacteria isolated from waters directly impacted by septic tank discharges as human. The results of this study indicated that ARA has potential for use as a BST method in subtropical waters as well. ARA is also now being applied in the development of bacterial TMDL's. An ARA study completed by Booth et al. (2003) was conducted on the fecally impaired Blackwater River in south-central Virginia. A library of 1,451 *Enterococci* isolates was developed from human, wildlife, and livestock. The ARCC was 85.3%. The results indicated that livestock was the major contributor to fecal loading, followed by wildlife and then human. The consequent TMDL and BMP plans will address all three sources. This study showed that source tracking works in a large watershed.

There is no one single method that has the ability to identify specific sources of fecal pollution with absolute certainty. Several issues, however, are common to all BST methods that must be addressed. Questions have arisen as to the appropriate number of samples and isolates required to obtain an accurate representation of the pollution sources in environmental waters. Recent research has suggested that the larger the number of isolates obtained from environmental samples the more representative the database will be, but little is known about the upper and lower limits that researchers can settle with and still have

confidence that the isolates are representative. Work by Alexandria Graves (2003) suggested that monthly sampling was adequate, and that between 24 and 48 isolates obtained from environmental samples were the limits that produced results that were representative of the pollution sources.

Other general issues of concern are the survival and regrowth of indicator organisms in the environment after deposition. Little is known about the extent of survival and regrowth for *Bifidobacterium, B. fragillis, Bacteroides,* and *Prevotella* (Scott et al. 2002). The survival and regrowth in the environment of enterococcus and *E.coli* has been documented in subtropical environments by Desmarais et al. (2002). The impact of survival and regrowth of indicator organisms can show artificial eleveations of fecal pollution above what is expected from the fecal impacts alone. The risk is also introduced for identifying sources in incorrect proportions as compared to what is actually being introduced into the system.

Library dependent and non-library dependent methods each have distinct advantages and disadvantages that have yet to be compared and resolved. Library dependent methods carry the question of adequate size and representativeness as well as geographic and temporal stability. Many methods have not incorporated tests for adequate library size and representativeness (Simpson et al., 2002) and thus interpretation of results remains loose at best. Wiggins et al. (2003) thoroughly described several means for determining library size and representativeness for ARA. The other major issues with library dependent methods are geographic and temporal stability. Not much information exists that have addressed these concerns (Simpson et al., 2002). Researchers want to know if isolates can successfully be used to identify sources of fecal pollution from watersheds outside of where collection took

place, and how rapidly isolate profiles (molecular or biochemical) are changing. Implications of geographic and temporal stability may very well impact the effectiveness of library dependent methods.

Non-library dependent methods provide much more rapid results that are helpful where human health risk assessments are of major concern. The advantage is that the time is not needed to culture and create large known source databases before testing of environmental isolates can begin. These methods, however, provide only qualitative information or a presences absence test for certain sources of fecal pollution. The question that has yet to be addressed is how to go from a non-database dependent method that just tests presence/absence to one that provides quantifiable source distributions.

Two relatively new and unaddressed issues for library dependent methods are appropriate statistical modeling and determining and setting a minimum similarity or matching probability limit that an isolate must meet or exceed in order to be confidently classified as a source. To date, several statistical analysis tools have been employed on library dependent methods in attempts to classify and quantify sources of fecal pollution. The most common statistical tools have been cluster analysis and discriminate analysis. These tools harbor certain assumptions that are now thought that microorganism populations may violate such as data sets for which pre-specified, well-defined groups exist, or group dispersions which are homogenous (McGarigal et al. 2000). Another result of discriminate analysis is that it forces all unknown isolates into a source category no matter how well the isolate actually matches any of the profiles. Some researchers have begun to create lower limit cutoffs for percent similarity or matching that an isolate must exceed in order to identify the isolate as a particular source. All isolates not matching any of the profiles by

that probability are labeled as unknowns or unidentifiable by the library. In a PFGE study Simmons et al. (2002) used a cutoff of 80% similarity and classified all other isolates as unknowns. As a result, almost half of the isolates were unidentifiable by the library. Appropriate statistical modeling and the creation of an unknown source category to hold isolates not well matched provide important information about library usefulness that have not been addressed much in the literature.

SUMMARY

Various BST methods have been shown as effective tools for identifying sources of fecal pollution, but clearly more work needs to be done on a methods comparison basis to determine which technique will provide information that can be used by regulatory agencies as an application for the TMDL program or even by local communities seeking to control and remediate contamination that has health risk implications. There are certain advantages associated with molecular and non molecular methods such as accuracy levels, cost per isolate, lab personnel training requirements, etc. that should be addressed under these studies, but the bottom line is, as of right now, no one method rises heads above all others. There are other issues at hand that also need to be addressed including appropriate number of samples and isolates, the relative temporal and geographical stability of the genotypic and phenotypic profiles, the fate of target organisms after deposition (Scott et al., 2002), appropriate statistical modeling, and analysis for similarity probabilities used for isolate identification. Dr. Charles Hagedorn (Hagedorn, 2003) suggests for now, the use of a "toolbox approach" which combines various methods to capitalize on the best features of each, which would lead to a stronger more reliable BST approach.

REFERENCES

Bernhard, A.E. and K.G. Field. 2000. Identification of Nonpoint Sources of Fecal Pollution in Coastal Waters by Using Host-Specific 16S Ribosomal DNA Genetic Markers from Fecal Anaerobes. Applied and Environmental Microbiology. 66: 1587-1594.

Carson, C.A., B.L. Shear, M.R. Ellersieck, and A. Asfaw. 2001. Identification of Fecal *Escherichia coli* from Humans and Animals by Ribotyping. Applied and Environmental Microbiology. 67 (4): 1503-1507.

Desmarais, T.R., H.M. Solo-Gabriele, and C.J. Palmer. 2002. Influence of Soil on Fecal Indicator Organisms in a Tidally Influenced Subtropical Environment. Applied Environmental Microbiology. 68 (3): 1165-1172.

Dombeck,. P.E., L.K. Johnson, S.T. Zimmerley, and M.J. Sadowsky. 2000. Use of Repetative DNA Sequences and the PCR to Differentiate *Escherichia coli* Isolates from Human and Animal Sources. Applied and Environmental Microbiology. 66 (6): 2572-2577.

Environmental Protection Agency. 2003a. Introduction to TMDLs. http://www.epa.gov/owow/tmdl/intro.html#definition [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003b. Overview of TMDLs. http://www.epa.gov/owow/tmdl/tptmdl/slide09.html [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003c. Overview of TMDLs. http://www.epa.gov/owow/tmdl/tptmdl/slide11.html [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003d. Overview of TMDLs. http://www.epa.gov/owow/tmdl/tptmdl/slide08.html [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003e. Overview of TMDLs. http://www.epa.gov/owow/tmdl/tptmdl/slide16.html [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003f. National Section 303 (d) List Fact Sheet. http://oaspub.epa.gov/waters/national_rept.control#TOP_IMP [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003g. 1998 Section 303(d) List Fact Sheet for Virginia. http://oaspub.epa.gov/waters/state_rept.control?p_state=VA [Online] Last accessed November 9, 2003.

Environmental Protection Agency. 2003. 1998 Section 303(d) List Fact Sheet for the District of Columbia. http://oaspub.epa.gov/waters/state_rept.control?p_state=DC#TPOL [Online] Last Accessed November 9, 2003.

Farnleitner, A.H., N.Kreuzinger, G.G. Kavka, S. Grillenberger, J.Rath, and R.L. Mach. 2000. Simultaneous Detection and Differentiation of *Escherichia coli* Populations from Environmental Freshwaters by Means of Sequence Variations in a Fragment of the β-D-Glucuronidase Gene. Applied and Environmental Microbiology. 66: 1340-1346.

Graves, Alexandria. 2003. Identifying Sources of Fecal Pollution in Water as a Function of Sampling Frequency Under Low and High Stream Flow Conditions. Dissertation for Doctorate of Philosophy, Virginia Tech.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneau Jr. 1999. Determining Sources of Fecal Pollution in a Rural Virginia Watershed with Antibiotic Resistance Analysis Patterns in Fecal Streptococci. Applied and Environmental Microbiology. 65 (12): 5522-5531.

Hagedorn, C. 2003. Bacterial Source Tracking: BST Methodologies. http://soils1.cses.vt.edu/ch/biol_4684/bst/BSTmeth.html. [Online] Last accessed November 9, 2003.

Hagedorn, C., J.B. Crozier, K.A. Mentz, A.M. Booth, A.K. Graves, N.J. Nelson, and R.B. Reneau Jr.. 2003. Carbon Source Utilization Profiles as a Method to Identify Sources of Faecal Pollution in Water. Journal of Applied Microbiology.94: 792-799.

Hager, M.C. 2001. Detecting Bacteria in Coastal Waters 1 and 2. Stormwater Magazine. 2 (3): 16-25. http://www.forester.net/sw_0105_detecting.html, and http://www.forester.net/sw_0106_detecting.html [Online] Last accessed November 9, 2003.

Harwood, V.J., J. Whitlock and B. Withington. 2000. Classification of Antibiotic Resistance Patterns of Indicator Bacteria by Discriminate Analysis: Use in Predicting the Source of Fecal Contamination in Subtropical Waters. Applied and Environmental Microbiology. 66 (9): 3698-3704.

Havelaar, A.H., W.M. Pot-Hogeboom, K. Furuse, R. Pot, and M.P. Hormann. 1990. F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. Journal of Applied Bacteriology. 69: 30-37.

Kariuki, S. C. Gilks, J. Kimari, A. Obanda, J. Muyodi, P. Waiyaki, and C.A. Hart. 1999. Genotype Analysis of *Escherichia coli* Strains Isolated from Children and Chickens Living in Close Contact. Applied and Environmental Microbiology. 65 (2): 472-476.

Kern, J., B. Petrauskas, P. McCelellan, V. Shanholtz, and C. Hagedorn. 2002. Bacterial Source Tracking: A Tool for Total Maximum Daily Load Development. Advances in Water Monitoring Research. 125-142.

MacDonald, I.A., V.D. Bokkenheuser, J. Winter, A.M. McLernon, and E.H. Mosbach. 1983. Degradation of fecal sterols in the human gut. Journal of Lipid Resources. 24: 675-694.

McGarigal, K., S. Cushman, S. Stafford. *Multivariate Statistics for Wildlife and Ecology Research*. Springer-Verlag. New York, Inc., 2000. pp. 129-180.

Muyzer, G., S. Hottentrager, A. Teske, and C. Waver. In Molecular Microbial Ecology Manual. F. de Brujin, Editor. Kluwer Academic Publishers: Boston, 1996. 3.4.4. 1-23.

Parveen, S., K.M. Portier, K. Robinson, L. Edmiston, and M.L. Tamplin. 1999. Discriminate Analysis of Ribotype Profiles of *Escherichia coli* for Differentiating Human and Nonhuman Sources of Fecal Pollution. Applied and Environmental Microbiology. 65 (7): 3142-3147.

Puig, A., N. Queralt, J. Jofre, and R. Araujo. 1999. Diversity of *Bacteroides fragillis* strains in their capacity to recover phages from human and animal wastes and from faecally polluted wastewater. Journal of Applied Microbiology. 65: 1772-1776.

Samadpour, M. and N. Chechowitz. 1995. Little Soos Creek Microbial Source Tracking: A Survey. University of Washington Department of Health, Seattle. P. 49.

Savelkoul, P.H., H.J.M. Aarts, J. de Haas, L. Dijkshoorn, B. Duim, M. Otsen, J.L.W. Rademaker, L. Schouls, and J.A. Lenstra. 1999. Journal of Clinical Microbiology. 37: 3083-3091.

Scott, T.M., J.B. Rose, T.M. Jenkins, S.R. Farrah, and J. Lukasik. 2002. Microbial Source Tracking: Current Methodology and Future Directions. Applied and Environmental Microbiology. 68 (12): 5796-5803.

Simmons, G.M., D.F. Waye, S. Herbein, S. Myers, and E.Walker. 2002. Estimating Nonpoint Source Fecal Coliform Sources Using DNA Profile Analysis. Advances in Water Monitoring Research. Tamim Younos, Editor. Water Resources Publications, LLC. 143-167.

Simpson, J.M, J.W. Santo Domingo, D.J. Reasoner. 2002. Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36 (24): 5279-5288.

Tartera, C., F. Lucena and J. Jofre. 1989. Human origin of *Bacteroides fragillis* bacteriophages in the environment. Applied and Environmental Microbiology. 55: 2696-2701

Virginia Department of Environmental Quality. 2001. State Water Control Board State Water Quality Standards. http://www.deq.state.va.us/wqs/WQS03.pdf [Online] Last accessed November 9, 2003.

Webster, E.J. 2000. Giardia, Soil Microbiology. http://soils1.cses.vt.edu/ch/biol_4684/Microbes/giardia.html. [Online] Last accessed November 9, 2003.

Wiggins, B.A., P.W. Cash, W.S. Creamer, S.E. Dart, P.P. Garcia, T.M. Gerecke, J. Han, B.L. Henry, K.B. Hoover, E.L. Johnson, K.C. Jones, J.G. McCarthy, J.A. McDonough, S.A. Mercer, M.J. Noto, H. Park, M.S. Phillips, S.M. Purner, B.M. Smith, E.N. Stevens, and A.K. Varner. 2003. Use of Antibiotic Resistance Analysis for Representativeness Testing of Multiwatershed Libraries. Applied and Environmental Microbiology. 69 (6): 3399-3405.

Wiggins, B.A., R.W. Andrews. R.A. Conway, C.L. Corr, E.J. Dobratz, D.P. Doughtery, J.R. Eppard, S.R. Knubb, M.C. Limjoco, J.M. Mettenburg, J.M. Rinehardt, J. Sonsino, R.L. Torrijos, and M.E. Zimmerman. 1999. Use of Antibiotic Resistance Analysis to Identify Nonpoin Sources of Fecal Pollution. Applied and Environmental Microbiology. 65 (8): 3483-3486.

Xiao, L., K.Alderisio, J.Limor, M.Royer, and A. Lal. 2000. Identification of Species and Sources of Cryptosporidium Oocysts in Storm Waters with a Small-Subunit rRNA-based Diagnostic and Genotyping Tool. Applied and Environmental Microbiology. 66 (12): 5492-5498.

CHAPTER III

ANTIBIOTIC RESISTANCE ANALYSIS: ISSUES RELATED TO LIBRARY SIZE, LIBRARY REPRESENTATIVENESS, AND APPROPRIATE STATISTICAL MODELING

INTRODUCTION

Antibiotic Resistance Analysis (ARA) using discriminate analysis was first reported by Wiggins (1996) as a means for determining sources of fecal pollution in surface waters. ARA is based on patterns of antibiotic resistance found in fecal bacteria from human and animal sources. The premise is that human fecal bacteria will demonstrate greater resistance to certain antibiotics, and other animals, specifically livestock, will be resistant to a different combination of antibiotics. The creation of a known source library is the first and most important aspect of ARA. The library must be both appropriate in size to minimize the chance for random grouping of isolates, and in representativeness to ensure that isolates exemplify a majority of the potential antibiotic resistance profiles that are present in the watershed being studied.

ARA has been shown to be an effective tool for BST by Wiggins et al. (1999), Hagedorn et al. (1999), and Harwood et al. (2000). Although there have been no method comparison studies published to date, and from a review of the literature, no single method rises above the others, Simpson et al. (2002) states that thus far, ARA appears to be the most practical approach for source tracking in small watersheds.

Building a known source library is the most essential part of ARA. The usefulness or success of the library is measured by the average rates of correct classification (ARCC). The ARCC is the average of the rates of correct classification for each source, determined by the percentage of known source isolates the statistical model places into the correct

source category (Wiggins, 1996). High ARCC's indicate that the library minimizes the chances for misclassifications as it forces each isolate pattern into a source category.

In Wiggins' (1996) first published report on ARA using discriminate analysis, four concentrations of five antibiotics were tested on 1,435 isolates collected from various sources. When the library was tested with six possible source categories the ARCC was 74%. Upon pooling the sources to four more general source categories, the ARCC rose to 84%, and pooling the sources to human vs. non-human resulted in an ARCC of 95%. To support the initial conclusion that ARA could be a useful tool for bacterial source tracking (BST), Wiggins et al. (1999) conducted another study using four larger data sets that were collected over a four-year period. The ARCC's for a four way source classification ranged from 64-78%.

Harwood et al. (2000) used ARA and discriminate analysis to create a known source library with thirty-two antibiotic/concentration combinations for both fecal coliforms and fecal streptococci. The project objective was to test the usefulness of ARA in predicting sources of fecal pollution in subtropical waters. When the library was split into six source categories, the ARCC was 62.3% for fecal streptococci and 69.3% for fecal coliforms. When the sources were pooled to human vs non-human, the ARCC increased to 74.0% for fecal streptococci and 73.9% for fecal coliforms. As expected, the fewer the source categories, the greater the predictive power of the library. Harwood et al. (2000) further enhanced the confidence in the predictive power of the library as the study included a hold-out cross validation test where isolates were randomly removed and reanalyzed as unknowns. The rates of correct classification (RCC's) of these isolates were not significantly different from those isolates in any source category, indicating that the library

was able to identify sources from isolates that were not part of the library. Field-testing ARA with unknown isolates was deemed to be feasible at that time, however, the interpretation of results with only a very small percentage of isolates attributed to a single source was difficult, because it was often unclear whether the isolates were assigned as a result of misclassification or if the source truly was represented by only a small portion of environmental the isolates. In order to have confidence in the predictions of the library, there was a need for a lower limit of detection, or a minimum percentage for which a source must be present to establish that the source was in fact a present in the environment. Harwood et al. (2000) suggested the use of the expected frequency of misclassification.

Whitlock et al. (2002) conducted a study to test the work previously reported by Harwood et al. (2000) to determine the major sources of fecal pollution in Stevenson Creek in southern Florida (in a highly urbanized watershed). Prior to field-testing ARA in an urban environment, the library was tested for appropriate size by randomly assigning the isolate fingerprints to source categories and comparing the ARCC's produced to the probabilities that an isolate would be assigned to a category by chance. If the probabilities were similar then it was said that the library was of appropriate size. Small libraries were found to yield much higher correct classification rates, but were determined to be unrepresentative of the population diversity of indicator organisms in fecal material as evidenced by high rates of artificial clustering. Artificial clustering is defined as random groupings of isolates rather than groupings based on true relationships. Determining appropriate library size was carried out by randomly assigning the ARA fingerprints of isolates to source categories. The assumption for appropriate size was that when discriminate analysis was carried out, the ARCC for the randomly assigned dataset should

be approximately equal to the probability that an isolate would be assigned to a category by chance. To test for library representativeness, Whitlock et al. (2002) compared known source isolates from a failing septic system against the library, but treated the isolates as unknowns. The majority of the isolates (52.8% in the creek, 89.6% in a puddle near the drainage site, and 91% in the soil) were classified as human indicating the library had good predictive power and that the known source isolates were in fact representative of environmental isolates.

Hagedorn et al. have proven ARA to be an effective tool for BST through numerous field tests. The most notable long term ARA study was conducted on a watershed improvement project on Page Brook in Clarke County, VA initiated in 1996 (Hagedorn et al., 1999). The library in this study was divided into three source categories (human, livestock, and wildlife) from 7,058 known source isolates, which yielded an ARCC of 88%. Only 892 known source isolates were obtained from within the Page Brook watershed. This high ARCC meant that the larger known-source database could be successfully used with isolates from different geographic regions. Hagedorn et al. (1999) study results indicated that cattle were the major contributor to fecal pollution, so cattle access to the stream was restricted. Post-fencing resulted in a reduction of fecal coliforms. Bringing fecal coliform levels to below recreational standards appeared to be an achievable goal. Another aspect of this study was to address issues associated with the known source library. Discriminate analysis (DA) was essential for the study of unknown stream sample isolates because it provided the proportions of fecal material contributed by each source. One consequence of DA, however, is that it forces all isolates into one of the designated source categories regardless of the probability of how well it actually fits into that source category. Cluster

analysis was employed to determine the levels of mismatching and was also useful in testing antibiotic concentrations and combinations to create the optimal source differentiation. To test library representativeness, Hagedorn et al. (1999) suggested regularly adding samples to the existing database and note changes in the ARCC's. If ARCC's do not change appreciably then the library should be representative.

The results of all of these major studies on ARA have reaffirmed work originally completed by Wiggins that antibiotic resistance patterns can be used to successfully identify sources of fecal contamination. This is an important conclusion because the next step is to use ARA for projects such as Total Maximum Daily Loads (TMDL) that will eventually call for load reductions and best management practices (BMP) that best address the major sources of fecal contamination.

The objective of chapter was to build and test a known source library based on an existing library that was both adequate in size and representativeness.

MATERIALS AND METHODS

Building the Known Source Library

Human source samples were collected from the Blue Plains Wastewater Treatment Plant and septic system pump-out companies in Northern Virginia and suburban Maryland. Horse and livestock samples were collected from the horse stables in Rock Creek and along riding trails in Rock Creek Park and the Chesapeake and Ohio Canal National Historic Park. Dog and cat samples were collected from pet-boarding facilities in Arlington and McLean, and dog samples from picnic areas and trails in Rock Creek Park, the Chesapeake and Ohio Canal National Historic Park, and East Potomac Park. The wildlife samples were collected wherever they could be found in Rock Creek Park, the Chesapeake and Ohio Canal National

Historic Park, the East Potomac Park, the Potomac Overlook Regional Park, and the Theodore Roosevelt Island. Cattle and additional horse samples were collected from farms near the Potomac River upstream of the metropolitan area. Individuals assisting in the known source collections were from organizations that included the Washington D.C. Metropolitan Council of Governments, National Park Services Police, Virginia Department of Game and Fish, students from the Virginia-Maryland Regional College of Veterinary Medicine (Middleburg Equine Lab), and the Virginia Cooperative Extension Service.

Isolation of *Enterococci*

Whenever fecal samples from the same source were collected at the same time, the samples were combined to create a composite sample. Isolation from known source samples was intiated by adding 10.0g of solid material or 10.0mL of liquid material (septic tank water) to 90mL of a phosphate buffer solution to create a 1:10 dilution. Successive dilutions of 1:100, 1:1000, and 1:10,000 were made from the original sample. Each dilution was spread plated onto mEnterococcus Agar (BBL, Cockeysville MD) in 15x100mm sterile petri dishes and incubated at 37^{o}C for 48 hours. A single colony of *Enterococci*, indicated by a burgundy color, was transferred via a steril toothpick to an individual well in a 96-micro well plate containing 0.2mL of Enterococcosel broth (BBL, Cockeysville MD). Ten to twelve colonies were transferred from each known source sample (24 to 48 colonies per sample) to the microwell plates. Microwell plates were incubated at 37^{o}C for 48 hours. The growth of enterococcus was confirmed by a change of broth color from yellow to black, caused when the enterococci isolates hydrolyze esculin.

Antibiotic Resistance Analysis

Thirty concentrations of nine antibiotics were used to develop antibiotic resistance profiles for the enterococci isolates. Stock solutions of each antibiotic were prepared from commercial powders obtained from Sigma Chemical Co. (St. Louis, MO). The stock solution preparations are listed below (Table 1).

Table 1: Antibiotic Stock Solution Preparations

Antibiotic	***Commerical Formulation***	***Solvent***	***Stock Concentration (mg/mL)***
Amoxicillin	Amoxicillin	1:1 water:methanol	2.5
Cephalothin	Cephalothin	Distilled water	10
Chlorotetracycline	Chlorotetracycline HCl	1 N NaOH	10
Erythromycin	Erythromycin	1:1 water:ethanol	10
Neomycin	Neomycin Sulfate	Distilled water	10
Oxytetracycline	Oxytetracycline HCl	1:1 water:methanol	10
Streptomycin	Streptomycin Sulfate	Distilled water	10
Tetracycline	Tetracycline HCl	Methanol	10
Vancomycin	Vancomycin Sulfate	1:1 water:ethanol	10

Trypticase Soy Agar (TSA, BBL, Cockeysville MD) was prepared in 250mL Erlenmeyer flasks, autoclaved, and cooled to 50C^{o}. The antibiotic stock solutions were then added to the TSA to create the appropriate concentrations (Table 2), swirled gently, and poured into 15x100mL plastic petri dishes, so that each dish contained one antibiotic at a specific concentration. A control plate was also poured, which consisted only of the TSA. Isolates from the 96-mirco well plate were then transferred 48 isolates at a time onto each of the thirty-one antibiotic/concentration plates via a 48 prong replica plater (Sigma Chemical Co, St Louis MO). The innoculant from the replica plater was allowed to soak into the media and dry to prevent smearing, and plates were incubated at 37C^{o} for 48 hours. Every isolate on every plate was then compared to corresponding isolate on the control and was scored for growth (a 1 was recorded) or no growth (a 0 was recorded).

Table 2: Antibiotic Concentrations after addition to TSA

Antibiotic	*Plate Concentrations (ug/L)*
Amoxicillin	2.5
Cephalothin	10, 15, 30, 50
Chlorotetracyclin	60, 80, 100
Erythromycin	10, 15, 30, 50
Neomycin	40, 60, 80
Oxytetracycline	20, 40, 60, 80, 100
Streptomycin	40, 60, 80, 100
Tetracycline	10, 15, 30, 50, 100
Vancomycin	2.5

Statistical Analysis

A total of 1,806 isolate antibiotic resistance profiles from known sources were entered into a spreadsheet in the statistical package JMP-In (Version 5.0 for Windows, SAS Institute, Inc.). Each isolate was assigned to the source from which it came (Bird, Human, Livestock, Pets, and Wildlife). Logistic Regression (LR) was run on the database to place each isolate in the correct category based on its profile in comparison to the similarity it has to every other isolate profile. From this information the RCC's were computed by dividing the number of isolates the model assigned to each category by the number of isolates that were originally assigned to each category.

It was also imperative to know and take into account the potential pollutant sources in each watershed. According to Wiggins (1996), classification rates can be improved further if one or more of the possible sources is excluded a *priori*. So in order to account for only localized sources of fecal contamination, two subset libraries were created. The original library was used for comparison of unknown stream samples taken from the Potomac River. The Anacostia River library contained only the sources Bird, Human, Pets, and Wildlife. Livestock was excluded from the library because land use surveys indicated that there are no livestock present in the Anacostia River watershed (D.C. Department of

Health et al., 2002). The Rock Creek library only contained the sources Bird, Horse, Human, Pets and Wildlife. All livestock except horse were excluded as a potential source in the Rock Creek library because land use surveys indicated that horse was the only livestock present in the Rock Creek watershed (D.C. Department of Health et al., 2002).

Testing for Adequate Library Size and Representativeness

To test the libraries for adequate size, all known source antibiotic resistance patterns were randomly assigned to source categories (bird, human, livestock, pets, or wildlife) as described by Whitlock et al. (2002). When logistic regression was performed, the ARCC for the randomly assigned dataset should be approximately equal to the probability that an isolate would be assigned to a category by chance. The latter number is calculated by dividing the number of source categories into 100. The chance probability that an isolate will be assigned to a source from the Potomac and Rock Creek libraries is 20% (five potential sources), while the chance probability that an isolate will be assigned to a source in the Ancaostia library is 25% (four potential sources). To test the libraries for representativeness, periodically, known source isolates were added to the library and treated as unknowns and the classification rates were recorded (Hagedorn et al., 1999).

RESULTS AND DISCUSSION

Analysis of the known source libraries

The first measures of the usefulness of a library are the RCC's. Using the logistic regression model, the Potomac River library produced RCC's of 95.6% for bird, 86.3% for human, 92.4% for livestock, 91.7% for pets, 79.7% for wildlife, and an ARCC of 89.1% (Table 3). The Anacostia River library produced RCC's of 95.9% for bird, 91.6% for human, 94.1% for pets, 90.4% for wildlife, and an ARCC of 93.0% (Table 4). The Rock

Creek library produced RCC's of 95.6% for bird, 97.5% for horse, 90% for human, 94.1% for pets, 86.6% for wildlife, and an ARCC of 92.8% (Table 6). The RCC's were well above what would be the expected distribution by random chance (20% for each source in the Potomac and Rock Creek, and 25% for each source in the Anacostia). The RCC's were comparable to, if not higher than those produced by other ARA studies performed by Hagedorn et al. (1999), Wiggins et al. (1999), and Harwood et al. (2000). In comparison of classification rates obtained by molecular BST methods, Dombeck et al. (2000) using Rep-PCR, obtained RCC's ranging from 78% to 100% with seven source categories, Carson et al. (2001) using Ribotyping, obtained RCC's ranging from 92% to 98% with three sources.

Table 3: Potomac River known source library

Predicted Source	Percent of known source isolates assigned to each source (Number of isolates)					
	Bird	Human	Livestock	Pets	Wildlife	Total Isolates Assigned to Source
Bird	**95.6** (237)	1.2 (5)	1.1 (8)	2.4 (4)	3.5 (9)	263
Human	1.2 (3)	**86.3** (371)	2.6 (28)	1.8 (3)	5.0 (13)	415
Livestock	1.6 (4)	4.2 (18)	**92.4** (646)	2.4 (4)	11.1 (29)	701
Pets	0.4 (1)	3.0 (13)	0.6 (4)	**91.7** (154)	0.8 (2)	174
Wildlife	1.2 (3)	5.4 (23)	2.3 (16)	1.8 (3)	**79.7** (208)	253
Total Isolates	248	430	699	168	261	1806

*ARCC 89.1%

Table 4: Anacostia River known source library

Predicted Source	Percent of known source isolates assigned to each source (Number of isolates)				
	Bird	Human	Pets	Wildlife	Total Isolates Assigned to Source
Bird	**96.0** (238)	0.47 (2)	2.4 (4)	2.7 (7)	251
Human	2.0 (5)	**91.6** (394)	3.0 (5)	4.6 (12)	416
Pets	0.8 (2)	3.7 (16)	**94.1** (158)	2.3 (6)	182
Wildlife	1.2 (3)	4.2 (18)	0.6 (1)	**90.4** (236)	258
Total Isolates	248	430	168	261	1107

*ARCC 93.0%

Table 5: Rock Creek known source library

Predicted Source	Percent of known source isolates assigned to each source (Number of isolates)					
	Bird	Horse	Human	Pets	Wildlife	Total Isolates Assigned to Source
Bird	**95.6** (237)	1.3 (3)	0.7 (3)	2.4 (4)	2.7 (7)	254
Horse	2.0 (5)	**97.5** (229)	1.9 (8)	0 (0)	3.8 (10)	252
Human	0.8 (2)	0.4 (1)	**90** (387)	3.0 (5)	4.6 (12)	407
Pets	0.8 (2)	0 (0)	3.7 (16)	**94.1**(158)	2.3 (6)	182
Wildlife	0.8 (2)	0.9 (2)	3.7 (16)	0.6 (1)	**86.6** (226)	247
Total Isolates	248	235	430	168	261	1342

*ARCC 92.8%

As Wiggins suggested (1996), manipulating the library to contain only the sources that are found in the watershed raised classification rates as seen by both the Anacostia and Rock Creek libraries.

From the tables above, expected frequencies of misclassification (EFMC) can be calculated. This provided important information about the lower limit of detection for which the library can predict a source with confidence. This suggestion for analysis was reported by Harwood et al. (2000). Expected frequencies of misclassification were the percentages of the known source isolates that were placed in the wrong source category by the model. The minimum detection level that samples had to meet to be classified as significant was calculated by adding the number of isolates for each category that were placed in the wrong category (non bold numbers in parentheses going from left to right) divided by the total number of all of the isolates in the source categories not in question. For example, the expected frequency of misclassification for wildlife in the Rock Creek library would be 26 (2+2+16+1) divided by 1,081 (248+235+430+168), which equals 1.9%. The expected frequencies for misclassification for the sources in the Potomac library were; 1.7% for bird, 3.4% for human, 5.0% for livestock, 1.2% for pets, and 2.9% for wildlife. The

expected frequencies for misclassification in the Anacostia library were; 1.5% for bird, 3.2% for human, 2.6% for pets, and 2.6% for wildlife. The expected frequencies for misclassification for the Rock Creek library were; 1.6% for bird, 2.1% for horse, 2.2% for human, 2.0% for pets, and 1.9% for wildlife. So on average, unknown sources with identified sources contributions below 2.8% in the Potomac, 2.5% in the Anacostia, and 2.0% in Rock Creek were considered insignificant. Upon calculation of the EFMC's, approximately 22% of the unknown stream isolates were below detection limits in the Potomac River (Appendix A, Table 4), 11% of the unknown stream isolates were below detection limits in the Anacostia River (Appendix A, Table 5), and 16% of the unknown stream isolates were below detection limits in Rock Creek (Appendix A, Table 6).

An alternative and more stringent determination of what is to be considered the minimum limit of detection is the minimum detectable percentage (MDP), described by Whitlock et al. (2002). This particular analysis was not used in the interpretation of results associated with the current project, only as an introduction and comparison of an alternative means for representing the lowest detection limits. The EFMC, described above, was determined by averaging the rates of misclassification for each source. The MDP, however, was calculated as four times the standard deviation of the mean rates of misclassification for each source added to the overall mean rate of misclassification. The MDP for the Potomac River Library was 29.5% (Appendix A, Table 1), the MDP for the Anacostia River Library was 16.9% (Appendix A, Table 2), and the MDP for the Rock Creek Library was 25.9% (Appendix A, Table 3). Whitlock et al. (2002) reported an MDP of 24.6%, so the MDP for each library in this study is within reason. Upon calculation of the MDP's, approximately 80% of the unknown stream isolates fell below the minimum detectable limit in the Potomac

River library (Appendix, Table 4), 50% of the unknown stream isolates fell below the minimum detectable limit in the Anacostia River library, and 70% of the isolates fell below the minimum detectable limit in the Rock Creek library (Appendix, Table 6).

