Development of Ecosystem Health Indexes For Non-Tidal Wadeable Streams and Rivers in the Chesapeake Bay Basin

Progress report prepared for the
CBP Non-Tidal Water Quality Workgroup

December 5, 2008

Katie Foreman (CBPO)
Andrea Nagel (ICPRB)
Claire Buchanan (ICPRB)

Contents

Summary

Staff of the Interstate Commission on the Potomac River Basin (ICPRB) and Chesapeake Bay Program Office (CBPO) expanded to the Chesapeake Bay basin a Benthic Index of Biotic Integrity (B-IBI) that was originally developed for the multi-jurisdictional Potomac River basin (Astin 2006, 2007). The index uniformly scores data collected by different programs monitoring non-tidal, wadeable streams and rivers. We merged comparable data collected by the states and river basin commissions between 2000 and 2006; appended post-1999 state, federal, and county Potomac data previously processed by ICPRB; verified Astin's scoring approaches for habitat and biological metrics in the Piedmont, Highland, and Valley physiographic regions; applied the Maryland Department of Natural Resources (MD DNR) scoring criteria to biological communities in the Coastal Plain region; developed scoring thresholds for the Northern Appalachian region; and analyzed and mapped the results. The results are mapped in several ways to facilitate CBP Non-Tidal Water Quality Workgroup (NTWQW) discussion. Non-tidal wadeable streams and rivers in approximately 49% of the Chesapeake Bay basin attain a passing grade. A passing grade is 3 or better on a 1 – 5 scale (or 21 or better on a 7 – 35 scale). The results of this analysis need to be reviewed before they are submitted to the CBP for its annual Chesapeake Bay Health and Restoration Assessment reports.

Data Acquisition

The ICPRB and CBPO acquired benthic, physical, and chemical data for non-tidal, wadeable streams from the Maryland Department of Natural Resources (MD DNR), Pennsylvania Department of Environmental Protection (PA DEP), Virginia Department of Environmental Quality (VA DEQ), West Virginia Department of Environmental Protection (WV DEP), the Susquehanna River Basin Commission (SRBC), and Delaware Department of Natural Resources and Environmental Control (DE DNREC) for the 2000 – 2006 time period. Staff processed, QA/QC'ed, and merged the data into the ICPRB MSAccess relational database structure developed by Astin (2006). These data were used to develop the Chesapeake Bay basin B-IBI. Additional, post -1999 data previously acquired and processed by ICPRB were subsequently appended to the database and used in the final evaluation of watershed status. These included data from PA DEP, the U. S. Forest Service's (USFS) Jefferson/George Washington National Forest Stream Assessment, the Fairfax County Stream Protection Strategy program (FC SPS), Montgomery County Department of Environmental Protection (MC DEP), and Prince George's County Department of Environmental Resources (PGC DER). All data sources and start/end years, are listed in Table 1.

Bioregions

One fundamental concept in evaluating stream and river health is to group data by ecoregion to reduce variability in natural communities caused by climate, geography, soil, and other natural factors. Astin (2006) aggregated USEPA level III ecoregions and Level IV subregions into four "bioregions" for the Potomac River basin: Southeastern Plains, Piedmont, Valleys, and Highlands. These four bioregions extend across most of

Table 1. Data sources. The asterisked data sets (*) were used to develop the B-IBI for the Chesapeake Bay basin. All of the data sets were used in portraying 2000-2006 status.

Collecting Agency	Collection Dates Start	End	Number of samples Benthic	Habitat
DE DNREC*	10/16/2001	11/14/2005	36	36
FC SPS	7/31/2001	12/31/2001	41	33
MC SPS	3/16/2000	9/29/2005	506	553
MD DNR	3/1/2000	4/26/2007	523	1,268
PA DEP*	4/6/2000	1/14/2008	500	301
PA DEP	5/24/2000	12/04/2003	153	153
PGC DER	3/6/2000	4/21/2004	161	209
SRBC*	4/14/1986	8/15/2008	947	710
USFS	5/18/2000	5/8/2003	7	
VA DEQ*	3/7/2000	12/11/2007	1,486	1,280
WV DEP*	3/28/2000	1/16/2008	503	512

the Chesapeake Bay basin. We kept Astin's Piedmont, Valley, and Highland bioregions; combined her Southeastern Plain bioregion with USEPA's Mid-Atlantic Coastal Plain ecoregion to form a Coastal Plain bioregion; and created a fifth bioregion, Northern Appalachian, in the Pennsylvania and New York portions of the Chesapeake basin (Table 2, Figure 1).

Habitat and Water Quality Conditions

<u>Piedmont, Valley, and Highland Bioregions</u>

We characterized reference and degraded sites in the Piedmont, Highland, and Valley bioregions using the suite of seven habitat and two water quality metrics originally identified in Austin (2006):

Habitat metrics:

- Anthropogenic Alternations (ANTHRO_ALT): evaluates evidence of human disturbance (i.e., grazing, aesthetics, remoteness)
- Bank Stability (BANK_STAB): measures stream bank erosion or potential for erosion
- Channel Alteration (CHAN_ALT): measures changes to the natural stream channel (e.g., straightening, dredging, artificial armor, bar development)
- Habitat Heterogeneity (HAB_HETERO): evaluates habitat diversity (e.g., riffle frequency, pool/glide/eddy quality)
- Instream Condition (INSTR_COND): evaluates the abundance of suitable substrate (e.g., cobble, gravel, woody debris)
- Riparian Zone (RIP_ZONE): evaluates the extent of natural vegetation of the stream edge and riparian zone
- Substrate Quality (SUB_QUAL): measures the degree to which the substrate is cover by silt, sand, or mud

Water quality metrics:

- pH (pH): evaluates acidity, Conductivity (CON): evaluates habitat suitability

Table 2. ICPRB/CBP Bioregions developed from EPA Level III Ecoregions and Level IV Subregions (www.epa.gov/bioiweb1/html/usecoregions.html).

Bioregion	EPA Level III Ecoregion		EPA Level IV Subregion	
Northern Appalachians - Subregions Not Classified	58	Northeastern Highlands	58unk	NY – Level IV unknown
	83	Eastern Great Lakes and Hudson Lowlands	83unk	NY – Level IV unknown
	60	Northern Appalachian Plateau and Uplands	60unk	NY – Level IV unknown
	62	North Central Appalachians	62unk	NY – Level IV unknown
Northern Appalachians	60	Northern Appalachian Plateau and Uplands	60a	Glaciated Low Plateau
			60b	Northeastern Uplands
	62	North Central Appalachians	62a	Pocono High Plateau
			62b	Low Poconos
			62c	Glaciated Allegheny High Plateau
			62d	Unglaciated Allegheny High Plateau
Coastal Plain	63	Middle Atlantic Coastal Plain	63b	Chesapeake-Pamlico Lowlands and Tidal Marshes
			63f	Delmarva Uplands
			63e	Mid-Atlantic Flatwoods
			63c	Swamps and Peatlands
			63unk	VA – Level IV unknown
			63d	Virginian Barrier Islands and Coastal Marshes
	65	Southeastern Plains	65n	Chesapeake Rolling Coastal Plain
			65m	Rolling Coastal Plain
Highlands	66	Blue Ridge	66a	Northern Igneous Ridges
			66b	Northern Sedimentary and Metasedimentary Ridges
	67	Ridge and Valley	67d	Northern Dissected Ridges and Knobs
			67c	Northern Sandstone Ridges
			67i	Southern Dissected Ridges and Knobs
			67h	Southern Sandstone Ridges
	69	Central Appalachians	69a	Forested Hills and Mountains
			69b	Uplands and Valleys of Mixed Land Use
	70	Western Allegheny Plateau	70c	Pittsburgh Low Plateau
Piedmont	45	Piedmont	45c	Carolina Slate Belt
			45e	Northern Inner Piedmont
			45f	Northern Outer Piedmont
			45g	Triassic Basins
	58	Northeastern Highlands	58h	Reading Prong
	64	Northern Piedmont	64d	Piedmont Limestone/Dolomite Lowlands
			64c	Piedmont Uplands
			64b	Trap Rock and Conglomerate Uplands
			64a	Triassic Lowlands
Valleys	67	Ridge and Valley	67e	Anthracite Subregion
			67a	Northern Limestone/Dolomite Valleys
			67b	Northern Shale Valleys
			67f	Southern Limestone/Dolomite Valleys and Low Rolling Hills
			67g	Southern Shale Valleys

