Analysis of Water Chemistry Data Collected Under Maryland's Municipal Separate Storm Sewer System (MS4) Permits: Database, Trends, Challenges, And Recommendations

Prepared for:

Maryland Department of the Environment

Prepared by:

Interstate Commission on the Potomac River Basin

March 2019

Revised June 2019

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

ANOVA	Analysis of Variance
AWQMS	Ambient Water Quality Monitoring System
BMP	Best Management Practice
CWP	Center for Watershed Protection
dt	Detection Limit
EMC	Event Mean Concentration
EMC0	Event mean concentration calculated using flow weighted averages of three discrete storm samples using zero (0) for any samples recorded at less than the detection limit.
EMCdt	Event mean concentration calculated using flow weighted averages of three discrete storm samples using the detection limit (dt) for any samples recorded at less than the detection limit.
EPA	U. S. Environmental Protection Agency
FWA	Flow Weighted Average
ICPRB	Interstate Commission on the Potomac River Basin
MDE	Maryland Department of the Environment
MDP	Maryland Department of Planning
MPN	Most probable number
MS4	Municipal Separate Storm Sewer System
NSQD	National Stormwater Quality Database
QA/QC	Quality Assurance/Quality Control
SHA	State Highway Administration

Units of Measurement

^{0}C	Degree Celsius
hr	hour
in	inch
l	liter
lb	pound
µg	microgram
mg	milligram
yr	year

Executive Summary

Since the 1990's, the Maryland Department of the Environment (MDE) has included a monitoring requirement in the Phase I Municipal Separate Storm Sewer System (MS4) permits they have issued to Baltimore City, nine Maryland counties (i.e., Anne Arundel, Baltimore, Carroll, Charles, Frederick, Harford, Howard, Montgomery, and Prince George's), and the State Highway Administration (SHA). MDE is interested in performing statistical analysis on what is now over 20 years of monitoring data collected under the MS4 program to better characterize stormwater discharges and evaluate watershed restoration activities. This report describes the first steps towards meeting those objectives undertaken by the Interstate Commission on the Potomac River Basin (ICPRB) on MDE's behalf.

Working closely with MDE staff, ICPRB designed a relational database in accordance with the principles of normalization to store the monitoring data submitted by Phase I MS4 jurisdictions and SHA under the requirements of their permits. The database is designed to hold water chemistry, habitat, biological, and physical data in separate tables, linked through other tables identifying sampling activities and monitoring locations. ICPRB staff populated the database with the available water chemistry data submitted by the permittees. All total, almost 97,500 records of chemical and flow parameters were included in the database, taken from nearly 5,000 sampling events at 69 monitoring locations. The following two challenges remain to be addressed:

Challenge 1: Missing data. Some data collected by the MS4 permittees may be missing from the database. Little or no data collected before 2005 were entered into the database from Howard, Montgomery, and Prince George's Counties. It also appears that data collected under the first round of permitting may be missing from other jurisdictions, because far fewer locations were reported than seem to be required to be monitored under first round permits.

Challenge 2: Lack of full QA/QC. The data were not subjected to a full Quality Assurance/Quality Control (QA/QC) review. The biggest QA/QC issue is the apparent large percentage of event mean concentrations (EMC) calculated using censored data (i.e., observations below laboratory detection limits). EMCs were intended to be calculated both by substituting zero for censored observations and by substituting the detection limit for these observations. The actual EMC should lie between these two values, but many EMCs were reported as zero, including those which were supposed to be calculated by substituting detection limits.

Preliminary descriptive statistics (mean, median, standard deviation, minimum, maximum, 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentiles) were calculated for the following chemical parameters at the 69 monitoring locations: BOD, *E. coli* or *Enterococci* species, hardness, nitrite plus nitrate, total kjeldahl nitrogen, total phosphorus, total suspended solids, total copper, total lead, total petroleum hydrocarbons, and total zinc.

The mean EMCs at outfalls monitored under the MS4 program in Maryland were compared to the mean EMCs reported in the National Stormwater Quality Database (NSQD) (Pitt *et al*., 2004) for outfalls in the northeast quadrant of the United States. Analysis of variance (ANOVA) was used to compare the parameter means of the two programs for statistical significance. Most of the Maryland means (i.e., BOD, nitrite plus nitrate, total copper, total kjeldahl nitrogen, total phosphorus, total lead, total suspended solids, total zinc) were significantly lower than those of the NSQD. The means of pH, total petroleum hydrocarbons, and water temperature differed by less than five percent, but the difference was nevertheless significant in the case of pH. Although the means of *E coli* and hardness differed greatly between the Maryland and NSQD, the differences were not significant.

A preliminary analysis was also made of the trends identified through the linear regressions of average EMCs against time. The analysis was restricted to locations with five or more years of monitoring data and to parameters reported at those locations with p-values for their estimated slopes less than 0.1 and less than 50% of the EMCs were censored. Most slopes meeting the criteria are negative, indicating that concentrations of the parameter are decreasing over time. Several long-term locations where data have been collected for more than ten years show decreasing trends for four or more parameters.

These results are preliminary and may change depending on the final treatment of censored data and the resolution of other QA/QC issues. The following actions need to be taken before proceeding to more complex analyses of the data:

1. Obtain any missing data and enter it into the database.
2. Fully QA/QC the water quality data, with a particular focus on evaluating EMCs reported as zero.
3. Where further analysis of the data indicates it is appropriate, adopt alternative statistical methods, such as the nonparametric seasonal Kendall test or maximum likelihood estimators for censored data, to calculate descriptive statistics and trends.

MDE initiated a pilot study in December 2018 which focuses on three watersheds sampled under the MS4 program. The pilot study will address some of the data quantity and quality issues raised in this report, and analyze the relationships among water quality parameters, BMP installation, and watershed restoration. It is hoped that the pilot study will serve as a model for analyzing all watersheds where enough monitoring data have been collected under the MS4 program, and yield recommendations for improving the monitoring component of MS4 permit requirements.

Introduction

Since 1993, the Maryland Department of the Environment (MDE) has included a monitoring requirement in the Phase I Municipal Separate Storm Sewer System (MS4) permits they have issued to nine Maryland counties (i.e., Anne Arundel, Baltimore, Carroll, Charles, Frederick, Harford, Howard, Montgomery, and Prince George's), Baltimore City, and the State Highway Administration (SHA). **Table 1** gives the current list of Phase I MS4 permittees.

Table 1: Maryland Phase I Municipal Separate Storm Sewer System Permittees

Abbreviation	Permittee
AACO	Anne Arundel County
BACI	Baltimore City
BACO	Baltimore County
CACO	Carroll County
CHCO	Charles County
FRCO	Frederick County
HACO	Harford County
HOCO	Howard County
MOCO	Montgomery County
PGCO	Prince George's County
SHA	State Highway Administration

The monitoring requirements have changed over time. In the first round of permits, the goal of the monitoring requirements was characterization of storm sewer discharges, particularly by dominant land use type. Permittees were required to monitor for a variety of water quality parameters at as many as five outfalls in their systems. Monitoring instream locations associated with the outfalls was also required. Around 2000, in the second round of permits, each permittee was required to monitor at only one outfall and one instream location downstream of the outfall, but in addition to water quality monitoring, biological monitoring, habitat assessment, and physical (geomorphic) monitoring were also required downstream of the outfall. The goal was still discharge characterization, with the variation in site characteristics occurring state-wide rather than within each permittee's jurisdiction. Physical monitoring was also required in a second small watershed to assess the effectiveness of Maryland's stormwater control regulations. Starting around 2004, in the third round of permits, while the monitoring requirements remained roughly the same, the goal of the monitoring was redirected to determining the effects of stormwater BMPs and watershed restoration on water quality, habitat, and the health of biological communities. Permittees were directed to monitor watersheds where watershed restoration was anticipated, and pre- and post-implementation conditions could be monitored. Subsequent monitoring permits specify roughly the same monitoring requirements for the same reason: the determination of the effectiveness of watershed restoration. **Table 2** lists the water quality parameters currently required to be monitored under the MS4 program.