Testing for Adequate Library Size and Representativeness

Adequate size and representativeness of the known source library was critical for establishing confidence in the results obtained for unknown environmental isolates. A library is representative when the sources are randomly assigned to the isolate profiles and the RCC's are approximately equal to the probability that an isolate would be assigned to a source category by chance. Random assignments to the Potomac library yielded an ARCC of 25.6% (Table 6), while the probability that an isolates would be assigned to a source category by chance was five (number of sources) out of 100 or 20%. Random assignments to the Anacostia library yielded an ARCC of 31.3%, while the probability that an isolates would be assigned by chance was 25% (Table 7). Random assignments to the Rock Creek library yielded an ARCC of 26.5%, while the probability that an isolate would be assigned by chance was 20% (Table 8). Each library was confirmed to be of appropriate size, as the chance for artificial clustering was minimal, demonstrated when the random assignments to source categories were approximately equal to the probability of assignment to a source categories by chance.

Table 6: Potomac Library RCC's from Random Assignments of Isolate to Sources

Predicted Source	*% Isolates Assigned to Source*
Bird	22.7
Human	30.2
Livestock	34.1
Pets	16.3
Wildlife	24.9
AVG	25.6

Table 7: Anacostia Library RCC's from Random Assignments of Isolate to Sources

Predicted Source	*% Isolates Assigned to Source*
Bird	27.8
Livestock	35.4
Pets	35.4
Wildlife	26.5
AVG	31.3

Table 8: Rock Creek Library RCC's from Random Assignments of Isolate to Sources

Predicted Source	*% Isolates Assigned to Source*
Bird	31.6
Horse	27.5
Human	25.0
Pets	29.5
Wildlife	19.0
AVG	26.5

Library representativeness as suggested by Hagedorn et al. (1999), can be established by periodically adding known sources to the library and treating them as unknowns, to observe the RCC's. If the library is able to classify the source approximately as well as the isolates from that source in the library, then the library can be considered representative of the population of fecal bacteria found in the watershed (Wiggins et al., 2003). In August 2002, horse and human samples were collected and analyzed as unknowns in the Potomac River Library to determine library the representativeness of these two sources during that period of the study. A second sample of human was analyzed as an unknown in June 2003 (Table 9). In the known source library the RCC for livestock was 92.4%, while the RCC for human was 86.3%. The library was able to correctly classify the known samples at high enough rates (between 70 and 85%) to confidently claim that the library is representative of the potential isolates that may be collected in environmental samples.

Table 9: Test Results for Library Representativeness

Sampling Date	*Known Source*	*% Correct Classification of Known Source Isolates*
8/16/02	Horse Composite	85.4%
8/21/02	Blue Plains Septage	79.2%
6/10/02	Blue Plains Septage	70.2%

SUMMARY

The libraries constructed for the Potomac River, the Anacostia River, and Rock Creek were found to be adequate for the BST study in the Washington D.C. area. The ARCC's of the library were high (89.1%, 93.0%, and 92.8% respectively), as Harwood et al. (2000) claim that regulatory agencies find ARCC's with as low as 60-70% to be adequate. The libraries were also thoroughly tested for adequate size and representativeness. The libraries were of proper size as the random assignment of isolates was within 5% of the probability that an isolate would be assigned to a category by chance. This minimizes the occurrence of artificial clustering. The libraries were also representative, with classification rates of these isolates close to those of the known source library for the respective sources. Human was tested for representativeness twice over the course of the study and the rate of correct identification decreased. This indicates that the library is most likely not temporally stable, but was adequate for the purposes of a one year study.

The results of the library analysis indicate that the library was representative of the isolates in the watersheds associated with each of the waterways. However the geographic and temporal stability were not analyzed in any detail as a part of this project. It is important to remember that not much information exists about the long-term stability of the isolates or the compatibility of isolates from one region to the next. Also, comparing environmental isolates to libraries in general have brought about some criticism regarding

population dynamics of indicator organisms after deposition. Despite these issues, the confidence of this method to predict sources of fecal contamination from environmental samples remains high.

REFERENCES

Carson, C.A., B.L. Shear, M.R. Ellersieck, A. Asfaw. 2001. Identification of Fecal *Escherichia coli* from Human and Animals by Ribotyping. Applied and Environmental Microbiology. 67 (4): 1503-1507.

Dombeck, P.E., L.K. Johnson, S.T. Zimmerley, M.J. Sadowsky. 2000. Use of Repetative DNA Sequences and the PCR to Differentiate *Eshcerichia coli* Isolates from Human and Animal Sources. Applied and Environmental Microbiology. 66 (6): 2572-2577.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneau Jr. 1999. Determining Sources of Fecal Pollution in a Rural Virginia Watershed with Antibiotic Resistance Analysis Patterns in Fecal Streptococci. Applied and Environmental Microbiology. 65 (12): 5522-5531.

Harwood, V.J., J. Whitlock and B. Withington. 2000. Classification of Antibiotic Resistance Patterns of Indicator Bacteria by Discriminate Analysis: Use in Predicting the Source of Fecal Contamination in Subtropical Waters. Applied and Environmental Microbiology. 66 (9): 3698-3704.

Simpson, J.M, J.W. Santo Domingo, D.J. Reasoner. 2002. Microbial Source Tracking: State of the Science. Environmental Science and Technology. 36 (24): 5279-5288.

Whitlock, J.E., D.T. Jones, V.J. Harwood. 2002. Identification of the Sources of Fecal Coliforms in and Urban Watershed Using Antibiotic Resistance Analysis. Water Research. 36: 4273-4282.

Wiggins, B.A. 1996. Discriminate Analysis of Antibiotic Resistance Patterns in Fecal Streptococci, a Method to Differentiate Human and Animal Sources of Fecal Pollution in Natural Waters. Applied and Environmental Microbiology. 62 (11): 3997-4002.

Wiggins, B.A., P.W. Cash, W.S. Creamer, S.E. Dart, P.P. Garcia, T.M. Gerecke, J. Han, B.L. Henry, K.B. Hoover, E.L. Johnson, K.C. Jones, J.G. McCarthy, J.A. McDonough, S.A. Mercer, M.J. Noto, H. Park, M.S. Phillips, S.M. Purner, B.M. Smith, E.N. Stevens, and A.K. Varner. 2003. Use of Antibiotic Resistance Analysis for Representativeness Testing of Multiwatershed Libraries. Applied and Environmental Microbiology. 69 (6): 3399-3405.

Wiggins, B.A., R.W. Andrews. R.A. Conway, C.L. Corr, E.J. Dobratz, D.P. Doughtery, J.R. Eppard, S.R. Knubb, M.C. Limjoco, J.M. Mettenburg, J.M. Rinehardt, J. Sonsino, R.L. Torrijos, and M.E. Zimmerman. 1999. Use of Antibiotic Resistance Analysis to Identify Nonpoint Sources of Fecal Pollution. Applied and Environmental Microbiology. 65 (8): 3483-3486.

CHAPTER IV

DETERMINING SOURCES OF FECAL CONTAMINATION IN THE POTOMAC RIVER, THE ANACOSTIA RIVER, AND ROCK CREEK USING ANTIBIOTIC RESISTANCE ANALYSIS

INTRODUCTION

Applying bacterial source tracking (BST) methods to projects that provide results that will become an integral part of Total Maximum Daily Load (TMDL) studies or use in human health risk assessments is the ultimate goal and direction of the source tracking community. Protecting the nation's water supply for drinking and recreation is imperative for human health. Though BST is not an official part of the TMDL program, some localities that are facing major water quality issues have taken a proactive stance and have begun to employ source tracking.

Antibiotic Resistance Analysis (ARA) has been proven to be a useful tool for determining sources of fecal contamination in several regions across the country (Wiggins et al. 1996, Parveen et al. 1997, Harwood et al. 2000). It is a BST method that has been field-tested with much success and research is currently underway to address issues of concern related to known source databases (Hagedorn et al. 1999, Whitlock et al. 2002, Booth et al. 2003). ARA is a method that requires simple laboratory techniques, basic equipment, and probably ARA's biggest advantage is the relatively low cost per isolate as compared to many other BST methods (Simpson et al., 2002).

Washington D.C. officials are proactively searching for answers to a variety of water quality issues. In 1996, the District's Department of Health created a list of impaired waters that do not or are not expected to meet water quality standards as required by section 303(d) of the Clean Water Act. The list was updated in 1998 and contained a priority list of those

waters that were the most polluted and was used to determine the waterbodies in critical need of immediate attention. The Potomac River, the Anacostia River, and Rock Creek were placed on this list as a high priority due to violations of water quality standards set for fecal coliforms. The city is required to create a TMDL plan and in addition, was interested in determining sources of fecal contamination for both load allocation purposes and for human health risk assessments (D.C. Department of Health, 2002). ARA was the BST method chosen by the Metropolitan Washington Council of Governments for this study in order to accomplish the objectives set forth for identifying pollutant sources, allocations, and human health risks.

The sources of fecal coliforms to these waterways are ubiquitous. Any means by which fecal matter can be transported to the receiving waters is a potential source. These sources include combined sewer overflows (CSOs), separate sanitary sewer overflows (SSOs), which can result from leaky or undersized sanitary sewer pipes, direct deposits from wildlife and livestock, and stormwater runoff, which includes overland flow and flow conveyed through storm sewer pipes.

The District of Columbia is currently serviced by a combined stormwater/sewer system. In the 1930's the Blue Plains Waste Water Treatment Plant (WWTP) was constructed to hold dry weather sewage flow. Unfortunately wet weather causes frequent exceedances of the transmission capacity to the WWTP, resulting in overflows. According to the Washington Post (Vasquez, 2002), approximately 75 times a year the system becomes overloaded and then flows into emergency pipes where it mixes with sewage from the Blue Plains sewage treatment facility. The emergency pipes empty the sewage/stormwater mixture directly into the Potomac River, the Anacostia River, and Rock Creek. This

amounts to roughly 3 billion gallons of unprocessed wastewater discharge into these waterways annually. The concern lies in the fact that all three of these waterways have designated beneficial uses of primary and secondary contact recreation, protection and propagation of fish, shellfish, and wildlife, protection of human health related to consumption of fish and shellfish, and navigation (D.C. Department of Health, 2002).

The main objective of this portion of the study was to apply ARA as a tool for determining the sources of fecal pollution in the Potomac River, the Anacostia River, and Rock Creek. The effect of seasonal variations and storm events were also addressed as a part of this objective.

This study was also an important test of the applicability of ARA in a different type of environment. It is one of the first comprehensive BST studies to take place in a highly urbanized region. Whitlock et al. (2002) published a study identifying sources of fecal coliforms in Stevenson Creek, in the highly urbanized area of Clearwater Florida. The results of that study found that, in general, wild animals were the dominant contributors to fecal contamination in the watershed when fecal coliform levels were elevated. When fecal coliform levels were low, however, other sources such as human and dog became dominant. Human was found to be a major contributor to low-level background contamination. No other information exists as to the distribution of contamination sources or the fate of the sources as a result of seasonal changes in an urban environment.

The results of this portion of the study were used to provide guidance to the Washington D.C. Department of Health as to the sources of fecal contamination and the relative risks to human health in these three recreational waterways. A TMDL is currently under development that will address current loading, load reductions, and load allocations.

The results of the present bacterial source tracking study will be used only to supplement the information gathered for the TMDL.

MATERIALS AND METHODS

Study Areas

The Potomac River

The Potomac River watershed covers approximately 29,940 square kilometers with drainage coming from Maryland, Virginia, West Virginia, and Pennsylvania. The river enters the District just below the fall line as it transitions between the Piedmont and Coastal Plain provinces. Only 0.5% of the river flows within the boundaries of Washington D.C.. Land use is predominantly agricultural and forested with the most significant development and population located in the Washington metropolitan region. The only combined sewer system is located on the river within the boundaries of the District. The District's Water and Sewer Authority's (WASA) National Pollution Discharge Elimination System (NPDES) indicates that there are 14 combined sewer outfalls (CSO) on the Potomac (D.C. Department of Health, 2002).

The Anacostia River

The Anacostia River watershed is approximately 456 square kilometers with 49% of the drainage area occurring in Prince George's County, 34% in Montgomery County, and 17% in the District of Columbia. Two thirds of the basin lies within the Coastal Plain, and the rest lies within the Piedmont. Land use is mostly residential with 30% park and forest lands evenly dispersed throughout the watershed. Park and open areas include the National Park Service, the National Arboretum, Greenbelt Park, and Beltsville Agricultural Research Center. The industrial manufacturing land use is mostly confined to the tidal area of the

basin, in the sub-watersheds of Hickey Run, Lower Beaverdam Creek, and Indian Creek. As much as 80% of this area is covered by impervious surfaces. The WASA NPDES permit indicates that there are 17 CSO outfalls located on the Anacostia (D.C. Department of Health, 2002).

Rock Creek

The Rock Creek watershed is approximately 161 square kilometers with 80% of the drainage area occurring in Montgomery County and 20% within the District. The creek is mainly in the Piedmont province, with only a small portion being tidally influenced. Land use is predominantly residential, commercial, and park/open land. Rock Creek Park is one of the oldest city parks, and is the site for many recreational activities. The United States Park Police maintain two horse stables within the Park and a private stable is located in Montgomery County just upstream from the District's border. The watershed is also home to the Smithsonian Institution's National Zoo. The WASA NPDES permit indicates that there are 29 CSO outfalls located along Rock Creek (D.C. Department of Health, 2002).

Sampling Regime

A total of fifteen sites were selected for sampling by District Health Officials; three on the Potomac River, six on the Anacostia River, and six on Rock Creek. The stretch of the Potomac River included in this study began at the Three Sister's Bridge, proceed to where Rock Creek enters the Potomac and ended at the Memorial Bridge (Appendix A, Figure 2). The stretch of the Anacostia River included in this study began as the river crossed the D.C. Maryland line, past Pennsylvania Avenue, and ended at Haines Point (Appendix A, Figure 1). The Rock Creek sampling stations began where Rock Creek passed the D.C. Maryland

line, through the area adjacent to the National Zoo, with the last sampling station near Pennsylvania Avenue (Appendix A, Figure 2).

Sampling began in July 2002 and officially ended in April 2003, with the addition of two storm-sampling events (October 2002) that followed a period of very dry weather. The sampling schedule is presented in Table 1.

Table 1: Sampling Schedule for the Potomac River, the Anacostia River, and Rock Creek

Sampling Date	*Sampling Type*
7/17/02	Monthly
8/7/02	Monthly
9/11/02	Monthly
10/9/02	Monthly
10/17/02	Storm 1
10/31/02	Storm 2
11/13/02	Monthly
12/10/02	Monthly
1/8/03	Monthly
2/12/03	Monthly
3/5/03	Monthly
4/2/03	Monthly

Monthly sampling was completed by personnel of the Occoquan Monitoring Lab (Civil and Environmental Engineering, Virginia Tech), contracted by the Metropolitan Washington Council of Governments to perform sampling events as well as the monthly collection of temperature and pH data on each waterway.

Analysis of Stream Samples

Sample Collection

Water samples were collected by Occoquan Monitoring Lab according to the schedule outlined above. Samples were collected in sterile polystyrene bottles, packed on ice and transported for processing within 24 hours.

Isolation of *Enterococci* from Unknown Water Samples

Enterococci counts were obtained by filtering various concentrations of each sample through 0.45 micron filters, and then placing the filters on mENT agar (BBL, Cockeysville, MD) according to Standard Methods (American Public Health Association, 1998). Plates were incubated at 37^{o}C for 48 hours. Results were recorded as Colony Forming Units (CFU's) per 100mL.

Twenty-four isolates from each sample were transferred via sterile toothpicks (Graves, 2003), each to an individual well in a 96-micro well plate containing 0.2mL of Enterococcosel broth (BBL, Cockeysville MD). Microwell plates were incubated at 37^{o}C for 48 hours. The growth of enterococcus was confirmed by a change of broth color from yellow to black, caused when the enterococci hydrolyze esculin.

Antibiotic Resistance Analysis

Thirty treatments of nine antibiotics were used to determine antibiotic resistance patterns for the unknown isolates. The plate concentrations are listed in Table 2.

Table 2: Antibiotic plate concentrations for analysis of unknown stream isolates

Antibiotic	*Plate Concentrations (ug/L)*
Amoxicillin	2.5
Cephalothin	10, 15, 30, 50
Chlorotetracyclin	60, 80, 100
Erythromycin	10, 15, 30, 50
Neomycin	40, 60, 80
Oxytetracycline	20, 40, 60, 80, 100
Streptomycin	40, 60, 80, 100
Tetracycline	10, 15, 30, 50, 100
Vancomycin	2.5

Isolates from the 96-mirco well plate were then transferred, 48 isolates at a time (two samples per plate), onto each of the thirty-one antibiotic/concentration plates via a 48 prong

replica plater (Sigma Chemical Co, St Louis MO). The innoculant was allowed to soak into the media and dry to prevent smearing. Plates were incubated at 37C^{o} for 48 hours. Each isolate on each plate was then compared to corresponding isolate on the control and was scored for growth (a 1 was recorded) or no growth (a 0 was recorded).

Statistical Analysis

The twenty-four isolates from each sample were added to the end of the known source library one sample at a time and logistic regression (LR) was run. In comparing the unknowns with the knowns, the model output included the number and percentage of the unknown isolates that matched each source category, and classified each isolate to its most likely source.

Physical Analysis

During each sampling event, personnel from the Occoquan Monitoring Lab recorded physical data for each sampling location. Field data included pH and temperature. Grab samples were also collected in sterile whirl pack bags, and were enumerated for fecal coliforms and enterococci by the Most Probable Number (MPN) procedure.

RESULTS AND DISCUSSION

Background Information

Fecal Coliform and Enterococci Counts

The monthly colony counts and violations of the water quality standards for fecal coliforms and enterococci were important pieces of background information regarding the need for source tracking and the starting point for locating patterns and trends that have a role in the dynamic state of each waterway. Waters are listed for pathogenic impairments if they exceed the standards set for *E.coli* and enterococci. The microbiological standards for

water quality have been changed since the beginning of this project based on new recommendations by the EPA. The old standard called for the detection of fecal coliforms, stating the organisms should not exceed 200 CFU's/100mL as a geometric mean of multiple samples taken within a thirty-day period or shall not exceed 1000CFU's/100mL at any time (D.C. Department of Health, 2002). The District of Columbia chose not to update bacterial standards, so the Occoquan Monitoring Lab enumerated fecal coliforms and this was the standard adhered to for the purposes of this discussion.

The EPA's newest standards call for no more than 126 *E.coli* per100mL or 33 enterococci per 100mL in freshwater based on a geometric mean of at least five samples taken within a 30-day period. Marine water recommendations call for no more than 35 enterococci per100mL as a geometric mean over a 30-day period (EPA, 2003). No standard existed for enterococci at the beginning of this study, so for discussion purposes, the new standard was used to determine the intensity of the violations of water quality standards for enterococci. Since sampling was not completed multiple times during a particular month as required by the EPA to calculate a geometric mean, the standards used for violation references for fecal coliforms are the maximum monthly limitations (1000CFU's/100mL). The comparison standard used for enterococci was the marine recreational standard of 35 CFU's/100mL. Of important consequence to the water quality in Washington D.C. in terms of total number of pathogenic impairments is a new motion by the EPA to force more stringent standards on states or territories that chose to continue to monitor for fecal coliforms. According to Jutta Schneider, the Director of the Virginia TMDL for the DEQ, the EPA has begun to recommend and will mandate by 2008, a standard of 400 fecal coliforms per 100mL (personal communication, 2003). This standard was not used in the

discussion to follow, but based on the number of violations obtained with the old standard for fecal coliforms verses the new standard for enterococci, the impact of this change will include many more pathogenic impairments than currently exist.

The Potomac River is the largest of the three waterways in this study, with an average flow of approximately 305.5 cubic meters per second. The large volume of the river and the high flow rates result in good dilution and flushing rates for pollutants that enter the waterway (D.C. Department of Health, 2002). Thus, enterococci and fecal coliform counts per sample were in most months significantly lower than in the Anacostia River or Rock Creek. Colony counts by month for the Potomac River are presented in Figures 3 and 4 of Appendix A. Violations of the water quality standards are listed in Table 3.

Table 3: Percent of sampling events violating water quality standards by station on the Potomac River

% Sampling Events Violating Standards at three Sampling Sites			
Indicator	FP01	FP02	FP03
Fecal Coliforms (old standard)	0.0	0.0	0.0
Enterococci (new standard for marine water)	50.0	58.3	66.7

The water quality standards for fecal coliforms were not exceeded during the course of this study, however, the old standard is not as stringent as the new standard recommended by the EPA. The station with the highest fecal coliform counts most consistently was FP03, followed by FP02. The highest observed enumerations occurred during the two storm events and in the month of November immediately following. Violations of the water quality standards for enterococci were fairly consistent between the three sampling sites. The common violations occurred during the two storm events as well as the months of November, January, and March, all correlating to times of significant precipitation or snow

melt. Station FP03 violated water quality standards for enterococci more often than the other two stations.

The Anacostia River is largely a tidal embayment of the Potomac. The volume of water results in moderate dilution, but low flow above the tidal zone causes poor flushing of contaminants. The mean annual stream flow for the Anacostia, as measured at the upstream flow gages, is 3.9 cubic meters per second (D.C. Department of Health, 2002). Colony counts by month are presented in Figures 5 and 6 of Appendix A. Violations of the water quality standards are detailed in Table 4.

Table 4: Percent of sampling events violating water quality standards by station on the Anacostia River

% Sampling Events Violating Standards at Six Sampling Sites						
Indicator	FA01	FA02	FA03	FA04	FA05	FA06
Fecal Coliforms (old standard)	50.0	25.0	33.0	20.0	27.3	10.0
Enterococci (new standard for marine water)	91.7	91.7	75.0	83.0	50.0	58.3

Site FA01 violated the water quality standards for both enterococci and fecal coliforms most often. Enterococci levels at station FA01 appeared to be most elevated with respect to the other sampling stations during the dry months of July through October. Sites FA02 and FA03 had the highest densities of enterococci during the storm events and the month of November. Enterococci and fecal coliform densities both appeared to become less significant at each progressive sampling station downstream, indicating the largest impact occurs further upstream. FA01 and FA02 appeared to have elevated levels of fecal coliforms during the dry months of July through October, while FA04 and FA05, with respect to the other stations, had the most substantial counts during the storm events and in the month of November. All sites violated water quality standards for fecal coliforms and enterococci during both storm events, and the months of November and January.

Rock Creek is fairly shallow and swift. The low volume in the creek results in poor dilution potential, but the swift flow rates allow for good flushing rates. The average flow rate in Rock Creek is approximately 1.8 cubic meters per second (D.C. Department of Health, 2002). Colony counts by month are presented in Figures 7 and 8 of Appendix A. Violations of the water quality standards are presented in Table 5.

Table 5: Percent of sampling events violating water quality standards by station on Rock Creek

% Sampling Events Violating Standards at Six Sampling Stations						
Indicator	FR01	FR02	FR04	FR05	FR06	FR07
Fecal Coliforms (old standard)	33.3	41.7	50.0	50.0	63.6	66.7
Enterococci (new standard for marine water)	91.7	100.0	83.3	91.7	91.7	100.0

There appeared to be no trend for enterococci densities as there was in the Anacostia and Potomac, but it did appear that at each progressive station downstream, the number of fecal coliform violations increased. Station FR04 had the fewest violations and the overall lowest levels enterococci. Sites FR06 and FR07 were most elevated, with respect to the other stations from December through March for both organisms and had the most significant levels of enterococci during the storm events and the month of November. The highest levels of enterococci that were detected during dry weather occurred at stations FR01 and FR02.

Colony counts for enterococci and fecal coliforms were correlated to the flow of each waterbody (Appendix A, Figure 9) to evaluate the dilution potentials. As was expected, the larger the flow, the lower the densities of the two indicator organisms. Enterococcus densities were linearly correlated to flow with a correlation coefficient of 1.0, while fecal coliform densities and flow were still linearly correlated, the correlation coefficient was 0.87.

Physical Data Analysis

Personnel from the Occoquan Monitoring Lab recorded pH and temperature data during each sampling event for each site (Tables 6, 7, and 8).

Table 6: Temperature and pH Data for the Potomac River by Sampling Station

	FP01		*FP02*		*FP03*	
	Temp C°	pH	Temp C°	pH	Temp C°	pH
Mean	13.7	7.4	13.6	7.4	13.7	7.5
Minimum	1.3	7.2	1.3	7.2	1.0	7.1
Maximum	28.9	7.9	29.0	7.6	29.0	7.8
Stdv	10.0	0.19	10.1	0.12	10.1	0.2

Table 7: Temperature and pH Data for the Anacostia River by Sampling Station

	FA01		*FA02*		*FA03*		*FA04*		*FA05*		*FA06*	
	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH
Mean	13.0	7.0	13.0	6.8	13.3	6.8	14.0	6.8	14.3	7.0	14.1	7.2
Minimum	1.3	6.9	1.1	6.5	1.9	6.5	2.4	6.4	3.0	6.4	2.5	6.9
Maximum	26.4	7.2	26.4	7.0	27.2	7.2	27.8	7.6	28.5	7.9	27.7	7.4
Stdv	9.3	0.1	9.5	0.2	9.6	0.2	9.5	0.3	9.8	0.4	10.0	0.2

Table 8: Temperature and pH Data for Rock Creek by Sampling Station

	FR01		*FR02*		*FR04*		*FR05*		*FR06*		*FR07*	
	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH	Temp C°	pH
Mean	11.3	7.4	11.3	7.3	11.5	7.4	11.5	7.4	11.6	7.3	11.6	7.1
Minimum	2.0	6.8	2.0	6.9	2.0	6.9	2.0	6.9	2.0	7.0	2.0	6.8
Maximum	23.0	7.9	23.0	7.7	23.5	8.0	24.5	8.0	24.5	7.6	24.0	7.2
Stdv	7.8	0.3	7.8	0.2	7.9	0.2	8.1	0.2	8.1	0.2	8.0	0.1

The temperature and pH were all within acceptable ranges for each waterway. Maximum temperatures were recorded in July on the Potomac and Anacostia, and in August on Rock Creek. Lowest temperatures were recorded in February in all three waterways. The mean pH levels in the Anacostia River were lower than those of the Potomac and Rock Creek.

Precipitation was recorded daily by the National Weather Service Forecast Office at the Ronald Reagan Washington National Airport. Precipitation totals were tabulated for the

sampling period and are presented in Figure 10 of Appendix A (National Weather Service Forecast Office, 2003).

Average colony counts for both enterococci and fecal coliforms recorded over the sampling period were compared to precipitation between sampling events (Appendix A, Figures 11-13) in order to determine where the highest indicator organism densities occurred in relation to precipitation. Spikes in colony counts for enterococci and fecal coliforms were pronounced from the time period between the first storm event and the month of November following the second storm event. Even though the precipitation appears much higher from the period of December through March, a majority of the precipitation was in the form of snow and did not have an immediate affect on runoff events as did the frequent rainfall that occurred during the latter half of October 2002.

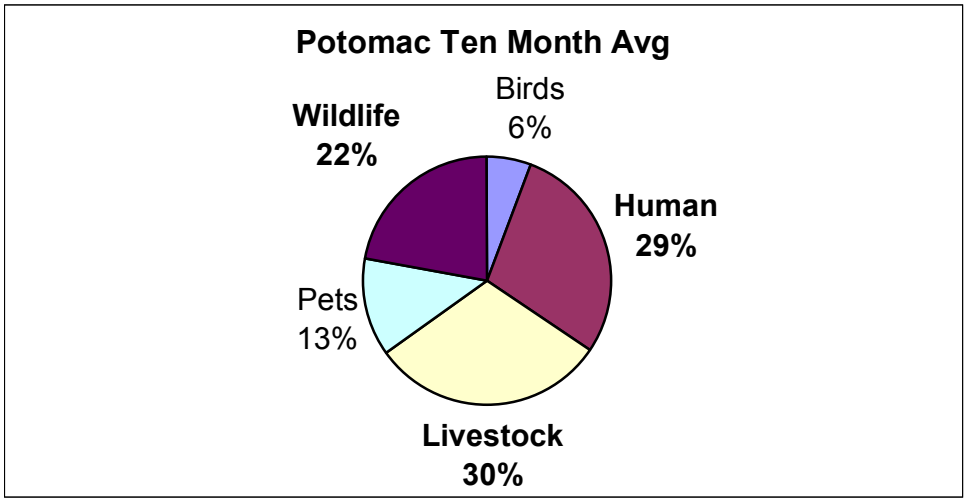
ARA Results

Whenever possible, twenty-four enterococci isolates from each water sample were analyzed using ARA and LR (during periods of high dilution, twenty four isolates may not be present in each sample). The results were analyzed by station and by waterway in an attempt to locate meaningful patterns for source contamination. The effect of storm events and seasonality trends were also determined. In all cases where a major or dominant source was determined, a pairwise comparison was computed for each combination of pairs. Significant differences were detected if the F value exceeded the critical value and if $p < 0.001$ (Whitlock et al. 2002). Sources were also considered significant contributors to the total fecal loading if the contribution was greater than the minimum detectable limit as determined for each waterway.

Potomac River

During regular monthly sampling (July 2002-April 2003), 621 isolates were analyzed from the Potomac River. The average monthly source contributions for the entire sampling area, excluding the two storm events, were determined first in order to obtain an accurate picture on what to expect as the major and minor sources (Figure 1). Dominant sources are highlighted in bold.

Figure 1: Ten Month Average of Source Contributions from all stations on the Potomac River



In order to determine the dominant signatures and relate the significant difference between the mean values of each source, a pairwise comparison (analysis of variance via the t-test) was computed for each combination of pairs. Significant differences were detected if $p < 0.001$ (Whitlock et al. 2002). A more common p value of 0.05 is typically used for this type of test for significance (Graves, 2003), however, a more stringent p value was used in this study in order to provide a greater level of confidence in the significant differences detected. All comparisons between source distributions for dominant signatures and significant differences were computed in this manner. The largest source contribution was attributed to livestock at 30%, but the human and wildlife contributions were not

significantly different. The test for signficance results are presented in Appendix B, Table 1. The predominant land use in the Potomac River watershed is agriculture and forested. This explains why livestock and wildlife were dominant signatures. The livestock signature was attributed to farms upstream of the District of Columbia. The dominant human signature was most likely a result of the 14 combined sewer overflow points on the Potomac River that discharge 1.1 billion gallons (4.8 billion liters) of combined sewage per year (D.C. Department of Health, 2002).

Minimum detection limits for each source were calculated in order to provide lower bounds for what was considered a significant contribution by a source (Table 9). The procedures were an adaptation of the expected misclassification frequency described by Harwood et al. (2000).

Table 9: Minimum Source Contributions for Significance in the Potomac

Source	*Minimum Detection Limit (%)*
Birds	1.7
Human	3.4
Livestock	5.0
Pets	1.2
Wildlife	2.9

The percent of sampling events that had source contributions below minimum detection limits are presented in Table 10, given by sampling station and source.

Table 10: Percent of Source Contributions Below Minimum Detection Limit in the Potomac

Station	*% Below Min. Detection*				
	Birds	*Human*	*Livestock*	*Pets*	*Wildlife*
FP01	33.3	8.3	0.0	50.0	8.3
FP02	50.0	0.0	0.0	25.0	8.3
FP03	33.3	0.0	16.7	33.3	8.3

On an average basis for the entire sampling region on the Potomac, bird and pets appeared to be the least significant sources of fecal pollution to the Potomac. These two sources each

had contributions below their respective minimum detection limits in over 36% of the sampling events. Source contributions fell below the detection limits for wildlife in 8.3% of the sampling events, livestock in 5.6% of the sampling events, and human in only 2.8% of the sampling events.

It was also important to look for major source contributions at each sampling location as an average over the sampling period to determine if there was a greater impact by one of the major sources in any particular area (Table 12). Brief site descriptions are presented in Table 11.

Table 11: Sampling Site Descriptions for the Potomac River

Station	Location	Description
FP01	Three Sister's Bridge	• Commercial shops and restaurants adjacent to river • Boat loading docks • No wooded areas • Numerous birds • Parking lots and sidewalks directly adjacent to river • Easy public access for pet walking • No banks, concrete walls or boat docks
FP02	Rock Creek	• Where Rock Creek enters Potomac • Adjacent to National Park (a nature conservancy island) • Wldlife and bird numerous
FP03	Memorial Bridge	• Wooded on both banks • Numerous birds • Abundant space for wildlife

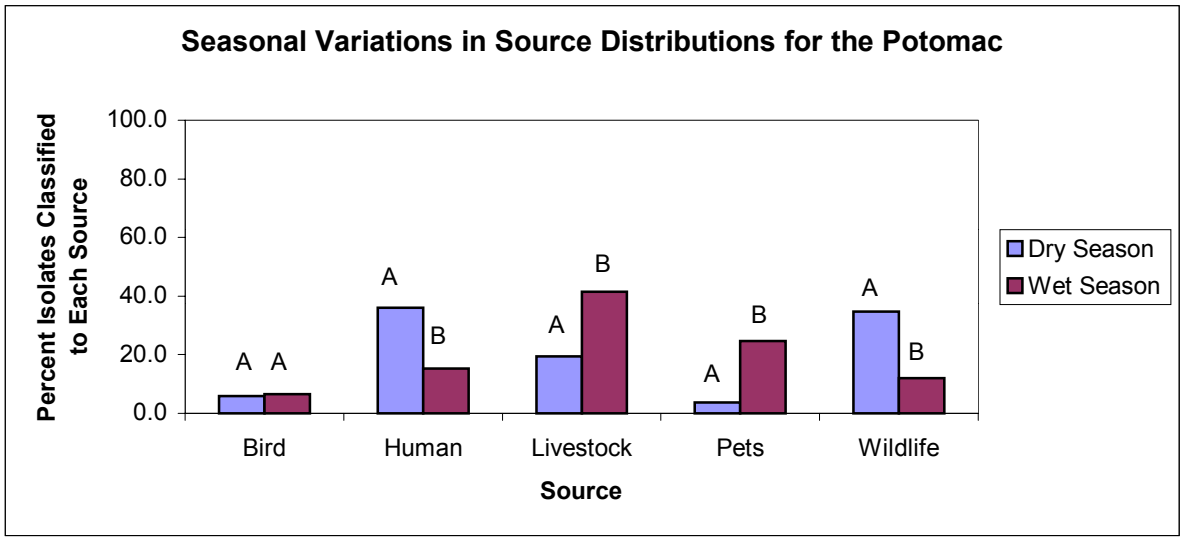
Table 12: Source Distributions at Each Sampling Station on the Potomac as a Ten-Month Average

Sampling Location	*% Bird*	*% Human*	*% Livestock*	*% Pets*	*% Wildlife*
FP01	8.4	21.7	34.3	15.6	20.0
FP02	2.9	32.3	30.2	9.2	25.4
FP03	5.8	32.3	27.4	13.6	20.9

The dominant source contributions at site FP01 was livestock followed by human and wildlife. As an average over the sampling period, all sources were found at concentrations that were considered significant according to Table 9. Human, wildlife, and livestock were all found to be the dominant sources at station FP02, with small but still significant contributions from bird and pets. At station FP03 human and livestock were found to be the major contributors, followed by wildlife. Bird and pets were also above the minimum detection levels. The human signature was most substantial at stations FP02 and FP03, while the pet and bird signatures were most substantial at stations FP01 and FP03. The wildlife signature was the most substantial at station FP02, adjacent to the National Park and the livestock signature was most substantial at FP01, the furthest station upstream, and closest to the agricultural input coming from Maryland farms.