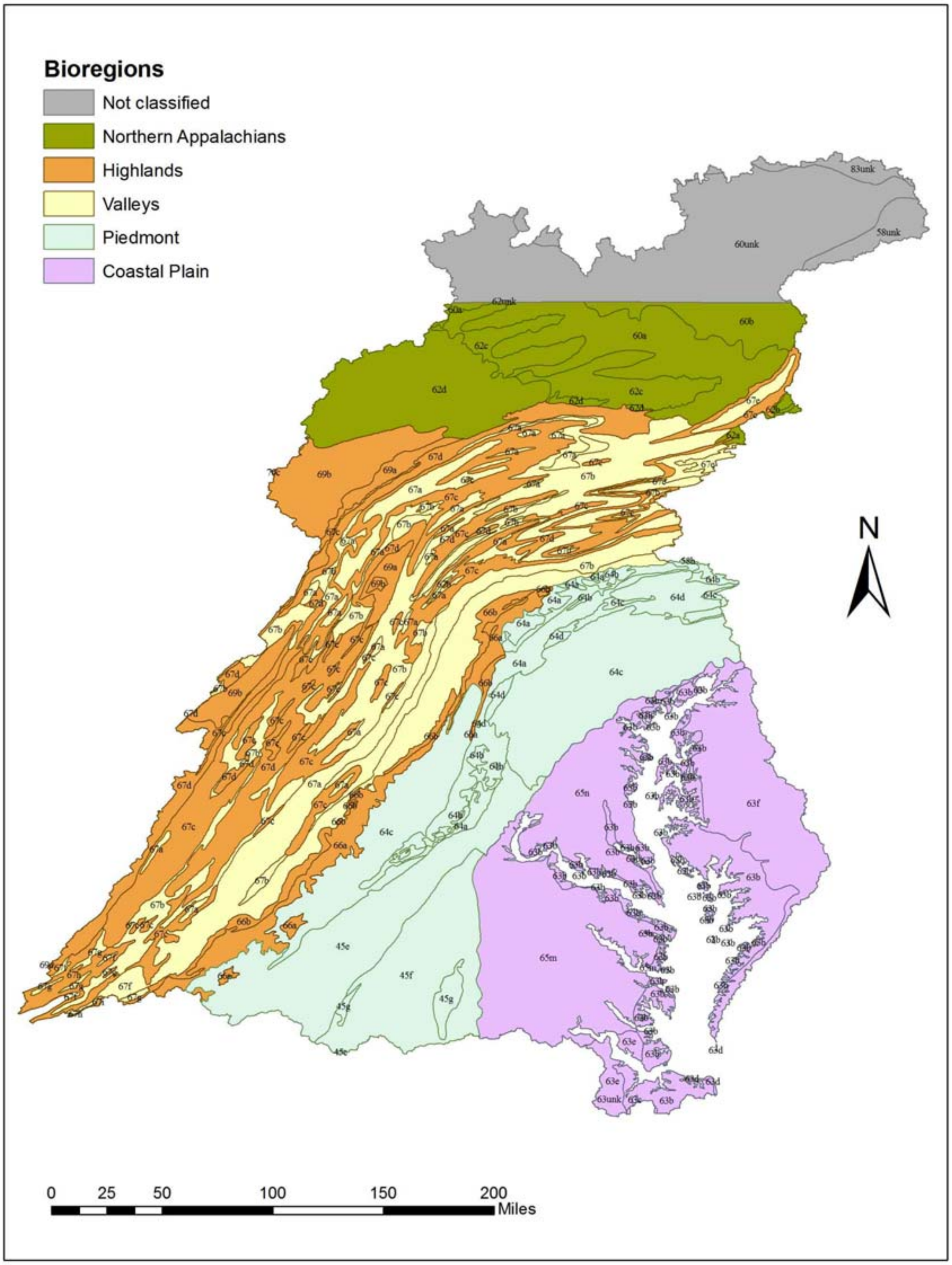

Figure 1. Chesapeake Bay basin bioregions. EPA Level IV subregions (see Table 2) are labeled.

Scores for BANK_STAB and CHAN_ALT, (all on a 0-20 scale) were obtained directly from the collecting agencies. Scores for ANTRO_ALT, HAB_HETERO, INSTR_COND, SUB_QUAL, and RIP_ZONE (all on a 0-20 scale) were calculated and/or converted from state-specific metrics to standard metrics for the purposes of comparability with the following equations: ANTHRO_ALT = GRAZ OR (AESTH + REMOT)/2; HAB_HETERO = RIFF OR POOL OR (RIFF + POOL)/2; INSTR_COND = EPI_SUB OR COVER OR (EPI_SUB + COVER)/2; SUB_QUAL = EMBED (If EMBED given as percent rather than score, convert to a score from 1 - 20); RIP_ZONE = RIP_Sc (If RIP is given as measurement rather than score, convert to a score from 1 - 20). The pH values, and the conductivity concentration in mg/liter were obtained directly from the collecting agencies. A site was classified as Reference when all metrics met or exceeded the reference site classification criteria (Table 3a). When three or more habitat metrics and/or one of the water quality metrics fall below the degraded site classification criteria, the location was classified as Degraded (Table 3b). All other locations are currently classified as Other.

The classification criteria were determined using percentiles that delineate the better and poorer ends of each habitat metric's values in each bioregion. We used the Piedmont, Valley, and Highland percentiles listed in Tables 5 and 6 of Astin (2006). These percentiles were applied to the pool of Chesapeake Bay basin samples within each bioregion, and their corresponding values became the classification criteria (Tables 3a, 3b). In all cases, these classification criteria derived from the larger Chesapeake database are less than 10% different from those determined by Astin (2007, Tables 2, 3) on the smaller Potomac database.

Table 3a. Reference site classification criteria (percentiles). All criteria must be met for a site to qualify as Reference. Coastal Plain criteria are from Stribling et al. 1998 Table II-1. *, if blackwater stream, pH < 6 and DOC ≥ 8 mg/liter.

Parameter	Piedmont	Valley	Highland	Northern Appalachian	Coastal Plain
ANTHRO_ALT	≥17 (70%)	≥15 (70%)	≥16.5 (70%)	≥ 19 (70%)	Optimal/suboptimal
BANK_STAB	≥16 (70%)	≥15 (50%)	≥16 (70%)	≥ 16 (70%)	-
CHAN_ALT	≥18 (70%)	≥15 (50%)	≥16 (70%)	≥ 16 (70%)	No channelization
HAB_HETERO	>8 (10%)	>9 (10%)	>10 (10%)	> 8.5 (10%)	
INSTR_COND	>9 (10%)	>9 (10%)	>11 (10%)	> 10 (10%)	Optimal/suboptimal
RIP_ZONE	≥16 (50%)	≥14 (79%)	≥19 (70%)	≥ 17 (70%)	≥ 15 meters
SUB_QUAL	≥16 (70%)	≥15 (70%)	≥16 (70%)	≥ 16 (70%)	-
pH	6-9	6-9	6-9	-	≥ 6*
CON	<500	<500	<500	-	-
ANC	-	-	-	-	≥ 50 µeq/liter
Diss. oxygen	-	-	-	-	≥ 4 mg/liter
Nitrate-N	-	-	-	-	≤ 4.2 mg/liter
Urban	-	-	-	-	≤ 20% area of catchment
Forested	-	-	-	-	≥ 25% area of catchment
Point sources	-	-	-	-	None

Table 3b. Degraded site classification criteria. Three or more habitat metrics and/or one of the water quality metrics must meet criteria for Piedmont, Valley, Highland, and Appalachian bioregions; two or more metrics must meet the criteria for the Northern Appalachian bioregion; and any criterion must be met in Coastal Plains. Coastal Plain criteria are from Stribling et al. 1998 Table II-1.

