Estuaries Vol. 25, No. 6A, p. 1231–1242 December 2002

An Estuarine Benthic Index of Biotic Integrity for the Mid-Atlantic Region of the United States. II. Index Development

ROBERTO J. LLANSÓ[1,*], LISA C. SCOTT[1], JEFFREY L. HYLAND[2], DANIEL M. DAUER[3], DAVID E. RUSSELL[4], and FREDERICK W. KUTZ[4]

[1] *Versar, Inc., 9200 Rumsey Road, Columbia, Maryland 21045-1934*
[2] *National Oceanic and Atmospheric Administration, National Ocean Service, 219 Fort Johnson Road, Charleston, South Carolina 29412-9110*
[3] *Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529*
[4] *U.S. Environmental Protection Agency, Region III, Environmental Science Center, 701 Mapes Road, Ft. Meade, Maryland 20755-5350*

ABSTRACT: A benthic index of biotic integrity was developed for use in estuaries of the mid-Atlantic region of the United States (Delaware Bay estuary through Albemarle-Pamlico Sound). The index was developed for the Mid-Atlantic Integrated Assessment Program (MAIA) of the U.S. Environmental Protection Agency using procedures similar to those applied previously in Chesapeake Bay and southeastern estuaries, and was based on sampling in July through early October. Data from seven federal and state sampling programs were used to categorize sites as degraded or non-degraded based on dissolved oxygen, sediment contaminant, and sediment toxicity criteria. Various metrics of benthic community structure and function that distinguished between degraded and reference (non-degraded) sites were selected for each of five major habitat types defined by classification analysis of assemblages. Each metric was scored according to thresholds established on the distribution of values at reference sites, so that sites with low scoring metrics would be expected to show signs of degradation. For each habitat, metrics that correctly classified at least 50% of the degraded sites in the calibration data set were selected whenever possible to derive the index. The final index integrated the average score of the combination of metrics that performed best according to several criteria. Selected metrics included measures of productivity (abundance), diversity (number of taxa, Shannon-Wiener diversity, percent dominance), species composition and life history (percent abundance of pollution-indicative taxa, percent abundance of pollution-sensitive taxa, percent abundance of Bivalvia, Tanypodinae-Chironomidae abundance ratio), and trophic composition (percent abundance of deep-deposit feeders). The index correctly classified 82% of all sites in an independent data set. Classification efficiencies of sites were higher in the mesohaline and polyhaline habitats (81–92%) than in the oligohaline (71%) and the tidal freshwater (61%). Although application of the index to low salinity habitats should be done with caution, the MAIA index appeared to be quite reliable with a high likelihood of correctly identifying both degraded and non-degraded conditions. The index is expected to be of great utility in regional assessments as a tool for evaluating the integrity of benthic assemblages and tracking their condition over time.

Introduction

In 1995, the U.S. Environmental Protection Agency implemented a research, monitoring, and assessment program to gather information on the extent and condition of natural resources in the mid-Atlantic region of the United States. This program, the Mid-Atlantic Integrated Assessment (MAIA), covers the watersheds of the Delaware Bay, Chesapeake Bay, the coastal bays of the Delmarva peninsula, and the Albemarle-Pamlico Sound. Since its establishment, MAIA has focused on a variety of issues that affect estuaries, streams, groundwater, and landscapes in the mid-Atlantic region (Bradley and Landy 2000). With regard to estuaries, MAIA has focused on various impacts resulting from population growth. Between 1950 and 1990, population in the MAIA region grew from 13 to 21 million (Culliton et al. 1990). The progressive urbanization and industrialization of watersheds in the MAIA region have caused a variety of impacts that have changed the quality of these estuaries over the years. The Delaware estuary, for example, is affected by the lack of water clarity and by toxic contaminants. It is one of the most nutrient-enriched estuaries in the world and has one of the highest levels of chemical contaminants in fish and shellfish in the nation (National Oceanic and Atmospheric Administration 1994; Pennock et al. 1994). The Chesapeake Bay suffers from nutrient over-enrichment, low dissolved oxygen, and reduced water clarity (Malone 1987; Tuttle et al. 1987; Harding 1994; Malone et al. 1996; Bricker et al. 1999). Albemarle-Pamlico Sound and neighbor-

* Corresponding author; tele: 410/964-9200; e-mail: llansorob@versar.com.

ing tributaries have been affected by multiple stressors, including low-oxygen conditions due to organic over-enrichment (Paerl et al. 1998) and sediment contamination (Hyland et al. 2000).

Regional assessments of environmental conditions are necessary for effective restoration of natural resources and maintenance of estuarine environmental quality. The objective is to use a common set of measurements that can be applied uniformly over the entire region. A recent approach to environmental characterization in estuaries has been the development and application of indices of biotic integrity that combine several measures of benthic community response (Weisberg et al. 1997; Adams et al. 1998; Engle and Summers 1999; Van Dolah et al. 1999; Paul et al. 2001). These indices focus on benthic communities because the benthos respond predictably to natural and anthropogenic stress (Pearson and Rosenberg 1978; Dauer 1993; Wilson and Jeffrey 1994). Two methods have been used. One method combines stepwise and linear discriminant analyses to produce a multi-variate index (Weisberg et al. 1993; Engle et al. 1994; Engle and Summers 1999; Paul et al. 2001). The discriminant approach produces a combination of measures that discriminate optimally between reference and degraded conditions. Components are normalized to account for the effects of natural variability, but the normalization process can be complex.

