

**UTILIZING DNA SEQUENCING AND LAND USE DATA FOR AN IMPROVED
UNDERSTANDING OF FECAL CONTAMINATION IN HUDSON RIVER
TRIBUTARIES**

A Final Report of the Tibor T. Polgar Fellowship Program

Elizabeth P. Farrell

Polgar Fellow

School for Earth and Environmental Sciences
CUNY Queens College
Flushing, NY 11367

Project Advisor:

Dr. Gregory D. O'Mullan
School for Earth and Environmental Sciences
CUNY Queens College
Flushing, NY 11367

Farrell, E. P. and G. D. O'Mullan. 2019. Utilizing DNA Sequencing and Land Use Data for an Improved Understanding of Fecal Contamination in Hudson River Tributaries, Section I: 1-25 pp. *In* S.H. Fernald, D.J. Yozzo, and H. Andreyko (eds.), Final Reports of the Tibor T. Polgar Fellowship Program, 2016. Hudson River Foundation.

ABSTRACT

Tributary mixing zones into the Hudson represent areas of both the highest frequency and magnitude of fecal indicator bacteria (FIB) contamination. The frequency and magnitude of contamination vary among tributaries and it is hypothesized that this indicates differing fecal sources. While EPA approved cultivation-based methods for the enumeration of FIB provide powerful tools for watershed monitoring, mitigation decisions require additional information related to the source of fecal bacteria. Data from cultivation based FIB and microbial community profiles based on high throughput DNA sequencing were analyzed in combination with land use patterns to better understand the sources of fecal contamination in six tributaries.

Land use patterns provided useful insights to begin understanding FIB patterns. On a watershed scale, forested areas were negatively correlated with FIB contamination, while developed areas had a positive correlation. The arrangement of site specific land use, ordered upstream to downstream, was often observed to influence the extent of FIB contamination. DNA sequencing data from a subset of sites was used to identify potential sewage and fecal contributions using a broader microbial community perspective for comparison to the patterns obtained from the commonly used, but taxonomically restricted cultivation-based fecal indicator, enterococci. The ratio of sewage to fecal microbial signatures varied among tributaries, possibly non-uniform sources, including fluctuating spatial contributions of human and animal sources. A diverse combination of monitoring tools should be developed and utilized, to provide complementary information toward improved differentiation of contamination sources and optimized mitigation actions for effective water quality management.

TABLE OF CONTENTS

Abstract.....	I-2
Table of Contents.....	I-3
Lists of Figures and Tables.....	I-4
Introduction.....	I-5
Methods.....	I-9
Fecal Indicators and Microbial Community Data.....	I-9
Land Cover Database and GIS Analysis of FIB Data.....	I-10
Results and Discussion	I-13
Conclusions.....	I-20
Acknowledgements.....	I-21
References.....	I-22

LIST OF FIGURES AND TABLES

Figure 1 – Sampling Sites Across Tributaries of the Hudson River.....	I-9
Table 1 – Simplification of Land Use Classifications	I-11
Figure 2 – GM ENT Levels by Tributary	I-13
Figure 3 – Percent Land Use by Tributary.....	I-13
Figure 4 – Percent Low to High Intensity Developed Land Use	I-14
Figure 5 – Percent Forested Land Use	I-14
Figure 6 – ENT, Land Use and DNA data in Sparkill: 3 Potential Zones of Impact	I-15
Figure 7 – ENT, Land Use and DNA data in Pocantico.....	I-16
Figure 8 – ENT, Land Use and DNA data in Wallkill.....	I-16
Figure 9 – ENT response to land use in Rondout (River Mile 0-28).....	I-17
Figure 10 – ENT response to land use in Rondout (River Mile 28-42).....	I-18
Figure 11 – ENT response to land use in Catskill	I-18

INTRODUCTION

Studying fecal contamination in the Hudson River tributaries is important for addressing potential public health risks, water resource management and habitat preservation. The Hudson's tributaries deliver water, nutrients, and sediment to the estuary while providing habitats for wildlife and for resident and migratory fish (NYS DEC 2015). Relatively little information is available regarding how land use in the watershed impacts the integrity and resiliency of the estuary. As the Hudson River Estuary Program action agenda endeavors to improve water quality by reducing pathogens (NYS DEC 2015), that connection between land use and the overall health of the estuary should be explored. Research that increases the understanding of important connections among land, tributaries, and the estuary can provide knowledge to better inform management actions that provide the greatest benefit to the health and resiliency of the estuary (NYS DEC 2015). Many Hudson tributaries contain very high levels of the Fecal Indicator Bacteria (FIB) enterococci (ENT) (Young et al. 2013; Suter et al. 2011), with 72% of citizen science samples (Riverkeeper 2015) exceeding the EPA's recommended Beach Action Value (BAV) of 60 ENT cells/100ml (USEPA 2012). Tributary mixing zones into the Hudson represent both the highest frequency (frequency of exceeding BAV) and magnitude (geometric mean by site) of FIB contamination, as compared to mid-channel, nearshore, and even wastewater treatment plant outfalls (Suter et al. 2011; Riverkeeper 2014). Understanding the sources of contamination in the tributaries is critical to deciphering patterns of fecal bacterial contamination of the Hudson River Estuary as a whole.

