An Analysis of Pooled Monitoring Data in Maryland to Evaluate the Effects of Restoration on Stream Quality in Urbanized Watersheds

Final Report

May 31, 2022

Project team:

ICPRB

Rikke Jepsen

Claire Buchanan

Andrea Nagel

CWP

Deb Caraco

Lisa Fraley-McNeal

Interstate Commission on the Potomac River Basin
30 West Gude Drive, Suite 450
Rockville MD 20850
www.potomacriver.org

Center for Watershed Protection
11711 E Market Place
Fulton, MD 207593
www.cwp.org

ICPRB Report 22-2

This is a draft report. When the final report is available, it can be downloaded from the Publications tab of the Commission's website, www.potomacriver.org. To receive hard copies of the report, please write:

Interstate Commission on the Potomac River Basin

30 West Gude Drive, Suite 450

Rockville, MD 20850

or call 301-984-1908

Acknowledgements

This project was funded by Maryland Department of the Environment (MDE) through PCA #40320, Purchase order # U00P1600515.

Suggested citation for this report:

Jepsen, R., Caraco, D., Fraley-McNeal, L, Buchanan, C., and Nagel, A. 2022. An Analysis of Pooled Monitoring Data in Maryland to Evaluate the Effects of Restoration on Stream Quality in Urbanized Watersheds: Final Report. ICPRB Report 22-2. Interstate Commission on the Potomac River Basin, Rockville, MD.

Table of Contents

List of Figures

List of Tables

Executive Summary

The Interstate Commission on the Potomac River Basin (ICPRB) and the Center for Watershed Protection (CWP) conducted a study to evaluate the effectiveness of stormwater management activities in central Maryland. A variety of data sources were used to characterize stream health, including the data collected by municipal separate storm sewer (MS4) permittees, National Pollutant Discharge Elimination System (NPDES) best management practice (BMP) databases, and ICPRB's macroinvertebrate and stream habitat database. Metrics of stream health that were analyzed represented instream water quality, habitat, and biology.

Three hypotheses were investigated:

1) Increasing impervious cover and declining forest cover are associated with degraded stream health.
2) Stormwater BMPs and restoration efforts can ameliorate the effects of urbanization.
3) Biological stream health, as measured by benthic health indices, can be predicted by hydrologic, geomorphic, and habitat water quality indicators.

Hypothesis 1 was evaluated by comparing impervious and forest cover with metrics of water quality (conductivity), habitat (bank stability and epifaunal substrate), and biology (BIBI index). As impervious cover increased and forest decreased, conductivity increased and benthic health decreased, which supports Hypothesis 1. The habitat metrics did not show any obvious correlation with impervious cover.

The primary objectives in evaluating Hypothesis 2 were to determine if changes in stream condition metrics, such as turbidity, bank stability, or BIBI index score, could be linked to stormwater BMP implementation or other restoration activities. This hypothesis was investigated by comparing data that characterized ten MS4 National Hydrography Dataset (NHD) catchments to a group of control and reference NHD catchments. Control catchments were selected based on similarity to each MS4 catchment in regard to land cover, size, and other general features, but had little or no BMP implementation. Reference catchments represented high quality watersheds that were largely forested with little impervious cover and no septic/sewer service areas. The analysis was mainly based on boxplots and scatterplots to visualize trends, with associated p-values generated using Dunn's test, following a Kruskal-Wallis comparison.

Reference catchments were clearly different from MS4 and control catchments, with higher quality instream habitat, water quality, and macroinvertebrate communities. Depending on the location, MS4 catchments tended to perform as well as or worse than the controls. A variety of factors influenced how MS4 catchments compared to controls, including whether stream conditions were very degraded before BMP implementation, the length of time since the disturbance of installing BMPs or performing stream restorations, and how much impervious cover versus forest was present in the local catchment.

Hypothesis 3 was supported by the finding that reference catchments, with uniformly better instream conditions than MS4 and control conditions, had high BIBI index scores. Further analyses are needed to explore the relationships between biological health and instream conditions when biology is poor, as was observed in MS4 and control catchments.

List of Abbreviations

BANKS	Bank Stability
BEHI	Bank Erosion Hazard Index
BMP	Best Management Practice
CART	Classification and Regression Tree Analysis
CC	Chesapeake Conservancy
CIC	Conservation Innovation Center
CP	Complexity Factor
CWP	Center for Watershed Protection
D.C.	District of Columbia
DO	Dissolved Oxygen
EMC	Event Mean Concentration
EPA	Environmental Protection Agency
EPI_SUB	Epifaunal Substrate
GAM	General Additive Model
HUC	Hydrologic Unit Code
IC	Impervious Cover
ICPRB	Interstate Commission on the Potomac River Basin
LNP	Lower Northern Piedmont
MBSS	Maryland Biological Stream Survey
MD	Maryland
MDE	Maryland Department of the Environment
MS4	Municipal Separate Storm Sewer System
NHD	National Hydrography Dataset
NLCD	National Land Cover Database
NP	Northern Piedmont
NPDES	National Pollutant Discharge Elimination System
OTH	Other
PA	Pennsylvania
SEP	Southeastern Plain
SPCOND	Specific Conductivity
SSO	Sanitary Sewer Overflow
UNP	Upper Northern Piedmont
USGS	United States Geological Survey
VA	Virginia
WSSC	Washington Suburban Sanitary Commission
XDED	Extended dry detention structure
XDPD	Detention structure (dry pond)
XFLD	Flood management area
XOGS	Oil grit separator

Introduction

This project builds on a pilot study (Jepsen & Caraco, 2020) that investigated the influence of MS4 controls on water quality in three Maryland areas: Baltimore City, Frederick County, and Carroll County. The scope of the current project was expanded to include more MS4 sites and examined whether stormwater management practices can lead to measurable differences in features other than water quality (e.g., stream hydrology, hydraulics, geomorphology, macroinvertebrate communities). MS4 catchments were compared to control catchments with similar land cover and few or no stormwater BMPs and to reference catchments with high forest cover and limited development.

Three hypotheses with associated sub-hypotheses were the project's focus:

1. Increasing impervious cover and declining forest cover are associated with declining stream health, as measured by the following indicators:
    a. Increasing streamflow flashiness and unit runoff volume (hydrology)
    b. Loss of floodplain connectivity (hydraulics)
    c. Declining stream stability and habitat quality, as measured by parameters such as the Bank Erosion Hazard Index (BEHI), stream embeddedness, and riparian buffer (geomorphology)
    d. Increasing instream pollutant concentrations during both storm and non-storm conditions (water quality)
    e. Degrading aquatic communities (biology)
2. Stormwater BMPs and infrastructure improvements can ameliorate the effects of urbanization, with resulting changes in the physical and chemical features of streams:
    a. Modern structural stormwater BMPs implemented outside of the riparian corridor are associated with hydrologic, geomorphic and water quality improvements.
    b. Infrastructure improvements that address sanitary sewer overflows or cross connections are associated with improved instream water quality.
3. Biological stream health, as measured by benthic health indices, can be predicted by hydrologic, geomorphic, habitat, and water quality indicators.

Methods

Study Area

The study area covers parts of central Maryland, which is comprised of several ecoregions (Figure 1). The following descriptions of the Environmental Protection Agency (EPA) level III ecoregions in the study area are based on Griffith (2010) and Woods et al. (1999). The averages and ranges of various features (e.g., elevation, slope) were derived using data from the United States Geological Survey's (USGS) ScienceBase catalog and EPA's StreamCat database.

Southeastern Plain

The Southeastern Plain (SEP) is a flat and sandy ecoregion in the Coastal Plain physiographic province, bordered on the east by the Chesapeake Bay and, on the west, the Piedmont and Northern Piedmont ecoregions. The Mid-Atlantic fall line forms SEP's western border and separates it from the harder metamorphic and igneous rocks of the Piedmont. The mean catchment elevation is 26 m, ranging from 0 m to 110 m, with slopes averaging 4% and ranging from 0% to 27%. Precipitation is evenly distributed throughout the year (average 1,358 mm;

range 1,140 mm to 1,520 mm). Streams in SEP are low to moderate in gradient, often have sandy bottoms, and can have swampy margins. Soils are loamy and naturally nutrient poor. The land cover in SEP is a mixture of forests, cropland, pasture, and developed areas. Upland forests are oak-hickory-pine dominated, and floodplains contain bottomland oaks, red maple, green ash, sweetgum, and (historically) American elm. Forest cover at the catchment scale averages 39% and cropland/hay/pasture averages 13%. SEP has pine plantations, largescale poultry and hog production, and crops including corn, cotton, soybeans, and peanuts. Larger cities include Richmond, VA; parts of Washington D.C.; Annapolis, MD; and parts of Baltimore, MD.

Northern Piedmont

The Northern Piedmont (NP) ecoregion is a transitional region of low hills, valleys, and plains between the Appalachian Mountains to the west and the Coastal Plain to the east. This ecoregion is separated from the Piedmont by a natural vegetation boundary between the Piedmont's oak-hickory-pine forests and Northern Piedmont's Appalachian oak forest.

Lower and Upper Northern Piedmont (LNP, UNP)

During development of ICPRB's macroinvertebrate index, discussed below, the Northern Piedmont ecoregion was separated into two bioregions to better distinguish benthic communities. The two bioregions, Lower Northern Piedmont (LNP) and Upper Northern Piedmont (UNP), were divided at the Susquehanna-Potomac River boundary, with LNP in the inland, non-Coastal Plain portion of the Potomac and Lower Chesapeake-James hydrologic subregions and UNP in the Susquehanna and Upper Chesapeake hydrologic subregions. For this study, data were originally separated by bioregion, but data from LNP and UNP were later combined to represent the overall Northern Piedmont ecoregion, which is aligned with how MDE uses data from Northern Piedmont.

LNP catchment elevation averages 137 m and ranges from 0 m to 470 m, and the average catchment gradient is 7%, ranging from 0% to 40%. Average annual precipitation is 1,097 mm (range is 930 mm to 1,250 mm). LNP is underlaid by metamorphic, igneous, and sedimentary rocks, and its soils are deep and fertile. Most of the streams are perennial and are low to moderate in gradient; some springs are found. The landscape of LNP is a mix of farms, urban development, and forests. Natural forest composition is dominated by red and white oaks, interspersed with hickory, ash, elm, and eastern redcedar; the average percentage of forest per catchment is 34%. Agricultural land use is an average of 39%, and cultivation includes soybeans, livestock feed crop, and Christmas trees. Notable cities and towns are Gettysburg, PA; Frederick, MD; Rockville, MD; Gainesville, VA; Culpeper, VA; and Charlottesville, VA.

Elevation in UNP is slightly lower than in LNP, with an average of 151 m and a range of 0 m to 326 m; the average gradient is 7%, ranging from 0% to 35%. UNP shares many features with LNP, including the bedrock, soils, streams, and land use. Average forest cover in catchments is 29%, and the average percent of agriculture is 41%. Larger urban areas include Columbia, MD; parts of Baltimore, MD; Hanover, PA; Harrisburg, PA; and Lancaster, PA.

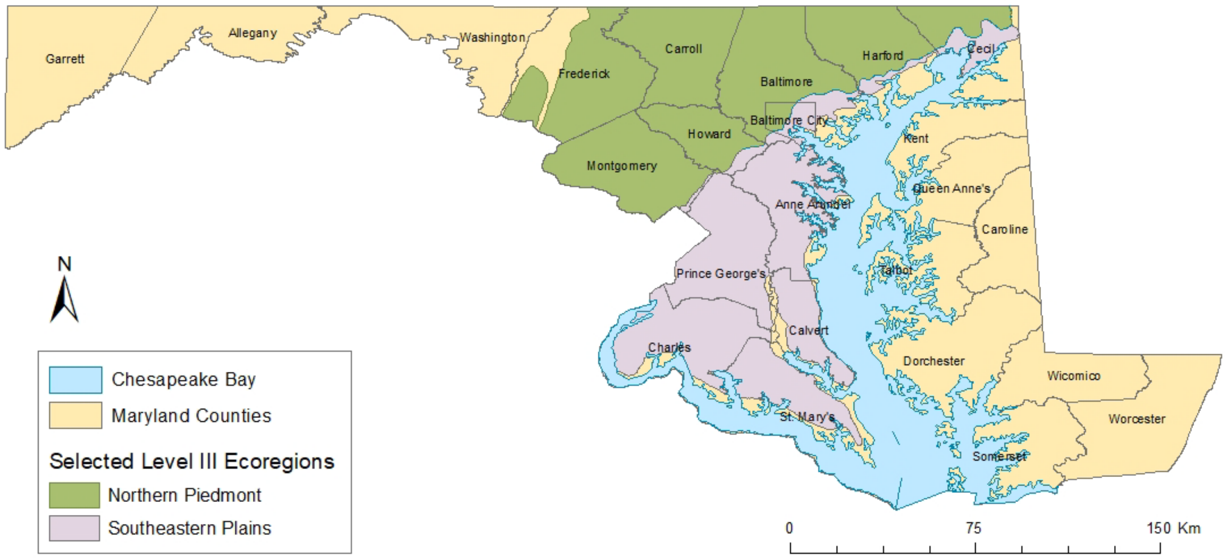

Figure 1. EPA level III ecoregions in the study area of central Maryland.

Data Assembly

The datasets needed to relate stream quality to stormwater management efforts were identified and then assembled. Datasets included water quality, geomorphology, benthic community metrics, meteorological data, land cover, and BMP implementation. Most of these datasets were already available at the start of the project or could be acquired quickly, but other datasets (i.e., records of sewer infrastructure repair) took longer to assemble and were not available until later in the project. Not all datasets were eventually used for analysis and some data were not universally available. The following section provides descriptions of the datasets that were included in analyses.

MS4 Monitoring Database

Data submitted by MS4 permittees were extracted from an updated version of the Microsoft Access MS4 monitoring database (version 5.3b) used in the pilot study. The database was developed by ICPRB staff and houses data that includes event mean concentrations (EMCs) of stormwater discharges from the outfall and instream sites; narrative and numeric information related to geomorphologic stream assessments; benthic and habitat indices/metrics; raw benthic counts; and station information. In-situ data were collected at the time of benthic sampling, and parameters include water temperature, dissolved oxygen (DO), pH, conductivity, and turbidity. Examples of habitat parameters in the database are physical habitat index, embeddedness, bank stability, and epifaunal substrate.

BMP Databases

MDE provided NPDES MS4 geodatabases from 2019 that contain updated design information for BMPs implemented for new development, redevelopment, and restoration as required by the MS4 permit, including the BMP location, type, and delineated drainage area. The geodatabases also contain alternative, non-structural BMPs that rely on a land cover change (e.g., tree planting or impervious cover elimination); non-stormwater practices such as septic practices; and municipal operations such as street sweeping and catch basin cleanouts. Stream restoration, shoreline stabilization, and outfall stabilization were also included in the geodatabases as alternative BMPs delineated as lines.

Chessie BIBI Database

The Chessie BIBI (Chesapeake Basin-wide Index of Biotic Integrity) database houses the ICPRB-developed Chessie BIBI (family-level, multimetric macroinvertebrate index; Smith et al. 2017) sample scores and narrative ratings for >29,000 benthic samples, as well as the metrics that underly the index. Additionally, most samples have associated habitat scores and water quality measurements. The most frequently sampled water quality parameters are specific conductance, dissolved oxygen, water temperature, pH, and turbidity. Habitat metrics include sinuosity, bank stability, epifaunal substrate, woody debris, sedimentation, and embeddedness.

Using selected habitat and water quality parameters from the Chessie BIBI database, ICPRB has developed a method to assign each sampled stream reach a condition of either reference, minimally degraded, mixed, moderately degraded, or degraded. This classification is based on three commonly sampled water quality parameters (pH, dissolved oxygen, specific conductivity) and eight commonly sampled habitat parameters (bank stability, bank vegetation, channel alteration, embeddedness, epifaunal substrate, flow, riffle/run/pool ratio, and sedimentation). Each sample with a Chessie BIBI score typically has an assigned stream condition category.

Habitat Metrics

Several habitat metrics are available in the MS4 monitoring and Chessie BIBI databases. Metrics are scored from 0 (worst condition) to 20 (best condition). The following list of the habitat metrics used in analyses describe a stream that would score high (16 – 20).

a. Bank stability – banks stable; evidence of erosion or bank failure absent or minimal; little potential for future problems. <5% of bank affected.
b. Bank vegetation – more than 90% of the streambank surfaces and immediate riparian zones covered by native vegetation, including trees, understory shrubs, or nonwoody macrophytes; vegetative disruption through grazing or mowing minimal or not evident; almost all plants allowed to grow naturally.
c. Channel alteration – channelization or dredging absent or minimal; stream with normal pattern.
d. Embeddedness – minimal extent (<25%) to which rocks (gravel, cobble, and boulders) and snags are covered or sunken into the silt, sand, or mud of the stream bottom.
e. Epifaunal substrate – a lot of substrate (~70%) is favorable for epifaunal colonization and fish cover (mix of snags, old submerged logs, undercut banks, cobble and other stable habitat).
f. Flow – water reaches base of both lower banks, and minimal amount of channel substrate is exposed.
g. Riffle – frequency of riffle habitat or diversity of habitat is high (heterogeneity).
h. Sediment – little or no enlargement of islands or point bars and less than 5% (<20% for low-gradient streams) of the bottom affected by sediment deposition.