Table 2: Water Quality Parameters Required to be Reported by Phase I MS4 Permits

Parameter	Description	Units
BOD	Biochemical Oxygen Demand (BOD5)	mg/l
DEPTH	Total rainfall depth	inch
DURATION	Length of storm event	hour
ECOCCI	Enterococci sp.	MPN/100
ECOLI	*Escherichia coli*	MPN/100
HARD	Hardness	µg/l
INTENSITY	Total depth/length of storm event	in/hr
NO23	Nitrite plus Nitrate	mg/l
pH	pH	std units
TCU	Total Copper	µg/l
TKN	Total Kjeldahl Nitrogen	mg/l
TP	Total Phosphorus	mg/l
TPB	Total Lead	µg/l
TPH	Total Petroleum Hydrocarbons	mg/l
TSQVOL	Total Storm Flow Volume	gallons
TSS	Total Suspended Solids	mg/l
TZN	Total Zinc	µg/l
WTemp	Water Temperature	°C

MDE is interested in using the more than 20 years of monitoring data collected under the MS4 program to perform statistical analysis and answer questions related to the water quality characterization of discharged stormwater and the effectiveness of BMPs and watershed restoration:

- Is water quality discharge by Maryland's MS4s improving over time?
- Can improvements in water quality, habitat, or biological health be correlated with watershed restoration?
- Does the dominant land use type in a catchment lead to differences in concentration in water quality parameters in discharged stormwater?
- Does low density development or minimizing untreated impervious surfaces lead to improved water quality and biological health?

In this project, MDE and ICPRB have accomplished the initial steps of preparing the MS4 monitoring data for statistical analysis, including

1. Designing a database to store the MS4 chemical, biological, and habitat data submitted by MS4 permittee in fulfillment of their permits;
2. Entering available water quality (chemical) data into the database;
3. Using the data to calculate preliminary basic descriptive statistics on the water quality data; and
4. Performing preliminary tests for linear trends in water quality data at each location where data were collected under the MS4 program.

The details of each of these steps are described in the rest of this document.

Database Design

Working closely with MDE staff, ICPRB designed a relation database to store the monitoring data submitted by Phase I MS4 jurisdictions and the SHA under the requirements of their permits. **Appendix A** contains a manual for the database showing the tables, fields in each table, and the relation between tables. The database is generally designed in accordance with the principles of normalization. The data is divided into tables of relatively homogeneous types of information, with an emphasis placed on minimizing storing duplicate information in multiple tables. **Table 3** shows the tables in the database and the type of information stored in them. The records in each table are defined by one or more fields that act as the primary keys or the identifiers for the record. The tables are linked together through these primary keys, as shown in the diagram at the end of the manual.

Table 3: Tables in the Maryland MS4 Database

Table Name	Function
Primary Tables	
ACTIVITY	Records the location, date and time, and type of sampling activities.
ACTIVITY_COMMENT	Tracks activity comments provided by collecting agencies or data manipulations performed by ICPRB.
ACTIVITY_EMC	Stores information about EMC data that may be censored.
ASSESSMENT_CHEMICAL	Includes information about chemical monitoring and event mean concentrations of stormwater discharges from MS4 outfall and instream monitoring locations.
ASSESSMENT_INSITU	Stores in-situ chemical data that was taken in conjunction with biohabitat assessments.
ASSESSMENT_PHYSICAL	Stores information related to geomorphologic stream assessments.
INDEX_BIOLOGY	Contains the index of biological integrity (BIBI) values calculated by the MS4 permittees to assess stream health.
INDEX_HABITAT	Contains habitat indices calculated by the MS4 permittees to assess stream health
MASTER_TAXA_LIST	Provides taxonomic information for all collected macroinvertebrates.
METRIC_BENTHIC	Stores benthic metrics calculated by the MS4 permittees to assess stream health.
METRIC_HABITAT	Stores habitat metrics calculated MS4 permittees to assess overall stream health.
MONITORING_SITES	Provides sampling location names and associated geographic attributes.
OUTFALLS	Provides information about outfalls associated with MS4 permit monitoring.
PROJECT	Provides ad description of the project purpose and/or a summary.
TAXA_COUNT	Contains raw benthic counts submitted by the MS4 permittees.
Domain Tables	
dAgency	Lists sampling agencies.
dActivity_Type	Provides information about the type of sample collected.
dFFGroup	Describes the functional feeding group designation of a benthic organism.
dHabit	Provides a description of the habit/behavior assignment of benthic organisms based on their locomotion or behavior in relation to their habitat.
dIndex_Biohab	Defines biohabitat indices
dLandUse	Provides Maryland Department of Planning (MDP) land use descriptions.
dLifeStage	Lists life stages of benthic organisms.
dMetrics_Benthic	Provides descriptions of benthic metrics calculated by the MS4 jurisdictions.
dMonitoringRequirement	Defines the specific monitoring requirement for an activity.
dOutfallMaterial	Provides information about the outfall material at monitoring locations.
dOutfallType	Provides information about the outfall type at monitoring locations.
dParameter_Chemical	Provides descriptions of chemical assessment parameters.
dParameter_Habitat	Clarifies the habitat assessment parameters.
dParameter_Physical	Describes the characteristics of the physical assessment.
dSiteCriteria	Defines the site selection criteria

Table Name	Function
dStatisticalBase	Describes methods used to calculate the values for the chemical assessment.
dStrata	Defines the physiographic stratum in which a site is located. The three strata used by MBSS for BIBI calculations are Coastal, Piedmont, and Highland.
dQuality	Provides qualitative description of the benthic or habitat sample

As shown in **Database Relationship Diagram** in **Appendix A**, the central table in the database is the Activity table, whose records are sampling events. Chemical, biological, habitat, and physical data are stored in separate primary tables and linked to the Activity table through the Activity_ID. These tables are linked to the corresponding monitoring location and its properties, stored in a separate table, through the Activity table. Many of these primary tables have fields containing codes that are described or defined in detail in related domain tables, whose names are prefix with "d."

MDE has begun to use the Ambient Water Quality Monitoring System (AWQMS) software to store its environmental data. One of the design criteria for the MS4 database is to make it as compatible with AWQMS as possible, without sacrificing the flexibility needed to facilitate data analysis.

Data Compilation

ICPRB populated the database with water quality data provided by MDE. Data were entered into the primary tables: ACTIVITY, CHEMICAL_ASSESSMENT, MONITORING_SITES, OUTFALL, and PROJECT. To summarize, 4,802 sample events were entered in the ACTIVITY table, while 97,478 observations of chemical and flow parameters were entered in the CHEMICAL_ASSESSMENT from the 69 monitoring locations recorded in MONITORING_SITES table. Data were also entered into PROJECT, OUTFALL, ACTIVITY_EMC and NOTES_EMC tables.