Fecal indicator bacteria (FIB), including ENT, are used to assess the combined extent of fecal contamination which can originate from numerous pollution sources, including human sewage, manure from livestock operations, wildlife, and urban runoff (Boehm et al. 2013). Ecosystems in developed areas possess a multitude of delivery mechanisms which often contain multiple fecal sources, making it extremely difficult to mitigate the pollution (Newton et al. 2013; O'Mullan et al. 2017). ENT measurements are not as useful when there is evidence of chronic contamination and sources need to be identified to address the problem (McLellan and Eren 2014). Human specific fecal pollution can originate from a variety of sources such as leaky or damaged sanitary sewer lines, faulty septic systems, illicit waste disposal, and sanitary/combined sewer overflows (Eaton et al. 2013). As a result, the characterization and management of human fecal pollution is closely linked with local waste management practices, adjacent land use, precipitation, and wet weather hydrology (Peed et al. 2011). Traditional culture-based methods, while commonly used to characterize fecal pollution, do not discern between human and other animal sources of fecal pollution. While FIB concentrations are essential tools for contamination assessment and the application of water quality regulations, management and mitigation efforts would benefit from the use of additional water quality research options.

Microorganisms that thrive within sewer systems may serve as useful adjuncts to fecal indicators for tracking sewage contamination because they could provide a signature of sewage pollution in surface waters (VandeWalle et al. 2012). These microbial sewage communities consist of a combination of human fecal microorganisms and non-fecal microorganisms which reside in the sewer infrastructure (Shanks et al.

2013). The advent of molecular methods allows for non-cultured organisms to be used as alternative fecal indicators (McLellan and Eren 2014), and these approaches have become popular and efficient methods for characterizing and tracking changes in the community structures of microbial populations (Bernhard and Field 2000). High throughput metagenomic DNA sequencing approaches can be used to evaluate the community signature from broader groups of fecal associated microbes. These broader groups of fecal associated microbes are what the traditional cultivation based FIB aim to indicate.

There are several factors that differ between the Hudson and its tributaries. One obvious dissimilarity is the scale of the bodies of water: in tributaries, a smaller volume of fecal or sewage input can have a larger spatial impact than in larger systems such as the Hudson, due to a reduction in the dilution of contaminants. Above the head of tide, the tributaries also have highly variable discharge rates and unidirectional flow, unlike the tributary mouths and main-stem Hudson, which are tidally influenced. These factors may lead to contamination remaining more localized in the Hudson, whereas in the tributaries the concentration of contaminants may be higher and contaminants may be transported larger distances downstream. In addition, sprawling development patterns can have negative consequences, including, but not limited to, the contribution of excess pollutants, nutrients and sediment to tributaries and the estuary (NYS DEC 2015).

Development can increase the amount of impervious surfaces and result in increases in stormwater flows (NYS DEC 2015). Impervious surface coverage has been positively correlated with fecal bacterial contamination in freshwater urban streams (Young and Thackson 1999) and tidal creek ecosystems (Mallin et al. 2000; Holland et al. 2004). A 1999 study showed that the concentration of certain pathogens, including ENT, were

directly related to the housing density, population, development, imperviousness, and apparent animal density (Selvakumar and Borst 2006). Surface runoff samples from more densely populated, sewerred areas regularly reflected higher bacterial counts when compared to runoff from less developed areas with septic tanks, which suggests that a relationship may exist between land use and potential bacterial loading (Young and Thackson 1999). The majority of larger rivers are influenced by hundreds of small streams draining from multiple watersheds which can make it particularly difficult to associate a specific land use scenario with poor water quality (Peed et al. 2011).

Prior monitoring activities have identified tributaries to be both hot spots of fecal contamination and an important determinant of water quality in the Hudson River itself. Ultimately, source identification and remediation of contamination rather than merely detection will protect public health and improve recreational opportunities afforded by our natural resources. In order to better understand the patterns and causes of tributary fecal contamination, the objectives of this study were to: 1) analyze existing culture based FIB data to identify spatial patterns in contamination; 2) analyze available DNA sequences from tributaries to investigate the spatial changes in the potential influence of “fecal” and “sewage infrastructure” microorganisms on tributary water quality; 3) assess land use patterns to determine if they can provide insight into patterns of FIB; and 4) examine the effectiveness of a combination of monitoring tools for improved information to differentiate between sources of contamination. As the frequency and magnitude of contamination was known to vary among tributaries, it was hypothesized that the sources of fecal contamination were also likely to differ among tributaries.

DNA was quantified and genes for 16S rRNA were amplified using bacterial primers 8F and 1492R, as described in O'Mullan et al. 2015. Amplified DNA was sent to Molecular Research DNA labs (www.mrdnalab.com, MRDNA, Shallowater, TX) for amplicon illumina sequencing. DNA sequence libraries were then used to estimate the percent representation of bacterial genera commonly found in fecal material and sewage infrastructure, (VandeWalle et al. 2012; Shanks et al. 2013; Newton et al. 2013) using bioinformatics analyses in the Quantitative Insights Into Microbial Ecology ver. 1.9.1(QIIME) software package (Caporaso et al. 2010). These quality control and data analysis steps, performed by lab member Roman Reichert, included removal of DNA barcodes, quality screening of sequences based on length and primer mismatches, De Novo chimera detection using USEARCH ver. 6.1 (Edgar 2010) in Qiime, and taxonomic classification relative to the SILVA 97% OTU database ver. 119 (Pruesse et al. 2007). The resulting data were used to calculate the percent representation (frequency relative to the total number of sequences) of both fecal and sewage infrastructure microbes from each sample.

