

Stream Biological Health in the Chesapeake Bay Watershed

February 10, 2023
Correction noted August 1, 2024

Report Prepared for the
Chesapeake Bay Program
Stream Health Workgroup



CORRECTION:

A computational error was subsequently discovered in the spatial analysis of the pre-baseline period (2000 - 2005) conducted with the Average Nearest Neighbor tool in ArcGIS. It caused a slight underestimate of the area-weighted and stream length-weighted %_nEGF in that period (Tables 6 and 7). The corrected value for stream length-weighted %_nEGF in Table 7 is 58.4%, not 57.1%.

ICPRB Report #ICP23-1

This report can be downloaded from the Publications tab of the ICPRB's website, www.potomacriver.org. To receive hard copies of the report, please write:

Interstate Commission on the Potomac River Basin
30 West Gude Dr., Suite 450
Rockville, MD 20850
Or call 301-984-1908
Or email info@icprb.org

Disclaimer

The opinions expressed in this report are those of the authors and should not be construed as representing the opinions or policies of the U. S. Government, the U. S. Environmental Protection Agency, the Potomac basin states of Maryland, Pennsylvania, Virginia, and West Virginia, and the District of Columbia, or the Commissioners to the Interstate Commission on the River Basin.

Acknowledgements

We thank representatives from the watershed's monitoring programs and Chesapeake Bay Program (CBP) Stream Health Workgroup for their constructive suggestions on the draft version of this report, including J. Champion (DOEE), B. Flaten (DNREC), J. Hill (VADEQ), K. Maloney (USGS), E. Perry (statistics consultant), G. Pond (USEPA), A. Santoro (MDDNR), D. Shull (PADEP), Z. Smith (NYDEC), P. Tango (USGS), R. Thompson (USGS), and S. Weglein (MDDNR). Support was provided by U. S. Environmental Protection Agency CB-96351301 (Chesapeake Bay Program) and the Interstate Commission on the Potomac River Basin.

Suggested citation for this report:

C. Buchanan, R. D. Jepsen, and M. E. Mallonee. 2023. Stream Biological Health in the Chesapeake Bay Watershed. ICPRB Report ICP23-1. Report Prepared for the Chesapeake Bay Program Stream Health Workgroup. Available online at: www.potomacriver.org

Front Cover

Close up of Chessie BIBI results by HUC12-bioregion units in the heavily sampled corridor between Harrisburg, PA and Richmond VA, 2000 – 2017 (Figure 4).

Stream Biological Health in the Chesapeake Bay Watershed

Report Prepared for the Chesapeake Bay Program Stream Health Workgroup

Claire Buchanan, Rikke D. Jepsen, and Michael E. Mallonee
Interstate Commission on the Potomac River Basin

February 10, 2023

Executive Summary

This report offers a numeric value for the 2008 Baseline referenced in the 2014 Chesapeake Agreement's stream health goal as well as evidence of a net improving trend in stream health in the Chesapeake watershed. The report demonstrates a process for tracking progress in achieving the stream health goal to *"improve health and function of ten percent of stream miles above the 2008 baseline."* The bioregion, family-level version of the Chesapeake Basin-wide Index of Biotic Integrity, or "Chessie BIBI," is used to quantify stream health. The index is calculated from macroinvertebrate data collected by state, federal, county, and volunteer monitoring programs with kick net methods and was developed specifically for 1st – 4th order streams in the Chesapeake watershed (Smith et al. 2017). The 2008 Baseline is the 2006 – 2011 period because it encompasses all sampling schedules of the watershed's state monitoring programs, most of which employ rotational sampling.

Gaps in the monitoring data's spatial and temporal coverage make it difficult to directly estimate percentages of healthy streams in the pre-baseline (2000 – 2005), baseline, and subsequent "first interval" (2012 – 2017) periods. Statistical analyses indicate approximately 61.7% (~89,317 miles) of non-tidal stream miles likely supported healthy macroinvertebrate communities in the baseline period. The percentage increased to 67.8% (~98,049 miles) in the first interval. Despite this roughly 6% net improvement, some areas of the watershed show degrading trends. The net improving trend, however, suggests the collective impact of multiple environmental stressors on streams may be slowly lessening in many parts of the Chesapeake watershed. Identifying which factors are responsible for the net improvement would be speculative at this point, although long-term efforts to conserve forests, preserve and restore riparian corridors and wetlands, mitigate acid rain and mine drainage, slow stormwater runoff, and reduce nutrients and sediment loads have all likely contributed. Metrics for a variety of environmental stressors are currently being explored and will help future investigations of stream macroinvertebrate responses to those stressors. They can help explain why the current trend is happening.

The purpose of this report is to present the monitoring-based results and provide CBP with a process for tracking progress in achieving the Chesapeake watershed's stream health goal. The process differs in some respects from those of the state agencies who use the data differently and for state regulatory purposes. We fully expect the Chessie BIBI results will also differ from state results at times, even though the underlying raw data are the same. The Chessie BIBI can be used for inter-jurisdictional, watershed-based planning and evaluation.

Introduction

The Chesapeake Bay Program (CBP) partnership established a goal in 2014 to “Continually improve stream health and function throughout the watershed and specifically, to improve health and function of ten percent of stream miles above the 2008 baseline for the Chesapeake Bay watershed” (CBP 2014). The reference-based Chesapeake Basin-wide Index of Biotic Integrity, or “Chessie BIBI” (Buchanan et al. 2011), was selected a year later as the initial CBP indicator of stream biological health (CBP 2015). At the time, the BIBI was the only stream health indicator available that could be applied watershed-wide. The index has subsequently been updated and refined (Smith et al. 2017a, 2017b), and resource managers and scientists selected the bioregion, family-level version of the index for CBP reporting purposes at a 2018 workshop (Buchanan et al. 2018).

Overall, ratings of the bioregion, family-level version of the BIBI index correctly distinguish high quality, or reference, from degraded instream conditions in 81.6% of samples (classification efficiency, Smith et al. 2017). The index is derived from macroinvertebrate samples routinely collected with kick nets by federal, state, county, and volunteer monitoring programs in non-tidal streams in the Chesapeake watershed. The index score is the average of five or more scored, family-level macroinvertebrate metrics identified as sensitive to instream degradation in each of the watershed’s twelve bioregions (Figure 1). Bioregions are geographic areas where the taxonomic composition of macroinvertebrate populations in undisturbed streams are like one another and dissimilar to those in neighboring geographic areas (Smith et al. 2017). The numeric BIBI scores reflect the natural geographic differences in stream macroinvertebrate populations, so the scores cannot be directly compared across bioregions. To make results comparable, scores are assigned a categorical rating of excellent, good, fair, poor, or very poor according to percentile scores at reference sites in their bioregion.

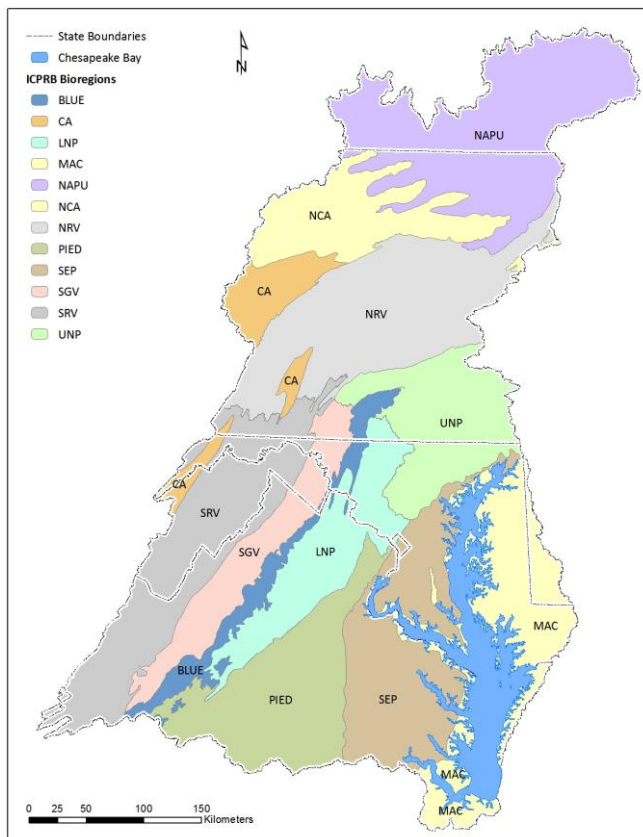


Figure 1. Bioregions for stream macroinvertebrates in the Chesapeake Bay watershed.

Monitoring coverage is sparse in some parts of the Chesapeake watershed. For unsampled streams, participants of the 2018 workshop recommended exploring the use of landscape-based models to fill the gaps. Maloney et al. (2018) developed a random forests model to predict Chessie BIBI ratings for small unsampled streams (<200 km² in drainage) using landscape features: spatial location, bioregion, land cover, soil, precipitation, number of dams. The 5-category Chessie BIBI ratings were reclassified into two categories – Poor and Fair/Good – to improve model prediction capabilities. Poor combines the poor and very poor (i.e., degraded) BIBI ratings; Fair/Good combines the fair, good, and excellent (i.e., healthy) BIBI ratings. Model accuracy varied by bioregion, but overall it correctly predicted 72% of samples in the training dataset and 73% in the test dataset. The model was recently updated to incorporate new data and better reflect landscape conditions affecting stream macroinvertebrates (Maloney et al. 2022), but more work is still needed before model results can be used to fill monitoring gaps.

The 6-year period from 2006 to 2011 was selected at the 2018 workshop as the 2008 Baseline for the CBP

goal, to be compared to successive 6-year periods. The baseline period encompasses all sampling designs of the Chesapeake watershed's various state monitoring programs, most of which employ rotational sampling based on a random stratified sampling design. Monitoring results are now available for the pre-baseline (2000 – 2005) and first post-baseline (2012-2017) intervals as well as the baseline. This report examines the proportions of healthy streams in the Chesapeake watershed for the three intervals and establishes a numeric 2008 Baseline in stream miles (1:24,000 NHD high resolution).

Methods

Data Preparation

The analysis dataset comprised 26,752 samples collected in the eighteen years between 2000 and 2017 by federal, state, and local agencies and other groups with similar kick net methods in streams and small rivers in the Chesapeake watershed. The raw data have been assembled and incorporated into the Chesapeake Bay Program's (CBP) water quality/living resources Chesapeake Environmental Data Repository (CEDR) database. The data are available for download via the CBP DataHub (<https://data.chesapeakebay.net/LivingResources>).

Prior to calculating the Chessie BIBI index, the raw data were normalized to minimize programmatic differences in field and laboratory protocols (Smith et al. 2017). Normalization involves a) excluding samples that were not collected with a kick-net or similar method, were collected in winter (December – February), or were collected in streams greater than 4th Strahler order; b) standardizing taxonomy by rolling up to the lowest common taxonomic resolution (e.g., genus rolled up to family-level); c) removing taxa that are counted by just a few programs (e.g., mites), are semi-aquatic taxa (e.g., skimmers), or are rarely found taxa (e.g., ostracods); and d) excluding samples with fewer than 70 individuals. Finally, samples with counts greater than 100 are rarified to a standard count of 100. All steps to normalize the data were performed using custom R-scripts within a larger R-program that calculates the Chessie BIBI. The scripts are available from the Chesapeake Bay Program, where there is also a [package](#) of documents that includes the raw data files and program outputs.

The bioregion, family-level Chessie BIBI assessment method described in Smith et al. (2017) and encoded in the custom R-program was applied to the normalized macroinvertebrate data. Established systems of geographic classification (hydrologic subregions, physiographic provinces, level III ecoregions) were used by Smith et al. (2017) to identify the twelve bioregions in the Chesapeake Bay watershed (Supplemental Table 1). Family-level macroinvertebrate metrics are used because in the early years many samples were enumerated at that taxonomic level. The five or more family-level metrics identified during index development as most sensitive to stream habitat and water quality degradation in each of the twelve bioregions are calculated by the R-program from the standardized, rarified data and scored on a scale of 0 to 100 in an identical manner. The metric scores are averaged to obtain a numeric Chessie BIBI index score for the sampling event.