Parameter	Piedmont	Valley	Highland	Northern Appalachian	Coastal Plain
ANTHRO_ALT	≤11.3	≤7.5	≤9	≤12	-
BANK_STAB	≤6	≤10	≤12	≤8	Poor *and* INSTR_COND poor
CHAN_ALT	≤10	≤11	≤10	≤10	Poor *and* INSTR_COND poor
HAB_HETERO	<8	<9	<10	≤8.5	
INSTR_COND	<9	<9	<11	<10	Poor *and* BANK_STAB poor
RIP_ZONE	≤2	≤2	≤4	≤5.6	-
SUB_QUAL	≤9	≤8	≤10	≤11	-
pH	<6 or >9	<6 or >9	<6 or >9	-	≤ 5 *and* ANC ≤ 0 μeq/liter
CON (mg/liter)	>1000	>1000	>1000	-	-
Diss. oxygen	-	-	-	-	≤ 2 mg/liter
Nitrate-N	-	-	-	-	≥ 7 mg/liter *and* DO ≤ 3 mg/liter
Urban	-	-	-	-	≥ 50% area of catchment *and* INSTR_COND poor

Northern Appalachian Bioregion

To classify reference and degraded habitats and develop biometric scoring criteria for the Northern Appalachian bioregion, we needed to work around several limitations in the data set (n=422). First, about one quarter of the sampling events were missing pH and/or conductivity data. Since only 8 of the remaining 301 events failed Astin's water quality criteria (pH < 6 and/or conductivity >1000 mg/liter), we decided at this stage to simply exclude water quality parameters as habitat selection criteria. Second, only 54 (12.8%) of the 422 records had data for all of the seven habitat metrics used by Astin (2006) to classify reference and degraded habitats; 81% had 5 of the 7 metrics and 6.2% had just 3 of the 7 metrics. Given the smaller number of available habitat metrics, we decided to identify Degraded sites as those meeting 2 or more of the classification criteria instead of 3 or more (Table 3b). Third, we decided for the purpose of classifying habitat condition to exclude those sampling events with just 3 habitat metric measurements. Exploration of the data showed that habitat condition scores based on just the 3 available habitat metrics (BANK_STAB, CHAN_ALT, HAB_HETERO) were poor predictors of habitat condition scores based on all 7 metrics. Finally, we excluded the 148 sampling events from the Northern Appalachian bioregion that did not have subregion classifications (see Table 2). Initial examination of all the Northern Appalachian data from ecoregions 58, 60, 62, 83 in the Chesapeake Bay basin suggested that variability in the New York portion of these ecoregions warrants dividing the Northern Appalachian bioregion into two separate bioregions. However, the current lack of subregion classes in New York did not allow for this separation. In this initial effort, we excluded the areas without subregion classifications (New York) from the biocriteria development process, and lumped areas with subregion classes (Pennsylvania) into the one bioregion, Northern Appalachian. For this analysis New York State Department of Environmental Conservation data were not available and we believe more New York water quality, habitat, and biological data need to be examined before these differences can be resolved.

A total of 266 sampling events, all from Pennsylvania, were thus used to create habitat classification criteria for the Northern Appalachian bioregion using similar methodology as was used to develop criteria for the Piedmont, Highland, and Valley bioregions. A site was classified as Reference when all metrics met or exceeded the reference site classification criteria (Table 3a). When two or more habitat metrics fall below the degraded site classification criteria, the location was classified as Degraded (Table 3b). All other locations are currently classified as Other.

<u>Coastal Plain Bioregion</u>

Maryland Department of Natural Resources (MD DNR), in conjunction with Tetra Tech, has identified metrics and selected criteria for classifying both habitat and biological condition in Coastal Plain non-tidal streams and wadeable rivers (Stribling et al. 1998). Faced with a paucity of Coastal Plain data in the Potomac River basin, Astin (2006, 2007) decided to use the MD DNR metrics and classification criteria for that region. We have adopted this approach. MD DNR's habitat metrics and criteria for the Coastal Plains bioregion are listed in Tables 3a and 3b, for the purpose of comparing them with the other bioregions. They were not independently tested in this study. MD DNR classifies a Coastal Plain site as Reference when all metrics meet or exceed the reference criteria and as Degraded when one or more of the impairment criterion is meet.

<u>Future steps</u>: A major problem in delineating Reference and Degraded habitat conditions in the Northern Appalachian bioregion is the fact that a significant number of sampling events do not have either water quality measurements or a full suite of habitat measurements. This prevents the identification of robust, clearly separated REF and DEG communities. Additional water quality, habitat and biological data, and further analysis of biological metric responses to habitat conditions, will help to better delineate REF and DEG communities in the Northern Appalachian bioregion. The bioregion may need to be subdivided if subregion differences prove significant.

For consistency, we opted in this study to use the same Coastal Plain scoring criteria that Astin (2007) used, namely the Coastal Plain scoring criteria defined in Stribling et al. (1998). These criteria have since been refined (Southerland et al. 2005). Other Chesapeake Bay basin jurisdictions have also refined their habitat metrics and scoring criteria since the Astin work was done. We will review and try to incorporate their new metrics into the Chesapeake Bay basin-wide index. All monitoring programs must be measuring the same or comparable habitat metrics in order to do this.

We are planning to develop a habitat quality index (HQI) to serve as a companion index to the B-IBI. The HQI will provide a numeric evaluation of the physical and chemical quality of the streams. If its results differ from those of the B-IBI at a station, it would indicate the impact of some unassessed stressor.

Biological Metrics Used in the B-IBI

The basin-wide index of benthic integrity (B-IBI) scores a suite of macro-invertebrate metrics with thresholds specific to each bioregion. The seven family-level metrics selected by Astin (2006, 2007) for the Highland, Valley, and Piedmont were retained for those bioregions and also applied to Northern Appalachian bioregion (Table 4a). The seven family-level metrics selected by MD DNR for the Coastal Plain were retained for that bioregion (Table 4b).

Table 4a. Family-level biological metrics used in the Highland, Valley, Piedmont, and Northern Appalachian bioregions.

Metric	Description	Measures
EPT Taxa	Number of Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) taxa	Taxonomic richness
TR	Taxa richness	Taxonomic richness
%EPT	Proportion of EPT individuals	Composition
FBI	Family-level Hilsenhoff Biotic Index: is based on the number of individuals in each tolerance class	Tolerance
%Dom1	Percent contribution of the dominant taxa	Tolerance
%CO	Proportion of collector taxa	Feeding
%CL	Proportion of clinger taxa	Habit

Table 4b. Family-level biological metrics used in the Coastal Plain bioregion.

Metric	Description	Measures
EPT Taxa	Number of Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) taxa	Taxonomic richness
TR	Taxa richness	Taxonomic richness
EphemTaxa	Number of Ephemeroptera taxa	Composition
%Ephem	Proportion of Ephemeroptera taxa	Composition
DipteraTaxa	Number of Diptera (true flies) taxa	Composition
SensitiveTaxa	Number of taxa considered intolerant	Tolerance
BecksBI	Beck’s Biotic Index: the weighted sum of intolerant taxa	Tolerance

Future steps: Several refinements would strengthen the accuracy of the Chesapeake B-IBI. Biological metrics representing a broader range of attributes in the benthic invertebrate community could possibly be derived from the merged data sets. If these additional metrics are discriminatory (sensitive to habitat conditions) they could be incorporated into the Chesapeake B-IBI. The lists of sensitive taxa used by the states differ somewhat from each other. These lists should be reviewed and brought to some consensus for each ecoregion.

Biological Metric Scoring Criteria

Scoring criteria for each metric were determined from percentiles of metric value distributions observed in each bioregion’s Reference community. Scoring criteria can be manipulated to a) favor identifying reference conditions correctly; b) favor identifying degraded conditions correctly; c) identify both conditions equally well; or d) maximize the B-IBI classification efficiency. In this initial effort to develop B-IBI's for the entire Chesapeake Bay basin, we opted to apply (a), the approach used in Astin (2007), to the Piedmont, Valley, Highlands, and Northern Appalachian bioregions. For metrics that decrease in response to stress: values > 50^{th} percentile scored 5; values between the 10^{th} and 50^{th} percentiles scored 3; and values < 10^{th} percentile scored 1. For metrics that increase in response to stress: values < 50^{th} scored 5; values between the 50^{th} and 90^{th} percentiles scored 3; and values > 90^{th} percentile scored 1. The shorthand label for this approach is "10%ile-50^{th}%ile(-90^{th}%ile)." The actual values of these percentiles for each bioregion are given in Table 5. For the Coastal Plain bioregion, we used the family level scoring criteria published in Table III-8 of Stribling et al. (1998).

The 10%ile-50^{th}%ile(-90^{th}%ile) method favors correctly identifying Reference conditions at the expense of correctly identifying Degraded conditions. In a separate analysis, we applied an approach using the 20^{th} and 80^{th} percentiles in place of the 10^{th} and 90^{th} percentile (20%ile-50^{th}%ile(-80^{th}%ile)) to see if this one modification would improve the discrimination efficiencies for Degraded sites without lowering the classification efficiencies of the B-IBI scores. Additional approaches can be tried in the future.

Metric Discrimination Efficiencies and B-IBI Classification Efficiencies

Each metric's ability to correctly identify a reference or degraded habitat is evaluated by calculating metric discrimination efficiencies (DE) for two pools of data: the Reference sites and the Degraded sites. A metric's DE for reference habitats is the number times the metric scores 3 or 5 in the pool of Reference sites, divided by the total number of Reference sites. A metric's DE for degraded habitats is the number of times the metric scores 1 in the pool of Degraded sites, divided by the total number of Degraded sites. A DE > 50% is desirable in each case. The higher the DE, the better the metric is in distinguishing the condition. The DE's for each metric are shown in Table 6a for the 10%ile-50%ile(-90%ile) method, and Table 6b for the 20%ile-50%ile(-80%ile) method.