Land Cover Database and GIS Analysis of FIB Data

The 2011 NLCD provides nationwide data on land cover and land cover change at the native 30-m spatial resolution of the Landsat Thematic Mapper (TM) (Homer 2015). Landsat 5 Thematic Mapper (TM) imagery provided the foundation for spectral change analysis, land cover classification, and imperviousness modeling for all NLCD 2011 products. All Landsat images were acquired from the USGS Earth Resources Observation and Science (EROS) Center Landsat archive, where they were radiometrically and geometrically calibrated (Homer 2015). The classification system

used by NLCD2011 is modified from the Anderson Land Cover Classification System, where detailed explanations of how the classification system was developed can be accessed (Anderson 1976).

GIS software was utilized to better observe trends in the spatial distribution of FIB data collected by citizen scientists in 2015 pertaining to the frequency and magnitude of contamination. Using GIS shapefiles, provided by Riverkeeper, the symbologies of each feature layer for each

watershed were modified in order to visually differentiate sample sites which had differing FIB geometric means. Specific intervals for the geometric mean were chosen according to relevant management guidelines. GIS software was then utilized to calculate percent land use from data obtained from the 2011 National Land Cover Database (NLCD) (MRLC 2016). New York contains 15 classes of land cover and for simplicity's sake, similar classes were combined when assigning land use categories (Table 1).

NLCD Land Cover Classification Legend		Combined Land Cover Classification
Open Water	→	Open Water
Developed, Open Space	→	Open Space Developed
Developed, Low Intensity	→	Low to High Intensity Developed
Developed, Medium Intensity		
Developed, High Intensity		
Barren Land (Rock/Sand/Clay)	→	Forested
Deciduous Forest		
Evergreen Forest		
Mixed Forest		
Woody Wetlands		
Emergent Herbaceous Wetlands		
Pasture/Hay	→	Agricultural
Cultivated Crops		

Table 1: Simplification of Land Use Classifications.

Initially, land use percentages were calculated on a watershed basis. This was achieved by using the “polygon to raster” tool in GIS and converting the polygon shape files of each individual watershed into a raster, maintaining the same grid size as the imported land use layer obtained from the NLDC. Next, the “zonal histogram” tool was used to create tables consisting of rows designated with each land use code and the number of cells within each category. Percentage of each land use was calculated using the outputs from the zonal histograms (number of cells in a land use classification compared to overall number of cells within the raster).

Next, land use was determined on a more “local” level. A 0.5 mile radius was created around each sampling site by using the “buffer” tool and creating a raster file, and again using the “polygon to raster” tool. Here, the GIS approach had complications for the intended analysis because some of the polygons overlapped where sampling sites were less than a half mile from one another. In order to obtain the most accurate information, overlapping polygons were identified and separate shape files were created to avoid overlap. Once this issue had been addressed, the “polygon to raster” tool was effectively utilized along with “zonal histograms” to create tables of different land use codes by site, along with the number of cells within each category. Here, many sites did not have one clear dominant land use and so categories were developed to differentiate between mixed land uses.

Five land use categories were developed: forested; forested/agricultural; forested/developed; forested/agricultural/developed; and developed. The criteria for a single dominant land use included having a minimum of 35% of total land use for that category, as well as being at least 10% higher in the dominant land use than any of the

other categories. For mixed use land categorizations, there had to be at least 20% of each of the included land uses.

RESULTS AND DISCUSSION

When the geometric mean (GM) of ENT was analyzed for each tributary, it was evident that contamination varies in both frequency and magnitude (Figure 2). The Wallkill, Pocantico and Sparkill showed the highest levels of contamination in both frequency and magnitude, while the Catskill, Esopus and Rondout were least contaminated. Similarly, within each tributary the frequency and magnitude of contamination varied among sample sites.

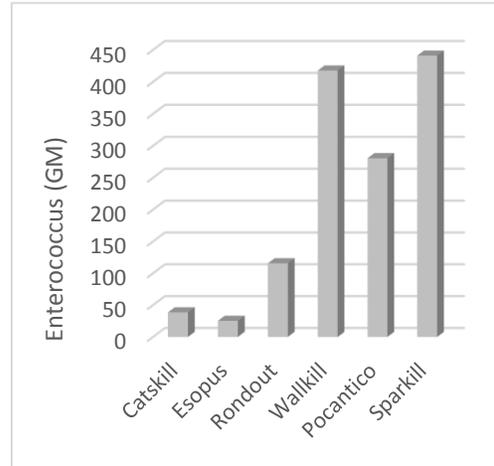


Figure 2: GM ENT Levels by Tributary.