As explained in Smith et al. (2017), the rarefaction step often results in a slightly different taxa assemblage each time the R-program is executed. This is due to the program's random selection of different rare taxa when rarefaction is repeated. Depending on which rare taxa are included or excluded, a sampling event's numeric BIBI score may change with repetition. To help remove some of this uncertainty, the R-program is run 100 times, and the median index score and its bioregion-specific categorical rating are used to represent the sampling event.

Index scores are assigned categorical ratings of excellent, good, fair, poor, or very poor based on their percentile rank when compared to the distribution of scores for reference sites in their bioregion. Reference sites in all bioregions are identified based on their specific conductivity, pH, and dissolved oxygen concentrations, as well as eight instream habitat metrics. The surrounding and upstream landscapes are not considered. For all sampled sites in a bioregion, an index score above the 50th percentile of scores in that bioregion's reference sites is rated excellent (E); an index score between the 25th and 50th percentile is good (G); between the 10th and 25th percentile, fair (F); below the 10th percentile, poor (P); and below half the value of the 10th percentile, very

poor (VP) (Table 5 in Smith et al. 2017).¹ The ratings are conservative and assume some reference site macroinvertebrate populations (~10%) will be impacted by unreported stressors. For purposes of CBP reporting, scores that rate E, G, and F indicate “healthy” stream macroinvertebrate populations.

Data Analysis

Station names for some locations have changed over time, so for this analysis a concatenation of each station’s latitude and longitude was used instead of name to identify unique locations. GPS accuracy also improved over time, so all recorded latitude and longitude coordinates were rounded to the fourth decimal point before concatenation to facilitate continuity.

Each unique sampling location was associated with a HUC12-bioregion unit in ArcMap (v 10.5). The 12-digit hydrologic units, or HUC12s, of the U. S. Geological Survey’s Watershed Boundary Dataset (WBD) (USGS 2016) superimposes a grid-like framework of 1,968 cells averaging 86.6 km²/cell on the Chesapeake Bay watershed. There are 534 HUC12s overlapping two or more bioregions, and these were further divided, resulting in 2,528 HUC12-bioregion units. HUC12-bioregion borders were used to map the BIBI results and their areas used to area-weight BIBI results in the first stages of the analysis.

We used fundamentally different approaches to geographically illustrate, or map, stream conditions and calculate the percentage of the Chesapeake watershed supportive of healthy streams. To create maps, the Chessie BIBI numeric scores for each sampling location were averaged by period, then the location averages are averaged by HUC12-bioregion unit and the unit’s average score rated according to the appropriate bioregion rating thresholds. The unit’s average score and its associated rating can mask inherent variability found in many units, especially in heterogeneous landscapes. To capture the diversity of stream conditions within a period and across each spatial unit, the ratings of the individual sampling locations in each unit are weighted by equal portions of the unit’s area. A singleton sample in one unit thus receives the same weight as multiple samples in an identically sized unit. The area-weighted ratings for E, G, F, P, and VP from sampled spatial units as well as the area of unsampled (unrated) spatial units are individually summed, and the sums divided by total Chesapeake watershed area to obtain estimated percentages of watershed area supporting each rating (%E, %G, %F, %P, %VP, %unrated). Chessie BIBI results can be analyzed by HUC unit alone once the average score for each location is assigned its appropriate bioregion-specific rating. The same approach of apportioning equal weight to each location in a HUC unit was employed.

The probability of sampling E, G, or F sites in the pool of *sampled* spatial units is calculated from the area-weighted percentages and that probability used to calculate the percent of the entire Chesapeake watershed area supporting healthy stream populations (%_nEGF). The probability is expressed as odds, or “1 in # number of sites.” In using the odds derived from sampled units to calculate %_nEGF of all units, we are assuming stream conditions in unrated units are generally comparable to those in sampled units.

Total area of the Chesapeake Bay watershed was computed using ArcMap (v 10.5). For the HUC12-bioregion units, the bioregion shapefile developed by ICPRB (ICPRB_Bioregions.shp) was first merged with the CBP shapefile for all tidal waters in Chesapeake Bay and its tributaries (Chesapeake_Bay_92_Segments.shp), and the Erase tool used to remove all tidal areas. The remaining 166,500.68 km² is the two-dimensional area of the Chesapeake non-tidal watershed. The Intersect tool was then used to combine the bioregion and tidal-erased HUC12 shapefiles (HUC12_bioregion_splits.shp), and the Calculate Geometry function was used to calculate the area (km²) of each HUC12-bioregion unit. The ecoregion III layer (Omernik 1987, Woods et al. 1999) used to create the bioregions does not precisely overlay the high resolution (1:24,000) National Hydrologic Database

¹ For comparison, the percentiles of the reference populations used by the National Rivers and Streams Assessment (USEPA 2016) to distinguish the three rating categories of their Macroinvertebrate Multi-metric Index (MMI) are: values greater than the 25th percentile, good; between the 5th and 25th, fair; and less than the 5th percentile, poor.

(NHDPlus2) layer containing HUC12 boundaries and stream lengths. Consequently, about 0.54% of the unrated Chesapeake watershed is hard-to-discern slivers of area created by the different boundaries.

Results

Spatial and Temporal Gaps

Maps of the Chessie BIBI monitoring results show spatial gaps in the pre-baseline (2000 – 2005), baseline (2006 – 2011), and first interval (2012 – 2017) periods (Figure 2). Over 7,000 samples were collected in each of the three periods, however sampling locations are unevenly distributed between and within jurisdictions. Locations are farther apart in rural and agricultural areas; they are closer together in the Baltimore-Washington metropolitan region and other urban areas where multiple agencies monitor streams. HUC12-bioregion units with no samples (unrated) comprised 38.8% of the total watershed area in pre-baseline, 30.4% in the baseline and 42.7% in the first interval (Table 1). Sampling frequency is also not uniformly distributed across the three time periods. Of the 2,528 HUC12-bioregion units, only 734 (29.0%) were sampled in all three periods and the majority of those were from the Baltimore-Washington metropolitan region. The large percentages of unrated area partly reflect the choice to analyze BIBI results by the relatively small HUC12-bioregion units (median size 61.6 km², range 0.12 – 173.7 km²). When data for all 18 years are combined, only 13.9% of the HUC12-bioregion units are unrated.

Replicate and Annual Variability

A macroinvertebrate sample is typically a composite created by sampling representative points in a stream reach. Field replicates are collected in the same stream reach at undisturbed points, usually upstream of where the first sample was collected (e.g., USEPA 1997). Replicate samples were collected in 739 of the 25,976 (2.84%) sampling events, or unique location-dates, by various monitoring programs. Ratings of the replicate Chessie BIBI scores agreed in 455 (61.6%) cases and disagreed in the other 284 (39.4%). When a location's average rating is

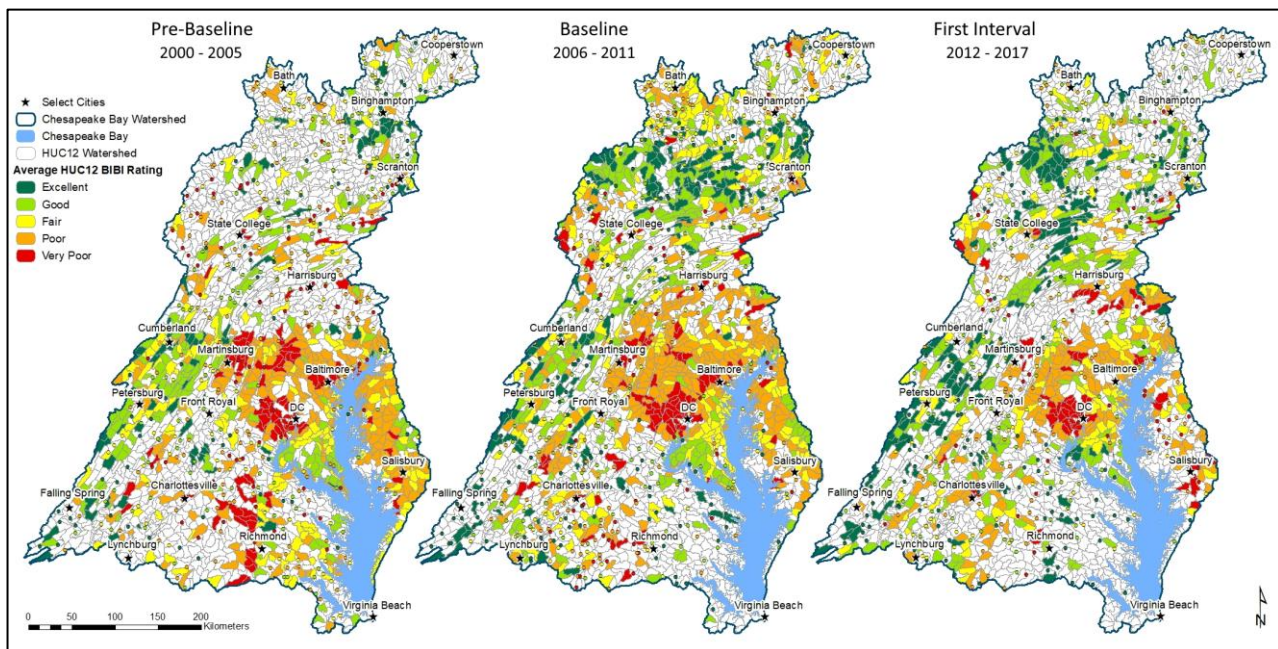


Figure 2. Spatial distribution of Chessie BIBI ratings in the pre-Baseline, baseline, and first interval periods by HUC12-bioregion unit. Replicates and multiple sampling events at each location are first averaged, then scores in units with multiple locations in each period are averaged. The bioregion-specific rating of the average score is then mapped. A solidly colored HUC12-bioregion unit contains two or more sampling locations and the color reflects the average result. A colored dot indicates only one sampling location in the unit. Unsampld units are left blank (white).

Table 1. Influence of averaging replicates and/or averaging multiple sampling event scores per location on subsequent calculations of each rating’s area-weighted percentages, by period. Analyses were done using HUC12-bioregion spatial units to group location results. Periods: pre-baseline, 2000 – 2005; baseline, 2006 – 2011; first interval, 2012 – 2017; combined, 2000 - 2017. Average: replicates averaged, then all scores at each location averaged. Unaveraged Reps: replicates not averaged before all scores at each location averaged. Last/Only Score: replicates averaged, then the last or only score at each location is selected. Ratings: excellent (E), good (G), fair (F), poor (P), and very poor (VP). Odds, the probability of each Chessie BIBI rating occurring in the sampled HUC12-bioregion units, by period. See text for details.

Period	%E	%G	%F	%P	%VP	%Unrated	%EGF	Odds
<u>Average</u>								
Pre-baseline	10.4%	12.7%	11.1%	16.0%	11.0%	38.8%	34.2%	1 in 1.791
Baseline	16.5%	14.9%	10.9%	17.0%	10.2%	30.4%	42.4%	1 in 1.642
First Interval	16.5%	12.5%	9.3%	11.6%	7.3%	42.7%	38.4%	1 in 1.493
Combined	20.1%	18.3%	14.6%	20.4%	12.6%	13.9%	53.0%	1 in 1.624
<u>Unaveraged Reps</u>								
Pre-baseline	10.4%	12.7%	10.9%	16.2%	11.0%	38.8%	34.0%	1 in 1.800
Baseline	16.5%	15.0%	10.9%	17.0%	10.2%	30.4%	42.4%	1 in 1.642
First Interval	16.5%	12.5%	9.4%	11.6%	7.3%	42.7%	38.4%	1 in 1.493
Combined	20.1%	18.3%	14.5%	20.5%	12.7%	13.9%	52.9%	1 in 1.627
<u>Last/Only Score</u>								
Pre-baseline	11.1%	12.1%	10.9%	16.0%	11.2%	38.8%	34.1%	1 in 1.797
Baseline	17.2%	14.3%	10.8%	16.8%	10.5%	30.4%	42.3%	1 in 1.646
First Interval	17.2%	12.0%	8.6%	11.9%	7.5%	42.7%	37.8%	1 in 1.513
Combined	21.3%	17.4%	14.1%	20.4%	13.0%	13.9%	52.7%	1 in 1.633

either excellent or very poor, the replicates’ ratings typically match. When the average rating is good, fair, or poor, however, there is a greater chance of the replicates’ ratings not matching each other or not matching the average score’s rating (Table 2). For an average rating of fair, for example, the replicate ratings only matched the average of the ratings 56.4% of the time.

