VERSAR, INC.*

Ecological Sciences and Applications
9200 Rumsey Road, Columbia, Maryland 21045

OLD DOMINION UNIVERSITY

Department of Biological Sciences
Old Dominion University, Norfolk, Virginia 23529

CHESAPEAKE BAY B-IBI RECALIBRATION

Prepared by

Principal Investigators:

Roberto J. Llansó*
Daniel M. Dauer
Michael F. Lane

Submitted to:

Cindy S. Johnson
Chesapeake Bay Monitoring Manager
Virginia Department of Environmental Quality
629 East Main Street
Richmond, Virginia 23219

August 2016

1.0 BACKGROUND AND OBJECTIVES

The Chesapeake Bay Benthic Community Restoration Goals (Ranasinghe et al. 1994) and the Benthic Index of Biotic Integrity (B-IBI) by which goal attainment is measured (Weisberg et al. 1997), are today standard tools in management decision making. The B-IBI is used to assess and monitor trends in Chesapeake Bay health, report condition for impaired waters assessments under the Clean Water Act (305b reports), support ambient water quality criteria development and assessment, and characterize benthic condition in tributary basins to assist in setting restoration goals. The B-IBI was last validated for tidal freshwater and oligohaline habitats by Alden et al. in 2002. However, the paucity of data available at that time made the index less robust in the tidal freshwater and oligohaline regions than in the more saline habitats of the Bay. In addition, some performance issues have been identified throughout the years, such as: (1) correct classification efficiencies for the B-IBI seem to be lower than those of the initial calibration effort for some regions of the Bay, (2) differences in pollution indicative and sensitive taxa metrics have been identified among the different salinity habitats which may affect index performance, and (3) high biomass is a desirable property in low salinity regions but excess biomass in the B-IBI is scored as degraded and thresholds may need modification. Large datasets that were unavailable to Weisberg et al. can be used today to test the performance of the B-IBI for the various salinity habitats, and used to recalibrate the B-IBI.

In this study we used the data available to Weisberg et al. (1997) and new data assembled from multiple sources and programs that were conducted in Chesapeake Bay from 1994 to the present. The aim of the study was to re-evaluate the metric thresholds. Classification efficiencies of samples classified a priori by biological, physical, and contaminant data were computed on the original Weisberg et al. thresholds and new thresholds. In addition, the scoring procedure for the biomass metric was re-evaluated, from a current scoring system (1,3,5,3,1) in which low biomass values (below the lower restorative threshold) and high biomass values (above the upper restorative threshold) are considered degraded, to a modified scoring system (1,3,5) in which only low biomass values are considered degraded. The study considered single replicate and means of replicate data, and post-1997 data separately because during the course of the project it became apparent that benthic conditions in Chesapeake Bay had changed from conditions on which the original calibration effort was based. Validation assessments were conducted for the following threshold iterations:

1. Original thresholds
2. New thresholds based on data available to Weisberg et al. and new data
3. New thresholds as above and modified biomass procedure
4. New thresholds based on means of replicate samples
5. New thresholds based on means of replicate samples and modified biomass procedure
6. New threshold based on post-1997 data

Weisberg et al. (1997) classification efficiencies were in the 80-90% range, but classification efficiencies on new data were somehow lower (70-75% range) in a subsequent study (Llansó et al. 2009). The present study addresses the question of whether adjusting thresholds using a larger dataset than that available to Weisberg et al. (iterations 2-6 above) produce better classification efficiencies than the baseline (iteration 1). The results of these validation assessments will be taken as the basis for accepting or rejecting the new thresholds.

2.0 DATA ASSEMBLAGE

2.1 Source

The datasets in this study were assembled from multiple sources (Table 1) and in a variety of formats either from: (1) data files downloaded from Internet websites maintained by the collecting agencies; (2) archived databases maintained by the participants of this study; (3) direct delivery via email transfer from the collecting agencies; and (4) data entry/cut and paste from electronic or hard copy versions of project final reports (Table 1). All samples met a set of selection criteria based on series of limitations that excluded observations based on geographic location, season of collection, and compatibility of sample processing.

All data selected were located strictly within the latitude and longitude boundaries of Chesapeake Bay and its contiguous tidal tributaries and were collected within the B-IBI index period (Weisberg et al., 1997). This period typically extends from July 1st through September 30th in any given year; however, additional samples from the first two weeks of October were included in this study to allow for samples collected later in the season due to storm events or other issues. With the exception of Virginia's National Coastal Condition Assessment (NCCA) data, all samples were collected using a 440-cm^2 surface sampling area Young grab. Virginia's NCCA data were collected using two ponar grabs per sample for a total sample area of 495 cm^2. All samples were sieved through a 0.5-mm mesh screen, and the organisms identified to the lowest possible taxonomic level. A comparison study between the Young grab and double ponar grab sampling approach for the Virginia NCCA samples indicated no significant differences in B-IBI metrics or benthic community dominant species at multiple stations in multiple habitat types, indicating that these data were compatible for combined analyses in this study (Dauer and Lane, 2005). Finally, no data were deemed acceptable for inclusion into the database unless they were accompanied by bottom salinity and dissolved oxygen measurements, estimates of the percentage of sediment silt-clay, sediment metal and contaminant concentrations, and 10 or 20-day endpoint amphipod toxicity test results for either *Ampelisca abdita*, *Leptocheirus plumulosus* or *Hyallela azteca*. This process resulted in a final dataset comprised of 1,831 samples (including replicates) collected at 1,051 separate sampling events throughout the length of the Chesapeake Bay tidal watershed.

Several websites in addition to those listed in Table 1 provided useful assistance in the construction of the final dataset. National Institute of Standards and Technology's website for CAS number searches was an extremely helpful tool for assisting with the standardization of chemical variable names and for help with using CAS numbers to identify chemical parameter names (and vice-versa) that were absent from data sets (see http://webbook.nist.gov/chemistry/cas-ser.html). The Integrated Taxonomic Information System website and the World Register of Marine Species were also helpful with resolving taxonomic issues (see http://www.itis.gov/ and http://www.marinespecies.org/about.php,respectively). Verification of station locations was made by visual inspection of maps created using freeware available at HamsterMap.com http://www.hamstermap.com/custommap.html.

2.2 Reference Site Selection

Prior to the calculation of new thresholds all sites were divided into two *a priori* stress categories, i.e. Degraded and Reference (non-degraded). Table 2 summarizes the Reference selection criteria for this study. All Reference site criteria needed to be met before a site could be classified as Reference while violation of only one of the criteria resulted in a site being classified as Degraded. If dissolved oxygen concentrations were greater than 3.0 ppm, no chemical contaminant concentration exceeded Long et al.'s (1995) effects range-median concentrations, no more than three chemical contaminants exceeded Long et al.'s (1995) effects range-low concentrations, the ERM quotient as defined by Hyland et al. (2003) did not exceed a value of 0.0440, and sediments were not toxic based on the amphipod toxicity test, sites were classified as Reference. Additionally samples with less than three species were classified as depauperate and therefore as being degraded under the assumption that some minimum number of species would be expected in reference conditions. These criteria were similar to those of previous studies (Weisberg et al. 1997; Van Dolah et al. 1999; Llansó, et al., 2002) but derived primarily from those of Llansó et al. (2009) with some modifications. Previous studies have included samples with toxicity tests conducted with *Ampelisca abdita*; however, this study has included many samples with survival endpoints for different species, specifically *Leptocheirus plumulosus* and *Hyallela azteca*.

Two thirds of the Reference dataset was randomly selected for the computation of new thresholds and scoring of metric and B-IBI values. This became the Calibration dataset. One third was reserved to conduct sensitivity and reliability tests, i.e., efficiencies based on a priori site impact classifications. This became the Validation dataset. The baseline, i.e., classification efficiencies based on the Weisberg et al. (1997) and Alden et al. (2002) thresholds, was conducted on the entire dataset, using both the Reference and Validation datasets.

Table 1. List of data sources and number of observations. An asterisk indicates that the probability-based monitoring program samples listed were combined with sediment chemistry data and sediment toxicity data that were collected separately as part of ambient sediment toxicity assessments for the Chesapeake Bay Program Ambient Toxicity Program.

Project	**Time Period**	**Number of Samples**	**Source of Biological and Dissolved Oxygen Data**	**Source of Chemistry and Toxicity Data**
Environmental Monitoring and Assessment Program Virginian Province Data (EMAP)	1990-1993	738	https://archive.epa.gov/emap/archive-emap/web/html/geographic.html	Same as Biological
Mid-Atlantic Integrated Assessment (MAIA)	1997-1998	370	Versar Inc.	Same as Biological
Chesapeake Bay Program Ambient Toxicity Program (AMTOX)	1997-2003	104	Versar Inc. and Old Dominion University Long-Term Databases	Data
National Oceanic and Atmospheric Administration National Status & Trends Program (NOAA NS&T)	1998-2001	191	https://products.coastalscience.noaa.gov/collections/ltmonitoring/nsandt/	Same as Biological
Maryland Chesapeake Bay Probability-Based Monitoring Program (MDRBP)*	1997-2010	55	www.chesapeakebay.net	AMTOX Reports
Virginia Chesapeake Bay Probability-Based Monitoring Program (VARBP)*	1997-2005	36	www.chesapeakebay.net	AMTOX Reports
National Coastal Condition Assessment (NCCA)	2005-2014	337	Donald Smith, Virginia Department of Environment Quality	Same as Biological
	Total	1831		

Table 2. Degraded and Reference site classification criteria based on number of species collected, dissolved oxygen, sediment chemistry, and sediment toxicity.

Criteria	**Degraded**	**Reference**
Number of Species Collected	≤3	>3
Bottom Dissolved Oxygen (psu)	≤2	>3
Effects Range Median Exceedances	Any	None
Effects Range Low Exceedances	>10	<3
Toxicity	<80% and significant difference from control	Not toxic
ERM Quotient	>0.044 (High and Very High Benthic Risk Level)	≤0.044 (Low to Medium Benthic Risk Level)

3.0 THRESHOLDS

3.1 Original Thresholds

Thresholds published in Weisberg et al. (1997) and Alden et al. (2002) were entered in the project database (Table 3) and used to score metrics and the B-IBI using current B-IBI protocols whenever data sources did not contain these data, or where the computations in these data sources were old (EMAP, MAIA) and did not employ the latest B-IBI methods.