Analyzing the metagenomic data for the representation of the fecal and sewage core, and the ratio of these two, provides some information about sources of contamination. The fecal core can represent either an animal or human fecal source, while the abundance of the sewage core is an indication of wastewater input. Therefore, when the ratio of sewage to fecal is high, it strongly suggests human fecal contamination, while a low ratio may indicate a non-human (or non-sewage) source of fecal contamination, such as wildlife, or manure from domesticated animals. Sparkill and

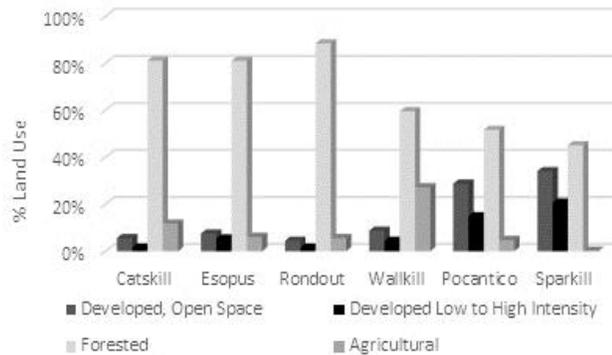


Figure 3: Percent Land Use by Tributary.

Pocantico both showed areas with high ratios of sewage to fecal core signatures. While in the Wallkill, the ratio of sewage to fecal core never rose above one, suggesting a lower relative signal of human fecal contamination.

Land use also exhibited differences amongst the tributaries (Figure 3). The Wallkill, Pocantico and Sparkill tributaries were less forested compared to the Catskill, Esopus and Rondout. The Wallkill had the highest percentage of agricultural land use. However, as was observed with GM of ENT, variation in land use also occurs by site within tributaries.

When considering both ENT and land use in unison, even on a large watershed scale, patterns were evident, including the correlations of forested and developed areas with the level of fecal contamination (Figures 4 and 5); however, to better

understand patterns within tributaries, land use and ENT levels were mapped site by site. Sources were more easily identified at the half mile radius surrounding each sampling site, and in order to utilize land use at this scale, upstream areas were evaluated for their impact on fecal contamination.

Examining site by site patterns within both Sparkill and Pocantico revealed increases in the ratio of sewage to fecal signatures as development increased. Sparkill had a lower ratio in forested areas, and interestingly, there was an increase in the ratio

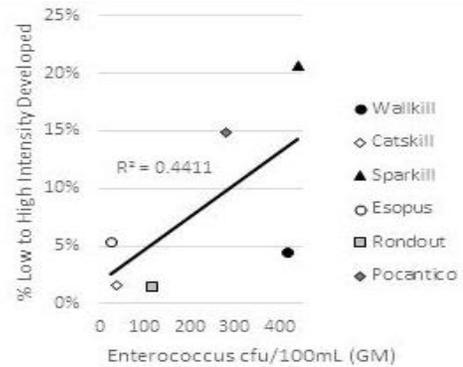


Figure 4: Percent Low to High Intensity Developed Land Use.

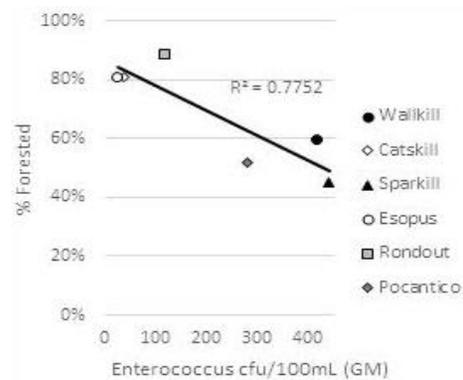


Figure 5: Percent Forested Land Use.

just upstream of the Orangetown waste water treatment plant (WWTP). In the Wallkill (forested, forested/agricultural, and forested/agricultural/developed), the ratio of sewage to fecal never rose above one across all sample sites.

When land use, DNA data, and ENT data were combined, it indicated the possibility of three zones in the Sparkill: a forested area reflecting low impact; a human wastewater input with increased development; and an input from potential non-human sources (Figure 6). In the Pocantico, densely developed areas with higher ENT levels showed a higher ratio of sewage to fecal core (Figure 7). The Wallkill exhibited a peak in the sewage to fecal ratio in a forested area which is located downstream of a developed area. In areas of forested/agricultural land use, there was no sewage signal present and while ENT values were lower, they still exceeded EPA guidelines for BAV. Downstream of areas with mixed forested/developed/agricultural land use, there were elevations in the sewage to fecal ratio (Figure 8).