Table 4 lists the location of the monitoring sites, whether they are outfall or instream locations, and their periods of record. Thirty-nine of the locations were instream while 24 were described as outfall or outflow. The remainder were labeled as "in pond," or "Inflow," or not labeled and entered as "unknown" into the database.

Table 4: Monitoring Locations where Water Quality Data were Collected under Maryland's MS4 Program

Jurisdiction	Project Location	Location Name	MDE Location ID	Local ID	Location Type	Date Range
Anne Arundel County	Parole Town Center	Parole Plaza	AA94MSI000008		Outfall	1999 - 2016
		Church Creek	AA94MSI000009		Instream	
Baltimore City	Moore's Run	Hamilton Avenue	BC16MS000I81		Outfall	1999 - 2016
		Radecke Avenue	BC16MS000I82		Instream	
Baltimore County	Long Quarter Branch		BA98MSI000LQ2	LQ2	Unknown	1998 - 2000
		Dulaney Valley Road	BA98MSI000LQ3	LQ3	Unknown	
	Powder Mill Run		BA07MSI000PM01	PM01	Instream	2007 - 2009
			BA07MSI000PM02	PM02	Instream	
			BA07MSI000PM03	PM03	Instream	2008 - 2009
			BA07MSI000PM03A	PM03A	Instream	2007
	Spring Branch	Near Potspring Rd	BA98MSI000SB2	SB2	Instream	1998 - 2002

Jurisdiction	Project Location	Location Name	MDE Location ID	Local ID	Location Type	Date Range
		In headwater	BA98MSI000SB3	SB3	Outfall	
	Scotts Level Branch		BA05MSI000SL00	SL00	Instream	2009 - 2010
			BA05MSI000SL01	SL01	Instream	2005 - 2010
			BA05MSI000SL02	SL02	Instream	
			BA05MSI000SL03	SL03	Instream	
			BA05MSI000SL04	SL04	Instream	
			BA05MSI000SL05	SL05	Instream	
			BA05MSI000SL06	SL06	Instream	
			BA05MSI000SL07	SL07	Instream	
			BA05MSI000SL08	SL08	Instream	
			BA05MSI000SL09	SL09	Outfall	
			BA05MSI000SL10	SL10	Instream	
Carroll County	Airpark Business Center	Trib to W. Br. Patapsco River	CR15MSI000003		Outfall	2000 - 2016
			CR15MSI000004		Instream	
Charles County	Zekiha Swamp		CH99MSI000096	96.5	Outfall	1999 - 2006
			CH99MSI000162	162.5	Instream	1999 - 2005
	Acton/Hamilton		CC15MSI000002	AH002	Outfall	2006 - 2016
			CC15MSI000001	AH001	Instream	
Frederick County	Peter Pan Run	Peter Pan Run	FR99MSI000058	PPAN-101	Instream	1999 - 2016
		Pond-R	FR02MSI000059	POND-R	Outfall	2003 - 2016
Harford County	Wheel Creek	Winters Run	HA99MSI000002	002	Outfall	1999 - 2009
			HA99MSI000001	001	Instream	
		Wheel Creek	HA10MSI0WC003	WC003	Outfall	2010 - 2016
			HA10MSI0WC002	WC002	Instream	
			HA10MSI0WC004	WC004	Instream	
			HA99MSI000001	001	Instream	
Howard County	Front Hill Tributary	Burnside	HO03MSI0000F3		Outfall	2003
		Carrigan	HO03MSI0000F2		Outfall	
		Font Hill	HO03MSI0000F1		Instream	
	Centennial Lake	Centennial Lake	HO07MSI0000CL		Instream	2007 - 2009
	Wilde Lake	Wilde Lake	HO07MSI000035	WL1	Instream	2007 - 2016
	Red Hill Branch	Meadowbrook Park Restoration Area	HO09MSI000036	MB1	Instream	2009 - 2016
		Dorsey Hall	HO14MSI000041	PT1	Instream	2015 - 2016
			HO14MSI000042	CR1	Instream	
		Bramhope Restoration Study Area/ Brampton Hills	HO10MSI000037	BH01	Outfall	2015 - 2016
			HO10MSI000038	BH02	Instream	
		Salterforth Pond BMP	HO10MSI000039	SF01	Outfall	2015 - 2016
		Retrofit	HO10MSI000040	SF02	Outfall	
Montgomery County	Lower Paint Branch / Stewart-April Lane Trib.	Stewart-April Lane	MO02MSI000104	PBPB104	Outfall	2006 - 2008
		Lower Paint Branch	MO02MSI000310	PBPB310A	Instream	
	Lower Paint Branch / Breewood	Breewood	MO09MSI000001	scbtoutfall1	Outfall	2009 - 2015
			MO09MSI000002	scbtinstream1	Instream	
Prince George's County	Bear Branch	Chapel Cove Drive	PG15MSI000013	PG005	Instream	2008 - 2016
		Contee Road	PG15MSI000012	PG003	Instream	
State Highway Administration	Pindell School Road, Howard County		SHA98MSI00PS01	PS01	Inflow	1998 - 2001
			SHA98MSI00PS02	PS02	Outflow	
	Dulaney Valley Rd, Baltimore County		SHA99MSI00DV01	DV01	Outfall	1999 - 2003
			SHA99MSI00DV02	DV02	Instream	

Jurisdiction	Project Location	Location Name	MDE Location ID	Local ID	Location Type	Date Range
	Long Draught Branch, Montgomery Co.		SHA06MSI00LD01	LD01	Instream	2006 - 2008
			SHA06MSI00LD02	LD02	Instream	
	Mt. Rainier LID Site, Prince George's Co.		SHA02MSI00MR01	MR01	Outflow	2002 - 2003
	Grass Swale LID Site		SHA04MSI00GS01	GS01	Outflow	2007 - 2009
			SHA04MSI00GS02	GS02	Outflow	
			SHA04MSI00GS03	GS03	Outflow	
	Wet Infiltration Basin LID Site		SHA08MSI00WI01	WI01	In Pond	2009 - 2012
			SHA08MSI00WI02	WI02	Inflow	
			SHA08MSI00WI03	WI03	Outflow	

Table 5 shows the years in which data were collected for each "project," which designates the monitoring requirement and/or catchment in which an outfall and instream monitoring location were sited. Some of these catchments, like Scotts Level Branch in Baltimore County or Front Hill Tributary in Howard Count, contain more than one outfall or instream monitoring location (see **Table 4**).

The longest periods of record, 18 years, were in Anne Arundel County (Parole Town Center) and Frederick County (Urbana). Baltimore City and Carroll County had locations with 17 years of record, in Moore's Run and near the Airpark Business Center, respectively. It does not appear that much data were made available from the first round of permits, at least not at all the locations monitored, and little data appears to be in the dataset from the second round in Howard, Montgomery, and Prince George's Counties.

Appendix B gives the sample count by year at each monitoring location. The samples are divided into stormflow and baseflow samples. Stormflow samples should have event mean concentrations (EMCs) calculated from three samples taken on the rising limb, peak, and falling limb of the storm hydrograph. Flow characteristics like rain depth, storm duration, and storm intensity, should also be reported. It was expected that baseflow samples would be simple grab samples but see discussion below.