Table 3. B-IBI thresholds derived by Weisberg et al. (1997) and further updated by Alden et al. (2002). Metrics: Shan = Shannon index, Abun = Abundance (#/m^2), Bmas = Biomass (g AFDW/m^2), OPA4 = Abundance of pollution indicative taxa (%), EQA4 = Abundance of pollution sensitive taxa (%), OPBM = Biomass of pollution indicative taxa (%), EQBM = Biomass of pollution sensitive taxa (%), CAAB = Abundance of carnivore and omnivores (%), DDAB = Abundance of deep-deposit feeders (%), OPA8 = Abundance of pollution indicative freshwater taxa (%), OPA = Abundance of pollution indicative oligohaline taxa (%), EQA8 = Abundance of pollution sensitive oligohaline taxa (%), SCOR = Tolerance Score, PCR = Tanypodinae to Chironomidae abundance ratio (%). Numbers after metric name indicate percentile threshold, 5^{th} to 95^{th}.

HABITAT	SHAN_05	SHAN_50	ABUN_05	ABUN_25	ABUN_75	ABUN_95	BMAS_05	BMAS_25	BMAS_75	BMAS_95
Tidal Freshwater			800	1,050	4,000	5,500				
Oligohaline			180	450	3,350	4,050				
Low Mesohaline	1.7	2.5	500	1,500	2,500	6,000	1	5	10	30
High Mesohaline Sand	2.5	3.2	1,000	1,500	3,000	5,000	1	3	15	50
High Mesohaline Mud	2	3	1,000	1,500	2,500	5,000	0.5	2	10	50
Polyhaline Sand	2.7	3.5	1,500	3,000	5,000	8,000	1	5	20	50
Polyhaline Mud	2.4	3.3	1,000	1,500	3,000	8,000	0.5	3	10	30

	OPA4_50	OPA4_95	EQA4_05	EQA4_50	OPBM_50	OPBM_95	EQBM_05	EQBM_50	CAAB_05	CAAB_50
Tidal Freshwater										
Oligohaline									15	35
Low Mesohaline	10	20	5	25			40	80		
High Mesohaline Sand	10	25	10	40					20	35
High Mesohaline Mud	20	50	10	30	5	30	30	60	10	25
Polyhaline Sand	10	40	25	50	5	15				
Polyhaline Mud	15	50	25	40	5	20	30	60	25	40

	DDAB_05	DDAB_50	DDAB_95	OPA8_50	OPA8_95	OPA_50	OPA_95	EQA8_05	EQA8_50	
Tidal Freshwater		70	95	39	87					
Oligohaline						27	95	0.2	26	
Low Mesohaline										
High Mesohaline Sand										
High Mesohaline Mud										
Polyhaline Sand	10	25								
Polyhaline Mud										

	SCOR_50	SCOR_95	PCR_05	PCR_50
Tidal Freshwater	8	9.35		
Oligohaline	6	9.05	64	17
Low Mesohaline				
High Mesohaline Sand				
High Mesohaline Mud				
Polyhaline Sand				
Polyhaline Mud				

3.2 New Thresholds

New thresholds were calculated for each metric using the Calibration dataset (Table 4). This dataset included data available to Weisberg et al. (EMAP data) as well as the new data specified in the data assemblage section of this report. Other threshold iterations included thresholds based on means of replicate samples, and thresholds based on post-1997 data, i.e., separating the older data (EMAP, MAIA) from the most current data (Ambient Toxicity, probability-based monitoring, NOAA NS&T, and NCCA).

Table 4. New thresholds derived with data assembled for this project, including data available to Weisberg et al. (EMAP data) and new data. See Table 3 for metric names and numbers after metric names.

HABITAT	SHAN_05	SHAN_50	ABUN_05	ABUN_25	ABUN_75	ABUN_95	BMAS_05	BMAS_25	BMAS_75	BMAS_95
Tidal Freshwater			1,409	2,864	6,773	9,817				
Oligohaline			432	1,318	3,977	16,318				
Low Mesohaline	1.5	2.4	750	1,886	3,682	11,932	0.128	0.445	1.6	6.2
High Mesohaline Sand	1.5	2.7	566	1,307	3,352	9,455	0.101	0.386	1.6	8.6
High Mesohaline Mud	1.6	2.7	523	1,068	2,318	5,999	0.143	0.303	0.909	1.8
Polyhaline Sand	1.4	3.2	909	1,778	4,932	9,591	0.119	0.505	5.1	14.9
Polyhaline Mud	1.6	3.0	682	1,776	6,175	9,636	0.202	0.524	2.3	33.7

	OPA4_50	OPA4_95	EQA4_05	EQA4_50	OPBM_50	OPBM_95	EQBM_05	EQBM_50	CAAB_05	CAAB_50
Tidal Freshwater										
Oligohaline									0.0	26.3
Low Mesohaline	5.5	71.7	0.94	18.3			4.4	26.8		
High Mesohaline Sand	16.3	75.8	0.72	22.3					3.4	23.2
High Mesohaline Mud	21.9	68.4	2.0	19.5	27.3	79.0	0.52	7.5	4.5	18.2
Polyhaline Sand	6.3	35.3	3.9	53.4	4.8	46.3				
Polyhaline Mud	19.9	73.1	5.7	33.8	16.5	62.0	0.57	20.6	2.8	29.7

	DDAB_05	DDAB_50	DDAB_95	OPA8_50	OPA8_95	OPA_50	OPA_95	EQA8_05	EQA8_50	
Tidal Freshwater		71.0	95.1	39	87					
Oligohaline						15.8	93.8	0	2.3	
Low Mesohaline										
High Mesohaline Sand										
High Mesohaline Mud										
Polyhaline Sand	0.16	20.3								
Polyhaline Mud										

	SCOR_50	SCOR_95	PCR_05	PCR_50
Tidal Freshwater	8.7	9.7		
Oligohaline	7.3	9.6	0	0
Low Mesohaline				
High Mesohaline Sand				
High Mesohaline Mud				
Polyhaline Sand				
Polyhaline Mud				

3.3 Comparison Among Thresholds

New thresholds derived with the reference dataset assembled for this project were lower than the original Weisberg et al. (1997) and Alden et al. (2000) thresholds for metrics for which low numbers indicate degraded conditions, and this difference was larger for the lower, 5th percentile threshold. These metrics include Shannon index (Figure 1), abundance and biomass of pollution sensitive taxa (Figures 5 and 7), abundance of carnivore and omnivores (Figure 8), abundance of deep-deposit feeders (Figure 9, but see below), and abundance of pollution sensitive oligohaline taxa (Figure 10).

For metrics for which high numbers indicate degraded conditions, the new thresholds were higher than the original thresholds (Figures 4, 6, and 11), except for abundance of pollution indicative taxa in the Polyhaline Sand habitat. For pollution indicative taxa, this difference was larger for the upper, 95th percentile threshold (Figures 4 and 6).

For abundance, for which low numbers and high numbers indicate degraded conditions, the new 5th percentile threshold was lower than the original 5th percentile threshold, and the new 95th percentile threshold was higher than the original 95th percentile threshold (Figure 2). This was true for the high salinity habitats, but for the low salinity habitats (Tidal Freshwater, Oligohaline, and Low Mesohaline), the new 5th percentile threshold was *higher* than the original threshold (Figure 2).

For biomass, for which low numbers and high numbers also indicate degraded conditions in the current B-IBI, the new 5th percentile threshold was lower than the original 5th percentile threshold; however, the new 95th percentile threshold was much lower (not higher, as with abundance) than the original 95th percentile threshold (Figure 3).

Deep-deposit feeder abundance is defined differently in the Polyhaline Sand habitat than in the Tidal Freshwater habitat. In the Polyhaline Sand habitat low numbers of deep-deposit feeders indicate degraded conditions whereas in the Tidal Freshwater, high numbers of deep-deposit feeders indicate degraded conditions. In the Tidal Freshwater habitat there was no difference between the new and the original thresholds for deep-deposit feeders (Figure 9). Also, in the same habitat there was little difference between the new 95th percentile threshold and the original 95th percentile threshold for pollution indicative taxa (Figure 10).

The above results can be interpreted as follows:

1. Lowered thresholds relative to those of Weisberg et al.'s effort indicate lower metric values in recent samples. Conversely, for metrics for which high numbers indicate degraded conditions, higher thresholds indicate higher metric values in recent samples.
2. Differences between the new and the original thresholds are larger at the 5th and 95th percentile thresholds than at the 25th, 50th, or 75th percentile thresholds, indicating increased depauperate conditions in Chesapeake Bay.
3. As thresholds are lowered (5th) or raised (95th), the number of samples in the validation dataset that score "1" for degraded conditions *decrease*, therefore increasing the B-IBI and giving the false impression that conditions in Chesapeake Bay have improved should these thresholds be adopted.

4. High biomass values (above restorative thresholds) have traditionally been viewed as indicating degraded conditions. However, lower values in recent samples for all biomass samples suggest that this concept needs revision.
5. The percentage of pollution tolerant organisms in the Tidal Freshwater (tubificid oligochaetes and many insect larvae) has not changed substantially in more recent samples, suggesting that conditions in this habitat have not changed.

As shown in the next section, classification efficiencies of the B-IBI using the new thresholds did not improve over the baseline or current condition using the Weisberg et al. (1997) and Alden et al. (2002) thresholds.