The variability found in replicate sample results suggests that averaging replicates could mask some of the natural biological variability in a stream, especially when the stream’s overall condition is fair. Area-weighted percentages of the five ratings watershed-wide in each of the three periods were calculated without averaging the replicates, effectively treating them as a separate sampling event at the location (Table 1, Unaveraged Reps). The resulting percentages of the ratings were then compared to those with the replicates averaged. Small differences occur in the good, fair, and poor percentages, but not in the excellent or very poor percentages. Averaging the replicates had little effect on the summed percentages of excellent, good, and fair (%EGFs) in each period.

Comparisons of BIBI scores and ratings at 186 frequently sampled locations (i.e., a location sampled in at least nine years of the 18-year period) demonstrate a similar pattern of variation in annual ratings around the

location’s overall average rating. These locations are sampled for specific purposes (e.g., sentinel sites, before-and-after restoration or construction) and are not representative of general stream conditions in the Chesapeake watershed. However, they give an indication of variability associated with the five narrative rating categories. Average scores close to zero (very poor) and close to 100 (excellent) exhibited the least variability in annual scores

Table 2. Ratings of replicate samples compared to the rating of their average numeric score.

Rating of Average Score	Ratings of Replicate Samples			
	Better	Same	Worse	N
Excellent	N/A	89.0%	11.0%	301
Good	19.1%	66.8%	14.1%	319
Fair	20.6%	56.4%	23.0%	243
Poor	13.8%	78.0%	8.2%	414
Very poor	5.9%	94.1%	N/A	238
Overall	12.0%	76.9%	11.1%	1515

and those in the middle of the numeric scale exhibited the most variability (Figure 3). Narrative ratings also reflect this pattern (Table 3, Supplemental Table 2). Locations with average ratings of very poor do not experience excellent or good annual ratings; locations with average ratings of excellent do not experience very poor annual ratings. Locations with overall ratings of good, fair, and poor, however, experience a wide range of annual ratings over time.

In the sampled HUC12-bioregion units, 696 (9.9%), 915 (11.4%), and 1,005 (21.3%) locations were sampled two or more times in the pre-baseline, baseline, and first interval periods, respectively. Averaging the scores for these locations (after replicates are averaged) avoids biasing the proportional weight assigned to each rating in a HUC12-bioregion unit. Annual variability found at repeat sampling locations (Table 3), however, suggests averaging location scores could mask some of the inherent biological variability in streams. As an alternative to averaging, the most recent index score in each period was used to represent locations with multiple sampling events (Table 1, Last/Only Score). When compared to results derived with location-averaged scores, percentages of excellent and very poor ratings increased slightly, percentages of good and fair decreased slightly, change in percentages of poor were mixed. The %EGFs decreased by 0.1% to 0.6%.

Percentage of Healthy Streams

Analyses from here on use ratings derived from the replicate-averaged and location-averaged BIBI scores. In the watershed’s sampled HUC12-bioregion units, the overall odds of sampling an area supporting healthy stream populations is decreasing, indicating the probability of sampling EGF populations is improving. Odds were 1 in

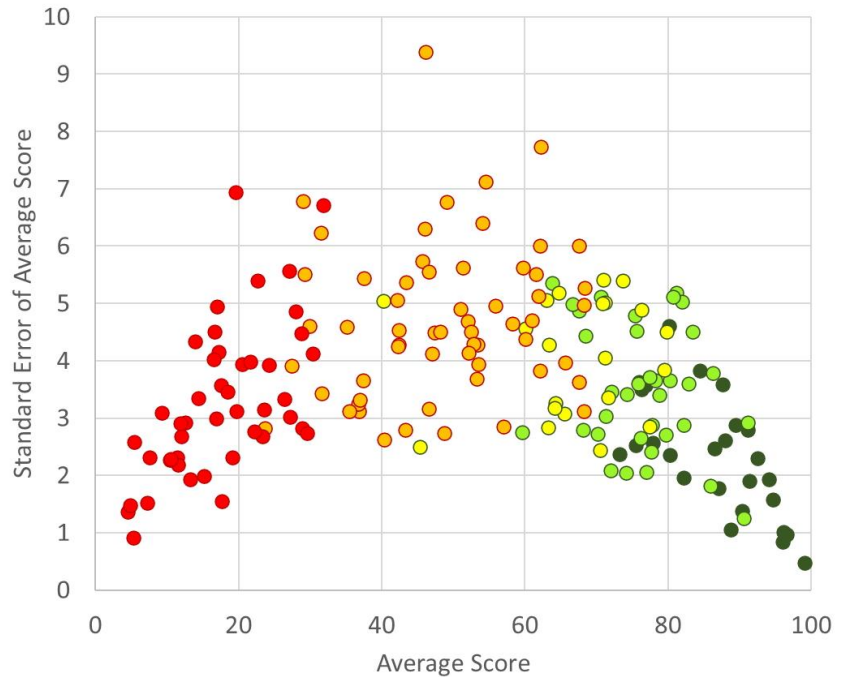


Figure 3. Variation (standard error) around the overall mean score at locations sampled nine or more times in the eighteen years between 2000 and 2017. Rating of each mean score is indicated with color.

Table 3. Percentages of each rating experienced annually at repeat monitoring locations as compared to the average score’s rating. Ratings: excellent (E), good (G), fair (F), poor (P), and very poor (VP). Repeat locations: number of locations sampled nine or more years in the 18-year study period. Location-years: number of location-year combinations. Annual rating: rating of each location’s average numeric score for individual years. Average rating: rating of the average of all annual numeric scores for a location. (See text for details.)

Average Rating	Annual Ratings					# Repeat Locations	# Location - Years
	%E	%G	%F	%P	%VP		
Excellent	79.9%	15.9%	3.5%	0.7%		26	283
Good	32.2%	39.2%	16.7%	11.6%	0.2%	35	413
Fair	9.9%	30.0%	28.5%	28.9%	2.8%	21	253
Poor	2.3%	8.8%	13.8%	53.8%	21.2%	59	725
Very poor			0.6%	16.1%	83.3%	45	479

Table 4. Odds of each area-weighted Chessie BIBI ratings in the pool of sampled HUC12-bioregion units with and without targeted locations. *, locations were removed from pre-baseline (n = 972) and first interval (n = 869) datasets if they were also sampled in baseline, on the assumption they are targeted sites and even if they were originally selected randomly. EGF, excellent, good, and fair ratings combined; % change, the change between pre-baseline and first interval and calculated directly from the odds. See text for details.

Period	Excellent	Good	Fair	Poor	Very Poor	EGF
All Sampled Units						
Pre-baseline	1 in 5.903	1 in 4.804	1 in 5.535	1 in 3.817	1 in 5.563	1 in 1.791
Baseline	1 in 4.211	1 in 4.664	1 in 6.360	1 in 4.100	1 in 6.802	1 in 1.642
First Interval	1 in 3.462	1 in 4.583	1 in 6.136	1 in 4.948	1 in 7.819	1 in 1.493
% Change	+11.9%	+1.0%	-1.8%	-6.0%	-5.2%	+11.2%
Targeted Locations Removed						
Pre-baseline*	1 in 5.854	1 in 5.137	1 in 5.714	1 in 3.763	1 in 5.160	1 in 1.850
Baseline	1 in 4.211	1 in 4.664	1 in 6.360	1 in 4.100	1 in 6.802	1 in 1.642
First Interval*	1 in 3.395	1 in 4.673	1 in 6.648	1 in 4.952	1 in 7.189	1 in 1.518
% Change	+12.4%	+1.9%	-2.5%	-6.4%	-5.5%	+11.8%

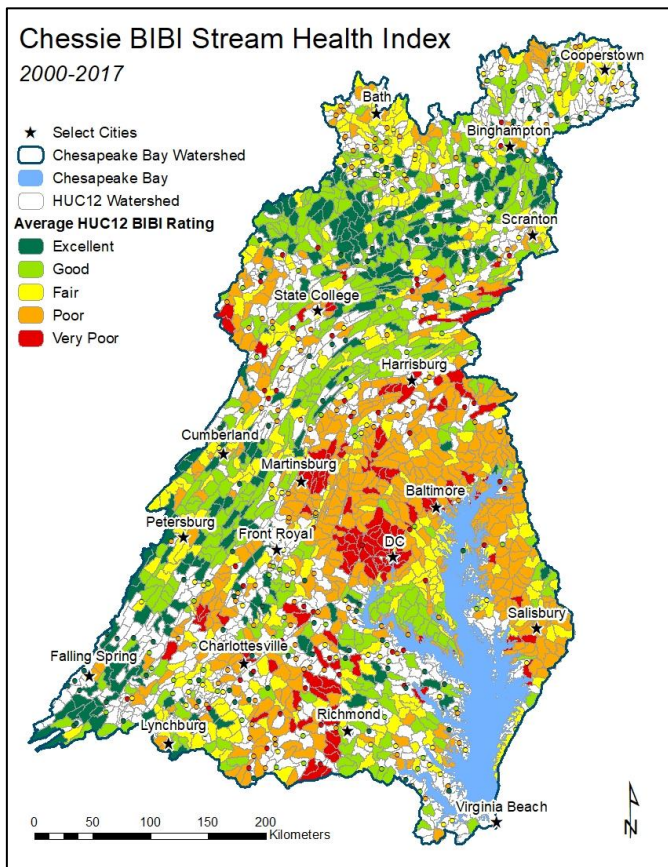


Figure 4. Spatial distribution of Chessie BIBI ratings for 2000 – 2017, by HUC12-bioregion unit. A solidly colored unit contains two or more sampling locations and the color reflects the average score. A colored dot indicates only one sampling location in the unit. Unsampled units are left blank (white).

1.791 in pre-baseline, 1 in 1.642 in baseline, and 1 in 1.493 in first interval (Table 1, Average). When the odds of each rating are considered separately, the odds of sampling excellent streams have improved 11.9% between pre-baseline and first interval, odds of sampling good and fair have not changed much, and odds of poor and very poor have changed to -6.0% and -5.2%, respectively (Table 4, All Sampled Units).

When data for the three 6-year periods are combined, the sampled HUC12-bioregion units cover most of the Chesapeake watershed (Figure 4). Scores at repeatedly sampled locations are averaged for the entire 2000 – 2017 period, and the location’s average score is rated with the appropriate bioregion rating thresholds. The percentage of unrated watershed area in this 18-year period is only 13.9% and the overall odds of sampling healthy streams is 1.624 (Table 1, Average). These odds are identical to those in the 6-year baseline period, which had 30.4% of the watershed area unrated. The concurrence of the odds for the 6-year baseline and the 18-year period strongly suggests about 61% - 62% of the Chesapeake watershed area supported healthy streams in the baseline period.

Many locations sampled in the baseline were also sampled in the pre-baseline (972 of 7,035, 13.8%) and/or the first interval (869 of 4,713,

18.4%). To test if these repeatedly sampled, or targeted, locations affect the overall odds of EGF, we identified and removed them from the pre-baseline and first interval groups, and recalculated the odds (Table 4, Targeted Locations Removed). Individual ratings and EGF odds changed slightly as did the overall %change in the ratings, and the slightly improving trend in EGF got 0.6% stronger.

