Analysis of Fecal Coliform Data from Onondaga Lake Tributaries

prepared for

Upstate Freshwater Institute

and

Onondaga County Department of Water Environment Protection

by

William W. Walker, Jr., Ph.D.

and

Jeffrey D. Walker, M.Eng.

Environmental Engineers

July 2013

1 Introduction

This report describes exploratory and statistical analyses of fecal coliform data collected by the Onondaga County Department of Water Environment Protection (OCDWEP) at six monitoring sites on four major tributaries to Onondaga Lake between 1999 and 2012. The analytical framework is similar to that applied to lakeshore bacteria data (Walker and Walker, 2011) and the Ambient Monitoring Program (AMP) statistical framework (Walker, 1998). Specific tasks include:

- Compile and develop inventories of data collected under various programs over the years (e.g., routine biweekly, wet weather, dry weather, storm events, and special project-related studies). This includes precipitation data that are important to support analysis of dry and wet weather bacteria data.
- Evaluate correlations between fecal coliform concentrations with season, precipitation, and flow.
- Test for long-term trends in annual geometric means under dry, wet and all weather conditions.
- Evaluate compliance with the New York State (NYS) Ambient Water Quality Standard (AWQS) (monthly geometric mean less than 200 cfu/100 ml based on at least 5 samples) and reductions needed to achieve compliance at each site. Because of the highly skewed distribution of fecal coliform data, the analysis is based upon a $\log_{10}$-transformation of the measured concentrations (200 cfu/100 ml corresponds to 2.3 expressed in $\log_{10}$ units).
- Using methods developed in the AMP statistical framework (Walker, 1998), develop estimates of variance components, precision of annual geometric means, and power for detecting trends or step changes towards achieving compliance with the monthly geometric mean standard following future implementation of CSO and other stormwater controls in the watersheds.
- Develop recommendations for monitoring program design and further data analysis.

The AMP tributary bacteria data are utilized for two primary purposes: (a) measuring status and trends in compliance and (b) measuring effectiveness of stormwater controls. These objectives require different data subsets and statistical models.

This report focuses on measuring frequency and trends in compliance, which is determined using periodic samples collected without regard to weather. The frequency of routine sampling for bacteria increased from biweekly (1999-2009) to weekly (5 samples per month, 2010-2012). Theoretically, the biweekly data could not be used for measuring compliance as regulations are based on a monthly geometric mean computed from a minimum of 5 samples. Increasing sampling frequency from biweekly to weekly would improve precision of the monthly and annual geometric means but would not change the average levels or excursion frequency. The historical biweekly data can still be used to calibrate statistical models and assess power for detecting trends with weekly sampling.

Measuring the effectiveness of CSO controls would logically focus on wet weather data derived from grab or high-intensity sampling during storm events. The most direct means of measuring CSO control effectives would be to measure the volume and concentrations in the discharges themselves, as opposed to the tributaries, which are subject to wide variations unrelated to the CSOs (e.g.

temperature, solar radiation, sediment re-suspension, among others). When feasible, direct monitoring of CSO discharges could supplement the tributary data to better evaluate the effective of CSO control measures. The statistical models and optimal sampling program design for the tributary sites would be different from those applicable to compliance measurement. Exploratory analysis of bacteria correlations with season, precipitation, and flow provides a basis for future development of a statistical framework for stormwater sampling. Prototype models for detecting wet weather trends can be developed by first separating dry weather and wet weather samples based upon antecedent precipitation, and then developing separate statistical models for dry weather and wet weather conditions.

2 Data Compilation & Inventory

2.1 Tributary Water Quality

Fecal coliform sampling data collected by the Onondaga County Department of Water Environment Protection (OCDWEP) were extracted from the AMP water quality database for the period 1999-2012 at seven stations on four tributaries to Onondaga Lake including Onondaga Creek, Harbor Brook, Ley Creek and Ninemile Creek (Table 1).

Site Name	Site Code	Waterbody
Ley Creek @ Park	PARK	Ley Creek
Ninemile Creek @ RT48	RT48	Ninemile Creek
Harbor Brook @ Velasko	VELASKO	Harbor Brook
Harbor Brook @ Hiawatha	HIAWATHA	Harbor Brook
Onondaga Creek @ Dorwin	DORWIN	Onondaga Creek
Onondaga Creek @ Kirkpatrick	KIRKPAT	Onondaga Creek
Onondaga Creek @ Spencer	SPENCER	Onondaga Creek

Table 1: Tributary Sampling Stations

The Kirkpatrick station on Onondaga Creek was supplemented with storm event samples collected at the Spencer station, which is located in close proximity to the Kirkpatrick station. The combined data at these two stations is referred to collectively as Kirkpatrick throughout this report.

Many of the statistical analyses developed for this report required that the samples were randomly collected without regard to flow, precipitation or other conditions. The sampling data thus were grouped by the study code defined in the AMP database to differentiate between the biweekly routine, weekly bacteria, storm events, and special studies. Calendar plots were used to inspect the periods of record and frequencies of sampling for each study (Appendix A).

Table 1 lists the study codes and associated study groups for the entire dataset, and indicates which studies were incorporated in each version of the dataset. Three versions of the dataset were defined using different combinations of the study codes for performing sensitivity analyses of the statistical test:

- **Primary:** includes all biweekly (1999-2009) and weekly (2010-2012) routine data as well as the geometric means computed for each storm event (see Section 2.2) and excludes all special studies, which are often project- or site-specific.
- **Biweekly**: includes only routine samples collected at approximately biweekly intervals and excludes all storm events, special studies, and weekly bacteria samples.
- **All:** includes all sampling data except individual storm event samples, which were represented as event geometric means.

Unless otherwise specified, all figures and results through the remainder of this report were performed using the Primary dataset. The Biweekly and All datasets were used to evaluate the sensitivity of the statistical analyses to different datasets.

It should be noted that considerable effort was required to infer whether or not some study codes were part of the routine sampling program. For example, samples assigned to the study code “High Flow” appeared to be part of the biweekly sampling effort as they filled in missing samples assigned to the study code “Routine (Biweekly)” in 2007-2010. The majority of samples collected between 1999 and 2008 were missing the study code (i.e. “(Blank)”) and were assumed to be part of the biweekly sampling program. The calendar plots in Appendix A provide a resource for understanding the frequency and periods of record for each study code.

Figure 1 shows time series of fecal coliform concentrations at each site based on the Primary dataset and indicates which values were censored (i.e. < or > a detection limit). Table 3 summarizes the fraction of censored values at each site, which were generally less than about 12% of the total samples. Because they were a relatively small fraction of the total samples, censored values were assigned to the value of the respective detection limit for use in all subsequent analyses.

Figure 2 shows the annual geometric mean fecal coliform concentration at each site with 95% confidence intervals and reference lines indicating the approximate date of completion for site-specific CSO remediation projects. Figure 3 shows the distributions by site and month indicating the seasonality of fecal coliforms, which tend to be higher in the summer months at all six sites. These figures include a reference line at 200 cfu/100 mL (2.3 in $\log_{10}$ units), which is the NYS AWQ standard for fecal coliform. Because this standard strictly applies to a monthly geometric mean based on a minimum of 5 samples, it is only included for numerical perspective.