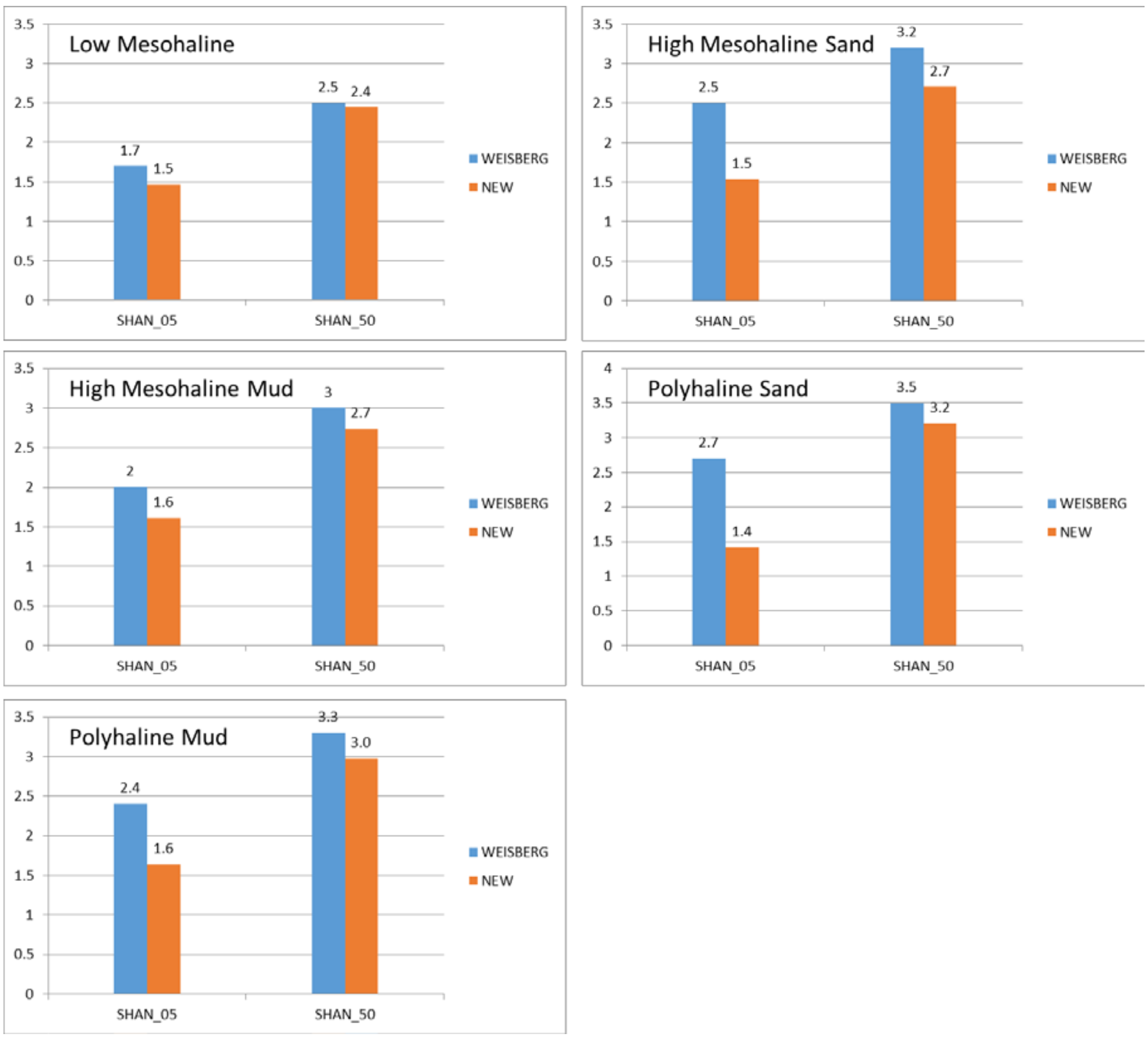

Figure 1. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for Shannon index (H')

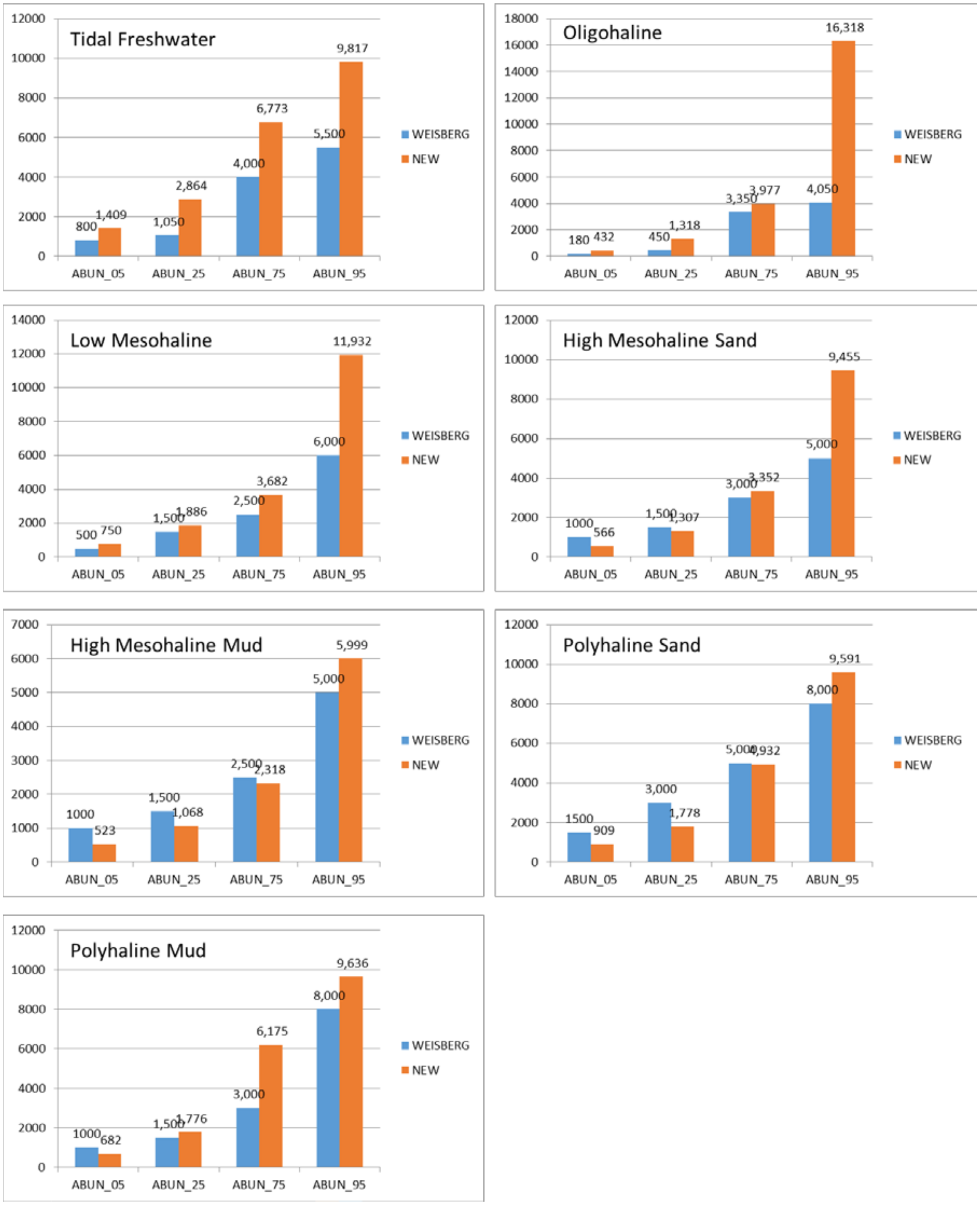

Figure 2. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance (#/m^2)

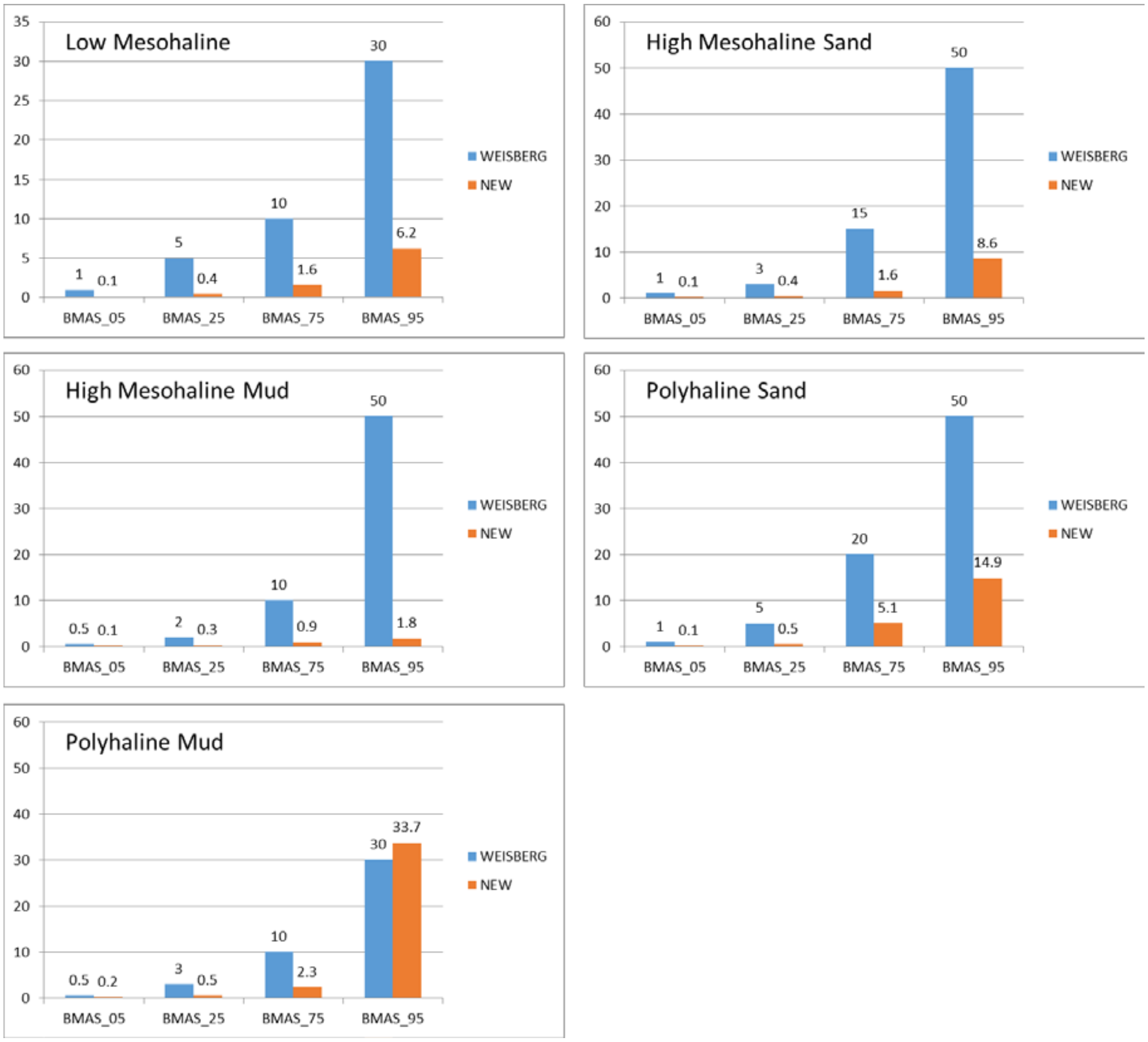

Figure 3. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for biomass (g AFDW/m^2)