New locations in later periods may have been selected non-randomly for different purposes (e.g., sentinel, reference), and could be biasing the overall %EGF. Comparisons of the targeted locations, or locations sampled in multiple periods, were examined to see if these subsets of locations also show slight improving trends. Comparisons were made for locations where 1) all three periods were sampled (n = 734), 2) only pre-baseline and baseline were sampled (n = 277), 3) only baseline and first interval were sampled (n = 200), and 4) only pre-baseline and first interval were sampled (n = 83). The odds in these four cases all show net improving trends over the three periods (combined, Table 5). If the results are further grouped by geographic region (Coastal, Piedmont, Ridge & Valley), the regional odds show net improvement or no change occurring in most cases. A net degrading trend was only found at the locations included in the Coastal region's 3-period comparison (n = 136) and those included in the Piedmont's pre-baseline to baseline comparison (n = 44). Locations used in this analysis are not always randomly distributed, and the results should be thought of as representing only the locations' HUC12-bioregion units. Results should not be extrapolated to entire regions. However, they support the finding that a generally improving trend may be occurring in stream health.

Table 6 shows the area-weighted results when locations are grouped and analyzed by HUC12 units. Results analyzed this way are nearly identical to those analyzed by HUC12-bioregion unit but the %unrated is lower. The

Table 5. Period comparisons of the odds of sampling a healthy (E, G, F) stream community at locations with sampling events in different periods. Odds are calculated from the sums of the area-weighted Chessie BIBI ratings in HUC12-bioregion units meeting the criteria of being sampled 1) in all three periods, 2) in only the pre-baseline and baseline, 3) in only the baseline and first interval, and 4) in only the pre-baseline and first interval. Area-weighted ratings are also grouped by three regional scales (Coastal, Piedmont, Ridge & Valley) and combined. Results represent just the locations' HUC12-bioregion units.

Period Region (# locations)	Pre- Baseline 2000-2005	Baseline 2006-2011	First Interval 2012-2017	Trend	%Change
All Three Periods					
Coastal (136)	1 : 1.81	1 : 1.66	1 : 1.89	Impr./Degr.	-2.36%
Piedmont (201)	1 : 2.75	1 : 2.76	1 : 2.57	Improving	2.57%
<u>Ridge&Valley (378)</u>	<u>1 : 1.52</u>	<u>1 : 1.41</u>	<u>1 : 1.28</u>	<u>Improving</u>	<u>11.98%</u>
Combined (734)	1 : 1.79	1 : 1.68	1 : 1.59	Improving	6.99%
Pre-Baseline and Baseline					
Coastal (86)	1 : 1.93	1 : 1.61		Improving	10.29%
Piedmont (44)	1 : 2.80	1 : 3.21		Degrading	-4.61%
<u>Ridge&Valley (147)</u>	<u>1 : 1.46</u>	<u>1 : 1.48</u>		<u>No change</u>	<u>-0.92%</u>
Combined (277)	1 : 1.73	1 : 1.68		Improving	1.46%
Baseline and First Interval					
Coastal (7)		1 : 1.06	1 : 1.09	No change	-2.65%
Piedmont (57)		1 : 1.78	1 : 1.76	No change	0.67%
<u>Ridge&Valley (136)</u>		<u>1 : 1.34</u>	<u>1 : 1.17</u>	<u>Improving</u>	<u>10.90%</u>
Combined (200)		1 : 1.43	1 : 1.30	Improving	7.08%
Pre-Baseline and First Interval					
Coastal (19)	1 : 1.74		1 : 1.33	Improving	17.66%
Piedmont (22)	1 : 2.91		1 : 1.73	Improving	23.40%
<u>Ridge&Valley (42)</u>	<u>1 : 1.48</u>		<u>1 : 1.35</u>	<u>Improving</u>	<u>6.65%</u>
Combined (83)	1 : 1.76		1 : 1.42	improving	13.64%

Table 6. Percentages of the area-weighted ratings for sampled HUC12 spatial units of the Chesapeake watershed and the odds of excellent, good, or fair ratings occurring in those spatial units, by period. %_nEGF, %EGF normalized to the entire watershed area and calculated directly from the odds (e.g., 1/1.778 = 56.2%). See Table 3 heading for details. (See correction notice on pg ii.)

Period	%E	%G	%F	%P	%VP	%Unrated	%EGF	Odds	% _n EGF
Pre-baseline	11.0%	13.4%	11.8%	16.6%	11.5%	35.8%	36.1%	1 in 1.778	56.2%
Baseline	17.5%	15.5%	11.5%	17.9%	10.6%	26.9%	44.6%	1 in 1.640	61.0%
First Interval	17.4%	13.2%	9.9%	12.2%	7.6%	39.7%	40.4%	1 in 1.491	67.1%
Combined	20.8%	18.7%	15.2%	21.3%	13.0%	11.0%	54.8%	1 in 1.625	61.6%

Table 7. Percentages of the stream length-weighted ratings for sampled HUC12 spatial units of the Chesapeake watershed and the odds of excellent, good, or fair ratings occurring in those spatial units, by period. A total 232,866 stream kilometers in the watershed is assumed. %_nEGF, %EGF normalized to the watershed's total stream miles and calculated directly from the odds. See Table 3 heading for details. (See correction notice on pg. ii.)

Period	%E	%G	%F	%P	%VP	%Unrated	%EGF	Odds	% _n EGF
Pre-baseline	11.1%	13.5%	12.0%	16.4%	11.1%	35.9%	36.6%	1 in 1.752	57.1%
Baseline	17.6%	15.2%	11.2%	17.2%	10.0%	28.7%	44.0%	1 in 1.620	61.7%
First Interval	17.1%	12.7%	9.8%	11.9%	6.9%	41.5%	39.7%	1 in 1.476	67.8%
Combined	20.5%	18.5%	15.2%	21.1%	12.6%	12.1%	54.3%	1 in 1.620	61.7%

odds for EGF are decreasing over time: 1 in 1.778 in pre-baseline, 1 in 1.640 in baseline, and 1 in 1.491 in first interval. The corresponding %_nEGFs indicate 56.2%, 61.0%, and 67.1% of the entire Chesapeake watershed area likely supported healthy populations in the three periods, respectively. Again, an improving trend is noticeable with +4.7% between pre-baseline and baseline, and +6.1% between baseline and first interval. For the full 18-year period, the odds of EGF are 1 in 1.625, corresponding to 61.6% of the watershed area supporting healthy streams.

Using the stream kilometers provided for each HUC12 in the NHD high resolution GIS layer (1:24,000), every location's rating can be weighted by an equal portion of its HUC12's total stream length. These values slightly inflate the lengths for 1st – 4th order streams in each HUC12 because they include the lengths of rivers as well as streams. However, stream length-based results (Table 7) very closely resemble the area-based results (Table 6) with %EGF odds of 1 in 1.752, 1 in 1.620, and 1 in 1.476 in pre-baseline, baseline, and first interval, respectively. The corresponding %_nEGFs for the three periods were 58.4% in pre-baseline, 61.7% in baseline, and 67.8% in first interval. Summing the stream lengths of all HUC12s in the watershed gives a total of 232,866 kilometers or 144,696.26 miles. Multiplying the number of total stream miles by %_nEGF yields estimates of healthy stream miles of 82,576 miles in pre-baseline, 89,317 miles in baseline, and 98,049 miles in first interval.

Discussion

Averaging the replicates collected during a sampling event is typically done to avoid giving extra weight to that sampling event's results relative to other sampling events at the location. Similarly, averaging the results of multiple sampling events for one location avoids giving extra weight to that location's results compared to other locations in the spatial unit. Chessie BIBI ratings for replicates (Table 2) and for repeat sampling locations (Table 3, Figure 3) show little variability when the average rating is excellent or very poor. Larger amounts of variation are found when the average rating is good, fair, and poor. The greater variability associated with good, fair, and poor ratings had little effect on the overall, large scale estimates of the percent of Chesapeake watershed supporting healthy streams (%EGF, %_nEGF). However, it is useful to be aware of it and should temper expectations of seeing meaningful trends occurring at individual locations over short time periods.

Different rotational sampling schedules and overlapping monitoring programs result in uneven densities of sampling locations across the watershed (Supplement Figure 1). Grouping and area-weighting Chessie BIBI

ratings in the grid-like HUC units minimizes to a large extent the biases created by this unevenness. Ratings in HUCs with multiple locations are weighted by equal proportions of the HUC's area while the rating for a HUC with just one location is weighted by the HUC's entire area. When summed to the Chesapeake watershed scale, the diversity in the ratings is preserved yet intensely sampled areas do not overwhelm sparsely sampled areas.

Almost two thirds of HUC12 units in the analysis dataset had two or more sampled locations in each period. While our approach of proportionally area weighting the locational Chessie BIBI ratings succeeds in capturing the diversity of stream conditions in these units, the approach has consequences that should be acknowledged. The overall percentages of excellent and very poor ratings are 4% - 7% higher and those of good, fair, and poor are 1% - 5% lower compared to when each unit's *average rating* is area weighted (Figure 5, Supplemental Figure 3). The effect on %_nEGF is slight but consistent, with %_nEGFs calculated from the proportionally area-weighted ratings 0.2% to 0.5% lower than those calculated from the averaged ratings.

Grouping and analyzing Chessie BIBI ratings by relatively small spatial units like HUC12-bioregions or HUC12s leaves large percentages of unrated areas. Fortunately, our objective to characterize stream health watershed-wide is facilitated by the probabilistic sampling designs employed by many monitoring programs. Statistical inferences can be made about unsampled populations if sample distributions are shown to be randomly or uniformly dispersed.

We assumed a large percentage of the locations inside each HUC12 were originally selected randomly because most of the monitoring organizations utilize random sampling designs and many stratify by Strahler stream order. When the locational Chessie BIBI ratings are grouped by HUC12s and area-weighted, the results become in a sense pixelated because HUC12 units rather than their sampling locations are the entities being tested (compare Supplemental Figures 1 and 2). The full complement of HUC12 units in the Chesapeake watershed is uniformly dispersed (Average Nearest Neighbor ratio 1.160, z-score 13.591, p-value < 0.001). The distributions of *sampled* HUC12s in each period are also uniformly dispersed (Supplemental Figure 2), which enables statistical inferences about unsampled HUC12s in the watershed.

When grouped by HUC12, the odds of sampling a stream with an excellent, good, or fair rating in the pool of sampled HUC12s were 1 in 1.778 in pre-baseline, 1 in 1.640 in baseline, and 1 in 1.491 in first interval (Table 6). These odds were used to estimate for the entire watershed the percentages of watershed area supporting healthy streams in each period. The normalized %EGF, or %_nEGF, was 56.2% in pre-baseline, 61.0% in baseline, and 67.1% in first interval. The decreasing odds (i.e., increasing probability of sampling an E, G, or F site in the pool of sampled spatial units) and the corresponding increase in %_nEGF over the three periods suggest watershed stream conditions may be improving. Changes in the odds of the individual ratings also support the perception that watershed stream health may be improving (Table 4). The odds of excellent ratings have become more likely, and the odds of poor and very poor ratings have become less likely. Improvement is not happening everywhere, and some areas appear to be degrading (Table 5), but the general trend is in the desired direction.

The baseline period's %_nEGF estimate of 61.0% of the watershed area supporting healthy streams (Table 6) is one value for a 2008 Baseline. We are reasonably confident in this estimate because the estimate for the three periods combined, which has a much lower %unrated, is very similar (61.6%). The 2014 Chesapeake Bay Agreement goal calls for improvement in *"...stream miles above the 2008 baseline ..."* (CBP 2014). Substituting HUC12 stream length (1:24,000) in place of area in our calculations, the estimated percentage of

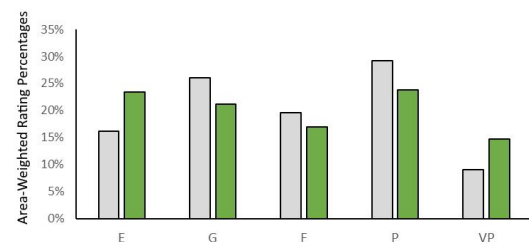


Figure 5. Percentages of ratings in the Chesapeake watershed when ratings of individual locations are area-weighted by equal portions of their HUC12-bioregion unit's area (green bars) and when the unit's average rating is area-weighted by the unit's area (gray bars), 2000 – 2017. Ratings: E, excellent; G, good; F, fair; P, poor; VP, very poor. See also Supplemental Figure 3.

healthy stream miles during the 2008 Baseline period was approximately 89,317 miles (143,743 km), or 61.7% of all stream miles (Table 7). The same calculation for the first interval indicates about 98,049 miles (157,795 km), or 67.8% of all stream miles, were healthy. Stream macroinvertebrate populations may have improved overall by roughly 6% between baseline and first interval.