Determining the effect of seasonal variations provides key information for remediation approaches. The average for the entire sampling area was analyzed for the impact of seasonality. The dry season was considered to be the months between July and September, while the wet season occurred from January to March (Appendix A, Figure 10). Source contributions in the wet and dry season for the Potomac River are presented in Figures 2.

Figure 2: Dry and Wet Season Source Contributions on the Potomac River



*Common letters denote means that were not statistically different (p>0.001)

Wildlife and human were the dominant sources during the dry season, while livestock was the dominant source during the wet season. One possible explanation for the seasonal pattern for wildlife is that during the hot dry months, wildlife moved to the river to obtain water and stay cool. The elevated human during the dry months was unexpected. This trend however, may indicate that human was a significant contributor to background levels of fecal contamination, quite possibly from a breakdown in the infrstructure of the sewer system, causing a continual leak of human fecal pollution. The human signature was also most likely then masked by other signatures such as livestock and pets that increased significantly during periods of high runoff. Whitlock et al. (2002) found similar results and suggested that human was most pronounced in the absence of other contributions that occurred during runoff events. The livestock signature was elevated during the wet months because of the increase in runoff from agricultural areas upstream on the Potomac River. On average for the entire sampling area, all sources were above their minimum detection limits, meaning all were significant sources of contamination. Significant differences between the

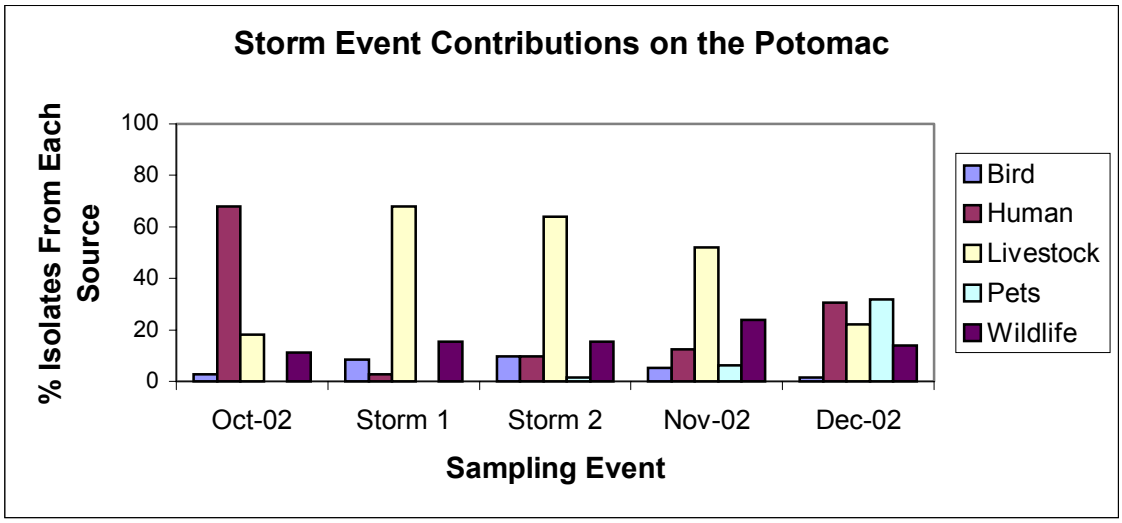
dry and wet season were found among all sources except bird, indicating the important role of seasonality on source distributions. Birds were a minor signature in both the wet and dry seasons, and appeared to be unaffected by seasonality.

Source contributions in the wet season and dry season for each sampling location are presented in Figures 1-3 in Appendix B. During the dry season, stations FP01 and FP02 had dominant human and wildlife signatures, while station FP03 had dominant signatures of human, wildlife, and livestock. The pet signature was below the minimum detection limit at sites FP01 and FP03, while bird was below the detection limit at site FP02. Wet season source distributions were drastically different at all three stations, indicating that there was a strong seasonal pattern. At station FP01, the dominant sources became pets and livestock, at station FP02 the dominant source became livestock, and at station FP03 livestock remained a dominant source, while the pet signature became dominant as well. At each station livestock and pet increased from the dry to wet season, while human and wildlife decreased. The wildlife signature was most substantial during the dry season at station FP02, which is adjacent to the National Park (Roosevelt Island). During the wet season, livestock was most substantial at FP01 and FP02, indicating that the impact of the livestock contribution to fecal contamination levels were influenced by runoff events from upstream on the Potomac.

Storm events play a significant although short-term role on source distributions. Samples were captured from two storm events as designated by the Occoquan Monitoring Lab. The samplings took place on October 17 and October 31, 2002. The designated storm events were captured with no regular monthly sampling between, and it was also the end of the regular dry season. In order to best show the impact of the storm events, source

distributions are provided in the month directly prior to the events and the two months following (Figure 3).

Figure 3: Average Source Contributions on the Potomac in the Months Surrounding Storm Events

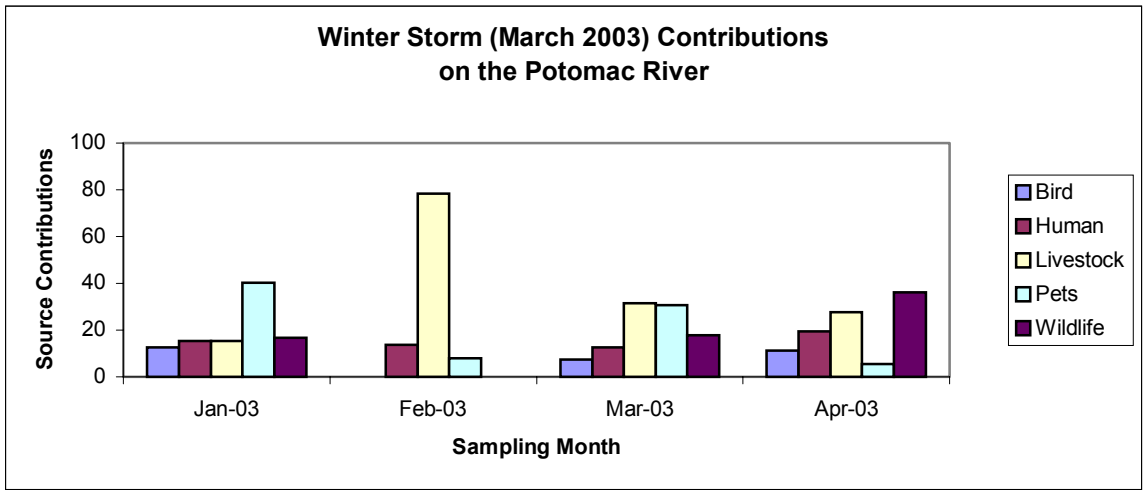


In October, prior to significant rainfall, human was by far the dominant signature at 68%. During the first storm event, the human signature dropped substantially to below the minimum detectable limit. During the second storm event, human rose slightly to 9.7%. Following the storms, the human contribution was on the rise as it reached 30.6% by December. Livestock was another source that was impacted by the storm events. The initial contribution prior to rainfall was 18.1%. During both storm events, livestock contributions were above 60%. As the rainfall ceased, livestock contributions dropped to 22% by December, just slightly above dry season levels. The bird signature increased slightly during each storm event, while the reaction for wildlife appeared to be delayed as major increases only occurred after the second storm event. The pet signature appeared unaffected by rainfall. Human was a major source in dry weather, but as rainfall began the human signature was diluted and possibly masked by the increase in the livestock signature. The

effects of the storm events were similar at all three sampling locations, and they followed the general trend that is shown by the average for the entire river (Appendix B, Figures 4-6).

The sampling month of March was designated as a winter (wet season) storm event in order to compare and contrast the effects of storm events within the different seasons. March was chosen because the precipitation totals in the days prior to the sampling event were 5.8 inches of snow, that quickly melted and 16.7inches of rain (Appendix A, Table 10). In order to best show the impact of the storm event, source distributions are provided in the two months directly prior to the event (the wet season) and the month following (Figure 4).

Figure 4: Average Source Contributions on the Potomac in the Months Surrounding a Winter Storm Event



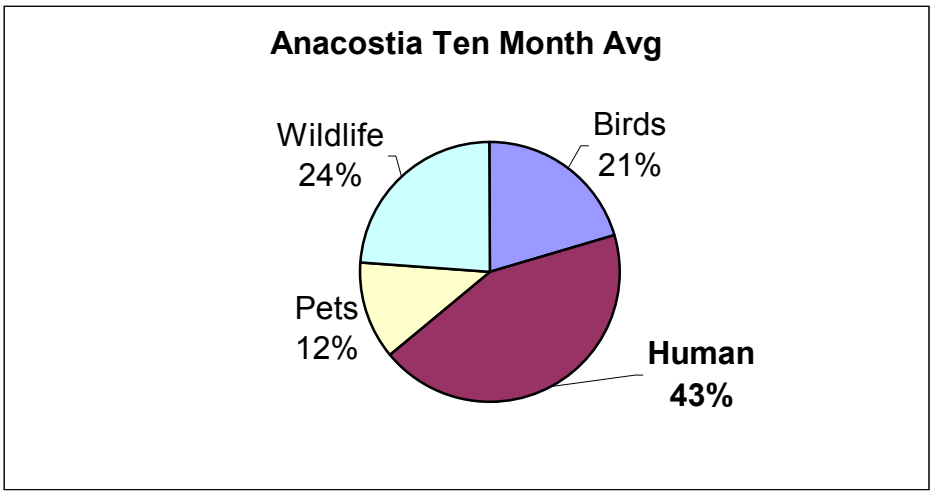
It was much more difficult to find a clear pattern for the effect the storm event had on source distributions during the wet season. Precipitation continuously fell between all sampling events shown in Figure 4. Unlike the storm events recorded in October, no extraordinarily large human signature was detected, and in March, all source contributions were more evenly distributed as compared to source distributions in the first two storm events. The month of February had the most recorded snowfall (20.3 inches), but the least amount of rain (2.9 inches) for the wet season. This may have something to do with the spike in

livestock contribution and the absence of bird and wildlife signatures in the February samples.

Anacostia River

During regular monthly sampling 1,483 isolates were analyzed from the Anacostia River. The average monthly source contributions, excluding the two storm events, for the Anacostia River were determined first in order to obtain an accurate depiction of what to expect as the major and minor sources (Figure 5). The dominant signature is highlighted in bold.

Figure 5: Ten-Month Average of Source Contributions from all stations on the Anacostia River



In order to determine the dominant signatures and relate the significant difference between the mean values of each source, a pairwise comparison (analysis of variance via the t-test) was computed for each combination of pairs. Significant differences were detected if $p < 0.001$ (Whitlock et al. 2002). A more common p value of 0.05 is typically used for this type of test for significance (Graves, 2003), however, a more stringent p value was used in this study in order to provide a greater level of confidence in the significant differences detected. All comparisons between source distributions for dominant signatures and significant differences were computed in this manner. The dominant signature in the

Anacostia River was human, contributing 43% of the total fecal loading. Birds, wildlife, and pets were also significant (above minimum detection limits). According to the District's Department of Health, the Anacostia River aquires the largest amount of annual combined sewer overflow of the three waterways, receiving 2.1 billion gallons (9.3 billion liters) per year (D.C. Department of Health, 2002). This helps to explain the dominance of the human signature in the Anacostia as well as why it was much more substantial than the levels detected in Rock Creek and the Potomac. The Anacostia is a tidal embayment of the Potomac, so the large contributions from wildlife and birds also make sense.

Minimum detection for each source were calculated in order to provide lower bounds for what is considered a significant contribution by a source. The lower limits for the Anacostia River library are presented Table 13.

Table 13: Minimum Source Contributions for Significance in the Anacostia

Source	*Minimum Detection Limit (%)*
Bird	1.5
Human	3.2
Pets	2.6
Wildlife	2.6

The percent of sampling events that had source contributions below minimum detection limits are given by sampling station and source (Table 14).

Table 14: Percent of Source Contributions Below Minimum Detection Limit in the Anacostia

	% Below Min. Detection			
Station	*Bird*	*Human*	*Pets*	*Wildlife*
FA01	20.0	0.0	58.3	0.0
FA02	8.3	0.0	41.7	0.0
FA03	0.0	0.0	0.0	8.3
FA04	8.3	0.0	8.3	8.3
FA05	8.3	0.0	33.3	0.0
FA06	8.3	0.0	8.3	16.7

Birds, and especially pets were regarded as undetectable in many of the samples. On average 25% of the samples had undetectable levels of pets, 8.3% of the samples had undetectable levels of birds, and 5.6% of the samples had undetectable levels of wildlife. Human was a significant contributor in all of the sampling events.

It was also necessary to examine the major source contributions at each sampling location as an average over the sampling period to determine if there was a greater impact by one of the major sources in any particular area (Table 16). Brief descriptions of each sampling location are outlined in Table 15.

Table 15: Sampling Site Descriptions for the Anacostia River

Sampling Site	Location	Description
FA01	D.C./Maryland Line	• Sparse brush adjacent to the river • Adjacent to large open field • Popular bird attraction • Some paved running paths along river • Lots of debris and garbage • Easy public access for dog walking
FA02	Benning Road	• Downstream of boat loading dock • Just downstream of Benning Road bridge • Adjacent to open field • Not much wildlife activity • Some trees adjacent to river
FA03	Pennsylvania Avenue	• Site for marinas and boat docks • More wooded adjacent to stream • Begins major boating area • Some birds • Not much wildlife
FA04	11th Street	• Adjacent to Naval District Headquarters • Some birds, little wildlife • Concreted wall on one side
FA05	South Capitol Street	• Adjacent to Anacostia Naval Station • No woods adjacent • Open field on one bank • Little wildlife • Some birds

FA06	Haines Point	• Where Anacostia enters the Potomac, just below FP03 • Adjacent to Anacostia Naval Station • Adjacent to East Potomac Park

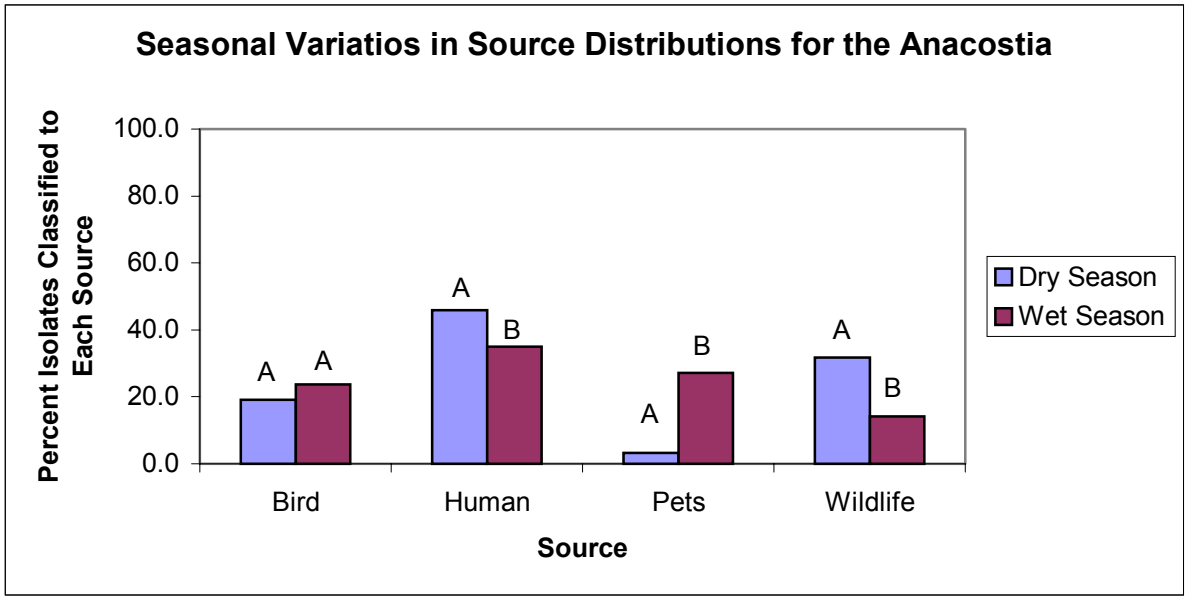
Table 16: Source Distributions at Each Sampling Station on the Anacostia as a Ten-Month Average

Sampling Location	*% Bird*	*% Human*	*% Pets*	*% Wildlife*
FA01	18.1	37.8	10.0	34.1
FA02	15.0	45.0	14.2	25.8
FA03	30.3	37.6	11.8	20.3
FA04	16.7	51.7	10.8	20.8
FA05	16.7	43.3	14.2	25.8
FA06	26.5	44.3	13.0	16.2

Stations FA02, FA04, FA05, and FA06 had dominant signatures of human only. Station FA01 had two dominant signatures of human and wildlife, and station FA03 had two dominant signatures of bird and human. Pets were the least substantial signature on average with less than 15% of the contribution to fecal contamination. The site most impacted by human was FA04, with an average of 51% over the ten-month sampling period. The site most impacted by wildlife was FA01. The most substantial bird signatures were detected at sites FA03 and FA06, where FA06 is adjacent to the East Potomac Park, a nature area that attracts birds. The most substantial pet signatures were detected at sites FA02, FA05, and FA06. These were areas easily accessible to the public and suitable for dog walking due to large open fields (FA02 and FA05) or the East Potomac Park (FA06). The impact from human was significant or dominant at all sites, but was hard to correlate with a particular point source because the location of many of the combined sewer overflows or other human waste discharges was unknown.

The average for the entire sampling area was analyzed for the impact of seasonal variations. The dry season was considered to be the months between July and September, while the wet season occurred from January to March (Appendix, Figure 10). Source contributions in the wet and dry seasons for the Ancaostia River are presented in Figure 6.

Figure 6: Dry and Wet Season Source Contributions on the Anacostia River



*Common letters denote means that were not significantly different (p>0.001)

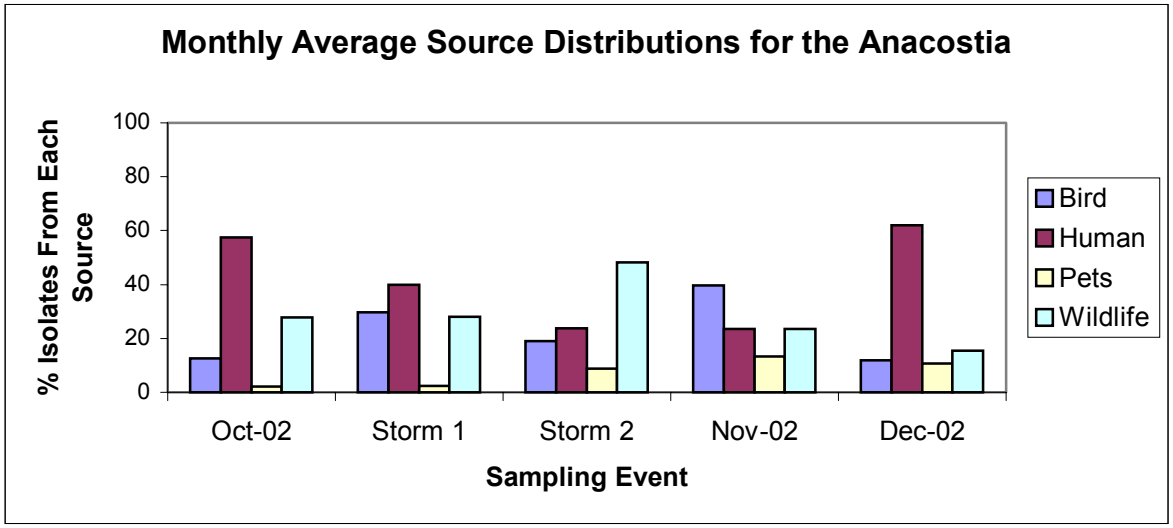
Human and wildlife were the dominant sources in the Anacostia during the dry season, followed by bird. Pet contributions were barely above the minimum detection limit. During the wet season however, the dominant sources were human (but not as substantial) and pets. The seasonal pattern for human was most likely due to the fact that human is a substantial contributor to background fecal pollution. During the wet weather, as in the Potomac, the human signature became diluted and possibly masked by other sources that increased as a result of runoff. The wildlife signature decreased from the dry to wet season, just as it did in the Potomac. This may be due to the fact that wildlife flock to the water during the dry hot months. Wildlife, human, and pet signatures were found to be significantly different between wet and dry season, while bird was not, therefore seasonality

was a controlling factor in all sources except bird. These trends were also observed in the Potomac River.

Source contributions in the wet season and dry seasons for each sampling location are presented in Figures 7-12 in Appendix B. At all stations for both the wet and dry seasons, human was a dominant source, though the percent contribution by human dropped during the wet season. The largest change in human occurred at station FA05, where during the dry season human comprised 64% and dropped to 24% during the wet season. At all stations, the pet signature increased significantly during the wet season. Wildlife decreased at all stations except FA06 (adjacent to East Potomac Park) during the wet season, while the bird signature varied by station and season. These seasonal patterns were also observed in the Potomac, where the wildlife signature decreased from dry to wet season and the pet signature increased, but unlike the Potomac, the human signature remained dominant through the wet and dry season at all stations, indicating the significance of the impact from human on this river.

Storm event samplings took place on October 17 and October 31, 2002. In order to best show the impact of the storm events, source distributions are provided in the month directly prior to the events (October) and the two months following (November and December, Figure 7).

Figure 7: Average Source Contributions on the Anacostia in the Months Surrounding Storm Events



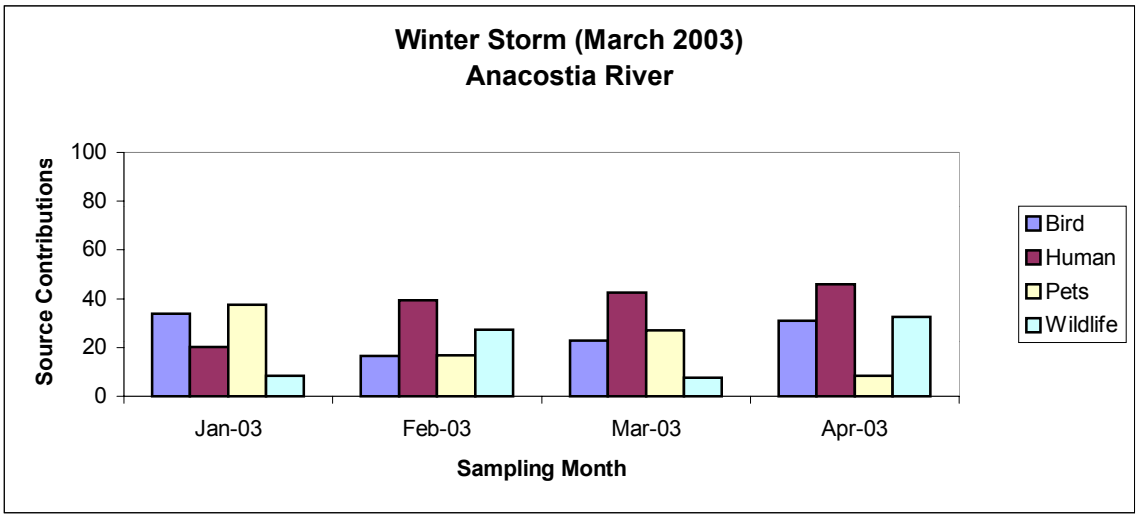
The human signature appeared to decrease as a result of the storm events. Prior to rainfall, the human contribution was approximately 58%, and after the second storm event the signature dropped to approximately 24%. By December, the human signature had returned to the pre-rainfall level. Pets and wildlife signatures increased as a result of the second storm event, while bird contributions did not appear to be correlated with the storm events at all, as seen by the fluctuation of the signature. The impact of rainfall on the human signature was similar to the impact recorded on the Potomac River. The decrease as a result of the storm events was most likely due to the fact that the human was a significant contributor to background levels of fecal contamination. As runoff increased, the human signature was masked. Pet and wildlife increases were most likely a result of increased runoff, though the large increases in contributions from pets and wildlife were delayed until the second storm. An increase in the wildlife signature was also delayed in the Potomac River.

The effects of the storm events at each sampling location showed similar patterns to those of the average effect of the storm events with the exception of the human signature at

stations FA03 and FA06 (Appendix B, Figures 13-18). At station FA06 the first storm event appeared to elevate the human signature, and then the human signature appeared to drop dramatically. At station FA03, the human signature was uncharacteristically low (33.3%) prior to the storm events, but as of December the human signature was as substantial as all the other stations (70.8%) (Appendix B, Figures 13-18).

The sampling month of March was designated as a winter (wet season) storm event in order to compare and contrast the effects of storm events within the different seasons. March was chosen because the precipitation totals prior to the sampling event were 5.8 inches of snow, that quickly melted and 16.7inches of rain (Appendix A, Table 10). In order to best show the impact of the storm event, source distributions are provided in the two months directly prior to the event (the wet season) and the month following (Figure 8).

Figure 8: Average Source Contributions on the Anacostia in the Months Surrounding a Winter Storm Event



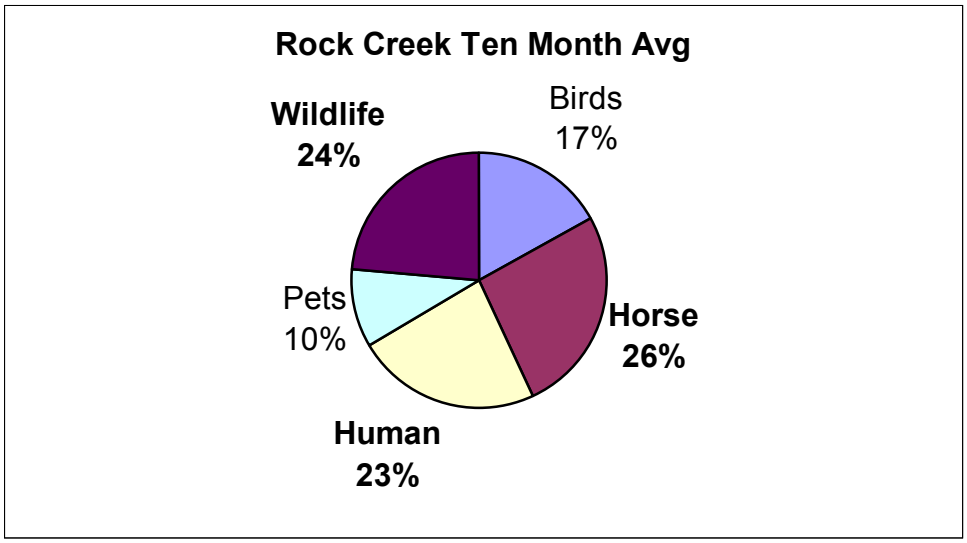
The effect of the designated storm event (March) was not obvious. There appeared to be no relationship between source distributions and this particular storm event. Precipitation fell continually over the sampling periods as either snow or rain. As a result, source distributions were more evenly distributed than were recorded for the storm events in

October. During the first two storm events a substantial decrease in the human signature was noted. During the winter storm event, the human signature did not increase significantly from the levels recorded in February.

Rock Creek

During regular monthly sampling 1,568 isolates were analyzed from Rock Creek. The average monthly source contributions, excluding the two storm events, for Rock Creek were determined first in order to obtain an accurate picture of what to expect as the major sources (Figure 9). Dominant signatures are highlighted in bold.

Figure 9: Average of Source Contributions from all stations on the Rock Creek



In order to determine the dominant signatures and relate the significant difference between the mean values of each source, a pairwise comparison (analysis of variance via the t-test) was computed for each combination of pairs. Significant differences were detected if $p < 0.001$ (Whitlock et al. 2002). A more common p value of 0.05 is typically used for this type of test for significance (Graves, 2003), however, a more stringent p value was used in this study in order to provide a greater level of confidence in the significant differences detected. All subsequent comparisons between source distributions for dominant signatures

and significant differences were computed in this manner. The dominant sources over the ten-month sampling period for Rock Creek were wildlife, horse, and human. Sources appeared more evenly distributed in Rock Creek than in the Anacostia and Potomac Rivers. The dominant horse signature was attributed to the presence of the Rock Creek Park Police Stables, the numerous horseback-riding trails present within park boundaries, and a private horse horse stable in Maryland just upstream of the D.C. border. Much of the area immediately adjacent to Rock Creek is park or open land that attracts wildlife. The dominance of the human signature was most likely a result of the 29 combined sewer outfalls that discharge 49 million gallons (215 million liters) of combined stormwater/sewage per year.

Minimum detection limits for each source were calculated in order to provide lower bounds for what is considered a significant contribution by a source. The lower limits for the Rock Creek library are presented Table 17.

Table 17: Minimum Source Contributions for Significance in Rock Creek

Source	*Minimum Detection Limit (%)*
Bird	1.6
Horse	2.1
Human	2.2
Pets	2.0
Wildlife	1.9

The percent of sampling events that had source contributions below minimum detection limits are presented in Table 18, given by sampling station and source.

Table 18: Percent of Source Contributions Below Minimum Detection Limit in Rock Creek

	% Below Min. Detection				
Station	*Bird*	*Horse*	*Human*	*Pets*	*Wildlife*
FR01	25.0	8.3	25.0	33.3	8.3
FR02	16.7	25.0	0.0	50.0	16.7
FR04	33.3	16.7	0.0	25.0	0.0
FR05	0.0	0.0	0.0	33.3	8.3
FR06	0.0	8.3	0.0	50.0	0.0
FR07	0.0	16.7	8.3	50.0	0.0

On average, pets fell below the minimum detection limit in 40.3% of the samples, bird and horse fell below minimum detection limits 12.5% of the samples, and human and wildlife fell below minimum detection limits in 5.6% of the samples.

It was also important to look for major source contributions at each sampling location as an average over the sampling period to determine if there was a greater impact by one of the major sources in any particular area (Table 20). Brief descriptions for each site are presented in Table 19.

Table 19: Sampling Site Descriptions for Rock Creek

Station	Location	Description
FR01	DC/Maryland Line	• Heavily wooded • Pedestrian bridge just upstream • Easy access for public use • Popular place for walking pets • Horse trails adjacent to stream • Private horse stable just upstream in Maryland
FR02	Tilden Street	• Location of an old abandoned mill • Wooded area, adjacent to stream, but open over the water surface • Easy public access • Popular place to walk pets • Parking area adjacent • Small dam, providing large pond area for bird activity

FR04	Porter Street	• Sampling location is under a pedestrian footbridge • Not wooded • Not as accessible, no nearby parking • Not as much pet walking • Very accessible to birds
FR05	Below the National Zoo	• Stream runs over an old submerged roadway • Heavily wooded • Horse trails located adjacent to stream • Significant signs of bird, and wildlife • Stormwater runoff pipes from adjacent parking lots
FR06	Massachusetts Avenue	• Adjacent to Rock Creek Parkway • Approximately one third of a mile downstream of Rock Creek Park Police Stables • Popular hiking area • Some horse trail access • Heavily wooded for wildlife • Popular area for birds • Many unknown pipes entering stream
FR07	Pennsylvania Avenue	• Sampling site under a bridge • No nearby parking, not easily accessible to public • Plenty of wildlife in wooded areas • Bird heavily nesting in rafters of bridge directly above sampling site • Old storm sewer drains

Table 20: Source Distributions at Each Sampling Station on Rock Creek as a Ten-Month Average

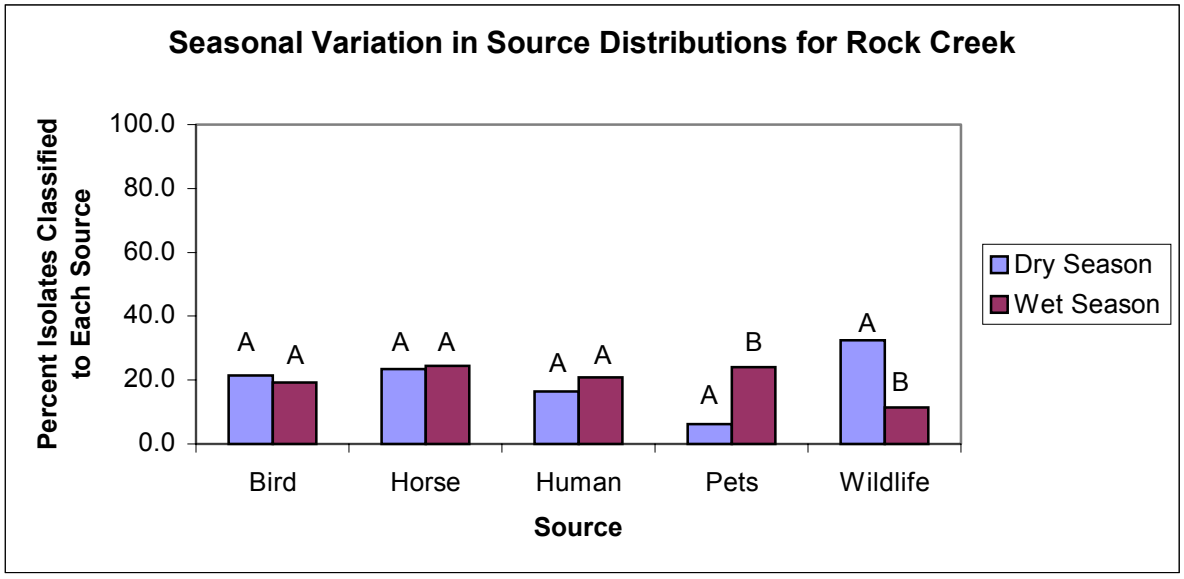
Sampling Location	*% Bird*	*% Horse*	*% Human*	*% Pets*	*% Wildlife*
FR01	13.7	28.8	22.0	14.0	21.5
FR02	15.8	27.1	26.2	8.8	22.1
FR04	12.5	24.0	24.6	13.9	25.0
FR05	13.7	33.0	16.8	7.9	28.6
FR06	23.3	25.5	20.5	7.1	23.6
FR07	19.3	24.6	20.0	10.8	25.3

Horse was the dominant signature at station FR01. Stations FR02, FR04, and FR07 had major source contributions of horse, human, and wildlife. Station FR05 was impacted less by human, as the dominant sources were only horse and wildlife. Station FR06 had dominant sources of bird, horse, human, and wildlife. Pets were the least significant source

contributions ranging from 8% to 14% between the six sampling stations. Birds had the largest impact on station FR06, horse had the largest impact on station FR05, human had the largest impact on stations FR02 and FR04, pets had the largest impact on stations FR01, FR04, and FR07, and wildlife had the largest impact on station FR05. All sites have dominant horse signatures, and all but FR01 have dominant wildlife signatures. The land use surrounding the creek had much to do with this distribution of source impacts. Rock Creek is a safe haven for wildlife and is a center for nature activities such as horseback riding, hiking, and recreation. The impact from human was still significant at most sites, but was hard to correlate with a particular point source because the location of many of the combined sewer overflows or other human waste discharges is unknown.