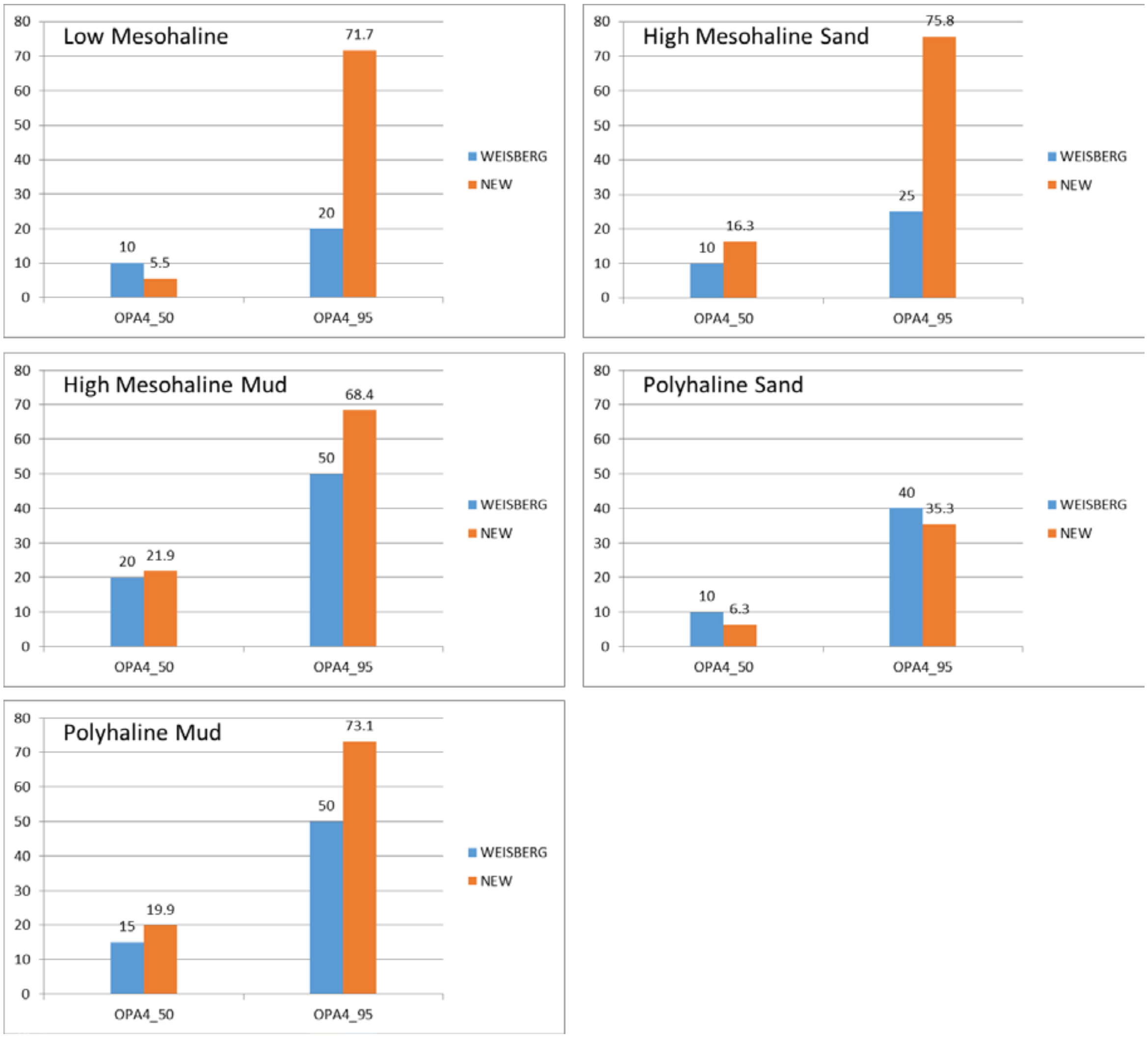

Figure 4. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance of pollution indicative taxa (%)

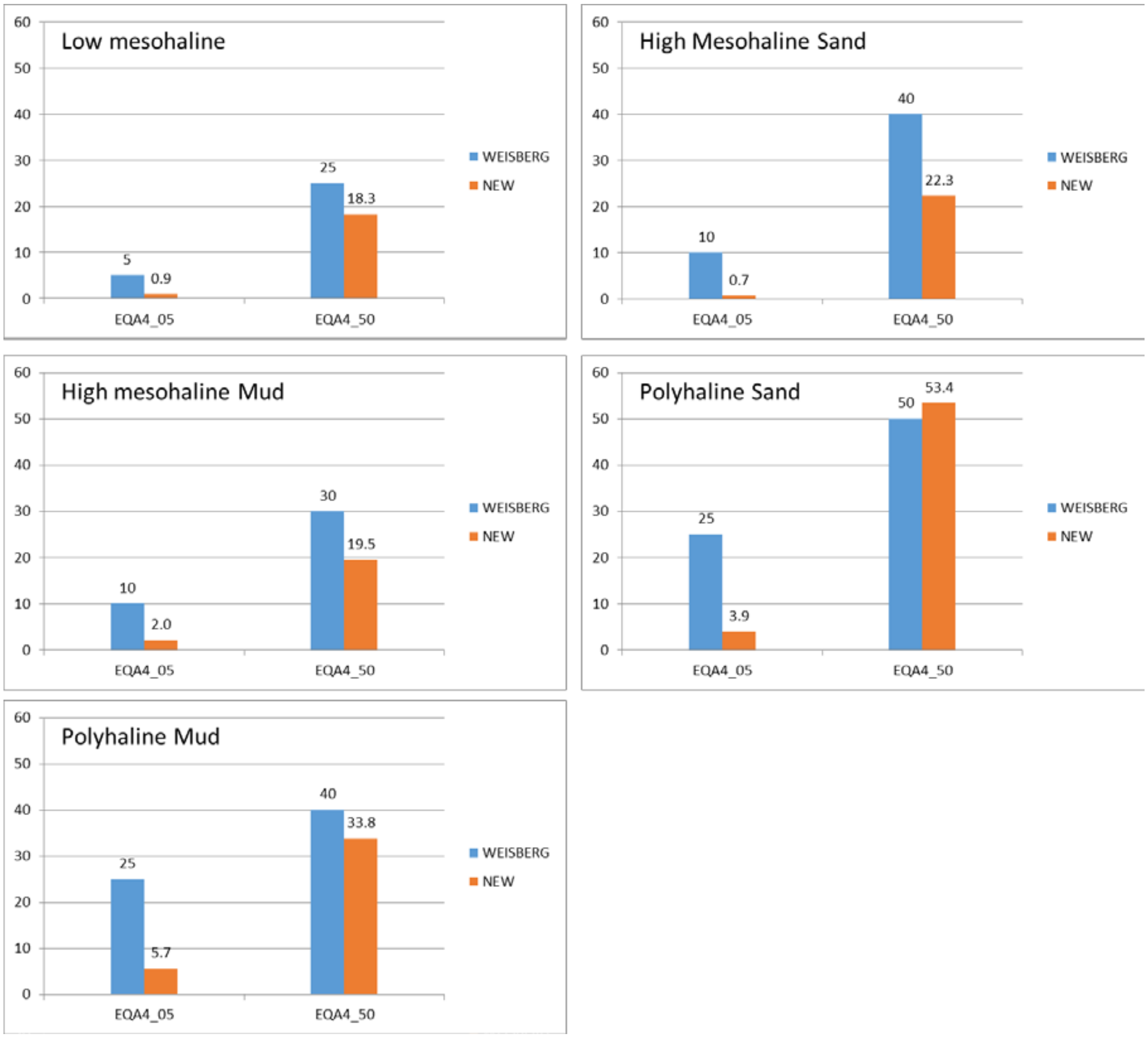

Figure 5. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance of pollution sensitive taxa (%)

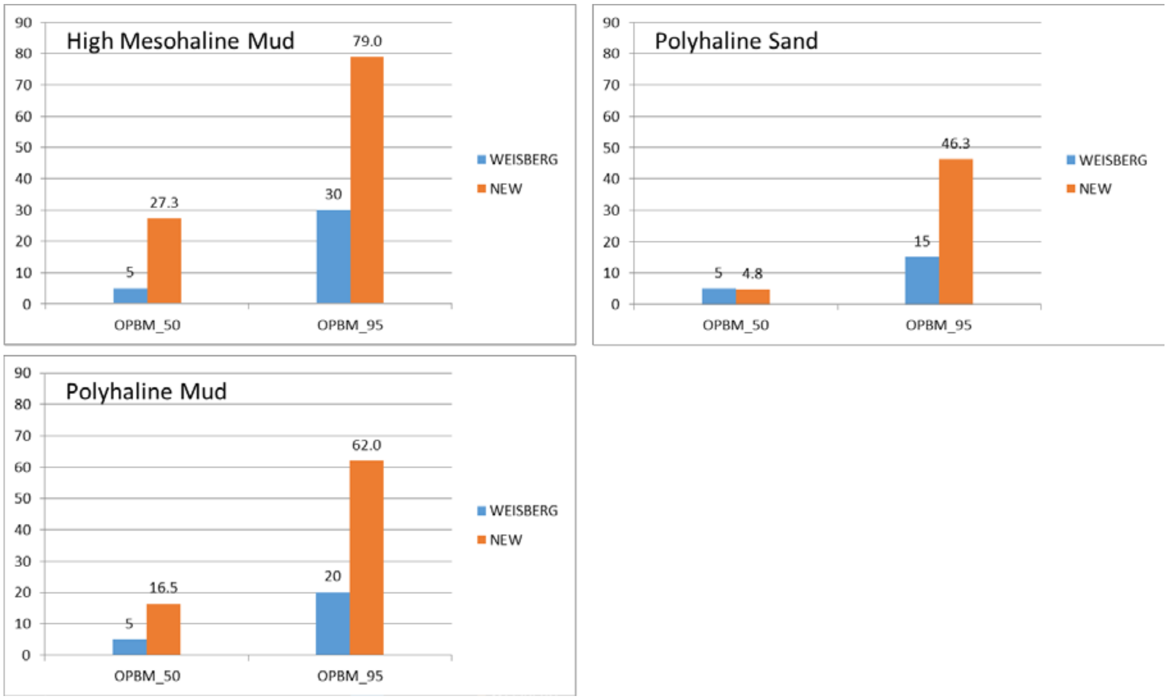

Figure 6. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for biomass of pollution indicative taxa (%)