The second method defines community characteristics expected at reference sites with no evidence of anthropogenic stress. A scoring system based on the distribution of values of key benthic attributes at these sites is then used to evaluate the condition of a site. Variability due to natural factors such as salinity or sediment type is accounted for by partitioning the data into habitat classes. This method was first applied to fish assemblages in freshwater environments by Karr (1981) and is known as the index of biotic integrity (IBI). Weisberg et al. (1997) modified the method to develop a benthic index of biotic integrity (B-IBI) for use in the Chesapeake Bay. The method has also been applied in the development of benthic indices for New York-New Jersey Harbor (Adams et al. 1998) and for estuaries of the southeastern U.S. (Van Dolah et al. 1999). The B-IBI approach is easy to understand and interpret, and an advantage over other methods resides in its ability to evaluate ecological condition of a sample by comparing values of key benthic attributes to reference values expected under non-degraded conditions in similar habitat types.

In this paper, we develop an index of benthic community condition for use in the MAIA region using the B-IBI approach. Results of initial steps to identify habitat types have been reported by Llansó et al. (2002). Here we describe the index development process and test the efficiency of the index at classifying samples as degraded vs. non-degraded.

Methods

The index development process consisted of seven steps: data compilation and standardization, classification of sites into reference (non-degraded) and degraded categories, partitioning of data into calibration and validation data sets segregated according to habitat types, testing of metrics for differences between reference and degraded sites for each of the major habitat types, calculation of reference-range thresholds and scoring of metrics, selection of metrics according to site classification criteria, and combination of metrics to represent the MAIA index. Each step is described in detail below.

Data Compilation and Standardization

Data from seven sampling programs were assembled for this study. Information on the sampling programs were presented in Llansó et al. (2002). Data from federal (MAIA, Environmental Monitoring and Assessment Program, National Status and Trends) and state (Maryland and Virginia) monitoring programs that have sampled estuaries in the region from as early as 1984 were assembled in a database. Only sites that included benthic invertebrate data and were sampled during the summer, defined as July 1 through October 7, were used in this study. Data for sites sampled more than once during the summer were averaged. All sampling programs used a 440-cm^2 Young grab (one to three grabs per site), sieved samples through a 0.5-mm mesh screen, and identified organisms to the lowest possible taxonomic level.

Information collected in conjunction with the benthic samples included water column measures (salinity, water depth, and bottom dissolved oxygen) and sediment quality measures (silt-clay content, percent organic carbon, contaminant concentrations, and sediment toxicity). Sediment toxicity was based on two laboratory bioassays, a 10-d acute amphipod bioassay using *Ampelisca abdita* (American Society for Testing Materials 1993) and the Microtox bioassay, which measures changes in bacterial luminescence as an indicator of acute sublethal effects in sediment elutriates (Bulich et al. 1981). Details on specific sampling protocols and measurements taken can be found in Chaillou et al. (1996), Dauer et al. (1998), Hyland et al. (1998, 2000), Strobel (1998), Paul et al. (1999), and Ranasinghe et al. (1999). Prior to analysis, data were standardized to ensure common species nomenclature and uniformity in station designation, variable designation, and units of measure. See Llansó et al. (2002) for the species standardization protocol. The database consisted of 1,999 sites, 19 of which were sampled in multiple years yielding a total of 2,083 sampling events. Because the majority of the sites were sampled only once, we refer to all sampling events as sites.

Classification of Sites

Reference (non-degraded) and degraded sites were selected according to dissolved oxygen, sediment contaminant, and sediment toxicity criteria. Sites were defined as reference if for all sampling events 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 two chemical contaminants exceeded Long et al.'s (1995) effects range-low concentrations, and sediments were not toxic in *Ampelisca* or Microtox bioassays. Amphipod bioassays followed guidelines provided in American Society for Testing Materials (1993) protocols, and were considered to indicate toxicity when mean test survival was significantly different from ($\alpha = 0.05$) and less than 80% of control survival. Microtox bioassays were considered to indicate toxicity when the EC50 of test sediments (sediment concentration that reduces bacterial light production by 50% relative to water controls) was $\leq 0.2\%$ for sediments with silt-clay content $\geq 20\%$, the EC50 was $\leq 0.5\%$ for sediments with silt-clay content $< 20\%$ (Ringwood et al. 1997), or the EC50 of test sediments was significantly different from controls. Sites were defined as degraded if any of the following criteria were met: dissolved oxygen concentrations were less or equal to 2.0 ppm, any chemical contaminant concentration exceeded Long et al.'s (1995) effects range-median concentrations, or sediments were toxic in the *Ampelisca* or Microtox bioassays. Sites that did not meet the classification criteria for either the reference or the degraded condition, or did not have associated dissolved oxygen or chemistry data, were classified as intermediate.

Partitioning of Data

The classification analysis of benthic assemblages distinguished nine habitat classes as a combination of six salinity classes, two sediment types, and the separation of North Carolina and Delaware-Chesapeake Bay polyhaline sites (Llansó et al. 2002). Based on these results, sites were partitioned into habitat types according to salinity, silt-clay content, and geographical location. Two thirds of the sites were selected randomly to represent the calibration data set. The remaining one third represented the validation data set. The calibration data set was used to develop the index. The validation data set was intentionally withheld and used to evaluate independently the performance of various possible indices resulting from different metric combinations (see below). Sites were classified into two data sets (calibration and validation) for each of three conditions (reference, degraded, and intermediate) and nine habitat groups (Table 1). The number of degraded sites available for both index calibration and validation were very few, particularly for polyhaline and euhaline habitats. For the North Carolina polyhaline sand habitat, no sites could be identified as degraded (Table 1). We confronted this problem by merging all high-salinity sites into one polyhaline-euhaline group. Although this solution was less than optimal, we felt it was reasonable given the high degree of overlap indicated between sediment types, and the lack of unique species representative of North Carolina estuaries (Llansó et al. 2002). The validation rate was nonetheless high for the combined euhaline and polyhaline sites (see Results). The final site classification consisted of two data sets, three conditions, and five habitat classes defined according to salinity: tidal freshwater (0–0.5‰), oligohaline (0.5–5‰), low mesohaline (5–12‰), high mesohaline (12–18‰), and polyhaline (≥ 18‰; Table 1). Intermediate sites were excluded from further consideration.