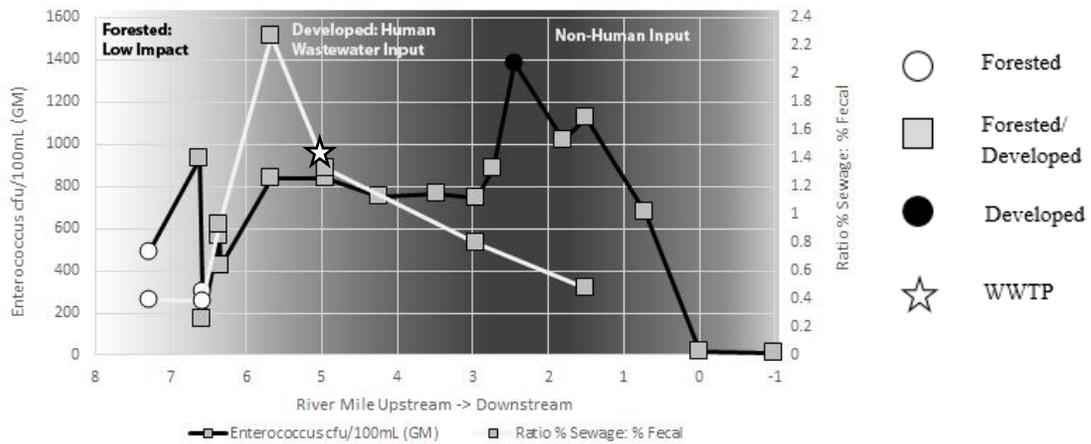


Figure 6: ENT, Land Use & DNA data in Sparkill: 3 Potential Zones of Impact.

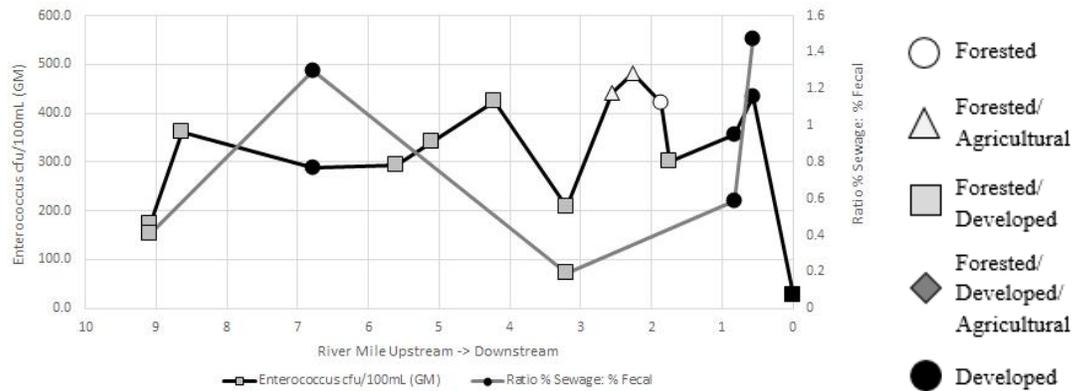


Figure 7: ENT, Land Use & DNA data in Pocantico.

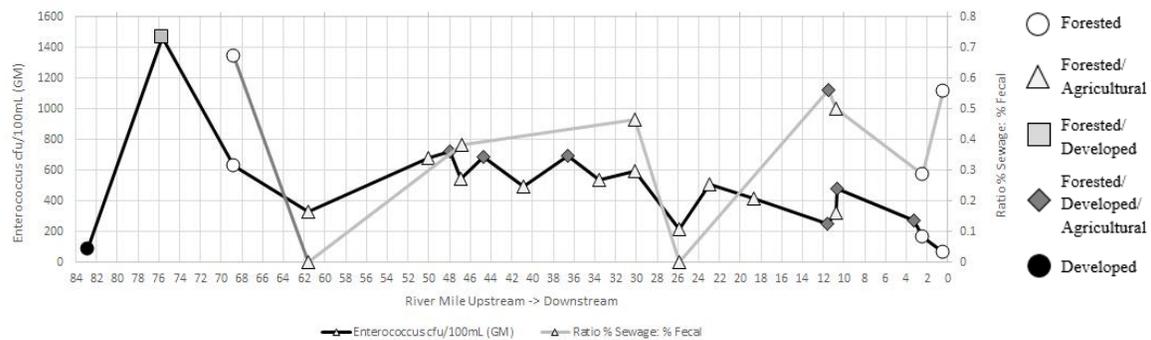


Figure 8: ENT, Land Use & DNA data in Walkkill.

While the culture based data provided by Riverkeeper supported the conclusion that tributaries have a fecal contamination problem, this widely used method for measuring fecal pollution does not differentiate the various possible sources of fecal contamination. This limitation makes it difficult to plan effective remediation efforts and, on its own, cannot specify whether fecal pollution originated from human waste management systems such as sewer lines and/or septic tanks, or other sources including local wildlife or livestock (Peed et al. 2011; Boehm et al. 2013). In order to supplement the information provided by ENT data, metagenomic sequencing can be a valuable tool for source tracking fecal contamination (McLellan and Eren 2014).

In past studies, land use correlations have indicated that the combination of an increase in urban development and subsequent intensified impervious surface coverage can lead to runoff that reaches surface waters with increased concentrations of FIB largely attributed to anthropogenic sources; therefore, their correlation with landscape characteristics confirms their effectiveness as indicators of urban pollution (Mallin et al. 2009). Population density, development age, and percent of residential development have also been shown to possibly be better at predicting levels of bacteria in urban stormwater runoff than factors such as rain intensity and antecedent dry period, among others (Glennie 1984; Chang 1999). Microorganism concentrations from high-density residential areas have been shown to be significantly higher than those associated with nearby low to moderate-density residential areas or landscaped commercial areas (Selvakumar and Borst 2006). Detailed land use characterization and the use of human-associated fecal source identification methods have allowed for the successful identification of septic systems as a key contributor of human fecal pollution (Peed et al. 2011); however, establishing a link between water quality and the adjacent landscape is often limited by sample site selection, spatial scale of catchment area, availability of associated runoff hydrology, and the accessibility of high-quality land use information.