The B-IBI score, which is the sum of the seven component metric scores, is similarly evaluated by calculating classification efficiencies (CE). CE's are shown in Tables 6a and 6b. The B-IBI's CE for reference habitats is the number of B-IBI scores ≥ 3.0 in the pool of Reference sites, divided by the total number of Reference sites. The B-IBI's CE for degraded habitats is the number of B-IBI scores < 3.0 in the pool of Degraded sites, divided by the total number of Degraded sites. The overall B-IBI classification efficiency is the average of the Reference and Degraded site CE's. The overall CE of the index was 85.4, 79.1, 86.2, and 77.7 in the Piedmont, Valley, Highland, and Northern Appalachian bioregions respectively when the 10%ile-50%ile(-90%ile) method was applied. The overall CE of the index was 86.0, 81.8, 77.9, and 84.1 in the Piedmont, Valley, Highland, and Northern Appalachian bioregions respectively when the 20%ile-50^{th}%ile(-80^{th}%ile) method was applied. As expected there was a noticeable drop in the Reference DE's and an increase in the Degraded DE's using the 20%ile-50%ile(-80%ile) method. The overall CE's went up in the Piedmont, Highland, and Northern Appalachian bioregions and went down in the Valley bioregion.

Table 5. Scoring criteria for biological metrics used in B-IBI based on application of the 10%ile-50th%ile(-90th%ile) method. Coastal Plain scoring criteria are from Stribling et al. 1998, Table III-8.

Parameter	Score	Piedmont	Valley	Highland	N. Appalachian	Coastal Plain
EPT Taxa	5	>7	>9	>12	>12	>6
	3	5-7	7-9	7.3-12	8-12	3-6
	1	<5	<7	<7.3	<8	<3
TR	5	>13	>16	>16	>19	>16
	3	10 - 13	12 - 16	12.3 - 16	13-19	8 - 16
	1	<10	<12	<12.3	<13	<8
%EPT	5	>79.1	>65.5	>75.8	>70.6	-
	3	49.8 – 79.1	65.5 -34.4	50.3 – 75.8	52.8-70.6	-
	1	<49.8	<34.4	<50.3	<52.8	-
FBI	5	<3.16	<4.07	<3.67	<3.79	-
	3	3.16 - 4.39	4.04 - 5.11	3.67 - 4.74	3.79-4.5	-
	1	>4.39	>5.11	>4.74	>4.5	-
%Dom1	5	<27.9	<29.4	<30.1	<29.1	-
	3	27.9 – 47.0	29.4 -47.6	30.1 – 52.7	29.1-43.9	-
	1	>47.0	>47.6	>52.7	>43.9	-
%CO	5	<53.5	<71.3	<52.6	<67.9	-
	3	53.5 – 77.3	71.3 – 87.3	52.6 -75.4	67.9-81.0	-
	1	>77.3	>87.3	>75.4	>81.0	-
%CL	5	>76.8	>63.2	>57.6	>60.1	-
	3	42.4 – 76.8	36.8 – 63.2	26.5 -57.6	43.4-60.1	-
	1	<42.4	<36.8	<26.5	<43.4	-
EphemTaxa	5	-	-	-	-	>2
	3	-	-	-	-	1-2
	1	-	-	-	-	<1
%Ephem	5	-	-	-	-	>11.4
	3	-	-	-	-	2.0-11.4
	1	-	-	-	-	<2.0
DipteraTaxa	5	-	-	-	-	>3
	3	-	-	-	-	2-3
	1	-	-	-	-	<2
SensitiveTaxa	5	-	-	-	-	>5
	3	-	-	-	-	2-5
	1	-	-	-	-	<2
BecksBI	5	-	-	-	-	>9
	3	-	-	-	-	4-9
	1	-	-	-	-	<4

Table 6a. Discrimination efficiencies (DE) of the biological metrics and classification efficiencies of the B-IBI, by bioregion, using the 50%ile-10%ile (50%ile-90%ile) scoring approach. Imp, pool of degraded condition sites; Ref, pool of reference condition sites.

Region	Data Pool	EPT Taxa	FBI Index	% Clingers	% Collectors	% Dom1	% EPT	Tax Rich	B-IBI
Piedmont	Imp	51.2	85.4	63.4	61.0	46.3	78.0	26.8	80.5
	Ref	91.3	89.1	89.1	89.1	90.2	89.1	90.2	90.2
Highland	Imp	69.6	65.2	59.4	52.2	56.5	63.8	68.1	75.4
	Ref	89.1	89.1	89.1	89.1	89.1	89.1	89.1	82.8
Valley	Imp	82.9	79.0	61.0	53.3	44.8	68.6	54.3	87.6
	Ref	90.6	89.4	89.4	89.4	89.4	89.4	91.8	84.7
N Appalachian	Imp	50.0	63.6	45.5	45.5	27.3	45.5	40.9	72.7
	Ref	91.3	87.0	87.0	87.0	87.0	87.0	91.3	82.6

Table 6b. Discrimination efficiencies (DE) of the biological metrics and classification efficiencies of the B-IBI, by bioregion, using the 50%ile-20%ile (50%ile-80%ile) scoring approach. Imp, pool of degraded condition sites; Ref, pool of reference condition sites.

Region	Data Pool	EPT Taxa	FBI Index	% Clingers	% Collectors	% Dom1	% EPT	Tax Rich	B-IBI
Piedmont	Imp	51.2	97.6	82.9	70.7	53.7	90.2	34.1	92.7
	Ref	91.3	79.3	79.3	79.3	79.3	79.3	82.6	79.3
Highland	Imp	78.3	68.1	76.8	59.4	82.6	69.6	71.0	87.0
	Ref	79.7	79.7	79.7	79.7	79.7	79.7	84.4	76.6
Valley	Imp	82.9	83.8	71.4	59.0	64.8	80.0	76.2	81.8
	Ref	90.6	80.0	80.0	80.0	81.2	81.2	81.2	73.9
N Appalachian	Imp	59.1	81.8	45.5	59.1	40.9	59.1	45.5	90.5
	Ref	78.3	78.3	78.3	78.3	78.3	78.3	78.3	77.6

Future steps: We opted to use the 10%ile-50%ile(-90%ile) method for this year's effort. The Non-Tidal Water Quality Workgroup may decide the 10%ile-50^{th}%ile(-90^{th}%ile) method too strongly favors the correct identification of Reference conditions at the expense of correctly identifying Degraded conditions. We encourage the Non-Tidal Water Quality Workgroup to review and discuss the alternative scoring methods.

Index Validation

One problem frequently encountered in IBI development is the lack of available data for independent validation (Seegert, 2000). In the development of this IBI, there were a limited number of reference sites in most bioregions, so data were not withheld for validation purposes. Instead, a model cross validation technique known as "jackknife with replacement" was employed to establish error estimates and better quantify performance of the bioregion B-IBIs (Snedecor & Cochran 1989). Jackknife validation is considered to produce conservative measurements of correct classification rates among the currently accepted cross validation techniques (Olden et al. 2002).

In each bioregion, the Reference data pool for each metric had one record at a time removed. The scoring criteria for that metric were recomputed and the record returned to the data set. The resulting sets of simulated metric scoring criteria were then used to rescore all available data. This process was repeated until each reference sample had been withheld. The root mean squared error, total error, and bias were calculated on the differences in scores between the original metric and simulated values for the overall IBI score and each of the IBI metrics (Table 7). A total error of less than 20% for a single metric is acceptable, considering that individual metrics and the composite index are scored on a scale from one to five and 20% is equivalent to approximately a 0.5 unit of score. Total error estimates ranged from 0 to 18.3 percent error across all bioregions. The error estimate generated by the jackknife simulation provides a measure of the accuracy of the scoring criteria.

Table 7. Jackknife validation results.