Land Cover Layers

Two land cover products were used. The first is the Chesapeake Conservancy (CC) Conservation Innovation Center (CIC) 1-meter, high resolution land cover layer for 2013/2014. Land cover is classified according to six classes: water, tree canopy and shrubland, herbaceous, barren, impervious (roads), and impervious (other). The *tabulate area* tool in ArcGIS (version 10.8) was used to calculate the areal extent of each land cover class in the catchments, after which the two impervious classes were combined so that there was only one value for impervious cover.

The second land cover product was the thirty-meter, medium-resolution National Land Cover Database (NLCD), aggregated at the catchment scale. Land cover is available for multiple years and includes more categories than the high- resolution land cover, including four classes of developed land (open, low, medium, high) and two wetland classes (woody and herbaceous); however, due to the lower resolution it was not the preferred land cover product.

Watershed Boundaries

Two watershed shapefiles were utilized, NHD catchments and HUC12 boundaries. NHD catchments are much smaller than HUC12 watershed and were derived from the 1:100,000-scale USGS NHDPlus V2.1 data product. HUC12 watershed shapefiles were acquired from the Watershed Boundary Dataset (WBD).

Septic and Sewer Service Areas

Shapefiles of sewer service and septic locations were provided by MDE. The sewer service layer outlines areas in Maryland serviced by the public sewer system in 2018 and areas that will be included in the sewershed in the future. Attributes include the polygon's jurisdiction code, associated wastewater treatment plant, and status (e.g., existing, under construction). The statewide septic layer shows point-locations of septic systems and provides the street address, designation (e.g., residential, agricultural, commercial), year built, and size of the system (ft^2). Parcel data obtained from the Maryland Department of Planning was used to help determine the septic and sewer service areas by catchment. All parcels that contained a septic point from the data layer provided by MDE were categorized as septic, and all parcels that were located within an existing sewer service area were categorized as sewer. In some instances, septic points were located within sewered areas, in which case the parcels were categorized as septic instead of sewer. The septic and sewer categorized parcels were then intersected with NHD catchments to calculate the area in acres of septic and sewer service areas within each catchment. These areas were then subtracted from the total catchment area and the remainder was noted as unclassified, which generally include areas such as agricultural fields, natural areas, and roadways.

Site Selection

Once the datasets were assembled, all MS4 outfall and instream monitoring stations from the MS4 database were evaluated for analysis suitability. Station coordinates were linked to NHD catchment and HUC12 shapefiles and the corresponding watershed IDs extracted. MS4 catchments were then selected for analysis based on data availability in the local catchment. Five MS4 catchments were excluded due to an overall lack of data or large data gaps in the MS4 monitoring and/or Chessie BIBI databases: Spring Branch in Baltimore County, Zekiah Swamp in Charles County, Winters Run and Wheel Creek in Harford County, and Wilde Lake in Howard County. Ten MS4s were retained for analysis (Appendix A Table A1).

Candidates for control catchments came from HUC12s containing MS4 catchments or HUC12s immediately adjacent to MS4 HUC12s. This resulted in 2,224 candidate catchments distributed among 40 HUC12s (Figure 2). The control catchments that were eventually selected from this pool had characteristics similar to individual MS4 catchments (e.g., ecoregion, stream order, percent forest, percent impervious cover, sewer/septic coverage) but had little or no BMP implementation.

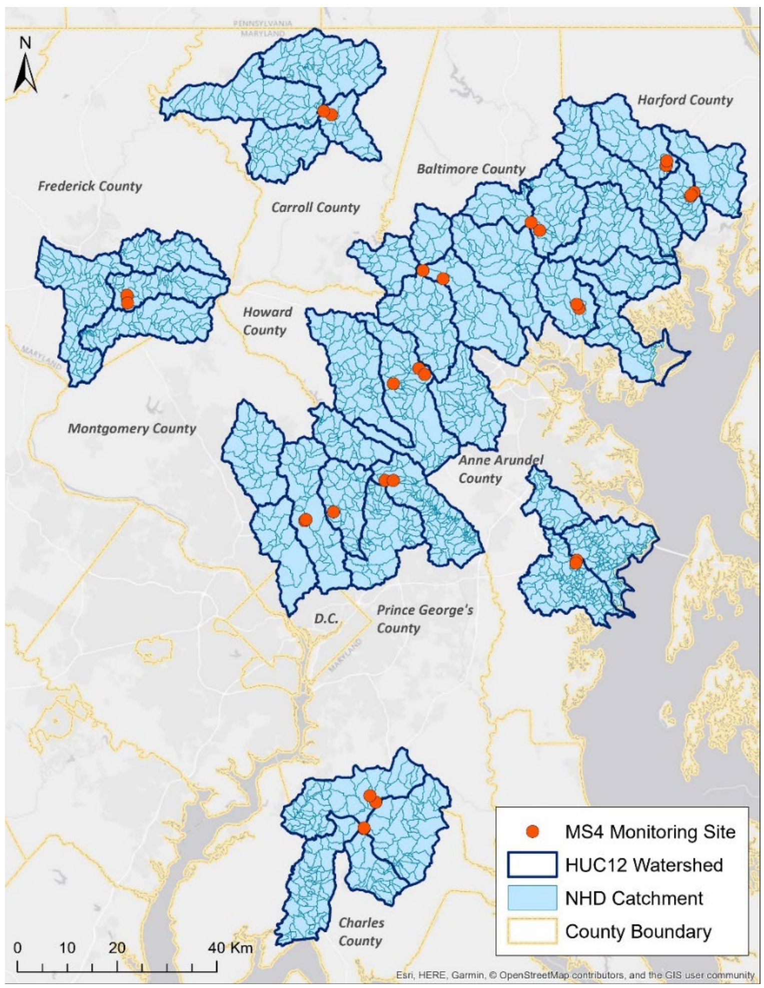

Figure 2. HUC12 watersheds with the candidate catchments that contain and surround the MS4 monitoring sites.

Candidates for the reference catchments came from a larger geographic area within the NP and SEP ecoregions, overlapping southern Pennsylvania and northern Virginia, because there were too few catchments meeting the criteria for reference in the MS4 and neighboring HUC12 watersheds. Reference catchments were first and second order streams in highly forested drainage areas with limited development in the local and upstream catchments.

Description of Filters

The process of selecting control and reference catchments to pair with MS4 catchments evolved over time to assure sufficient analysis power. Catchments with no water quality, habitat, or benthic samples were discarded. Specific filters were applied to the list of candidate catchments to exclude catchments that did not fit the criteria for either a control or a reference. Control and reference candidates for SEP were also filtered to only include for consideration those catchments with less than 10% woody wetlands in order to more closely match the MS4 catchments in SEP.

Control Catchment Selection

Control catchments were individually selected for each drainage area containing an MS4 BMP. Candidates were excluded if they were in a different ecoregion, outside of ±10% of the MS4 drainage area's percent of impervious cover and outside of ±20% of the MS4 drainage area's percent of forest (based on the 1-m Chesapeake Conservancy layer). Next, catchments were filtered according to Strahler stream order, which was acquired from an attribute table accompanying the NHD shapefile. Controls with the same stream order as the MS4 drainage area were preferred but not required. Of the ten MS4 drainage areas, two are second order and eight are first order, with seven of those eight being headwaters, meaning they contain the origin of a stream segment. For the headwaters, a control that was also a headwater was preferred where possible. For the two second order catchments, land cover in the *entire* drainage area (i.e., including both the MS4 drainage area and the contributing upstream catchment) was determined before controls were selected. This included identifying contributing catchments and summarizing total impervious cover, drainage area, and forest/shrub areas in the upstream drainage area. When the upstream land cover had been calculated for each non-headwater catchment, the upstream percent impervious cover and percent tree/shrub were filtered on using the same ranges as the local land cover filters. Finally, control candidates with greater than one percent BMP implementation were removed. Septic and sewer coverage, which were not filtered on, were compared to the MS4 catchment's values.

Reference Catchment Selection

Reference catchments were also selected among first and second Strahler order streams. To identify least-disturbed, or reference, drainage areas in the NP and SEP ecoregions, candidate catchments with less than 5% impervious cover and/or greater than 70% trees/shrubs were identified. These numbers are loosely based on a report by the U.S. Army Corps of Engineers, The Nature Conservancy, and ICPRB (2014), which found that stream flow metrics become altered when impervious cover exceeds 0.4% to 2.1%, when urban land cover exceeds 5% to 15%, or when percent forest falls below 46% to 75%. In the SEP ecoregion, reference catchments were further filtered to include only catchments with less than 10% woody wetland, representing the 95th percentile of woody wetland cover in 1st and 2nd order stream drainage areas in the Northern Piedmont. This wetland filter was applied because highly forested, low impervious cover watersheds in the SEP tended to be wetland-dominated, in contrast to the treatment (MS4) watersheds, making these watersheds an inappropriate reference condition.

BMP and Land Cover Change Analysis

The BMP and land cover change analysis was conducted within the instream and outfall MS4 monitoring drainage area boundaries for all sites except Breewood and Stewart-April Lane. The

Breewood MS4 monitoring drainage area is substantially smaller than the catchment it is located within, so the NHD catchment boundary was used instead. For Stewart-April Lane, the MS4 monitoring location did not have corresponding monitoring drainage areas provided in Montgomery County's MS4 GIS database. Furthermore, this monitoring location is situated on a second order stream with a drainage area that includes thirteen upstream NHD catchments, which were used as the boundary for this analysis.

The previous study (Jepsen & Caraco, 2020) included a detailed analysis of BMP and land cover change over time for Airpark Business Center, Urbana, and Moores Run which was used as part of this analysis. For the remaining MS4 locations, a similar analysis was conducted, where impervious cover was used as an indicator of land use change over time. The acres of impervious cover within the BMP drainage areas were compared to impervious cover in the monitoring drainage areas they were located within to determine the portion of the monitoring drainage areas treated by BMPs. Both BMP coverage and impervious cover were analyzed for the following years: 2003, 2008, 2014, and 2019.

The extent of BMP coverage was obtained from the NPDES BMP geodatabases, which included information about locations and types of BMPs in MS4 jurisdictions, as well as built year and drainage areas. The individual MS4 BMP drainage areas were merged into one layer with all pertinent BMP information, and the following BMP drainage areas were removed:

- BMP types noted as other (OTH), oil grit separator (XOGS), and flood management area (XFLD).
- Those without corresponding BMP information.
- One large regional wet pond in Bear Branch that was located downstream and outside of the monitoring drainage area.

The BMPs areas were then filtered by type: detention (extended dry detention structure; XDED and extended dry pond; XDPD) and water quality (structural BMPs that have a required water quality treatment volume; include wet ponds, bioretention, infiltration, wetlands, sand filters). The drainage areas for BMPs that existed during each of the analysis years (2003, 2008, 2014, and 2019) were exported to new layers such that a layer was created for each combination of detention/water quality BMP type and year. Drainage areas within each of the layers were dissolved to remove overlapping areas. The water quality BMP coverage layer for each year was then erased from the detention BMP coverage layer for that year to remove areas where the two overlap. The result is one layer of BMP coverage for each analysis year that represents areas treated by water quality BMPs and areas only treated by detention BMPs. The BMP coverage layers were intersected with the monitoring drainage areas to calculate the extent of BMP coverage for each MS4 monitoring location over time.

Impervious cover change over time was analyzed according to the same analysis years as the BMP coverage. Impervious (roads) and impervious (other) from the Chesapeake Conservancy 2013/2014 land cover data were merged into a separate layer representing impervious cover during the 2014 analysis year and then clipped to the monitoring drainage area boundaries. This layer was then reviewed with aerial imagery from MD iMAP and Google Earth to identify areas of impervious cover change over the years. The 2014 impervious cover layer was used as a base layer from which changes in impervious cover were delineated from the aerial imagery and saved into new layers for the 2003, 2008, and 2019 analysis years. Note that only larger areas of impervious cover change were delineated (e.g., industrial buildings, main roadways, subdivisions) as opposed to smaller areas of change (e.g., individual houses).

The impervious cover layers were then intersected with the corresponding BMP coverage layers for each year to calculate the following metrics for characterizing land use and BMP change over time for each monitoring drainage area: total impervious cover, impervious cover captured by detention ponds, and impervious cover captured by other water quality practices. Tables showing the results of this analysis for each of the MS4 catchments are included in the results section.

Metric Selection for Drainage Area Comparisons

Not all of the available parameters in the MS4 monitoring and Chessie BIBI databases were used in the study. Water quality was limited to DO, pH, conductivity, and turbidity. Eight habitat metrics were studied, but data were relatively sparse for some metrics. Epifaunal substrate and bank stability typically had the greatest amount of data. Almost sixty biological metrics are in the Chessie BIBI database, so a subset of metrics was selected based on a simple analysis where correlation coefficients were calculated between impervious cover and the metric. Metrics most sensitive to impervious cover were used. The BIBI database was also limited to first and second Strahler order streams in each ecoregion.

For SEP, biological metrics included in analyses were:

- Average Score Per Taxon (ASPT_MOD)
- Margalef's Index (MARGALEFS)
- Number of moderately tolerant families (RICH_MODTOL)
- Percent of burrowers (PCT_BURROW)
- Percent of intolerant families with tolerance values of 0-4 (PCT_INTOL_0_4)
- Percent of Trichoptera excluding family Hydropsychidae (PCT_NON_HYDROP_TRICHOPTERA)
- Percent of shredders (PCT_SHRED)

For NP, biological metrics included in analyses were:

- Number of clinger families (RICH_CLING)
- Number of Ephemeroptera families (RICH_EPHEMEROPTERA)
- Number of Plecoptera families (RICH_PLECOPTERA)
- Percent of Ephemeroptera (PCT_EPHEMEROPTERA)
- Percent of Ephemeroptera excluding family Baetidae (PCT_EPHEMEROPTERA_NO_BAETID)
- Percent Ephemeroptera + Plecoptera + Trichoptera (PCT_EPT)
- Percent of intolerant families with tolerance values 0-3 (PCT_INTOL_0_3)
- Percent of Plecoptera (PCT_PLECOPTERA)

Monitoring Data Assembly

The Chessie BIBI database and the MS4 monitoring database were the sources of data for the analyses. Chessie BIBI data attributes include both an individual station ID and an NHD catchment feature ID, and MS4 monitoring data are associated with a specific MS4 monitoring drainage area. Datasets were derived from these two sources, with different rules for including data in each analysis.

Data Assembly to Test Hypothesis 1 (Impacts of Land Cover on Stream Health)

To test the hypothesis that stream health was negatively correlated with changes in impervious and forest cover, data from the entire Chessie BIBI database were extracted and filtered to include NHD catchments that were first or second order and in only the SEP or NP ecoregions. Data from the ten MS4 drainage areas selected as a part of this study were then added to the Chessie BIBI dataset.

Data Assembly for MS4 vs. Control and Reference Catchments

The bulk of the analysis focused on comparing metrics of stream health between MS4 (treatment), control, and reference conditions. Control and reference catchments were characterized by all of the Chessie BIBI data associated with that the catchment ID. For treatment drainage areas, selected data depended on the area of the NHD catchment covered by the MS4 instream drainage area and the number of catchments draining to the MS4 drainage area, such that:

- For most MS4 drainage areas, the MS4 was similar to (represented at least 2/3 the area of) the NHD catchment were it was located. In these cases, the treatment dataset included all Chessie BIBI data from the entire catchment NHD catchment where the MS4 drainage area was located.
- In two cases (Airpark Business Center and Parole Plaza/Church Creek), the MS4 drainage area represented only a small portion of the overall MS4 NHD catchment. In these catchments, Chessie BIBI data from lower reaches were not included in the analysis.
- Three MS4 drainage areas (Bear Branch, Stewart-April Lane/Paint Branch, and Urbana/Peter Pan Run) overlapped with two or more upstream NHD catchments. For these areas, all of the Chessie BIBI points from catchments in the contributing drainage area were included in the treatment dataset.
- In Breewood, the MS4 drainage area was extremely small and no comparable controls could be found. The MS4 monitoring area was on the whole not representative of the larger catchment. Therefore, the entire NHD catchment was used to represent the treatment condition and Chessie BIBI data from the entire catchment was included.
- The MS4 data from instream stations were included in the treatment dataset for all MS4 watersheds.