Testing of Metrics

Twenty-three attributes of benthic community structure and function (referred to as metrics) were tested with the calibration data set to determine those that differed significantly between reference and degraded sites within each of the habitat classes (Table 2). The candidate metric list included a variety of measures of productivity, diversity, species composition, and trophic composition. The list included metrics that were found to be useful indicators of benthic condition in other index development efforts (Weisberg et al. 1997; Adams et al. 1998; Engle and Summers 1999; Van Dolah et al. 1999; Paul et al. 2001), as well as some novel metrics. The productivity and diversity metrics are widely used benthic community metrics. Percent dominance, defined as 100 minus the percent abundance contribution of the top two numerically-dominant taxa in the community, was selected in the development of the B-IBI for southeastern U.S. estuaries (Van Dolah et al. 1999). Sensitive taxa that showed strong relationships to sediment contaminant levels in that study were also used in the present analysis. Pollution-sensitive North Carolina taxa (Group C, Van Dolah et al. 1999) were Ampeliscidae, Haustoriidae, Tellinidae,

TABLE 1. Number of sites available for index development (calibration data set) and validation for each of three conditions (reference, degraded, and intermediate) and nine habitat groups. In the final site classification, sites in groups (5) through (9) were combined into one polyhaline (≥18‰) group (see text). Intermediate sites did not meet the classification criteria for either the reference (non-degraded) or the degraded condition, and were excluded from analysis. Intermediate sites also include sites without either dissolved oxygen or chemistry information. NC = North Carolina; DE-CBAY = Delaware-Chesapeake Bay; Sand = 0–10% silt-clay content (particles <62 μ) by weight; Silt-clay = 10–100% silt-clay content by weight.

		Habitat									
Data Set	Condition	(1) Tidal Freshwater	(2) Oligohaline	(3) Low Mesohaline	(4) High Mesohaline	(5) NC Polyhaline Sand	(6) NC Polyhaline Silt-clay	(7) DE-CBAY Polyhaline Sand	(8) DE-CBAY Polyhaline Silt-clay	(9) DE-CBAY Euhaline	(5)–(9) Combined Polyhaline
Calibration	Reference	24	25	37	87	21	13	18	30	45	127
	Degraded	21	15	37	65	0	3	2	18	4	27
	Intermediate	87	97	230	241	1	7	18	72	138	236
Validation	Reference	12	13	19	44	10	7	9	15	22	63
	Degraded	11	8	19	33	0	1	1	9	2	13
	Intermediate	45	52	100	116	1	3	10	30	69	113

Lucinidae, Hesionidae, Cirratulidae, *Cyathura polita,* and *Cyathura burbancki.* Two other pollution-indicative and sensitive species lists were modified from those used in the Chesapeake Bay B-IBI (Weisberg et al. 1997) and were based either on evidence in the literature or on statistical testing conducted during the Chesapeake Bay B-IBI development. The trophic categories were those used by Weisberg et al. (1997).

One recently published metric (North Carolina Sensitivity Index) and two previously unpublished metrics (Tolerance Score and Tanypodinae-Chironomidae percent abundance ratio) were tested as well. The North Carolina Sensitivity Index (Eaton 2001) is a weighted abundance average for taxa classified according to sensitiveness to pollution. Tolerance values for > 200 estuarine taxa were used as weights in the index (range = 1–5). The Tolerance Score is very similar in construction to the North Carolina Sensitivity Index, and was expanded from Lenat (1993) to include freshwater taxa from piedmont and coastal streams in the MAIA region. The Tanypodinae-Chironomidae percent abundance ratio is a measure of the relative contribution of the Tanypodinae to all the other midges found in a sample. The Tanypodinae are considered tolerant to pollution (Lenat 1993), and the ratio is expected to increase in perturbed areas. Similar ratios have been used in other studies (Barbour et al. 1996).

A metric is defined as an attribute of the benthic assemblage that changes in some predictable way with increased human influence (Karr et al. 1986). For a metric to be useful, differences between reference and degraded sites should be statistically significant and the differences should be in a direction that is consistent with established ecological principles. These principles are based on paradigms which predict changes in stressed ecosystems, such as reductions in the abundance and diversity of organisms and shifts in dominance from pollution sensitive to pollution tolerant species (Pearson and Rosenberg 1978; Odum 1985; Dauer 1993). Based on consistency with these ecological principles and Mann-Whitney *U* tests for differences in means between reference and degraded sites, 21 metrics were selected for further testing (Table 2). A probability level of 0.1 was used to reduce the risk of Type II errors of declaring degraded sites as non-degraded. To avoid additional statistical variance that may reduce the power of the test, sites with no fauna (azoic) were excluded from testing.