Bioregion	Metric/Index	n	Mean Difference	Uncorr. Sum of Squares	Mean Score	Total Error (%)	Bias
Piedmont	Taxa Richness	105892	0	0	3.67	0	0
	EPT Taxa	105892	0	0	3.13	0	0
	% EPT	105892	0.006081668	6536	2.48	10.03	0.245528
	FBI	105892	-0.000113323	632	2.15	3.59	-0.00527
	% Clinger	105892	-0.0042685	4504	2.83	7.29	-0.15085
	% Collector	105892	-0.015374155	9400	2.43	12.27	-0.63312
	% Dominant1	105892	0.003475239	1472	3.05	3.87	0.114058
	B-IBI	105892	-0.010199071	22968	19.73	2.36	-0.05168
Highland	Taxa Richness	56320	-0.109303977	12312	3.25	14.37	-3.3585
	EPT Taxa	56320	-0.119424716	13452	2.70	18.13	-4.4306
	% EPT	56320	-0.018288352	3868	2.92	8.98	-0.6267
	FBI	56320	-0.006214489	2148	2.92	6.70	-0.21312
	% Clinger	56320	-0.007457386	1608	3.30	5.11	-0.22567
	% Collector	56320	-0.008735795	3888	2.77	9.50	-0.31584
	% Dominant1	56320	-0.009765625	7020	3.41	10.36	-0.28646
	B-IBI	56320	-0.279190341	51832	21.26	4.51	-1.31299
Valley	Taxa Richness	110245	0	0	2.91	0	0
	EPT Taxa	110245	0	0	2.35	0	0
	% EPT	110245	-0.004970747	2072	2.77	4.94	-0.17913
	FBI	110245	-0.011574221	3344	2.53	6.87	-0.4567
	% Clinger	110245	-0.021751553	7972	3.03	8.89	-0.71877
	% Collector	110245	-0.013968887	7496	3.01	8.67	-0.4642
	% Dominant1	110245	-0.016454261	3964	3.01	6.31	-0.54735
	B-IBI	110245	-0.06871967	24904	19.61	2.42	-0.35039
Northern Appalachian	Taxa Richness	1035	0	0	3.04	0	0
	EPT Taxa	1035	-0.092753623	192	2.87	15.02	-3.23559
	% EPT	1035	-0.08115942	240	3.09	15.59	-2.62746
	FBI	1035	-0.083091787	244	2.78	17.48	-2.9913
	% Clinger	1035	-0.115942029	328	3.18	17.72	-3.64853
	% Collector	1035	-0.050241546	152	3.13	12.23	-1.60345
	% Dominant1	1035	-0.059903382	124	3.36	10.32	-1.7852
	B-IBI	1035	-0.483091787	1360	21.44	5.35	-2.25276

Narrative Condition Ratings

B-IBI scores for each sampling location in the Piedmont, Valley, Highland, and Northern Appalachian bioregions were converted to narrative condition ratings as follows: "Excellent" for B-IBI scores ≥ 50th%ile of Reference community scores; "Good" for 25th%ile to < 50th%ile of Reference scores; "Fair" for 10th%ile to < 25th%ile of Reference scores; "Poor" for ≥ 50th%ile of Degraded scores to <10th%ile of Reference scores; and "Very Poor" for < 50th%ile of Degraded scores. This approach deviates from that of Astin (2007) who used just 3 rating categories. We decided 5 categories would provide more resolution in the gradient of ratings and it would allow us to better match the rating categories to that of the individual states.

The threshold percentiles outlined above have similar, if not identical, values on the B-IBI scoring scale of 7 – 35 used in the Piedmont, Valley, Highland, and Northern Appalachian bioregions (Table 8). Variability in the 10th percentile values is due to the fact that they are derived from the tails of the Reference community B-IBI distributions. We decided for simplicity to use the average value of the scoring thresholds rather than the actual values in each individual bioregion. The rating categories for the B-IBI scores were thus determined by the values listed in the Average column of Table 8.

MD DNR developed the Coastal Plain bioregion B-IBI on a 1 – 5 scoring scale and rated the B-IBI scores according to 3 categories, not 5. To make the qualitative rating approach comparable in all 5 Chesapeake bioregions, we converted the average of the scoring thresholds used in the Piedmont, Valley, Highland, and Northern Appalachian to equivalent values on a 1-5 scale (Table 8). These equivalent values were used to rate the Coastal Plain B-IBI scores.

Table 8. The B-IBI values corresponding to the percentiles of the Reference and Degraded B-IBI scores that were identified as scoring thresholds. The Piedmont, Valley, Highland, and Northern Appalachian bioregions are scored on a 7 – 35 scale; the Coastal Plain bioregion is scored on a 1 – 5 scale (see text for details).

Rating Category	Piedmont	Valley	Highland	N. Appalachian	Average	Coastal Plain equivalent
Excellent	≥27	≥27	≥27	≥27	≥27	≥3.86
Good	23 - <27	23 - <27	23 - <27	23 - <27	23 - <27	3.29 - <3.86
Fair	21 - <23	19 - <23	17 - <23	19 - <23	19 - <23	2.71 - <3.29
Poor	15 - <21	13 - <19	13 - <17	15 - <19	14 - <19	2.00 - <2.71
Very Poor	<15	<13	<13	<15	<14	≤2.00

Percent Attainment of Passing Grade

The Chesapeake Bay Program (CBP) has adopted the approach of presenting indicators as a "percent of restoration goal achieved." We explored the possibility of evaluating the Chesapeake basin B-IBI results this way. A passing grade of "an index value greater than or equal to a value of 21 on a scale of 7 - 35" (or "3 on a scale of 1 – 5" in the Coastal Plain bioregion) was selected. This goal is directly comparable to the estuarine phytoplankton and benthic IBI restoration goals of "index values greater than or equal to

3 on a scale of 1 – 5." The logic underlying this approach is that 90% of the values of each metric in the Reference communities are scored either 3 or 5. Therefore, a minimum B-IBI score of 3, or 21 when the 7 metric scores are summed instead of averaged, should reflect attainment of a "passing grade." The HUC8 percentages can be weighted by the area of their respective HUC8's and then "rolled up" to an overall percent of restoration goal achieved for the Chesapeake basin (minus the New York portion of the basin and several underrepresented HUC8's in this case).

Results

Post -1999 data previously acquired and processed by ICPRB but not used in development of the Chesapeake Bay basin B-IBI were appended and all available data were used in an evaluation of watershed status for 2000 - 2006. For mapping purposes, the coordinates of each sampling station (in decimal degrees) were rounded to five decimals, and all sampling stations with the same coordinates were lumped into one sampling location. The classification was based on the average B-IBI score when a location was sampled more than once. Figures 2a and b depict the narrative condition ratings of the Chesapeake Bay Basin B-IBI by individual location and by HUC8. Narrative condition ratings are summarized in Table 9 for each bioregion. Figure 3 shows the percent of passing grades (i.e., 3 on a scale of 1 – 5) achieved by HUC8 for the Chesapeake Bay basin. Figure 4 depicts this basin-wide percentage based on the 2000 – 2006 data we used in this analysis. The data suggest that approximately 49% of the Chesapeake Bay basin non-tidal wadeable streams and rivers are achieving the restoration goal above.

<u>Future steps</u>: Due to time limitations, we were not able to confirm in this analysis that the cumulative frequency distributions of the B-IBI scores for random and non-random sites overlay each other in the HUC8 units and/or bioregions. It is possible the evaluations of some of the HUC8's shown in Figures 2b and 3 are biased by the non-random approaches used to select the majority of sampling sites.

Figure 2a shows that sampling events can sometimes be heavily concentrated in some areas of a HUC8 unit and sparse in other areas of the same unit due to different sampling intensity by neighboring jurisdictions. The HUC8 status and % of passing grades achieved will thus be biased by results from the heavily sampled areas. This is especially true in the Washington, DC metropolitan area. One way to deal with this imbalance might be to parse selected HUC8's into HUC11 units when the density of sampling events is high. Another approach would be to go completely over to the HUC11 scale and expect many units to show gray ("insufficient data"). (The original plan was to illustrate the analysis results by HUC11 units but the scarcity of data in many HUC11's convinced us to use HUC8 units until more data are accumulated.) The Non-Tidal Water Quality Workgroup should discuss this issue and decide if a change from the current HUC8 presentation is required.

Figure 2a. Average condition rating by location.