Statistical and Numeric Methods

This section summarizes the numeric and statistical methods used. Although some methods were only presented visually (e.g., time plots of metrics), most techniques reported p-values. Throughout the study, p-values of 0.05 or less were considered significant.

Impervious and Forest Cover vs. Stream Health Metrics

This analysis was completed by answering three sub-questions: 1) Is watershed impervious cover negatively correlated with stream health?; 2) Is watershed forest cover positively correlated with stream health?; and 3) With a given impervious cover, is the forest cover in the *remaining pervious area* correlated positively with stream health?

To answer the three sub-questions, four metrics of stream health were focused on: the Chessie BIBI score to represent biological health; specific conductivity (SPCOND) to represent water quality, and two habitat parameters, epifaunal substrate (EPI_SUB) and bank stability (BANKS).

This analysis used an expanded data set, which included MS4 data from the treatment watersheds, plus data from the Chessie BIBI database for all first and second order streams in the SEP and NP ecoregions.

Measures of impervious cover and forest cover were represented as a percent of the drainage area, but were slightly modified to account for drainage areas with a high amount of open water coverage, such that:

$$(F,I)_{adj} = \frac{F,I}{W} \times 100$$

Where:

$(F,I)_{adj}$	=	Adjusted Forest or Impervious Cover (%)
F,I	=	Forest or Impervious Cover (%)
W	=	Open Water Cover (%)

For simplicity, these adjusted parameters were referred to simply as “Forest Cover” or “Impervious Cover” in graphs and tables.

The relationship between impervious cover and various stream health metrics was evaluated using quantile regression with the R package *quant-reg* (Koenker, 2022). For SPCOND and EPI_SUB, the predictor variables (Forest Cover or Impervious Cover) were related to the 90^{th} percentile value of the metric (tau-0.9). For SPCOND, the “rq” function was used to generate the regression for the median (tau-0.5) versus the predictor variables. P-values for these regressions were calculated using the “anova.rq” function with nid standard error calculation. Smoothing lines in plots of impervious cover versus were created using the same tau values, and using default quantile regression methods contained in the R *ggplot2* package (Wickham, 2016).

The last sub-question related the “Forested Pervious Cover” to stream health metrics. This metric was chosen to reflect the effects of forest as a supplement to impervious cover. It was used in these combined models to minimize the effect of multicollinearity; impervious cover and forest cover are highly (inversely) correlated, since land development results in both forest clearing and an increase in impervious cover. Forested pervious cover is calculated as:

$$FP = \frac{F_{adj}}{100-I_{adj}} \times 100$$

Where:

FP = Forested Cover in Pervious Land (%)

The effect of forested pervious cover is developed using a model of the form $E = I_{adj}+FP$. These regressions used quantile-regression and the same tau values as the regressions of forest and impervious cover alone. P-values and slopes were calculated for both impervious cover (I_{adj}) and forested pervious cover (FP).

BMP Implementation vs. Stream Health

A method of evaluating the effects of BMPs was created using a metric that described level of implementation as the percent of impervious cover captured by water quality BMPs:

$$BMP_{WQ} = \frac{A_I}{A_{I-BMP}} \times 100$$

Where:

BMP_{WQ} = BMP Implementation (%)
A_I = Area of impervious cover in the catchment (acres)
$A_{I\text{-}BMP}$ = Area of impervious cover captured by water quality BMPs (acres)

Since both BMP implementation and impervious cover changed slightly over time in treatment watersheds, both the impervious cover and BMP level was assigned a "treatment era" based on the year associated with a monitoring event, such that:

- Observations from earlier than 2006 were assigned impervious cover and BMP implementation calculated for 2003
- Observations between 2006 and 2010 were assigned 2008
- Observations between 2011 and 2016 were assigned 2014
- Observations after 2016 were assigned 2019

This analysis was restricted to treatment and control drainage areas; reference data had much less impervious cover than either the treatment or control drainage areas, and also had no BMP implementation. These two characteristics effectively made these points influential outliers in evaluating BMP implementation.

Relationships were evaluated using quantile regression and the same stream health metrics, R packages, and tau values as the "Impervious and Forest Cover vs. Stream Health" analysis described above. BMPs were also evaluated in combination with impervious cover, using an equation of the form: $E=I+WQ_{BMP}$.

Statistical Comparisons Between Control, Reference, and Treatment Catchments

To determine if there was a significant difference between the catchment types, the non-parametric, rank-based Kruskal-Wallis test from the *stats* package (R Core Team, 2021) was utilized, followed by the Dunn's post-hoc test from the *FSA* package (Ogle et al., 2022) to identify which conditions were significantly different from each other.

Smoothing Methods for Time Curves

Scatterplots of the metrics showing change over time were smoothed using one of two methods depending on the number of points. Smoothing was first conducted using LOESS (locally weighted smoothing) curves, but plots with fewer than 20 points produced curves far outside reasonable bounds. Instead, linear regression was used to depict trends on curves with fewer than 20 points, and loess curves were used for datasets with at least 20 points. Both smoothing curves used the default methods included in the *ggplot2* package (Wickham, 2016).

Effects of Water Quality and Habitat on Biological Health

The final hypothesis was that biological health is dependent on both water quality and habitat quality. This hypothesis was tested using an expanded data set, which included all the MS4 data from the treatment watersheds, plus data from the Chessie BIBI database for all first and second order streams in the SEP and NP ecoregions. Quantile regression was again used to relate metrics with the 90th percentile BIBI score.

Classification and Regression Tree (CART) Models

CART models were developed to evaluate interactions and support other statistical analyses, with one model developed for each hypothesis; all three modeled the effects on the Chessie BIBI score. The first model evaluated the combined effects of impervious cover and forested pervious cover (Evaluating Hypothesis 1) using the expanded dataset . The second model evaluated the combined effects of impervious cover and BMP implementation, and used data from control, treatment, and reference catchments (Evaluating Hypothesis 2). Finally, the third model evaluated the impacts of habitat and water quality on biology, using conductivity, BANKS, and EPI_SUB as predictors. All three models were developed using the R package *rpart* (Therneau et al., 2022). The models used defaults from the rpart program, with the exception of the complexity factor (cp) which was set to 0.015 for all three models. The value was set slightly higher than the program default, as the trees developed using the default cp value resulted in trees with many branches that appeared to overfit the data.

Results and Discussion

Hypothesis 1: Effects of Land Cover on Stream Health

Hypothesis 1 states that increasing impervious cover and corresponding declines in forest are associated with reduced stream health, as measured by biological, habitat, and water quality stream health indicators.

With the exception of the BANKS habitat indicator, impervious cover had a significant negative impact (decreasing BIBI and EPI_SUB and increasing SPCOND) on stream health in both the coastal (SEP) and inland (NP) regions (Appendix B Table A5; Appendix A Figures A1, A3 and A5). For both the BIBI score and EPI_SUB, the 90th percentile quantile regression was utilized.

As Appendix A Figures A1 and A5 illustrate, poor BIBI and habitat scores are seen even at low levels of impervious cover. This result suggests that, while increasing impervious cover limits the potential for good biological health and epifaunal substrate, low impervious cover does not necessarily guarantee this result. SPCOND was variable but did not show the same wedge-shaped pattern (Appendix A Figure A3). The BANKS parameter was the only measure with results contrary to the original hypothesis, with a significant *improvement* associated with impervious cover in the Inland region, and no significant relationship in the coastal region (Appendix B Table A5; Appendix A Figure A5).

Results for regressions of forest cover and the same metrics show essentially the reverse results, with forest cover positively correlated with BIBI scores and EPI_SUB but having no significant relationship with the BANKS score (Appendix B Table A6; Appendix A Figures A2, A4, and A6). The forest effects have a similar pattern to imperviousness in that streams with high levels of forest cover may not necessarily have good stream health, but forest results are somewhat more variable. For example, some high BIBI scores were recorded at very low levels of forest cover (Appendix A Figure A2). The BANKS parameter showed no significant correlation with forest cover in either the inland or coastal region.

Finally, the combined effects of impervious cover and forest (measured as the forest percentage of pervious area) were investigated. The results (Appendix B Table A7) suggest that forested pervious cover (after accounting for imperviousness), has a significant positive effect on stream health every metric in the Inland region. Interestingly, the BANKS parameter continues to have an improvement associated with impervious cover, with an enhanced improvement associated with greater forest in the pervious areas. In the coastal region, however, the benefit of forest in pervious areas is not as clear. In that region, impervious cover alone appears to result in lower BIBI and EPI_SUB scores and increasing conductivity (all signs of degradation), and continues to have no impact on the BANKS score, increasing forest cover in pervious areas is associated with declining biological health (BIBI score), no significant impact on either habitat indicator, and a positive impact (lower conductivity values) on water quality.

Overall, the results generally support hypothesis 1, in that relationships of declining stream health were broadly associated with increasing impervious cover, and improvements associated with increasing pervious cover, and especially forest cover. The high variability, particularly at low levels of impervious cover and high levels of forest cover, suggests that factors other than these simple land cover metrics impact stream health. This is not surprising since macroinvertebrate community health can be impacted by a multitude of factors. One cause of these low BIBI scores could be influences from upstream catchments that are highly impervious, which would result in poor water quality in the local catchment where the macroinvertebrate sample was taken. Instances of high conductivity in less developed catchments such as a large application of road salt during a winter storm in an agricultural or forested catchment could explain some of the biological degradation. Also, if the local catchment was agricultural with low levels of impervious cover, the stream would be impacted by pesticides, herbicides, and/or fertilizer in the runoff from fields, or from animals given stream access. Highly forested watersheds can also have degraded water quality if sources of pollution are nearby, such as mines or road crossings.

Results for the BANKS scores were somewhat surprising, in that the parameter appeared to be uninfluenced by impervious cover (except for an apparent improvement in inland catchments). One possible explanation for this effect is that stream banks may actually be more stable at very high levels of impervious cover, either due to streams having reached equilibrium after development has occurred, or possibly due to armoring of the stream bank.

The other somewhat surprising result was the response to forest pervious cover (in combined impervious/forest models) in the coastal region. Forest cover was originally derived from the percent “Tree and Shrub” in each drainage area. It is possible that coastal watersheds are more heavily influenced by wetland cover.

Hypothesis 2: Effects of BMPs on Stream Health

To evaluate the impacts of stormwater BMPs on stream health (Hypothesis 2), a variety of comparisons were made between benthic, habitat, and water quality metrics representing the treatment, control, and reference drainage areas. The next two sections present a series of summaries, one for each MS4 site. These results are divided into the Inland/NP region and the Coastal/SEP region to further facilitate comparisons. ‘Treatment’ is used interchangeably with MS4 areas.

Northern Piedmont

Airpark Business Center (UNP)

Airpark Business Center is in Westminster, Carroll County near the Carroll County Regional Airport. The outfall monitoring station (WPU01) drains a large wet detention pond, and the instream monitoring station (WPU02) is located in a first order tributary to the confluence of the West Branch North Patapsco River (Figure 3). The catchment is primarily agricultural with pockets of forest and development (Table 1).

Table 1. Selected attributes of the Airpark Business Center MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
11.20 km^2	2.25 km^2	0.84 km^2	1	22.64	63.47	12.98	4.1 km^2	3.4 km^2

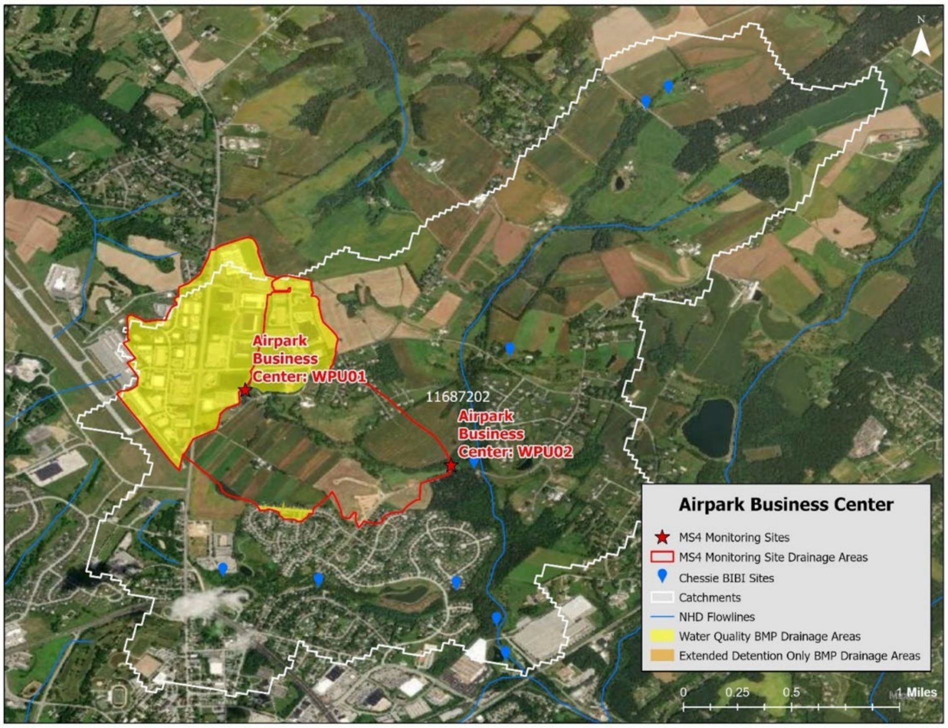

Figure 3. Airpark Business Center BMP coverage through 2019.

Between 2003 and 2014, impervious cover in the MS4 instream drainage area increased 8% from 17% to 25% (Table 2). This increase was offset by a large BMP retrofit, which converted an existing dry flood control pond to a wet pond water quality practice in October 2008. As a result, while only 18% of the watershed impervious cover was captured by water quality BMPs in 2003, nearly all of the impervious cover was treated by a water quality BMP by 2008 and onwards, which resulted in a marked decrease in metals and nitrogen species (Jepsen & Caraco, 2020).

Table 2. Impervious cover (IC) analysis of the instream drainage area at Airpark Business Center.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only
2003	17%	18%	0%
2008	21%	100%	0%
2014	25%	99%	0%
2019	25%	99%	0%

Airpark Business Center's MS4 instream drainage area is much smaller than the overall catchment and is more developed, so only data that were collected within the instream drainage area or just downstream of it were used in the analysis. Therefore, data used were primarily those collected as a part of the MS4 monitoring program. No BIBI stations were located within the MS4 drainage itself, but data were incorporated from one station directly below the MS4 monitoring station. As a result, relatively little data were available to characterize water quality, biology, and habitat.

Of the eight habitat metrics analyzed for the various MS4s, Airpark Business Center only had data for epifaunal substrate and riffles (riffle/run/pool ratio) (Appendix A Figure A9). Both metrics scored lower than the pooled NP reference and control catchments, but this difference was only significant between the MS4 and references (Appendix B Table A3). From 2008 and 2019, there appears to be a slight improvement in the MS4 watershed in the habitat scores (Appendix A Figure A10).

There were no water quality measurements in the MS4 outfall or instream drainage area or at the BIBI site just downstream, so water quality cannot be assessed at Airpark Business Center.

The health of macroinvertebrate communities at Airpark Business Center was worse than those in Northern Piedmont reference catchments, and similar to or slightly worse than controls. This is supported by the overall MS4 BIBI score and every metric being significantly worse than the reference communities (Appendix A Figure A7, Appendix B Tables A1 and A2). The BIBI score for the treatment was not significantly different from the control, but the treatment catchment was statistically worse for some metrics, including percent Plecoptera, clinger richness, and Plecoptera richness. Over time, the BIBI score at Airpark may have improved slightly after 2008, but the same apparent trend occurred in the control watersheds, and the BIBI score was consistently lower at Airpark than in either the control or reference watersheds (Appendix A Figure A8). The same apparent patterns are evident in individual metrics, except that the richness metrics appeared to decline, while some percent metrics improved, notably percent Ephemeroptera and percent EPT (Ephemeroptera + Plecoptera + Trichoptera).

Breewood (LNP)

Breewood is in Kemp Mill, Montgomery County. The outfall (scbtoutfall1) is located in Breewood Neighborhood Park at the outlet of a storm sewer in a wooded area approximately 175 feet south of University Boulevard. The instream station (scbtinstream1) is located on the downstream end of a culvert underneath Sligo Creek Parkway, approximately 50 feet upstream of the confluence with Sligo Creek, a first order tributary to the Anacostia River (Figure 4). The catchment is primarily suburban neighborhoods with mature trees and local parks (Table 3).