In this study, when considering land use and ENT levels on a scale of a half mile radius around each sampling site, there was an association among adjacent sites where areas upstream subsequently

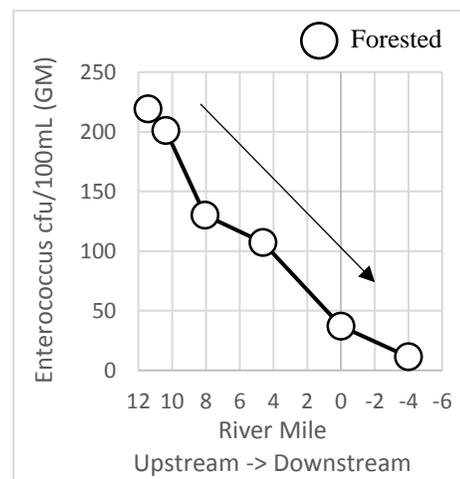


Figure 9: ENT response to land use in Rondout (River Mile 0-28).

affected ENT levels downstream. These patterns, in response to land use, included declines in ENT levels in areas of persistent forested land use (e.g. Figure 9). Alternatively, areas of transition from forested to agricultural or developed areas often showed increases in levels of ENT (Figure 10), and when land use shifted from agricultural or developed to forested, levels typically dropped (Figure 11). When sites are ordered upstream to downstream, a scale of impact from land use emerges and there is structure that appears to be very important. Areas where a forested land use site which was nestled between other land uses, behaved more like the surrounding land use than a forested area, indicating a more regional impact beyond the smaller scale of the half mile radius being considered (Figure 10). The magnitude and frequency of contamination fluctuates between tributaries, yet areas within each individual tributary showed similarities in the ENT response to land use to other tributaries despite their differences in overall land use and levels of contamination.

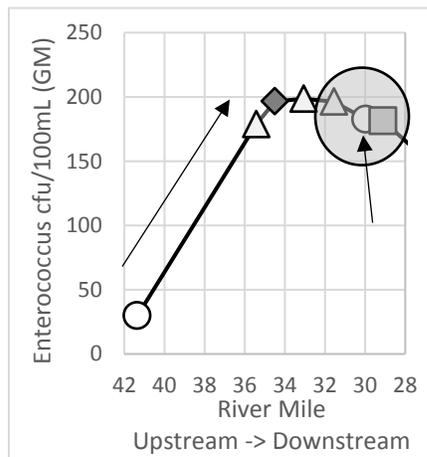


Figure 10: ENT response to land use in Rondout (River Mile 28-42).

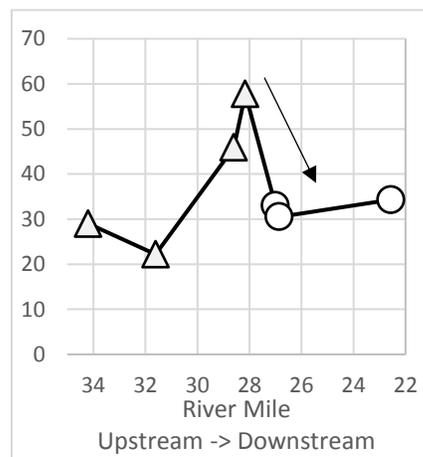


Figure 11: ENT response to land use in Catskill.

- Forested
- △ Forested/Agricultural
- ◻ Forested/Developed
- ◆ Forested/Developed/Agricultural

When DNA data was taken into consideration along with land use and ENT levels, the combination of information provided a clearer overall picture of potential

sources of contamination. The increase in ratio of sewage to fecal signal upstream of a WWTP in the Sparkill (Figure 6) suggested that the increase was not in response to an unpermitted discharge from the WWTP, but instead it is more likely that there is a different input causing the increase in ratio of sewage to fecal, such as a leaking sewage line or septic source. Further downstream in the Sparkill, the decrease in ratio of sewage to fecal suggests a greater non-human input such as wildlife near the duck pond. In the Pocantico, developed areas had increased ratios of sewage to fecal signal, suggesting potential human wastewater input associated with increased development. As there are no WWTP in the Pocantico watershed, this increase in the ratio of sewage to fecal signal is likely due to leaking sewage infrastructure or illicit discharges. In the Wallkill, the elevation in sewage to fecal ratio in areas of mixed forested/developed/agricultural land use areas indicated a relationship between increases in the level of development and an increase in the sewage signal. The behavior of the sewage to fecal ratio in the Wallkill suggested that, while there were fluctuations in the ratio, the potential for sewage to deliver contamination to the system is lower than the other two tributaries, which insinuates the potential of a non-human fecal input. This hypothesis could possibly be supported by the land use data, considering that the Wallkill has the highest percentage of agricultural land use compared to all other tributaries (Figure 3). In addition, neither the Sparkill nor the Pocantico had 0% sewage signal however, there was 0% sewage signature detected at two of the sample sites in the Wallkill. A major difference between the Wallkill and the other two tributaries is land use. Despite the fact that the Wallkill has similar levels of ENT as the Sparkill and the Pocantico, it is not as developed and has more agricultural land use, indicating increased importance of non-human sources.