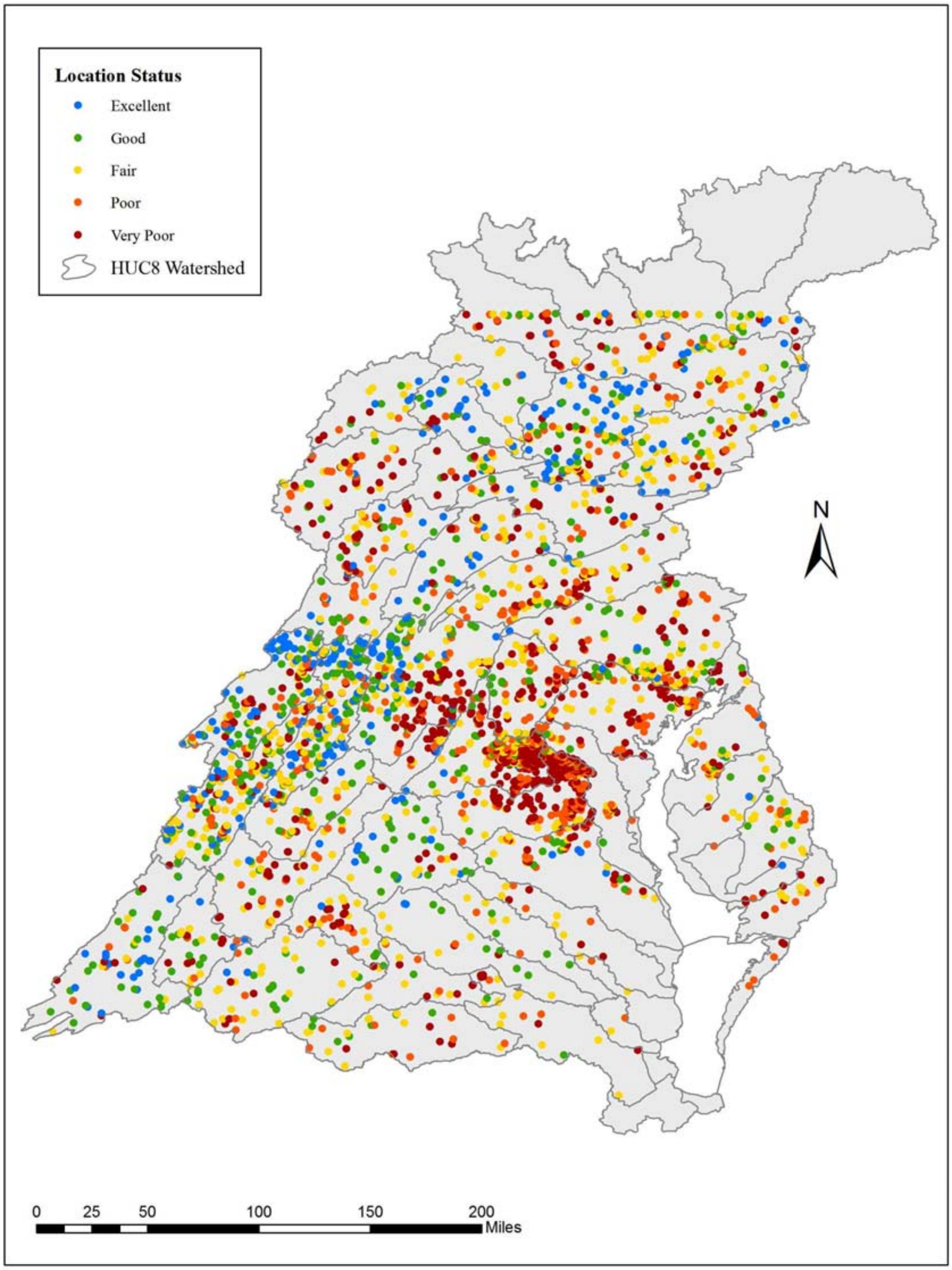

Figure 2b. Average condition rating by HUC8. Gray areas indicate a) less than 10 non-tidal data points were available, or b) a B-IBI method has not been developed yet.

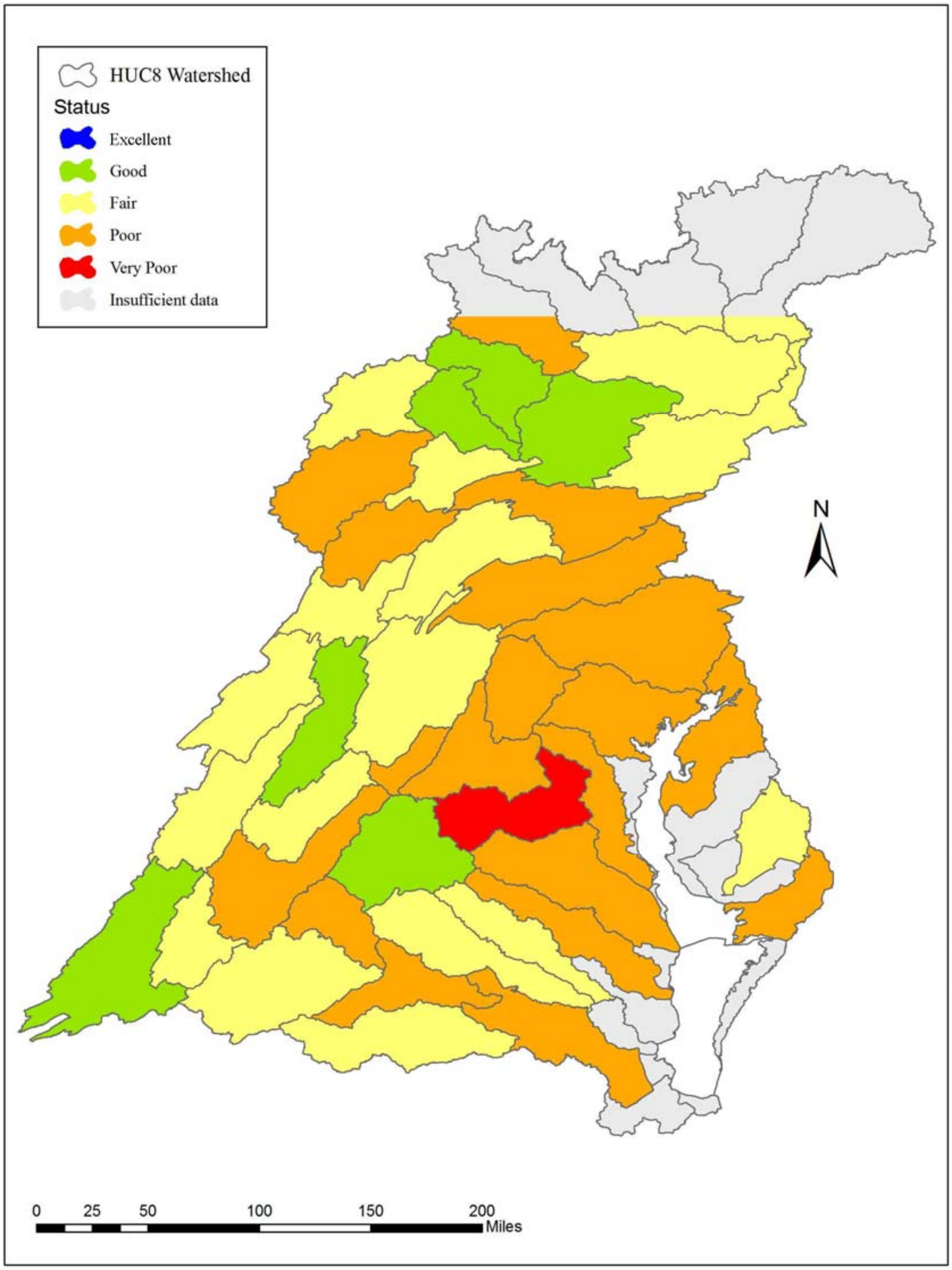

Figure 3. Percent of non-tidal, wadeable streams and rivers achieving the Chesapeake basin restoration goal, by HUC8. The restoration goal is "all index values are greater than or equal to a value of 21 on a scale of 7 - 35" (or "3 on a scale of 1 – 5" in the Coastal Plain). Gray areas indicate a) less than 10 non-tidal data points were available, or b) a B-IBI method has not been developed yet.