Calculation of Thresholds and Scoring of Metrics

The next step in the index development process was the calculation of thresholds for defining ref-

TABLE 2. Candidate metrics and mean values at reference (top number) and degraded (bottom number) sites in the Mid-Atlantic Integrated Assessment (MAIA) region by habitat class. Bold numbers indicate pair is different ($p < 0.1$) by Mann-Whitney *U* test. Asterisks indicate metrics selected for further testing. See text for metric definition. na = not applicable.

Metric	Tital Freshwater	Oligohaline	Low Mesohaline	High Mesohaline	Polyhaline
Productivity					
Abundance* (no. m^{-2})	**3,937.5**	2,758.2	**2,890.3**	**3,609.1**	**4,343.6**
	13,365.8	2,355.6	**908.3**	**1,108.2**	**1,430.9**
Biomass* (g AFDW m^{-2})	7.2	6.0	**2.0**	**3.2**	**4.7**
	2.2	8.4	**3.5**	**1.1**	**5.5**
Diversity					
Shannon-Wiener diversity* (log_2)	1.9	**2.0**	**2.3**	**2.6**	**3.0**
	1.9	**1.4**	**0.7**	**1.2**	**1.6**
Number of taxa*	10.0	**8.0**	**10.0**	**12.2**	**19.2**
	8.9	**5.8**	**2.6**	**4.5**	**8.8**
Percent dominance*	22.9	**28.1**	**35.2**	**40.9**	**47.0**
	25.3	**16.8**	**15.2**	**19.4**	**30.0**
Species Composition					
Percent abundance of pollution-indicative taxa*	**58.8**	**35.6**	**16.1**	**21.0**	9.9
	80.3	**71.3**	**43.4**	**50.8**	22.2
Percent abundance of pollution-sensitive taxa*	**2.7**[a]	**27.7**	22.8	**25.3**	**42.6**
	3.4	**12.6**	23.0	**12.3**	**23.8**
Percent abundance of pollution-sensitive NC taxa*	**1.4**[a]	**7.8**	**26.4**	**17.0**	**16.7**
(Group C, Van Dolah et al. 1999)	**2.7**	**1.8**	**11.0**	**9.8**	**12.7**
Percent abundance of pollution-indicative oligohaline taxa*	na	**49.8**	na	na	na
		83.3			
Percent abundance of pollution-sensitive oligohaline taxa*	na	**21.0**	na	na	na
		6.3			
North Carolina Sensitivity Index*	**1.8**	1.8	**1.7**	**1.6**	**1.7**
	1.3	1.7	**1.2**	**1.3**	**1.3**
Tolerance Score*	**8.3**	**6.2**	na	na	na
	9.2	**8.5**			
Tanypodinae-Chironomidae percent abundance ratio*	**27.2**	**18.5**	na	na	na
	50.8	**52.1**			
Percent abundance of *Limnodrilus*	13.6	5.6	na	na	na
	14.8	13.4			
Percent abundance of Tubificidae*	**52.5**	**33.1**	7.6	**3.4**	0.4
	72.2	**61.8**	11.8	**14.5**	0.6
Percent abundance of Spionidae*	1.2	**21.8**	16.9	15.1	11.3
	0.9	**7.8**	33.1	31.3	24.1
Percent abundance of Capitellidae	0.0	**4.0**[b]	**12.8**[b]	**16.5**[b]	**17.9**[b]
	0.1	**0.0**	**9.5**	**9.3**	**12.1**
Percent abundance of Amphipoda*	**0.5**	5.7	**11.5**	**5.5**	**7.7**
	0.1	11.4	**7.0**	**4.6**	**2.6**
Percent abundance of Bivalvia*	15.5	8.1	**20.0**	**21.3**	**14.3**
	9.3	5.0	**13.6**	**10.7**	**3.5**
Trophic Composition					
Percent abundance of carnivore-omnivores*	**29.4**	**26.0**	**21.8**	**22.2**	28.4
	16.6	**13.4**	**8.6**	**12.7**	37.2
Percent abundance of interface feeders*	**1.9**	29.7	48.3	**33.4**[c]	28.6
	1.1	19.7	49.1	**43.4**	30.2
Percent abundance of deep-deposit feeders*	**53.2**	**37.3**	25.7	29.7	**25.9**
	72.9	**62.0**	34.2	35.6	**19.2**
Percent abundance of suspension feeders*	14.5	7.0	3.1	**14.2**	**16.9**
	9.3	4.9	4.8	**7.1**	**6.2**

[a] Result contrary to ecological principle.
[b] Cannot interpret metric direction; the species *Mediomastus ambiseta* and the genus *Capitella* respond differently to pollution.
[c] Cannot interpret metric direction; interface feeders include spionids and some bivalves (e.g., *Macoma*) that respond differently to pollution.

erence-value ranges for each metric-habitat combination and the application of a scoring system to these ranges. Procedures were based on modifications of those described by Weisberg et al. (1997). Threshold values were established as the 10th and 50th (median) percentile values of reference sites for each metric-habitat combination. This procedure differed from Weisberg et al. (1997), who used 5th and 50th percentiles. For each metric, a value falling below the 10th percentile threshold was considered to deviate strongly from values at the best reference sites in the same habitat. An upper threshold corresponding to the 90th percentile was used for some metrics (e.g., percent abundance of pollution-indicative taxa) because the direction of the response is such that higher percentages are expected in degraded sites than in reference sites. Fifth and 95th percentile thresholds as used by Weisberg et al. (1997) were also calculated, but a high percentage of misclassifications for various metrics were found when these thresholds were applied.