CONCLUSIONS

Tributaries are hot spots of ENT contamination and vary in regard to frequency and magnitude of contamination. ENT data has proven to be a very valuable tool, but cannot directly provide information as to its source and it can be difficult to identify the best management actions to reduce fecal contamination. Land use patterns provide useful information to begin understanding ENT patterns, and potential interactions can be observed at both watershed and single site scales. There is evidence for regional impact of land use which can be observed when comparing ENT levels and percent land use (Figures 4 and 5). Models are increasingly connecting water quality to land use types and benefitting from remotely gathered data and GIS-based data handling (Selvakumar and Borst 2006). This approach requires an understanding of the concentration and load from a given area based on land use (Selvakumar and Borst 2006); however, land use data alone doesn't provide enough information due to the variability in frequency and magnitude of contamination among tributaries.

DNA sequencing helps to constrain sewage and fecal contributions within watersheds and is important because, unlike ENT, it is not limited to a single indicator. Metagenomic sequencing can be used in the future to look at the community signatures, however, sequencing is not fully quantitative and, therefore, on its own is not as informative as it can be when combined with ENT data. Additional DNA tools, such as quantitative PCR (Bernhard and Field 2000) (Chern et al. 2009; Wade et al. 2006) (Gentry et al. 2007; Noble et al. 2006) could help to further differentiate fecal sources.

A combination of monitoring approaches will be very useful in differentiating fecal sources. Culture based FIB, GIS analysis of land use, and DNA sequencing appear

to work well in conjunction with one another to better constrain sources of contamination in the tributaries. Understanding the sources of contamination is important for effective water quality management as it can help to identify types of mitigation actions needed in tributaries. Improved management of the tributaries will, in turn, have a positive impact on tributary habitat as well as Hudson River water quality management.

ACKNOWLEDGEMENTS

We would like to acknowledge David Yozzo, Sarah Fernald, and Helena Andreyko from the Hudson River Foundation for their guidance, Dan Shapley and Jen Epstein from Riverkeeper for their support of the project and the contributions of the following individuals: A. Montero, R. Reichert, and D. Mondal.

REFERENCES

- Anderson, J. R. 1976. A land use and land cover classification system for use with remote sensor data. USGS Circular: 671.
- Bernhard, A., and K. 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.
- Boehm, A. B., L. C. Van De Werfhorst, J. F. Griffith, P.A. Holden, J. A. Jay, O. C. Shanks, and S. B. Weisberg. 2013. Performance of forty-one microbial source tracking methods, a twenty-seven lab evaluation study. *Water Research* 47(18): 6812-6828.
- Caporaso, J., J. Kuczynski, J. Stombaugh, K. Bittinger, F. Bushman, E. Costello, and G. Huttley. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*: 335-336.
- Chang, G. 1999. personal communication. (T. R. Schueler, Interviewer) Austin TX Environmental and Conservation Dept. Austin.
- Chern, E., K. Brenner, L. Wymer, and R. A. Haughland. 2009. Comparison of fecal indicator bacteria densities in marine recreational waters by QPCR. *Water Quality, Exposure and Health*: 203-214.
- Eaton, T., G. D. O'Mullan, and A. A. Rouff. 2013. Assessing continuous contamination discharge from a combined sewer outfall (cso) into a tidal wetland creek: bacteriological and heavy metals indicators. *Annals of Environmental Science* 7: 79-92.
- Edgar, R. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*: 2460-2461.
- Environmental Protection Agency. 2012. 2012 Recreational water quality criteria. doi:EPA-HQ-OW-2011-0466.
- Gentry, R., A. Layton, L. McKay, J. McCarthy, D. Williams, S. Koirala, and G. Sayler. 2007. Efficacy of bacteroides measurements for reducing the statistical uncertainty associated with hydrologic flow and fecal loads in a mixed use watershed. *Journal of Environmental Quality*: 1324-1330.
- Glenne, B. 1984. Simulation of water pollution generation and abatement on suburban watersheds. *Journal of American Water Resources*: 211-217.
- Holland, A. F., D. M. Sanger, C. P. Gawle, S. B. Lerberg, M. S. Santiago, G. H. Riekerk, L. E. Zimmerman, G. I. Scott. 2004. Linkages between tidal creek ecosystems and the landscape and demographic attributes of their watersheds. *Journal of Experimental Marine Biology and Ecology*: 151-178.