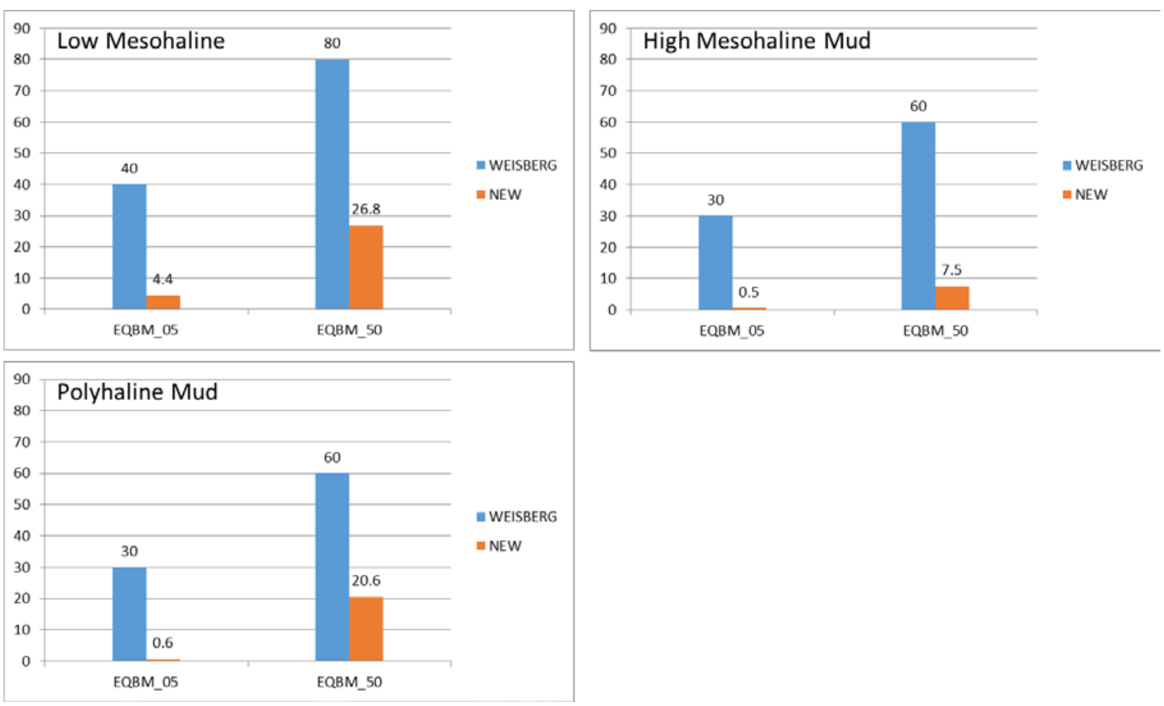

Figure 7. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for biomass of pollution sensitive taxa (%)

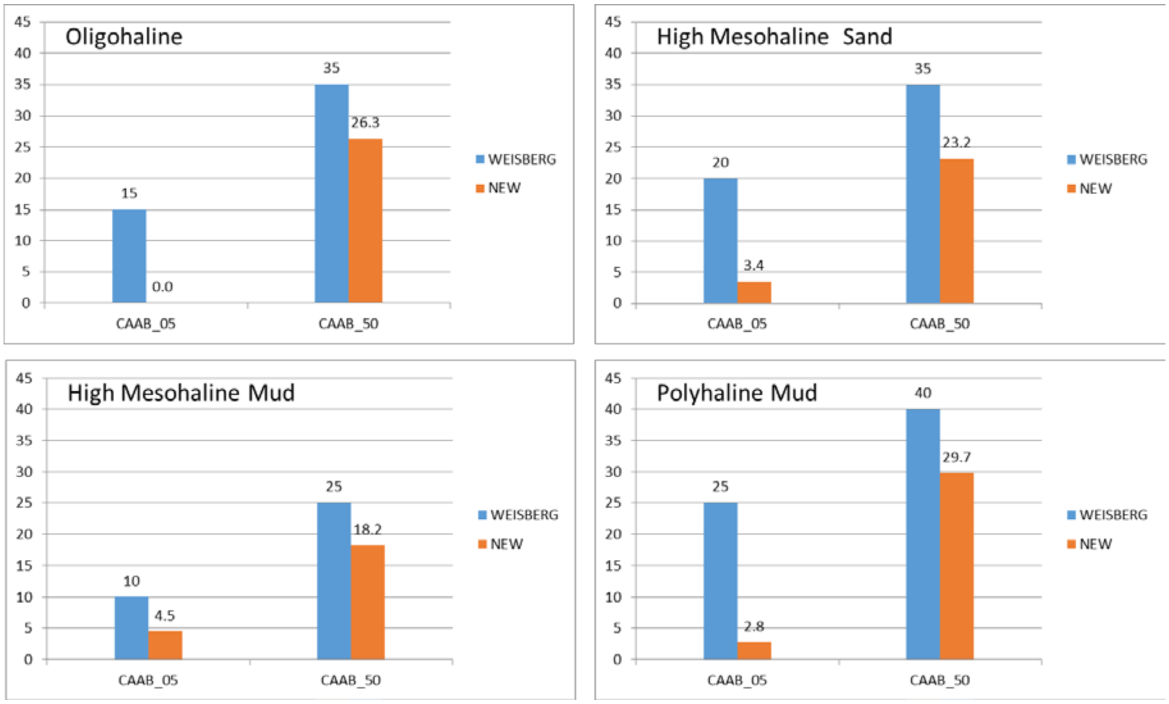

Figure 8. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance of carnivore and omnivores (%)

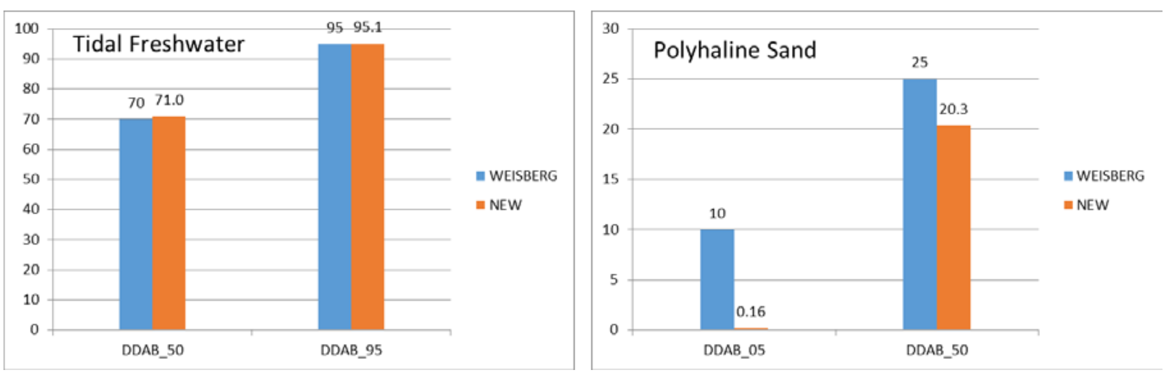

Figure 9. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance of deep-deposit feeders (%)

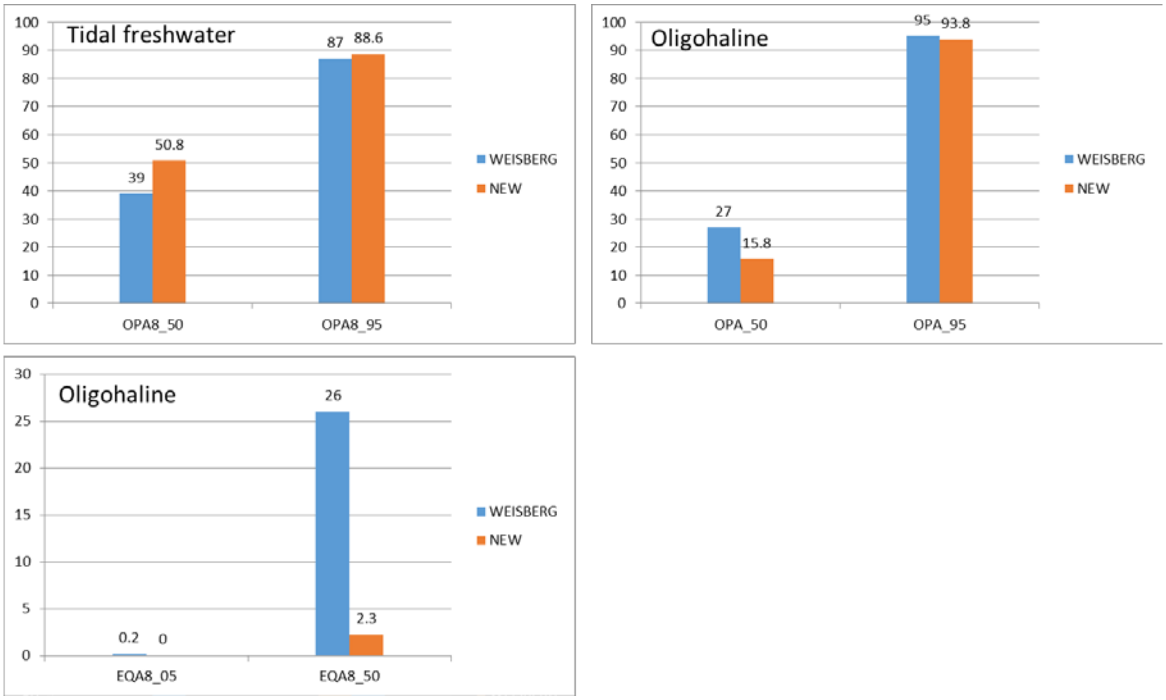

Figure 10. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for abundance of pollution indicative freshwater and oligohaline taxa (%, upper panel), and abundance of pollution sensitive oligohaline taxa (%, lower panel)

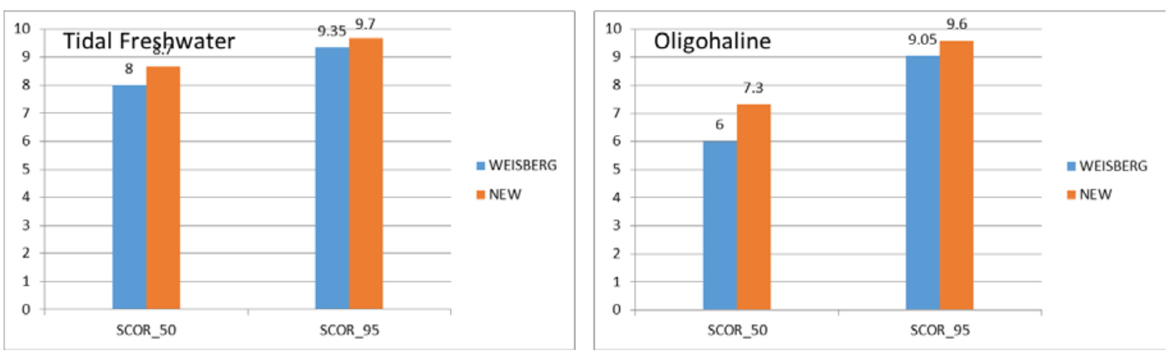

Figure 11. Comparison of thresholds between the original Reference dataset of Weisberg et al. (1997) and the new Reference dataset assembled for this project for Tolerance Score

4.0 VALIDATION ASSESSMENTS

4.1 Original Thresholds

Validation assessment of the original B-IBI thresholds developed by Weisberg et al. (1997) and later updated by Alden et al. (2002) showed classification efficiencies ranging from a minimum of 45% correct classification in the Tidal Freshwater habitat type to a maximum of 81% correct classification in the Polyhaline Mud habitat (Table 5). Classification efficiencies for Low Mesohaline, High Mesohaline Sand, High Mesohaline Mud, and Polyhaline Mud habitat types were higher for Degraded sites than for Reference sites ranging from 55% to 92% (Table 5). Classification efficiencies were higher for Reference sites for Tidal Freshwater, Oligohaline, and Polyhaline Sand habitats in the 68-73% range (Table 5).