The objective of this report was to group the assembled monitoring-based index ratings into the uniformly dispersed spatial units (HUC12s) while retaining each unit's observed diversity in stream biological conditions, and then statistically estimate the percent of healthy streams in the Chesapeake watershed. We are not attempting to characterize individual HUC12 units. Single locations are typically insufficient to fully characterize units the size of a HUC12, especially if the unit's landcover is heterogeneous and locations fall in a mix of urban, agricultural, and forested catchments. HUC12 020700110104, for example, has eleven sampled locations with ratings ranging from poor to excellent.

A landscape-based, Random Forests model has been developed to predict Chessie BIBI ratings at the very small spatial scale of EcoShed catchment (Maloney et al. 2018). These catchments average 1.55 km², and a tighter relationship between a stream segment and the surrounding landscape is expected. The model was able to predict ratings for 91.9% of the Chesapeake watershed area, leaving only 8.1% unrated. Currently, model predictions do not agree well enough with monitoring results in the same catchments to be merged with the monitoring results. This could be due in part to the model being derived from landscape features in the watershed whereas the BIBI ratings are derived from the apparent responses of stream macroinvertebrates to habitat and water quality features in their stream reach and adjacent stream corridor. More work is needed to refine the model, but it should eventually be very useful in filling the monitoring gaps and addressing management questions such as "Can restoration efforts improve stream health?" and "What efforts (e.g., protections, BMPs) are most effective for improving stream health?" For now, the monitoring results appear to be more than adequate in addressing the questions inherent in the stream health goal of the 2014 Chesapeake Bay Agreement: "Is stream health and function improving throughout the watershed?" and "What is the 2008 Baseline?"

Conclusions

We estimated 61.7% of the non-tidal stream miles in the watershed, or ~89,317 miles, supported healthy macroinvertebrate populations during the 2008 Baseline period of 2006 – 2011. Gaps in monitoring coverage occur most often in rural or forested areas which are usually less impacted by anthropogenic factors and tend to have more better quality streams. So, the actual 2008 Baseline could be slightly higher. Modeling tools to determine a more accurate baseline are not yet available.

The analyses point to a likely net improvement in stream biological health watershed-wide. The percent of watershed area supporting healthy streams has increased roughly 6% since the baseline period, extending an earlier positive trend from the pre-baseline period. There is no guarantee this improving trend will persist or even be detectable in the upcoming 6-year interval (2018 – 2023), especially given pandemic-related interruptions in state monitoring schedules during 2020 and 2021. However, the trend appears to be moving in the desired direction.

The apparent improvement over the three 6-year periods suggests the collective impact of environmental stressors on streams may be slowly lessening, at least in parts of the Chesapeake watershed. Biological communities are considered the definitive measures of stream health because their responses integrate the impacts of multiple stressors (Law et al. 2015). If a stream's Chessie BIBI index score is high or improving over time, it is likely the combined influence of multiple environmental stressors in the stream is low or abating. Identifying which environmental factors are responsible for the net improvement would be speculative at this point, although long-term efforts to conserve forests, preserve and restore riparian corridors and wetlands,

mitigate acid rain and mine drainage, slow stormwater runoff, and reduce nutrients and sediment loads have all likely contributed. Studies have demonstrated the Chessie BIBI index's sensitivity to a few stressors, including altered stream flows (Buchanan et al. 2013, Maloney et al. 2021), elevated nitrogen and phosphorus concentrations (Mandel et al. 2011), and different landcovers and climate change (Maloney et al. 2020). Metrics for a variety of other environmental stressors are currently being explored (e.g., Fanelli et al. 2022) and used to characterize Chesapeake watershed health (e.g., Roth et al. 2020) and will help future investigations of the index's responses to those stressors.

This report offers an initial numeric value for the 2008 Baseline referenced in the 2014 Chesapeake Agreement as well as some evidence for a net improving trend in stream health in the Chesapeake watershed. The purpose of the report is to present monitoring-based results and provide CBP with a process for tracking progress in achieving the 2014 Agreement's stream health goal. The process differs in some respects from processes used by state agencies monitoring the Chesapeake watershed. We fully expected that our results would also differ somewhat from the state results at times, even though the underlying raw data are the same. The Chessie BIBI results should only be used for watershed-based planning and evaluation and not for regulatory assessments.

This work will be continued with a data call to monitoring organizations in late 2024 requesting raw macroinvertebrate counts and associated water quality and habitat data collected from 2018-2023, the second post-baseline interval. We will also continue working with USGS scientists to further refine the random forest model with the intention of providing a companion analysis that uses modeled Chessie BIBI scores to more completely characterize biological conditions in the Chesapeake watershed. We also will be developing confidence intervals around the pre-baseline, baseline, and first interval monitoring-based estimates of healthy stream miles.

Citations

- Buchanan, C., K. Foreman, J. Johnson, and A. Griggs. 2011. Development of a basin-wide benthic index of biotic integrity for non-tidal streams and wadeable rivers in the Chesapeake Bay watershed: Final report to the Chesapeake Bay Program Non-Tidal Water Quality Workgroup. Interstate Commission on the Potomac River Basin, [Report ICP11-01](#).
- Buchanan, C., H. Moltz, H. C. Haywood, J. Palmer, and A. Griggs. 2013. A test of the Ecological Limits of Hydrologic Alteration (ELOHA) method for determining environmental flows. *J. Freshwater Biology* 58:2632–2647. [DOI:10.1111/fwb.12240](https://doi.org/10.1111/fwb.12240).
- Buchanan, C., K. Maloney, Z. Smith, A. Nagel, and J. Young. 2019. Creating a stream health baseline for the Chesapeake basin from monitoring and model data. Report prepared for Chesapeake Bay Program. Interstate Commission on the Potomac River Basin [Report ICP19-7](#).
- Chesapeake Bay Program (CBP). 2014. [Chesapeake Bay Watershed Agreement](#).
- Chesapeake Bay Program (CBP). 2015. Stream Health Outcome Management Strategy, v. 1.0. Prepared by the CBP Stream Health Workgroup.
- Fanelli, R. M., M. J. Cashman, and A. J. Porter. 2022. Identifying key stressors driving biological impairment in freshwater streams in the Chesapeake Bay watershed, USA. *Environmental Management* (2022) 70:926–949. <https://doi.org/10.1007/s00267-022-01723-7>.
- Interstate Commission on the Potomac River Basin (ICPRB). 2018. Workshop to develop a 2008 Baseline for the CBP Stream Health Outcome, Cacapon Resort State Park, WV, 5-6 April 2018. Workshop report prepared for Chesapeake Bay Program. Interstate Commission on the Potomac River Basin [Report ICP18-6](#).

- Law, N., B. Stack, R. Starr, and E. Yagow. 2015. Designing sustainable stream restoration projects within the Chesapeake Bay watershed. Science and Technical Advisory Committee (STAC), [Publication Number 15-003](#), Edgewater, MD. 50 pp.
- Mandel, R., C. Buchanan, A. Griggs, A. Nagel, and O. Devereux. 2011. Data analysis to support development of nutrient criteria for Maryland free-flowing waters. Final report prepared for Maryland Department of the Environment. Interstate Commission on the Potomac River Basin [Report ICP11-2](#).
- Maloney, K. O., Z. M. Smith, C. Buchanan, A. Nagel, and J. A. Young. 2018. Predicting biological conditions for small headwater streams in the Chesapeake Bay watershed. *Freshwater Science* 37(4):795-809. [DOI: 10.1086/700701](#).
- Maloney, K. O., K. P. Krause, C. Buchanan, L. E. Hay, G. J. McCabe, Z. M. Smith, T. L. Sohl, and J. A. Young. 2020. Disentangling the potential effects of land-use and climate change on stream conditions. *Global Change Biology* 26(4):2251-2269. [DOI: 10.1111/gcb.14961](#).
- Maloney, K. O., D. M. Carlisle, C. Buchanan, J. L. Rapp, S. H. Austin, M. J. Cashman, and J. A. Young. 2021. Linking altered flow regimes to biological condition: an example using benthic macroinvertebrates in small streams of the Chesapeake Bay watershed. *Environmental Management* 67:1171–1185. [DOI: 10.1007/s00267-021-01450-5](#).
- Maloney, K. O., C. Buchanan, R. D. Jepsen, K. P. Krause, M. J. Cashman, B. P. Gressler, J. A. Young, and M. Schmid. 2022. Explainable machine learning improves interpretability in the predictive modeling of biological stream conditions in the Chesapeake Bay Watershed, USA. *J. Environmental Management*. 322. [DOI: 10.1016/j.jenvman.2022.116068](#).
- Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77:118–125.
- Roth, N., C. Wharton, B. Pickard, S. Sarkar, and A. R. Lincoln. 2020. Chesapeake healthy watersheds assessment: assessing the health and vulnerability of healthy watersheds within the Chesapeake Bay watershed. Tetra Tech, Inc. Available online at: https://d18lev1ok5leia.cloudfront.net/chesapeakebay/chesapeake_healthy_watersheds_assessment_report.pdf
- Smith, Z. M., C. Buchanan, and A. Nagel. 2017a. Refinement of the basin-wide benthic index of biotic integrity for non-tidal streams and wadeable rivers in the Chesapeake Bay watershed. Interstate Commission on the Potomac River [Report ICP17-2](#) (report).
- Smith, Z. M., C. Buchanan, and A. Nagel. 2017b. Refinement of the basin-wide benthic index of biotic integrity for non-tidal streams and wadeable rivers in the Chesapeake Bay watershed. Interstate Commission on the Potomac River [Report ICP17-2](#) (appendices).
- US Environmental Protection Agency (USEPA). 1997. Field and laboratory methods for macroinvertebrate and habitat assessment of low gradient nontidal streams. Prepared by the Mid-Atlantic Coastal Streams Workgroup, Environmental Services Division, Region 3, Wheeling, WV.
- US Environmental Protection Agency (USEPA). 2016. National Rivers and Streams Assessment 2008 – 2009; a collaborative study. Office of Water and Office of Research and Development. EPA/841/R-16-007. Washington, DC. Available online at: <http://www.epa.gov/national-aquatic-resource-surveys/nrsa>
- U. S. Geological Survey (USGS), National Geospatial Program. 2016. Water Boundary Dataset, v2.3.1. Available online at <https://www.usgs.gov/national-hydrography/watershed-boundary-dataset>.
- Woods, A. J., J. M. Omernik, and D. D. Brown. 1999. Level III and IV ecoregions of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. US Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Corvallis, Oregon.

Supplemental Materials

Supplemental Table 1. Bioregions in the Chesapeake Bay watershed.

Code	Bioregion Name	Corresponding USEPA level III ecoregion and hydrologic subregion
BLUE	Blue Ridge	66
CA	Central Appalachians	69
LNP	Lower Northern Piedmont	64 in Potomac and Lower Chesapeake-James HUC4 subregions
MAC	Mid-Atlantic Coastal Plain	63
NAPU*	Northern Appalachian Plateau & Uplands	60
NCA	North Central Appalachians	62
NRV	Northern Ridge & Valley	67 in Susquehanna HUC4 subregion <i>minus</i> the Great Appalachian Valley portion of 67
PIED	Piedmont	45
SEP	Southeastern Plain	65
SGV	Southern Great Valley	67 in Potomac and Lower Chesapeake-James HUC4 subregions <i>and</i> excluding SRV
SRV	Southern Ridge & Valley	67 in Potomac and Lower Chesapeake-James HUC4 subregions <i>and</i> excluding SGV
UNP	Upper Northern Piedmont	64 in Susquehanna and Upper Chesapeake HUC4 subregions <i>and</i> the Great Appalachian Valley portion of 67

* Includes slivers of level III ecoregion Eastern Great Lakes and Hudson Lowlands (83). Earlier bioregion maps (e.g., Smith et al. 2017) show slivers of Northeastern Highlands (58) in NAPU, however updates to the EPA level III ecoregion shapefiles changed the extent of 58 and it no longer overlaps the Chesapeake Bay watershed. See text for details.