				Stations						Dataset		
Study Code	**Study Group**	**No. Samples**	**Period**	**PARK**	**RT48**	**VELASKO**	**HIAWATHA**	**DORWIN**	**KIRKPAT**	**Primary**	**Biweekly**	**All**
(Blank)	Unknown	1,246	1999-2008	●	●	●	●	●	●	●	●	●
Clinton Phase 2A	Special	18	2009					●				●
Clinton Phase I	Special	17	2008-2009					●				●
Enhanced Trib	Special	5	2012					●				●
High Flow	Routine	132	2005-2011	●	●	●	●	●	●	●	●	●
High Flow (Special)	Routine	4	2008			●	●	●	●	●	●	●
High Flow Event	Routine	6	2006	●	●	●	●	●	●	●	●	●
HillcrestPS FM Break	Special	4	2007			●	●					●
Midland RTF Dewater	Special	70	2005-2008					●				●
Onondaga Creeks	Routine	48	1999-2000	●	●	●	●	●	●	●	●	●
Onondaga Lake	Routine	6	1999	●	●	●	●	●	●	●	●	●
Quarterly	Routine	95	2007-2011	●	●	●	●	●	●	●	●	●
Quarterly Creek Event	Routine	12	1999-2000	●	●	●	●	●	●	●	●	●
Routine (Biweekly)	Routine	362	2007-2011	●	●	●	●	●	●	●	●	●
Routine Biweekly-Dry	Routine	48	2010	●	●	●	●	●	●	●	●	●
Spill Response	Special	3	2003	●								●
Storm Event	Storm	1,296	1999-2009	●	●	●	●	●	●			
Storm Event (Geomean)*	Storm	56	1999-2009	●	●	●	●	●	●	●		●
Trib Bacti ID Study	Special	106	2008-2012	●		●	●	●				●
Trib Bacti Only	Bacteria	329	2011-2012	●	●	●	●	●	●	●		●
Trib Biweekly	Routine	141	2011-2012	●	●	●	●	●	●	●	●	●
Trib Biweekly Dry	Routine	35	2011	●	●	●	●	●	●	●	●	●
Trib Biweekly HF	Routine	42	2011-2012	●	●	●	●	●	●	●	●	●
Trib Quarterly	Routine	27	2011-2012	●	●	●	●	●	●	●	●	●
Trib Quarterly Dry	Routine	12	2011	●	●	●	●	●	●	●	●	●
Tributary Bacteria	Bacteria	176	2010-2011	●	●	●	●	●	●	●		●
Tributary Bacti-Dry	Bacteria	96	2010-2011	●	●	●	●	●	●	●		●
Wet Weather	Routine	6	2001	●	●	●	●	●	●	●	●	●

* Storm event geometric means computed from storm event samples

Table 2: Summary of Study Codes and Dataset Definitions

Rows highlighted by Study Group.

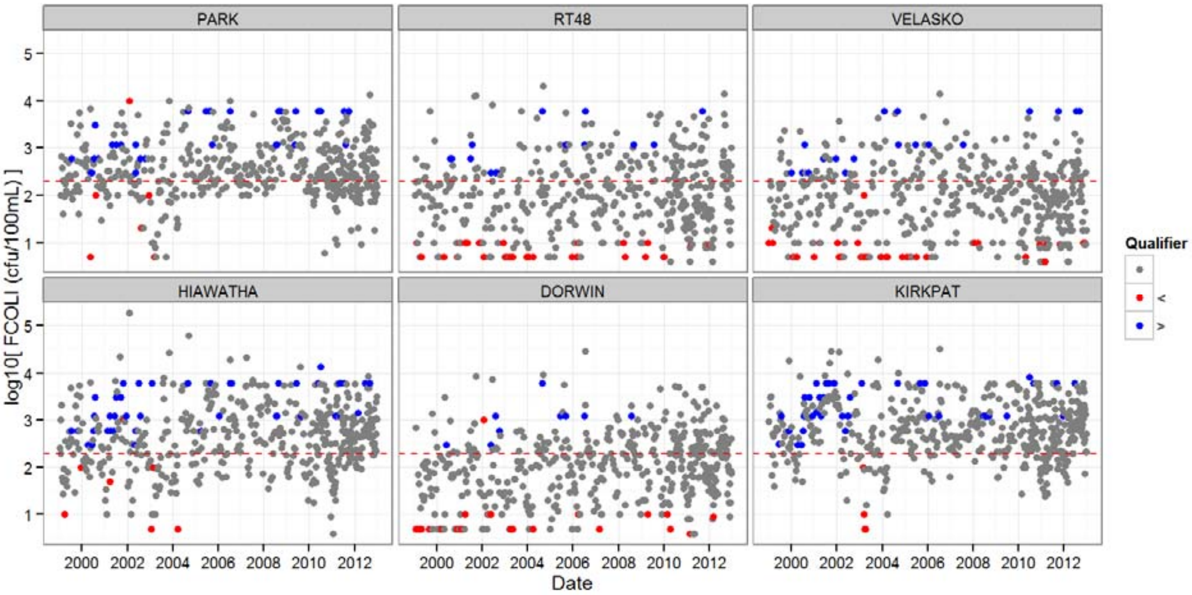

Figure 1: Time Series of Fecal Coliform Sample Data

Dashed red line indicates 200 cfu/100mL reference.

Site	Left Censored (< DL)	Right Censored (> DL)	Not Censored
PARK	1.5%	6.3%	92.2%
RT48	7.2%	3.3%	89.5%
VELASKO	8.5%	3.7%	87.8%
HIAWATHA	1.5%	9.2%	89.4%
DORWIN	5.8%	1.9%	92.4%
KIRKPAT	0.8%	8.3%	90.9%

Table 3: Fraction of Censored Samples by Site

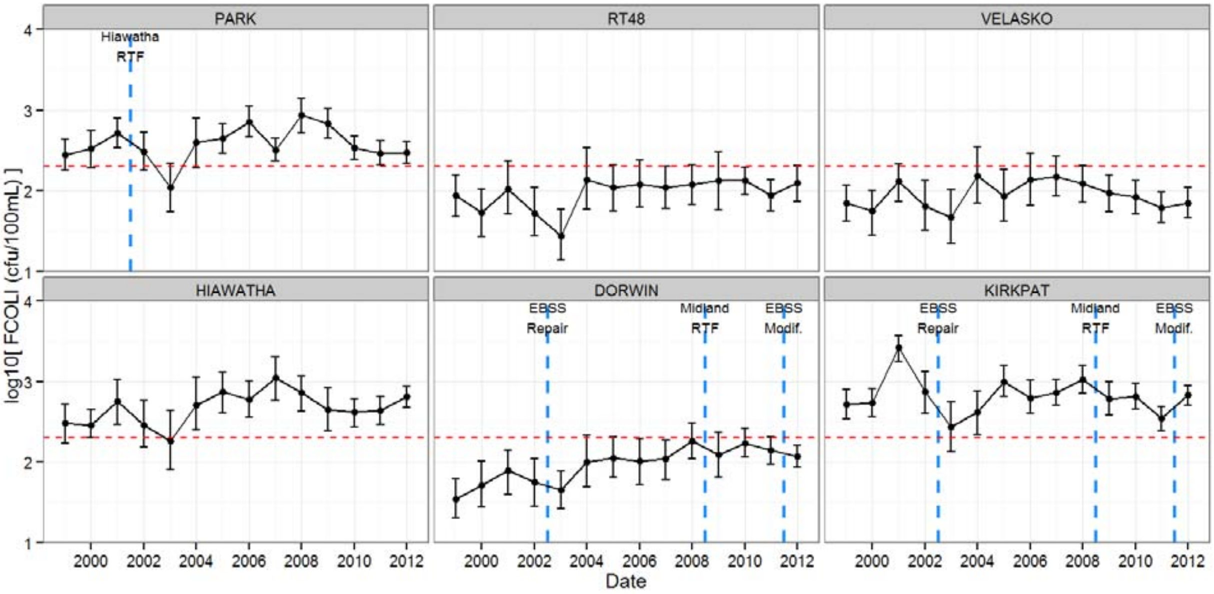

Figure 2: Annual Geometric Mean Fecal Coliform Concentration

Error bars represent 95% confidence intervals. Dashed red line indicates 200 cfu/100mL reference.
Dashed blue lines indicate completion dates of CSO remediation projects.