Table 3. Selected attributes of the Breewood MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
14.43 km^2	0.25 km^2	0.07 km^2	1	56.68	17.79	25.21	0 km^2	10.8 km^2

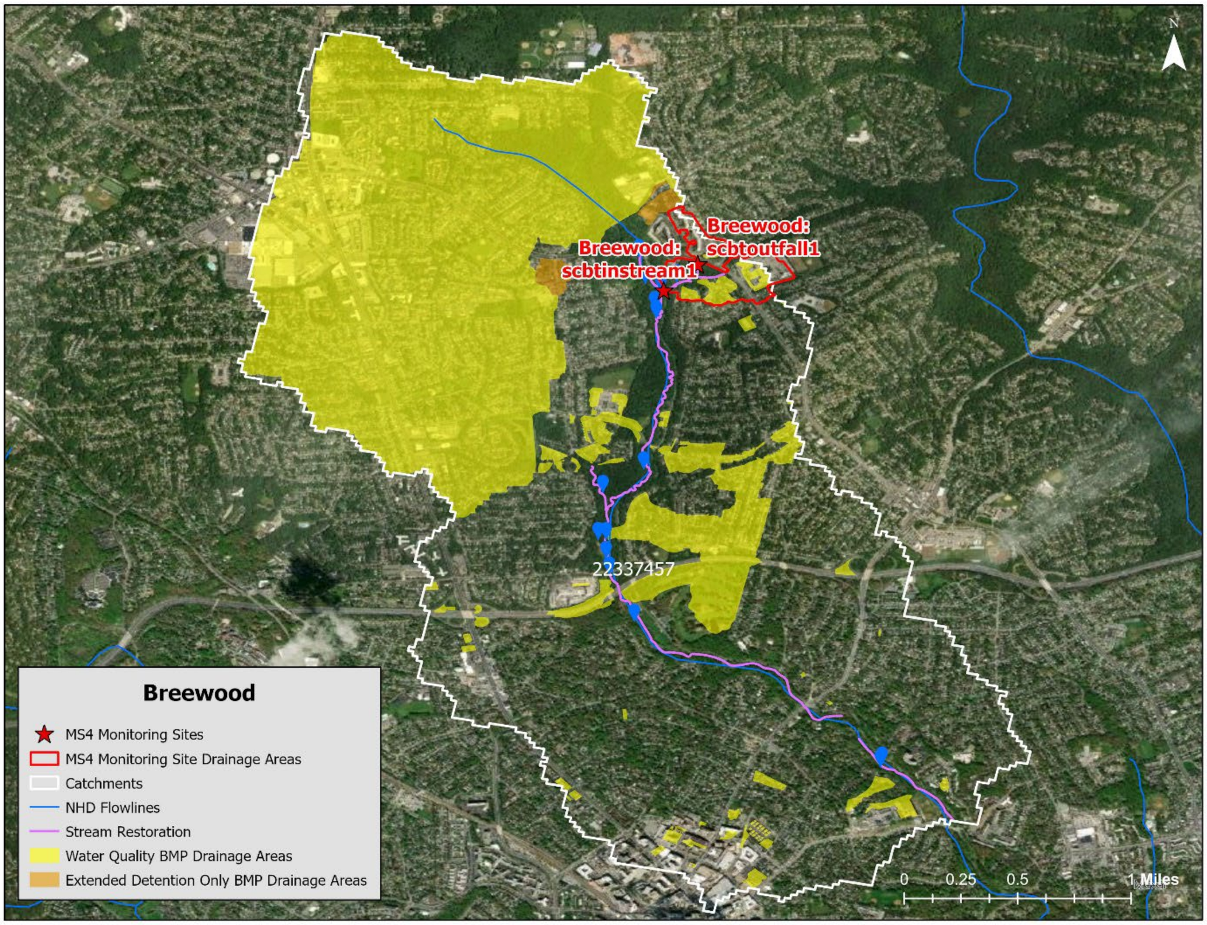

Figure 4. Breewood BMP coverage through 2019.

Breewood was different from most of the other MS4 catchments studied because the MS4 monitoring area for this watershed represented a very small fraction of the catchment area (< 2%); however, the BMP coverage and land cover were very similar between these areas. Therefore, it was decided that data collected from the MS4 stations were comparable to data collected from downstream stations in Sligo Creek, so these additional data points were used for analysis.

BMP implementation and impervious cover were constant in the overall catchment between 2003 and 2019 (Table 4). Impervious cover started at 25% and ended at 26%, and the percent of impervious cover captured by water quality BMPs increased only 4%, from 44% to 48%. The largest change to the watershed was a substantial project that restored more than half of the stream length in the catchment over the monitoring period, with the majority occurring between 2003 and 2008.

Table 4. Impervious cover (IC) and stream restoration analysis of the entire catchment drainage area at Breewood.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	25%	44%	1%	3,890.4
2008	25%	45%	1%	19,278.4
2014	25%	47%	1%	19,278.4
2019	26%	48%	1%	20,482.1

Water quality was well-characterized for conductivity, DO, and pH, but turbidity only had one data point in the MS4 catchment and few for the control and reference catchments. The MS4 catchment had significantly lower DO and higher conductivity than both the reference and control conditions, suggesting potentially degraded water quality. Treatment pH was significantly higher than the reference condition, as was the control (Appendix A Figure A17, Appendix B Table A4).

As the time series indicates, conductivity steadily increased at Breewood from 2000 to 2020, but the same trend was not apparent in reference or control conditions (Appendix A Figure A18). The control trend appeared to be heavily influenced by four extremely low values observed between 2010 and 2020. DO also seemingly increased slightly over time for both treatment and control watersheds.

Habitat data were available for each of the eight habitat metrics (Appendix A Figure A15). The control catchment had the fewest observations, with about half the metrics being represented by fewer than ten data points. Compared to the control condition, the treatment had significantly worse habitat than the control for bank stability, embeddedness, and sedimentation. Data for every habitat parameter was statistically lower in the MS4 catchment than the reference pool. Although the reference condition consistently had higher scores than the control, the only significant differences were for bank vegetation, epifaunal substrate, and flow.

The habitat time series showed a similar pattern to water quality, with the reference condition being uniformly better than the treatment or controls (Appendix A Figure A16). Breewood was similar to the control for both epifaunal substrate and flow metrics, but otherwise more degraded. Embeddedness and sedimentation, which are related to instream sediment transport, showed a similar pattern in which degradation occurred between 2004-2005 and 2012-2013, with improvements after. This initial decline is roughly aligned with large amount of stream restoration in Breewood that occurred between 2003 and 2008.

Breewood has thirty-two Chessie BIBI scores and accompanying metrics, but the control has only five data points (Appendix A Figure A13). Although the median BIBI score for the treatment is slightly lower than the control, neither the score nor any individual metric was significantly different between MS4 and control conditions (Appendix B Tables A1 and A2). The reference BIBI score, as well as every individual metric, was significantly better than both the reference and the control. No discernable trends were visible for the biological metrics or

BIBI scores (Appendix A Figure A14). Although the BIBI score for the control appeared to decline, it is important to note that that pattern is based on only five observations.

Red Hill Branch (UNP)

Red Hill Branch is in Ellicott City, Howard County. The outfall (BH01) is located at the upstream extent of an unnamed tributary to Red Hill Branch, approximately 100 feet downstream of the outfall of the stormwater network originating from Middlesboro Court in the Brampton Hills neighborhood. The instream station (BH02) is located in Meadowbrook Park approximately 250 feet upstream of the confluence with the first order Red Hill Branch (Table 5). There are some industrial areas in the catchment, but most of the area is occupied by neighborhoods and forested areas (Figure 5).

Table 5. Selected attributes of the Red Hill Branch MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
6.74 km^2	5.00 km^2	0.20 km^2	1	38.04	36.73	24.40	0.2 km^2	4.1 km^2

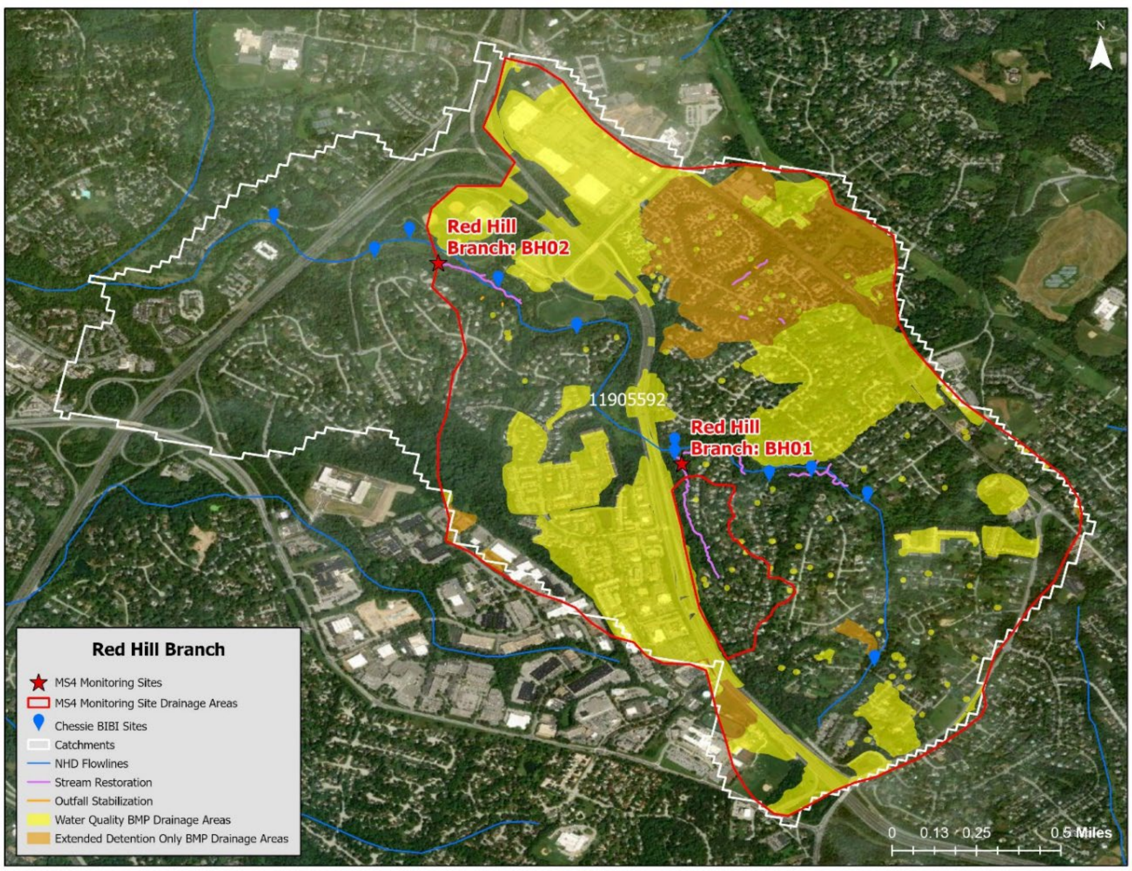

Figure 5. Red Hill Branch BMP coverage through 2019.

Impervious cover in the watershed increased from 23% in 2003 to 26% in 2008, and then remained constant after that point (Table 6). The fraction of watershed imperviousness captured by BMPs steadily increased from 2003 to 2009, increasing from 42% to 53%, while impervious cover captured by detention only decreased slightly over that time period, suggesting that stormwater retrofits were actively implemented and would likely have helped to mitigate runoff from the small amount of development. Some stream restoration occurred in this watershed as well, totaling 8,779 feet of restoration. The majority of the restoration took place between 2008 and 2014, when over 8,000 feet of stream were restored.

Table 6. Impervious cover (IC) and stream restoration analysis of instream drainage area at Red Hill Branch.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	23%	42%	16%	0
2008	26%	47%	14%	100.2
2014	26%	52%	13%	8,199.3
2019	26%	53%	12%	8,779.0

Red Hill Branch has poor data richness for water quality, with only three data points for turbidity and four for all other parameters. While the full suite of habitat metrics was present, only three parameters (embeddedness, epifaunal substrate, and riffles) were represented by more than five data points, and the control condition only had data for five metrics. Biology data were more plentiful, with twenty-six data points in the MS4 catchment.

Although Red Hill Branch had a relatively small amount of data for water quality, results indicate that levels of DO in the MS4 catchment are significantly higher than the control pool, but different from reference conditions (Appendix A Figure A23; Appendix B Table A4). Although conductivity appeared to be highest in the treatment watershed, conductivity was not statistically different between the control and MS4 conditions, but both treatment and controls had significantly elevated conductivity compared to the reference group. Due to the limited data availability in the MS4 catchment, conclusions cannot be drawn about water quality trends over time (Appendix A Figure A24).

Habitat scores in the treatment watershed were not statistically different from the control pool, but the treatment watershed was statistically worse for every metric except channel alteration compared to the reference pool (Appendix A Figure A21; Appendix B Table A3). A similar pattern is present for embeddedness, riffles, and epifaunal substrate in the MS4 catchment over time, where scores dipped around 2010 and then increased (Appendix A Figure A22).

Red Hill Branch displayed consistently worse biological health than the control watersheds as measured by both the BIBI score and every individual metric (Appendix A Figure A19; Appendix B Tables A1 and A2). The MS4 BIBI scores remain about the same over time, always lower than the control and reference conditions (Appendix A Figure A20).

Scotts Level Branch (UNP)

Scotts Level Branch is in Randallstown, Baltimore County. The outfall (SL02) is located off of Tiverton Road, and the instream monitoring station (SL03) is in Scotts Level Branch, a first order tributary to Gwynns Falls (Figure 6). Much of the catchment is residential neighborhoods with local parks and some sports fields (Table 7).

Table 7. Selected attributes of the Scotts Level Branch MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
10.45 km^2	5.84 km^2	0.06 km^2	1	39.76	29.93	30.06	0.2 km^2	8.0 km^2

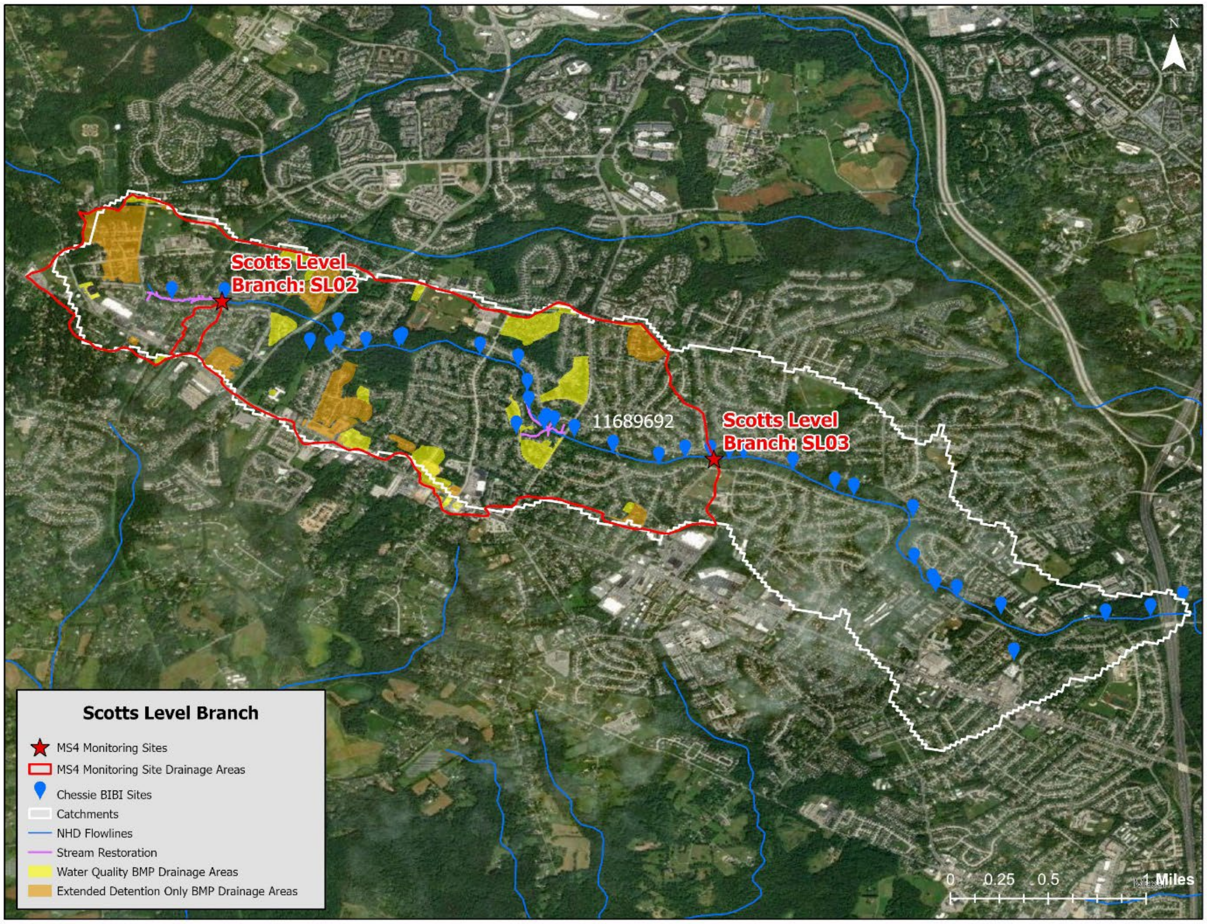

Figure 6. Scotts Level Branch BMP coverage through 2019.