Table 5: Summary of Monitoring Activities Included in the Maryland MS4 Database.
The color ramp (dark green to yellow) provides a visual interpretation of the number of years data was collected for each project.

Jurisdiction	Project Location	Count (Years)	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Anne Arundel County	Parole Town Center	18		y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
Baltimore City	Moore's Run	17		y	y	y	y	y	y	y	y	y	y	y	y	y	y		y	y	y
Baltimore County	Spring Branch	5	y	y	y	y	y														
	Scotts Level Branch	6								y	y	y	y	y	y						
	Long Quarter Branch	3	y	y	y																
	Powder Mill Branch	3										y	y	y							
Carroll County	Airpark Business Center	17			y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
Charles County	Zekiha Swamp	8		y	y	y	y	y	y	y	y										
	Acton/Hamilton	9									y	y	y	y	y	y		y		y	y
Fredrick County	Peter Pan Run	18		y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
Harford County	Winters Run	11		y	y	y	y	y	y	y	y	y	y	y							
	Wheel Creek	7													y	y	y	y	y	y	y
Howard County	Front Hill Tributary	1						y													
	Red Hill Branch / Meadowbrook Park	8												y	y	y	y	y	y	y	y
	Red Hill Branch / Dorsey Hall	2																		y	y
	Red Hill Branch / Bramhope	2																		y	y
	Red Hill Branch / Salterforth Pond	2																		y	y
	Centennial Lake	3										y	y	y							
	Wilde Lake	10										y	y	y	y	y	y	y	y	y	y
Montgomery County	Lower Paint Branch/Stewart-April Lane	3									y	y	y								
	Lower Paint Branch/Breewood	7												y	y	y	y	y	y	y	
Prince Georges County	Bear Branch	9											y	y	y	y	y	y	y	y	y
State Highway Administration	Pindell School Road, Howard Co.	4	y	y	y	y															
	Dulaney Valley Rd, Baltimore Co.	5		y	y	y	y	y													
	Long Draught Branch, Montgomery Co.	3									y	y	y								
	Mt. Rainier LID Site	2					y	y													
	Grass Swale LID Site	3										y	y	y							
	Wet Infiltration Basin LID Site	4												y	y	y	y				

Data Quality

The data were not subject to a full QA/QC review, but several problems emerged while entering the data, including the following:

- **Event Mean Concentration**: Two EMCs were required to be calculated by the jurisdictions via the flow weighted averages (FWA) of three discrete storm samples taken on the rising limb, peak, and falling limb of the storm hydrograph using a) zero (0) and b) the detection limit value (dt) for any discrete samples recorded at less than the detection limit, denoted as EMC0 and EMCdt respectively in the Access database. Any deviations from this method noted by a jurisdiction was recorded in the in the ACTIVITY_COMMENTS table within the MS4 database.
- **Units:** At times, the units were inconsistent during a given collection period and often not provided. This was especially true for water temperature and metals measurements. Any water temperatures that were recorded at 30 degrees or higher for a sampling season, were deemed to be in Fahrenheit and converted to Celsius. This action was noted in the ACTIVITY_COMMENTS table.
  The units for monitored metals were reported either in mg/l or µg/l. When the unit was not provided, the data for an entire sampling season were examined to determine whether conversions were necessary. All metals are reported in µg/l in the database, and any conversions were recorded in the ACTIVITY_COMMENTS table.
- **Zero Values:** There were instances were rain depth, storm duration, storm intensity, and total storm flow volume were reported as “zero” for baseflows. These zero values were deleted. On the other hand, there were also occurrences were values were reported as greater than zero during baseflows. These entries were retained, but it is uncertain whether they are valid. As stated in the Chemical Monitoring section of the MS4 permits “*If extended dry weather periods occur, baseflow samples shall be taken at least once per month at the monitoring stations if flow is observed*.” In other words, because baseflow sampling should occur during dry weather periods, there should not be any storm related entries.
  There were also a few cases were rain depth, storm duration, storm intensity, and total storm flow volume were reported as “zero” for storm flows. These values were retained but are likely invalid.
  In addition, there were cases were parameter values for the detection limit, EMC0, and EMCdt were reported as “zero” for an entire year. It was assumed that the parameter had not been monitored for that year and these entries were deleted so that these zero values would not bias statistical and trend analyses. The action was recorded in the ACTIVITY_COMMENTS table.
- **Monitoring Locations**: The current database contains 69 monitoring locations. Of these 20 are lacking coordinates, and two have incorrect coordinates.
- **Outfalls**: There are a total of 18 monitoring sites of type “outfall” in the database. Of these, only four had outfall identifiers that allowed them to be matched to the outfall data in the MDE MS4 geodatabase.

Table 6 documents additional issues specific to permittees and projects and/or locations.

Table 6: Problems with Water Quality Data Specific to Permittees

Jurisdiction	Location	Year(s)	Issue
Anne Arundel County	Parole Town Center	2010	No EMCs. The county entered three values for each water chemistry parameter per storm event (presumably at the rising limb, peak, and falling limb of the hydrograph), but only the total storm flow volume. Therefore, EMCs could not be calculated retroactively.
		Prior to 2006	Missing baseflow data.
Baltimore City	Moore's Run	2013	Missing data.
Baltimore County	LQ		Only two years of data. The location, flow type, and purpose were not stated by the county.
	PM		Location was not reported.
	Spring Branch	1998 - 2001	Flow type, detection limits, and units were not stated.
		2003 - 2004	Statement by county: "*There are only three storms in the dataset and therefore this is not being reported this year.*" (See READ ME.doc in data folder supplied by MDE: Baltimore County/2005/Databases/)
		2008 - 2016	Monitoring data exists, but were received as raw data (i.e., not as EMCs), as trends and/or regressions.
Charles County	Acton/Hamilton	2011, 2013	Missing data.
Harford County	Winters Run	1999 - 2002	EMC were reported at ½ the detection limit and at zero. To keep entries consistent in the database, EMCs at ½dt were not included.
		2009	Data were extracted from Appendix A: Water Quality and Rainfall Data of the 2010 *Chemical Data Analysis Ambient Station / Unnamed Tributary to Winters Run Harford County, Maryland* report.
Howard County		1999	Missing EMC data (only raw data were provided).
		Prior to 2007	Missing data.
	Centennial Lake	2007 - 2016	Only one site.
	Wilde Lake		Only one site.
Montgomery County		Prior to 2006	Missing data.
State Highway Administration		2004, 2005 2009 - 2016	Missing data.

Censored Data

The identification of censored data can illustrate the QA/QC issues with the water chemistry data. As described above, the permittees were supposed to report storm event concentrations as two types of EMCs: (1) EMC0, where zero was used for values below the detection limit, and (2) EMCdt, where the detection limit itself was used for values below the detection limit. If these rules were followed, we would expect the following:

1. The data used in the EMC calculation are uncensored if EMC0 = EMCdt;
2. At least one observation was below the detection limit if EMCdt > EMC0. In this case, the data is censored, and the EMC itself is below the value reported by EMCdt; and
3. EMCdt is never less than EMC0.

There are some cases in which the data submitted does not conform to these rules. Sometimes no data was entered for either EMC0 or EMCdt. Sometimes a zero was entered, which for EMC0 could be a legitimate value if all the data used in the EMC calculation were below the detection limit; however, it is not a legitimate value for EMCdt. Because some software produces zeroes as default data entry, it is also possible that the zeros are null values.