The average for the entire sampling area was analyzed for the impact of seasonal variations. The dry season was considered to be the months between July and September, while the wet season occurred from January to March. Source contributions in the wet and dry season as for the Ancaostia River are presented in Figure 10.

Figure 10: Dry and Wet Season Source Contributions on Rock Creek



*Common letters denote means that are not significantly different (p>0.001)

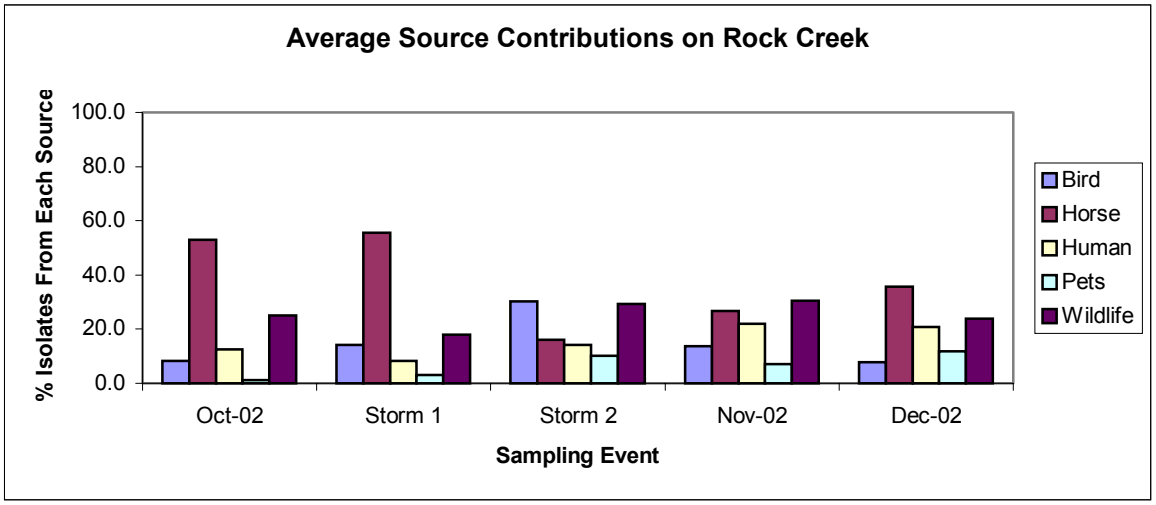
The dominant source during the dry season was wildlife, followed by bird and horse. During the wet season, major source contributions changed to horse, human, and pets. The most substantial and only statistically significant changes were noted in wildlife that decreased and pets that increased from the dry to wet season. During the dry season, wildlife had the most substantial impact because they moved to the creek. The increase in the pet signature during the wet season was most likely due to the greater potential for runoff. No statistical change was noted for human, bird, and horse from dry to wet season. Like the other two waterways, the bird signature did not show a correlation to seasonality. The horse and human signatures also did not show any correlation to seasonality. The behavior of the human signature may likely be due to the fact that Rock Creek has a much lower dilution potential than the Potomac, so human contributed to background fecal contamination in dry weather, but was not as easily masked when a high volume of human was input from the combined sewers during wet weather.

Source contributions in the wet season and dry seasons for each sampling location are presented in Figures 19-24 of Appendix B. The dominant sources for wet and dry seasons varied by station. Wildlife was the only dominant source during the dry season at stations FR01, FR02, and FR04. Wildlife was also one of the dominant sources at all other stations except FR07 during the dry season. Horse was the dominant source during the dry season at station FR07, and one of the dominant sources at stations FR05 and FR06. Station FR06 had a diversity of dominant sources including bird, horse, human, and wildlife during the dry season. During the wet season, pets were the dominant signature at station FR01, and one of the dominant signatures at stations FR02 and FR04. Bird was a dominant source at stations FR02 and FR06 and FR07, human was a dominant source at stations FR06 and

FR07, and horse was a dominant signature at stations FR04 and FR05 during the wet season. At all stations, pets increased from the dry to wet season. At all stations except FR07, wildlife decreased substantially from the dry to wet season. The human signature increased from the dry to wet season at stations FR01, FR02, FR06 and FR07. Bird and horse did not show consistent trends to relate the signatures to season. Comparing these trends to the site descriptions for each station in Table 19, show that most of the source distributions and site to site and seasonal variations make practical sense.

The storm event samplings took place on October 17 and October 31, 2002. In order to best show the impact of the storm events, source distributions are provided in the month directly prior to the events (October) and the two months following (November and December, Figure 11).

Figure 11: Average Source Contributions on Rock Creek in the Months Surrounding Storm Events

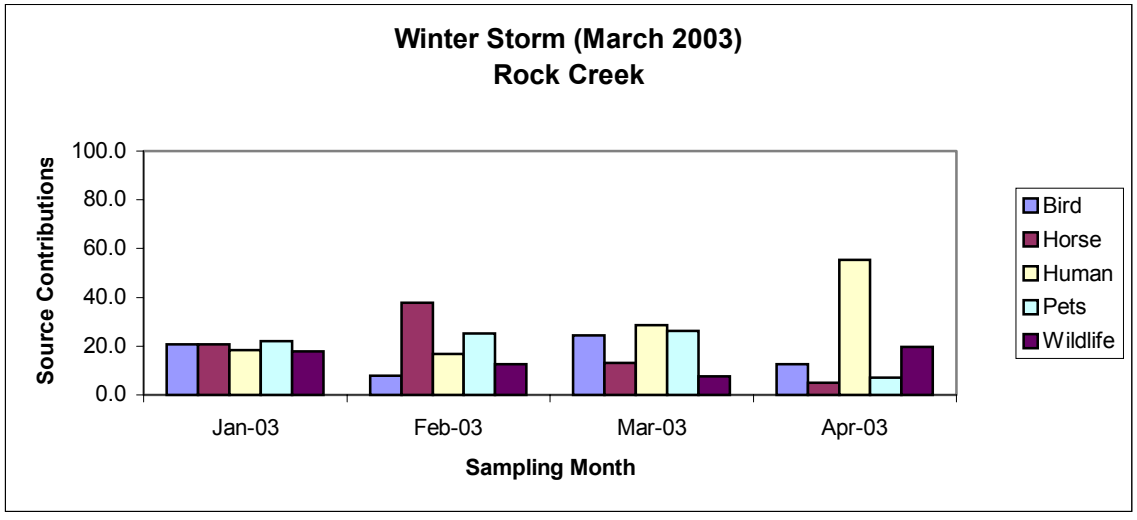


The bird and pet signatures increased as a result of the successive storm events, while bird contributions fell to pre rainfall levels, pets continued to rise through November. The most substantial change in source contributions appeared with horse after the second storm event, dropping from 56% to 16%. The human signature fell slightly after the first storm event, but

then appeared to increase slowly through the second storm event and the months immediately following. The wildlife signature also dropped as a result of the first storm event, but increased after the second storm event and through November. The effects of the storm events were similar at all the sampling locations, and they followed the general trend that was shown by the average for the sampling area (Appendix B, Figures 25-30).

The sampling month of March was designated as a winter (wet season) storm event in order to compare and contrast the effects of storm events within the different seasons. March was chosen because the precipitation totals in that month prior to the sampling event were 5.8 inches of snow, that quickly melted and 16.7inches of rain (Appendix A, Figure 10). In order to best show the impact of the storm event, source distributions are provided in the two months directly prior to the event (the wet season) and the month following (Figure 12).

Figure 12: Average Source Contributions on Rock Creek in the Months Surrounding a Winter Storm Event



Human increased with each sampling event from January to April. A similar pattern was noted in the October storm events, but the human was a much more substantial portion

of the source distributions during the wet season. Unlike the earlier storm events, the horse signature decreased during the winter storm sampling event, but spiked in the month of February immediately preceding the storm event. A similar pattern was noted in the livestock signature of the winter storm sampling event on the Potomac River. The large snowfall event probably played a role in the runoff levels of these agricultural signatures. The wildlife signature decreased consistently from January to March, when a bulk of the precipitation fell, indicating the wildlife were not consistently inhabiting areas close to the creek.

SUMMARY

Violations of the water quality standards provided key information as to which stations were potentially impacted the most by fecal pollution. As expected, there was a negative correlation between the flow and both fecal coliform and enterococci densities, indicating the impact that dilution had on the presence and detection of fecal bacteria. Of important consequence was the fact that the old standard (1000 fecal coliforms/100mL) was not violated nearly to the extent as the new standard (35 enterococci/100mL). The District of Columbia chose not to adopt the new standard. Based on the number of violations that occurred with each standard, the old standard was much less stringent. Consequently, the District of Columbia will report fewer bacterial impairments and remediation plans will require less of a reduction in total and source specific fecal loading. Though this seems to be in Washington D.C.'s favor not to switch standards, according to Jutta Schneider, the EPA is now beginning to recommend and will eventually mandate a more stringent standard for fecal coliforms. This change will force states or territories that chose not to adapt the new recommendations to monitor for pathogenic impairments with a comparable standard.

Washington D.C. will consequently be faced with the listing of many more pathogenic impairments and TMDL's in the near future.

The Potomac River, as it was the largest waterway in this study had the highest dilution rates. Fecal coliform standards were not violated at any of the sampling stations on the Potomac River during the course of the study, while enterococci standards were violated regularly, but not nearly to the magnitude as was in the Anacostia and Rock Creek. The largest impact of fecal loading on the Potomac appeared to occur at site FP03, the furthest downstream. The Anacostia had an intermediate flow, with the potential for some dilution, and as expected, the number of violations fell between those of the Potomac and Rock Creek. The station receiving the greatest impact, or the largest amount, from fecal pollution appeared to be station FA01, with a lesser impact at each successive sampling station downstream. Rock Creek had the smallest flow rate, and the poorest potential for dilution. This was reflected in the very high percentage of samples that violated water quality standards for both enterococci and fecal coliforms. The station impacted by fecal loading the most appeared to be FR07, the furthest site downstream. The impact by fecal coliforms decreased at each successive station upstream.

The enumeration of bacteria provides important information about the overall microbial quality of each waterway, but the ARA results provided important information regarding the actual sources of fecal contamination and the relative distributions. In the Potomac, three dominant sources were detected over the course of the study (human, livestock, and wildlife). The human signature was of greatest concern and was found to have the largest impact at stations FP02 and FP03, while the livestock signature had the greatest impact FP01, the station most upstream. The dominant signature detected on the

Anacostia belonged to human comprising over 40% of the total isolates sampled. These results were easily interpreted and understood because the largest annual discharge of combined sewer overflow occurs on the Anacostia River (Vasquez, 2002). The site most impacted by human fecal pollution, of the sampling stations on the Anacostia River, was FA04, with a contribution of 51% of the total fecal loading. Rock Creek had three dominant sources (horse, human, and wildlife) and the source contributions appeared more uniformly distributed than in the Potomac or Anacostia. Human had the largest impact on stations FR02 and FR04, while horse and wildlife most impacted station FR05 near the zoo, adjacent to a heavily wooded area with public riding trails.

Seasonal variations were as important for determining the impacts from certain sources as were overall averages. During the dry season (July through September), dominant sources on the Potomac and Anacostia were wildlife and human and the dominant source on Rock Creek was wildlife. In all three waterways, wildlife appeared most pronounced during the dry months. Human appeared to be a substantial contributor to background fecal contamination in the Potomac and Anacostia as evidence by the high levels detected during the dry season and the substantial decrease in the signature in the wet season, as it became masked by other sources affected more by runoff. The significant human signature in the dry season also indicated that there might be a substantial infrastructure failure in the sewer system, causing continual leakage into all three waterways, but was most pronounced in the Anacostia and Potomac Rivers, which have the greatest potential for dilution during the wet season. During the wet season (January through March), livestock was the dominant signature on the Potomac, human and pets were dominant in the Anacostia, and horse, human, and pets were dominant on Rock Creek. The

signatures changed dramatically from the dry to wet seasons for most sources, indicating the impact of seasonal changes on source distributions. Bird was the only signature not affected by seasonality in any of the waterways. Human and horse were unaffected by seasonality in Rock Creek. The behavior of the human signature in Rock Creek is most likely due to the low dilution potential, where the amount of fecal contamination from runoff is not enough to mask the human signature during wet weather as most likely occurred in the Anacostia and Potomac. Seasonal variation of source distributions occurred because of changes in rainfall and runoff events that greatly impact the influence of sources not deposited directly into the waterways. Source distributions also varied by season as sources such as wildlife, and livestock changed the amount of time spent in or near the waterways.

Storm events provided a finer snapshot of the short-term effects of precipitation on source distributions. In the Potomac the most dramatic changes in source distributions occurred with human as it dropped substantially and with livestock as it increased substantially as a result of the storm events. In the Anacostia, a dramatic decrease also occurred in the human signature as a result of the storm events. In Rock Creek, the pet signature appeared to increase as a result of the storm events, while human and wildlife decreased during the first storm, but then increased. The affect the storm events had on Rock Creek was much more variable and less predictable than on the Potomac and Anacostia. Though only two official storm events were sampled, these were only representative of conditions immediately following the dry season. The effects of storm events in the wet season were drastically different, indicating that there was no consistent predictive capability for source distributions following a major precipitation event. Seasonality appeared to provide more meaningful and definitive trends that were consistent

between the waterways and even predictable. This is perhaps a better way to analyze the overall effects of precipitation than is the sampling of individual storm events.

The results of this project support findings of a study conducted by Whitlock et al. (2002), where it was found that in a highly urbanized environment, human was a major contributor to low-level background fecal contamination. The findings suggest that human fecal pollution may become masked by other sources, especially when enterococci and fecal coliform densities are high such as occurs in major precipitation events. This study also corroborated findings by Whitlock et al. (2003) that there was greater temporal variability in the sources of fecal pollution in an urban watershed, which may not necessarily be as significant in rural watersheds. For example, Booth et al. (2003) conducted a study on the Blackwater River in a rural watershed. In that study the dominant signature was cattle. As sample months were divided into warm and cool seasons, it was found that livestock dominated nearly an equal proportion of the samples in each season, while wildlife was more prevalent in the warm season samples and human was more prevalent in the cool season. There was a trend for some seasonal effects, but the dominant signature (livestock) was prevalent throughout the entire sampling period. The results of this study indicate that ARA was a viable BST technique when applied to urban watersheds based on the detection of a substantial human signature, a masking of the human signature during wet weather, and a significant variation of source distributions by season.

REFERENCES

American Public Health Association, American Water Works Association, and Water Environment Federation. 1998. M. A. Franson, Editor. Standard Methods for the Examination of Water and Wastewater, 20th Edition. American Public Health Association, Washington D.C. pp.9-74-9-78.

Booth, A.M., C. Hagedorn, A.K. Graves, S.C. Hagedorn, K.H. Mentz. 2003. Sources of Fecal Pollution in Virginia's Blackwater River. Journal of Environmental Engineering. 129 (6): 547-552.

Department of Health, Environmental Health Administration, Bureau of Environmental Quality Water Quality Division, Water Quality Control Branch. 2002. Total Maximum Daily Loads, District of Columbia, Fecal Coliforms. Unpublished Report.

Environmental Protection Agency. 2003. Bacterial Water Quality Standards for Recreational Waters. http://www.epa.gov/waterscience/beaches/local/statreptac.pdf [Online] Last Accessed November 10, 2003.

Graves, Alexandria. 2003. Identifying Sources of Fecal Pollution in Water as a Function of Sampling Frequency Under Low and High Stream Flow Conditions. Dissertation for Doctorate of Philosophy, Virginia Tech.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneau Jr. 1999. Determining Sources of Fecal Pollution in a Rural Virginia Watershed with Antibiotic Resistance Analysis Patterns in Fecal Streptococci. Applied and Environmental Microbiology. 65 (12): 5522-5531.

Harwood, V.J., J. Whitlock and B. Withington. 2000. Classification of Antibiotic Resistance Patterns of Indicator Bacteria by Discriminate Analysis: Use in Predicting the Source of Fecal Contamination in Subtropical Waters. Applied and Environmental Microbiology. 66 (9): 3698-3704.

National Weather Service Forecast Office. 2003. Baltimore/Washington. http://www.erh.noaa.gov/er/lwx/climate.htm [Online] Last November 10, 2003.

Parveen, S., K.M. Portier, K. Robinson, L. Edmiston, and M.L. Tamplin. 1999. Discriminate Analysis of Ribotype Profiles of *Escherichia coli* for Differentiating Human and Nonhuman Sources of Fecal Pollution. Applied and Environmental Microbiology. 65 (7): 3142-3147.

Vasquez, M. August 15, 2002. A Problem Between Rain and The Rivers, City Seeks U.S. Funds To Cut Waste Spillage. *Washington Post*. Section J, p T3.

Virginia Department of Environmental Quality. 2003. Memorandum Commonwealth of Virginia Department of Environmental Quality, Water Division. http://www.deq.state.va.us/waterguidance/pdf/032007.pdf. [Online] Last Accessed November 10, 2003.

Whitlock, J.E., D.T. Jones, V.J. Harwood. 2002. Identification of the Sources of Fecal Coliforms in and Urban Watershed Using Antibiotic Resistance Analysis. Water Research. 36: 4273-4282.

Wiggins, B.A. 1996. Discriminate Analysis of Antibiotic Resistance Patterns in Fecal Streptococci, a Method to Differentiate Human and Animal Sources of Fecal Pollution in Natural Waters. Applied and Environmental Microbiology. 62 (11): 3997-4002.

CHAPTER V
EXPANDING THE LIMITS OF ARA

INTRODUCTION

Expanding the limits of previous ARA studies is an integral part of progress being made to enhance not only the science of ARA, but the science of BST and its application to TMDL's, remediation efforts, and human health risk assessments. Two important pieces of information not addressed in the literature that would provide some confidence and direction to the science of ARA are the creation of an unknown source category to hold all isolates not classified to a source category by a specified acceptable margin, and a test to see if accepted statistical procedures such as discriminate analysis (DA) provide the most appropriate statistical model for ARA. Another important topic that is addressed in the literature and is the center of some debate is the elimination of duplicate isolate patterns from the library. Work needs to be done in all of these areas in order to fully evaluate and expand the understanding of not only ARA, but all library-based methods.

The currently used statistical models (DA and logistic regression [LR]) force isolates to be classified into one of the given source categories no matter how well the isolate actually matches any of the known profiles of that source. This can give a false impression of the source contributions if a majority of the isolates do not match the model assigned source category at a high percentage similarity. Some molecular methods have taken this into consideration by creating an acceptable lower limit of similarity (or matching) that an isolate must numerically rank above in order to be classified as a particular source with confidence. A PFGE study conducted by Kariuki et al. (1998) found that with cluster analysis, isolates that produced patterns showing greater than 60% similarity were closely related. In another PFGE study, Simmons et al. (2002) used a cutoff of 80% similarity and

found that approximately 46% of their isolates could not be successfully classified (46% of the isolates fell below and 80% probability of a match). The 80% similarity and 50% similarity cutoffs were used for the unknown isolates in this study. It was the expectation that nearly all of the unknown isolates will be classified in the correct category at the 50% probability level in order to have confidence that the library is representative of the isolates found in the environment, and that approximately half of the isolates will be classified with patterns matching the known source library with at least an 80% similarity. Applying the same stringent standards that some molecular methods have insisted on is one way to assert that ARA is producing results that are as consistent and as confident as results produced by molecular methods. Also, if source distributions are not significantly different with the remaining isolates that met the >80% matching probability vs no minimum matching standard, then it may signify that the library does an adequate job identifying the sources regardless of the level at which the isolates match the intended source.

Another important issue that needs to be addressed is appropriate statistical modeling. The statistical analysis model now widely used and accepted among researchers of ARA is discriminate analysis. It was first described as a tool for ARA by Wiggins (1996). The statistical tool was appealing for ARA because of the fact that it created a classification rule for all of the isolates, and based on that rule, classifies each individual isolate into one of many possible sources. Some important characteristics of DA are its ability to describe maximum differences among prespecified groups of sampling units based on a variety of discriminating characteristics, and its ability to predict the group membership of future samples from unknown groups based on that same suite of classification characteristics (McGarigal et al., 2000). An underlying assumption, however, is that it

operates on data sets for which prespecified, well-defined groups exist, or has group dispersions which are homogenous (McGarigal et al., 2000). Antibiotic resistance characteristics of fecal bacteria are the sample units from which the data sets are comprised. The assumptions for DA may be violated when used to model ARA because antibiotic patterns of bacterial groups from specific sources are neither well-defined nor homogenous. There is an alternative to DA that has not been applied to ARA, but was explored as an appropriate model for this type of application, called logistic regression (LR). LR is a statistical tool that can also describe the relationship between and outcome variable and a set of independent variables. The underlying assumptions of LR are not the same as the assumptions for DA, and thus could potentially model antibiotic resistance patterns of bacteria in a more meaningful and justified way. Important assumptions of the logistic regression model are a non-linear relationship between the independent and dependent variables, no requirement for normality of the variables, and no assumption of equal variances. It is also a flexible and easily used model (Hosmer et al., 2000). In fact, it is a popular modeling procedure to use when the dependent variable or variables are dichotomous (binary) (Kleinbaum, 1994). These assumptions fit in very nicely with the analysis of microbial populations such as the reaction (growth or no growth [1or 0]) obtained with antibiotic resistance profiling where the population is not necessarily uniform and reactions are not necessarily normally distributed.

All library-based methods are faced with questions of representativeness. Researchers of BST have often argued about what constitutes an appropriate test of representativeness. The DA or LR statistical models already do a test of sorts for representativeness as the models go through a process of resubstituting each isolate back

into the library in order to compare it to all known source patterns in the library, including its own, and classify it to its most likely source. The results of the resubstitution analysis are the RCC's. Wiggins et al. (2003) believe that the ARCC, when computed through resubstitution analysis, may actually overestimate the library representativeness. The suggestion is to use cross validation analysis (jackknife). For cross validation analysis all isolates that came from the same sample are pulled from the library and RCC's are computed on a sample basis and not an isolate basis. When removing all isolates from a sample, the probability that the sample being compared to the rest of the library will have a matching or similar isolate patterns becomes much smaller. To go even further, Wiggins suggests that all duplicate patterns should be eliminated from the library, leaving only unique patterns against which to compare unknown isolates.

RESU LTS AND DISCUSSION

Addition of Unknown Source Category

The statistical model, LR, outputs the probability of how well each isolate matches any of the known isolate patterns for each source. The isolates are then classified as the source that had the highest similarity probability. An unknown source category was created to hold isolates that did not match any source category by at least 50% and 80%.

The addition of an unknown source category to hold isolates that do not match any isolate profiles in the known source library by at 50% and 80% showed a significant loss in the isolates that could subsequently be used to determine source distributions. In the Potomac River, approximately 12% of the isolates were unrecognizable based on a similarity of less than 50% to any of the known isolate profiles. With a lower limit of 80% matching probability, that number increased to 44% categorized as unknowns (Table 1). In

the Anacostia River, fewer isolates were lost at the 50% similarity cutoff than in the Potomac (5.7%), but approximately the same percentage of isolates were lost with an 80% cutoff (42.7%) (Table 2). In Rock Creek approximately 6.5% of the isolates were unrecognizable at the 50% similarity level. With an 80% similarity criterion, 41% of the isolates were classified as unknown (Table 3). These results are not unlike results obtained with several molecular methods. Simmons et al. (2002) and Samadpour et al. (1995) performed this type of analysis with PFGE and ribotyping, respectively, and had to classify approximately half of the isolates as unknowns based on an 80% matching probability.

Table 1: Source Distributions (%) with Unknown Categories for the Potomac River

Source	*Any % Similarity*	*> 50% Similarity*	*> 80% Similarity*
Bird	5.7	5.6	4.7
Human	28.8	11.3	24.0
Livestock	30.6	32.5	14.2
Pets	12.8	12.5	4.2
Wildlife	22.1	26.1	8.7
Unknown	0.0	12.0	44.2

Table 2: Source Distributions (%) with Unknown Categories for the Anacostia River

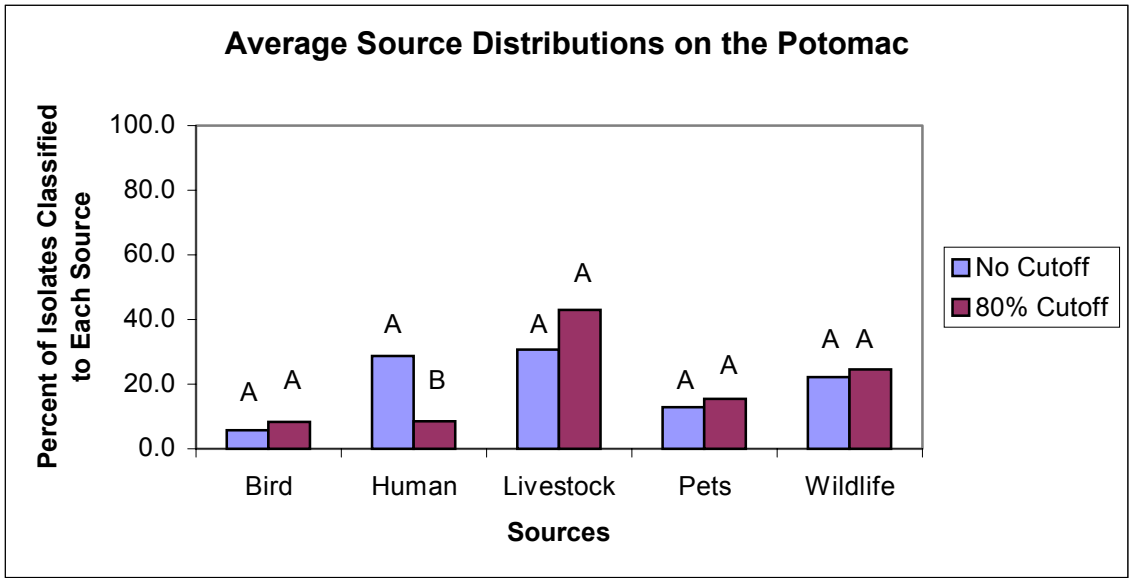
Source	*Any % Similarity*	*> 50% Similarity*	*> 80% Similarity*
Bird	20.6	20.6	15.3
Human	43.3	38.5	19.3
Pets	12.3	10.6	8.4
Wildlife	23.8	24.6	14.3
Unknown	0.0	5.7	42.7

Table 3: Source Distributions (%) with Unknown Categories for Rock Creek

Source	*Any % Similarity*	*> 50% Similarity*	*> 80% Similarity*
Bird	16.4	15.6	11.6
Horse	27.2	25.9	15.1
Human	21.6	19.5	12.3
Pets	10.4	10.4	8.7
Wildlife	24.4	22.1	11.2
Unknown	0.0	6.5	41.1

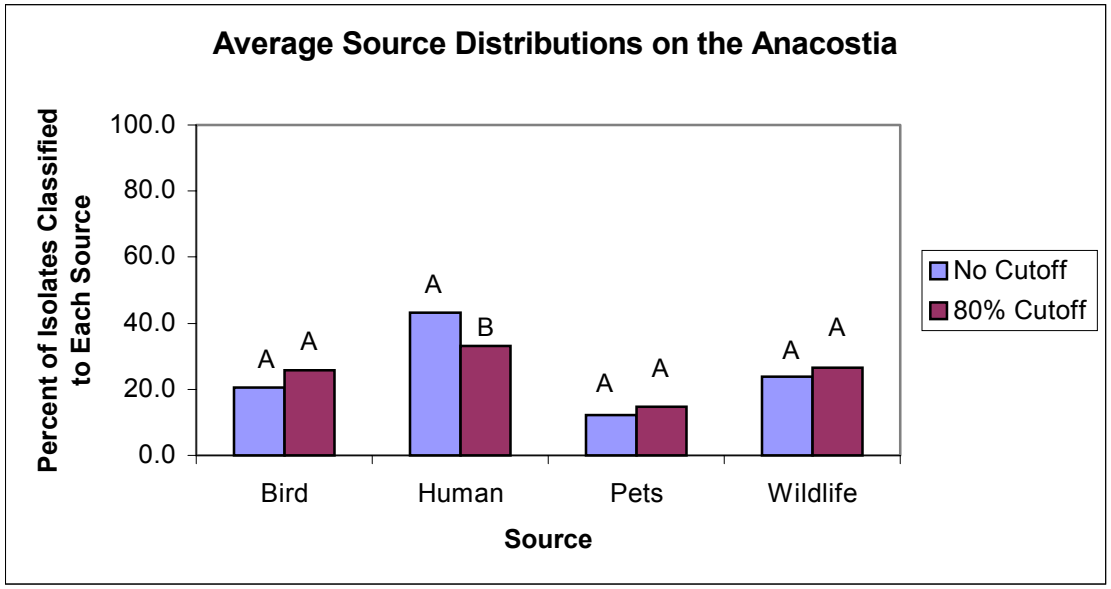
A comparison of the source distributions with no unknown source category and source distributions with only the isolates that match the model-classified source by at least an 80% similarity are presented in Figures 1-3.

Figure 1: Source Distributions on the Potomac Before and After Exclusion of Isolates with < 80% Similarity to Known Source Profiles



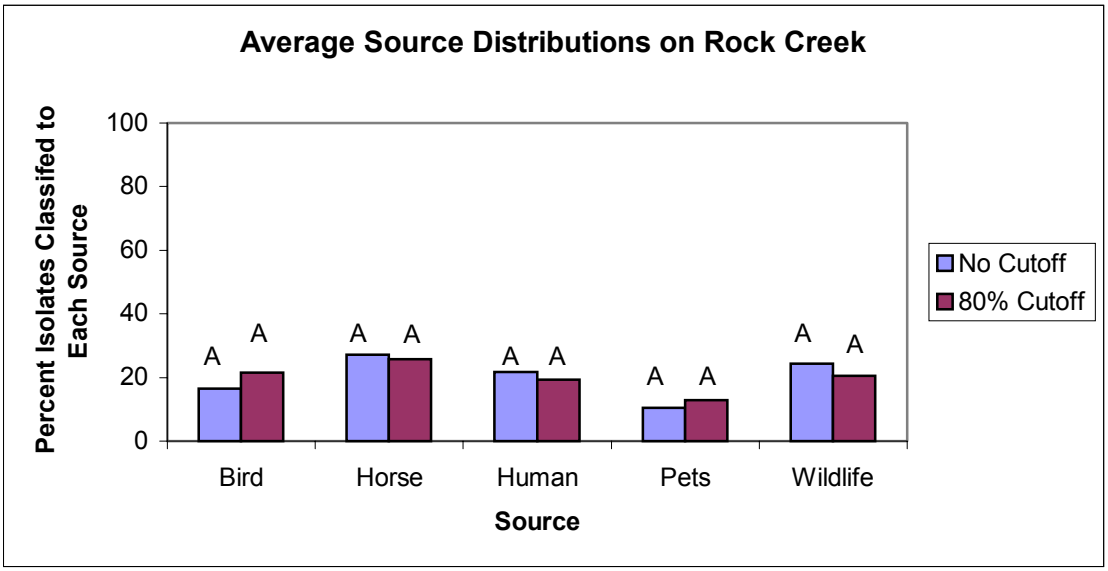
*Common letters denote means that are not significantly different (p>0.001)

Figure 2: Source Distributions on the Anacostia Before and After Exclusion of Isolates with < 80% Similarity to Known Source Profiles



*Common letters denote means that are not significantly different (p>0.001)

Figure 3: Source Distributions on Rock Creek Before and After Exclusion of Isolates with < 80% Similarity to Known Source Profiles



*Common letters denote means that are not significantly different (p>0.001)

Exclusion of the isolates that did not match with at least 80% probability did not significantly affect source distributions of the unknown isolates for any of the sources in Rock Creek, and only affected human in the Anacostia and Potomac. In both cases where the human signature was significantly different, the signature decreased as the isolates were required to match known isolate profiles at the 80% or greater level. This signifies that for the Anacostia and Potomac libraries, the majority of the human isolates were lost (classified as unknowns) at a greater rate than the other sources meaning the human isolates that comprise the library may not be adequately representative of the human isolates found in the environment. This finding was not suprising considering there are 2.5 million people that inhabit the metropolitan region. Creating the 80% cutoff and categorizing the isolates that do not meet that standard as unknowns had no significant affect on source distributions (except for human), indicating that it may not be necessary to perform this analysis, and still have confidence in the results obtained with no cutoff.