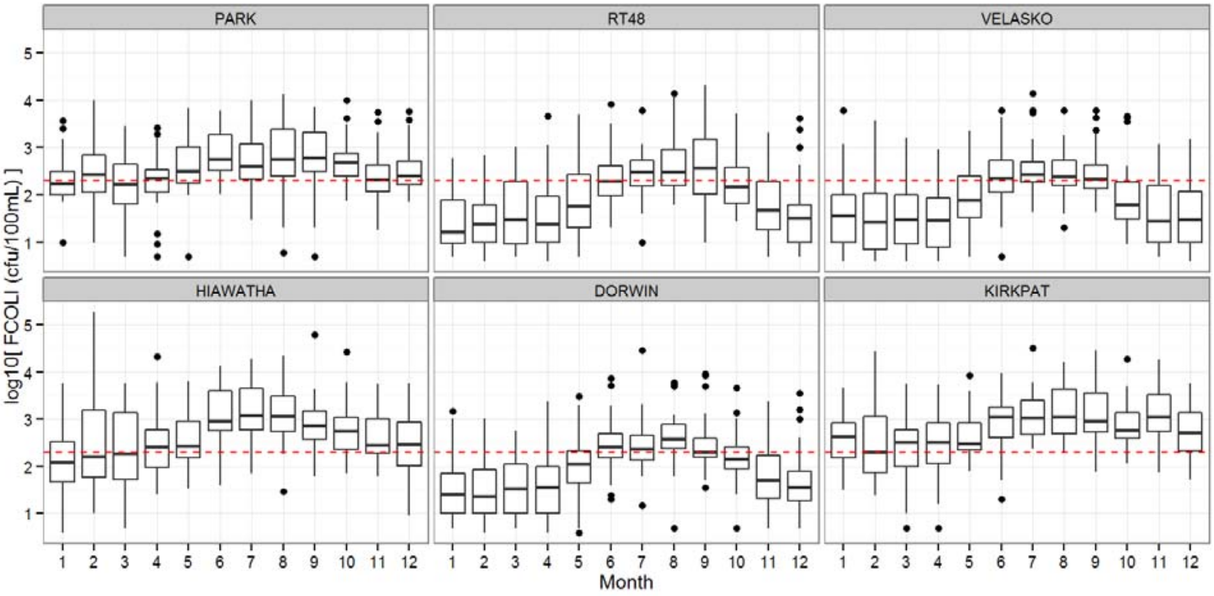

Figure 3: Boxplots of Fecal Coliform by Site and Month Showing Seasonality

Dashed red line indicates 200 cfu/100mL reference.

2.2 Storm Event Samples

The Primary dataset included 54 storm events where a series of samples were collected at short intervals (usually 3 hours) over multiple days during a wet weather event. The storm event data were primarily collected in 1999-2003, with three additional events in 2008-2009 collected at the Onondaga Creek stations and one event in 2009 at the Hiawatha station. Including all of the individual storm event samples in the statistical analyses would bias the dataset towards wet weather conditions. Therefore, the geometric mean concentration was computed for each event by calculating the mean of the $\log_{10}$-transformed concentrations. The geometric means of the storm events were then assigned a study code of "Storm Event (Geomean)" and added to the dataset for inclusion in subsequent analyses.

Unlike the non-storm grab samples, the storm event data were comprised of multiple samples collected over the course of an event. For some events, the precipitation fell before sampling began, while for others precipitation fell while samples were being collected making it difficult to assign a single value for the antecedent precipitation. To account for the differing time of precipitation before or during an event, the antecedent precipitation associated with each storm event geometric mean was assigned to the maximum antecedent precipitation over the sampling period. To document the precipitation data associated with each sampling event, a series of plots were generated showing the individual fecal coliform samples, the geometric mean of the event, the corresponding daily flow, and the hourly precipitation (Appendix B). These figures highlight which precipitation data were used to represent the antecedent precipitation of the event geometric mean.

2.3 Streamflow Data

Daily streamflow data for each sampling site were collected by the USGS and extracted from the AMP water quality database.

2.4 Precipitation Data

Hourly precipitation data collected at Hancock International Airport (COOP station 308383) was obtained from the National Climatic Data Center (NCDC) Climate Data Online (CDO). This dataset was compared to daily precipitation records from the Global Historical Climatology Network – Daily (GHCND) dataset at the same location (Station USW00014771) to ensure the hourly data were complete and similar to the reported daily total precipitation. Hourly precipitation collected at the Metro Wastewater Treatment Plant was also provided by OCWEP, but was not utilized after initial reviews of the dataset indicated missing data and insufficient quality at hourly time steps (A. Deskins, personal communication). Appendix C provides a series of figures that compare the hourly airport, daily airport, and hourly Metro datasets. The extent of missing data in the Metro dataset is evident in the annual total precipitation, which was on average about 5 inches less than the two airport datasets.

3 Correlations with Precipitation and Flow

The response of each tributary to wet weather conditions has been evaluated by comparing the relationships between fecal coliform concentrations and both antecedent precipitation and flow.

3.1 Precipitation Correlation

Figure 4 shows the correlation between fecal coliform concentrations and 48-hour antecedent precipitation. This figure includes a locally-weighted scatterplot smooth (LOESS; Helsel and Hirsch, 2002) with 95% confidence intervals to highlight the non-linear patterns in the data (weak relationship for dry weather samples and strong relationship for wet weather). The samples are classified as dry, wet or storm where dry and wet weather samples were defined as less than or greater than 48-hour precipitation of 0.1 inches, respectively, and storm points indicate storm event geometric means. Both variables were $\log_{10}$-transformed to facilitate visualization of the relationship. Data points with antecedent precipitation of 0 inches were assigned to 0.005 inches, which is half the detection limit for hourly precipitation of 0.01 in, to allow plotting on the log-scale.

The antecedent period of 48 hours was chosen for the wet/dry classification as it exhibited a strong relationship between fecal coliforms and precipitation under wet weather and a weak relationship under dry weather. The 0.1 inch threshold was chosen as it represented an approximate break point above which levels started to increase with precipitation. Appendix D provides correlations for alternative antecedent periods (1-5 days) and thresholds as well as correlations based on seasonal subsets of the data. Sensitivity of the trend analyses to varying antecedent periods and thresholds are discussed in Section 4.2. Regardless of time period or threshold, levels in increase by 1-2 log units (10-100 fold) under wet weather conditions.

Time series of the individual sample data classified by dry weather, wet weather or storm event geometric mean are shown in Figure 5. Time series of the annual geometric means for the dry weather and wet weather as well as all weather conditions are shown in Figure 6, which includes reference lines indicating the approximate completion data of site-specific CSO remediation projects.

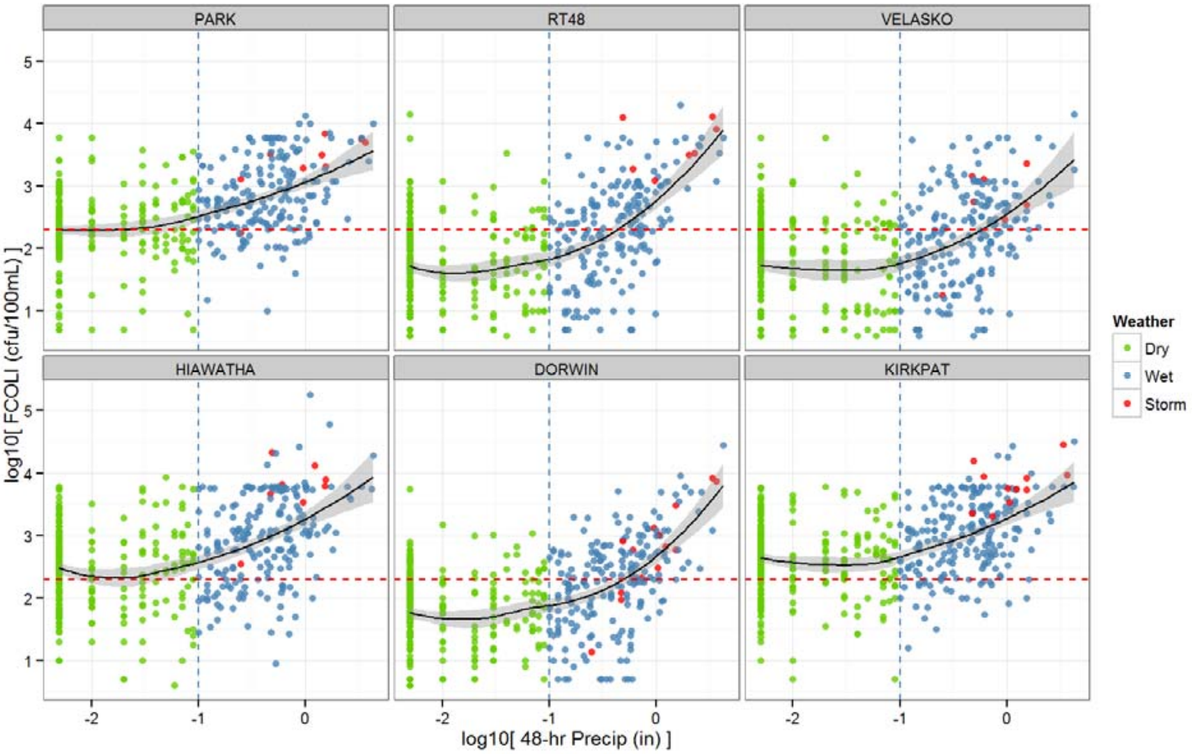

Figure 4: Correlation Between Fecal Coliform and 48-hour Antecedent Precipitation

Black line is LOESS smooth with 95% confidence intervals. Dashed red line indicates 200 cfu/100mL reference. Dashed blue line indicates dry/wet threshold of 0.1 inches. Storm samples indicate storm event geometric means. Antecedent precipitation values of 0 were assigned to 0.005 inches to facilitate plotting on the log-scale.