Table 5. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using thresholds defined in Weisberg et al. (1997) and Alden et al. (2002) and the entire calibration and validation datasets assembled for this project. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and a priori impact classifications. Overall classification efficiency for this B-IBI is provided in **bold**.

Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Correctly Classified Percentage
Tidal Freshwater	Reference	55	40	72.7
	Degraded	161	58	36.0
	Total	216	98	45.4
Oligohaline	Reference	24	17	70.8
	Degraded	111	70	63.1
	Total	135	87	64.4
Low Mesohaline	Reference	92	51	55.4
	Degraded	214	156	72.9
	Total	306	207	67.6
High Mesohaline Sand	Reference	189	91	48.2
	Degraded	58	32	55.2
	Total	247	123	49.8
High Mesohaline Mud	Reference	106	30	28.3
	Degraded	309	241	78.0
	Total	415	271	65.3
Polyhaline Sand	Reference	240	163	67.9
	Degraded	46	23	50.0
	Total	286	186	65.0
Polyhaline Mud	Reference	47	18	38.3
	Degraded	179	164	91.6
	Total	226	182	80.5
	Overall	**1831**	**1154**	**63.0**

4.2 New Thresholds With and Without Modified Biomass Scoring

Validation of the B-IBI scored using new thresholds developed from old (data available to Weisberg et al.) and new probability-based data showed total classification efficiencies ranging from a minimum of 31% correct classification in the Oligohaline habitat type to a maximum of 68% correct classification in the Polyhaline Sand habitat (Table 6). Classification efficiencies for Reference sites were substantially higher than for Degraded sites (Table 6) ranging from 56% in the Oligohaline to 100% correct classification in Polyhaline Sand. Classification efficiencies for Degraded sites were less than 50% in all habitat types (Table 6). Modification of the procedure for scoring biomass using the same thresholds resulted in little and often no change in classification efficiency for all of the habitat types for both Reference and Degraded sites (Table 7).

Table 6. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using new thresholds and the validation dataset assembled for this project. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and a priori impact classifications. Overall classification efficiency for this B-IBI approach is provided in **bold**.

			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Tidal Freshwater	Reference	22	15	68.2
	Degraded	161	49	30.4
	Total	183	64	35.0
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Oligohaline	Reference	9	5	55.6
	Degraded	111	32	28.8
	Total	120	37	30.8
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Low Mesohaline	Reference	33	25	75.8
	Degraded	214	101	47.2
	Total	247	126	51.0
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Sand	Reference	65	53	81.5
	Degraded	58	18	31.0
	Total	123	71	57.7
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Mud	Reference	39	32	82.1
	Degraded	309	159	51.5
	Total	348	191	54.9
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Sand	Reference	81	77	95.1
	Degraded	46	9	19.6
	Total	127	86	67.7
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Mud	Reference	15	15	100
	Degraded	179	70	39.1
	Total	194	85	43.8
	Overall	**1342**	**660**	**49.2**

Table 7. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using new thresholds, the validation dataset assembled for this project, and a modified procedure for scoring biomass. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and a priori impact classifications. Overall classification efficiency for this B-IBI approach is provided in **bold**.

			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Tidal Freshwater	Reference	22	15	68.2
	Degraded	161	49	30.4
	Total	183	64	35.0
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Oligohaline	Reference	9	5	55.6
	Degraded	111	32	28.8
	Total	120	37	30.8
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Low Mesohaline	Reference	33	28	84.9
	Degraded	214	96	44.9
	Total	247	124	50.2
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Sand	Reference	65	56	86.2
	Degraded	58	18	31.0
	Total	123	74	60.2
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Mud	Reference	39	34	87.2
	Degraded	309	149	48.2
	Total	348	183	52.6
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Sand	Reference	81	76	93.8
	Degraded	46	9	19.6
	Total	127	85	66.9
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Mud	Reference	15	14	93.3
	Degraded	179	72	40.2
	Total	194	86	44.3
	Overall	**1342**	**653**	**48.7**

4.3 New Thresholds Based on Means With and Without Modified Biomass Scoring

Classification efficiencies obtained for the B-IBI based on thresholds developed using a calibration dataset of mean replicate values were, in general, similar to those obtained from thresholds developed using a calibration dataset of individual replicate values (presented above), although the overall classification efficiency improved slightly (Table 8). Total classification efficiencies by habitat type ranged from a minimum of 32% correct classification in the Oligohaline habitat type to a maximum of 71% correct classification in Polyhaline Sand habitat (Table 8). In general, classification efficiencies for

Reference sites were substantially higher within habitat types than for Degraded sites (Table 8). Modification of the procedure for scoring biomass resulted in almost no changes in classification efficiencies with the exception of a slight improvement in the classification of Degraded sites within the Polyhaline Sand habitat (Table 9).

Table 8. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using new thresholds (developed from mean replicate values) and the validation dataset assembled for this project. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and *a priori* impact classifications. Overall classification efficiency for this B-IBI approach is provided in **bold**.

Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Tidal Freshwater	Reference	7	4	57.1
	Degraded	84	46	54.8
	Total	91	50	54.9
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Oligohaline	Reference	5	4	80.0
	Degraded	55	15	27.3
	Total	60	19	31.7
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Low Mesohaline	Reference	15	13	86.7
	Degraded	107	37	34.6
	Total	122	50	41.0
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
High Mesohaline Sand	Reference	39	37	94.9
	Degraded	32	9	28.1
	Total	71	46	64.8
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
High Mesohaline Mud	Reference	18	16	88.9
	Degraded	181	85	47.0
	Total	199	101	50.8
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Polyhaline Sand	Reference	52	49	94.2
	Degraded	26	6	23.1
	Total	78	55	70.5
Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Polyhaline Mud	Reference	11	11	100
	Degraded	109	40	36.7
	Total	120	51	42.5
	Overall	**741**	**372**	**50.2**

Table 9. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using new thresholds (developed from mean replicate values), the validation dataset assembled for this project, and a modified procedure for scoring biomass. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and *a priori* impact classifications. Overall classification efficiency for this B-IBI approach is provided in **bold**.

			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Tidal Freshwater	Reference	7	4	57.1
	Degraded	84	46	54.8
	Total	91	50	54.9
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Oligohaline	Reference	5	4	80.
	Degraded	55	15	27.3
	Total	60	19	31.7
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Low Mesohaline	Reference	15	14	93.3
	Degraded	107	35	32.7
	Total	122	49	40.2
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Sand	Reference	39	36	92.3
	Degraded	32	9	28.1
	Total	71	45	63.4
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
High Mesohaline Mud	Reference	18	16	88.9
	Degraded	181	84	46.4
	Total	199	100	50.3
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Sand	Reference	52	49	94.2
	Degraded	26	6	23.1
	Total	78	55	70.5
			Correctly Classified	
Habitat	*a priori* Classification	Sample #	Number	Percentage
Polyhaline Mud	Reference	11	11	100
	Degraded	109	45	41.3
	Total	120	56	46.7
	Overall	**741**	**374**	**50.5**

4.4 Thresholds based on data collected after 1997

Classification efficiencies obtained for the B-IBI based on thresholds developed using a calibration dataset of mean replicate values and only post-1997 data (i.e., EMAP and MAIA datasets excluded) were slightly lower overall to those obtained using other methods (Table 10). Total classification efficiencies by habitat type ranged from a minimum of 24% correct classification in the Low Mesohaline habitat to a maximum of 73% correct classification in the Polyhaline Sand habitat (Table 10). In general, classification efficiencies for Reference sites were substantially higher within habitat types than for Degraded sites.

Table 10. Classification efficiencies within habitat type and across all habitat types for both Reference and Degraded sites based on B-IBI values scored using new thresholds (developed from mean replicate values of post-1997 samples), the validation dataset assembled for this project, and a modified procedure for scoring biomass. Provided are the total number of validation samples (Sample #) and the number and percentages of samples correctly classified within each habitat type and *a priori* impact classifications. Overall classification efficiency for this B-IBI approach is provided in **bold**.

Habitat	*a priori* Classification	Sample #	Correctly Classified	
			Number	Percentage
Tidal Freshwater	Reference	7	7	100
	Degraded	50	18	36
	Total	57	25	43.9
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
Oligohaline	Reference	4	4	100
	Degraded	27	10	37
	Total	31	14	45.2
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
Low Mesohaline	Reference	10	9	90
	Degraded	56	7	12.5
	Total	66	16	24.2
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
High Mesohaline Sand	Reference	30	26	86.7
	Degraded	20	10	50
	Total	50	36	72
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
High Mesohaline Mud	Reference	20	17	85
	Degraded	120	54	45
	Total	140	71	50.7
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
Polyhaline Sand	Reference	41	39	95.1
	Degraded	18	4	22.2
	Total	59	43	72.9
Habitat	*a priori* Classification	Sample #	Correctly Classified Number	Percentage
Polyhaline Mud	Reference	12	12	100
	Degraded	74	12	16.2
	Total	86	24	27.9
	Overall	**489**	**229**	**46.8**

4.5 Summary

Overall, modifications to the original thresholds of Weisberg et al. (1997) and Alden et al. (2002) based either on changes to the datasets used or the procedure for scoring biomass resulted in decreases in overall classification efficiencies (Figure 12). A closer examination of classification efficiencies within habitat types and a priori impact classification groups indicates that the B-IBI based on new thresholds (Iteration 1), in general, had higher classification efficiencies for Reference sites while the B-IBI based on original thresholds (the baseline) had higher classification efficiencies for Degraded sites for most habitat types (Figure 13). Similar results were obtained for other iterations, including modifications to the existing biomass scoring procedure (see Tables 7 to 10). These results indicate that additional datasets or modifications to existing procedures did not improve the classification efficiency of the B-IBI in any of the habitats to a degree that would warrant adoption of any of the iterations here examined.