Appropriate Statistical Modeling

Discriminate analysis as described by Wiggins (1996) has been adopted as the main statistical analysis tool for ARA, and other BST methods as well. In this study, logistic regression was proposed as possibly a more realistic model for the antibiotic resistance profiles because the assumptions for the data do not include a homogenous or normal data set. The RCC's were compared between the DA and LR models for each library to determine if there were substantial differences, and if the differences were substantial, it would provide evidence to suggest there was a need for a more appropriate statistical analysis tool (LR) for ARA. The RCC's computed from the LR model were higher than the RCC's computed from the DA model for each source in each library (Tables 4-6). A match comparisons test was performed in order to calculate the standard error between the RCC's of each method for each library. The standard error for the Potomac library was 1.28, the standard error for the Anacostia library was 1.02, and the standard error for the Rock Creek library was 0.72. Though the RCC's were significantly different (having greater than 1 standard error between) in the Anacostia and Potomac libraries, the averages were very close to the one standard error cutoff for significant difference and it cannot be reasonably suggested at this time, to choose either DA or LR as both provided similar results.

Table 4: Potomac Library Comparison of RCC's from DA and LR Models

Predicted Source	*Percent of known source isolates assigned to each source by DA (Percent of known source isolates assigned to each source by LR)*				
	Bird	Human	Livestock	Pets	Wildlife
Bird	**94.0 (95.6)**	3.3 (1.2)	1.1 (1.1)	5.4 (2.4)	5.8 (3.5)
Human	1.2 (1.2)	**77.9 (86.3)**	2.6 (2.6)	3.0 (1.8)	6.5 (5.0)
Livestock	2.4 (1.6)	8.4 (4.2)	**88.8 (92.4)**	0.6 (2.4)	10.7 (11.1)
Pets	2.4 (0.4)	5.8 (3.0)	2.2 (0.6)	**90.5 (91.7)**	0.8 (0.8)
Wildlife	0 (1.2)	4.9 (5.4)	5.3 (2.3)	0.6 (1.8)	**76.3 (79.7)**

* DA ARCC 85.5%
*LR ARCC 89.1%

Table 5: Anacostia Library Comparison of RCC's from DA and LR Models

Predicted Source	*Percent of known source isolates assigned to each source by DA (Percent of known source isolates assigned to each source by LR)*			
	Bird	Human	Pets	Wildlife
Bird	**94.0 (96.0)**	3.0 (0.47)	4.8 (2.4)	7.3 (2.7)
Human	2.8 (2.0)	**85.8 (91.6)**	4.2 (3.0)	6.5 (4.6)
Pets	2.0 (0.8)	6.7 (3.7)	**91.1 (94.1)**	1.9 (2.3)
Wildlife	1.2 (1.2)	4.4 (4.2)	0.0 (0.6)	**84.3 (90.4)**

* DA ARCC 88.8%
*LR ARCC 93.0%

Table 6: Rock Creek Library Comparison of RCC's from DA and LR Models

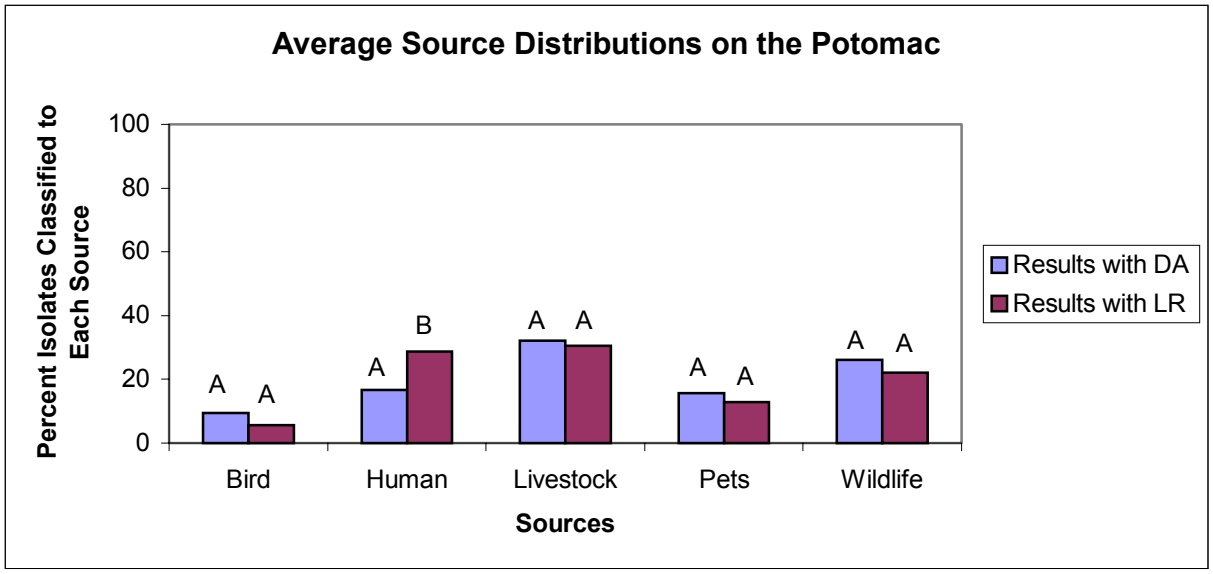
Predicted Source	*Percent of known source isolates assigned to each source by DA (Percent of known source isolates assigned to each source by LR)*				
	Bird	Horse	Human	Pets	Wildlife
Bird	**91.5 (95.6)**	3.0 (1.3)	3.0 (0.7)	5.4 (2.4)	6.9 (7)
Horse	4.4 (2.0)	**94.5 (97.5)**	3.7 (1.9)	1.8 (1.8)	6.1 (3.8)
Human	1.6 (0.8)	0.0 (0.4)	**82.6 (90.0)**	3.6 (3.0)	5.0 (4.6)
Pets	2.4 (0.8)	0.0 (0.0)	5.8 (3.7)	**89.3 (94.1)**	0.4 (2.3)
Wildlife	0.0 (0.8)	2.6 (0.9)	4.9 (3.7)	0.0 (0.6)	**81.6 (86.6)**

* DA ARCC 87.9%
*LR ARCC 92.8%

No significant differences were detected between the ARCC's of each model for each library. The two statistical methods were analyzed in more detail with a comparison of

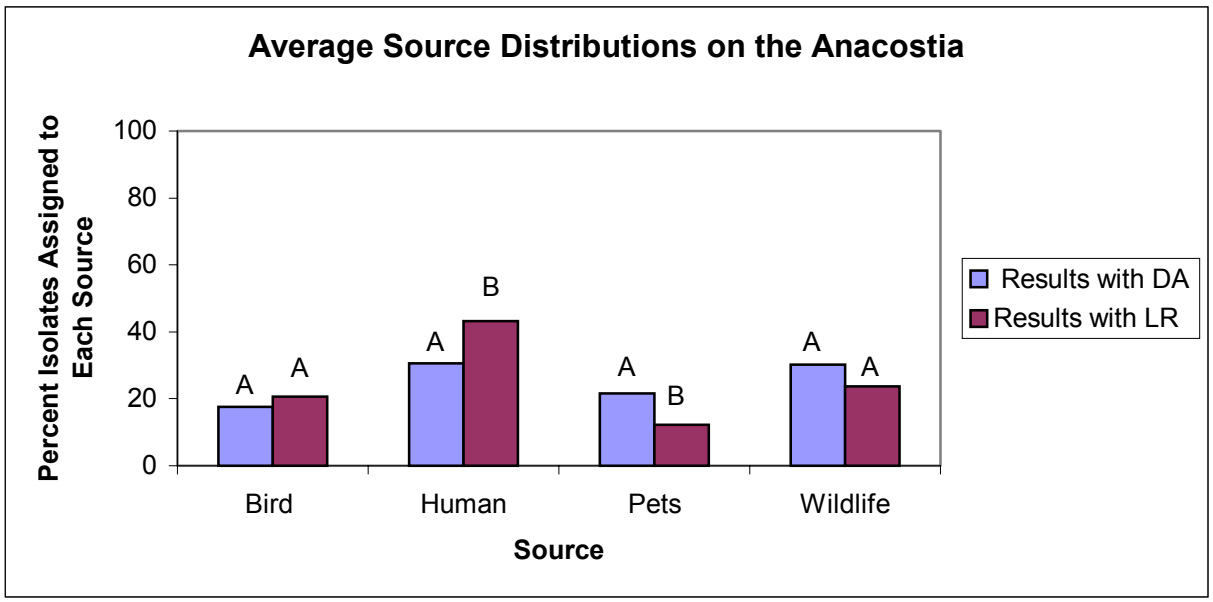
results for the unknown stream isolates produced by each statistical model. The results for DA are detailed for each waterway in Tables 2,4, and 6 of Appendix B. Source distributions and statistical differences between the two models are presented in Figures 4-6.

Figure 4: Potomac Source Distributions with DA and LR



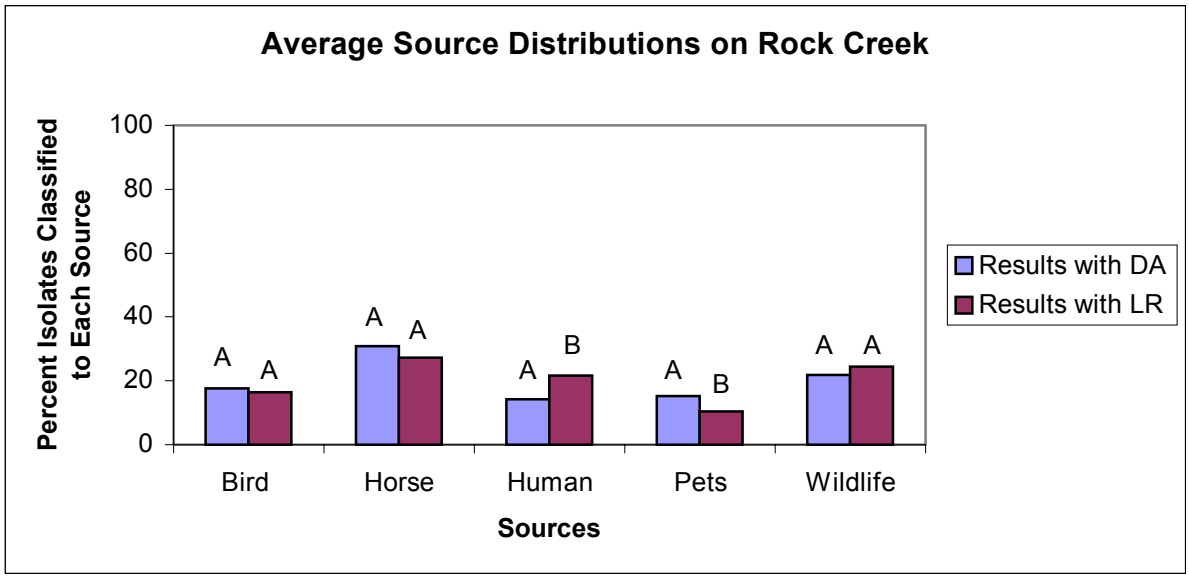
*Common letters denote means that are not significantly different (p>0.001)

Figures 5: Anacostia River Source Distributions with DA and LR



*Common letters denots means that are not significantly different (p>0.001)

Figures 6: Rock Creek Source Distributions with DA and LR



*Common letters denote means that are not significantly different (p>0.001)

Though the ARCC's produced by each model were not significantly different, some of the dominant signatures in each waterway differed between the two models (Appendix B, Tables 2,4,6). The human signature was lost as dominant in the Potomac and Rock Creek models run with DA. The livestock and horse signatures remained dominant with both models, while the wildlife signature was lost in Rock Creek and gained as dominant in the Anacostia. The human signature was consistently and significantly larger with the LR model, while the pet signature was consistently larger with the DA model, except for in the Potomac. Bird, horse, wildlife, and livestock did not show significant differences when run with either model. Making a suggestion as to which model is most appropriate or provides the best results was not feasible with this limited comparison. Both models produced statistically similar correct classification rates and proved to be useful as statistical analysis tools for ARA, but the major source contributions were varied. The results in comparing the distributions for each source calculated by both models showed some significant differences.

Understanding the differences and determining the point of change may provide some clue as to the more appropriate statistical model.

Cross Validation Analysis and the Elimination of Duplicate Patterns

The ARCC is also a measure of library representativeness, as it is a result of resubstitution analysis. Each known source isolate is classified based on the patterns of the entire library, including its own. Wiggins et al (2003) suggest that the ARCC may overestimate representativeness of the library because each isolate is compared against its own pattern, and proposes the use of cross validation analyses ("jack knife analyses") where an entire sample (all isolates taken from the same sample) is removed from the library, one at a time, and compared to the remaining isolate patterns. A new ARCC is then calculated. This analysis required the use of SAS System 8.2. Dr. Bruce Wiggins and colleagues wrote the SAS code to run the cross validation program. All previous studies have been completed in discriminate analysis (DA), therefore the program was written to run DA and not logistic regression. Results of the cross validation analysis are presented in Tables 7, 8, and 9. For comparison purposes, the results of the resubstitution analysis are in parenthesis.

Table 7: Anacostia River Library Cross Validation Analysis

Predicted Source	Percent of Known Source Isolates Assigned to Each Source Cross Validation (Resubstitution)			
	Bird	Human	Pets	Wildlife
Bird	**57.3 (94.0)**	4.4 (2.8)	24.2 (2.0)	14.1 (1.2)
Human	4.9 (3.0)	**67.7 (85.8)**	11.4 (6.7)	16.1 (4.4)
Pets	39.3 (4.8)	15.5 (4.2)	**35.1 (91.1)**	10.1 (0.0)
Wildlife	23 (7.3)	13.4 (6.5)	2.7 (1.9)	**60.9 (84.3)**

* Cross Validation ARCC 55.3%
* Resubstitution ARCC 88.8%

Table 8: Potomac River Library Cross Validation Analysis

Predicted Source	Percent of Known Source Isolates Assigned to Each Source Cross Validation (Resubstitution)				
	Bird	Human	Livestock	Pets	Wildlife
Bird	**57.7 (94.0)**	2.8 (1.2)	12.9 (2.4)	24.2 (2.4)	2.4 (0.0)
Human	4.0 (3.3)	**34.7(77.9)**	21.6 (8.1)	10.0 (5.8)	29.8 (4.9)
Livestock	2.2 (1.1)	4.3 (2.6)	**74.8 (88.8)**	3.7 (2.2)	15.0 (5.3)
Pets	50.0 (5.4)	5.4 (3.0)	19.1 (0.6)	**16.7 (90.5)**	8.9 (0.6)
Wildlife	22.6 (5.8)	8.1 (6.5)	19.9 (10.7)	0.8 (0.8)	**48.7 (76.3)**

*Cross Validation ARCC 46.5%
*Resubstitution ARCC 85.5%

Table 9: Rock Creek Library Cross Validation

Predicted Source	Percent of Known Source Isolates Assigned to Each Source Cross Validation (Resubstitution)				
	Bird	Horse	Human	Pets	Wildlife
Bird	**57.3 (91.5)**	11.7 (4.4)	2.8 (1.6)	22.6 (2.4)	5.7 (0.0)
Horse	3.8 (3.0)	**91.1 (94.5)**	2.1 (0.0)	0.0 (0.0)	3.0 (2.6)
Human	3.5 (3.0)	22.3 (3.7)	**51.9 (82.6)**	10.5 (5.8)	11.9 (4.9)
Pets	41.1 (5.4)	5.4 (1.8)	12.5 (3.6)	**31.6 (89.3)**	9.5 (0.0)
Wildlife	23.4 (6.9)	11.5 (6.1)	11.5 (5.0)	1.5 (0.4)	**52.1 (81.6)**

*Cross Validation ARCC 56.8%
*Resubstitution ARCC 87.9%

Pulled-sample cross validation was a way of simulating the testing of new isolates. This analysis removed all isolates from each sample one at a time from the library and reclassified them. Comparing the ARCC from the resubstitution analysis and the cross validation analysis provided a means of quantifying library representativeness. If the difference was small (less than 5%) then the library could be considered representative (Wiggins et al., 2003). Upon taking out an entire sample and comparing it to the rest, the likelihood of having identical or nearly identical patterns in the library associated with isolates collected from the same sample was much smaller. The differences obtained for the Anacostia Library was 37%, 46% for the Potomac Library, and 35% for the Rock Creek Library, meaning each library could be considered unrepresentative of the isolates in their respective watersheds. Wiggins et al. (2003) performed pulled sample cross validation

analyses on six libraries and found new ARCC's ranging from 46% to 61%, (similar to those obtained in this study), while some of the libraries failed the representativeness test, only the larger libraries were actually considered representative due to lower correct classification rates obtained with resubstitution analysis.

Another suggestion by Wiggins et al. (2003) that is actually popular in the molecular community is the elimination of duplicates from the library, so that each unknown isolate is only compared to unique isolate patterns. This eliminates the possibility of isolates being compared to themselves when performing resubstitution analysis, and according to Wiggins (personal communication, 2003), this is the most accurate way to maintain a representative library. The duplicates were removed from the libraries using SAS System 8.2. The libraries were then rerun using DA.

Table 10: Elimination of Duplicate Patterns in the Anacostia Library

Predicted Source	Percent of Known Source Isolates Assigned to Each Source No Duplicates (With Duplicates)			
	Bird	Human	Pets	Wildlife
Bird	**76.2 (94.0)**	6.4 (2.8)	7.9 (2.0)	9.5 (1.2)
Human	4.1 (3.0)	**74.2 (85.8)**	11.1 (6.7)	10.5 (4.4)
Pets	10.0 (4.8)	10.0 (4.2)	**77.5 (91.1)**	2.5 (0.0)
Wildlife	5.6 (7.3)	8.2 (6.5)	0.9 (1.9)	**85.4 (84.3)**

*ARCC with No Duplicates 78.3%
*ARCC with Duplicates 88.8%

Table 11: Elimination of Duplicate Patterns in the Potomac Library

Predicted Source	Percent of Known Source Isolates Assigned to Each Source Cross Validation (Resubstitution)				
	Bird	Human	Livestock	Pets	Wildlife
Bird	**76.2 (94.0)**	1.6 (1.2)	6.4 (2.4)	6.4 (2.4)	9.5 (0.0)
Human	4.1 (3.3)	**68.6 (77.9)**	9.3 (8.1)	10.5 (5.8)	7.6 (4.9)
Livestock	2.3 (1.1)	7.5 (2.6)	**72.4 (88.8)**	6.3 (2.2)	11.5 (5.3)
Pets	10.0 (5.4)	12.5 (3.0)	2.5 (0.6)	**72.5 (90.5)**	2.5 (0.6)
Wildlife	5.5 (5.8)	8.2 (6.5)	17.3 (10.7)	0.9 (0.8)	**68.2 (76.3)**

*ARCC with No Duplicates 71.6%
*ARCC with Duplicates 85.5%

Table 12: Elimination of Duplicate Patterns in the Rock Creek Library

Predicted Source	Percent of Known Source Isolates Assigned to Each Source Cross Validation (Resubstitution)				
	Bird	Horse	Human	Pets	Wildlife
Bird	**73.0 (91.5)**	9.5 (4.4)	1.6 (1.6)	7.9 (2.4)	7.9 (0.0)
Horse	7.0 (3.0)	**90.7 (94.5)**	2.3 (0.0)	0.0 (0.0)	0.0 (2.6)
Human	2.9 (3.0)	8.1 (3.7)	**69.2 (82.6)**	11.1 (5.8)	8.7 (4.9)
Pets	10.0 (5.4)	0.0 (1.8)	10.0 (3.6)	**75.0 (89.3)**	5.0 (0.0)
Wildlife	5.5 (6.9)	10.0 (6.1)	8.2 (5.0)	0.9 (0.4)	**75.5 (81.6)**

*ARCC with No Duplicates 76.7%
*ARCC with Duplicates 87.9%

The elimination of the duplicate patterns substantially decreased the ARCC's of each library, specifically in the Potomac library (Tables 10-12). The results of the cross validation analysis and removal of the duplicates indicate that there were many similar or identical antibiotic resistance patterns associated with isolates that were collected from the same sample. The Anacostia library was created with 1,107 isolates. Removing the duplicates left only 385 isolates with unique patterns. The Potomac library was created with 1,806 isolates, and removal of the duplicate patterns left only 559 isolates. The Rock Creek library contained 1,342 isolates, and removal of the duplicate patterns left only 428 unique isolate patterns.

The pulled-sample cross validation analysis and removal of duplicate patterns is a topic of some debate in the source tracking community for several reasons. Taking out a high number of similar isolates may actually underestimate library representativeness because these are the isolates (and their patterns) that represent the most common patterns found in the environment. Leaving only unique isolates to compare unknowns against then becomes an issue of matching and is no longer a function of population ecology or a representation of the isolates found in the environment in the distribution of the antibiotic resistance patterns that exist. The other argument is that a common isolate resistance profile

should receive more weight in the analysis than a resistance pattern that occurs very little. Another very important issue related to these procedures is cost. If over half of all the isolates that are collected are duplicates, then a larger number of isolates are going to have to be sampled in order to maintain a library that is both sufficiently large and representative. Much time is then wasted on collection and analysis of isolates that will have to be discarded.

SUMMARY

This chapter entitled "Expanding the Limits of ARA" is as important to the science of ARA as were the results of source contributions of fecal contamination on the Potomac River, the Anacostia River, and Rock Creek. Field-testing the method of BST, especially in an urban environment, is a crucial part of progress to find methods that can effectively identify sources of pollution, but there are many issues that have not been addressed that play an important role in determining the effectiveness of ARA or any other method. Three important issues addressed here were; what happens when a lower limit was placed on the probability that an isolate matches any of the known source profiles, was the currently accepted statistical model (DA) the most appropriate model for ARA, and how does eliminating duplicate antibiotic resistance profiles affect the library.

Adding a criterion that accounts for the probability of how well an isolate matches the isolates in the known source database provided a better understanding of how well the library was able to classify isolates and showed that the interpretation of dominant signatures should be aired on the side of caution. From this data, the isolates that were classified as human appeared to be the most likely to be included in the unknown source category, indicating that many of those isolates did not match any of the known human

signature patterns to the lower limit that was specified that an isolate should meet in order to classify it as a particular source with confidence (80%). Human was a dominant signature in the monthly averages in all three waterways, and after the 80% confidence limit was used, human was a dominant signature only in the Anacostia River.

Creating an unknown source category to hold all isolates that do not match a known source category with enough of a similarity was not something previously discussed or explored in the literature for ARA, but has been discussed briefly in some molecular BST studies. In a PFGE study, Simmons et al. (2002) suggested and used an 80% cutoff and found that 46% of the isolates were classified as unknown. In a ribotyping study, Samadpour et al. (1995) used an 80% cutoff and found that between 48% and 67% of the water ribotypes could not be matched. In this ARA study between 41% and 44% of the isolates in the libraries were classified as unknowns. This clearly indicated that ARA of environmental samples provided results with similar classification abilities to those of more time consuming, expensive, and precise molecular methods. When the isolates in the unknown category were excluded from the calculations of source distributions, it was found that only the human signature was statistically different in the Potomac and Anacostia when compared to source distributions calculated with all of the isolates. This indicated that applying the stringent cutoff may remove a significant portion of the unknown isolates that can be used in the analysis, but it did not significantly change average source distributions. As of now, the process to create an unknown source category had to be completed manually and was very tedious and time consuming. More research needs to be done to corroborate the findings of both molecular and non-molecular methods after the addition of an unknown source category, but the results of this analysis provided sufficient evidence that source

distributions do not change if an unknown category is implemented to hold isolates that do not match signatures in the known source library by a certain margin.

Appropriate statistical modeling for ARA is something that has not been addressed in previous research. DA has been widely accepted and applied as the statistical tool for BST projects. The underlying assumptions that validate the analysis, however may be violated when applied to create a statistical model for the classification of microbes. DA requires prespecified, well-defined groups, and enteric bacteria from environmental samples do not necessarily fit this profile. LR was employed as well as DA to determine if LR provided a better statistical model, based on rates of correct classification. There was a no significant difference between the ARCC's produced by the two models with any of the libraries, therefore no suggestion was readily made as to which statistical tool provides a better model of the microbial population's antibiotic resistance profile. The models were run on the unknown samples, and produced statistically similar results for a majority of the sources, with the exception of human in all three waterways and pets in the Anacostia and Rock Creek. Though distributions were similar for most sources, dominant signatures differed in all three waterways. The underlying assumptions for LR apply better for this type of analysis, but DA is the commonly used and accepted statistical tool, so more research is needed before any recommendation can be made to utilize LR instead of DA.

There has been some question and dispute as to when a library can be considered representative of the isolates found in the environment. It was suggested by Wiggins et al. (2003) that a library be tested for representativeness via pulled sample cross validation analysis and that a library should contain no duplicate antibiotic resistance patterns because they bias the RCC's of the library when resubstitution analysis is run. These tests were run

on each library and it was found that the ARCC's from the pulled sample and removed duplicates libraries were significantly different than the ARCC's produced by the resubstitution analysis. According to Wiggins, these libraries should be considered unrepresentative. These methods are not yet approved by the entire source tracking community and there is considerable debate about whether these analyses are even appropriate. Removing duplicates and even performing pulled sample cross validation takes out the patterns that appear most often and weights that pattern the same as a pattern that occurs very rarely. In fact, between 61% and 65% of the isolate patterns in each library were thrown out as duplicates, leaving the remaining portion of the library as small and with significantly lower classification rates. The additional time and cost that would be required to build subsequent libraries with only unique isolate patterns has not been justified to the entire microbial source tracking community based on the recent research that has been completed on this subject, but not yet published. To date there is no substantiating evidence that creating a large library based exclusively on unique patterns would provide higher and more useful RCC's of known source isolates and more accurate classifications of unknown source isolates. Until more research is done on to answer questions and provide justifications for issues such as how many unique isolates are required to bring up the RCC's and how much more time and money will these library based methods require to create libraries of unique isolate patterns, there will be no real progress or application of these analyses.

REFERENCES

Hosmer, D. W., and S. Lemeshow. *Applied Logistic Regression, 2nd Edition.* New York: John Wiley & Sons, 2000.

Kariuki, S. C. Gilks, J. Kimari, A. Obanda, J. Muyodi, P. Waiyaki, and C.A. Hart. 1998. Genotype Analysis of *Esherichia coli* Strains Isolated from Children and Chickens Living in Close Contact. Applied and Environmental Microbiology. 65 (2): 472-476.

Kleinbaum, D.G., K. Dietz, M. Gail, K. Krickeberg, B. Singer (Editors). *Logistic Regression, A Self-Learning Text.* Springer-Verlag New York Inc., 1994. pp. 1-5.

McGarigal, K., S. Cushman, S. Stafford. *Multivariate Statistics for Wildlife and Ecology Research*. Springer-Verlag New York, Inc., 2000. pp. 129-180.

Samadpour, M., N. Chechowitz. 1995. Little Soos Creek Microbial Source Tracking, A Survey. University of Washington Department of Environmental Health.

Simmons, G.M., D.F. Waye, S. Herbein, S. Myers, and E.Walker. 2002. Estimating Nonpoint Source Fecal Coliform Sources Using DNA Profile Analysis. Advances in Water Monitoring Research. Tamim Younos, Editor. Water Resources Publications, LLC. 143-167.

Wiggins, B.A. 1996. Discriminate Analysis of Antibiotic Resistance Patterns in Fecal Streptococci, a Method to Differentiate Human and Animal Sources of Fecal Pollution in Natural Waters. Applied and Environmental Microbiology. 62 (11): 3997-4002.

CHAPTER VI

SUMMARY

The Need for Antibiotic Resistance Analysis

The number and nature of pathogenic impairments in Washington D.C. waterways has left the District with a need to determine the sources of fecal contamination for major recreational waters. The primary objectives of this study were to build a library of antibiotic resistance patterns from known sources that was of appropriate size and representativeness and to use that library to determine sources and source distributions of unknown fecal environmental isolates deposited in the Anacostia River, the Potomac River, and Rock Creek. These three waterways were identified as exceeding water quality standards set for fecal coliform levels and were designated by the District of Columbia to the Environmental Protection Agency's 303 (d) impaired waters list. The initial sources of fecal contamination were thought to come from combined sewer overflows, direct deposits of feces from wildlife, and separate sanitary sewer overflows, which can result from leaky or undersized sanitary sewer pipes, and stormwater runoff, which includes overland flow and flow conveyed through storm sewer pipes.

The Known Source Library

A library profile of 1,806 enterococcus isolates from known sources was built based on antibiotic resistance patterns from thirty concentrations of nine antibiotics. These sources included human, cattle, chicken, horse, goat, sheep, deer, raccoon, muskrat, goose, seagull, coyote, duck, wild turkey, dog, and cat. To account for localized sources of contamination, a sub library was built with isolates and sources specifically found in each watershed. The Average Rate of Correct Classification (ARCC) for the Potomac River library was 89.1%,

the ARCC of the Anacostia River library was 93.0%, and the ARCC of the Rock Creek library was 92.8%. The library was tested and approved for appropriate size and representativeness. Over the sampling period, new known source isolates were added to the database and changes in the ARCC's of the libraries were noted. One important implication in this test for representativeness was that over the sampling, the ARCC's declined, though not enough to classify the library unrepresentative. The decline was within reason for the purposes of a one-year study, but does indicate that the libraries were not temporally stable. This is an issue raised by many regarding library based methods, but has not been thoroughly addressed in the literature.

Antibiotic Resistance Analysis Results

Antibiotic profiles were characterized for 24 unknown enterococci isolates on each of 198 samples (38 samples from the Potomac River, 79 samples from the Anacostia River, and 81 samples from Rock Creek) collected periodically from July 2002 through April 2003 as well as two major storm events. These isolate profiles were compared to the known source library using logistic regression. Three dominant sources of fecal pollution were detected in the Potomac River; livestock (30%), human (29%), and wildlife (22%). Three dominant signatures were also detected in Rock Creek; horse (26%), human (26%), and wildlife (24%). Human was the only dominant source detected in the Anacostia River, averaging 43% over the sampling period. Site to site variations existed that differed in some instances from the source distributions found as an average for each waterway. Site to site variations were explained in many instances by determining the sources adjacent to each site and evaluating their potential impact. All sources except human were more easily related to variations among sites. Human proved difficult to interpret because the location of many of

the discharge points of human wastewaters was unknown. Seasonal and storm related variations were also determined for each waterway. Storm variations approved to be more difficult to interpret and even varied between wet and dry seasons, making predictability of source distributions in relation to precipitation from single storm events difficult. Seasonality on the other hand provided some consistent and predictable patterns among the three waterways. Human was found to be a major contributor in the dry season in the Anacostia and Potomac. Wildlife was also found to be a dominant signature in all three waterways during the dry season. Horse, livestock, and pets were all major source contributors in the wet season, most likely as a result of increased runoff. Bird was the only source consistently unaffected by seasonal variations.

The results of this study indicated that human was a substantial contributor to the fecal contamination problems, especially in the Anacostia River, but there were significant agricultural and wildlife contributions as well. The major variability of the source distributions by season was another finding that was unique to highly urbanized environments and will need to be considered in remediation efforts. The results of this project will aid the Metropolitan Washington D.C. Council of Governments in making important management decisions to help improve the water quality in and around the Washington D.C. area.

Expanding the Limits of ARA

Expanding the limits of ARA was also an integral part of this research. Three new and even controversial analyses were run on the data collected from this project in an attempt to provide confidence and direction in the results of this study. The first was a comparison of the more commonly used statistical analysis model discriminate analysis

(DA) with a possibly more appropriate statistical model, logistic regression (LR). No significant differences were found between the output of the two models for the known source libraries. When the unknown isolates were run with each model, source distributions and dominant signatures were statistically different for some of the samples, and without further research to expound upon the differences or confirm no statistical difference between the output of each model, no suggestion could be made in favor of one model over the other.

Another analytical test of the data was the introduction of a standard imposed on the data output by the requiring isolates to meet a minimum of 80% similarity to the known source profiles it is classified to. With the 80% cutoff, between 41% and 44% of the isolates could not be classified to any source and were placed in an unknown category. These results are similar to results obtained by several molecular methods that have used this same criterion for unknown source classifications. This indicated that ARA classifies isolates on par with the more precise, expensive, and time cosuming molecular methods, thus enhancing its appeal as a BST tool. The remaining isolates were used to recalculate source distributions and it was found that these source distributions were not statistically different than those calculated with no restriction for isolate similarity for matching. This indicated that the unknown isolates could be analyzed without this restriction and the results could be upheld with confidence.

The last major test of the data was the analysis of the library for representativeness via pulled sample cross validation and the exclusion of all duplicate patterns from the known source library. These analyses did not confirm the representativeness of the databases because the ARCC's calculated with these analyses were greater than 5% different than the original ARCC's of the library. These analyses are not accepted by the entire source

tracking community and are still the center of some controversy. Those in opposition claim that eliminating the dominant signatures from the library or even all duplicates creates an analysis for isolate matching and is no longer a test of population ecology by no longer representing the isolates in the proportion that they are found in the environment. The other downside of these suggestions is that in order to create a library that is of appropriate size and representativeness, thousands of isolates will be needed and the time and supplies required for finding only isolates that are unique will greatly increase the costs associated with library dependent methods.