Impervious cover remained constant at 29% in the instream catchment over the period from 2003 to 2019, but the water quality BMPs constructed increased impervious capture from 1% in 2003 to 8% in 2019 (Table 8). The area captured by extended detention practices declined slightly from 15% to 12% in 2019, suggesting that some water quality BMPs were constructed within the drainage area of extended detention practices or an extended detention practice was converted to a water quality BMP. A total of 5,120 feet of stream were restored over the project period, with restoration occurring between 2008 and 2019.

Table 8. Impervious cover (IC) and stream restoration analysis of instream drainage area at Scotts Level Branch.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	29%	1%	15%	0
2008	29%	1%	15%	0
2014	29%	5%	15%	2,734.0
2019	29%	8%	12%	5,120.2

Water quality parameters were generally well-characterized in the treatment watershed, with at least 10 data points for every parameter, except for turbidity which had just one data point. Habitat data were available for five out of the eight metrics, but fewer than 10 data points are available for every metric in the treatment watershed and for all but one (embeddedness) in the control. Finally, benthic data are plentiful in Scotts Level Branch, with 137 points in the treatment condition.

Dissolved oxygen was significantly higher in the treatment watershed than in either the reference or control conditions, which is surprising as the reference catchments would be expected to have the highest DO (Appendix A Figure A29; Appendix B Figure A4). The treatment catchment had the highest median conductivity, but it was only significantly different from reference conductivity measurements. pH in the MS4 catchment was significantly more basic than the reference and control catchments. For the short timeframe that MS4 and control data are available, both conductivity and DO appeared to be increasing for treatment and control conditions. The time series confirms that the observed DO and conductivity for the treatment condition are higher than both control and reference for every observed point (Appendix A Figure A30).

Although median habitat scores were generally lower for the treatment than the control conditions, there was no statistical difference between these conditions for any habitat parameter (Appendix A Figure A27; Appendix B Table A3). Reference habitat scores were generally higher than the MS4, significantly so for embeddedness, epifaunal substrate, and riffles. With the limited data available, apparent trends in treatment conditions are difficult to discern, although the habitat scores are lower for the MS4 in more recent years (Appendix A Figure A25).

Both the Chessie BIBI score and scores for each BIBI metric suggests that biological health in the MS4 watershed is more degraded than in the control, and the reference pool has the best biological health (Appendix A Figure A25; Appendix B Tables A1 and A2). There was potentially an improvement in biological health over time in the MS4 for the BIBI score and some metrics including percent EPT and clinger richness, along with either no change or an apparent decline in the reference and treatment catchments (Appendix A Figure A26).

Peter Pan Run/Urbana (LNP)

Peter Pan Run is in Urbana, Frederick County. The outfall (Pond-R) is at a wet detention pond in the Villages of Urbana development, and the instream station (Peter Pan Run) is in Peter Pan Run, a second order tributary to Bush Creek (Figure 7). This is the only MS4 where the outfall and instream stations are in different catchments. The Pond-R catchment has several neighborhoods, but also some areas of forest and agriculture, whereas the Peter Pan Run catchment is primarily agricultural with some forested areas (Table 9).

Table 9. Selected attributes of the Urbana MS4 catchment.

Site	Catchment Area	Upstream Area	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
Peter Pan Run (instream)	1.39 km^2	6.61 km^2	2	29.79	63.76	6.04	0.8 km^2	0 km^2
Pond-R (outfall)	1.90 km^2	n/a	1	27.46	41.02	30.79	0.5 km^2	0.7 km^2

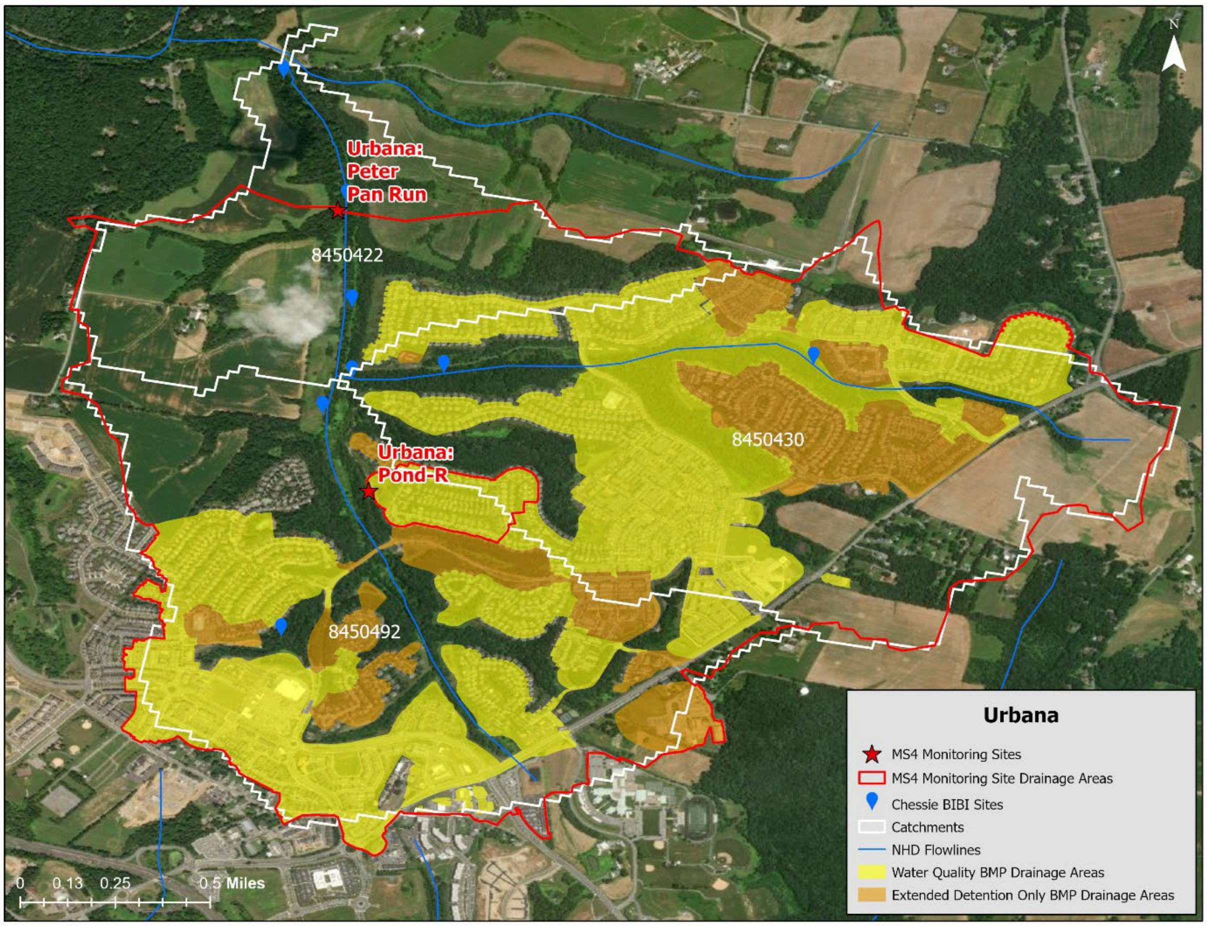

Figure 7. Urbana BMP coverage through 2019.

The analysis and selected controls for Urbana focused on the instream catchment, Peter Pan Run, which included data from Peter Pan Run and the neighboring Pond-R catchment. Urbana experienced substantial development of residential neighborhoods between 2003 and 2008, with watershed impervious cover increasing from 9% to 21% (Table 10). This growth, however, was accompanied by stormwater improvements, so the fraction of impervious cover captured by BMPs increased over time. Between 2003 and 2014, the fraction of watershed impervious area captured by water quality BMPs increased from 6% to 46%, and the amount captured by extended detention practices increased from 0% to 26%. Urbana is the only MS4 to experience both substantial growth and increased stormwater treatment concurrently.

Table 10. Impervious cover (IC) analysis of instream drainage area at Urbana (Peter Pan Run).

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only
2003	9%	6%	0%
2008	21%	22%	19%
2014	25%	46%	26%
2019	25%	46%	25%

Urbana is data rich, with greater than 10 water quality data points for every parameter-catchment combination except for turbidity, where seven and eight data points are present for the control and reference conditions, respectively, but over 100 are present for the treatment condition. Similarly, habitat data are plentiful in the MS4 catchment, with more than 10 data points for every habitat metric. At least 25 data points are available for each benthic metric, with 78 observed points in Urbana.

Although median conductivity in the treatment catchment was higher than the control value, this difference was not significant, but both treatment and control values are elevated compared to the reference. The median turbidity value was not significantly different from reference conditions, but it was significantly higher than the control. Although pH was significantly higher than the reference, pH values do not indicate degradation in either drainage area (Appendix A Figure A35; Appendix B Table A4). Conductivity increased in the treatment watershed over time, without a comparable increase in the control or reference (Appendix A Figure A36).

Habitat scores were generally lower in the treatment than in the control condition, with significant differences for embeddedness, epifaunal substrate, and riffles (Appendix A Figure A33). Trends over time were difficult to interpret, especially for bank vegetation, channel alteration, flow, and sedimentation because there were just a few years of data (Appendix A Figure A34).

Biological health in the treatment catchment was generally more degraded than both the reference and control conditions, and both control and treatment catchments had worse benthic measures compared with reference (Appendix A Figure A31). The BIBI scores for both the treatment and control were significantly lower than the reference. The same was true for most metrics, with the exception of percent Plecoptera and Plecoptera richness, both of which are not statistically different between the control and reference sites (Appendix B Table A1 and A2). While each individual metric was statistically worse in the treatment than in the control condition, the BIBI score difference was not quite significant (p-value of 0.055). Trends in BIBI metrics were static for control and reference catchments; however, Chessie BIBI scores appear better after 2010 (Appendix A Figure A32).

Stewart-April Lane/Paint Branch (LNP)

Stewart-April Lane is in White Oak, Montgomery County. The outfall monitoring station (PBPB104) is located in the Stewart-April Lane tributary, and the instream station (PBPB310A) is located in Paint Branch, a second order stream (Figure 8). Much of the catchment is comprised of suburban neighborhoods and forested areas (Table 11). The instream station (Paint Branch) captures runoff from a much larger contributing drainage area, with a total of 13 upstream catchments.

Table 11. Selected attributes of the Stewart-April Lane MS4 catchment.

Catchment Area	Upstream Area	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
5.43 km^2	33.97 km^2	2	57.01	19.57	22.69	0 km^2	4.7 km^2

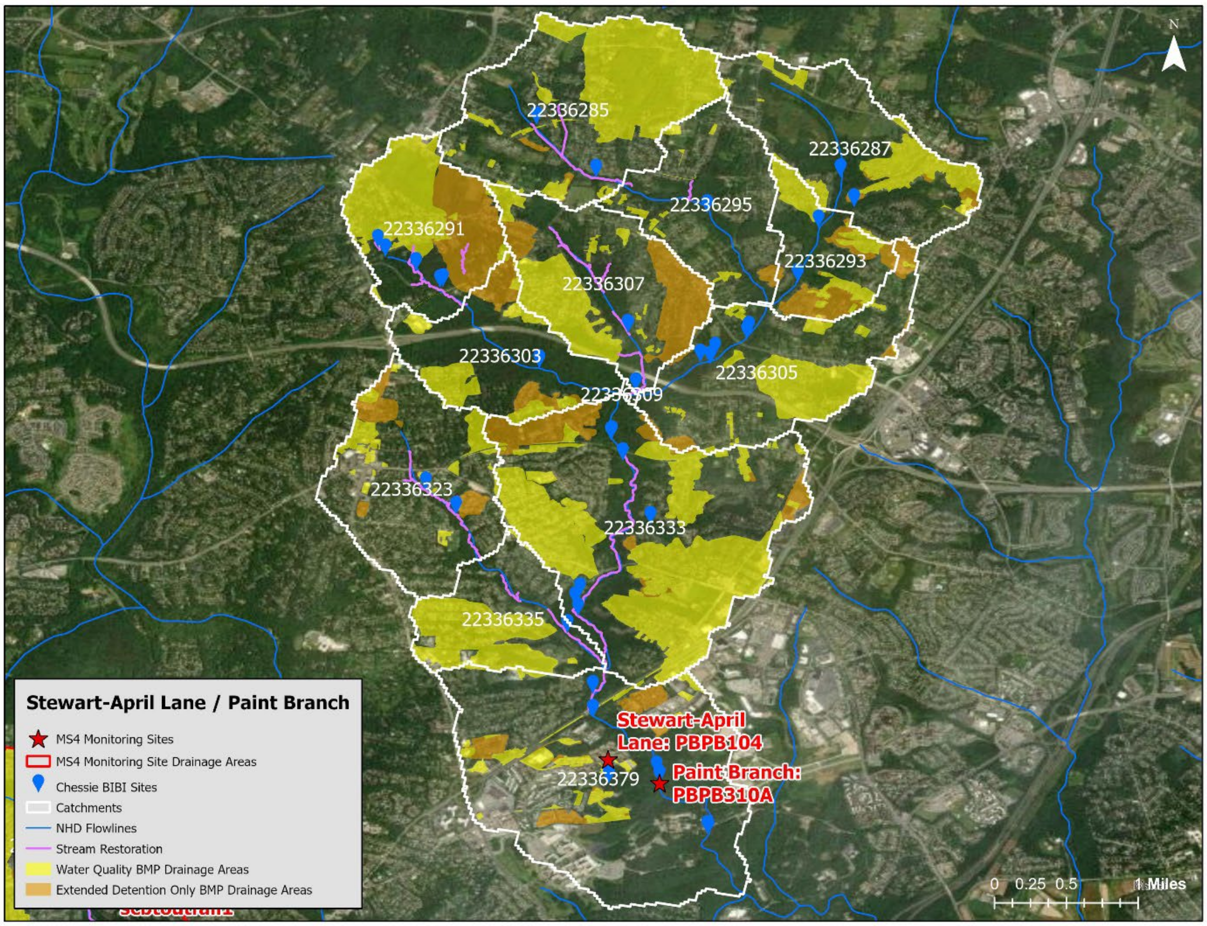

Figure 8. Stewart-April Lane BMP coverage through 2019. Upstream catchments included.

The amount of impervious cover in the entire drainage area did not change much between 2003 and 2019, increasing only 2%, from 14% to 16% (Table 12), but substantial BMP implementation and stream restoration occurred over that time. In 2003, only 16% of the watershed impervious cover was captured by water quality BMPs, but that increased to 34% over the same time period, with the majority of the BMP implementation occurring between 2003 and 2008. In addition, over eight miles (43,611 feet) of stream restoration have been implemented in this watershed, including over 28,000 feet between 2003 and 2019. The initial analysis for this drainage area evaluates both the drainage area characteristics and water quality for the entire drainage area, but a separate analysis (Appendix A Table A26) evaluates current conditions for each catchment separately.

Table 12. Impervious cover (IC) and stream restoration analysis of instream drainage area at Stewart-April Lane.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	14%	16%	9%	15,539.0
2008	15%	28%	10%	16,260.6
2014	15%	32%	10%	28,725.5
2019	16%	34%	9%	43,610.8

Since this MS4 includes a large drainage area with several Chessie BIBI monitoring stations, a very large dataset was available to characterize the instream station. Unfortunately, only one control catchment matched the characteristics of the MS4 catchment, so the control conditions are not well characterized, with only four events for water quality, habitat and benthic parameters or metrics.

Although the treatment had many favorable DO concentrations, median DO is significantly lower than the reference, but not the control. Conductivity was statistically higher in the treatment than both the reference and control conditions, but no other significant differences in water quality were found, and the general range of water quality metric values are comparable to the reference and control conditions (Appendix A Figure A41). Despite some outliers, especially in MS4 conductivity, water quality conditions appeared relatively stable over time (Appendix A Figure A42).

Habitat scores in the MS4 catchment were higher than the control condition, but only statistically higher for embeddedness (Appendix B Table A3). Compared to reference conditions, treatment habitat scores were significantly worse (Appendix A Figure A39). The scatterplots in Appendix A Figure A40 suggest a gradual decline in bank vegetation, bank stability, embeddedness, and epifaunal substrate. Over this time period, large lengths of stream restoration were likely beginning or recently adjusting to restoration projects.

Chessie BIBI scores in the MS4 and control catchments were not statistically different, but both were significantly worse than reference scores, although some very good BIBI scores were found in the treatment catchment (Appendix A Figure A37; Appendix B Table A1). For the benthic metrics, treatment and control watersheds were not significantly different for any metric other than percent Plecoptera, which was higher in the control. Compared to reference benthic communities, the treatment scored worse than for every biological metric. No obvious trends were observed over time (Appendix A Figure A38).