Chesapeake Stream Biological Health

Supplemental Table 2. Annual Chessie BIBI ratings from frequently sampled locations (i.e., nine or more years sampled over eighteen years between 2000 and 2017). HUC12, 12-digit hydrologic code, x indicates more than one location in the HUC12. Location ID, concatenated 4-digit latitude and longitude. N, number of years sampled. Ratings: E, excellent; G, good; F, fair; P, poor; and VP, very poor. Mean, rating of the average numeric score. SE, standard error of the numeric scores. See Supplemental Table 1 for bioregion names.

HUC12	Location ID	N	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Mean	SE	Bio-region
20600030302	39.6761-76.7717	10									E	E	E	E	E	E	E	E	E	E	E	1.06	UNP
20600060507	38.5639-76.7601	12	E	E	E	E	E	E	E	G	G	G	E							G	E	2.47	SEP
20700010102	38.6271-79.5688	12		E						E	E	E	G	G	E	E	E	E	E	E	E	2.36	SRV
20700010104	38.8494-79.4869	12		E					E	G	F	G	E	E		G	E	E	E	E	E	3.51	SRV
20700010302	38.4353-79.5214	10	E	E	E	G	E	E	E	E				E						F	E	3.64	SRV
20700010509	39.0131-78.9567	11							F	E	G	E	E	E	E	E	E	E	E		E	3.57	SRV
20700030301	39.6867-78.4557	11	G	G	E	E	G	G	E	E	E	E	G								E	2.37	SRV
20700030702	39.042-78.598	9			E	E	F	G	G	E	E	E	E								E	4.61	SRV
20700040201	39.4996-78.2997	11	G	E			E			G		E	G			E	E	E	E	E	E	2.57	SRV
20700060701	38.7294-78.5333	9	E		G	E	E	E	E				E			E			E		E	1.97	SGV
20700090505	39.6097-77.4156	9	G	E	E	E	E	E	E	E	E										E	1.01	LNP
20700090702	39.2739-77.3006	12				E	E	E	G	G	E				E	E	G	G	E	F	E	1.90	LNP
x 20700110104	38.5763-77.3776	13					F	E	E		E	E	E	E	E	E	E	E	E	E	E	1.78	PIED
x 20700110104	38.5775-77.382	13					P	F	E		E	E	E	E	E	E	E	E	E	E	E	3.59	PIED
x 20700110104	38.5764-77.375	13					F	E	P		E	G	E	E	G	E	E	G	G	E	E	3.83	PIED
20700110403	38.4994-76.9035	10	E	E	E	G	E	E	E	E	E		E								E	2.80	SEP
20700110701	38.3666-76.7372	10	E	E	E	E	E	E		E	E	E	E								E	1.93	SEP
20801030301	38.6543-78.1311	9					E	E	E	E	G			E	E	E	G				E	1.57	LNP
20801030401	38.5222-78.1717	9	E	E	E	E	E	E		E	E				E						E	0.97	LNP
20801030402	38.6031-78.2528	13	E	E	E	E	E	E	E	E	E		E	E	E	E					E	0.47	LNP
20801030701	38.4378-78.3678	13	G	E	E	E	E	E	E	E	G		E	E	E	E					E	0.85	BLUE
20801030906	38.3361-78.0908	10		E	G	E	E	E	E	E	G			E					F		E	2.31	LNP
20801060401	38.1619-78.0383	13	E		E	E	E	E	G	E	E		E	E	E	E	E				E	1.39	PIED
20802010701	38.0778-79.6594	10	E	E	E	G	G	G	G	E	E		E								E	2.53	SRV
20802040504	37.8451-78.2296	10	E	E	G	E	F	E	E	E	E			E							E	2.88	PIED
20802060903	36.9203-76.6514	9	E	E	E	E	E	E		E	E	G									E	2.61	MAC
20600020103	39.6106-76.0461	11	G	G	E	F	G	E	F	G	F	E	E								G	2.42	UNP
20600030302	39.6619-76.781	14				P		F	F	E	E	G	G	E	G	E	G	G	E	E	G	2.65	UNP
x 20600030402	39.6056-76.6424	13						P	F	P	G	G	E	F	E	F	G	G	G	G	G	3.04	UNP
x 20600030402	39.5898-76.5863	11				P		E	G	E	P	G	F	G	G	E	P				G	3.45	UNP
x 20600030405	39.4805-76.6983	10	G	E	F	F	G	E		G	E	E	G								G	2.06	UNP
x 20600030405	39.4799-76.6792	11				P		E	G	G	E	G	E	F	E	F	E				G	4.52	UNP
20600030805	39.4406-76.8642	16		G	P	G	E	G	P	G	G	E	G		F	F	G	E	E	G	G	2.05	UNP
20600030901	39.417-76.8258	13				VP				G	G	P	G	E	G	G	G	G	E	E	G	5.12	UNP

Chesapeake Stream Biological Health

HUC12	Location ID	N	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Mean	SE	Bio-region
20600060105	39.1695-76.972	12	G	E	F	P	E	G	G	P	G	F	G							G	G	2.09	UNP
20700010202	38.9941-79.1118	10							F	F	G	G	F	G		F	F	E		G	G	2.76	SRV
20700041004	39.6583-77.5453	9	G	G									E		E	E	F	G	G	E	G	2.92	BLUE
20700050203	38.1572-78.9719	12	P	F	G	G				F	E			P	G	G	E	P	G		G	2.80	SGV
20700050504	38.3928-78.9797	9	G	G	P			E			E			P	E	G				P	G	4.88	SGV
20700060305	38.7553-78.6706	9	G	G	F	G	G	P			E			F	E						G	2.73	SGV
20700060403	38.8422-78.6025	9	G	E	P		E	P			G				G	P		F			G	4.99	SGV
20700080505	39.0136-77.6997	11		E	G	G	E	E		P	G	F	G					G	P		G	3.61	LNP
x 20700080801	39.2194-77.3154	17		E	G	G	G	E	G	P	G	G	F	G	P	E	G	F	F	P	G	2.88	LNP
x 20700080801	39.2186-77.3175	12				P				G	G	G	G	G	G	E	F	P	G	P	G	3.66	LNP
x 20700080801	39.2251-77.3107	13				G	E		G	G		G	E	G	G	E	E	P	P	P	G	5.18	LNP
20700090505	39.5864-77.4915	9	F		E	F	F	E		G	G	E	E								G	3.79	BLUE
20700100101	39.1506-77.1245	11					G	F	G	F		G	G		G	G		E	G	E	G	1.82	LNP
x 20700100803	38.6593-77.1786	13					G	E	G		G	G	E	E	F	E	E	E	E	E	G	5.03	SEP
x 20700100803	38.6586-77.1796	13		P			G	E	G		F	F	G	G	F		E	G	G	F	G	5.35	SEP
20700110102	38.5823-77.0977	11	E	E	F	F	E	E		E	G	E	G					E			G	5.12	SEP
x 20700110104	38.5843-77.4137	12					P	E	F		E	E	F	G	E	E	G		E	F	G	2.88	PIED
x 20700110104	38.5768-77.3667	13					G	E	P		E	P	E	F	G	E	F	F	F	P	G	3.42	PIED
x 20700110104	38.6018-77.3666	13					G	E	P		G	P	G	P	E	E	E	E	F	G	G	3.60	PIED
x 20700110104	38.5724-77.3482	13					F	E	P		G	F	F	E	E	E	E	G	G	F	G	3.66	PIED
x 20700110104	38.5676-77.3432	10									F	P	E	G	E	E	E	F	G	G	G	3.72	PIED
x 20700110104	38.5859-77.3792	13					P	G	P		E	E	E	P	F	E	E	F	F	E	G	4.79	PIED
20700110302	38.4839-77.0859	10	G	E	G	G	E	E		E	G	G	E								G	4.51	SEP
20700110304	38.5111-77.0291	11	E	G	G	G	G	E	E	G	G	G	E								G	3.40	SEP
20801030203	38.5828-77.8758	9	G	G	G	G	E		G	E	E			G							G	1.25	LNP
20801110402	37.9125-75.5917	18	E	E	E	E	G	P	F	P	E	G	G	G	G	G	E	F	G	F	G	4.44	MAC
20802030505	37.7019-79.0278	12	E	E	G	E	E	F	G	F	F			G			F		G		G	2.71	PIED
20501030203	41.9897-76.0006	9			E		F	G	F			F	F	F	F	F					F	2.49	NAPU
20600030402	39.6052-76.6728	14				VP		P	G	P	G	G	F	P	F	F	F	G	G	G	F	4.29	UNP
20600031201	39.4395-76.6984	12	P	F	F	E	G	E	P	P	P	P	F					P			F	3.08	UNP
20600060401	39.0905-76.8703	11								F	P	F	F	F	F	P	VP	F	G	G	F	5.04	SEP
20700050704	38.0569-78.9078	13	F	G	G				P	P	E			P	F	P	G	P	F	G	F	2.83	SGV
20700060204	38.6756-78.6405	9	G	P							G			E	G	P	E	P	G		F	5.19	SGV
20700060302	38.6184-78.7957	11		G	P	P	G				F			P	E		E	P	G	P	F	5.06	SGV
20700060502	38.9261-78.4247	10	P	G	P	F	P			G	F			E	G			F			F	3.26	SGV
20700080301	39.1464-77.7322	9		F	G	F		G			F	P	P		G	E					F	3.84	LNP
x 20700080801	39.2256-77.3108	12				G				F	G	P	G	G	P	G	F	F	P	P	F	2.85	LNP
x 20700080801	39.2329-77.308	12				P				P	P	F	P	P	F	E	F	G	F	P	F	4.06	LNP
x 20700080801	39.2296-77.3111	14				G		G	G	G	E	G	G	F	VP	E	F	G	G	P	F	4.51	LNP