The bottom line is that ARA proved to be a very valuable tool for the examination of sources and source distributions of fecal contamination in a highly urbanized area. The source distributions and seasonal patterns are consistent with what was expected based on prior knowledge of each watershed. Seasonal variations proved to be complex in that human was found to be a significant contributor in the dry season instead of the wet season when most of the combined sewer overflows occur. This was also found in another study of a highly urbanized region, and is most likely explained by dilution potentials in the larger waterways (Anacostia and Potomac) and the presence of unknown discharges of human wastes.

APPENDIX A

Figure 1: Sampling Locations for Rock Creek and the Potomac River.

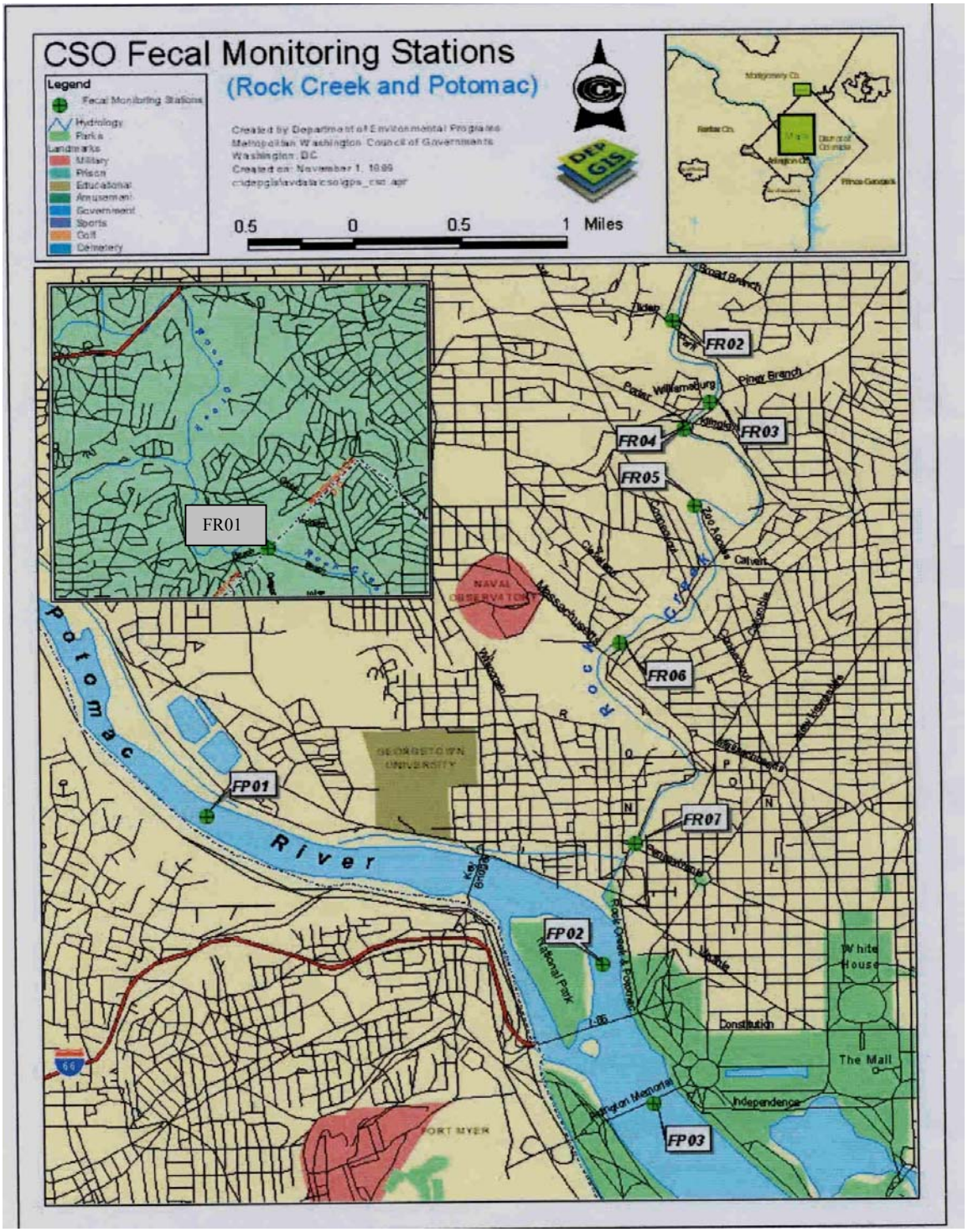


Figure 2: Sampling Locations for the Anacostia River.

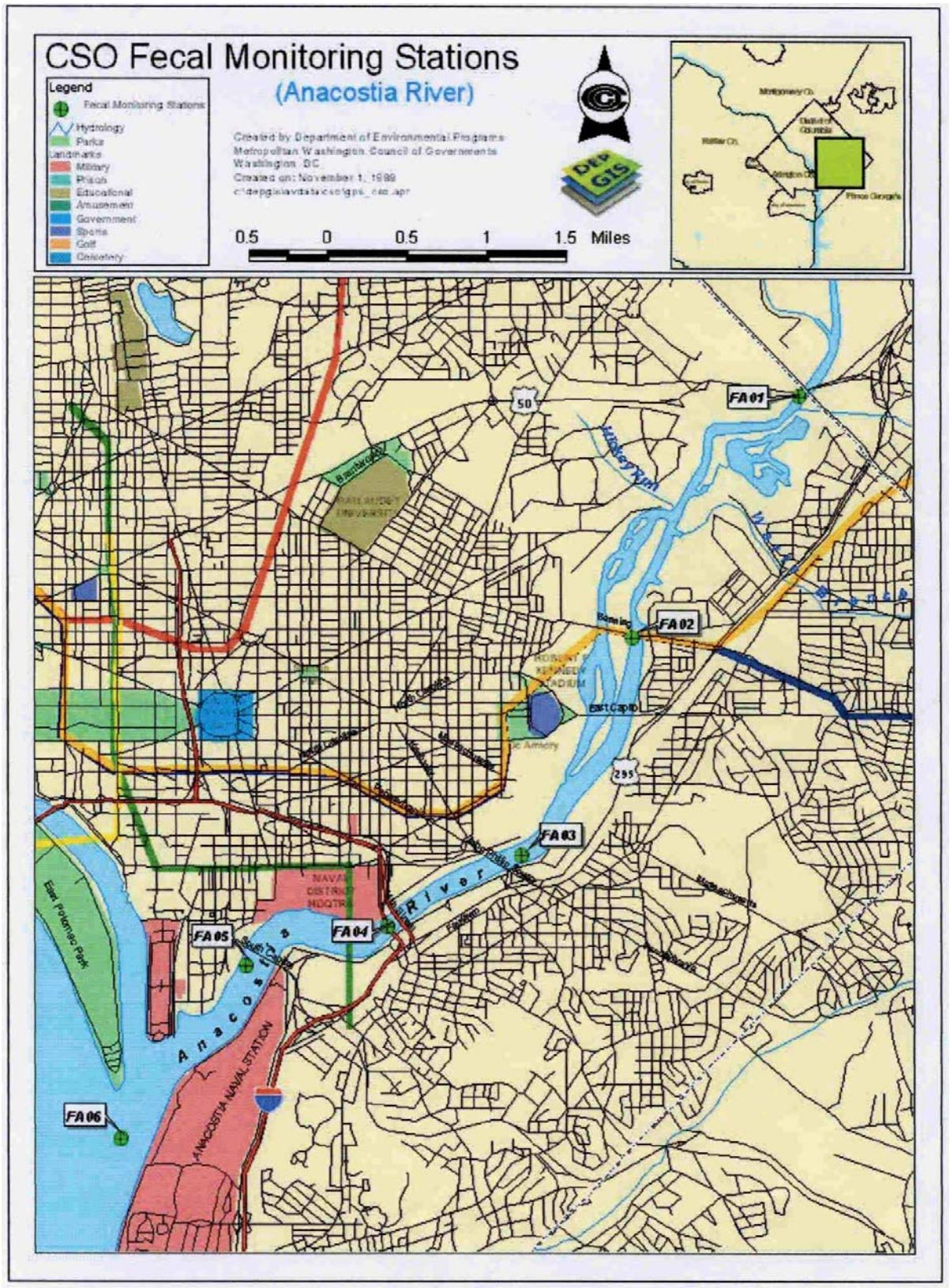


Figure 3: Monthly and Storm Event *Enterococci* Counts for the Potomac River

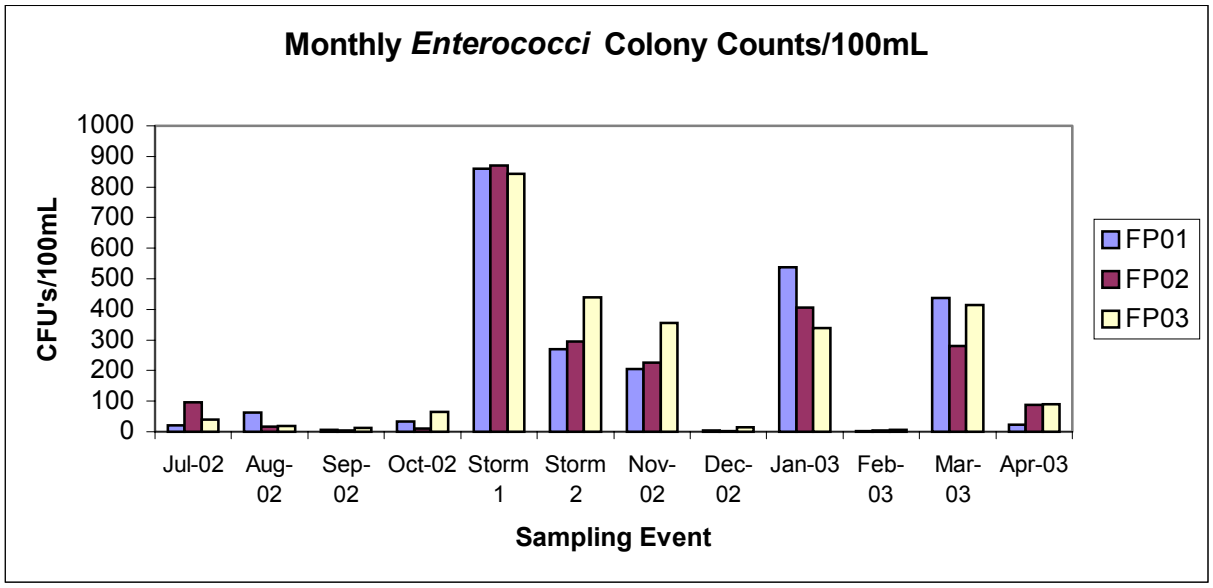


Figure 4: Monthly and Storm Event Fecal Coliform Counts for the Potomac River

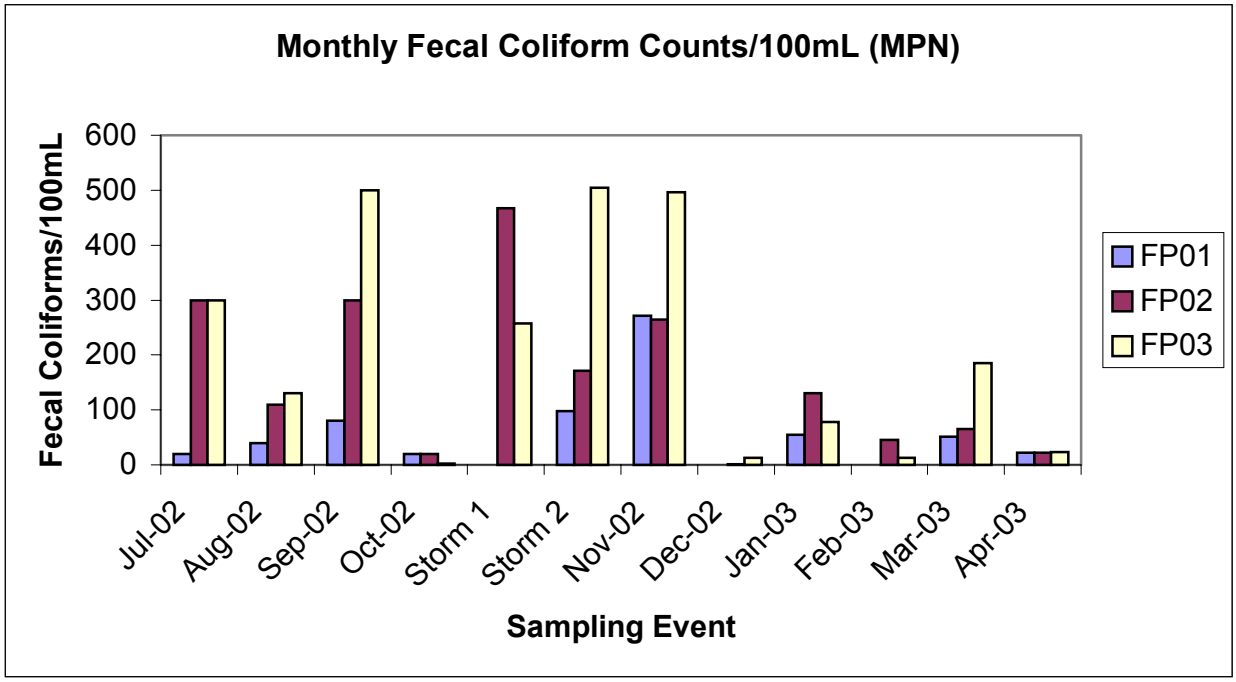


Figure 5: Monthly and Storm Event *Enterococci* Counts for the Anacostia River

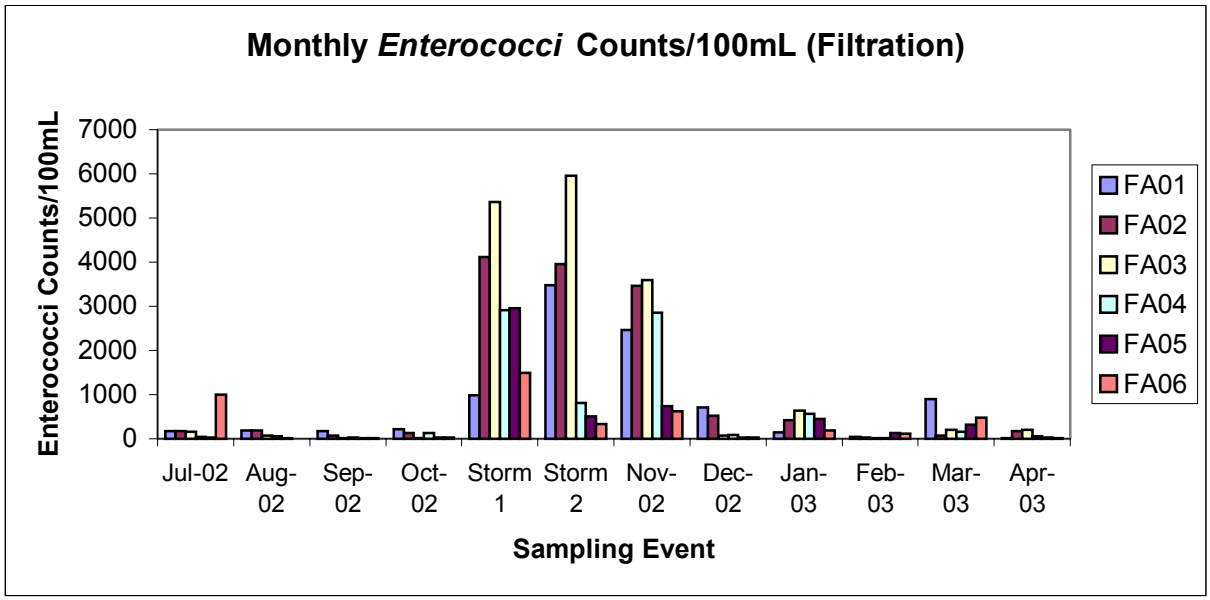


Figure 6: Monthly and Storm Event Fecal Coliform Counts for the Anacostia River

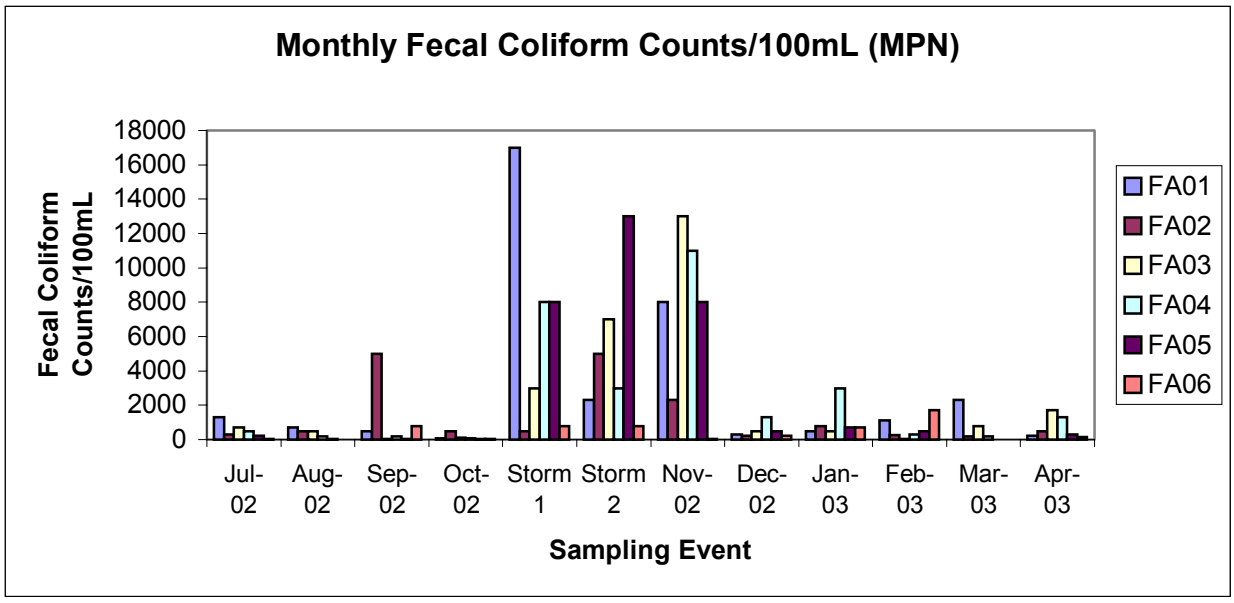


Figure 7: Monthly and Storm Event *Enterococci* Counts for Rock Creek

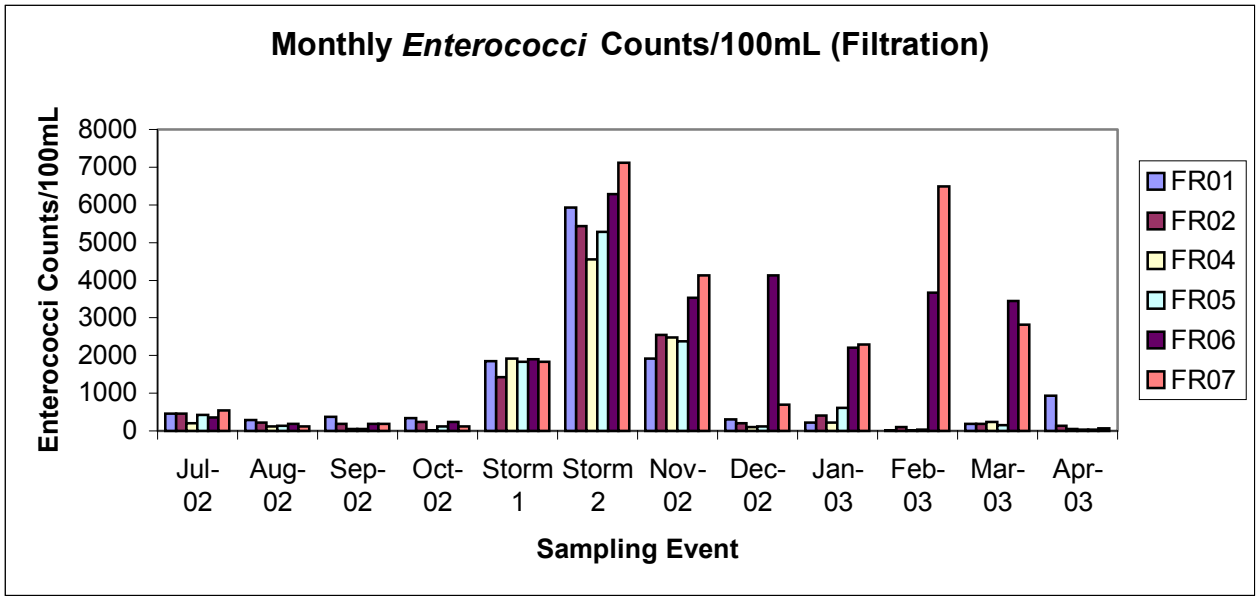


Figure 8: Monthly and Storm Event Fecal Coliform Counts for Rock Creek

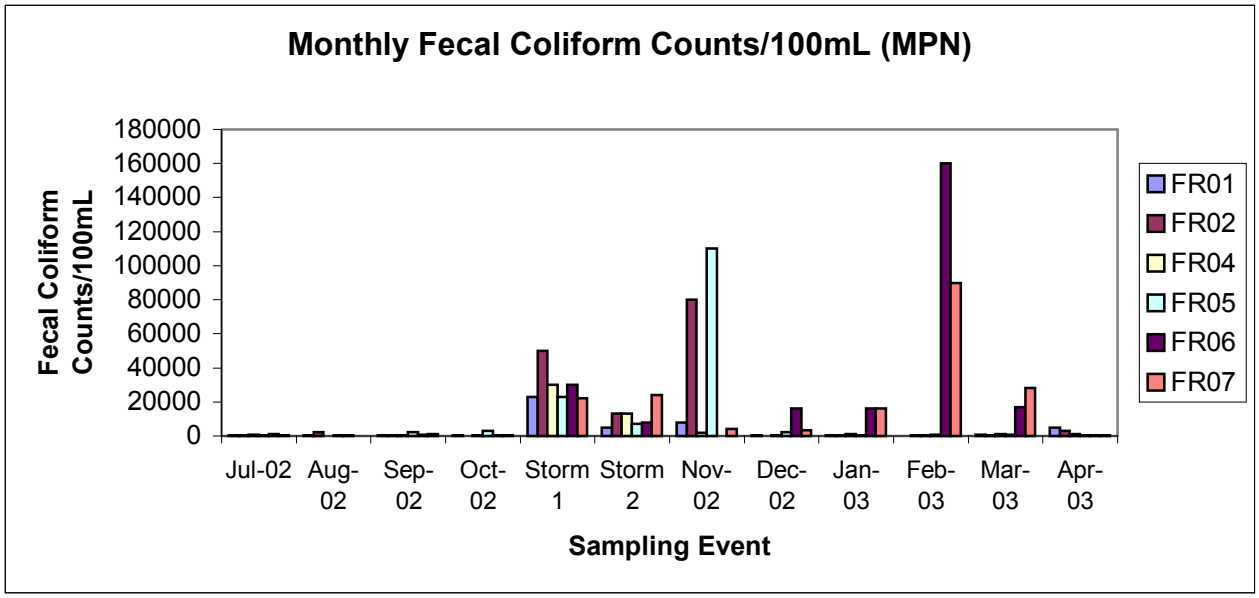


Figure 9: Relating Flow to Indicator Organism Densities

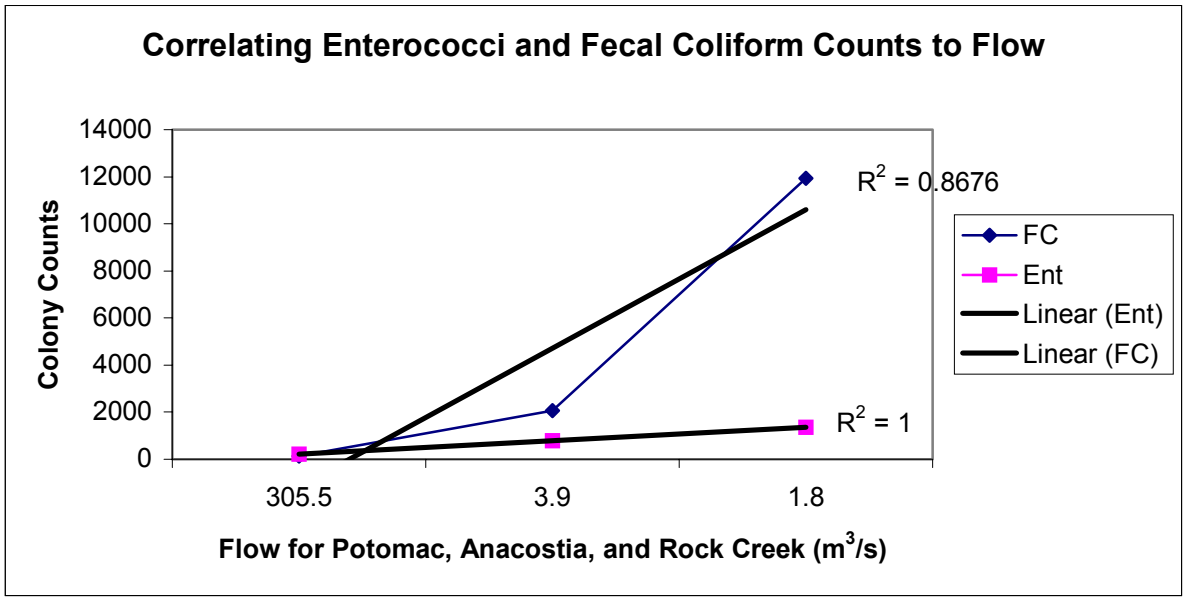


Figure 10: Precipitation Totals at Regan National Airport

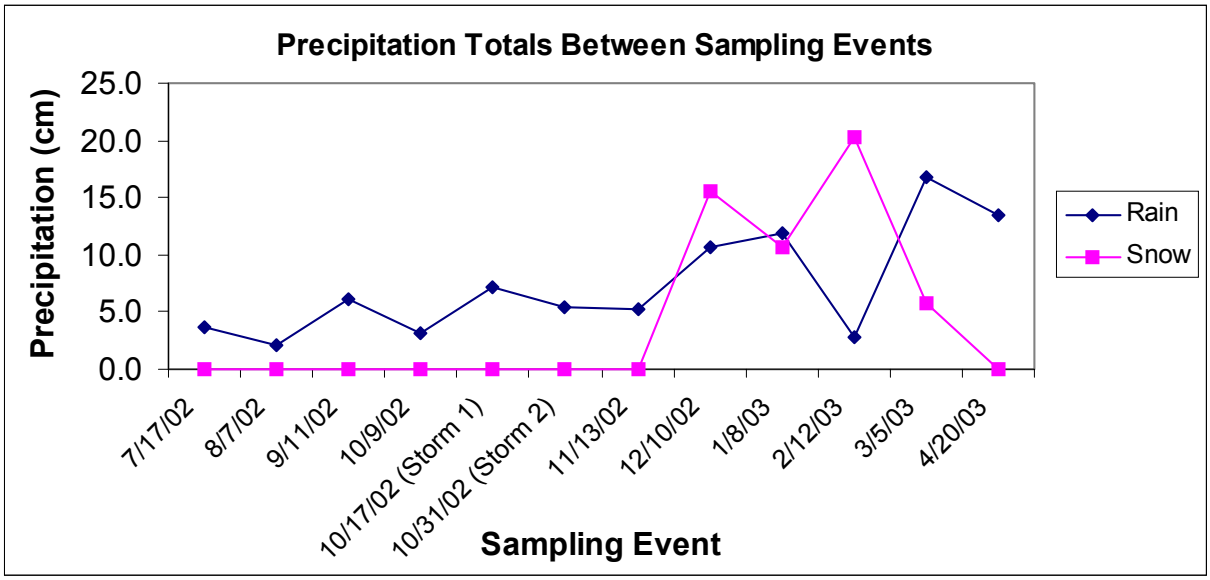


Figure 11: Fecal Coliform and *Enterococci* Densities in the Potomac as a Function of Precipitation Between Sampling Events

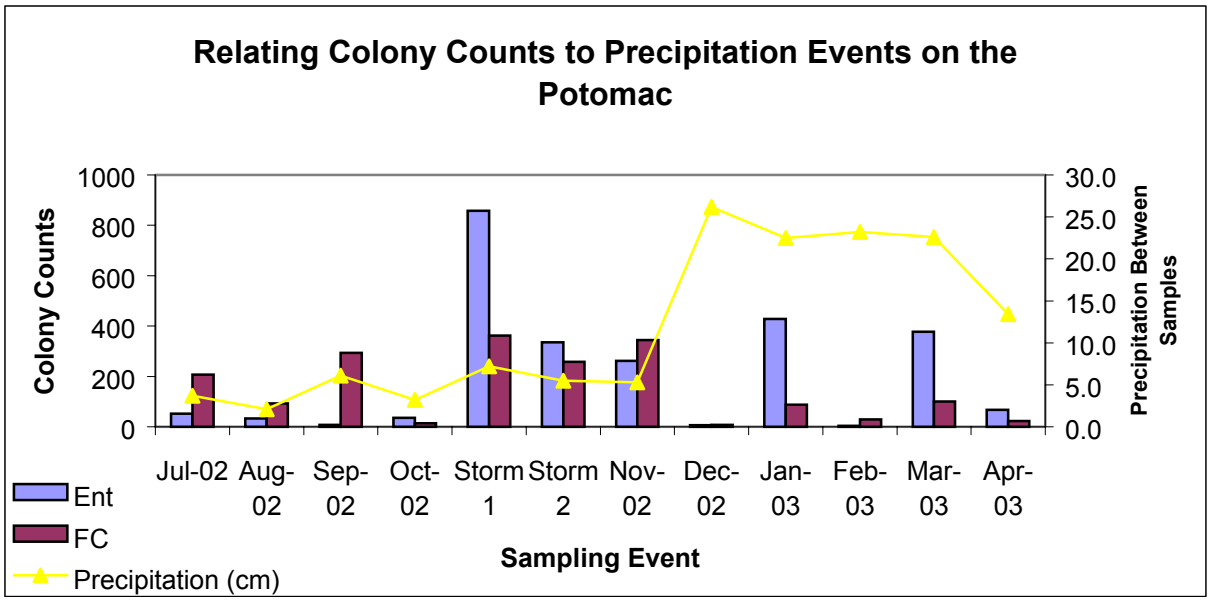


Figure 12: Fecal Coliform and *Enterococci* Densities in the Anacostia as a Function of Precipitation Between Sampling Events

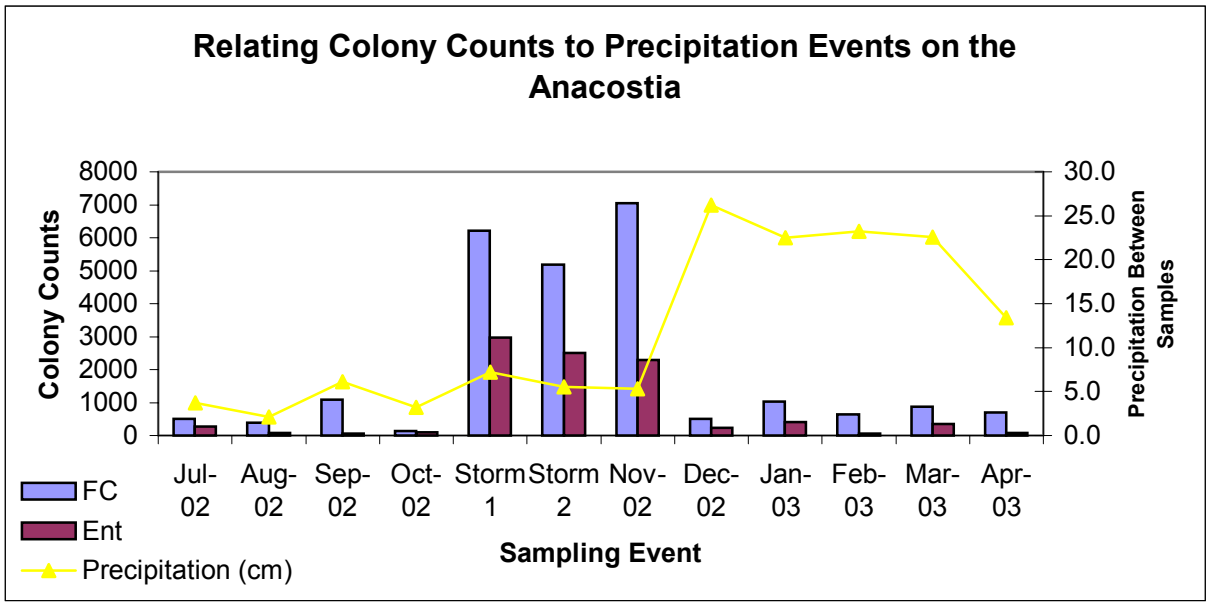
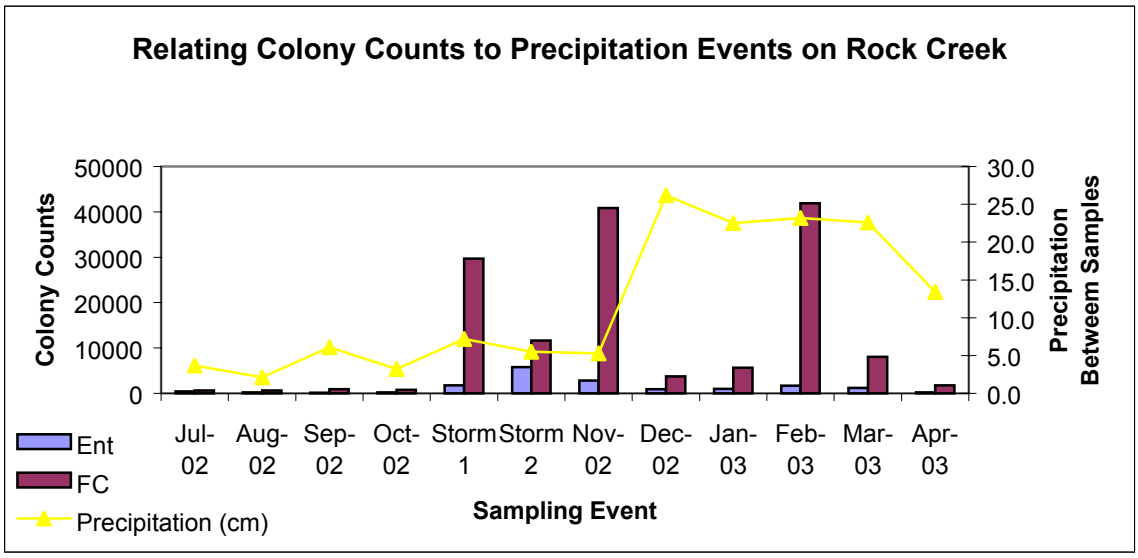