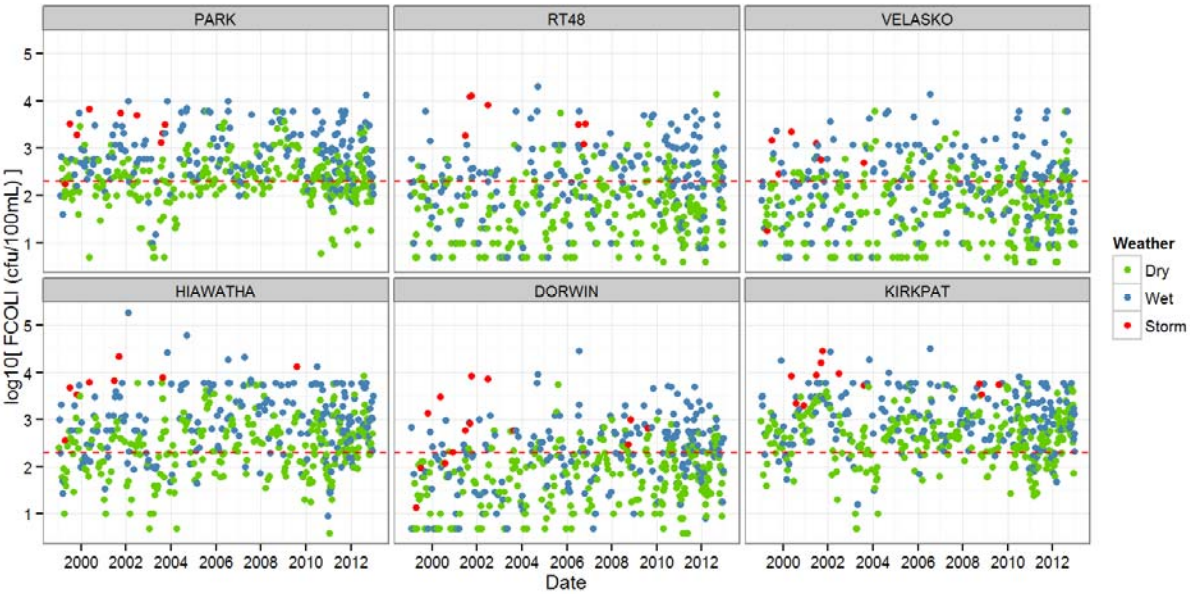

Figure 5: Time Series of Fecal Coliform Samples Classified by Weather Condition

Dashed red line indicates 200 cfu/100mL reference. Dry and wet samples based on a 48-hour precipitation of 0.1 inches. Storm samples indicate storm event geometric means.

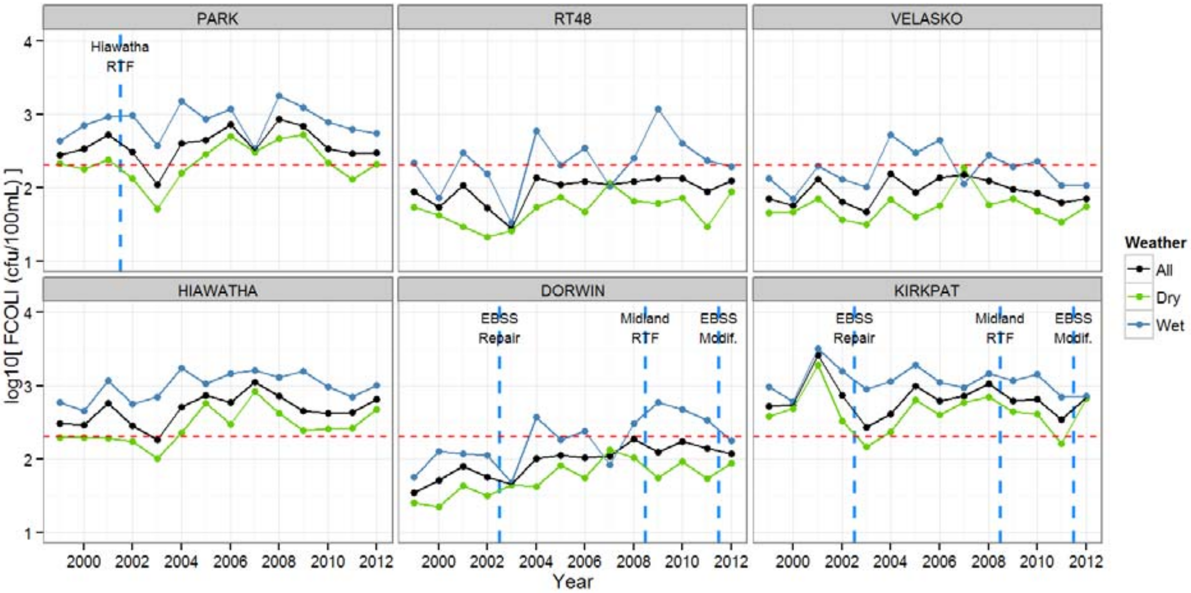

Figure 6: Annual Geomean of Fecal Coliform for All Data and by Weather Condition

Dashed red line indicates 200 cfu/100mL reference. Dry and wet samples based on a 48-hour precipitation of 0.1 inches. Dashed blue lines indicate completion dates of CSO remediation projects.

3.2 Flow Correlation

In addition to antecedent precipitation, dry and wet weather conditions can also be indicated by streamflow. Figure 7 shows the correlation between fecal coliform concentration and daily flow at each station including a LOESS smooth with 95% confidence intervals. The dry and wet weather classifications are based on a 48-hr precipitation of 0.1 inches, as discussed in Section 3.1. The relationships of fecal coliform against flow are not as strong as the relationship against precipitation. The LOESS smooth lines indicate the lower concentrations in the middle of the range of flows and higher concentrations at the lower and higher ends of the range. This pattern is likely due to the seasonality of streamflow, where the lowest flows tend to occur in the summer months when fecal coliform concentrations are higher (Figure 3). In general, the wet weather and storm samples (blue and red points) tend to be greater than the dry weather samples (green points) across the entire range of flows at each site. The relationships between fecal coliform and flow at each site for different seasons are shown in Appendix D.