Chesapeake Stream Biological Health

HUC12	Location ID	N	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Mean	SE	Bio-region	
x	20700080801	39.2065-77.2887	15		G	F	G	G	P	P		G	G	G	VP	G	P	F	P	P	F	5.01	LNP	
	20700080802	39.2153-77.2284	10		F		G	G	F	G	P	F	P	G		G					F	4.89	LNP	
x	20700100101	39.1741-77.1078	14					P	F	G	E	P	F	P	F	G	F	F	P	P	F	3.37	LNP	
x	20700100101	39.1241-77.1066	14					E	E	G	E	G	G	G	G	P	P	P	P	P	F	5.39	LNP	
	20700100202	39.1019-76.9528	14		VP	G	P	VP	F	G	G		G	F	P	F	F			G	G	F	5.42	LNP
	20700110104	38.5932-77.3759	13					P	P	P		F	G	F	F	E	F	G	F	P	P	F	3.17	PIED
	20801060101	38.0919-78.19	9					P	G	G	G	F	G						F	F	F	F	2.44	PIED
	20801110402	37.9028-75.5792	18	G	G	G	E	G	F	P	VP	E	F	P	F	F	F	E	F	F	F	F	4.57	MAC
	20802030504	37.7153-78.9817	10	E	P	P	E				E			F			F	P	P	E	F	5.01	PIED	
	20502010309	40.97-78.4072	9			P						P		E	P	P	F	F	P	P	P	3.95	CA	
	20600020401	39.2877-75.7836	10					P				VP	VP	P		VP	VP	P	P	P	P	2.82	MAC	
	20600020503	39.1939-76.2186	10	F	P	P	VP	VP	VP		VP	VP	P	VP							P	3.91	MAC	
	20600030401	39.656-76.6425	14					VP		P	F	P	G	E	G	G	P	E	F	F	P	G	5.13	UNP
	20600030406	39.4614-76.5898	9						VP			P	P	P	P	P	P			P		4.49	UNP	
	20600030702	39.3445-76.5166	9										VP	VP	P	VP	F	P	G	F	F	6.79	SEP	
	20600031201	39.4293-76.7188	11	G	F	G	P	G	G	P	P	P	P	P								P	3.82	UNP
	20600050304	38.7241-75.9606	9									P	VP	F		P	P	VP	VP	VP	G	6.23	MAC	
	20600060401	39.0891-76.8899	10									P	P	F	G	F	P	P	VP	VP	VP	5.51	SEP	
	20700010302	38.4153-79.5767	9	F	F	F			F	VP	F			G	P						P	4.54	SRV	
	20700040905	39.3458-78.0674	9								VP	P	VP	VP	VP		P	P	P		P	3.13	SGV	
	20700050101	38.0819-79.2467	9			VP		VP	VP					VP	P	P	VP	P	P		P	4.61	SGV	
x	20700050503	38.4517-78.9872	10	VP			VP	P		P	P			G	E	G	E	P			P	6.40	SGV	
x	20700050503	38.486-78.96	9			VP	P	P	P	P	P	P		P							P	2.79	SGV	
x	20700050603	38.3725-78.9342	12	P	P	VP	VP		VP	P				P	VP	VP	P	P	G		P	4.60	SGV	
x	20700050603	38.3543-78.9394	9							P	P	P		VP	VP	P	P	P	G		P	4.29	SGV	
	20700050606	38.3161-78.8189	10	VP	VP	VP		P	F	P	G	G							G	P		6.31	SGV	
	20700050804	38.4111-78.7125	10	P	P		P	P	P	P				P			P	F			P	2.74	SGV	
	20700060301	38.6325-78.8428	9	G	F					P	P			P	F	F	P	P			P	2.86	SGV	
	20700070105	39.0656-78.0042	9				P		P	P	P				P	P	P			P	P	2.62	SGV	
x	20700080403	39.1864-77.4492	14		G	P	P	G	E	F	F	F			P	VP	VP	P	P	G	P	5.26	LNP	
x	20700080403	39.1969-77.4628	14		P	P	VP	P	P	VP	VP	F			P	P	P	VP	VP	VP	P	5.07	LNP	
x	20700080801	39.2224-77.2505	12		F		G		G	G				G	G	F	P	VP	P	VP	VP	7.73	LNP	
x	20700080801	39.2119-77.2848	9										F	F	P	P	F	F	G	P	P	6.01	LNP	
x	20700080801	39.2319-77.2762	10						E	F				G	F	P	P	P	F	P	VP	6.00	LNP	
x	20700080801	39.2389-77.2714	17		P	E	VP	VP	VP	VP	VP	P	F	P	P	F	P	P	P	VP	P	5.73	LNP	
x	20700080801	39.2316-77.2558	17		G	G	F	F	P	VP	VP	P	P	P	VP	P	P	VP	P	P	P	5.63	LNP	
x	20700080801	39.2264-77.2574	15		F		E	G	P	VP		F	VP	G	P	P	P	P	P	P	VP	5.63	LNP	
x	20700080801	39.2287-77.26	17		G	G	VP	VP	P	VP	VP	F	P	P	P	P	P	P	P	VP	P	5.55	LNP	
x	20700080801	39.2226-77.2501	16		E		F	F	G	P	F	P	G	G	P	P	P	VP	P	P	VP	5.52	LNP	

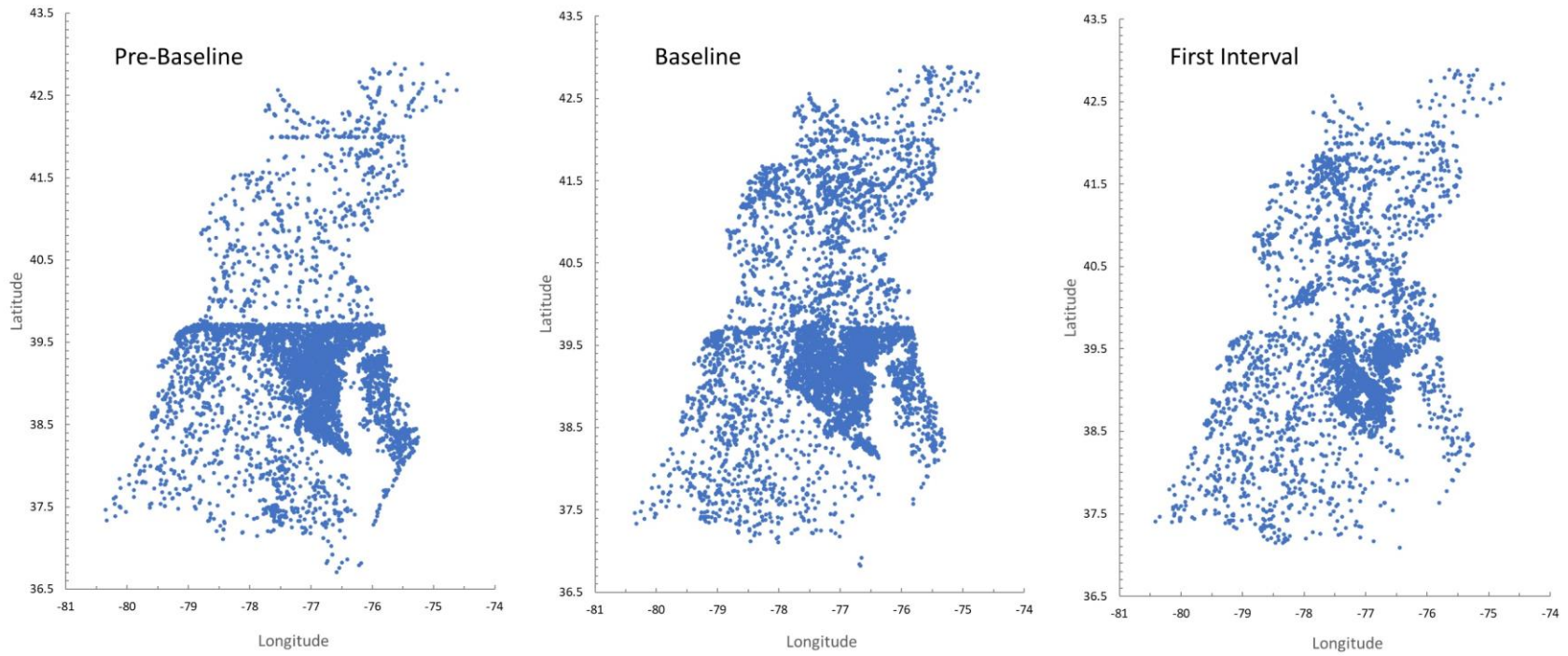
Chesapeake Stream Biological Health

HUC12	Location ID	N	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Mean	SE	Bio-region
x	20700080801	39.2151-77.2909	16		F	G	P	G	G	P	F	G	E	P	VP	P	P	P	F	P	P	4.98	LNP
x	20700080801	39.2217-77.2583	13	VP	F	G	F	F	F					P	P	P	P	P	P	P	P	4.70	LNP
x	20700080801	39.2317-77.3076	12			P				P	VP	F	G	P	F	P	P	P	P	P	P	4.65	LNP
x	20700080801	39.2341-77.2503	16	E		F	F	F	P	P	F	P	P	P	P	P	P	P	P	VP	P	4.38	LNP
x	20700080801	39.2114-77.2777	17	P	F	P	P	P	P	P	F	P	F	P	VP	P	VP	P	P	P	P	4.28	LNP
x	20700080801	39.2226-77.2594	17	F	F	P	P	P	VP	P	F	P	P	P	P	P	P	P	VP	P	P	3.68	LNP
x	20700080801	39.234-77.2927	17	P	P	P	VP	VP	VP	VP	VP	P	P	P	P	P	P	VP	P	P	P	3.24	LNP
	20700081002	39.0499-77.224	17	P	P	P	F	P	P	G	P	P	P	P	VP	VP	VP	VP	P	P	P	4.90	LNP
x	20700081004	38.9101-77.3377	9									P	P	F	P	P	P	P	P	VP	P	4.50	PIED
x	20700081004	38.9095-77.3376	9									P	P	VP	VP	VP	P	P	P	VP	P	3.43	PIED
	20700090701	39.3425-77.3497	9	VP	E	VP	P	F	P	VP	P	VP										9.39	LNP
	20700100101	39.1522-77.1272	11				P	P	P	P		P	P		P	P		P	VP	VP	P	3.16	LNP
	20700100101	39.165-77.099	13				P	G	P	F		G	G	F	P	F	P	P	P	P	P	3.12	LNP
	20700100201	39.1394-77.0267	9	P			VP					F		P	P	P	P	VP	P		P	6.78	LNP
x	20700100202	39.1049-76.9788	14	F	VP	G		P	VP			P	VP	VP	P	P	VP	VP	P	VP	P	5.36	LNP
x	20700100202	39.0891-76.9634	15		VP	P	VP	P	VP	P	P	P		P	P	P	E	P	P	P	P	4.69	LNP
x	20700100202	39.0944-76.9585	17	VP	G	P	P	F	P	P	P	P	P	F	P	P	VP	VP	P	P	P	4.51	LNP
x	20700100202	39.0857-76.9845	15		E	P	VP	F	P	P	P	P	P	P	P			P	P	P	P	4.30	LNP
x	20700100202	39.0834-76.9748	16	G	F	P	P	P	VP	VP	P	P	F	P	P	P		VP	P	P	P	4.14	LNP
x	20700100202	39.1049-76.9542	15	P	G	G	P	G	P	P	P	P	G	G	P	P			P	F	P	3.97	LNP
x	20700100202	39.0785-76.9773	17	P	P	P	VP	P	VP	VP	P	P	VP	P	VP	VP	P	P	P	VP	P	3.65	LNP
x	20700100202	39.0893-76.9756	15	F	G	F	VP	G	F	P	F	P	P	F	F	P			P	P	P	3.63	LNP
x	20700100202	39.1013-76.9676	15	P	VP			VP	VP	P	P	VP	VP	VP	VP	P	VP	VP	VP	P	P	3.12	LNP
	20700100705	38.7885-77.3806	9									G	P	P	F	P	P	P	F	VP	P	4.96	PIED
	20700110104	38.5894-77.383	13				VP	P	VP		F	P	P	VP	F	F	P	P	P	P	P	4.13	PIED
	20801100301	38.4142-75.5914	11	VP	P	P	P	F	F	P	F	P	P	F								3.32	MAC
	20801110101	38.2633-75.4625	9		F	P	VP	P	VP	F	G	P	VP									5.44	MAC
	20802020202	37.9499-79.4592	12	P	F	F		P	VP	P				G	P	P	F	P	G		P	4.25	SRV
	20802020303	37.7794-79.4561	10				VP	F	P	P	P				P	P	E		G	E	P	7.13	SGV
x	20600030502	39.4215-76.5389	12						VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	P	P	VP	4.51	UNP
x	20600030502	39.4149-76.5013	9								VP	P	P	P	VP	VP	P		VP	P	VP	4.12	UNP
x	20600030502	39.4114-76.5544	11					P	VP	VP		VP	VP	VP	VP	VP	VP	VP	VP		VP	3.12	UNP
x	20600030502	39.4025-76.5761	9					VP	VP	VP	VP	VP	VP	VP	VP	VP					VP	2.31	UNP
x	20600030702	39.3124-76.5546	15			F	VP	P	VP	VP	P	VP	VP	VP	VP	VP	VP	VP	VP	P	VP	3.09	SEP
x	20600030702	39.3308-76.5351	12	VP			VP	P		P		VP	VP	VP	VP	VP	P	P		P	VP	2.92	SEP
x	20600030702	39.3368-76.5397	15	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP		VP	VP	VP	1.37	SEP
x	20600030902	39.3051-76.6873	14	VP	VP	VP	VP	VP	VP			VP	VP	VP	VP	VP	VP	P	VP	P	VP	3.95	UNP
x	20600030902	39.364-76.7654	10							VP	VP	P	VP	VP	VP	VP	VP	VP	VP		VP	3.58	UNP
x	20600030902	39.3603-76.7479	10							VP	P	VP	VP	VP	VP		P	P	VP	P	VP	2.83	UNP