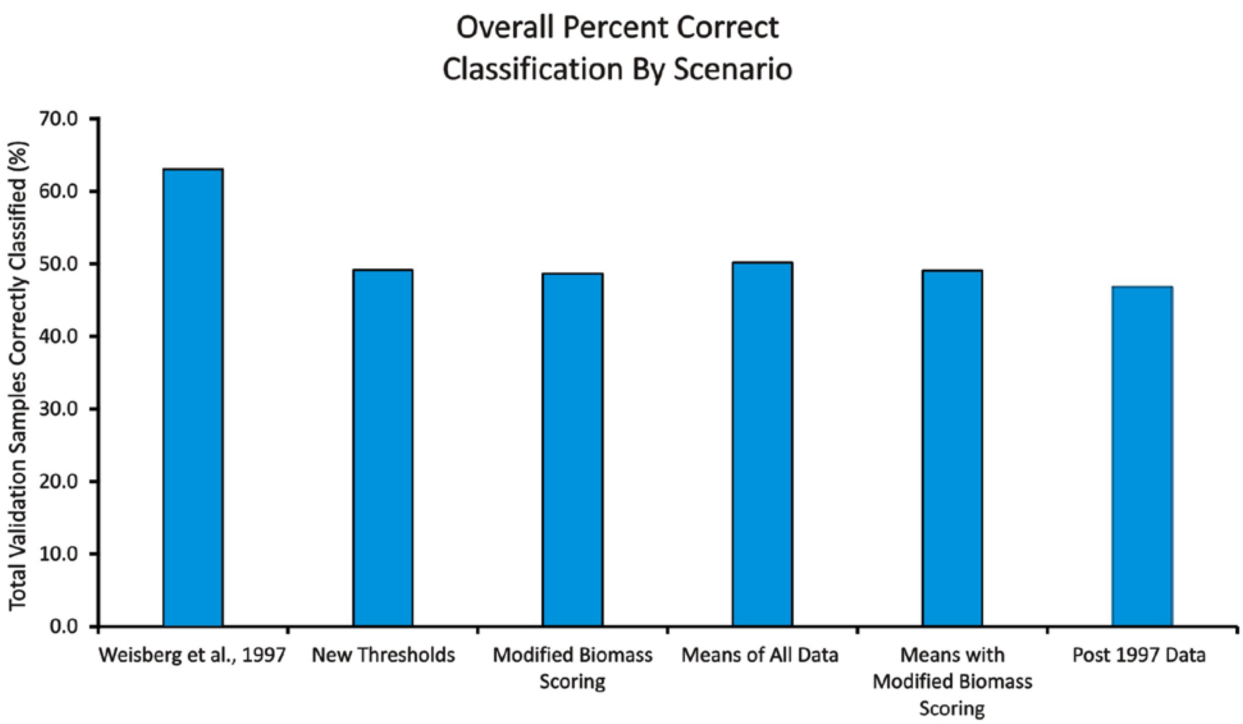

Figure 12. Overall classifications efficiencies for data assembled in this project using the original Weisberg et al. (1997) and Alden et al. (2002) thresholds (the baseline), and new thresholds with or without further modifications to datasets or biomass scoring procedure

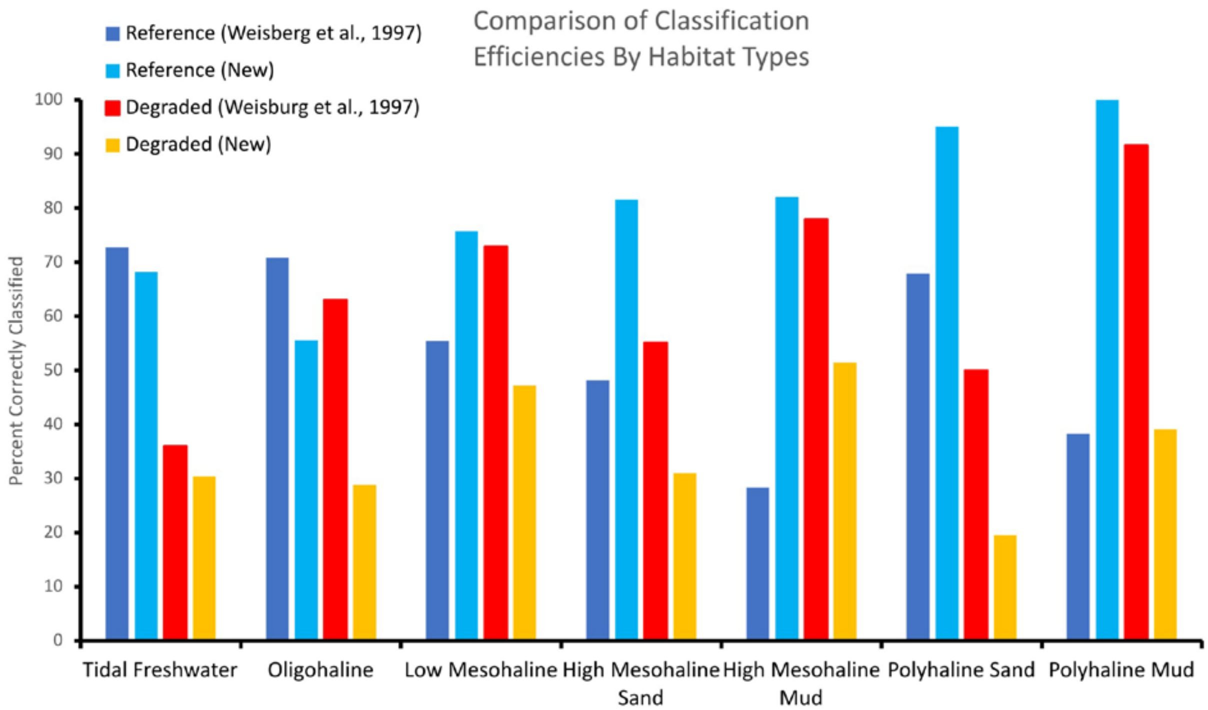

Figure 13. Classification efficiencies for validation dataset for Reference and Degraded sites by habitat type obtained using the original (Weisberg et al., 1997) thresholds and new thresholds for the B-IBI

5.0 WATER DEPTH ANALYSIS

The validation assessments indicated that additional data did not improve the classification efficiency of the B-IBI. Further, the new calibration data included many more depauperate samples than the data of the initial calibration effort. Also, the new data did not improve the challenges in the low salinity habitats. When the calibration data were segregated by depth, it was noted shallow versus deep differences among the values of a metric. For data after 1996 (i.e., excluding the EMAP samples), the lowest values in the calibration dataset below the lower 5th percentile threshold (or the highest values above the upper 95th percentile threshold) were on average in shallow water for most metrics (Table 11). Some of the differences were statistically significant. The Polyhaline Sand and Polyhaline Mud habitats have a water depth boundary of about 3-4 m (Table 12). This corresponds to Reilly's 4 m boundary, an area of "maximum interaction between human activities and biological resources" (Reilly 1996). Thus, water depth may be a surrogate for nearshore anthropogenic effects.

Table 11. Average water depth (m) of calibration samples for metrics below the original Weisberg et al. (1997) 5th percentile threshold, or above the 95th percentile threshold. Data after 1996. Numbers in bold and underlined are significantly different by t-test at the 0.05 probability level. Shaded cells indicate depths that are, on average, lower for "bad" values of a metric (values below the 5th percentile threshold or above the 95th percentile threshold) than for "good" values. Blanks denote metrics that are not attributes of the B-IBI in that habitat. * >95th percentile threshold.

Habitat	Shannon (H')		Abundance		Biomass		Pollution Indicative Abundance		Pollution Sensitive Abundance		Pollution Indicative Biomass	
	Low	High	Low/High	Medium	Low	High	High*	Low	Low	High	High*	Low
Polyhaline Sand	4.0	4.8	**3.2**	**5.1**	**3.7**	**5.5**	.	.	4.5	4.5	3.4	5.1
Polyhaline Mud	**3.1**	**7.1**	3.6	6.1	4.4	6.0	.	.	.	.	4.5	7.1
High Mesohaline Mud	3.8	2.7	3.3	2.6	2.4	3.4	.	.	.	.	2.8	2.9
High Mesohaline Sand	2.4	2.7	2.2	2.8	2.8	2.4	2.8	2.4	**1.9**	**2.9**	.	.
Low Mesohaline	1.9	2.1	3.2	1.9	1.8	2.5	2.1	2.1	.	.	.	.
Oligohaline	.	.	4.1	2.3	.	.	3.2	2.5	2.0	3.2	.	.
Tidal Freshwater	.	.	3.0	3.0	.	.	n/a	3.0	.	.	.	.

Habitat	Pollution Sensitive Biomass		Carnivore and Omnivore Abundance		Deep Deposit Feeder Abundance		Tolerance Score		Tanypodinae/ Chironomidae Abundance Ratio	
	Low	High	Low	High	Low	High	High*	Low	High*	Low
Polyhaline Sand	.	.	.	.	4.8	4.3	.	.	.	.
Polyhaline Mud	4.0	7.3	**3.2**	**8.7**	.	.	.	.	.	.
High Mesohaline Mud	2.7	3.5	2.1	3.1	.	.	.	.	.	.
High Mesohaline Sand	.	.	2.1	3.0	.	.	.	.	.	.
Low Mesohaline	1.9	2.5	.	.	.	.	.	.	.	.
Oligohaline	.	.	2.0	2.9	.	.	2.4	2.7	1.5	2.8
Tidal Freshwater	.	.	.	.	3.0	n/a*	5.5	2.8	.	.

Table 12. Comparisons of B-IBI component metrics between shallow (<=4 m) and deep (>4 m) samples in the Polyhaline Sand Reference habitat. Provided are descriptive statistics, results of two-tailed Student's t-tests and equality of variance tests for each metric. If test for equality of variance is significant, then t-test provided is for unequal variances.