Figure 13: Fecal Coliform and *Enterococci* Densities in Rock Creek as a Function of Precipitation Between Sampling Events



APPENDIX A

Table 1: Potomac River Library Source Classification Numbers

Potomac										
	Bird		Human		Livestock		Pets		Wildlife	
	CC	MC	CC	MC	CC	MC	CC	MC	CC	MC
Bird	237			5		8		4		9
Human		3	371			28		3		3
Livestock		4		18	646			4		29
Pets		1		13		4	154			2
Wildlife		3		23		16		3	208	
Total Isolates	248		430		699		168		261	
% CC	95.6		86.3		92.4		91.7		79.7	
% MC		4.4		13.7		8.0		8.3		16.5
Average % MC	10.2									
Standard Deviation of %MC	4.8									
MDP	29.5%									

(CC-correctly classified)
(MC-misclassified)

Table 2: Anacostia River Library Source Classification Numbers

Anacostia								
	Bird		Human		Pets		Wildlife	
	CC	MC	CC	MC	CC	MC	CC	MC
Bird	238			2		4		7
Human		5	394			5		12
Pets		2		16	158			6
Wildlife		3		18		1	236	
Total Isolates	248		430		168		261	
% CC	96.0		91.6		94.0		90.4	
% MC		4.0		8.4		6.0		9.6
Average % MC	7.0							
Standard Deviation of % MC	2.5							
MDP	16.9%							

(CC-correctly classified)
(MC-misclassified)

Table 3: Rock Creek Library Source Classification Numbers

Rock Creek										
	Bird		Horse		Human		Pets		Wildlife	
	CC	MC	CC	MC	CC	MC	CC	MC	CC	MC
Bird	237			3		3		4		7
Horse		5	229			8		0		10
Human		2		1	387			5		12
Pets		2		0		16	158			6
Wildlife		2		2		16		1	226	
Total Isolates	248		235		430		168		261	
% CC	95.6		97.4		90		94.0		86.6	
% MC		4.4		1.4		10		6.0		13.4
Average % MC	7.0									
Standard Deviation of % MC	4.7									
MDP	25.9%									

(CC-correctly classified)
(MC-misclassified)

Table 4: Percent of Isolates Below Detection Limits in Potomac River Library as Determined by Expected Frequency of Misclassification (EFMC) and Minimum Detectable Percentage (MDP)

Potomac									
Bird		Human		Livestock		Pets		Wildlife	
EFMC	MDP	EFMC	MDP	EFMC	MDP	EFMC	MDP	EFMC	MDP
47%	100%	3%	70%	7%	67%	43%	87%	10%	77%

Table 5: Percent of Isolates Below Detection Limits in Anacostia River Library as Determined by Expected Frequency of Misclassification (EFMC) and Minimum Detectable Percentage (MDP)

Anacostia							
Bird		Human		Pets		Wildlife	
EFMC	MDP	EFMC	MDP	EFMC	MDP	EFMC	MDP
10%	55%	0%	17%	27%	73%	5%	53%

Table 6: Percent of Isolates Below Detection Limits in Rock Creek Library as Determined by Expected Frequency of Misclassification (EFMC) and Minimum Detectable Percentage (MDP)

Rock Creek									
Bird		Horse		Human		Pets		Wildlife	
EFMC	MDP	EFMC	MDP	EFMC	MDP	EFMC	MDP	EFMC	MDP
13%	85%	15%	53.00%	5%	73%	40%	87%	8%	50%

APPENDIX B

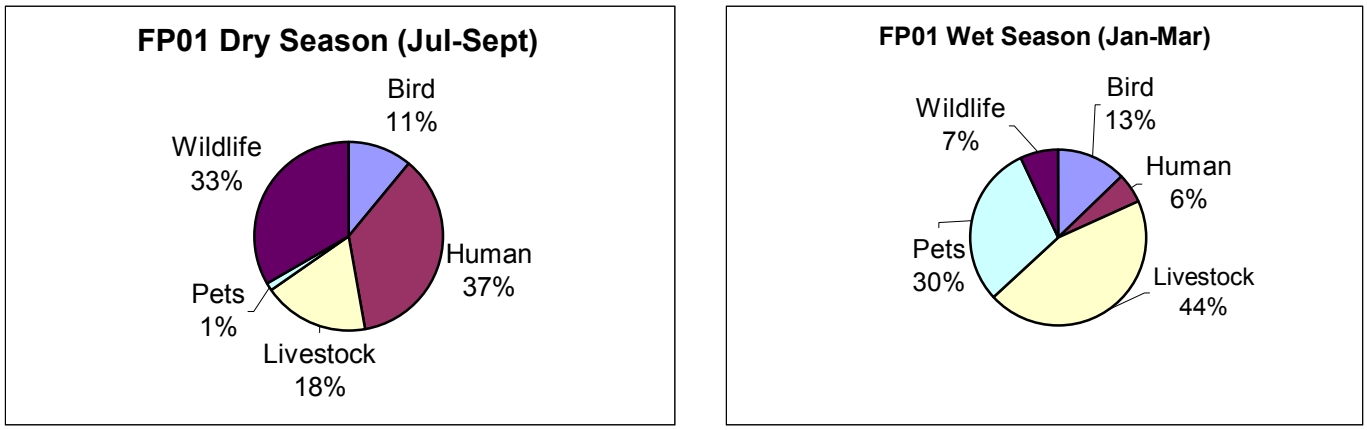
Potomac River

Table 1: Significant Differences of Mean Source Distributions on the Potomac

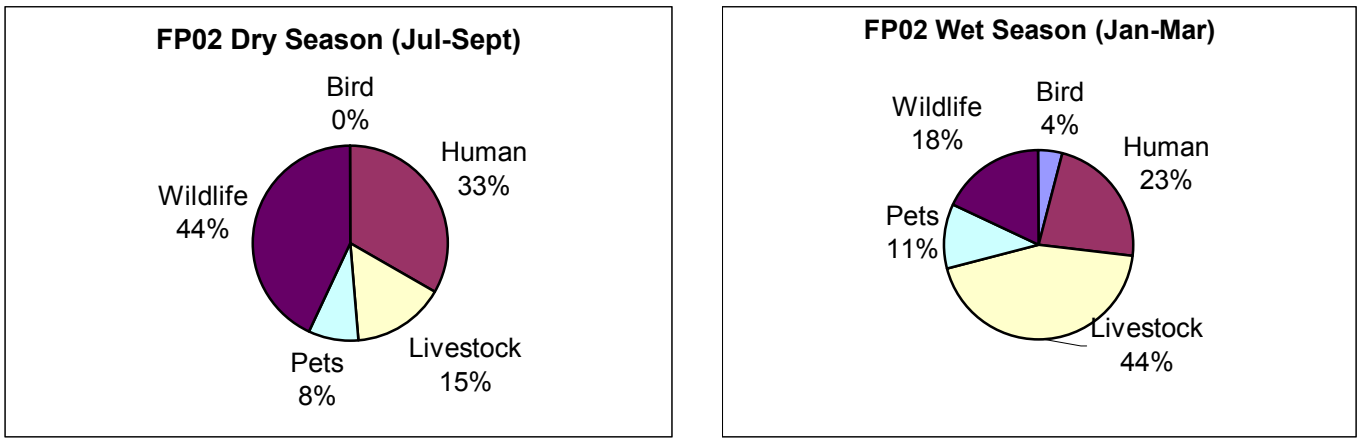
Station	*Bird*	*Human*	*Livestock*	*Pets*	*Wildlife*
FP01	8.4	21.7	34.3	15.6	20.0
FP02	2.9	32.3	30.2	9.2	25.4
FP03	5.8	32.3	27.4	13.6	20.9
Mean	5.7 C	28.8 A	30.6 A	12.8 B	22.1 A

*Common letters denote means that were not statistically different (p>0.001)

Figure 1: Dry and Wet Season Source Contributions for site FP01



Figures 2: Dry and Wet Season Source Contributions for site FP02



Figures 3: Dry and Wet Season Source Contributions for site FP03

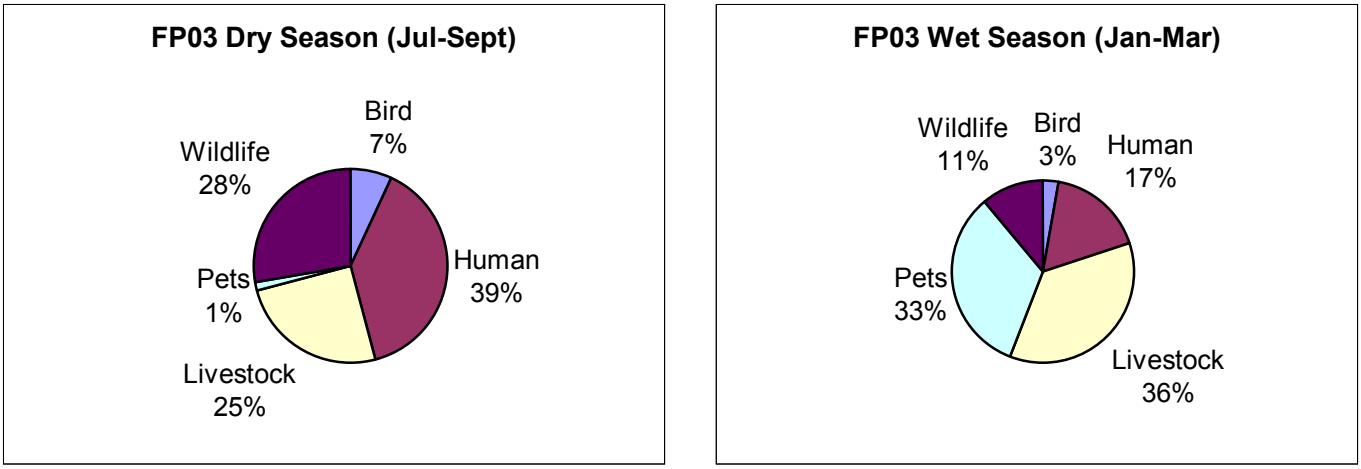


Figure 4: Source Contributions at Sampling Site FP01 in the Months Surrounding Storm Events

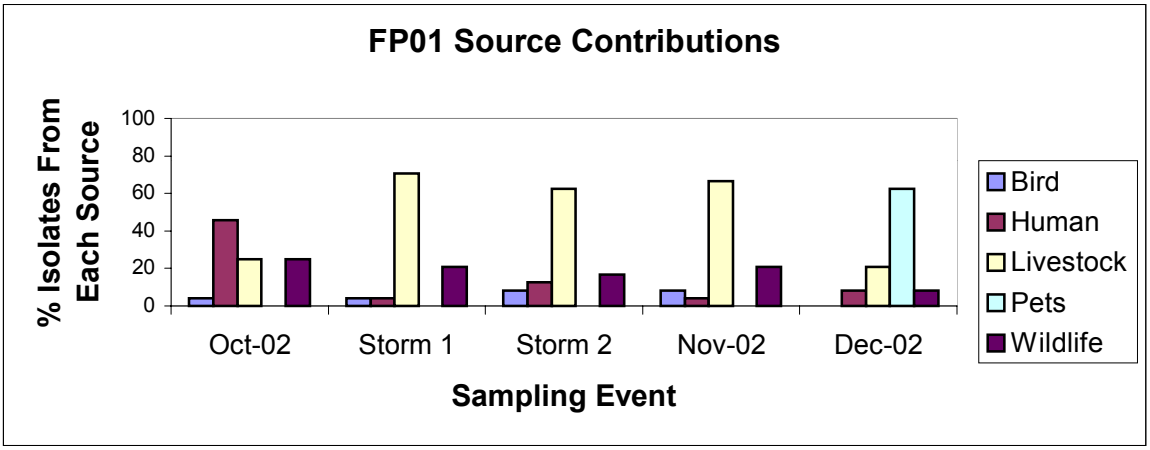


Figure 5: Source Contributions at Sampling Site FP02 in the Months Surrounding Storm Events

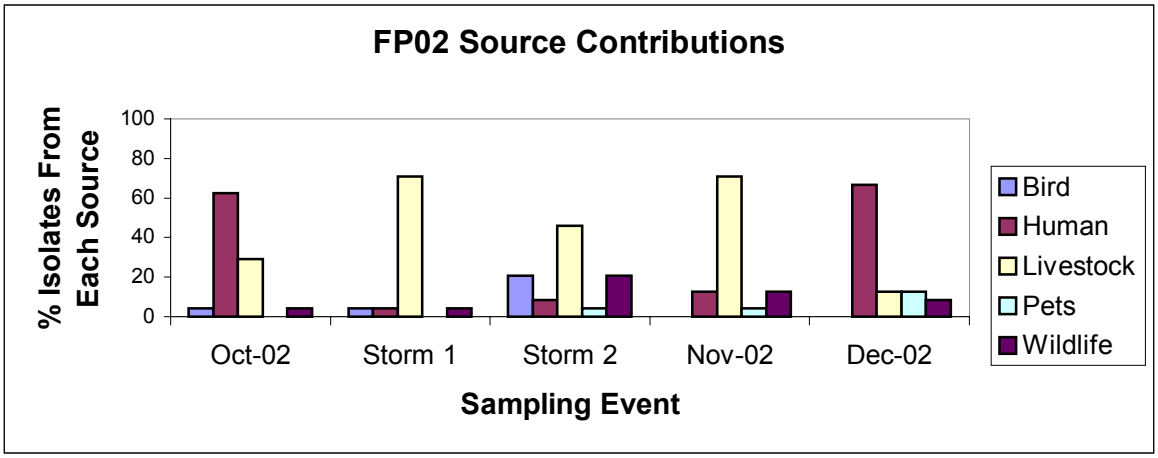


Figure 6: Source Contributions at Sampling Site FP03 in the Months Surrounding Storm Events

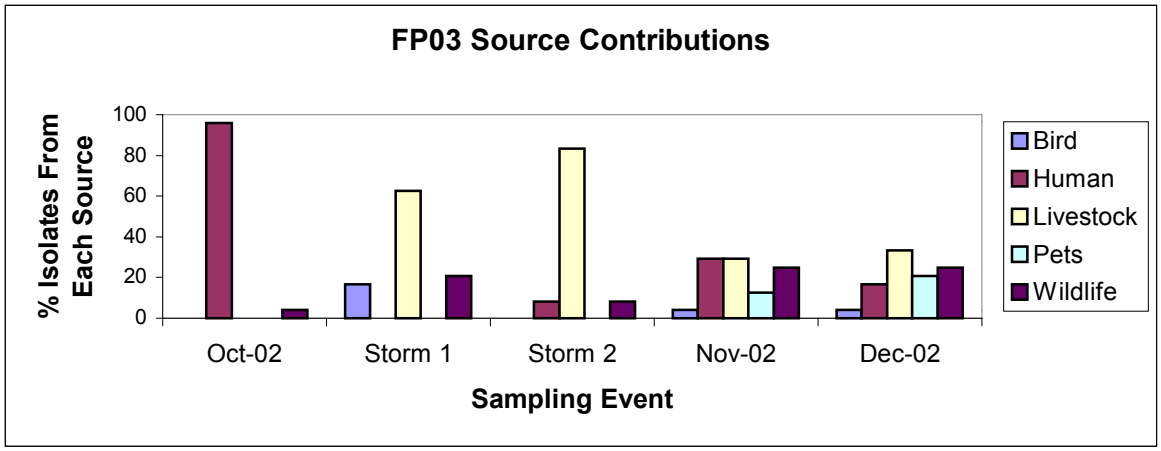


Table 2: DA Significant Differences of Mean Source Distributions on the Potomac

Station	*Bird*	*Human*	*Livestock*	*Pets*	*Wildlife*
FP01	13.1	13	36.6	22.1	15.3
FP02	7.8	17	33.8	12.9	28.5
FP03	7.2	19.9	26.3	12.1	34.6
Mean	9.3 C	16.6 B	32.2 A	15.7 BC	26.1 AB

*Common letters denote means that are not significantly different ($p>0.001$)

Anacostia River

Table 3: Significant Differences of Mean Source Distributions on the Anacostia

Station	*Bird*	*Human*	*Pets*	*Wildlife*
FA01	18.1	37.8	10	34.1
FA02	15	45	14.2	25.8
FA03	30.3	37.6	11.8	20.3
FA04	16.7	51.7	10.8	20.8
FA05	16.7	43.3	14.2	25.8
FA06	26.5	44.3	13	16.2
Mean	20.6 B	43.3 A	12.3 C	23.8 B

*Common letters denote means that were not statistically different (p>0.001)

Figure 7: Dry and Wet Season Source Contributions for site FA01

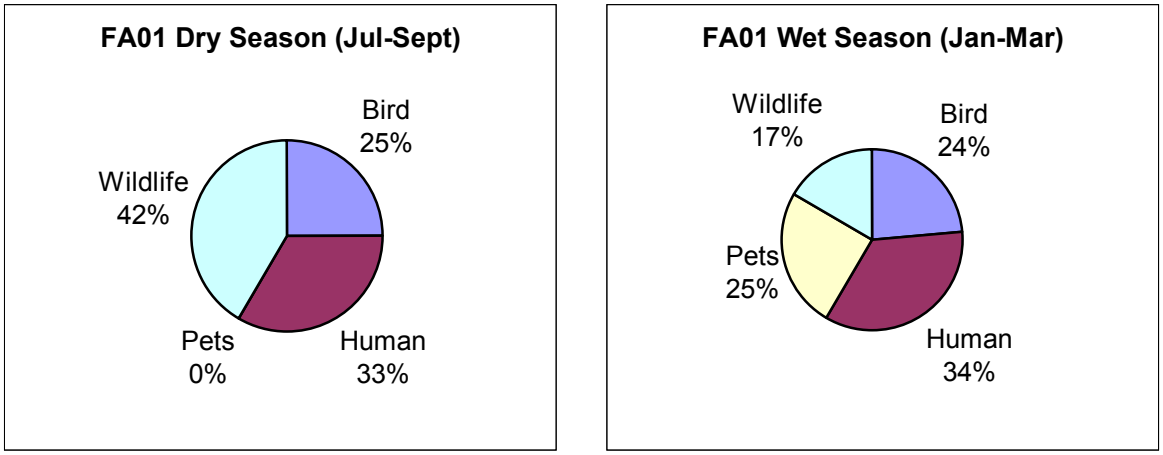


Figure 8: Dry and Wet Season Source Contributions for Site FA02

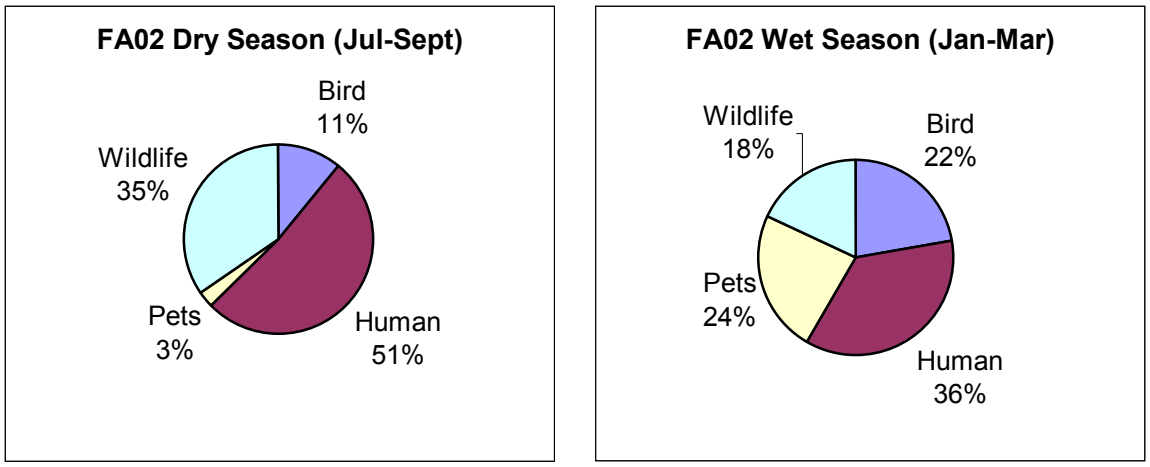


Figure 9: Dry and Wet Season Source Contributions for Site FA03

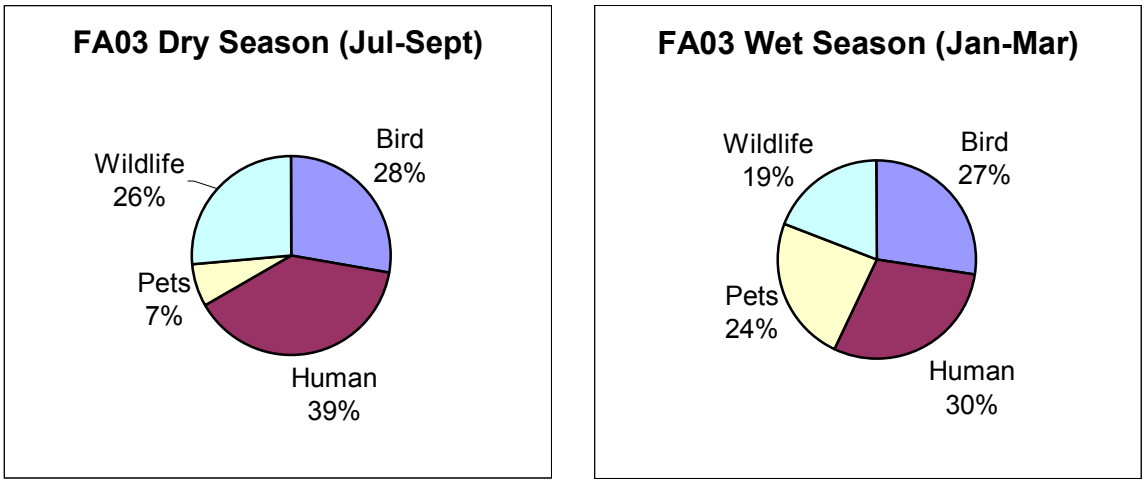


Figure 10: Dry and Wet Season Source Contributions for Site FA04

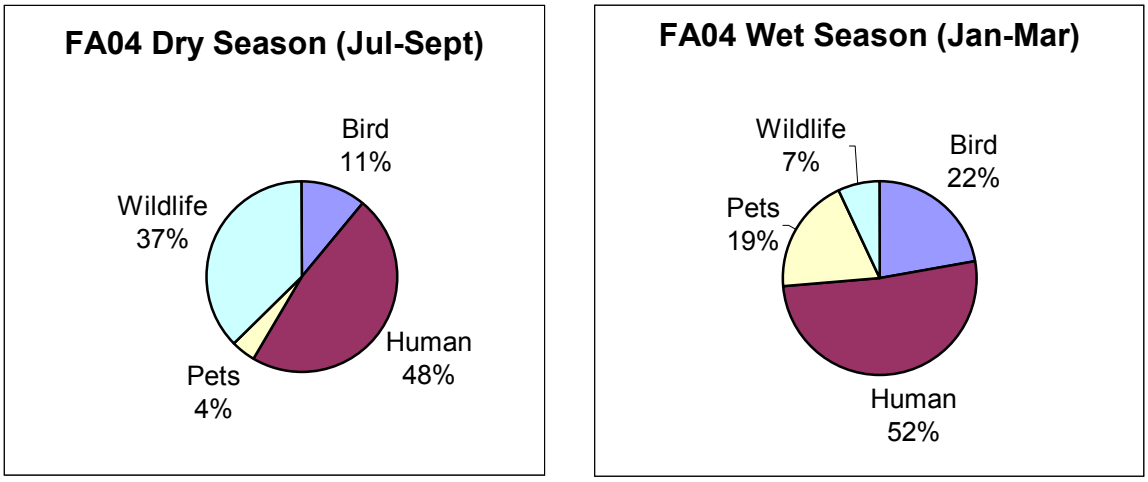


Figure 11: Dry and Wet Season Source Contributions for Site FA05

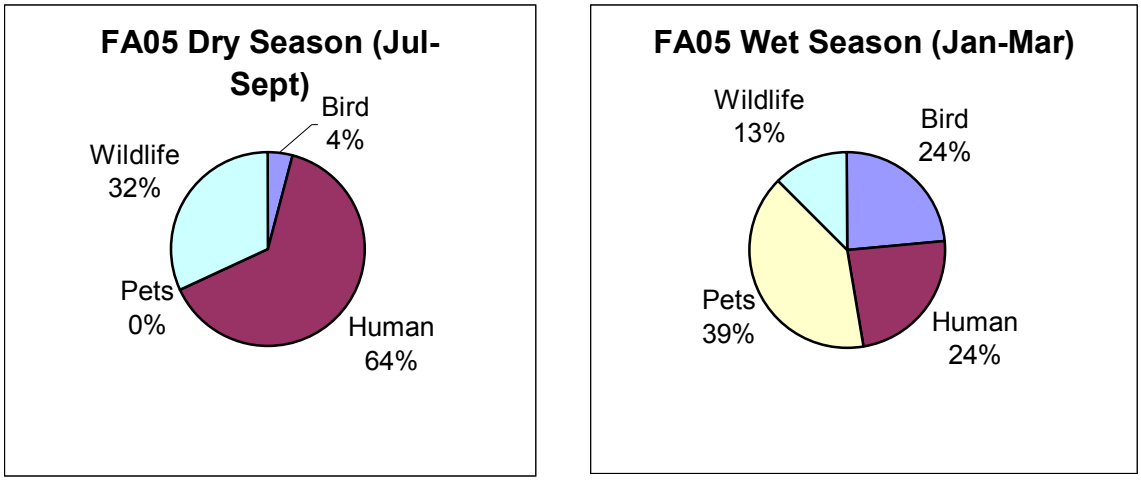


Figure 12: Dry and Wet Season Source Contributions for Site FA06

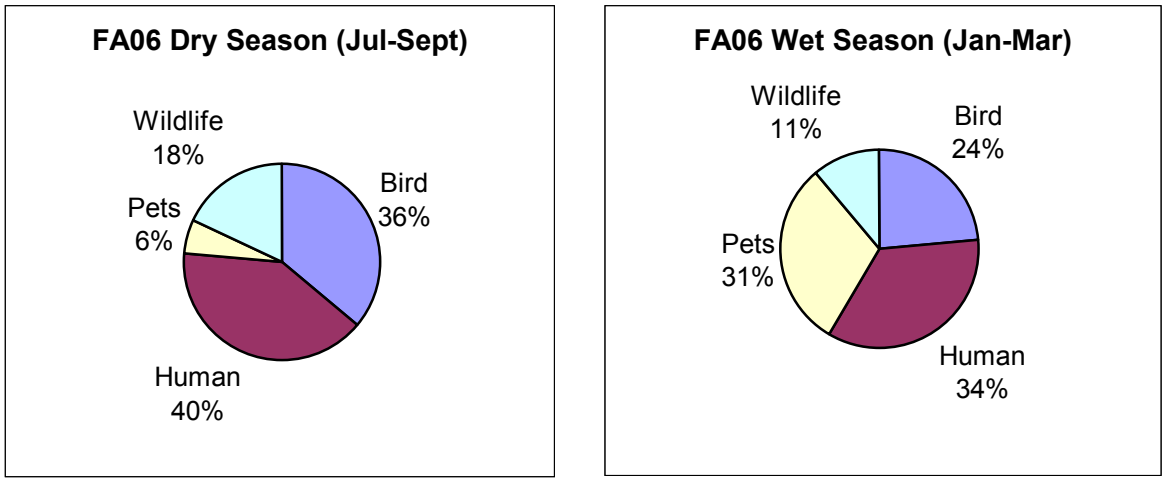


Figure 13: Source Contributions at Sampling Site FA01 in the Months Surrounding Storm Events

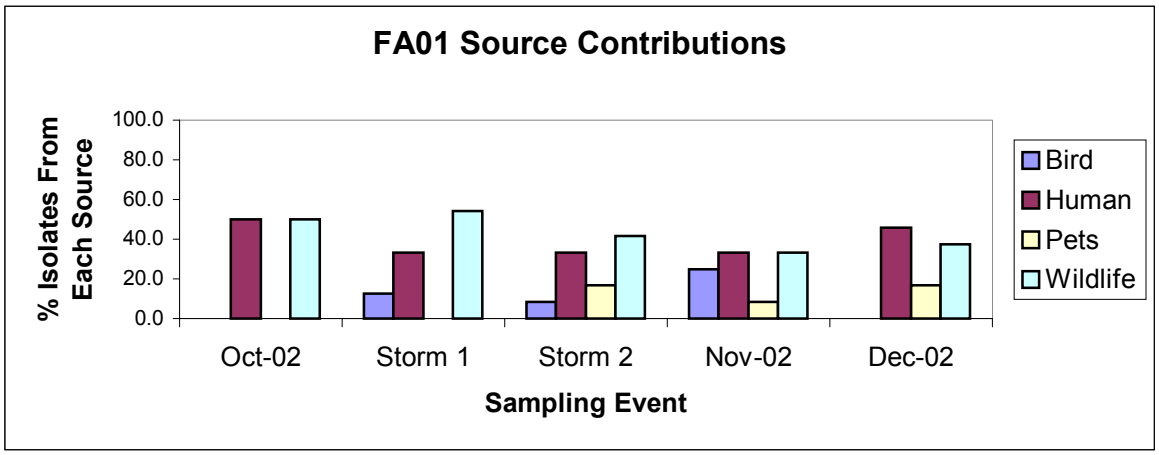


Figure 14: Source Contributions at Sampling Site FA02 in the Months Surrounding Storm Events

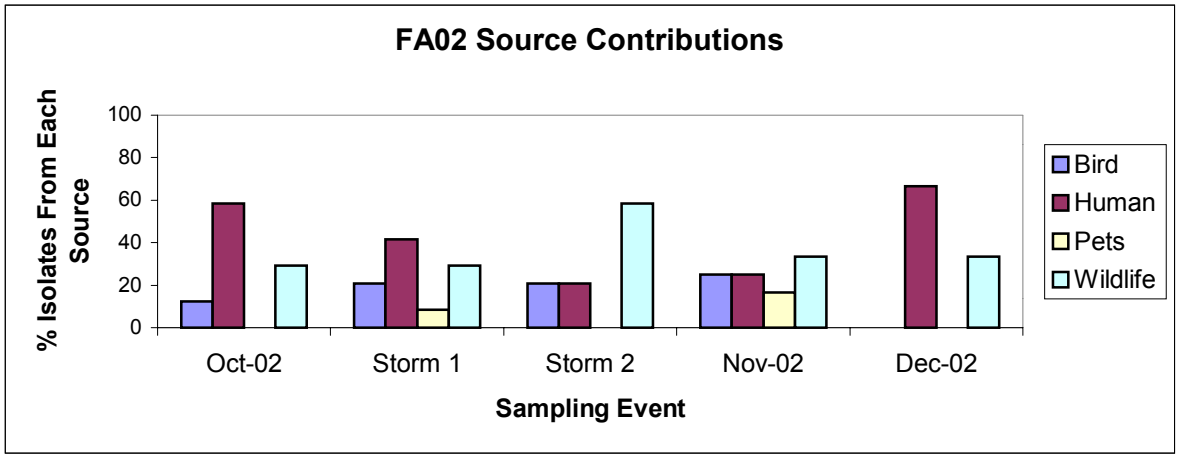


Figure 15: Source Contributions at Sampling Site FA03 in the Months Surrounding Storm Events

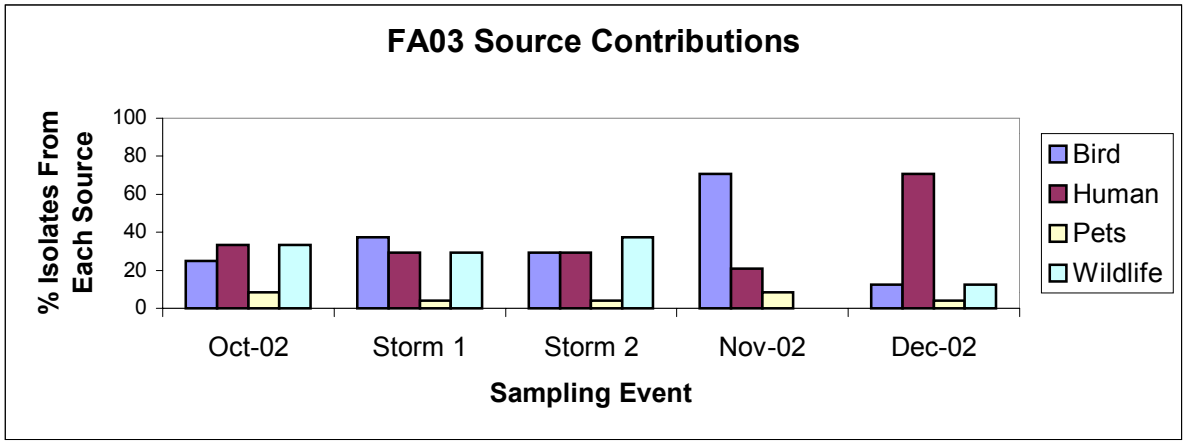


Figure 16: Source Contributions at Sampling Site FA04 in the Months Surrounding Storm Events