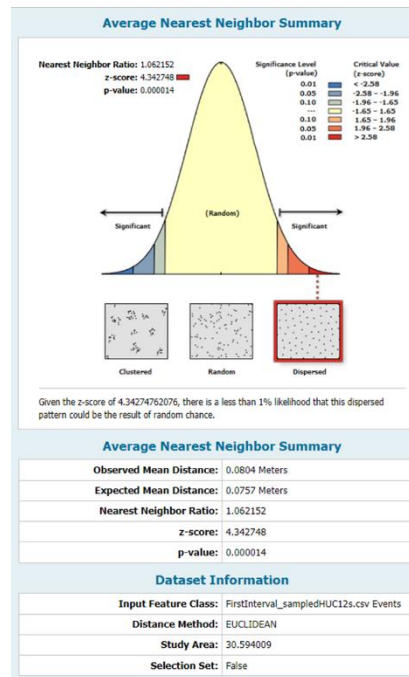
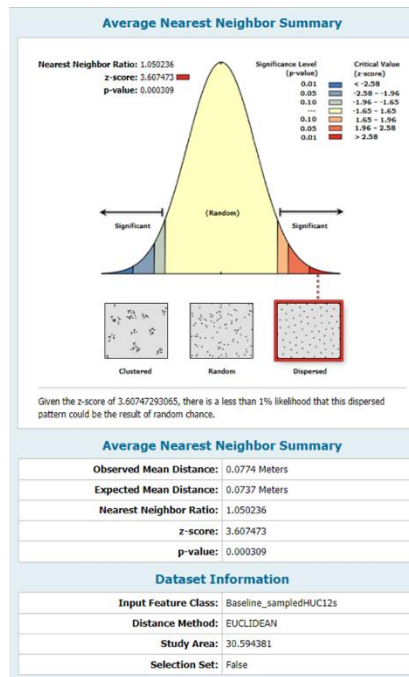
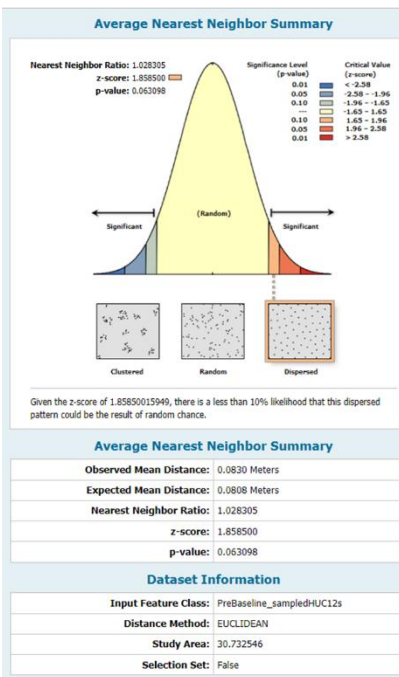
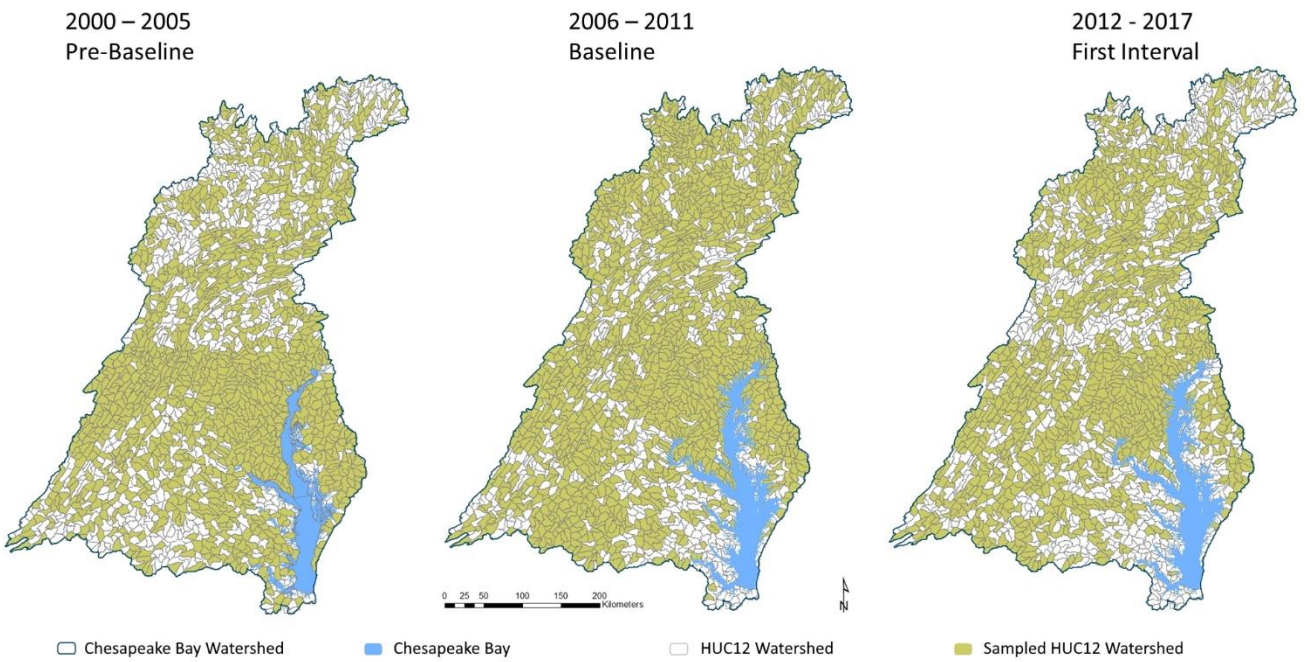
Chesapeake Stream Biological Health

HUC12	Location ID	N	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Mean	SE	Bio-region
x	20600030902	39.2789-76.6927	9			VP							VP	VP	VP	VP	VP	VP	VP	VP	VP	2.68	UNP
x	20600030902	39.3796-76.7974	11						VP		VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	2.31	UNP
x	20600030902	39.3719-76.7791	11							VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	1.99	UNP
x	20600030902	39.38-76.8058	13					VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	1.48	UNP
x	20600031201	39.3268-76.6252	14			VP	VP	VP		VP	VP	VP	VP	P	VP	VP	VP	P	VP	VP	VP	3.35	UNP
x	20600031201	39.4-76.7277	9					VP	VP	VP	P	P	VP	VP	P					P	VP	2.74	UNP
	20700010302	38.4167-79.5728	9	VP	VP	VP		VP	VP	VP				P	VP						VP	2.19	SRV
	20700040503	39.6484-78.2126	10							VP	VP	VP	P	VP	P	P	VP	VP		VP	VP	3.00	SRV
	20700040909	39.5175-77.8894	9							VP	VP	VP	VP	VP	VP		VP	P		P	VP	2.68	SGV
x	20700050602	38.3734-78.9331	12		P	P	VP		VP	P	VP			VP	VP	P	VP	VP	P		VP	4.48	SGV
x	20700050602	38.4186-78.8894	13	VP	VP	VP	VP		VP	VP	VP			VP	VP	VP	VP	P	P		VP	2.76	SGV
	20700050803	38.405-78.745	9	VP			VP	P	VP	VP	P			VP			VP	P			VP	3.92	SGV
	20700050804	38.4114-78.6994	9	VP	VP	P		VP	VP	VP	VP			P				P			VP	4.86	SGV
	20700060303	38.6339-78.7486	11	VP			VP		VP	VP		VP	VP	VP	VP	VP	VP	VP			VP	1.56	SGV
x	20700081002	39.0872-77.208	15		P	F		VP	VP	VP		VP	VP	VP	VP	VP	VP	VP	VP	P	VP	5.40	LNP
x	20700081002	39.0823-77.2114	9		P	P	VP	VP	VP	VP	VP	VP									VP	4.94	LNP
x	20700081002	39.0885-77.2079	12		P	P		VP	VP	VP	VP	VP	VP	VP	VP	VP					VP	3.46	LNP
x	20700081002	39.088-77.2057	12		VP	P		VP	VP	VP	VP	VP	VP	VP	VP	VP					VP	3.15	LNP
x	20700081002	39.0619-77.2086	15		VP	P	VP	VP	VP	VP	VP	P	P	VP		VP	P	VP	VP	P	VP	3.03	LNP
x	20700081004	38.8845-77.3328	9									P	VP	VP	VP	VP	VP	VP	P	VP	VP	3.34	PIED
x	20700081004	38.9842-77.2874	9									VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	2.32	LNP
x	20700081004	38.938-77.2568	9									VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	2.28	LNP
	20700081005	38.9596-77.1762	9									VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	1.52	LNP
x	20700090701	39.3492-77.3499	9		VP	F	VP	VP	VP	VP	VP	VP									VP	6.94	LNP
x	20700090701	39.3452-77.3497	9		P	P	VP	VP	P	P	VP	P	VP								VP	6.72	LNP
x	20700090701	39.3327-77.3524	9		VP	VP	VP	VP	VP	VP	VP	VP									VP	2.90	LNP
x	20700100201	39.1175-77.0224	9		P			VP				VP		VP	VP	VP	VP	VP	VP		VP	5.56	LNP
x	20700100201	39.0481-77.0157	9		VP	VP		VP			VP		VP				VP	VP	VP		VP	1.93	LNP
	20700100202	39.0955-76.9969	13			P	VP	VP	VP	VP		VP	VP	VP	VP	VP			VP	VP	VP	3.99	LNP
	20700100307	38.7431-77.0684	9										VP	VP	VP	VP	VP	P	VP	P	VP	2.58	SEP
x	20700100704	38.8831-77.4311	9										VP	VP	VP	P	VP	VP	VP	VP	VP	4.15	LNP
x	20700100704	38.8684-77.4067	9										VP	VP	VP	VP	VP	VP	VP	VP	VP	2.29	LNP
x	20700100704	38.881-77.4303	9										VP	VP	VP	VP	VP	VP	VP	VP	VP	0.92	LNP
	20700100705	38.797-77.3486	9										VP	VP	VP	VP	VP	P	VP	VP	VP	4.34	PIED
	20801110102	38.2494-75.4931	11	P	VP	VP	P	VP	VP	VP	P	VP	P	VP							VP	4.03	MAC

Chesapeake Stream Biological Health

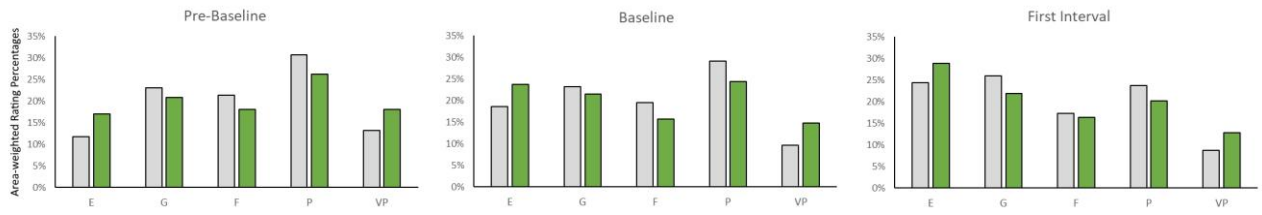


Supplemental Figure 1. Sampling locations in the pre-baseline (2000 – 2005), baseline (2006 – 2011), and first interval (2012 – 2017) periods.



Supplemental Figure 2. Distribution of sampled HUC12s in each period. Spatial analysis conducted with the Average Nearest Neighbor tool in ArcGIS (version 10.5).

Chesapeake Stream Biological Health



Supplemental Figure 3. Estimated watershed percentages of the Chessie BIBI ratings when ratings at individual locations are area-weighted by equal portions of their HUC12-bioregion unit's area (green bars), and when ratings are averaged and the average's rating is area-weighted by the unit's area (gray bars). Pre-baseline, 2000 – 2005; baseline, 2006 – 2011; first interval, 2012 – 2017. Ratings: E, excellent; G, good; F, fair; P, poor; VP, very poor. See also Figure 5.