Legitimate or otherwise, there are three possible types of values for EMC0 and EMCdt that were used in the submitted data: (1) blank, (2) zero, or (3) a number larger than zero, which means that there are nine possible pairs of value types. **Table 7** divides the sampling events with submitted EMC data among the types. If both EMC fields were blank, the record does not appear in the database, so that outcome is not shown in **Table 7**.

Table 7: Characterization of Paired EMC0 and EMCdt Values

Case	Condition	Valid Entry[1]	Censor Type	Record Count
1	EMC0 = NULL AND EMCdt = 0	No	Considered uncensored	2
2	EMC0 = NULL AND EMCdt > 0	No	Considered uncensored	51
3	EMC0 = 0 AND EMCdt = NULL	No	Inconclusive	41
4	EMC0 > 0 AND EMCdt = NULL	No	Considered uncensored	1,776
5	EMC0 = 0 AND EMCdt = 0	No	Uncensored	2,430
6	EMC0 = 0 AND EMCdt > 0	Yes	Censored	11,240
7	EMC0 > 0 AND EMCdt = 0	No	Considered uncensored	121
8	EMC0 = EMCdt (Both > 0)	Yes	Uncensored	22,267
9	EMC0 < EMCdt (Both > 0)	Yes	Censored	3,462
10	EMC0 > EMCdt (Both > 0)	No	Considered uncensored	292
			Total	**41,682[2]**

[1]Null and EMCdt = 0 are invalid and data need to be requested from the data provider.
[2]There is one additional record probably introduced by using exact equality in Case 7.

The rules are clearly not followed in about 11% of the records. These include all records where only one EMC is entered (Cases 1 through 4), and the cases where EMCdt is zero but EMC0 is either zero (Case 5) or a number larger than zero (Case 7). Case 6 is a legitimate case, which can occur when all samples used in the EMC calculation are below the detection limit; however, the presence of zeros in records that don't conform to the rules may indicate that some records in Case 6 are really examples of records where only one EMC value was entered for the event.

Table 7 also distinguishes three cases in which both fields were values: Case 8: EMC0 = EMCdt, which is the uncensored case; Case 9: EMC0 < EMCdt, which implies that the EMC is below the reported values and is thus censored; and Case 10: EMC0 > EMCdt, which should not occur. Case 8 was implemented using strict equality, which is inappropriate for floating point number calculations but in this case serves as a test for whether the same numerals were entered for EMC0 and EMCdt. Many of the records in Cases 9 and 10 are likely cases of equality in which one of the two EMCs was truncated or rounded off. A visual inspection suggests that most of the illegitimate Case 10 records are examples of truncation or rounding. Differences in EMC values are small both in an absolute sense and relative to the size of the values, and it is difficult to devise a numerical test to determine which of the records where both values are present but not strictly equal should be considered uncensored. For example, in about in 50 records in Case 10, the difference in the values is less than 0.01%, but many records both above and below this threshold appear to be cases of truncation or rounding.

Keeping these issues in mind, the percent of censored EMCs (EMCs below the value reported in EMCdt) were calculated for each parameter and monitoring location. An EMC was considered censored if it belonged to either Case 6 or Case 9. Case 1 was considered uncensored, because it would be same as Case 5 if EMC0 value was zero, or it would mimic Case 7 if the EMC0 value was greater than zero. It is impossible to assign a potential censor type to Case 3. If the EMCdt value was equal to zero, it would be the same as Case 5 (uncenscored), but if the EMCdt value was greater than zero, it would be the same as Case 6 (censored). For the purpose of this study, all records where either EMC0 or EMCdt had values greater than zero were included in the analysis, including Cases 2, 4, 7, and 10, where none of these records were counted as censored.

Table 8 shows the results. Because some EMCs in Case 6 and 9 may be examples of truncation, rounding, or automated entry of zeros for missing data, the amount of censoring is likely overstated. Nevertheless, it indicates that potentially more than a third of the data for many parameters at many locations may be censored and require the use of special statistical methods for censored data. These parameters include not only metals like copper and lead, but conventional pollutants like BOD, TKN, TP, and TSS.

Table 8: Estimated Percent of EMCs Calculated with Censored Observations.

Location Id	BOD	ECOCCI	ECOLI	HARD	NO23	TCU	TKN	TP	TPB	TPH	TSS	TZN
AA94MSI000008	23%		5%	0%	6%	21%	50%	8%	62%	75%	6%	1%
AA94MSI000009	25%		3%	2%	4%	40%	48%	12%	57%	77%	6%	2%
BA05MSI000SL00	100%			0%	0%	100%	100%	100%	100%		75%	0%
BA05MSI000SL01	71%		0%	0%	8%	42%	33%	55%	78%	40%	66%	30%
BA05MSI000SL02	100%			0%	0%	33%	17%	67%	100%		72%	33%
BA05MSI000SL03	83%			0%	0%	33%	39%	61%	94%		72%	22%
BA05MSI000SL04	100%			0%	6%	44%	33%	72%	100%		67%	28%
BA05MSI000SL05	94%			0%	0%	33%	44%	50%	78%		78%	17%
BA05MSI000SL06	94%			0%	0%	35%	24%	53%	100%		88%	12%
BA05MSI000SL07	100%			0%	0%	64%	27%	82%	100%		64%	27%
BA05MSI000SL08	100%			0%	0%	41%	47%	76%	94%		82%	29%
BA05MSI000SL09	55%		0%	0%	10%	46%	31%	48%	68%	67%	66%	22%
BA05MSI000SL10	100%			0%	0%	41%	53%	71%	94%		65%	18%
BA07MSI000PM01	80%			0%	0%	30%	40%	50%	80%		50%	40%
BA07MSI000PM02	100%			0%	0%	57%	71%	71%	100%		86%	43%
BA07MSI000PM03	100%			0%	0%	50%	25%	75%	100%		100%	50%
BA07MSI000PM03A	100%			0%	0%	0%	100%	50%	100%		100%	0%
BA98MSI000LQ2	13%				0%	0%	10%	0%	40%		0%	0%
BA98MSI000LQ3	60%				8%	0%	0%	0%	75%		50%	0%
BA98MSI000SB2	70%				0%	13%	6%	22%	67%	100%	59%	23%
BA98MSI000SB3	52%				0%	13%	0%	13%	43%	100%	33%	0%
BC16MS000I81	31%	100%	0%	0%	0%	15%	11%	0%	45%	85%	9%	29%
BC16MS000I82	34%	0%	0%	0%	0%	16%	9%	2%	49%	84%	4%	35%
CC15MSI000001	41%		0%	0%	51%	65%	14%	20%	92%	88%	22%	4%
CC15MSI000002	46%		0%	0%	44%	61%	19%	20%	80%	91%	28%	2%
CH99MSI000096	19%				10%	41%	20%	4%	87%	98%	35%	9%
CH99MSI000162	26%				19%	38%	3%	1%	100%	98%	25%	15%
CR15MSI000003	38%		15%	0%	12%	32%	34%	3%	93%	94%	4%	25%
CR15MSI000004	61%		1%	0%	0%	48%	53%	6%	87%	94%	7%	44%
FR02MSI000059	42%		0%	5%	10%	19%	43%	2%	95%	90%	14%	16%
FR99MSI000058	80%		0%	1%	13%	61%	68%	12%	87%	90%	17%	47%