Abundance and biomass respond bimodally to pollution (Pearson and Rosenberg 1978; Weisberg et al. 1997). An increase in the abundance and/or biomass of organisms is expected at polluted sites when the stress from pollution is moderate, such at sites where there is organic enrichment of the sediment. A decrease in the abundance and biomass of organisms is expected at sites with high degrees of stress from pollution. For these two metrics, an upper threshold corresponding to the 90th percentile value of reference sites was established in addition to the lower threshold corresponding to the 10th percentile.

A score of 1 was used if the value of the metric for the site being evaluated was below the 10th percentile of corresponding reference values, a score of 3 was used for values between the 10th percentile and the median, and a score of 5 was used for values above the median. For metrics with upper thresholds (see above), the scoring was reversed so that a score of 1 was used for values above the 90th percentile, a score of 3 for values between the 90th percentile and the median, and a score of 5 for values below the median of the corresponding reference values. For abundance and biomass, a score of 1 was used if the value of the metric for the site being evaluated was below the 10th percentile or above the 90th percentile, a score of 3 for values between the 10th and 25th or between the 75th and 90th percentiles, and a score of 5 for values between the 25th and 75th percentiles of corresponding reference values. A score of 1 was considered to indicate impaired condition.

In some cases, two modifications to the above scoring system were introduced. These procedures also differed from procedures described by Weisberg et al. (1997). First, metrics based on percentages or tolerance values (all of the metrics listed in Table 2 except abundance, biomass, Shannon diversity, and number of taxa) were not scored when there was no fauna. We did this because the inclusion of azoic sites would have exaggerated the response of metrics that are based on percentages. Second, pollution-sensitive taxa metrics (i.e., percent abundance of pollution-sensitive taxa, percent abundance of pollution-sensitive NC taxa, and percent abundance of pollution-sensitive oligohaline taxa) that would score 3 or 5 when the overall abundance at a site is low (i.e., $\leq 20\%$ of the lower abundance threshold) were not scored. This was done to avoid high scores due to the presence of a few individuals of pollution-sensitive species found among a small number of total individuals within a sample (Weisberg et al. 1997).

Selection of Metrics

Using the scoring system described above, classification efficiencies (percent sites correctly classified) for each of the metrics selected during statistical testing were examined to determine which metrics performed best (Table 3). Most metrics classified non-degraded sites very well, typically better than 90% correct classifications, but degraded sites were generally classified at much lower levels. Those metrics with at least 50% correct classifications for the degraded sites were selected for final evaluation. This approach reduced the risk of Type II errors (i.e., degraded stations classified as non-degraded). The exception to this rule were metrics in the tidal freshwater habitat, for which this criterion was relaxed due to the generally lower classification efficiencies observed for this habitat. Percent abundance of pollution-indicative oligohaline taxa and percent abundance of Tubificidae which met this rule were considered redundant of other metrics in the oligohaline habitat and were excluded from further consideration. Twelve metrics, some performing well in more than one habitat (Table 3), were selected to determine which combination of metrics represented the best index.

Combination of Metrics into an Index

In the final phase of the index development process, the calibration and validation data sets were used to evaluate which combination of metrics best distinguished between degraded and non-degraded conditions for each of the five habitats. For each habitat, metrics were combined in all possible ways and the resulting combinations of metrics were arranged in order according to how well they correctly classified sites as degraded and non-degraded in the calibration and validation data sets. Four criteria of acceptance were then established. The metric combination that correctly classified sites in the calibration data set within 5% of the most efficient combination, correctly classified sites in the validation data set within 10% of the most efficient combination, had the most number of metrics, and had a variety of functional categories, was selected to represent the final index. Only combinations that did not include biomass were considered. Biomass was excluded because it was not determined by all the sampling programs in the MAIA region,

TABLE 3. Classification efficiencies (percent of sites correctly classified) on calibration data for each of the metrics that distinguished between reference and degraded sites in a direction consistent with established ecological principles. Asterisks indicate metrics selected for final testing in various combinations (see text).