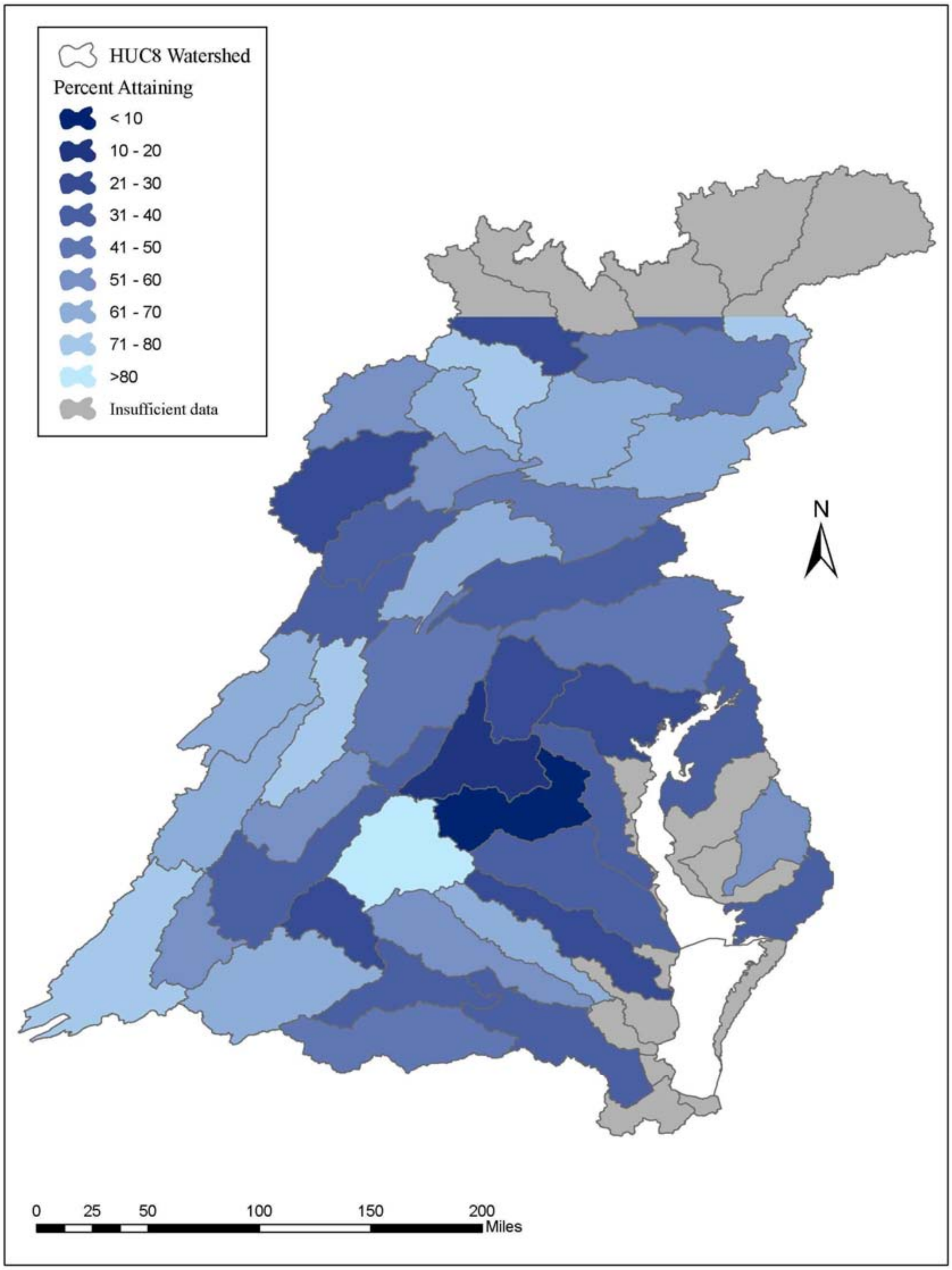

Table 9. Percent of B-IBI Condition Ratings by Bioregion for 2000 – 2006.

Bioregion	Excellent	Good	Fair	Poor	Very Poor	n
Piedmont	2.4	13.4	22.4	19.5	42.3	906
Valley	18.6	17.7	24.8	15.8	23.1	860
Highland	18.4	24.7	26.5	14.8	15.5	683
N. Appalachian	21.1	18.6	29.5	14.4	16.5	285
Coastal Plain	2.1	10.3	24.8	36.2	26.6	428

Biological Condition Gradient

The EPA's Office of Water, Office of Science and Technology, has developed the Biological Condition Gradient (BCG) conceptual model to evaluate biological structure and function in streams relative to a baseline condition of "minimally disturbed" or "as naturally occurs" (US EPA 2005, Davies and Jackson 2006). The BCG model posits that ecological systems follow a predictable trajectory in response to generally increasing anthropogenic stress (Figure 5). It provides "a common framework for interpreting biological information regardless of methodology and geography. When calibrated to a regional or state scale, States and Tribes can use the model to more precisely evaluate the current and potential biological condition of their waters and use that information to better define their aquatic life uses" (US EPA 2005).

The Chesapeake basin-wide B-IBI and habitat scores appear to corroborate the BCG model's trajectory (Figure 6) despite differences in the number and type of metrics used. The Chesapeake basin-wide B-IBI (representing the BCG's biological condition) decreases as the habitat index score (representing BCG's generalized stressor gradient) decreases. The BCG biological condition consists of 10 biological "attributes" whereas the Chesapeake B-IBI is presently comprised of 7 biological metrics (Tables 4a, 4b). The BCG stressor gradient is composed of metrics from 6 stressor categories (flow regime, habitat structure, water quality, toxics & bioengineered chemicals, energy source, biotic interactions). The Chesapeake bioregion-specific habitat indexes presently depend on 5 or more metrics of physical structure and water quality (Tables 3a, 3b). In spite of these differences, the general descriptions for Levels 2-6 biological condition (Figure 5) correspond roughly to biological communities in the habitat scores groupings >17-20 to < 8, respectively (Figure 6).

Future steps: The Chesapeake basin-wide B-IBI index relies on federal, state, county, and volunteer monitoring programs to measure a common set of biological metrics with comparable collection methods. The current suite of 7 metrics comprising the B-IBI could potentially be expanded to include metrics representing more of the 10 BCG "attributes" if additional common metrics are identified. A similar examination of the habitat and water quality data sets could potentially increase the number and type of metrics presently included in the habitat index. Incorporating metrics of actual (gauged) or modeled flow regime in the habitat index might make the B-IBI's sensitive to flow alterations due to surface and groundwater withdrawals.

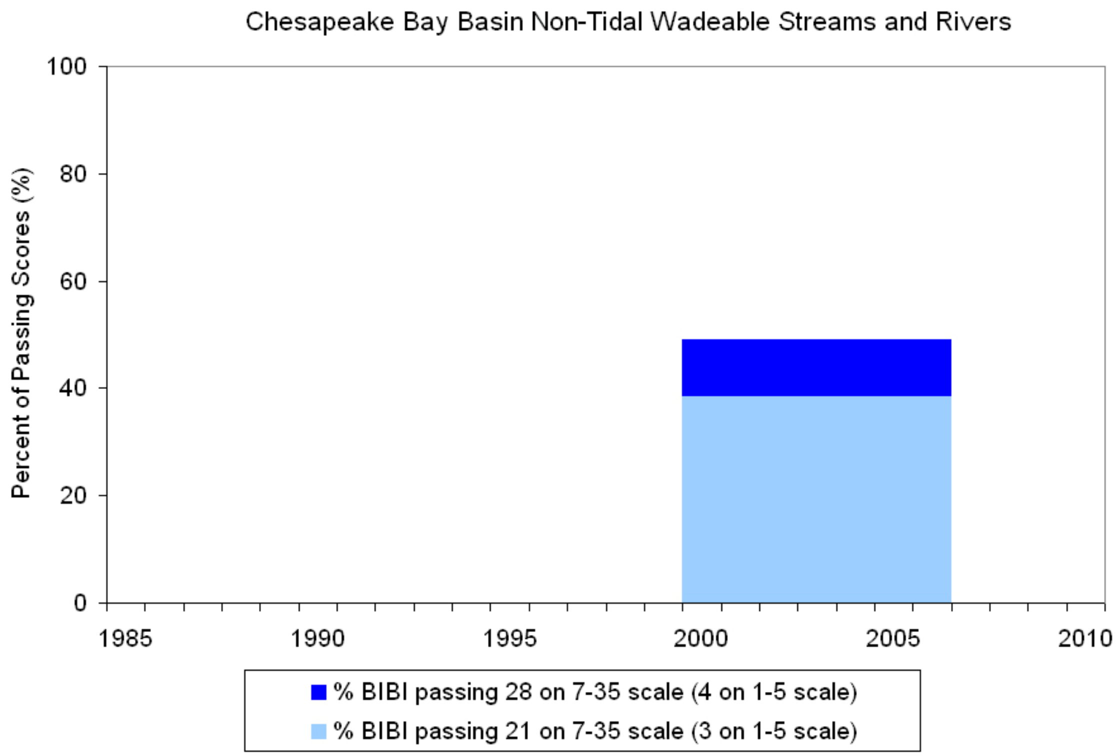

Figure 4. Percent of non-tidal, wadeable streams and rivers achieving a passing grade. Percentages for the individual, evaluated HUC8's are area-weighted and summed to obtain the basin percentage shown in the bar graph.

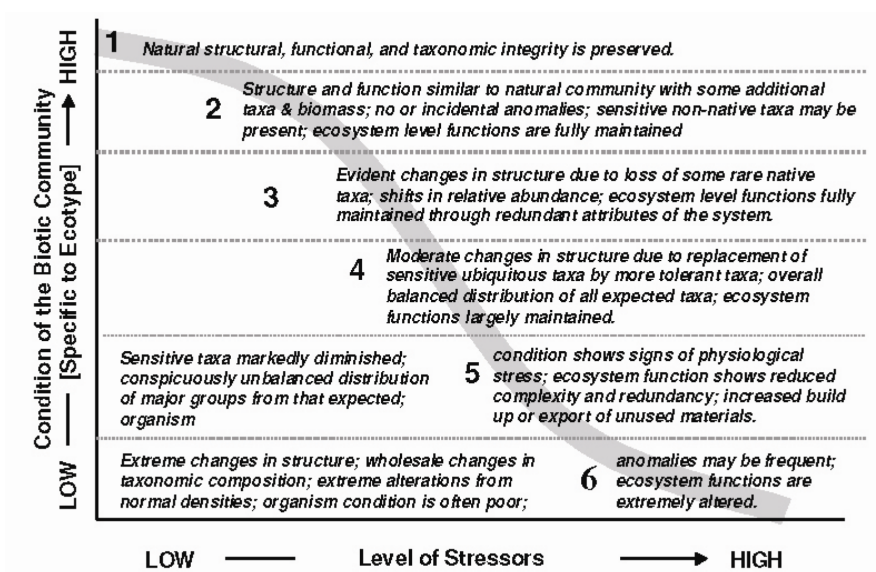

Figure 5. Conceptual model of the Biological Condition Gradient (BCG).