Location Id	BOD	ECOCCI	ECOLI	HARD	NO23	TCU	TKN	TP	TPB	TPH	TSS	TZN
HA10MSI0WC002	70%		0%	0%	1%	38%	41%	60%	82%	97%	60%	8%
HA10MSI0WC003	75%		0%	0%	2%	40%	38%	57%	80%	96%	49%	8%
HA10MSI0WC004	78%		0%	0%	4%	31%	38%	51%	78%	97%	61%	7%
HA99MSI000001	46%		1%		14%	28%	53%	38%	44%	59%	37%	21%
HA99MSI000002	46%		0%		20%	31%	49%	33%	44%	60%	38%	10%
HO03MSI0000F1	25%				11%	33%	78%	0%	67%	100%	0%	33%
HO03MSI0000F2	0%				33%	0%	78%	0%	89%	100%	0%	0%
HO03MSI0000F3	29%				38%	13%	88%	0%	88%	100%	0%	13%
HO07MSI000035	24%		0%	43%	4%	33%	59%	26%	50%	100%	17%	22%
HO07MSI0000CL	67%				0%	100%	83%	89%	100%	100%	56%	89%
HO09MSI000036	27%		0%	0%	4%	24%	57%	17%	55%	100%	9%	7%
HO10MSI000037					0%		91%	9%			9%	
HO10MSI000038					0%		100%	9%			9%	
HO10MSI000039					71%		71%	0%			0%	
HO10MSI000040					43%		71%	0%			0%	
HO14MSI000041					44%		78%	0%			0%	
HO14MSI000042					0%		100%	0%			0%	
MO02MSI000104	74%			0%	3%	2%	25%	87%	48%	89%	37%	5%
MO02MSI000310	70%			0%	0%	3%	41%	75%	40%	93%	45%	5%
MO09MSI000001	44%	5%		0%	35%	25%	26%	95%	84%	63%	0%	13%
MO09MSI000002	79%	0%		0%	13%	26%	65%	99%	93%	86%	14%	8%
PG15MSI000012	39%		10%	0%	10%	30%	8%	40%	65%	95%	7%	12%
PG15MSI000013	39%		9%	0%	15%	17%	10%	40%	67%	99%	14%	18%
SHA02MSI00MR01					100%	100%	100%	100%	100%		100%	100%
SHA04MSI00GS01						100%	100%	100%	100%		100%	100%
SHA04MSI00GS02						100%	100%	100%	100%		100%	100%
SHA04MSI00GS03						100%	100%	100%	100%		100%	100%
SHA06MSI00LD01	15%		0%		0%	85%	0%	11%	100%	80%	32%	20%
SHA06MSI00LD02	20%		0%		0%	85%	10%	5%	100%	80%	26%	30%
SHA08MSI00WI01					100%	100%	100%	100%	100%		100%	100%
SHA08MSI00WI02					100%	100%	100%	100%	100%		100%	100%
SHA08MSI00WI03					100%	100%	100%	100%	100%		100%	100%
SHA98MSI00PS01	15%				0%	19%	38%	0%	100%	92%	0%	15%
SHA98MSI00PS02	42%				3%	58%	39%	0%	100%	100%	13%	42%
SHA99MSI00DV01	6%				0%	3%	6%	0%	78%	91%	0%	0%
SHA99MSI00DV02	63%				2%	63%	33%	22%	92%	100%	39%	51%
Median	**54%**	**3%**	**0%**	**0%**	**3%**	**35%**	**41%**	**33%**	**88%**	**93%**	**35%**	**20%**

Descriptive Statistics

The following descriptive statistics were calculated for each water quality parameter and monitoring:

- Mean
- Standard deviation
- Maximum and Minimum
- Median
- 10th, 25th, 75th and 90th percentiles

For many parameters, both EMC0 and EMCdt were available for a sampling event, and in these cases the two values were averaged. Both baseflow and stormflow samples were included in the location statistics. No attempt was made to take into account the censoring of the data in the EMC calculations. The results are found in **Appendix C**. Given the discussion both of the QA/QC issues and the high frequency of data below the detection limit in the previous section, these results are only preliminary.

The average value of the sampling events for each parameter and monitoring location are displayed graphically in two ways. **Appendix D** shows boxplots of the average EMC values by year and **Appendix E** shows scatterplots of the data versus time. The scatter plots also show the estimated linear trend with time, which is discussed in the next section. Both appendices are in pdf format but are searchable by parameter or monitoring location. The figures are useful for identifying outliers. See, for example, the BOD figure for CH99MSI000096 or the Total Phosphorus figure for AA94MSI000009 in **Appendix D**.

To put the descriptive statistics in context, at least for monitoring locations at outfalls, the average values from each outfall location in the MS4 database were compared against the outfall averages from the National Stormwater Quality Database (NSQD) for the Environmental Protection Agency's (EPA) Rainfall Regions 1 and 2 (Pitt et all, 2004). The combined area of these regions covers the Northeastern United States, from Maine to Minnesota and north from Tennessee and North Carolina, as shown in **Figure 1**. Maryland is in Zone 2. To make the data comparable, the detection limit was used for NSQD data, when data was below the reported value, while only EMCdt values were used from the Maryland MS4 database. Only stormflow data were used in the Maryland calculations. Contrasting boxplots of the site averages from the NSQD and Maryland MS4 database are given for each parameter in **Appendix F**.

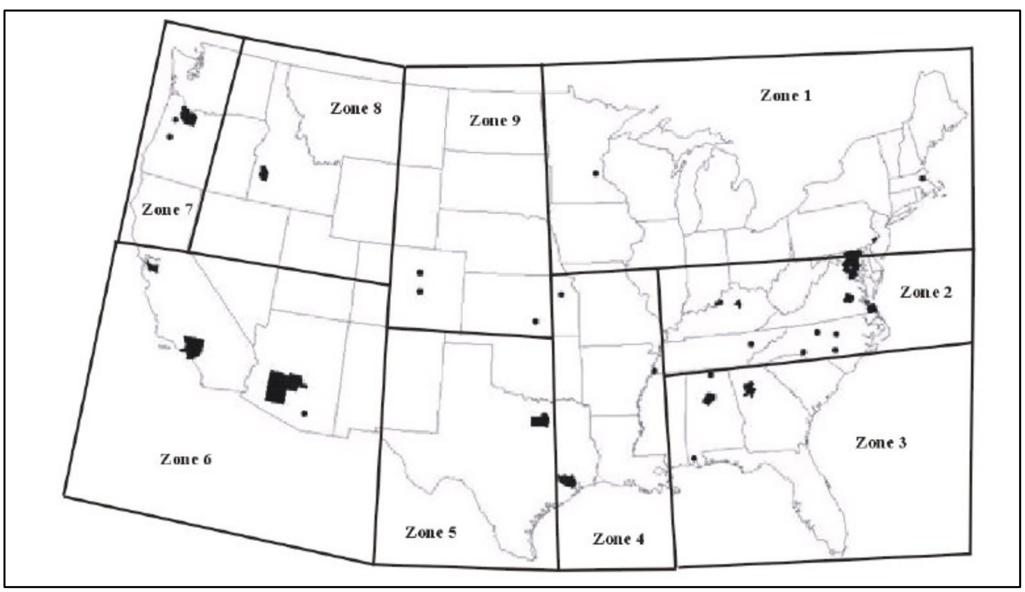

Figure 1: Communities Contributing Monitoring Data to the NSQD by EPA Rain Zone (from Pitt et all, 2004).