Habitat/Metric	Percent Sites Correctly Classified		
	Non-degraded	Degraded	Total
Tidal Freshwater			
Abundance*	79.2	76.2	77.8
Percent abundance of pollution-indicative taxa	91.7	14.3	55.6
North Carolina Sensitivity Index*	91.7	38.1	66.7
Tolerance Score	91.7	23.8	60.0
Tanypodinae-Chironomidae percent abundance ratio*	91.7	38.1	66.7
Percent abundance of Tubificidae	91.7	28.6	62.2
Percent abundance of Amphipoda	100.0	0.0	53.3
Percent abundance of carnivore-omnivores	91.7	28.6	62.2
Percent abundance of interface feeders	100.0	0.0	53.3
Percent abundance of deep-deposit feeders*	91.7	38.1	66.7
Oligohaline			
Shannon-Wiener diversity	92.0	46.7	75.0
Number of taxa	92.0	26.7	67.5
Percent dominance*	92.0	53.3	77.5
Percent abundance of pollution-indicative taxa*	92.0	53.3	77.5
Percent abundance of pollution-sensitive taxa	91.7	46.7	74.4
Percent abundance of pollution-sensitive NC taxa	91.7	46.7	74.4
Percent abundance of pollution-indicative oligohaline taxa	92.0	53.3	77.5
Percent abundance of pollution-sensitive oligohaline taxa	100.0	0.0	61.5
Tolerance Score	92.0	46.7	75.0
Tanypodinae-Chironomidae percent abundance ratio*	92.0	53.3	77.5
Percent abundance of Tubificidae	92.0	60.0	80.0
Percent abundance of Spionidae	100.0	0.0	62.5
Percent abundance of carnivore-omnivores	92.0	46.7	75.0
Percent abundance of deep-deposit feeders*	92.0	60.0	80.0
Low Mesohaline			
Abundance*	81.1	73.0	77.0
Biomass*	82.1	94.3	88.9
Shannon-Wiener diversity*	91.9	81.1	86.5
Number of taxa*	91.9	86.5	89.2
Percent dominance*	91.9	70.0	84.2
Percent abundance of pollution-indicative taxa*	91.9	50.0	77.2
Percent abundance of pollution-sensitive NC taxa	91.9	47.4	76.8
North Carolina Sensitivity Index*	91.9	70.0	84.2
Percent abundance of Amphipoda	100.0	0.0	64.9
Percent abundance of Bivalvia	91.9	40.0	73.7
Percent abundance of carnivore-omnivores	91.9	40.0	73.7
High Mesohaline			
Abundance*	80.5	55.4	69.7
Biomass *	80.0	69.0	75.0
Shannon-Wiener diversity*	90.8	72.3	82.9
Number of taxa*	90.8	67.7	80.9
Percent dominance*	90.8	62.8	80.4
Percent abundance of pollution-indicative taxa	90.8	43.1	73.2
Percent abundance of pollution-sensitive taxa*	90.7	62.5	80.6
Percent abundance of pollution-sensitive NC taxa	90.7	49.0	75.6
North Carolina Sensitivity Index*	90.8	51.0	76.1
Percent abundance of Tubificidae	90.8	25.5	66.7
Percent abundance of Amphipoda	100.0	0.0	63.0
Percent abundance of Bivalvia	90.8	39.2	71.7
Percent abundance of carnivore-omnivores	90.8	41.2	72.5
Percent abundance of suspension feeders	100.0	0.0	63.0
Polyhaline			
Abundance*	80.3	66.7	77.9
Biomass*	79.4	70.8	77.6
Shannon-Wiener diversity*	90.6	59.3	85.1
Number of taxa*	92.9	66.7	88.3
Percent dominance	90.6	47.8	84.0
Percent abundance of pollution-sensitive taxa	90.4	35.3	83.8
Percent abundance of pollution-sensitive NC taxa	90.3	38.9	83.8
North Carolina Sensitivity Index	90.6	43.5	83.3
Percent abundance of Amphipoda	100.0	0.0	84.7
Percent abundance of Bivalvia*	90.6	69.6	87.3
Percent abundance of deep-deposit feeders	90.6	26.1	80.7
Percent abundance of suspension feeders*	90.6	52.2	84.7

and therefore some data sets did not include biomass measurements.

The index value for a site was computed by averaging the scores of the individual metrics (range: 1–5). Index values < 3.0 were considered to indicate stressed benthic assemblages indicative of degraded conditions.

Results and Discussion

The index selected for use in the MAIA region consisted of three metrics in each of the tidal freshwater and oligohaline habitats, five metrics in each of the low mesohaline and high mesohaline habitats, and four metrics in the polyhaline habitat. Table 4 lists the selected metrics and thresholds. Two metrics were percent abundance contributions of species indicators: percent abundance of pollution-sensitive taxa (Table 5) and percent abundance of pollution-indicative taxa (Table 6). One metric was the percent abundance contribution of deep-deposit feeders: all the Oligochaeta and the polychaetes *Heteromastus filiformis* and *Mediomastus* spp. This last metric provided information about the trophic composition of the community.

The index selected for the MAIA region correctly classified 83% of the sites in the calibration data set and 82% of the sites in the validation data set. The lowest overall classification efficiency by habitat was for the oligohaline using the calibration data set (77% of the sites correctly classified) and for the tidal freshwater using the validation data set (61% of the sites correctly classified; Table 7). Although we used both calibration and validation data in the metric combination selection process, validation data were independent of those used to derive the metric thresholds. The usefulness of the index should then be evaluated according to how well it classifies sites of known sediment quality in the independent data set. The index correctly classified a high percentage of sites in the independent data set, but classification efficiencies varied substantially across habitats. Classification efficiencies of sites during the validation phase were higher in the high-salinity habitats (low mesohaline, high mesohaline, and polyhaline; range = 81–92%) than in the low-salinity habitats (tidal freshwater and oligohaline; range = 61–71%). Chesapeake Bay B-IBI estimates were also found to be low for sites in tidal freshwater and oligohaline habitats (Weisberg et al. 1997; Alden et al. 2002). The development and application of biocriteria is most difficult in the tidal freshwater and oligohaline regions because of the high variability in benthic community composition in those habitats (Dauer 1993). This variability may be partially due to the various natural stresses in low-salinity habitats, which make the application of biological indices less reliable than those established for more saline habitats. The classification performance for single metrics in the tidal freshwater and oligohaline habitats was relatively poor, but it improved when the metrics were combined. Classification efficiencies on validation data in these habitats were low for the degraded sites (Table 7), about what it would be expected from chance alone. One possible factor contributing to the relatively low classification efficiency in the low-salinity habitats is sample size. The number of samples available for selection of thresholds in the tidal freshwater and oligohaline habitats was particularly low. Better results may be achieved if a larger array of samples were used for the reference distributions, which might improve threshold definition and reduce the number of misclassifications.