Abundance (#/m^2)											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	7,876	21359.8	2962.1	52.13	-1.64	0.11	51.00	42.00	109.42	<0.001
Deep	43	2,997	2042.0	311.4							
Biomass (g AFDW/m^2)											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	3.19	16.01	2.22	52.42	-0.68	0.50	51.00	42.00	86.66	<0.001
Deep	43	1.67	1.72	0.26							
Shannon Index (H')											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	2.58	0.90	0.12	93.00	1.85	0.07	51.00	42.00	1.03	0.93
Deep	43	2.92	0.89	0.14							
Pollution Sensitive Abundance (%)											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	48.94	27.73	3.85	93.00	-1.29	0.20	51.00	42.00	1.24	0.47
Deep	43	41.87	24.89	3.80							
Pollution Indicative Biomass (%)											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	18.96	18.22	2.53	90.73	-2.97	**<0.001**	51.00	42.00	2.03	0.02
Deep	43	9.48	12.78	1.95							
Deep Deposit Feeder Abundance (%)											
	Descriptive Statistics				t Test				Equality of Variance Test		
Class	N	Mean	Std. Dev.	Std. Error	DF	t Value	P > t	Num DF	Den DF	F	P>F
Shallow	52	23.82	20.69	2.87	87.88	-1.92	0.06	51.00	42.00	2.44	<0.001
Deep	43	17.07	13.25	2.02							

6.0 POLLUTION TOLERANCE

The pollution tolerance of benthic species typically is categorized based on their life-history characteristics (Dauer 1993). However, in some cases life-history characteristics are inconsistent with pollution tolerance (Seitz and Schaffner 1995). Capitellid polychaetes, for example, have been identified as opportunistic (high reproductive output, rapid growth), but in Chesapeake Bay a key species, *Mediomastus ambiseta*, was found in higher numbers in reference sites than in degraded sites (Weisberg et al. 1997). This species is currently listed as pollution sensitive in the B-IBI. One concern with this species listed as pollution sensitive is its dominance throughout the Elizabeth River. Of special concern is its dominance in the Southern Branch and some smaller creeks of the Southern Branch, all considered highly anthropogenically stressed. In addition this species has increased greatly in dominance over the years. Based on these concerns, *Mediomastus ambiseta* classification as pollution sensitive in the B-IBI was reconsidered during the present effort by testing for differences in abundance among sites in the more recent data.

Results of t-test comparing the abundance of *M. ambiseta* between Reference and Degraded sites indicated no significant differences in means for any of the habitat types (Table 13). However, results

Table 13. Summary of two sample comparisons of mean and median abundance of *Mediomastus ambiseta* between Degraded and Reference sites for High Mesohaline Sand, High Mesohaline Mud, Polyhaline Sand and Polyhaline Mud habitat types. Provided for each habitat type are descriptive statistics, t-test results, equality of variance tests (all indicating significantly different variances), Wilcoxon two-sample, and Kolmogorov-Smirnov comparisons of distributions.

High Mesohaline Sand	Descriptive Statistics				t-Test (unequal Variances)			Equality of Variance Test		Wilcoxon Two Sample test			Kolmogorov-Smirnov Test	
Impact Classification	N	Mean	Std. Dev.	Std. Error	D.F.	t Value	P > t	F	P > F	W Value	Z	P>Z	KS_a	$P>KS_a$
Degraded	56	27.7	184.40	24.64	57.78	0.47	0.641	12.99	<0.001	5506	-2.43	0.008	1.49	0.023
Reference	172	16.0	51.16	3.90										

High Mesohaline Mud	Descriptive Statistics				t-Test (unequal Variances)			Equality of Variance Test		Wilcoxon Two Sample T-test			Kolmogorov-Smirnov Test	
Impact Classification	N	Mean	Std. Dev.	Std. Error	D.F.	t Value	P > t	F	P > F	W Value	Z	P>Z	KS_a	$P>KS_a$
Degraded	264	7.3	49.69	3.06	211.85	-1.21	0.224	1.47	0.027	22307	6.52	<0.001	2.72	<0.001
Reference	99	13.5	40.98	4.12										

Polyhaline Sand	Descriptive Statistics				t-Test (unequal Variances)			Equality of Variance Test		Wilcoxon Two Sample T-test			Kolmogorov-Smirnov Test	
Impact Classification	N	Mean	Std. Dev.	Std. Error	D.F.	t Value	P > t	F	P > F	W Value	Z	P>Z	KS_a	$P>KS_a$
Degraded	41	28.3	36.40	5.69	78.38	-0.09	0.933	1.93	0.016	4602	-0.13	0.4463	0.8	0.540
Reference	185	28.9	50.55	3.72										

Polyhaline Mud	Descriptive Statistics				t-Test (unequal Variances)			Equality of Variance Test		Wilcoxon Two Sample T-test			Kolmogorov-Smirnov Test	
Impact Classification	N	Mean	Std. Dev.	Std. Error	D.F.	t Value	P > t	F	P > F	W Value	Z	P>Z	KS_a	$P>KS_a$
Degraded	152	26.4	63.26	5.13	78.62	-1.58	0.119	1.74	0.045	5338	4.78	<0.001	2.35	<0.001
Reference	40	40.8	47.97	7.58										

were more complicated when examined using nonparametric procedures and distribution tests. Wilcoxon two-sample tests for both High Mesohaline Mud, High Mesohaline Sand, and Polyhaline Mud habitats indicated median abundances of *M. ambiseta* significantly higher in Reference than in Degraded sites. Additionally there were significant differences in the distribution of this species between Reference and Degraded sites for these habitat types (Table 13). In the Polyhaline Sand, no significant differences between medians or distributions were observed.

These results indicate that *M. ambiseta* could not be consistently characterized as being strictly representative of either Reference or Degraded sites. This species has been referred to as opportunistic and pollution indicative based both on ecological surveys (Grassle and Grassle, 1974; Boesch, 1977; Billheimer et al., 1997) and experimental results (Shaffner, 1990). Given the evidence from the literature in combination with the results of this study, it is likely that retaining *M. ambiseta* as either pollution sensitive or pollution indicative for the purposes of the B-IBI calculation is likely to result in sample misclassifications, and is therefore unwarranted.

7.0 CONCLUSIONS

1. Additional data did not improve the classification efficiency of the B-IBI in any of the habitats.
2. Data did not improve the challenges in the lower salinity habitats. This is a global issue with no obvious solution.
3. Samples meeting the reference criteria included enough samples with low diversity, abundance, biomass, and numbers of pollution indicative and sensitive taxa to bias the data toward too many false positives of undegraded condition.
4. There are at least two hypotheses relative to the lowered thresholds and unacceptable correct classification efficiencies compared to the baseline:
    a. Anthropogenic stress criteria not accounted for by this study might better classify samples into Reference and Degraded categories. However, the same criteria that were used in the initial calibration effort were used in this study. Water depth as a surrogate for nearshore anthropogenic effect is one possible new criterion.
    b. There is a subtle deterioration of water quality in the Bay that has resulted in false positives in our calibration dataset.
5. A reasonable next step is a best professional judgement approach to determining biological criteria, similar to that of Weisberg et al. (2008).

8.0 LITERATURE CITED

Alden, R.W. III, D.M. Dauer, J.A. Ranasinghe, L.C. Scott, and R.J. Llansó. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. Environmetrics 13:473-498.

Billheimer, D., D.T. Cardoso, E. Freeman, P. Guttorp, H. Ko, and M. Silkey. 1997. Natural variability of benthic species composition in the Delaware Bay. Environmental and Ecological Statistics. 4:95-115.

Boesch, D.F. 1977. A new look at the zonation of benthos along the estuarine gradient. Pp. 245-266 *In* B.C. Coull (ed.), Ecology of Marine Benthos, University South Carolina Press, Columbia, SC.

Dauer, D.M. 1993. Biological criteria, environmental health and estuarine macrobenthic community structure. Marine Pollution Bulletin 26:249-257.

Dauer, D.M. and M.F. Lane. 2005. Side-by-side comparison of Young grab and composite petite ponar grab samples for the calculation of the Benthic Index of Biological Index (B-IBI). Report to the Virginia Department of Environmental Quality by Old Dominion University, Norfolk VA. 46 pp.

Grassle, J. F. and Grassle, J. P. 1974 . Opportunistic life histories and genetic systems in marine benthic polychaetes. Journal of Marine Research 32:253-284.

Hyland, J.F., W.L. Balthis, V.D. Engle, E.R. Long, J.F. Paul, J.K. Summers and R.F. Van Dolah. 2003. Incidence of stress in benthic communities along the U.S. Atlantic and Gulf of Mexico coasts within different ranges of sediment contamination from chemical mixtures. Environmental Monitoring and Assessment 81:149-161.

Llansó, R.J., L.C. Scott, D.M. Dauer, J.L. Hyland, and D.E. Russell. 2002. An estuarine benthic index of biotic integrity for the Mid-Atlantic region of the United States. I. Classification of assemblages and habitat definition. Estuaries 25:1219-1230.

Llansó R.J., J.H. Vøstad, D.M. Dauer and J.R. Dew. 2009. Assessing benthic community condition in Chesapeake Bay: does the use of different benthic indices matter? Environmental Monitoring and Assessment 150:119-127.

Long, E.R., D.D. MacDonald, S.l. Smith, and F.D. Calder. 1995. Incidence of adverse environmental effects within ranges of chemical concentrations in marine and estuarine sediments. Environmental Management 19:81-97.

Ranasinghe, J.A., S.B. Weisberg, D.M. Dauer, L.C. Schaffner, R.J. Diaz, and J.B. Frithsen. 1994. Chesapeake Bay Benthic Community Restoration Goals. Report prepared for the U.S. Environmental Protection Agency Chesapeake Bay Program Office, the Governor's Council on Chesapeake Bay Research Fund, and the Maryland Department of Natural Resources by Versar, Inc., Columbia, Maryland.

Reilly Jr., F.J., R.J. Spagnolo, and E. Ambrogio. 1996. Marine and estuarine shallow water science and management. Estuaries 19:166–168.

Schaffner L.C. 1990. Small-scale organism distributions and patterns of species diversity: evidence for positive interactions in an estuarine benthic community. Marine Ecology Progress Series 61:107-117.

Seitz, R.D. and L.C. Schaffner. 1995. Polulation ecology and secondary production of the polychaete *Loimia medusa* (Terebellidae). Marine Biology 121:701-711.

Van Dolah, R.F., J.L. Hyland, A.F. Holland, J.S. Rosen, and T. R. Snoots. 1999. A benthic index of biological integrity for assessing habitat quality in estuaries of the Southeastern USA. Marine Environmental Research 48:269-283.

Weisberg, S.B., J.A. Ranasinghe, D.M. Dauer, L.C. Schaffner, R.J. Diaz, and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for the Chesapeake Bay. Estuaries 20:149-158.

Weisberg, S.B., B. Thompson, J.A. Ranasinghe, D.E. Montagne, D.B. Cadien, D.M. Dauer, D. Diener, J. Oliver, D.J. Reish, R.G. Velarde, and J.Q. Word. 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. Ecological Indicators 8:389-394.