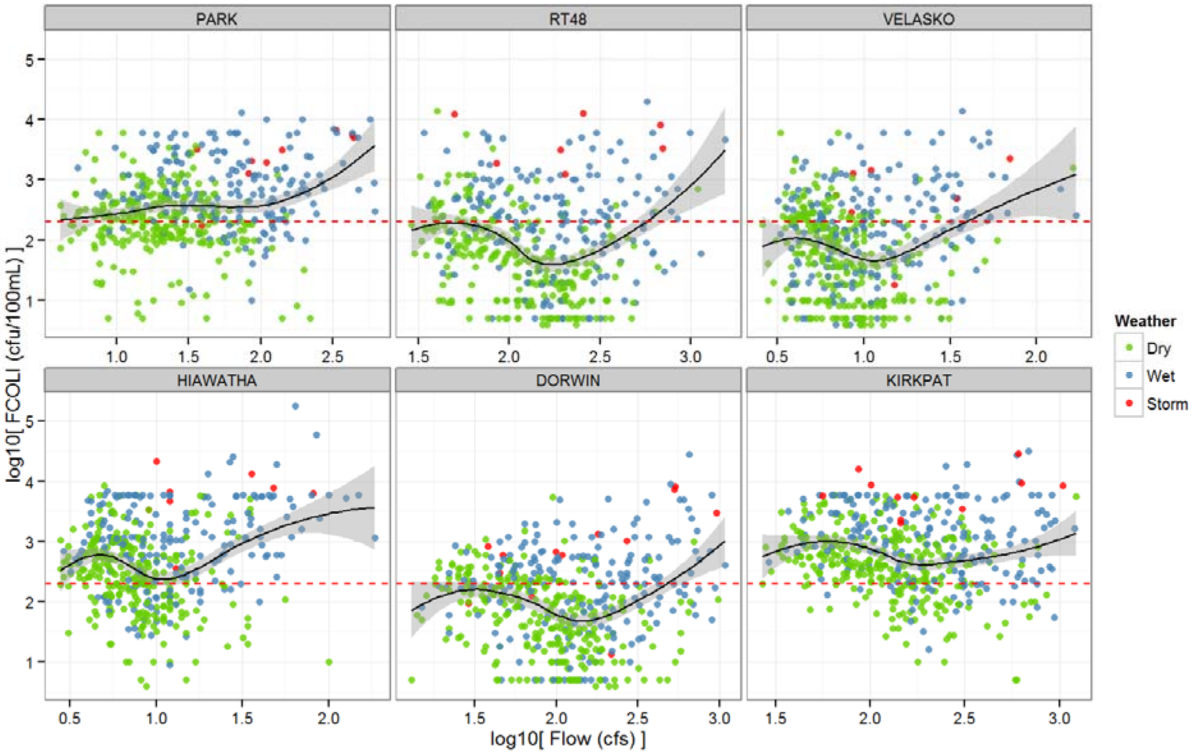

Figure 7: Correlation Between Fecal Coliform and Daily Flow

Black line is LOESS smooth with 95% confidence intervals. Dashed red line indicates 200 cfu/100mL reference.
Dry and wet samples based on 48-hour precipitation of 0.1 inches.
Storm samples indicate storm event geometric means.

4 Trend Analysis

Long-term trends in fecal coliform concentrations were evaluated using the seasonal Kendall test (Helsel and Hirsch, 2002). The seasonal Kendall test is a non-parametric statistical test based on the ranks of observations instead of their specific values. If observations near the end of a period have generally higher (or lower) ranks than observations at the start of the period, then this is evidence that there is a change in magnitude over time. The slope of the trend is computed using the Sen slope, which is the median slope between all pairwise observations. The intercept is then computed as the median intercept when this slope is applied to all data points. Appendix E provides results of two alternative trend tests: linear regression and the Mann-Kendall test (Helsel and Hirsch, 2002). The linear regression trend test is a parametric test based on the annual geometric mean as a function of year. The Mann-Kendall test is a non-parametric trend test that provides the basis for the seasonal Kendall test, but is only applied to the annual geometric mean as function of year.

The trend tests were performed using the $\log_{10}$-transformed fecal coliform concentrations. For each site, the monthly mean of the log-transformed concentrations were first computed, which correspond to the geometric means of the untransformed concentrations. For the linear regression and Mann-Kendall tests, the annual geometric mean was computed as the annual mean of the monthly geometric means in order to reduce potential bias resulting from uneven distribution of wet and dry weather samples at different times of the year (e.g. more wet samples collected in spring than winter).

4.1 Trend Results for Primary Dataset

The trend test was performed on the primary dataset, which includes biweekly routine samples, weekly bacteria samples, and storm event geometric means, and excludes site-specific special studies and individual storm event samples (see Table 2). Figure 8 summarizes the trend slope and significance level (p-value) at each site using all data and for subsets of dry and wet weather samples based on a 48-hour antecedent precipitation threshold of 0.1 inches. Table 4 lists the number of samples, overall geometric mean, and slope magnitudes and significance levels for each site and weather condition using all three trend tests. Note that trends of 0.02 and 0.04 log10[cfu/100ml]/yr correspond to approximately 5%/yr and 10%/yr, respectively, in the long-term geometric mean.

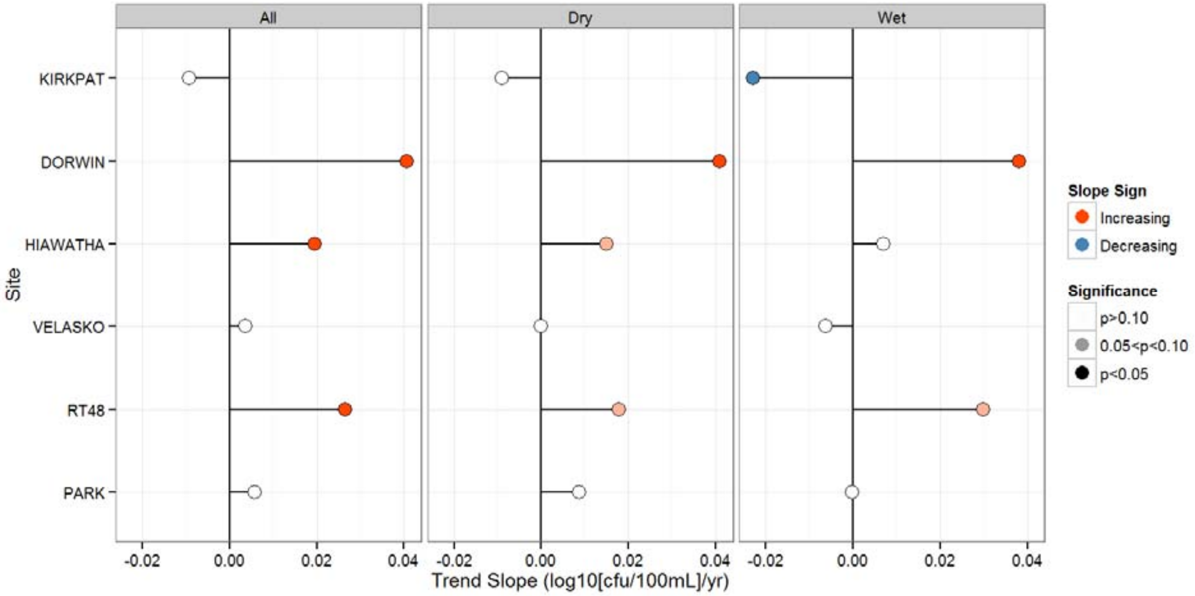

Figure 8: Trend Slopes and Significance of Fecal Coliform for All Data and Weather Condition

				Seasonal Kendall			Mann Kendall			Linear Regression		
Site	Weather	No. Samples	Geomean (cfu/100mL)	Slope (log[cfu/100mL]/yr)	Slope (%/yr)	p-value	Slope (log[cfu/100mL]/yr)	Slope (%/yr)	p-value	Slope (log[cfu/100mL]/yr)	Slope (%/yr)	p-value
PARK	**All**	475	370	0.006	1.3%	0.51	0.002	0.5%	0.58	0.010	2.4%	0.52
	Dry	273	226	0.009	2.0%	0.21	0.007	1.6%	0.66	0.014	3.4%	0.44
	Wet	202	786	0.000	0.0%	0.87	0.012	2.7%	0.83	0.007	1.7%	0.67
RT48	**All**	457	87	0.026	6.3%	0.01	0.012	2.9%	0.23	0.025	5.8%	0.06
	Dry	261	46	0.018	4.2%	0.06	0.019	4.6%	0.15	0.021	4.9%	0.16
	Wet	196	230	0.030	7.1%	0.09	0.024	5.8%	0.19	0.035	8.5%	0.17
VELASKO	**All**	485	91	0.004	0.8%	0.67	-0.004	-0.9%	0.83	0.002	0.5%	0.86
	Dry	278	50	0.000	0.0%	0.94	0.002	0.5%	0.83	0.004	0.8%	0.82
	Wet	207	185	-0.006	-1.4%	0.56	0.006	1.3%	0.83	0.007	1.6%	0.70
HIAWATHA	**All**	480	476	0.019	4.6%	0.04	0.017	4.1%	0.15	0.021	4.8%	0.15
	Dry	274	274	0.015	3.5%	0.10	0.020	4.6%	0.23	0.022	5.3%	0.18
	Wet	206	1031	0.007	1.6%	0.35	0.017	4.0%	0.38	0.020	4.7%	0.16
DORWIN	**All**	486	88	0.041	9.8%	0.00	0.038	9.2%	0.00	0.043	10.5%	0.00
	Dry	273	53	0.041	9.9%	0.00	0.042	10.1%	0.00	0.042	10.1%	0.00
	Wet	213	185	0.038	9.2%	0.00	0.047	11.6%	0.06	0.052	12.6%	0.02
KIRKPAT	**All**	496	643	-0.009	-2.1%	0.43	0.004	0.8%	0.91	-0.007	-1.6%	0.67
	Dry	280	408	-0.009	-2.0%	0.18	0.008	1.8%	0.66	-0.005	-1.2%	0.80
	Wet	216	1199	-0.023	-5.1%	0.02	-0.005	-1.1%	0.74	-0.005	-1.2%	0.70

Table 4: Summary of Trend Tests for Primary Dataset

Slopes and p-value columns shaded for slope direction (orange: increasing, blue: decreasing) and significance (darker: p<=0.05, lighter: 0.05<p<=0.1, white: p>0.10).