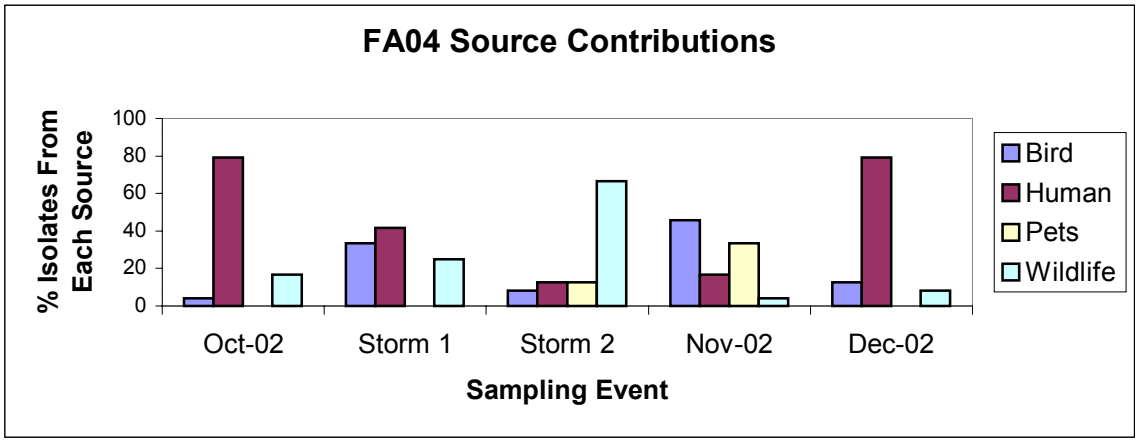


Figure 17: Source Contributions at Sampling Site FA05 in the Months Surrounding Storm Events

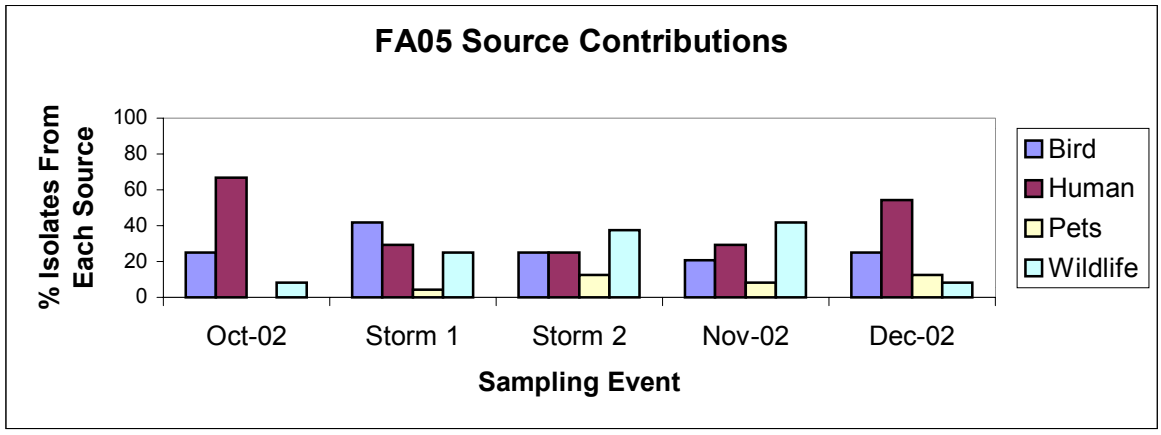


Figure 18: Source Contributions at Sampling Site FA06 in the Months Surrounding Storm Events

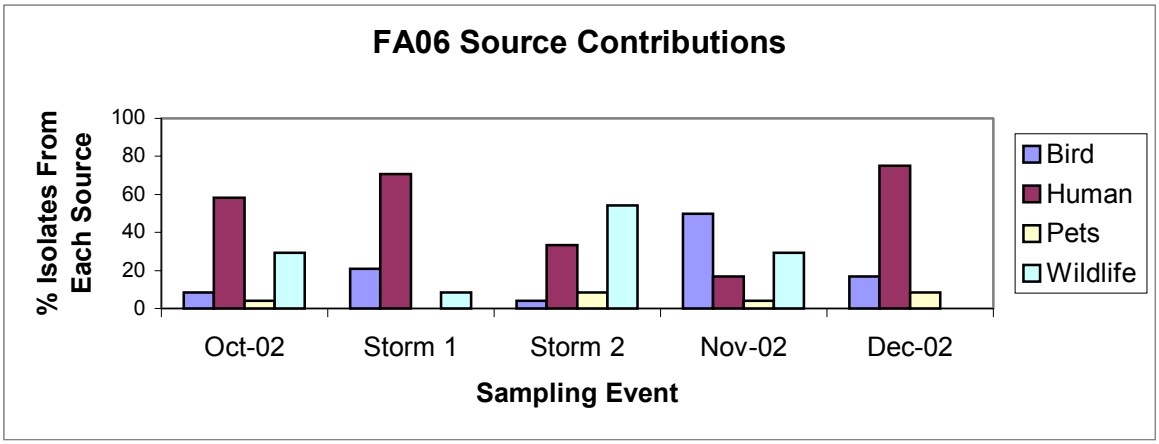


Table 4: DA Significant Differences of Mean Source Distributions on the Anacostia

Site	*Bird*	*Human*	*Pets*	*Wildlife*
FA01	20.7	30.1	12.1	37.2
FA02	11.1	30.9	24.1	33.9
FA03	22.6	28.4	20.4	28.6
FA04	15.8	34.8	20.6	28.8
FA05	18.6	29.5	21.7	30.2
FA06	16.7	30.1	30.6	22.5
Mean	17.6 B	30.6 A	21.6 B	30.2 A

*Common letters denote means that are not significantly different (p>0.001)

Rock Creek

Table 5: Significant Differences of Mean Source Distributions on Rock Creek

Station	*Bird*	*Horse*	*Human*	*Pets*	*Wildlife*
FR01	13.7	28.8	22	14	21.5
FR02	15.8	27.1	26.2	8.8	22.1
FR04	12.5	24	24.6	13.9	25
FR05	13.7	33	16.8	7.9	28.6
FR06	23.3	25.5	20.5	7.1	23.6
FR07	19.3	24.6	20	10.8	25.3
Mean	16.4 B	27.2 A	21.7AB	10.4 C	24.4 A

*Common letters denote means that are not significantly different (p>0.001)

Figure 19: Dry and Wet Season Source Contributions for Site FR01

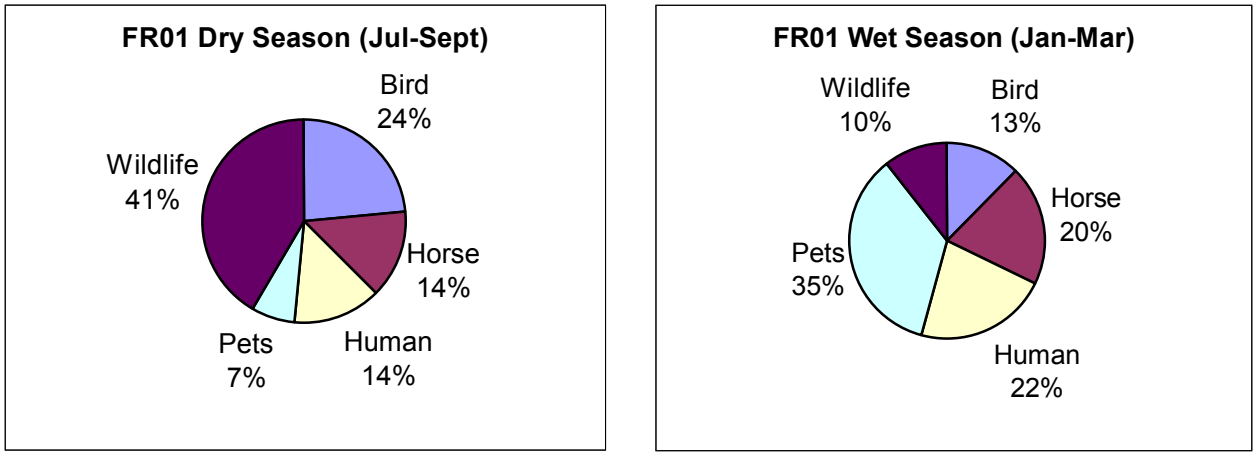


Figure 20: Dry and Wet Season Source Contributions for Site FR02

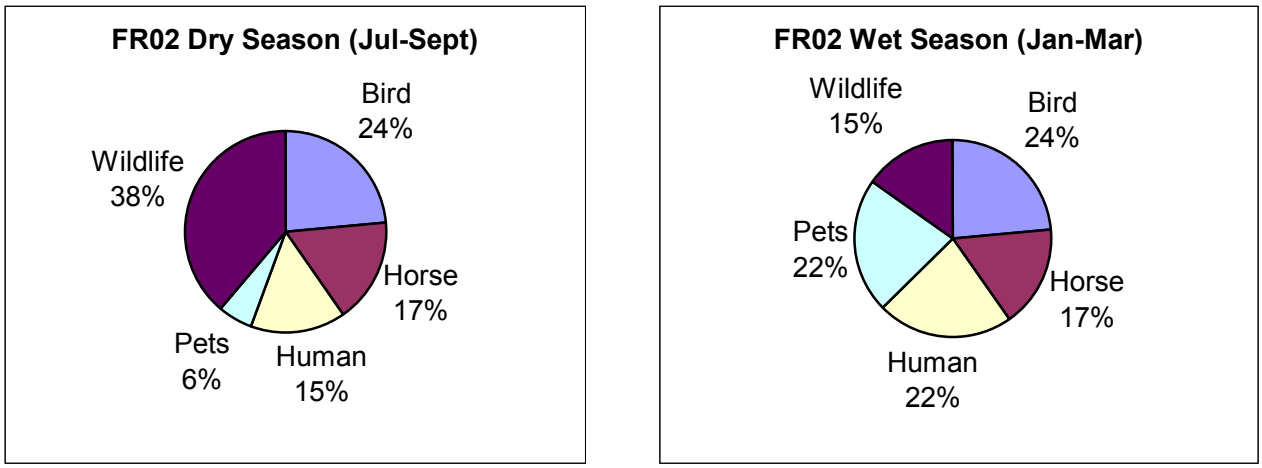


Figure 21: Dry and Wet Season Source Contributions for Site FR04

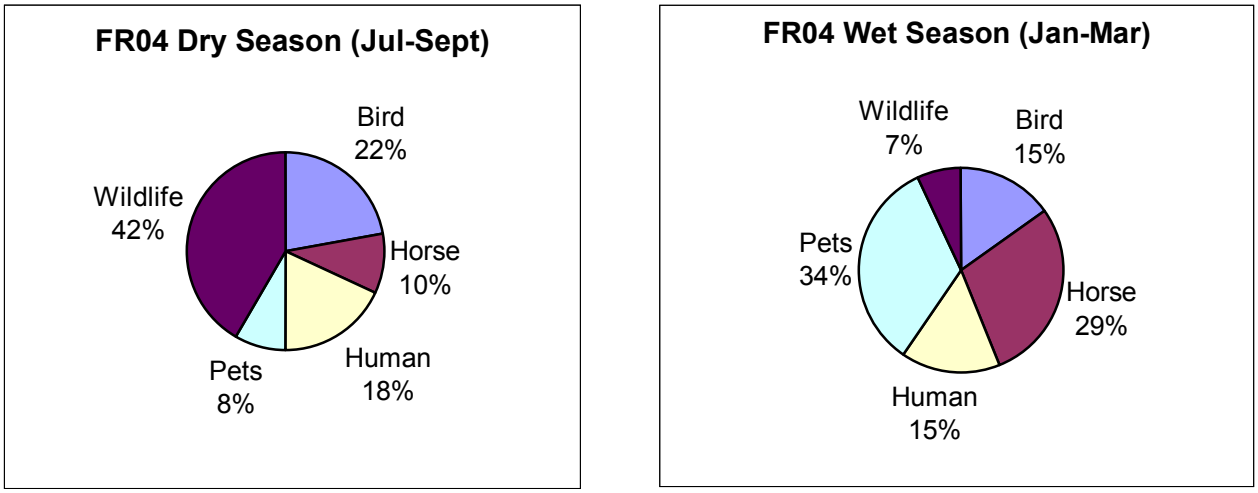


Figure 22: Dry and Wet Season Source Contributions for Site FR05

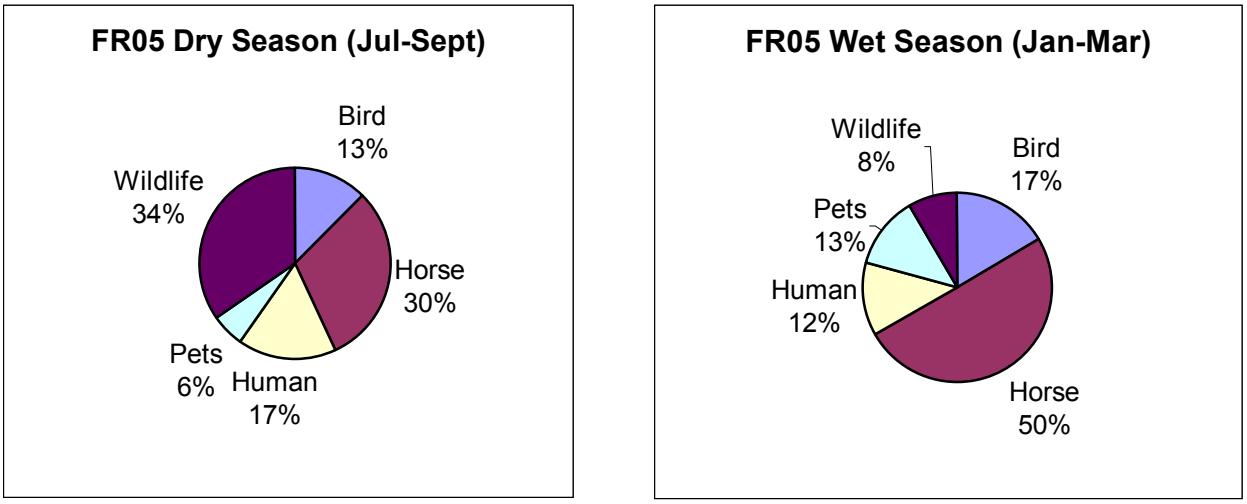


Figure 23: Dry and Wet Season Source Contributions for Site FR06

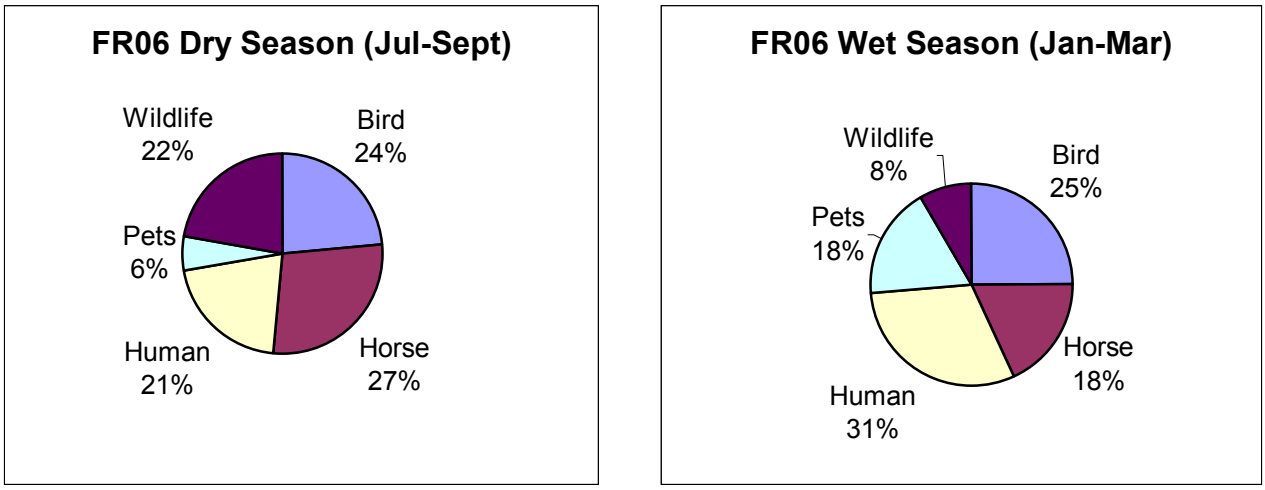


Figure 24: Dry and Wet Season Source Contributions for Site FR07

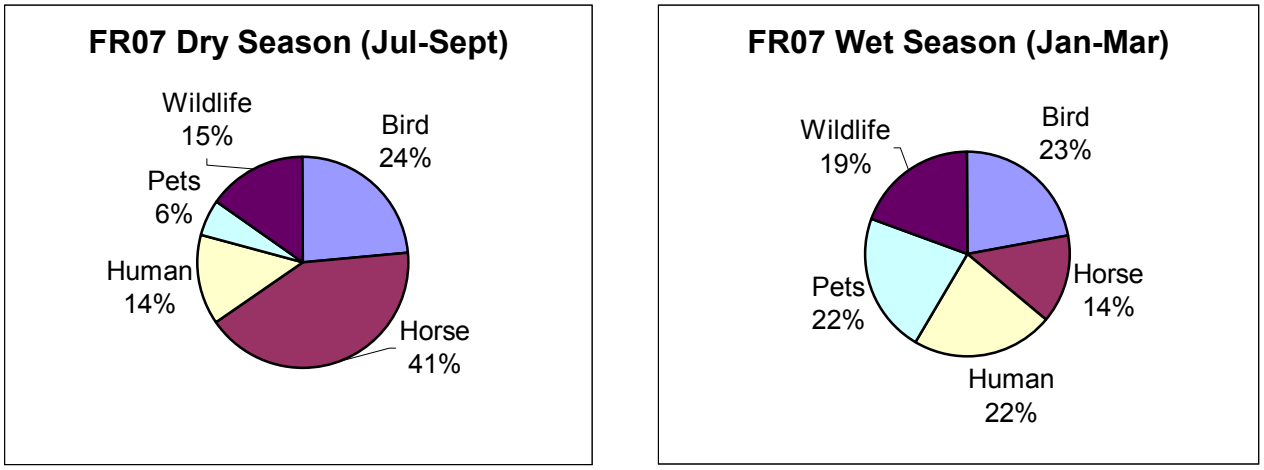


Figure 25: Source Contributions at Sampling Site FR01 in the Months Surrounding Storm Events

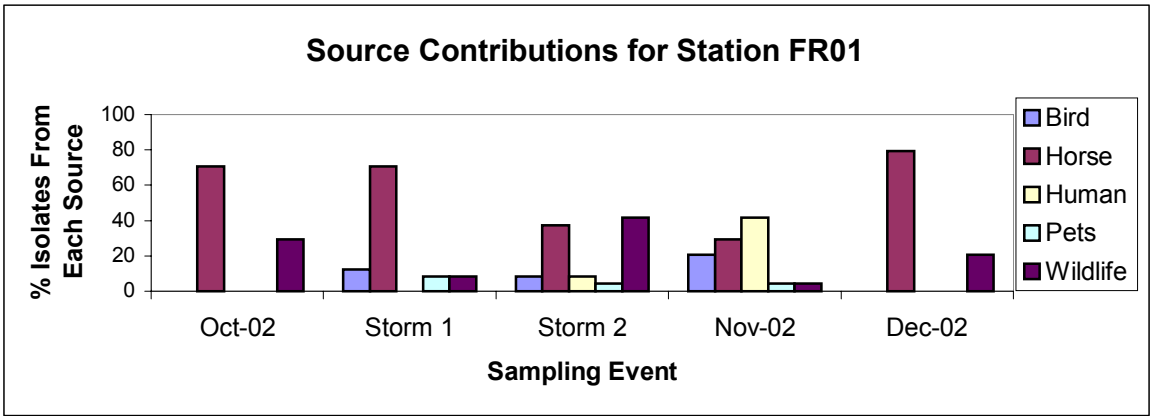


Figure 26: Source Contributions at Sampling Site FR02 in the Months Surrounding Storm Events

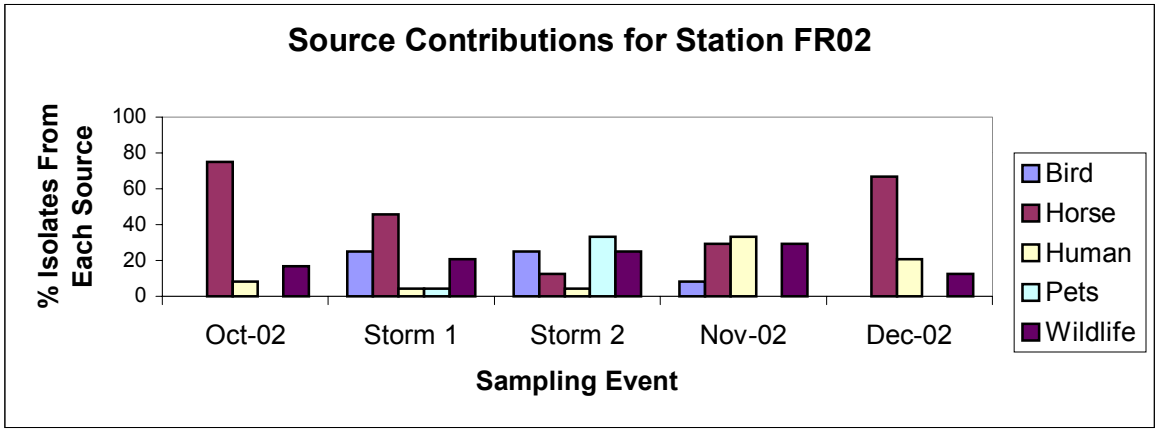


Figure 27: Source Contributions at Sampling Site FR04 in the Months Surrounding Storm Events

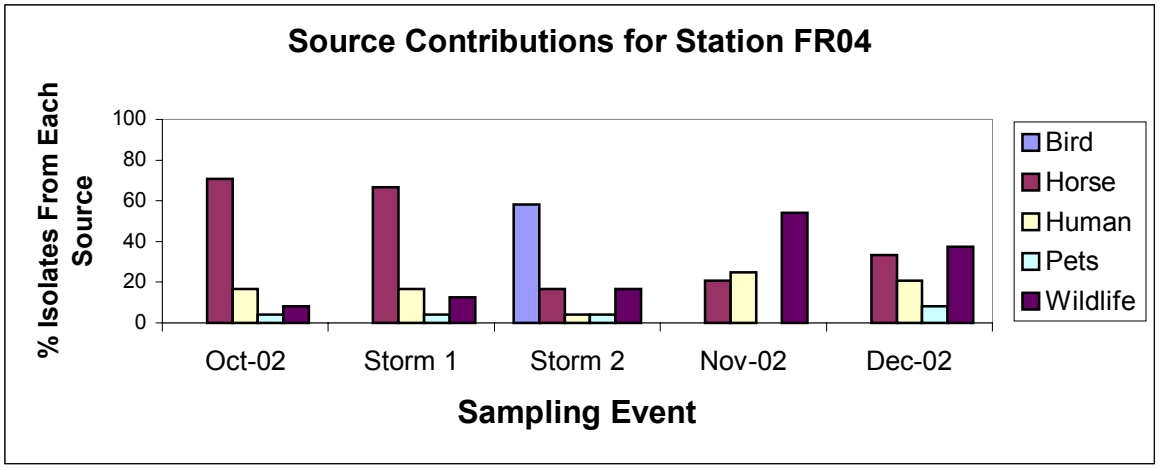


Figure 28: Source Contributions at Sampling Site FR05 in the Months Surrounding Storm Events

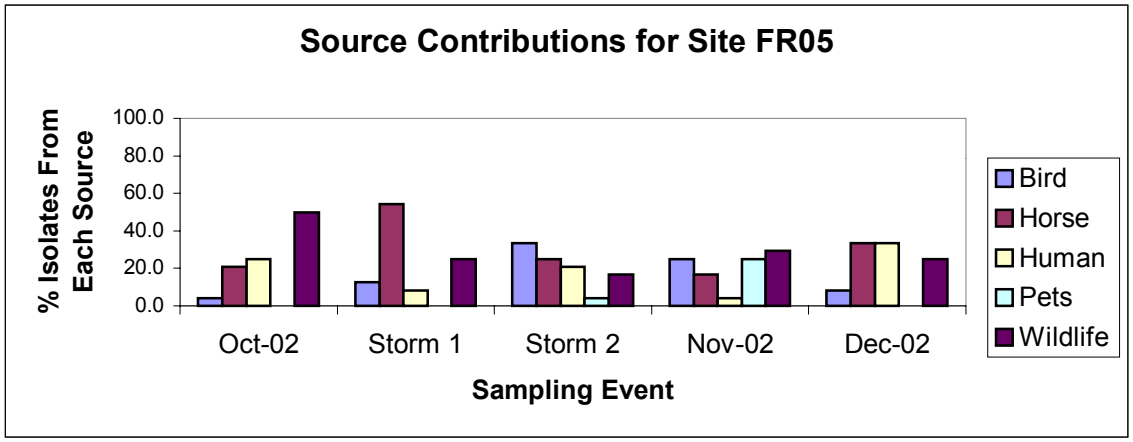


Figure 29: Source Contributions at Sampling Site FR06 in the Months Surrounding Storm Events

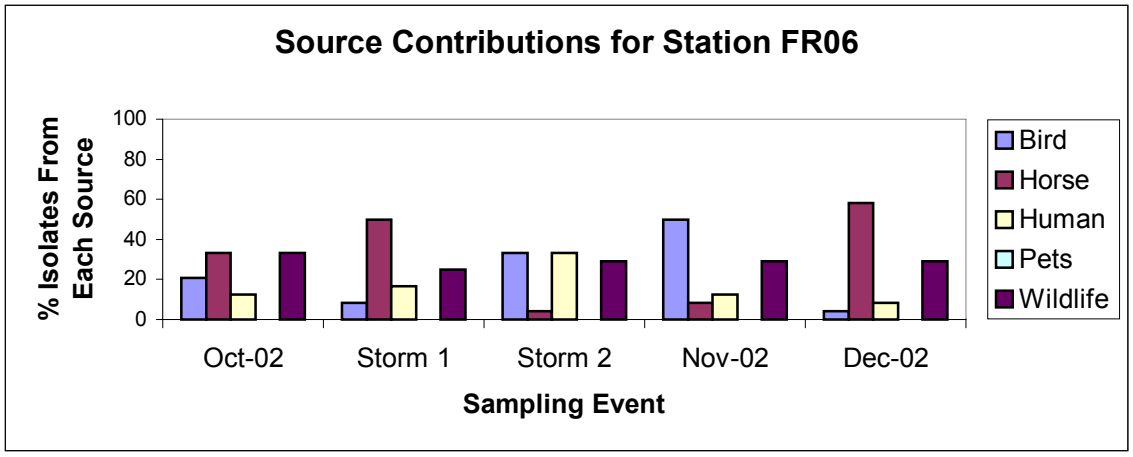


Figure 30: Source Contributions at Sampling Site FR07 in the Months Surrounding Storm Events

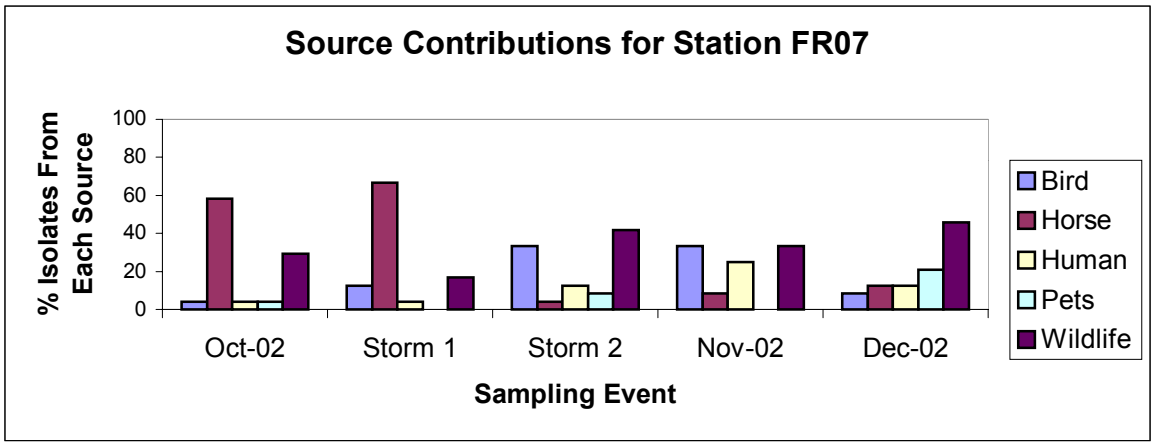


Table 6: DA Significant Differences of Mean Source Distributions on Rock Creek

Station	*Bird*	*Horse*	*Human*	*Pets*	*Wildlife*
FR01	13	35.2	14.6	17.7	19.5
FR02	19.2	30	16.3	12.9	21.7
FR04	19.2	29	15	16	20.8
FR05	12.5	38.3	10.4	14.6	24.2
FR06	22	25.1	14.4	12.7	25.8
FR07	19.6	27.9	15.2	18.1	19.3
Mean	17.6 C	30.9 A	14.3 C	15.3 C	21.9 B

*Common letters denote means that are not significantly different (p>0.001)

Vita

Kimberly R. Porter
kporter@vt.edu

EDUCATION

B.S., Environmental Science
December 2002
Virginia Tech, Blacksburg, VA
Concentration: Aquatic Resources
Minor: Chemistry
Cumulative GPA 3.91/4.0

M.S., Environmental Science and Engineering
December 2003
Virginia Tech, Blacksburg, VA
Cumulative GPA: 3.9/4.0

RELEVANT COURSEWORK

Undergraduate
Fundamentals of Environmental Science
Basic Soils/Lab
Soil and Groundwater Pollution
Physics of Pollution
Intro to Environmental Engineering
Groundwater Hydrology

Graduate
Principles of Environmental Engineering
Environmental Chemistry
Water and Wastewater Treatment and Design
Aquatic Ecotoxicology
Environmental Microbiology

EXPERIENCE

Virginia Polytechnic Institute and State University, Thesis Research
Blacksburg, VA
January 2003-present
Washington DC Area Urban Watershed Microbiological Study: *Enterococcus* Antibiotic Resistance Analysis.

- Utilized Antibiotic Resistance Analysis to determine sources of fecal pollution in Rock Creek, the Anacostia River, and the Potomac River in Washington D.C.
- Collected and analyzed the microbial quality of environmental samples
- Provided the Metropolitan Washington D.C. Council of Governments with information needed to make management decisions addressing the health risk assessments, and restoration efforts for issues related to fecal contamination
- Modeled database with statistical tools using JMP 5.0 software
- Created antibiotic stock solutions and media for various microbial analyses

Map Tech, Inc.
Blacksburg, VA
October 2003-present

- Worked on an EPA sponsored Small Business Innovation Research (SBIR) grant to establish the practical usefulness of Carbon Utilization Profiling for determining sources of fecal pollution

Research Intern, Worldwide Performance and Innovation
Blacksburg, VA
2001- 2002

- Assisted consultants with Brownfield's Program to determine contamination levels of abandoned industrial sites
- Collected samples from groundwater and soil contaminated with hazardous materials
- Organized team for sampling events
- Compiled list and contacted potential clients for Brownfield's Program
-

Research Assistant, Virginia Water Resources and Research Center
Blacksburg, VA
2000-2002

- Participated in a Total Maximum Daily Load (TMDL) study for receiving waters from trout farm effluent
- Aided in the creation and implementation of stream corridor assessment protocol for small watersheds
- Updated Access database with stream corridor assessment information
- Tracked location of environmental problem conditions using GPS system
- Analyzed water quality data, and reported on the condition of the Upper Powell River

Member, International Standards of Operations (ISO) 14001 Team, Town of Blacksburg
Blacksburg, VA
Spring 2002

- Analyzed and applied the ISO 14001 Standard for the municipality of Blacksburg
- Conducted Gap Analysis on Town of Blacksburg's Environmental Management System

TEACHING EXPERIENCE

Teaching Assistant, Environmental Microbiology, Virginia Tech
Dr. Charles Hagedorn, Professor
Spring 2003

- Facilitated the laboratory learning experience for senior and master's level students
- Emphasized the isolation and characterization of microorganisms from natural and engineered systems
- Addressed microbial processes in the environment, microbial communities, and microbial interactions

PUBLICATIONS

Porter, K., C. Hagedorn, A. Chapman. 2003. Sources of Fecal Pollution in Washington D.C. Waterways. *Proceedings 2003 American Society for Microbiology Conference*, Abstract, Washington D.C.

Porter, K. *Interpretation of Powell River Monitoring Data*. May 2001. Virginia Water Resources and Research Center Technical Report.

Porter, K., M. Stoughton, M. Waltham, S. Garman, and R. de Leon. June 2001. *Stroubles Creek Corridor Assessment Protocol.* Presented at the Virginia Service Training for Environmental Progress (STEP)Workshop.

Walker, J. L., K. R. Porter, and T. Younos. 2002. Monitoring Needs to Meet Benthic TMDL Requirements. *Proceedings of the 2002 National Monitoring Conference.* May 19-23, 2002. Madison, WI.

Walker, J., T. Younos, J. Anderson, and K. Porter. 2001. Challenges in Preparing TMDL Reports for Stream Segments Impaired by Trout Farm Effluent. *Proceedings Virginia Water Research Symposium 2001--Protecting Our Water Resources for the Next Generation: Where Do We Go From Here?* November 14-16, 2001. Charlottesville, VA. Virginia Water Resources Research Center Publication P7-2001. pp 81-83.

Walker, J. L., T. Younos, J. L. Anderson, and K. R. Porter. 2001. TMDL Development for Benthic Impairments from Trout Farm Effluent. *Total Maximum Daily Load Environmental Regulations Proceedings of the March 11-13, 2002 Conference.* ASAE Publication 701P0102. pp 449-453.

PROFESSIONAL MEMBERSHIPS

American Society for Microbiology, 2003

ACTIVITIES

Phi Eta Sigma National Honor Society, 2000-2002
University Honors Program, 1999-2002