Southeastern Plain

Acton/Hamilton (SEP)

Acton/Hamilton is in Waldorf, Charles County. The outfall (AH001) is located off Westdale Court, and the instream station (AH002) in in a first order tributary to Mattawoman Creek (Figure 9). The middle portion of the catchment is forested, and the rest is a mix of suburbs and businesses (Table 13).

Table 13. Selected attributes of the Acton/Hamilton MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler Order	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
3.73 km^2	2.56 km^2	0.82 km^2	1	38.92	26.21	34.33	0.01 km^2	2.8 km^2

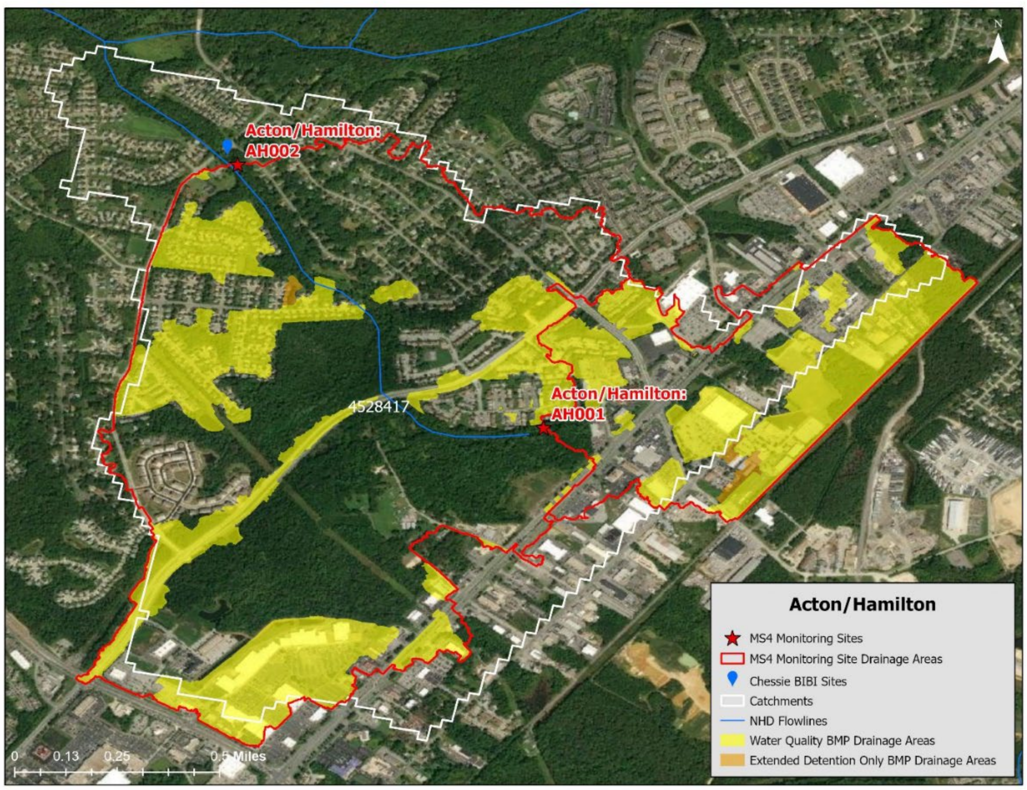

Figure 9. Acton/Hamilton BMP coverage through 2019.

This MS4 monitoring area experienced a slight but steady increase in impervious cover from 32% in 2003 to 35% in 2019 (Table 14). Simultaneously, water quality BMPs and retrofits were implemented such that that the fraction of watershed impervious cover treated by stormwater BMPs gradually increased over time. One notable feature in this watershed is the large, contiguous forest area bordering the stream mainstem, which serves as a buffer between development and Mattawoman Creek.

Table 14. Impervious cover (IC) analysis of instream drainage area at Acton/Hamilton.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only
2003	32%	37%	1%
2008	34%	47%	1%
2014	35%	49%	1%
2019	36%	47%	1%

The water quality boxplots for the treatment and control catchments appear very similar and are not statistically different for any metric (Appendix A Figure A47; Appendix B Table A4). Compared to reference conditions, the MS4 catchment has significantly higher conductivity and pH, but not significantly different levels of turbidity or DO. Although the pH values are significantly higher than reference conditions, they are still within the optimal range of 6-8.

Although only two habitat parameters were available for the MS4 catchment (bank stability and epifaunal substrate), both habitat measures scored highly. Bank stability was significantly higher than reference conditions, and epifaunal substrate was significantly higher than the controls and comparable to reference (Appendix A Figure A45; Appendix B Table A3).

Chessie BIBI scores were significantly better in the reference catchments compared to the MS4 and control catchments, but MS4 scores were not significantly different from the controls (Appendix A Figure A43; Appendix B Table A1). Biological metrics were not statistically different between control and treatment conditions with the exception of percent burrowers, which was greater in the Action/Hamilton treatment watershed (Appendix A Figure A43; Appendix B Table A2). The BIBI score in the treatment watershed appears to improve slightly over the monitoring period, but individual metrics remain relatively constant (Appendix A Figure A44).

The more favorable stream conditions at Acton/Hamilton likely reflect the benefit of the large forest buffer in the MS4 catchment and that almost half of the impervious cover is treated by water quality BMPs after 2008 as turbidity, DO, bank stability, and epifaunal substrate are similar to that of reference conditions.

Bear Branch (SEP)

Bear Branch is in Laurel, Prince George’s County. There is no outfall station. Rather, instream stations were established in 2007 to monitor the results of restoration efforts aimed at reducing sediment in the downstream Laurel Lakes. The stations are Contee Road/003 and Chapel Cove Drive/005. Both stations are on the mainstem of Bear Branch, with Contee Road in the middle of the watershed adjacent to Contee Road, and Chapel Cove located approximately 2,400 feet downstream of Contee Road, ~200 feet from the dead end of Chapel Cove Drive (Figure 10). A fairly equal mix of residential, forest, and business/industrial areas comprise the catchment (Table 15). Although this catchment is first order, it has one upstream catchment that contains the stream origin.

Table 15. *Selected attributes of the Bear Branch MS4 catchment.*

Catchment Area	Upstream Area	Instream Drainage (Contee)	Instream Drainage (Chapel)	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
3.12 km^2	4.96 km^2	2.91 km^2	4.41 km^2	1	39.19	29.56	28.40	0.1 km^2	2.0 km^2

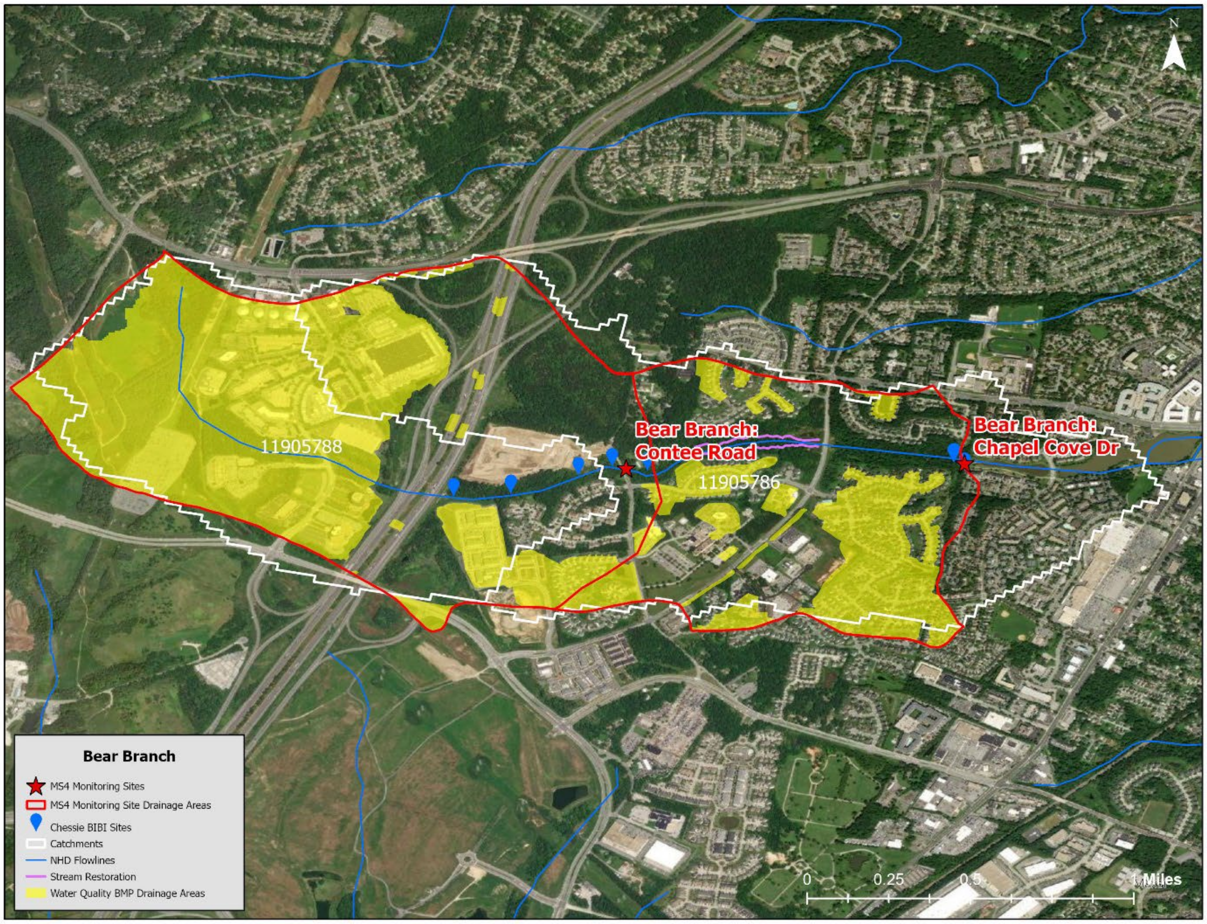

Figure 10. Bear Branch BMP coverage through 2019.

Impervious cover the Bear Branch drainage increased over time, with development occurring between 2003 and 2008 and again between 2014 and 2019 (Table 16). The additional impervious cover was treated by water quality BMPs, thus increasing the total percent of impervious captured by water quality practices. The stream restoration aimed at reducing sediments in Laurel Lakes was implemented between 2008 and 2014 downstream of the Contee Road Station.

Table 16. Impervious cover (IC) and stream restoration analysis of instream drainage area at Bear Branch.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	20%	54%	0%	0
2008	25%	61%	0%	0

2014	25%	65%	0%	3,620.2
2019	28%	65%	0%	3,620.2

A large amount of water quality data were available for the treatment, control, and reference conditions, except for some of the habitat metrics. Conductivity was higher in the treatment and control watersheds than in the reference group, but not significantly different from one another. Surprisingly, DO was higher in the treatment than in either the reference or control groups (Appendix A Figure A53; Appendix B Table 4). Conductivity increased in both the control and treatment watersheds between 2000 and 2015, with a few low measurements in the treatment watershed towards the end of the time period. In the MS4 catchment, the disturbance caused by the stream restoration may account for the elevated conductivity above the control (Appendix A Figure A54).

Habitat results show that treatment catchments are significantly worse than controls for bank stability, epifaunal substrate, flow, and sedimentation. Although the treatment condition is generally more degraded than reference, the only significant differences were for channel alteration and sedimentation (Appendix A Figure A51; Appendix B Table A4). The time series in Appendix A Figure A52 indicate a loss in habitat quality over time in Bear Branch for each habitat metrics. Continued monitoring after the Laurel Lakes restoration project would be needed to determine its effectiveness.

The Chessie BIBI score in the treatment condition was significantly better than in the control condition, but results were more mixed for individual metrics. For most metrics, the control and treatment metrics were not significantly different. Richness of moderately tolerant taxa and percent burrowers were significantly different. (Appendix A Figure A49; Appendix B Tables A1 and A2).

Moores Run (SEP)

Moores Run is located in northeast Baltimore City. The outfall station (Hamilton) is located off Hamilton Avenue, and the instream station (Radecke) is located just north of Radecke Avenue in the first order Moores Run, a tributary to the Back River (Figure 11). Much of the catchment is residential and commercial with few local parks and a narrow forest buffer around the stream (Table 17).

Table 17. Selected attributes of the Moores Run MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
12.86 km^2	9.10 km^2	7.52 km^2	1	33.17	24.23	42.17	0 km^2	9.7 km^2

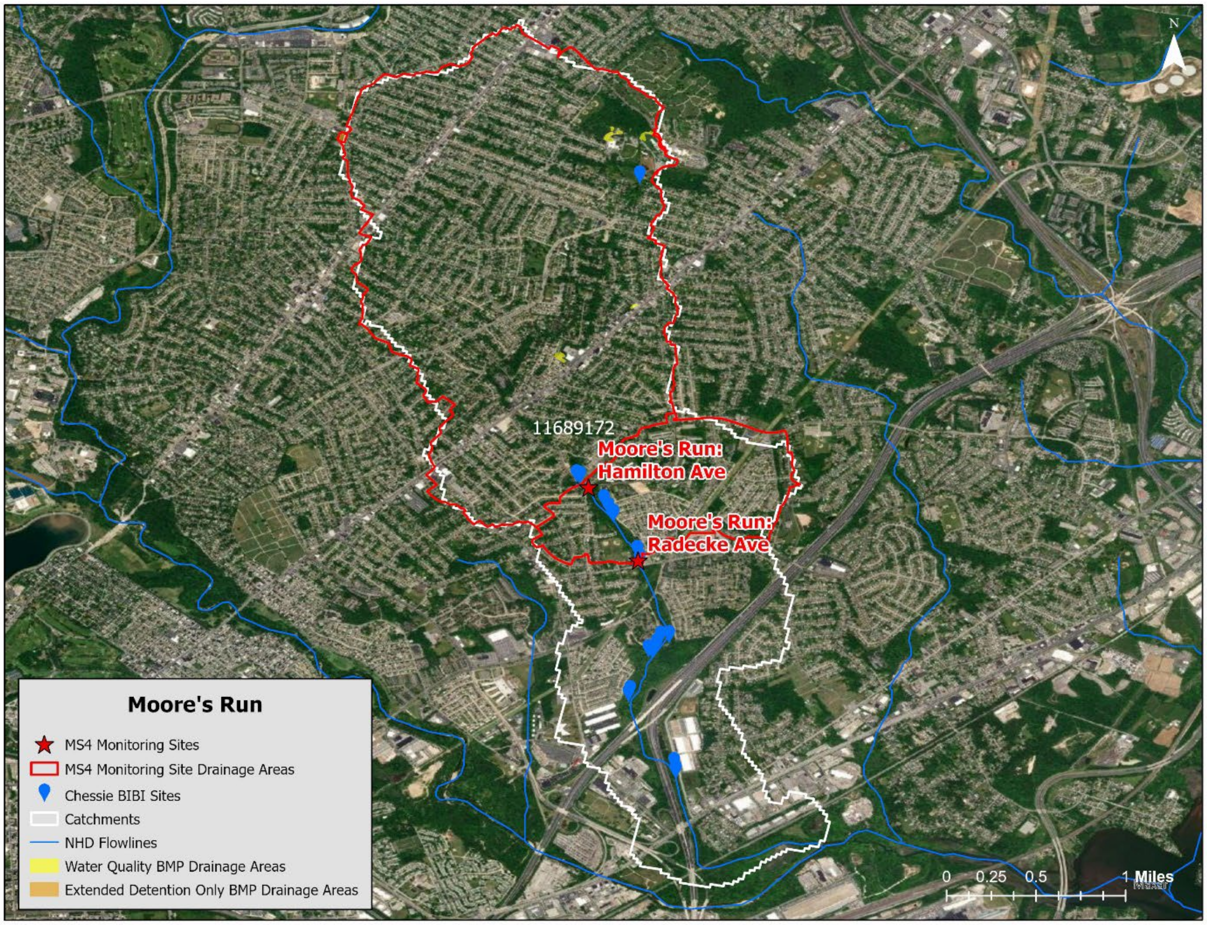

Figure 11. Moores Run BMP coverage through 2019.

Impervious cover in the Moores Run catchment remained constant at 46% from 2003 to 2019 (Table 18). Although Moores Run is considered a treatment watershed because it is an MS4 location, it had almost no BMP implementation, with the majority being non-traditional BMPs like street sweeping and catch basin cleanouts. There were, however, significant repairs made to the sewershed between 2003 and 2008 with almost 24,000 linear feet of gravity pipes replaced (Jepsen & Caraco, 2020). These improvements were linked to significant decreases in stormflow *E. coli* and NO_{23} and baseflow copper and zinc.