The seasonal Kendall trend tests indicate strongly significant ($p<0.05$) increasing trends of 0.041, 0.041, and 0.038 $\log_{10}$[cfu/100mL]/yr (9.8, 9.9, 9.2 %/yr) at the Dorwin station on Onondaga Creek based on all weather, dry weather and wet weather subsets, respectively. These annual trends primarily reflect significant increasing trends in both the winter months of January-March and summer months of June-August (see Appendix E).

A strongly significant increasing trend of 0.026 $\log_{10}$[cfu/100mL]/yr (6.6%/yr) was found at the RT48 station on Ninemile Creek based on all data, while the dry and wet weather subsets showed weakly significant ($0.05 < p <= 0.10$) increasing trends of 0.018 and 0.030 $\log_{10}$[cfu/100mL]/yr (4.2%/yr and 7.1%/yr), respectively. These trends were driven by increasing trends primarily in during December and April.

Similarly, the Hiawatha station on Harbor Brook showed a strongly significant increasing trend of 0.019 $\log_{10}$[cfu/100mL]/yr (4.6%/yr) based on all data, and a weakly significant trend of 0.015

$\log_{10}$[cfu/100mL]/yr (3.5%/yr) under dry weather only. The Hiawatha trends were driven by increasing trends in March, April and June.

The Kirkpatrick station on Onondaga Creek was the only station that showed a strongly significant ($p < 0.05$) decreasing trend under wet weather of -0.023 $\log_{10}$[cfu/100mL]/yr (-5.1%/yr) primarily during January, although this trend was not observed using only the Biweekly dataset (see Appendix E). It also did not occur if the storm event geometric mean samples were removed from either the Primary or All datasets (see Section 4.3). This suggests that inclusion of storm-event samples, which were collected in the earlier part of the period of record (2000-2003), may be causing a bias towards extreme wet weather conditions in the earlier years.

With the exception of Dorwin, most of the observed trends were driven by changes during the winter months (Nov-Apr). Dorwin was the only site that showed significant increasing trends throughout the summer months (May-Oct), although Park and Hiawatha showed significant increasing trends during the month of June using all weather data.

For the Dorwin site, the two alternative trend tests (linear regression and Mann-Kendall), which are based upon annual geometric means as a function of year, yielded results similar to those yielded by the seasonal Kendall test. The alternative tests indicated no significant trend at Hiawatha and a less significant trend at Rte 48. This may reflect the fact that the seasonal Kendall test is more powerful when there are strong seasonal variations in the datasets.

Appendix E provides summary and diagnostic plots, and summary tables of the trend magnitudes and significance using all three tests at each site and for each weather condition using all three datasets. The diagnostic plots show individual monthly trends computed for the seasonal Kendall tests as well as time series of the monthly and annual geometric mean concentrations for each site and weather condition.

4.2 Sensitivity to Antecedent Precipitation

The sensitivity of the seasonal Kendall trend test was evaluated by using different antecedent precipitation periods and thresholds for classifying wet and dry samples. The results of these tests are shown in Appendix F and indicate that the increasing trends at RT48, Dorwin, and Hiawatha and the decreasing trend at Kirkpatrick are reasonably stable for the range of antecedent periods and thresholds when using the Primary dataset. It should be noted that at higher thresholds, the number of wet weather samples decreases which could result in gaps in the monthly time series that could affect the trend test. The fraction of wet/dry samples for each period and threshold are also included in Appendix F for reference.

4.3 Sensitivity to Dataset

The trend tests were performed on all three versions of the dataset (Primary, Biweekly, All; see Table 2) to evaluate the effect of including or excluding data from different study codes. Table 5 compares the seasonal Kendall trend tests at each site and weather condition across the three datasets.

		Primary				Biweekly				All			
Site	**Weather**	**No. Samples**	**Geomean (cfu/100mL)**	**Slope (log[cfu/100mL]/yr)**	**p-value**	**No. Samples**	**Geomean (cfu/100mL)**	**Slope (log[cfu/100mL]/yr)**	**p-value**	**No. Samples**	**Geomean (cfu/100mL)**	**Slope (log[cfu/100mL]/yr)**	**p-value**
PARK	**All**	475	370	0.006	0.51	362	358	0.014	0.13	500	369	0.003	0.74
	Dry	273	226	0.009	0.21	213	229	0.016	0.07	298	225	0.008	0.27
	Wet	202	786	0.000	0.87	149	775	0.019	0.15	202	786	0.000	0.87
RT48	**All**	457	87	0.026	0.01	360	81	0.029	0.02	457	87	0.026	0.01
	Dry	261	46	0.018	0.06	212	45	0.016	0.11	261	46	0.018	0.06
	Wet	196	230	0.030	0.09	148	206	0.049	0.00	196	230	0.030	0.09
VELASKO	**All**	485	91	0.004	0.67	374	87	0.004	0.62	531	95	0.009	0.33
	Dry	278	50	0.000	0.94	217	49	-0.001	0.80	324	54	0.008	0.47
	Wet	207	185	-0.006	0.56	157	184	0.007	0.65	207	185	-0.006	0.56
HIAWATHA	**All**	480	476	0.019	0.04	374	454	0.025	0.01	502	470	0.015	0.06
	Dry	274	274	0.015	0.10	218	270	0.014	0.12	296	276	0.014	0.09
	Wet	206	1031	0.007	0.35	156	939	0.023	0.07	206	1031	0.007	0.35
DORWIN	**All**	486	88	0.041	0.00	371	81	0.044	0.00	616	85	0.042	0.00
	Dry	273	53	0.041	0.00	214	50	0.033	0.00	352	52	0.037	0.00
	Wet	213	185	0.038	0.00	157	173	0.061	0.00	264	192	0.038	0.00
KIRKPAT	**All**	496	643	-0.009	0.43	381	609	-0.005	0.57	496	643	-0.009	0.43
	Dry	280	408	-0.009	0.18	220	404	-0.009	0.21	280	408	-0.009	0.18
	Wet	216	1199	-0.023	0.02	161	1136	-0.005	0.47	216	1199	-0.023	0.02

Table 5: Comparison of Seasonal Kendall Trend Results by Dataset

Slopes and p-value columns shaded for slope direction (orange: increasing, blue: decreasing) and significance (darker: p<=0.05, lighter: 0.05<p<=0.1, white: p>0.10).

The Biweekly dataset yielded similar results as the Primary dataset at Dorwin, although the magnitude of the trend under wet weather increased from 0.038 to 0.061 $\log_{10}$[cfu/100mL]/yr (9.1 to 15.1 %/yr) and the trend under dry weather decreased from 0.041 to 0.033 $\log_{10}$[cfu/100mL]/yr (9.9 to 7.8 %/yr). The Biweekly dataset did not yield the decreasing trend at Kirkpatrick under wet weather nor increasing trends at Hiawatha and RT48 under dry weather, which were all found using the Primary dataset, but it did yield a weakly significant increasing trend at Park under dry weather, which was not found using the Primary dataset.

In general, the Primary and All datasets yielded similar results at all stations indicating that inclusion of the site-specific special studies did not have a major effect on the analysis. Comparing the number of samples in each dataset shows that the special studies were primarily dry weather samples, with the exception of Dorwin where there 51 wet weather samples added to the All dataset, which did not have a significant effect on the trend results.

5 Compliance

Figure 9 summarizes the status of compliance with the fecal coliform standard at each site based upon the biweekly dataset. Compliance is measured using periodic samples collected without regard to weather. Routine sampling frequency increased from biweekly (1999-2009) to weekly (5 samples per month, 2010-2012). Theoretically, biweekly data would not be used for measuring compliance, since regulations are based upon monthly geometric means computed from at least 5 samples. Increasing sampling frequency from biweekly to weekly would be expected to improve the precision of monthly geometric means but not change average levels or excursion frequency. Given the ranges of historical data, the biweekly data are sufficient to assess compliance status and variability in the data.