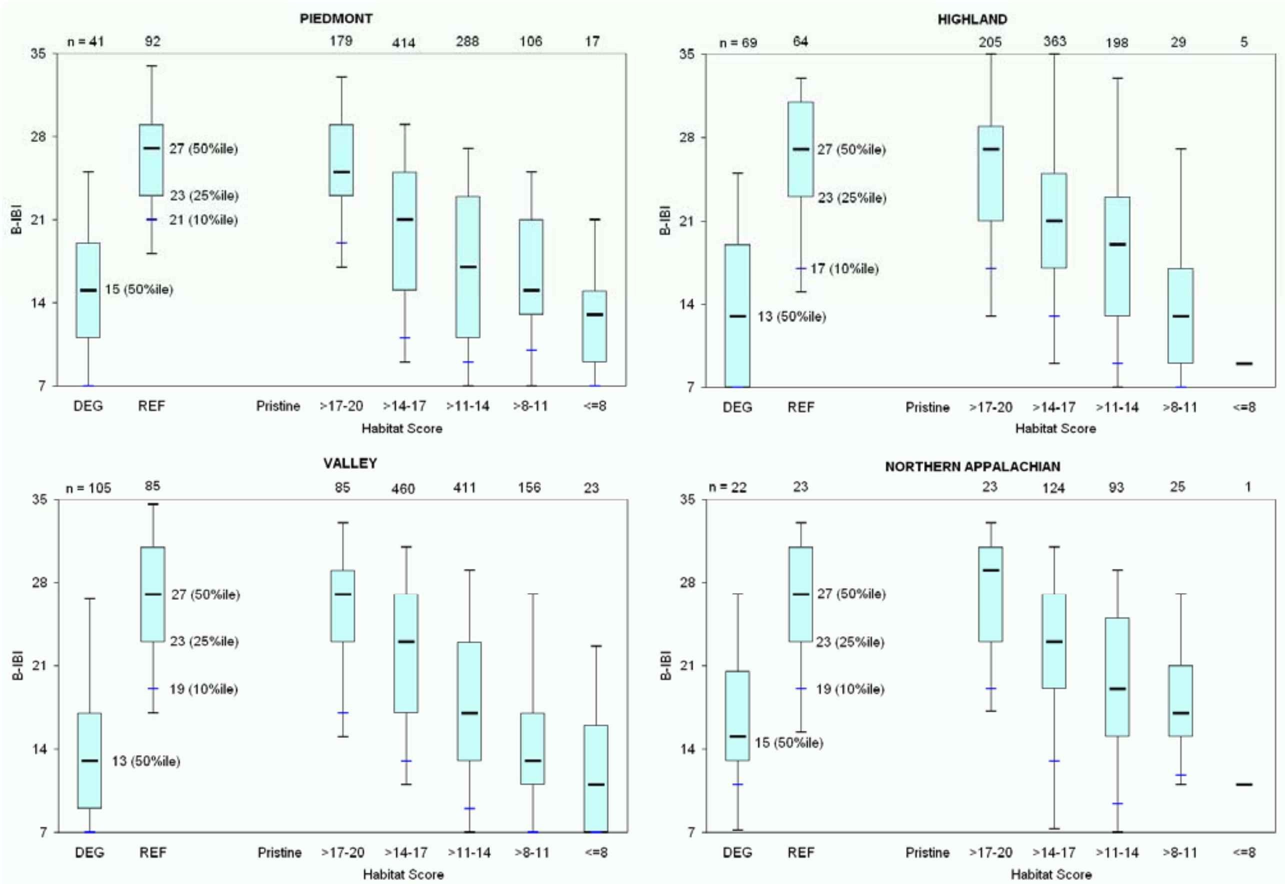

Figure 6. Distributions of B-IBI scores in the Degraded (DEG) habitat sites and Reference (REF) habitat sites, and in five habitat classes that are based on average habitat scores (see text for details). The box-and-whiskers indicate the 95th, 75th, 50th, 25th, 10th, and 5th percentiles. Habitat scores can range from 0 – 20; B-IBI scores can range from 7 – 35.

Comparisons with Individual State Assessments

The Chesapeake basin-wide B-IBI assessment generally agrees with the basin states' assessment derived from the same pool of data as shown in Table 10. The Chesapeake basin-wide B-IBI agrees most often with the state's ratings in the "best" and "worst" categories with 75%, 71%, 83%, and 84% agreement in the "best" categories and 81%, 100%, 83%, and 75% agreement in the "worst" categories for Maryland, West Virginia, Pennsylvania, and Virginia, respectively. There were not enough data points to do this comparison with Delaware's data. There is variability in how comparable the Chesapeake basin-wide B-IBI and states categories are between the two "worst" and "best" extreme rankings with percentages ranging from 16% in Maryland to 47% agreement in Virginia. Only a few locations (3% in Maryland, 6% in Pennsylvania, and 1% in Virginia) were rated as good by the B-IBI where the states considered benthic conditions as poor. Some locations were rated as poor by the Chesapeake basin-wide B-IBI where states rated them as good, however in all cases the percent mismatch was always less than 10%.

Table 10. Comparison of Chesapeake basin-wide B-IBI and State Narrative Condition Ratings for All Seasons, 2000 – 2006

WV DEP	B-IBI Condition Rating (%) Very Poor	Poor	Fair	Excellent/ Good	n
Unimpaired	0	4	25	71	124
Grey zone	0	42	32	26	19
Impaired	65	32	3	0	37
Severely Impaired	100	0	0	0	1

MD DNR	B-IBI Condition Rating (%) Very Poor	Poor	Fair	Excellent/ Good	n
Good/Un-impaired	2	8	15	75	61
Fair	20	38	27	16	45
Poor	80	16	0	4	55
Very Poor	81	16	0	3	31

PA DEP	B-IBI Condition Rating (%) Very Poor/ Poor	Fair	Good	Excellent	n
Special protection	0	4	12	83	89
Attaining/Un-impaired	4	10	21	65	188
Grey zone	22	38	31	9	68
Impaired	83	11	5	1	94

VA DEQ	B-IBI Condition Rating (%) Very Poor	Poor	Fair	Good	Excellent	n
Excellent	0	0	1	15	84	279
Good	0	1	14	47	38	479
Fair	2	17	45	32	5	260
Moderately Impaired	23	44	27	5	0	313
Severely Impaired	75	20	5	1	0	376

Regardless of the rating discrepancies, the standardized procedures used to derive Chesapeake basin-wide index of benthic integrity make it more useful than the various state-specific indexes for a Chesapeake basin-wide assessment of stream health and can lead to a better understanding of benthic and habitat issues in the basin.

Acknowledgements

The Cheaspeake basin-wide index of benthic integrity could not have been developed without the expertise, guidance, and contributions from the following individuals: Jackie Johnson (ICPRB) and LeAnne Astin (Fairfax County)

Literature Cited

Astin, L.E. 2006. *Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River basin, USA.* Ecological Indicators 6: 664-685.

Astin, L. E. 2007. *Developing biological indicators from diverse data: The Potomac Basin-wide Index of Benthic Integrity (B-IBI).* Ecological Indicators 7: 895-908.

Davies, S. P., and S. K. Jackson. 2006. The Biological Condition Gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251-1266.

Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2002. *Predictive models of fish species distributions: A note on proper validation and chance predictions*. Transactions of the American Fisheries Society 131:329–336.

Seegert,G. (2000). The development, use and misuse of biocriteria with an emphasis on the index of biotic integrity. Environmental Science and Policy, 3, S51-S58.

Snedecor,G.W. and W. G. Cochran. 1989. Statistical Methods, 8th edition. The Iowa State University Press, Ames, Iowa.

Stribling J.B., B.K. Jessup, J.S. White, D. Boward, and M.Hurd. 1998. *Development of a Benthic Index of Biotic Integrity for Maryland Streams.* Prepared by Tetra Tech, Inc., Owings Mills, MD and Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Program. CBWP-MANTA-EA-98-3.

US Environmental Protection Agency. 2005. *Use of Biological Information to Better Define Designated Aquatic Life Uses in State and Tribal WQS: Tiered Aquatic Life Uses.* EPA-822-R-05-001.