In the tidal freshwater and oligohaline habitats, the fauna is often dominated by tubificid oligochaetes, and a variety of midges are present and can be abundant. In the MAIA index and other index development efforts (Weisberg et al. 1993, 1997; Engle et al. 1994; Eaton 2001; Paul et al. 2001) tubificids and many chironomids have been categorized as indicative of organic and toxic pollution. This does not mean that low salinity habitats are of lesser quality than high salinity habitats where tubificids are less abundant and chironomids usually absent. What matters is the relative percentage of these taxa in a sample. Indicator species usually dominate heavily the community in situations of organic enrichment or where toxic pollution eliminates other species. In the MAIA index, the relative abundance of pollution-indicative taxa in the oligohaline habitat must be > 75% for this metric to indicate impaired conditions, and in the tidal freshwater the abundance of deep-deposit feeders (mostly tubificids) must be > 90%. The classification of most oligochaetes and many chironomids as indicative of pollution may add to the low efficiency problem of biological indices in low salinity environments. Although a wide range of sensitivities to pollution has been recognized for insect larvae (Hilsenhoff 1987; Lenat 1993; Eaton 2001), a better understanding of the biology of oligochaetes and chironomids in relation to pollution might help improve assessments in estuaries.

The combinations of metrics that were used to represent the index were among those with the highest percentages of correct classifications. They were not the best performing in terms of highest number of sites correctly classified as degraded or non-degraded. Top classification efficiencies on validation data were 70% in the tidal freshwater, 76% in the oligohaline, 92% in the low mesohaline, 84% in the high mesohaline, and 92% in the polyhaline habitat. Combinations with slightly low-

TABLE 4. Thresholds used to score each metric of the Mid-Atlantic Integrated Assessment (MAIA) index.

Habitat/Metric	Scoring Criteria		
	1	3	5
Tidal Freshwater			
Abundance (no. m^{-2})	<1,401.5 or ≥7,068.2	1,401.5–1,772.7 or ≥5,253.8–7,068.2	≥1,772.7–5,253.8
Tanypodinae-Chironomidae percent abundance ratio	>69.8	20.9–69.8	≤20.9
Percent abundance of deep-deposit feeders	>90.6	61.7–90.6	≤61.7
Oligohaline			
Percent dominance	<10.4	10.4–30.0	≥30.0
Percent abundance of pollution-indicative taxa	>75.8	35.5–75.8	≤35.5
Percent abundance of deep-deposit feeders	>64.3	38.9–64.3	≤38.9
Low Mesohaline			
Abundance (no. m^{-2})	<818.2 or ≥5,151.5	818.2–1,590.9 or ≥3,659.1–5,151.5	≥1,590.9–3,659.1
Shannon-Wiener diversity (log_2)	<1.5	1.5–2.4	≥2.4
Number of taxa	<6.3	6.3–10.3	≥10.3
Percent dominance	<18.7	18.7–35.8	≥35.8
Percent abundance of pollution-indicative taxa	>39.5	7.5–39.5	≤7.5
High Mesohaline			
Abundance (no. m^{-2})	<636.4 or ≥6,909.1	636.4–1,174.2 or ≥4,159.1–6,909.1	≥1,174.2–4,159.1
Shannon-Wiener diversity (log_2)	<1.7	1.7–2.6	≥2.6
Number of taxa	<6.0	6.0–11.7	≥11.7
Percent dominance	<22.4	22.4–41.9	≥41.9
Percent abundance of pollution-sensitive taxa	<2.9	2.9–21.3	≥21.3
Polyhaline			
Abundance (no. m^{-2})	<647.7 or ≥8,719.7	647.7–1,454.5 or ≥5,704.5–8,719.7	≥1,454.5–5,704.5
Shannon-Wiener diversity (log_2)	<2.0	2.0–3.1	≥3.1
Number of taxa	<7.0	7.0–19.3	≥19.3
Percent abundance of Bivalvia	<0.44	0.44–7.5	≥7.5

TABLE 5. Taxa defined as pollution-sensitive in the Mid-Atlantic Integrated Assessment (MAIA) index. Taxa listed at the genus level include all species.

Annelida: Polychaeta	*Listriella clymenellae*
Bhawania heteroseta	Mollusca
Chaetopterus variopedatus	*Acteocina canaliculata*
Clymenella torquata	*Anadara ovalis*
Diopatra cuprea	*Anadara transversa*
Glycera americana	*Ensis directus*
Glycinde solitaria	*Macoma balthica*
Loimia medusa	*Mercenaria mercenaria*
Macroclymene zonalis	*Rangia cuneata*
Marenzelleria viridis	*Spisula solidissima*
Mediomastus ambiseta	*Tagelus divisus*
Mesochaetopterus taylori	*Tagelus plebeius*
Nephtys picta	*Tellina agilis*
Sabaco elongatus	Unionidae (all species)
Spiochaetopterus costarum	Phoronida
Spiophanes bombyx	*Phoronis* spp.
Arthropoda	Echinodermata
Cyathura polita	*Microphiopholis atra*
Ampelisca verrilli	

TABLE 6. Taxa defined as pollution-indicative in the Mid-Atlantic Integrated Assessment (MAIA) index. Taxa listed at the genus level include all species.