Figure 9 shows the percent of months when the standard would be met as a function of hypothetical reductions in the long-term geometric mean for each site over the 1999-2012 period. This calculation was performed by re-scaling historical biweekly data (i.e. reducing each sample by a fixed percentage) and re-computing the percent of months with geometric means less than 200 cfu/100 ml. Calculations assume that variability would not change as the longterm geometric mean is reduced. Results for 0% reduction reflect the average compliance status over the 14-year dataset and are approximate because they are based upon biweekly instead of 5/month sampling.

Upstream/rural sites shown on the left (Velasko, Rt48, Dorwin) met the standard in ~ 70% of the months, whereas downstream/urban sites (Kirkpatrick, Ley, Hiawatha) met the standard in 15-25% of the months. Achieving the standard in 90% of the months would require reductions of ~60% at the upstream/rural sites and ~90% at the downstream/urban sites. Results for other frequencies can be read from the graphs. Achieving higher compliance frequencies would require significantly greater reductions. This analysis provides approximate estimates of overall reductions required to achieve the standard. Results for May-October are also shown in Figure 9. As expected, compliance frequencies are lower because bacteria levels are higher during those months. Refinements to the analysis could differentiate dry-weather vs. wet-weather reductions.

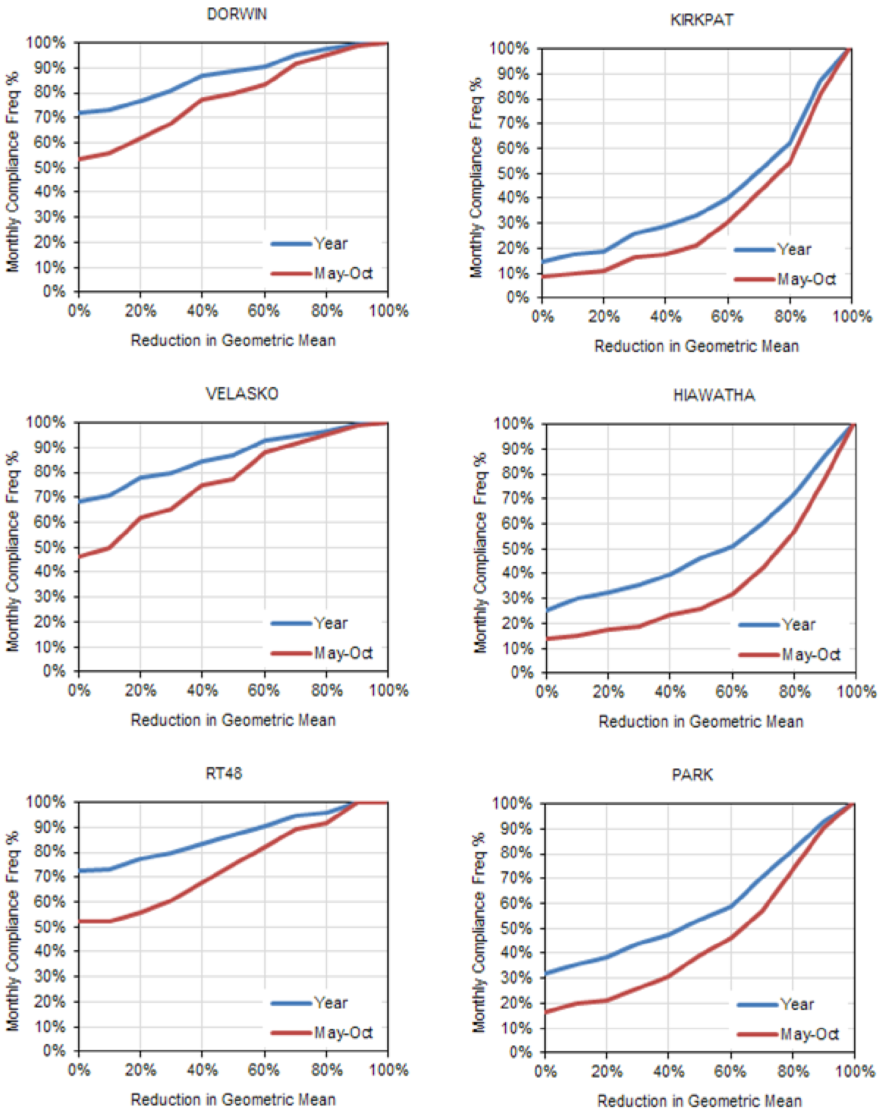

Figure 9: Monthly Compliance Frequency vs. Hypothetical Reductions in Long-Term Geometric Mean

6 Variance Components & Power

Estimation of variance components provides a basis for evaluating the precision of the yearly geometric means and power for detecting long-term trends for different monitoring program designs (e.g., duration and sampling frequency). Details are described in the AMP statistical framework reports (Walker, 1998). While the math is more complicated, the variance of log-transformed values can generally be partitioned as follows:

$$V(\text{Total}) = V(\text{Season}) + V(\text{Precipitation}) + V(\text{Trend}) + V(\text{Year}) + V(\text{Residual})$$

For example, V(Season) is that portion of the total variance explained by seasonal variations. Analysis of Variance (ANOVA) is used to estimate Seasonal, Trend, Precipitation, and Year variance components in the historical data. The Residual term reflects the combination of random sampling error, natural variability, and effects of other factors that are not listed. Residual terms are generally estimated based on difference between the other variance terms (Total – Season – Precipitation – Trend – Year). The Trend term would reflect any long-term trend or change related to management measures or other anthropogenic factors.

Once the variance components are calibrated to historical data, power for detecting future trends or step changes depends on the year-to-year variability in the measured geometric means:

Vm(Year) = V(Year) + (V(Precipitation) + V(Residual))/Ns [Ns= number of samples / yr]

The year-to-year variability in the measured log-mean (Vm) is greater than the actual variability (V(Year), first term) because of sampling variability (second term). So by increasing sampling intensity (Ns), we can reduce Vm(Year), reduce sampling variability, and increase power for trend detection. Power is essentially the likelihood of detecting a trend (rejecting the null hypothesis of no trend, $p<0.1$) when in fact a trend exists. Increasing power decreases the risk of Type II error in the statistical test (probability of a false negative).

Routine biweekly data are best suited for estimating variance components of the compliance measurements. Reasonably uniform sampling frequency across years and months is desirable for developing accurate estimates of variance components. Compliance measurements are based upon periodic samples that are random with respect to precipitation. The Precipitation and Residual variance components are included in the above equation because they would both contribute to the random variability in the biweekly data, after adjusting for seasonal and year-to-year variations.

Results of the variance component and power analysis using the biweekly data are summarized in Figure 10. Power (probability of detecting change) is assessed for a range of sampling frequencies (monthly, biweekly, weekly), durations (3 – 10 years), and statistical tests (step change vs. trend). Several different combinations of these parameters could be tested. The values selected are intended to illustrate general magnitudes, variations across sites, sensitivity to monitoring frequency, and sensitivity to duration. The sites are sorted in order of upstream/rural sites (RT48, Velasko, Dorwin) to downstream/urban sites (Kirkpatrick, Hiawatha, Park). Results tend to cluster along these lines in many respects. The figure includes the following metrics:

- Log10-Means and Geometric Means +/- one standard deviation of the yearly geometric means. Geometric means range from ~100 cfu/100 ml for upstream sites to 300-600 cfu/100 ml for the downstream sites.

- Variance components (year, trend, seasonal, precipitation, random) expressed in $[\log10(\text{cfu}/100\text{mL})]^2$ units and as a percent of the total variance at each site. While upstream sites have significantly lower geometric means, they tend to have higher seasonal and total variance. The higher variance in the upstream sites may partially reflect lower precision of data in the lower concentration range when values approach the lower detection limits. The downstream sites tend to have higher random year-to-year variance, which largely determines power for trend detection. Precipitation (wet vs. dry classification based on 48-hour precipitation of 0.1 inches) explains about 20-30% of the total variance across all sites.

- Precision of the yearly geometric mean (standard error) with sensitivity to sampling frequency (monthly, biweekly, weekly) expressed as a percent of the geometric mean ($10^s - 1$, where s = standard error of the log-10 mean). These generally range from ~20-25% for weekly sampling to ~50-60% for monthly sampling).