Table 18. Impervious cover (IC) analysis of instream drainage area at Moores Run.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only
2003	46%	0%	0%
2008	46%	1%	0%

2014	46%	1%	0%
2019	46%	1%	0%

The Moores Run catchment has at least 10 data points for every water quality and biology parameter or metric, but is more data-limited for habitat, with only embeddedness, epifaunal substrate, and riffles having more than ten data points. Control data are lacking in water quality and habitat, with no data for turbidity and some habitat metrics.

No significant differences were found between the treatment and control watersheds for any water quality parameter or between the control and reference watersheds (Appendix A Figure A59; Appendix B Table A4). The MS4 watershed had elevated DO and conductivity compared to the reference, but unlike most catchments studied, there was no apparent increase over time in conductivity in Moores Run (Appendix A Figure A60). The pH was significantly higher in the treatment watershed than in the reference pool. Although the pH values are not necessarily associated with degradation, the high conductivity values suggest degradation, although the DO values are favorable for aquatic health (Appendix A Figure A59).

Habitat scores are in general quite good in the treatment watershed, and not statistically different from the reference condition. The range of values is also smaller than the reference and controls except for epifaunal substrate and riffles (Appendix A Figure A57, Appendix B Table A4).

Biology in the treatment catchment is similar to controls overall, with no significant difference in the Chessie BIBI scores, and the control only has better biological metric values for Margalefs and richness of moderately tolerant taxa. Both the treatment and control catchments are significantly worse than the reference for the BIBI and every biological metric (Appendix A Figure A55, Appendix B Tables A1 and A2). There was no apparent change in biological health over time at Moores Run (Appendix A Figure A56).

Parole Town Center/Church Creek (SEP)

Parole Town Center is in Annapolis, Anne Arundel County. The outfall station (Parole Plaza) is located at the head of the Parole Tributary to Church Creek at the Southwest corner of Forest Drive and MD State Highway 2. The instream station (Church Creek) is 2,000 feet downstream of the outfall in first order Church Creek, a tributary to South River (Figure 12). The catchment land use is divided between residential, forest, and a large shopping center (Table 19).

Table 19. Selected attributes of the Parole Town Center MS4 catchment.

Catchment Area	Instream Drainage	Outfall Drainage	Strahler	%Trees, Shrubs	%Herbaceous	%Impervious	Septic	Sewer
3.63 km^2	1.14 km^2	0.43 km^2	1	41.85	17.40	39.36	0.3 km^2	2.3 km^2

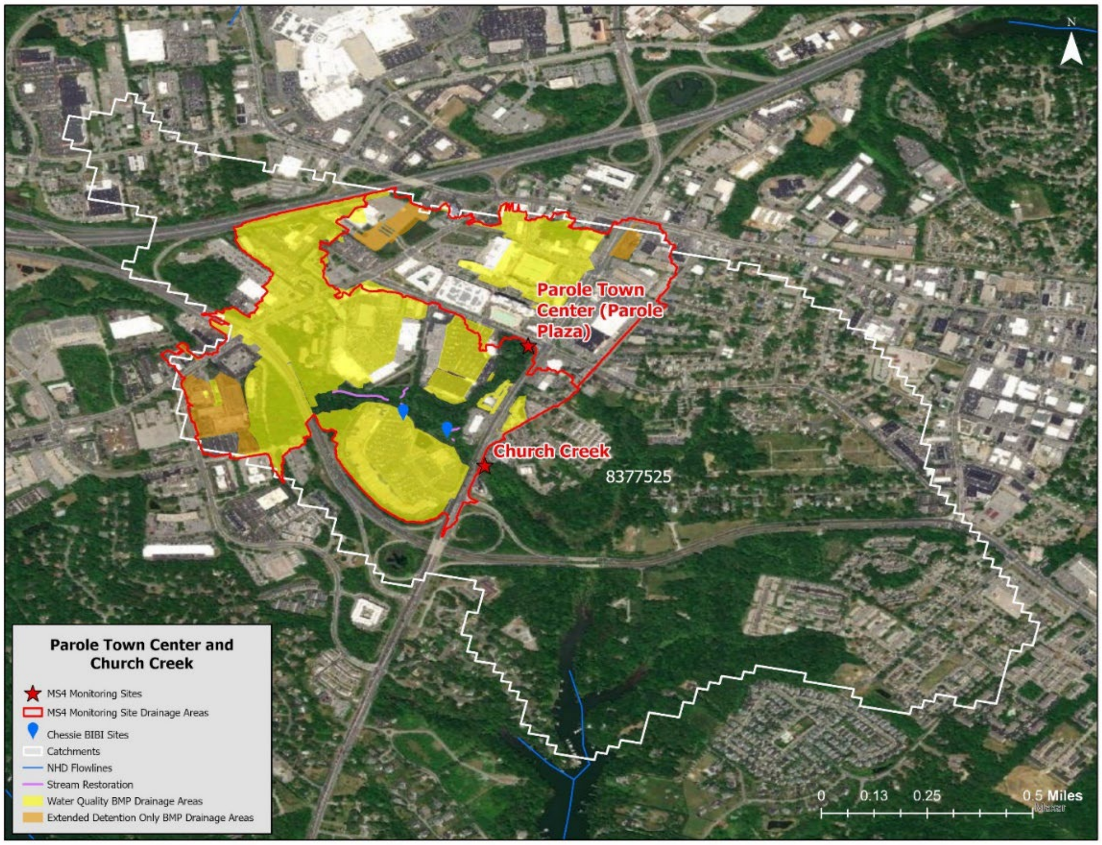

Figure 12. Parole Town Center BMP coverage through 2019.

The impervious cover in the MS4 drainage area remained relatively constant, and treatment by water quality BMPs increased over time as retrofits were implemented between 2008 and 2019 (Table 20). In addition, a small length of Church Creek was restored between 2014 and 2019.

Table 20. Impervious cover (IC) and stream restoration analysis of instream drainage area at Parole Town Center.

Year	%IC	%IC Captured by Water Quality BMPs	%IC Captured by Detention Only	Cumulative Feet of Stream Restoration
2003	68%	33%	6%	0
2008	66%	34%	6%	0
2014	66%	42%	6%	0
2019	66%	50%	5%	827.4

In the Church Creek MS4 catchment, conductivity and turbidity were significantly increased compared to the reference catchments. Dissolved oxygen levels were statistically the same between all three catchment types. There was no pH data for the treatment watershed (Appendix A Figure A65; Appendix B Table A3). Data were sparse for the controls, so it was difficult to see a trend over time. Also, the reference and treatment data did not overlap much, so comparisons are hard to make about comparative change over time (Appendix A Figure A66).

Epifaunal substrate was the only habitat metric with data in the MS4 catchment, and it was significantly lower than the reference condition, although the range of scores for references was large (Appendix A Figure A66; Appendix B Table A4).

The BIBI scores and all biological metrics except percent burrowers were not significantly different between treatment and control conditions, both of which were significantly worse than the reference for every metric except richness of moderately tolerant taxa and shredder richness (Appendix A Figure A61; Appendix B Tables A1 and A2 A3). Although the data from the treatment condition do not represent the entire analysis time period, there were no apparent trends over time (Appendix A Figure A62).

Overarching Patterns

The initial concept for this project was based on the stream functions pyramid (Figure 13), which describes the building blocks needed for optimal stream health. The five tiers are, from the foundation, hydrology, hydraulics, geomorphology, physiochemical, and biology. Biology, the top level, is influenced by each of the underlying tiers, and good biological health should be an indication that all of those tiers are in good condition. The following three subsections discuss the general patterns observed in the individual MS4 analyses relating to the biology, habitat, and water quality sections of the pyramid. Appendix A Figure A43 also summarizes the findings.

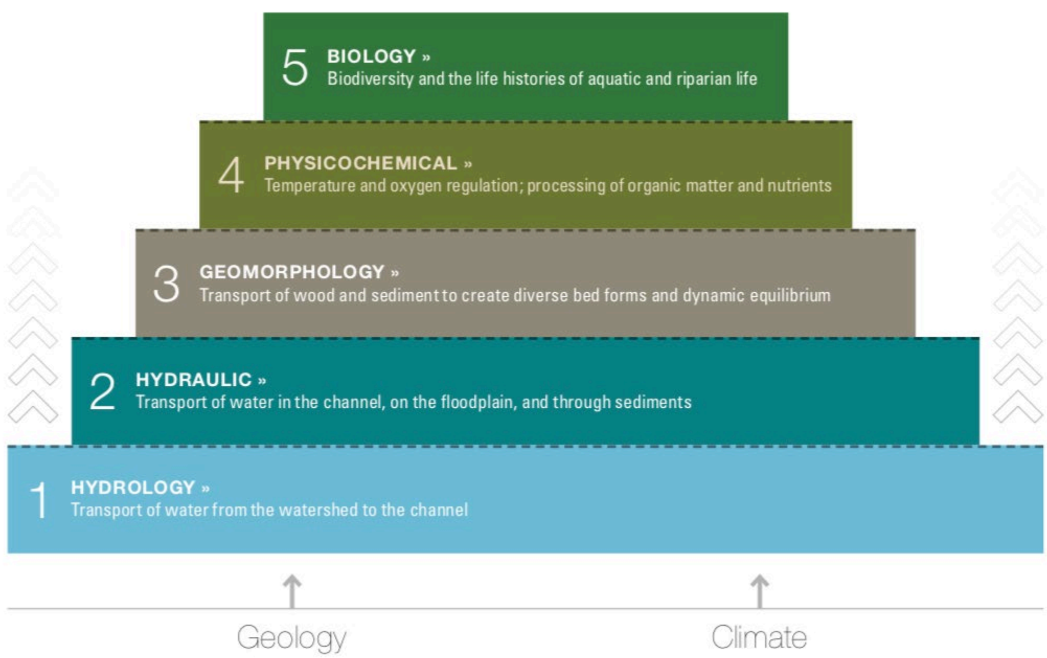

Figure 13. Stream functions pyramid. Reprinted from EPA, Stream Mechanics.

Habitat

Data for the hydrology and hydraulics steps were not widely available (see the Limitations section for more detail). The habitat parameters used to represent the physical habitat tier often had lower data coverage compared to water quality and biology, more so for control and MS4 catchments. There was also a tendency for these parameters to have a large range in the boxplots, which likely reflects both the variety of habitat conditions present in Maryland streams and the subjective nature of habitat scoring which is based on a visual assessment.

The habitat metric boxplots in the preceding section show that reference catchments typically had favorable habitat, with scores of 20 being common. MS4 and control catchments most often scored lower than the reference pool. Compared to the controls, treatment catchments scored better, the same, and worse depending on the metric and MS4 site, so there was not a uniform pattern of habitat conditions in MS4 watersheds. Furthermore, there does not seem to be an immediate response of the physical habitat to MS4 implementation as habitat scores did not increase measurably over time in any MS4 catchment, except possibly in Airpark Business Center.

Four MS4 catchments (Bear Branch, Peter Pan Run, Scotts Level Branch, and Stewart-April Lane) scored a 20 on at least one habitat metric. Bear Branch had scores of 20 for channel alteration around 2011, but the scores decreased after that, which is probably related to the Laurel Lakes restoration project. The best score at Peter Pan Run was for bank stability and occurred earlier in time with other higher bank stability scores, between 2000 and 2003/2004. During this time, the Villages of Urbana were being developed, which explains the decrease in bank stability and other habitat scores. It does not seem like enough time has passed for the stream habitat to recover. Scotts Level Branch was scored 20 for bank stability around 2017, but in the same year it also scored 4 and 6, so the 20 may be an error. At Stewart-April Lane, although there is a wide range of scores, instream habitat appears closer to the reference pool than the other MS4 sites, and scores of 20 were given to bank vegetation, channel alteration, epifaunal substrate, and riffles, though the conditions appeared to gradually be on the decline over the years. This is an interesting finding because Stewart-April Lane has thirteen upstream catchments so more degraded conditions would not be surprising; however, there has been a great deal of stream restoration in these upstream catchments, and impervious cover is relatively low both in the MS4 catchment and upstream.

Water Quality

The water quality tier of the stream functions pyramid is represented in this study by DO, pH, turbidity, and conductivity. One of the functions of BMPs is to settle out sediments, and the success of sediment removal is best measured by turbidity. Unfortunately, turbidity was often the parameter with the least amount of data; Airpark Business Center and Bear Branch had no turbidity measures. A proxy for sedimentation could be the instream habitat, particularly embeddedness and epifaunal substrate.

Conductivity was consistently elevated in both the treatment and control catchments compared to reference conditions, which is expected because reference conditions have less imperviousness, less road salt application, and fewer disturbances that cause increased runoff. In most MS4 treatment watersheds, conductivity was higher compared to controls, which was unexpected. It was thought that conductivity would be comparable between MS4s and controls because they have similar levels of imperviousness. Since this was not the case, it is possible that treatment catchments have higher conductivity because they are often in areas with more recent land disturbance (e.g., BMP installation, land development) or perhaps MS4 watersheds are disproportionately located in areas with wider streets that are more heavily salted in the winter. This could partially explain the lower BIBI scores found in many MS4 catchments compared to their controls.

pH levels in treatment and control catchments were often the same or slightly higher (i.e., more alkaline) than in reference areas. Alkalinization of fresh water in concurrence with salinization is a known process that occurs in developed watersheds due to inputs of basic salts and carbonates from anthropogenic sources and accelerated weathering of rock (Kaushal et al., 2018).

Levels of dissolved oxygen were typically higher in the reference pool compared to controls and treatments. Interestingly, in some MS4 catchments there were instances of DO being greater than references (e.g., Bear Branch, Moores Run, and, in particular, Scotts Level Branch). Health of aquatic organisms declines quickly when DO concentrations are below 5 mg/L in freshwater. Reassuringly, DO measured in the reference, control, and treatment catchments rarely were lower than 5 mg/L. A few suspiciously low values at the SEP reference sites appear to be outliers, as reference DO often was measured at and above 10 mg/L. Of the MS4s, Breewood, Peter Pan Run, and Stewart-April Lane had a few DO values under 5 mg/L.

Biology

The Chessie BIBI index and various individual macroinvertebrate metrics were used to assess the biological tier. Reference conditions always had better median BIBI scores than either the treatment or control catchments, a reflection of the relatively good water quality and undisturbed habitat conditions in reference systems (Figure 14). Patterns in the individual metrics were similar, with reference communities having higher family richness and higher percentages of taxa associated with good water quality (e.g., Plecoptera, Ephemeroptera, Trichoptera, shredders, clingers).

The lower BIBI scores in the MS4 and control catchments suggest these areas are experiencing other stresses, even long after BMP installation, and are not likely to improve even with significant improvements in one tier of the stream pyramid. For example, an undocumented stressor such as increased temperatures or toxin contamination would still degrade the community after BMP installation. Dominance of pollution-tolerant taxa in the communities may also need time to be overcome as sensitive taxa migrate into the area.

Stream communities could also still be recovering from the disturbance and degradation caused by recent (2014 – 2019) – as opposed to early (2003 – 2008) or mid (2008 - 2014) period – BMP construction, continuing development, and/or stream restoration efforts. BIBI scores in the Peter Pan Run and Scotts Level Branch MS4 catchments were sharply lower than those in their control catchments (Figure 14). Both experienced recent (2014-2019) BMP construction and Scotts Level Branch also experienced recent stream restoration. Both are showing significantly improving BIBI trends. The Red Hill Branch, Airpark Business Center, and Stewart-April lane MS4 catchments also had much lower BIBI scores than their controls. Red Hill Branch had ongoing BMP construction and development; Airpark Business Center had ongoing development; and Stewart-April Lane had a large ongoing stream restoration project.