- Homer, C. D. 2015. Completion of the 2011 national land cover database for the conterminous united states-representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81-5: 345-354.
- Idexx. 2016. "Enterolert," Idexx Laboratories.
<https://www.idexx.com/water/products/enterolert.html> (accessed June 3, 2016.)
- Mallin, M. A., V. L. Johnson, and S. H. Ensign. 2009. Comparative impacts of stormwater runoff on water quality of an urban, a suburban, and a rural stream. *Environmental Monitoring and Assessment*: 475-491.
- Mallin, M. A., K. E. Williams, E. C. Esham, and R. P. Lowe. 2000. Effect of human development on bacteriological water quality in coastal watersheds. *Ecological Applications*: 1047-1056.
- McLellan, S., and A. M. Eren. 2014. Discovering new indicators of fecal pollution. *Trends in Microbiology* 22(12): 697-706.
- MO BIO Laboratories. 2016. "PowerWater DNA Isolation Kit. Retrieved from MO BIO Laboratories," MO BIO Laboratories.
<https://mobio.com/media/wysiwyg/pdfs/protocols/14900-S.pdf> (accessed February 12, 2016).
- Newton, R., M. Bootsman, H. Morrison, M. Sogin, and S. McLellan. 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of lake michigan. *Microbial Ecology* 65(4): 1011-1023.
- Noble, R., J. Griffith, D. Blackwood, J. Fuhrman, J. Gregory, X. Hernandez, X. Liang, A. Bera, K. Schiff. 2006. Multitiered approach using quantitative PCR to track sources of fecal pollution affecting Santa Monica Bay, California. *Applied and Environmental Microbiology*: 1604-1612.
- NYS DEC. 2015. "The State of the Hudson 2015," NYS DEC.
http://www.dec.ny.gov/docs/remediation_hudson_pdf/hresoh15all.pdf (accessed Jan 15, 2015).
- NYS DEC. 2015. "Hudson River Estuary Action Agenda 2015-2020," NYS DEC.
http://www.dec.ny.gov/docs/remediation_hudson_pdf/dhreaa15.pdf (accessed Jan 15, 2015).
- O'Mullan, G., M. E. Dueker, K. Clauson, Q. Yang, K. Umemoto, N. Zakharova, J. Matter, S. Martin, T. Takahashi, and D. Goldberg. 2015. Microbial stimulation and succession following a test well injection simulating CO₂ leakage into a shallow Newark basin aquifer. *PloS ONE* 10(1): e0117812.
- O'Mullan, G., M. E. Dueker, and A. R. Juhl. 2017. Challenges to managing microbial fecal pollution in coastal environments: extra-enteric ecology and microbial exchange among water, sediment, and air. *Current Pollution Reports* 3(1): 1-16.

- Peed, L. A., C. T. Nietch, C. A. Kelty, M. Meckles, T. Mooney, M. Sivaganesan, and O. C. Shanks. 2011. Combining land use information and small stream sampling with pcr-based methods for better characterization of diffuse sources of human fecal pollution. *Environmental Science and Technology*: 5652-5659.
- Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. Glockner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*: 7188-7196.
- Riverkeeper. 2014. "Quality assurance project plan: citizen science water quality testing program," Riverkeeper. https://www.riverkeeper.org/wp-content/uploads/2009/06/Riverkeeper-Citizen-Science-Water-Quality-QAPP_2014.pdf (accessed January 14, 2016).
- Riverkeeper. 2015. "How's the water 2015," Riverkeeper. http://www.riverkeeper.org/wp-content/uploads/2015/06/Riverkeeper_WQReport_2015_Final.pdf (accessed January 15, 2016).
- Riverkeeper. 2015. "Water quality program," Riverkeeper. <http://www.riverkeeper.org/water-quality/testing/> (February 1, 2016).
- Selvakumar, A., and M. Borst. 2006. Variation of microorganism concentrations in urban stormwater runoff with land use and seasons. *Journal of Water and Health*: 109-124.
- Shanks, O., R. Newton, C. Kelty, S. Huse, M. Sogin, and S. McLellan. 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Applied and Environmental Microbiology* 79(9): 2906-2913.
- Suter, E., A. Juhl, and G. O'Mullan. 2011. Particle association of *Enterococcus* and total bacteria in the lower Hudson River Estuary, USA. *Journal of Water Resource and Protection*: 715-725.
- MRLC. 2011. "Multi resolution land characteristics consortium. retrieved from national land cover database 2011," MRLC. http://www.mrlc.gov/nlcd11_data.php (accessed May 6, 2016).
- VandeWalle, J., G. Goetz, S. Huse, H. Morrison, M. Sogin, R. Hoffman, K. Yan, and S. McLellan. 2012. *Acinetobacter*, *Aeromonas* and *Trichococcus* populations dominate the microbial community within urban sewer infrastructure. *Environmental Microbiology* 14(9): 2538-2552.
- Wade, T., R. Calderon, E. Sams, M. Beach, K. Brenner, A. Williams, and A. Dufour. 2006. Rapidly measured indicators of recreational water quality and swimming-associated illness. *Environmental health perspectives*: 24-28.

- Young, K. D., and E. L. Thackson. 1999. Housing density and bacterial loading in urban streams. *Journal of Environmental Engineering*: 1177-1180.
- Young, S., A. Juhl, and G. O'Mullan. 2013. Antibiotic resistant bacteria in the Hudson River Estuary linked to wet weather sewage contamination. *Journal of Water and Health* 11(2): 297-310.