- Detectable decrease vs. sampling frequency for 10 years of baseline data and 3-years of data after a hypothetical intervention (e.g. operation of CSO control, etc.). For a decrease of the computed magnitude, there would be >80% likelihood that a t-test (or other statistical test) applied to the data would indicate a significant change (p<.10). For a weekly sampling frequency, the detectable decreases range from 45-55% at the upstream sites to 55%-65% at the downstream sites. These values increase by ~10% with a monthly frequency. Results are relatively insensitive to sampling frequency because power is controlled more by random year-to-year variability than by precision of the yearly geometric means. These detectable decreases indicate that the existing monitoring program (biweekly or weekly) is sufficient for measuring long-term progress towards achieving compliance, which would require >60% decreases at the upstream sites and >90% reductions at the upstream sites.

- Detectable decrease vs. number of years of post-intervention monitoring (3, 5, & 10 years) for 10 years of baseline data and biweekly sampling frequency. For a weekly sampling frequency, these range from 38-43% of the long-term geometric mean at the upstream sites to 45-50% at the downstream sites. Values increase by ~6% with biweekly monitoring and ~12% with monthly monitoring.

- Probability of detecting decreases of 20, 40, and 60% in the long-term geometric means with 10 years of baseline and 3-years of post-intervention monitoring. Typical probabilities for the upstream sites are 15%, 45%, and 85%, respectively. Typical probabilities for the downstream sites are 12%, 35%, and 75%, respectively. The probabilities are slightly lower at the downstream sites because they have higher year-to-year variability.

- Detectable linear trend (%/yr) vs. sampling frequency for 10 years of data. This would be a test for gradual decreases (or increases) not tied to specific interventions or projects. For a trend of this magnitude, there would be >80% likelihood that a linear regression applied to the annual geometric means of the data would indicate a significant trend (p<.10). For a weekly sampling frequency, the detectable trends range from 12-14%/yr at the upstream sites to 16-18 %/yr at the downstream sites. These values increase by ~5% with a monthly frequency. Again, several different combinations of these parameters could be tested. The values selected are intended to illustrate magnitudes and variation across sites and sensitivity to frequency.

- Probability of detecting trends of +/- 5%, 10 %, and 15%/yr over 10 years with biweekly samples. Typical probabilities for the upstream sites are 20%, 50%, and 80%, respectively. Typical probabilities for the downstream sites are 15%, 40%, and 60%, respectively.

Appendix G contains similar results for each dataset (biweekly vs. all) with wet, dry, and all weather samples analyzed separately. Future review and discussion of the results can provide a basis for refining the monitoring program for measuring the effectiveness of specific stormwater controls. Such a program would obviously require a focus on wet-weather sampling and exploration of more complex models to represent relationships between fecal coliform levels, antecedent precipitation, and other factors.

Figure 10: Variance Components and Power Estimates for Biweekly Program

7 Conclusions

Conclusions derived from statistical analyses of AMP monitoring data between 1999 and 2012 are summarized below:

- Fecal coliform concentrations are significantly lower at upstream/rural sites (Dorwin, Velasko, Rte48), as compared with downstream/urban sites (Kirkpatrick, Hiawatha, and Park). Long-term geometric means are <100 cfu/100 ml upstream vs. 300-600 cfu/100 ml downstream.

- Compliance with the 200 cfu/100 ml standard was observed in 70% of the months at the upstream sites, as compared with 15-25% at the downstream sites.

- Achieving compliance with the standard in 90% of the months would require >60% reductions at the upstream sites and >90% reductions at downstream sites. Achieving higher compliance rates would require significantly higher reductions.

- Fecal coliform concentrations exhibit strong seasonality with higher concentrations during the summer months at all stations. Seasonal variations tend to be higher at the upstream sites. Year-to-year variations, which control power for detecting trends, tend to be higher at the downstream sites.

- Fecal coliform levels at all sites are strongly correlated with antecedent precipitation. Levels tend to be 10- to 100-fold higher during wet weather as compared with dry weather. Similar results were observed in previous analysis of bacteria data at lakeshore monitoring sites.

- The relationship between fecal coliform concentrations and antecedent precipitation is stronger than that with flow likely due to the seasonality of flow response, which tends to be lower in the summer when evapotranspiration is higher.

- There is strong evidence of increasing trends at a rate of about 10%/yr under both dry and wet weather conditions at the Dorwin station on Onondaga Creek. Alternative trend test methods and datasets all yielded similar results.

- There is some evidence of increasing trends in long term geometric mean fecal coliforms at RT48 on Ninemile Creek (6.3 %/yr) and Hiawatha at Harbor Brook (4.6 %/yr) based upon data from all weather conditions. The trends become less or not significant for dry or wet weather subsets at both stations.

- Although a significant decreasing trend of -5.1%/yr was observed at the Kirkpatrick station under wet weather, this trend was driven primarily by the inclusion of storm event geometric means, which are likely biasing the dataset towards more extreme wet weather conditions in the earlier part of the record when the majority of storm events samples were collected.

- Trend analysis results are reasonably robust to data subset (biweekly, primary, all) and statistical method.
|
- Variance component estimates provide a basis for evaluating the power of the historical biweekly program for detecting trends or step changes in the data that could reflect implementation of stormwater controls or other anthropogenic factors. Power is relatively insensitive to sampling frequency (weekly, biweekly, monthly) because power is controlled more by random year-to-year variability than by precision of the yearly geometric means.

- The power analysis indicate that the existing monitoring program (biweekly or weekly) is sufficient for measuring long-term progress towards achieving compliance, which would require >60% decreases at the upstream sites and >90% reductions at the upstream sites in order to achieve compliance in >90% of the months.

8 Recommendations

Based on this analysis, we provide the following recommendations for consideration in future sampling efforts:

- The classification of sampling data in the AMP dataset according to study codes causes some confusion and should be reviewed. A number of study codes are very similar and could be combined (e.g. “Quarterly” and “Trib Quarterly”). In addition, the assignment of wet and dry weather events in the study code should be reconsidered as this definition depends on the analysis (e.g. “Trib Biweekly HF” and “Trib Biweekly Dry” could be assigned to “Trib Biweekly”)

- While storm event samples are useful for characterizing extreme wet weather conditions, they must be collected uniformly over time for use in long-term trend analyses to avoid biasing dataset to the periods when events were sampled.

- Refinements to the Metro hourly precipitation measurements are recommended if that is information is to be used in the future to interpret wet-weather tributary data.

- Future review and discussion of the results can provide a basis for refining the monitoring program for measuring the effectiveness of specific stormwater controls. Such a program would obviously require a focus on wet-weather sampling and exploration of more complex models to represent relationships between fecal coliform levels, antecedent precipitation, season, and other factors.

- The historical biweekly program is more than adequate for tracking progress towards achieving the standard, which will require significant reductions in the current levels at all sites. While ultimately needed to assess compliance, the recent (2010-2012) increase from biweekly to weekly (5/month) sampling provides very little additional information, given the current levels. Until significant reductions are achieved, the biweekly data are more than sufficient to assess compliance status. Targeting the additional sampling to wet weather (grab and event-oriented) would be a more cost-effective use of monitoring resources.

9 References

Walker, W.W. (1998). A Statistical Framework for the Onondaga Lake Ambient Monitoring Program. Prepared for Onondaga County Department of Water Environment Protection. July 13, 1998. 14 pages. http://wwwalker.net/onondaga/Statistical%20Framework%20July%2098.PDF
Other statistical framework reports: http://www.wwwalker.net/onondaga

Walker, W.W. and J.D. Walker (2011). Trends in Onondaga Lake Nearshore Monitoring Data Onondaga Lake Ambient Monitoring Program – 2011 Report. Prepared for Ecologic, LLC and Onondaga County Department of Water Environment Protection. July 2011. 19 pages. http://wwwalker.net/onondaga/olamp_nearshore_trends_www_draft_070611.pdf

Helsel, D.R. and R. M. Hirsch (2002). Statistical Methods in Water Resources, Techniques of Water Resources Investigations, Book 4, Chapter A3. U.S. Geological Survey. 522 pages.

10 List of Appendices

Appendix A: Calendar Plots of Study Codes

Appendix B: Storm Event Data

Appendix C: Comparison of Precipitation Datasets

Appendix D: Correlations with Precipitation and Flow

Appendix E: Trend Analysis Results

Appendix F: Trend Analysis Sensitivity

Appendix G: Variance Components