Generally, the Maryland location means are lower than or comparable to the NSQD. The differences in the mean values for each parameter as well as the significance levels of one-way ANOVA tests are provided in **Table 9**. As can be seen, eight of the thirteen means for the MS4 parameters were significantly lower than those of the NSQD, with a difference of greater than five percent. Although the difference in the mean values for *E. coli* and hardness were large, the ANOVA tests were not significant. The difference between the mean for total petroleum hydrocarbons and water temperature were low and not significant. Even though the difference between the mean for pH was small (3%), it was highly significant. Again, these results are preliminary, given the caveats that the Maryland MS4 data have not been fully reviewed and a more careful treatment of censored data is required because of the high degree of censoring in the dataset.

Table 9: Comparison of Parameter Averages between the Maryland MS4 and NSQD Data.

	MS4 Data		NSQD		Percent Difference between Averages	ANOVA		
Parameter	n	Mean	n	Mean		F	Pr (>F)	Significance Level
BOD	781	9.78	3,127	14.03	-44%	24.6	7.35E-07	0.001
ECOLI	262	24,353	25	6,897	72%	0.523	0.47	
HARD	222	58.09	280	74.70	-29%	3.098	0.079	0.1
NO23	821	0.73	3,628	0.92	-26%	10.08	1.51E-03	0.01
PH	710	7.24	1,031	7.42	-3%	12.7	3.76E-04	0.001
TCU	792	14.51	2,901	28.46	-96%	73.1	< 2e-16	0.001
TKN	850	1.35	3,724	1.84	-36%	42.68	7.15E-11	0.001
TP	851	0.24	6,192	0.32	-31%	16.58	4.71E-05	0.001
TPB	751	10.98	2,460	27.17	-147%	51.82	7.53E-13	0.001
TPH	612	2.36	39	2.45	-4%	0.05	0.823	
TSS	849	59.10	4,199	114.17	-93%	66.73	3.91E-16	0.001
TZN	818	80.87	3,303	158.68	-96%	41.76	1.15E-10	0.001
WTEMP	527	15.28	461	15.07	1%	0.232	0.63	

Preliminary Trend Analysis

A preliminary trend analysis was performed using simple linear regression. For each parameter and monitoring location in the dataset used to calculate descriptive statistics, a linear regression was performed with the parameter as the dependent variable and time as the independent variable. When both EMC0 and EMCdt were available for a sampling event, the two values were averaged. Fitted values are graphed in the scatter plots in **Appendix E**. The slope of the regression line and the p-value of the slope are also shown on each figure. The time variable used in R, where these regressions were calculated, is in seconds, so the slopes were converted to units/year to make the values more intelligible. **Appendix G** also reports these statistics as well as the regression's coefficient of determination (R^2), the estimated intercept of the line, and the p-value of the intercept. The same caveats apply to these regressions that applied to the descriptive statistics: the data has not had a full QA/QC review, and the censoring of the data has been ignored.

The trend analysis was simplified by analyzing slopes for monitoring locations and parameters that met the following criteria:

- Five or more years of monitoring data
- Slope p-value less than 0.1
- 50% or less of the data is censored

These criteria are probably less stringent than would be required by a more critical analysis. In particular, using special methods for censored data instead of ordinary least squares would certainly be recommended even if the rate of censoring was much lower than 50%, but given the uncertainties in the level of censoring that was discussed previously, it seemed appropriate for a screening analysis to use a less stringent criterion.

The results are shown in **Table 10**, where negative slopes are highlighted in green and coefficients of determination (R^2) of 0.5 or greater are indicated in red font. The level of censoring for each location and parameter is provided in **Table 8**. If the goal of the analysis is to demonstrate that conditions are improving, the results are encouraging, although decreases in hardness probably do not indicate improving conditions. As **Table 10** shows, for the data that met the criteria, the majority of the slopes were negative. Except for hardness and BOD, each parameter had negative slopes at three quarters or more of the locations that met the criteria. Several long-terms locations, such as AA94MSI000009, FR02MSI000059, and HA99MSI000002, had five or more parameters (excluding hardness) showing negative trends. Given the caveats already expressed about both the dataset and the analysis, it is worth restating that these results would have to be confirmed by a more careful analysis of data, as discussed in the next section.

Table 10: Linear Regression Slopes (units/year) for Parameters with Five or More Years of Data, p-values Less than 0.1, and Less than 50% Censoring.[1]

		Significant	BOD (mg/l)		ECOLI (MPN/100)		HARD (µg/l)		NO23 (mg/l)		TCU (µg/l)		TKN (mg/l)		TP (mg/l)		TSS (mg/l)		TZN (µg/l)	
Location	**Years**	**Parameters**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**	**Slope**	**R2**
AA94MSI000009	18	6	-0.503	0.080	-9,983	0.070							-0.050	0.034	-0.079	0.026	-4.332	0.105	-2.480	0.020
BA05MSI000SL02	6	1									-0.002	0.247								
BA05MSI000SL03	6	1									-0.001	0.339								
BA05MSI000SL04	6	1									-0.002	0.188								
BA05MSI000SL06	6	1									-0.003	0.533								
BA05MSI000SL07	5	1							-0.045	0.318										
BA05MSI000SL08	6	2							-0.294	0.179	-0.002	0.568								
BA05MSI000SL10	6	1									-0.002	0.524								
BA98MSI000SB2	5	2									6.822	0.382							3.929	0.504
BA98MSI000SB3	5	2									9.949	0.514							9.356	0.565
BC16MS000I81	17	5	0.558	0.059	-10,788	0.066					-0.375	0.015					1.970		-1.574	0.012
BC16MS000I82	17	4			-8,849	0.070	-22	0.095			-0.232	0.010							-1.771	0.020
CC15MSI000001	9	3	0.688	0.053			-6	0.204									3.614	0.185		
CC15MSI000002	9	2					7	0.274	-0.042	0.083										
CH99MSI000096	8	3							0.140	0.073					-0.040	0.062			-10.217	0.063
CH99MSI000162	7	2											-0.138	0.183	-0.033	0.088				
CR15MSI000003	17	6					-272	0.321	-0.032	0.037	-0.535	0.112	0.027	0.035	-0.010	0.075	-3.168	0.044		
CR15MSI000004	17	2					-76	0.237			-0.386	0.173								
FR02MSI000059	14	7			-640	0.277	7	0.085	-0.037	0.093	-0.327	0.070			-0.012	0.054	-1.705	0.039	-1.260	0.040
FR99MSI000058	18	2					3	0.311	-0.048	0.049										
HA10MSI0WC004	7	1					-54	0.089												
HA99MSI000001	11	2									-0.616	0.177							-0.660	
HA99MSI000002	11	5									-0.512	0.198	-0.027	0.031	-0.007	0.072	-2.022	0.022	-1.925	0.119
HO07MSI000035	10	4	0.381	0.029			119	0.675	-0.048	0.066	0.298									
HO09MSI000036	8	4	-0.591	0.062	3,260	0.765	136	0.587	-0.066	0.067										
MO09MSI000002	7	4					10	0.131	-0.264	0.119							-8.185	0.024	1.960	0.023
PG15MSI000012	9	2									-0.382	0.024	-0.180	0.167						
PG15MSI000013	9	5									-1.027	0.052	-0.206	0.181	-0.012	0.086	-24.882	0.058	-3.876	0.106
SHA99MSI00DV01	5	5							-0.251	0.244	-5.897	0.123	-1.212	0.158			-14.378	0.133	-43.559	0.094
SHA99MSI00DV02	5	1											0.187	0.072						
Percent negative slope			**40%**		**80%**		**45%**		**91%**		**84%**		**75%**		**100%**		**78%**		**75%**	