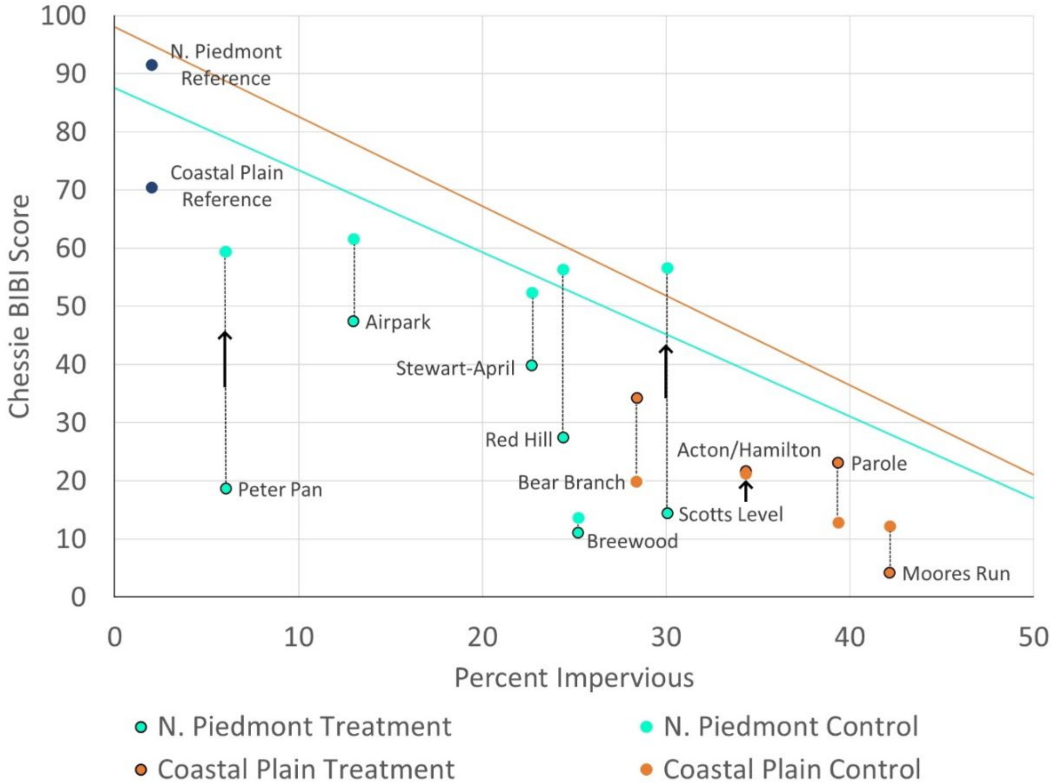

Figure 14. Effect of percent imperviousness on Chessie BIBI scores in reference, treatment, and control catchments. Trend lines are the 95th percentiles from Appendix A Figure A1. Up arrows indicate significant improvement occurring during the study's 2003 – 2019 period.

Poorer BIBI scores in the controls compared to the reference were expected and observed, particularly as percent impervious cover increased. BIBI scores calculated from raw taxa counts in the MS4 treated catchments were also always lower than the reference. Of the 107 samples in the ten MS4 catchments, only one scored as "Good," with 89% of the samples scoring "Poor" or "Very Poor" and the remainder scoring "Fair." While scores in the controls and MS4 treated catchments both declined as percent impervious cover increased, there were no consistent differences between MS4 and control scores (Figure 14). Eight MS4 catchments had median BIBI scores the same as or *lower than* their controls; Bear Branch and Parole Town Center had median BIBI scores higher than their controls. It was somewhat surprising how often MS4 treated catchments scored the same as or worse than their controls since the controls had few or no BMPs installed during the study period. BIBI scores in the remaining five MS4 catchments were similar to or better than their control catchments. Acton/Hamilton, Parole Town Center, and Bear Branch were experiencing recent stream restoration efforts and Parole Town Center had some ongoing BMP construction. However, development and/or BMP construction in these five catchments occurred for the most part in the early period (2003 – 2008) or before. Acton/Hamilton showed an increasing trend in BIBI scores, but the other four MS4 catchments did not, suggesting they had stabilized.

Until this point, comparisons between treatment, reference, and control catchments have not factored in the *level* of BMP implementation. Combined effects models of BMP implementation and impervious cover were used to quantify the effects of BMP implementation across all treatment and control watersheds (Appendix A Figures A67-69, Appendix B Table A8). The analysis also indicated that BMPs can be associated with a significant decline in EPI_SUB and an increase in conductivity but may have no significant effect on either the BIBI score or the BANKS parameter. These results corroborate the paired-watershed results which also showed mixed, and more often worse, stream health associated with BMP implementation.

These results suggest that biological "lift" may take time to achieve after BMP installations or certain levels of stream restorations are done but given time BIBI scores in MS4 catchments can reach and sometimes exceed the scores in their controls. Absent extreme measures, the amount of biological improvement that can be achieved will always be limited by the level of impervious cover.

Hypothesis 3: Effects of Water Quality and Habitat on Biological Health

The third hypothesis states that biological stream health can be predicted by hydrologic, geomorphic, habitat, and water quality indicators. In cases where there is a convergence of predominantly forested land cover, high-quality instream habitat conditions, good water quality, (i.e., reference conditions) and no fish stocking or presence of invasive species, the likelihood of a healthy, resilient aquatic community - represented by "Excellent" and "Good" BIBI scores - is very high. In cases where land cover is largely developed, with high levels of impervious cover and degraded water quality and/or instream habitat, aquatic communities are almost always in poor condition. This is clearly shown in Figure 14. Outside of these two extreme cases, however, it is usually difficult to predict biological health from water quality and instream habitat. This can be due to a lack of information about *other* aspects of the stream's ecosystem represented in the five tiers of the stream pyramid (Figure 13).

When sufficient data for known stressors are available, the CART analysis method can be used to determine when individual stressors begin to degrade stream biological health. Models are built to test the relative importance of two or more factors in a group of catchments. The results provide a somewhat more nuanced understanding of the paired watershed comparisons. Three examples are described below and illustrated in Appendix A Figures A70-A72.

The first model (Appendix A Figure A70) was developed using SPCOND, as well as the two habitat parameters (BANKS and EPI_SUB). SPCOND alone was the first factor to separate good and bad BIBI scores, with conductivity above 125 associated with the poorest BIBI scores, and below 89 associated with the best scores. Although the quantile regressions of these parameters suggested that higher habitat scores were significantly associated with declining BIBI scores, the effect of water quality (conductivity) appears to be more predictive of biology overall, even overriding the habitat parameters.

A second model (Appendix A Figure A71) relates the BIBI score to impervious cover and water quality BMP implementation. Although regressions suggested that water quality BMPs had no significant effect on biology, the model suggests that there may be a beneficial effect at high (above 25%) levels of impervious cover *and* high (above 59%) levels of BMP implementation.

Finally, a third model (Appendix A Figure A72) relates the BIBI score to land cover metrics (impervious cover and forest cover). The data suggest that impervious cover affects biological health of streams, with apparent breaks at 14% and 22% impervious cover. Unlike BMPs, which appear to have the greatest impact at the highest levels of imperviousness, forest is most effective at low imperviousness (less than 14%), with worse biological health in catchments with less than 36% forested pervious cover.

In layman's terms, planting trees in large backyard lawns will improve neighboring stream health when impervious cover in a catchment is relatively low. When impervious cover is relatively high, installing more BMPs will improve stream health in the catchment.

Limitations

Data

When assembling datasets, the original goal was to incorporate metrics from each level of the stream functions pyramid (Figure 13), however, this proved to be difficult. While the MS4 monitoring and Chessie BIBI databases provided data for the biology, physiochemical, and geomorphology tiers of the pyramid, they had limited data to represent the hydraulic and hydrology tiers, measures of which include surface/groundwater exchange and floodplain connectivity for stream hydraulics, and flow duration, precipitation/runoff ratio, and flood frequency for hydrology. USGS flow gages are an excellent source of hydrology measures, but only Moore's Run is currently gaged, and a proxy for flow was not found for the other MS4 catchments. Under new MS4 permits, jurisdictions will be required to monitor flow, which will be useful for future projects.

Additionally, some datasets that were assembled were not incorporated into the analyses. One of these datasets was records of repairs to sewer infrastructure. Only a few of the contacted jurisdictions were able to provide data, mainly due to time constraints because detailed documentation can be difficult to obtain in useful formats. Next, the stream category classification associated with BIBI samples was not integrated into an analysis. Another dataset not used was precipitation, which, without associated metrics, is less useful. A final unused source of data was the Water Quality Portal database. Stream data were downloaded and matched to the catchments of interest, but it proved to be challenging to ensure that no duplicate records were present since the station names in the BIBI database and Water Quality Portal differed. There were also questions of data quality. For example, some chemistry values had no associated units or analysis methods, and very low (<1) pH values were present in the dataset. Despite not using the Water Quality Portal, there was sufficient data to produce plots and run statistical tests once the control and reference data were pooled.

Methods, Metrics, and Analyses

There are a number of caveats relating to the inherent features and selection of treatment (MS4), control, and reference catchments. First, MS4 monitoring watersheds were generally targeted by MDE with the goal of implementing restoration practices. Some watersheds would have been selected because of apparent degradation, such as Bear Branch, which served as a monitoring watershed after a restoration project to reduce sediment influx into the downstream Laurel Lakes. Degraded watersheds purposefully selected for MS4 permitting would presumably have poor habitat, water quality, and biological communities, so the potential positive impacts of BMPs could be obscured when compared to MS4s in less degraded systems. Furthermore, stormwater BMPs may not be enough to noticeably improve a severely degraded watershed. A restoration may first be needed to lower eroded banks or otherwise stabilize the channel.

Next, in terms of the filtering process, controls were selected based on catchment-scale impervious cover and having very low reported BMP implementation, but areas with at least moderate impervious cover and no BMPs would most likely have been developed before the late 1990s, with no or limited development since then. It is possible that the development characteristics in these catchments was different than the MS4 treatment catchments in some other way that was not accounted for and that would change the interpretation of results, such as maturity of vegetation in older neighborhoods leading to increased rainfall capture, small-scale practices implemented by homeowners (e.g., raingardens, stream buffer plantings), and/or the likelihood of larger contiguous forests, like neighborhood parks or abandoned fields that returned to forest.

The last set of filters, namely for reference catchments, were based on few characteristics, i.e., largely forested with no septic/sewer service areas and very little impervious cover. Catchments with these characteristics can still have small pockets of development, roads, or agriculture that can negatively impact stream health, so the reference pool may have catchments that are not of true reference quality. In spite of this, the selected reference catchments overall had much better health than control and MS4 catchments, with only some poor metric scores observed.

For the paired watershed analysis, the timeframe during which data were collected was considered to ensure that comparisons were not biased due to non-overlapping time series of the treatment and reference/control. For example, data in Parole Plaza/Church Creek were collected later in time than the control and reference pools, potentially skewing the results if only considering boxplots. To compensate for that, the scatterplots with trend lines were generated, but data were not otherwise adjusted to account for variation in space. When comparing data using very large datasets, such as the entire BIBI database, the effect of the monitoring location within each catchment is effectively balanced out due to the large sample size, but the majority of this study was focused on individual treatment areas and a smaller number of corresponding controls. It is also possible that biases were introduced due to the specific location of monitoring stations, such as stations being upstream versus downstream of BMPs; local conditions such as shade and road crossings; and land cover characteristics.

Many of the regressions showed unexpected or insignificant results for habitat parameters. The methods used in these regressions selected only two habitat metrics, largely due to limited data availability across habitat metrics. Selecting single metrics is not an ideal reflection of the effect of habitat, for which several metrics can combine to collectively reflect habitat quality. In addition, even after screening for the most frequently occurring habitat metrics, much more data were available for conductivity and BIBI scores. This imbalance in data availability could have influenced some findings.

The CART analyses developed for this project are somewhat limited in that they potentially overfit the observed dataset. Thus, the "break points" observed, particularly in the lower branches of the trees, do not necessarily reflect the population of possible instream effects, but rather the observed patterns in the available datasets. In particular, the models that used combined indicators (habitat and water quality) were somewhat limited by imbalance in the number of points and lack of paired data. The default methods for the CART analysis for missing values may be one factor that resulted in eliminating any habitat parameters from the CART model presented in Appendix A Figure A70.

Stormwater BMPs are primarily designed to reduce peak discharges, reduce runoff volume, and filter or settle pollutants. Unfortunately, most of the widely available measures of stream health are not highly influenced by these measures. For example, conductivity, which proved to be the best consistent measure of water quality, is likely not impacted by BMPs, which are unable to remove salts from stormwater, and are generally not highly effective at removing soluble pollutants. Consequently, it might not be surprising that BMPs were not generally associated with improved instream conditions.

Additional Analyses

The following supplemental analyses would strengthen results of this study and further investigate some of the questions that arose:

- In the long term, repeating the individual MS4 BMP analyses and expanding to drainage areas that were not necessarily monitored as a part of the MS4 monitoring strategy would allow for a more robust analysis. This analysis would require a deeper investigation into the BMP database, for catchments that were not included in this study.

- The relatively high BIBI scores in some control watersheds suggest that there may be inherent differences between control and MS4s beyond level of BMP implementation and land cover. A future analysis from a broader pool of control catchments could investigate other watershed factors that could lead to these differences, such as age of development and the type of impervious cover (e.g., residential, industry, airports). Age of development in particular may have been a confounding effect in this study, since developed watersheds without BMPs, are necessarily older.

- The BMP analysis incorporated change over time, by assigning impervious cover to specific time "windows" but this analysis did not account for the lag of these effects. Data were insufficient for this type of analysis in this study, but an expanded view of other available data could allow for a more rigorous look into the effects of recovery from watershed development and restoration.

Conclusions

Increasing impervious cover correlates with poorer stream biological health (Hypothesis 1), and the amount of impervious cover in a stream catchment effectively limits how much biological "lift" can be achieved with implementation of BMPs and other restoration activities (Figure 14).

Consistently better stream biological health, water quality, and habitat in MS4 versus control catchments were not always found. BMPs are primarily designed to remove sediments and control flows during storm events, but the water quality and habitat metrics most widely available may not have reflected BMP performance. For example, a reduction in turbidity would be an excellent way to gauge sediment removal, but data coverage for this parameter was inconsistent and often poor. While some habitat parameters related to sedimentation and stream stability could reflect BMP effectiveness (e.g., embeddedness, bank stability), these habitat impacts tend to lag behind BMP implementation, so it is difficult the judge BMP effectiveness in the short term. Metrics that would be best at assessing flow reduction by BMPs were not widely available at the MS4 sites. A general recommendation is to monitor the same parameters consistently over time across all MS4s so that a larger dataset will be available for future analysis projects. Additionally, water quality and biology metrics may be more reliable than habitat scores because their measurement is not as subjective as habitat scores. MS4 permittees may want to consider providing their field crews with regular training with the Maryland Biological Stream Survey (MBSS) in scoring habitat. This will help keep habitat scores more consistent.

Due to the mixed results and limited data for some metrics, Hypothesis 2 is not fully supported and is dependent on several factors, like watershed land cover, the number and types of BMPs, and, ultimately, expectations of what is defined as improved stream quality. The MS4 watersheds that performed the best compared to controls tended to be located in watersheds with little active development, have ≥50% of the BMP drainage areas treated by water quality BMPs, and retain pockets of contiguous forest to buffer the stream.

Following large amounts of stream restoration, there is often an apparent decline in stream health. This was noted at Bear Branch, Breewood, and Stewart-April Lane, which had stream restorations and experienced declines in bank stability, bank vegetation, and/or epifaunal substrate in the following years. In MS4 watersheds with ongoing stream restoration or BMP implementation, it would be beneficial to install monitoring stations located throughout the watershed and to monitor for several years after to allow the channel to equilibrate after the disturbance before assessing whether the practice was a success or not.

Stream biological health is difficult to predict from underlying components of the stream functions pyramid except in extreme instances such as reference or severely degraded conditions (Hypothesis 3). However, the CART analysis method can be used to determine thresholds of when individual stressors start to degrade stream biological health.

Finally, measures of hydrology and hydraulics were not identified and used in the study. Flow especially would have been useful to incorporate into analyses for evaluating how BMPs regulate stream flow. Habitat metrics were used as a surrogate, but this is not ideal and a complete understanding of stream flow in BMP-regulated streams would help to assess success of BMPs.

References

Griffith, G. (2010). CEC Level III description – United States. Prepared for the North American Commission for Environmental Cooperation (www.cec.org).

Kaushal, S.S., Likens, G.E., Pace, M.L., Utz, R. M., Haq, S., Gorman, J., and Grese, M. 2018. Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences. 115*(4). DOI: 10.1073/pnas.1711234115

Koenker, R. (2021). quantreg: Quantile Regression. R package version 5.86. https://CRAN.R-project.org/package=quantreg

Ogle, D.H., Doll, J.C., Wheeler, P., Dinno, A. (2022). FSA: Fisheries Stock Analysis. R package version 0.9.3, https://github.com/fishR-Core-Team/FSA.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Smith, Z. M., C. Buchanan, and A. Nagel. 2017. Refinement of the Basin-Wide Benthic Index of Biotic Integrity for Non-Tidal Streams and Wadeable Rivers in the Chesapeake Bay Watershed. ICPRB Report 17-2.

U.S. Army Corps of Engineers, The Nature Conservancy, and Interstate Commission on the Potomac River Basin. 2014. Middle Potomac River Watershed Assessment: Potomac River Sustainable Flow and Water Resources Analysis. Final Report. 107 pp. and 11 appendices.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Woods, A. J., J. M. Omernik, and D. D. Brown. 1999. Level III and IV ecoregions of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. U. S. Environmental Protection Agency. Available online at: http://www.fwspubs.org/doi/suppl/10.3996/022017-JFWM-017/suppl_file/10.3996022017-jfwm-017.s11.pdf