[1]Negative slopes are highlighted in green and coefficients of determination (R^2) of 0.5 or greater are indicated in red font

Recommendations and Next Steps

This project has taken the first steps toward an analysis of the water quality data collected under Maryland's MS4 program. These steps need to be reviewed and finalized by MDE and ICPRB. In particular, the following actions need to be taken before proceeding to more complex analyses of the data:

1. **Obtain available missing data.** Little or no data were entered into the database from Howard, Montgomery, and Prince George's Counties collected before 2005 (**Table 5**). It also appears that data collected in the first round may be missing from other jurisdictions, because far fewer locations were reported than seem to be required to be monitored under first round permits.
2. **Fully QA/QC water quality data**. The data have not been systematically checked for outliers and suspicious values; these need to be identified and corrected if necessary. A systematic approach to treating differences between EMC0 and EMCdt introduced by rounding and truncating one of the two EMC values needs to be implemented. The status of zero values for EMC0 and EMCdt needs to be determined, particularly for sampling events in which both EMCs have zero values. In the latter case, it may be necessary to contact the county or verify the values. It may not be necessary to determine if the differences between EMCs are truly due to calculations with observations below detection limits, as opposed to data input methods, as long as there is reasonable assurance that classification of differences will not impact analyses performed with the data.
3. **Use alternative statistical methods where appropriate**. A more detailed analysis of the data is required to select the most appropriate statistical analyses to apply to the data. Preliminary analysis suggests that a significant number of observations below detection limits were used in the calculation of EMCs. Should that be confirmed, then the statistical methods recommended by Helsel (2012) for treating censored data should be used, not only for trend analysis, but for calculating descriptive statistics as well. The preliminary trend analysis described in this report estimated the trend by the slope of a linear regression of concentration versus time and tested the significance of the slope by its p-value. A full application of this method would require determining whether the standard conditions on the residuals are met, i.e., the residuals are normally-distributed, independent, and homoscedastic. When these conditions are not met, non-parametric methods, like the Mann-Kendall trend test in particular or the seasonal Kendall test, if there are seasonal effects, are appropriate alternatives. Linear regression and the Kendall test are appropriate if there is an increasing trend with time. If a major effort at watershed restoration is confined to a narrow time period, estimating a step-trend, which measures conditions before and after a change at a point in time, may be more appropriate. Step trends

may be estimated using parametric or non-parametric methods; Helsel (2012) describes variations on these methods for use with censored data.[1]

The preliminary data analysis performed on the water quality dataset has deliberately stopped short of addressing questions concerning stormwater characterization or watershed restoration effectiveness. To address these topics, it is necessary to introduce additional sets of independent variables in the analysis. For example, to characterize whether copper concentration in stormwater vary by land use or percent impervious cover, it necessary to associate monitoring locations with land use type and percent impervious cover. Similarly, to determine if watershed restoration has decreased concentrations of total phosphorus at a monitoring location, it is necessary to have some variable that represents the degree of watershed restoration. Land use information and other variables will have to be calculated from available information before analysis oriented towards stormwater characterization or evaluation of watershed restoration can take place.

MDE, ICPRB, and the Center of Watershed Protection (CWP) are currently working on a pilot study of the data from three watersheds where outfall and instream monitoring has been performed in compliance with Phase I MS4 permits. The qualifying watersheds are expected to have the following characteristics:

1. Have a long (ten years or more) monitoring record at fixed stations with regular frequency and minimum gaps;
2. Have relatively complete data sets;
3. Be relatively free of data quality issues;
4. Have documented changes in watershed restoration or BMP implementation over the monitoring period; and
5. Show preliminary evidence of trends, as reported in this project, consistent with the changes in watershed restoration.

It is hoped that the pilot study will serve as a model for analyzing all watersheds in which sufficient monitoring data have been collected under the MS4 program, and yield recommendations for improving the monitoring component of MS4 permit requirements.

In a follow-up project to this one, MDE and ICPRB are working together to enter stream habitat and biological metric scores and indices, and raw benthic taxa data into the MS4 database. Analyses similar to the ones described in this report, including descriptive statistics and basic trend analyses for biological and habitat metrics and indices are in progress.

[1] Helsel and Hirsch (2002) provide guidelines for choosing the appropriate methods for descriptive statistics and trend analysis. Buchanan and Mandel (2015) demonstrate the application of Helsel and Hirsch's recommendations for trend analysis on 43 years of ambient water quality monitoring data in West Virginia.

Summary

A relational database has been designed to house the water chemistry, habitat, biological, and physical monitoring data collected by under Maryland's Phase I MS4 program. The database has been populated with almost 97,500 records of water chemistry and flow parameters taken from nearly 5,000 sampling events at 69 monitoring locations. Missing data, the high degree of censored data, and other data quality issues have been identified but not yet resolved. Descriptive statistics of EMCs were calculated by water quality parameters and monitoring location. The distributions of mean EMCs from Maryland outfalls tend to be comparable or somewhat lower than the distribution of mean EMCs from the northeast quadrant of the United States in the NSQD for all parameters except *E.coli*. Trends in water quality parameters were identified through linear regressions of average EMCs against time. Several long-term locations where data have been collected for more than ten years show decreasing trends for four or more parameters. These results are preliminary pending resolution of data quality issues and a more rigorous analysis of the choice of statistical methods to apply to the data. A pilot study is being planned to (1) select three watersheds monitored under the MS4 program; (2) resolve outstanding data quality issues effecting the data collected in those watersheds; (3) quantify BMP implementation in the catchments above each monitoring location; and (4) use more rigorous statistical methodology to determine trends in water quality parameters and their relation to watershed restoration efforts. The pilot study will provide a template for analysis of the other watersheds monitored in the program and provide recommendations for improving the monitoring component of MS4 permit requirements.

References

Buchanan, C. and R. Mandel. 2015. Water Quality Trend Analysis at Twenty-Six West Virginia Long Term Monitoring Sites. Interstate Commission on the Potomac River Basin (ICPRB): Rockville, MD. ICPRB Report14-6. https://dep.wv.gov/WWE/watershed/wqmonitoring/Pages/waterquality.aspx

Helsel, D. R. and R. M. Hirsch. 2002. Statistical Methods in Water Resources. Techniques of Water Resource Investigations of the United States Geological Survey. USGS. Available online at http://pubs.usgs.gov/twri/.

Helsel, D. R. 2012. Statistics for Censored Environmental Data Using Minitab and R. Second Edition. John Wiley and Sons. Hoboken, NJ.

Pitt, R., A. Maestre, and R. Morquecho. 2004. The National Stormwater Quality Database (NSQD), Version 1.1. A Compilation and Analysis of NPDES Stormwater Monitoring Information. Available at http://unix.eng.ua.edu/~rpitt/publications/stormwater%20characteristics/nsqd%20epa.pdf