Annelida: Polychaeta	Mollusca
Eteone heteropoda	*Mulinia lateralis*
Leitoscoloplos fragilis	Arthropoda: Insecta
Paraprionospio pinnata	*Ablabesmyia parajanta*
Streblospio benedicti	*Chaoborus* spp.
Annelida: Oligochaeta	*Chironomus* spp.
Branchiura sowerbyi	*Clinotanypus* spp.
Dero spp.	*Coelotanypus* spp.
Enchytraeidae (all species)	*Cryptochironomus* spp.
Ilyodrilus templetoni	*Dicrotendipes* spp.
Isochaetides spp.	*Endochironomus* spp.
Limnodrilus spp.	*Glyptotendipes* spp.
Quistradrilus multisetosus	*Harnischia* spp.
Telmatodrilus vejdovskyi	*Kiefferulus* spp.
Tubifex spp.	*Microchironomus* spp.
Tubificidae immature with capilliform chaetae	*Parachironomus* spp.
Tubificidae immature without capilliform chaetae	*Procladius* spp.
Tubificoides spp.	*Tanypus* spp.

er classification efficiencies but with larger number of components and categories of biotic response were favored over combinations with fewer metrics. We reasoned that the use of more metrics was better than fewer metrics because the benthos often show a staged response to stress in which the various components of the assemblage respond differently to different levels of disturbance (Pearson and Rosenberg 1978). Categories that indicate different aspects of benthic community structure and function provide more robust information about the developmental stage of the community.

While applying the final selection criteria, we selected against metric combinations that included biomass. Biomass measurements are very useful for various reasons. For example, biomass data provide information about the presence of large-sized species in the community, which is often indicative of a past history of good sediment quality conditions (Warwick 1986; Dauer 1993). Because biomass data were unavailable for North Carolina estuaries or Delaware Bay, the reference range distribution of biomass values excluded a large portion of the MAIA region. This would represent a problem if a MAIA index that included biomass as a metric were used in a region for which the reference range distribution had been inappropriately estimated. If the index included biomass and the biomass component were dropped in a subsequent field application, the resulting partial index might end up having an unacceptable high rate of misclassifications. Good alternative metric combinations with high classification efficiencies were available and were used in the selection process. Biomass was one of the least effective metrics at distinguishing between degraded and non-degraded conditions. Inclusion of biomass in the final index would have resulted only in small improvements. Most effective metrics on validation data were number of taxa and Shannon diversity in the high salinity habitats, deep-deposit feeding taxa in the tidal freshwater, and pollution-indicative taxa in the oligohaline.

One aspect of the index development process that has been under emphasized in other studies is the classification performance for the degraded sites. In the MAIA index development, special attention was given to attaining percent correct classifications for the degraded sites of at least 50%. Classification efficiencies for the reference sites in the calibration phase of the index were expected to be high around 90%, as the thresholds used to score the metrics were based on the distribution of values at reference sites. Classification efficiencies for the degraded sites are usually much lower be-

TABLE 7. Classification efficiencies (percent of sites correctly classified) of the Mid-Atlantic Integrated Assessment (MAIA) index for sites from the calibration and validation data sets within each habitat class.

		Percent Sites Correctly Classified		
Data Set	Habitat/Metric	Non-degraded	Degraded	Total
Calibration	Tidal Freshwater	91.7	71.4	82.2
	Oligohaline	92.0	53.3	77.5
	Low Mesohaline	83.8	89.2	86.5
	High Mesohaline	86.2	78.5	82.9
	Polyhaline	85.0	70.4	82.5
Validation	Tidal Freshwater	66.7	54.6	60.9
	Oligohaline	84.6	50.0	71.4
	Low Mesohaline	89.5	94.7	92.1
	High Mesohaline	84.1	75.8	80.5
	Polyhaline	88.9	76.9	86.8

cause not all values of a metric at a degraded site would be expected to fall below the 10th percentile of corresponding reference values, especially in communities that are stressed by natural factors and show responses to natural disturbance that cannot be distinguished from those due to pollution. If only overall classifications were examined, it would be quite possible to select a metric that correctly classified most reference sites while failing to identify every single degraded site. This method would be neither sensitive nor useful to environmental managers since they would be most interested in knowing where the problems occurred. Although classification efficiencies for the oligohaline and tidal freshwater habitats were not as high as we had hoped, the MAIA index appears to be quite reliable in the more saline habitats, with a high likelihood of correctly identifying both degraded and non-degraded conditions. Caution should be used when applying the index to samples collected from oligohaline and tidal freshwater environments.

The MAIA index provides an uniform measure with which to make comparisons of benthic condition in different estuaries within the mid-Atlantic region of the U.S. Development of this new index is based on benthic data from several estuaries in the region, thereby widening its potential application beyond that of any index previously developed in the mid-Atlantic. The application of indices of biotic integrity beyond the region for which they were developed may be compromised by the large natural variability inherent in estuarine systems. Index development efforts should incorporate data that are representative of the full range of biological, physical, chemical, and hydrodynamic characteristics of the region of application. Existing indices (e.g., Weisberg et al. 1993, 1997; Van Dolah et al. 1999; Paul et al. 2001) do not cover all parts of the MAIA region. The MAIA index has incorporated geographic variation by using data collected by several state and federal monitoring programs that have sampled benthic assemblages throughout the region and in a wide variety of estuarine environments. The usefulness of the MAIA index is that it combines wide geographical application with the simplicity of approach, easiness of calculation and interpretation, and effectiveness of the B-IBI method. The index is expected to be of great utility in regional assessments as a tool for evaluating the integrity of benthic assemblages and tracking their condition over time.

Acknowledgments

The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development, funded and managed the research described here under contract number 68-D-98-002 to Technology Planning and Management Corporation via a subcontract with Versar, Inc. Participants of a benthic workshop helped clarify issues and provided valuable input in the index development process. We especially thank Richard Batiuk (Chesapeake Bay Program), Ananda Ranasinghe (Southern California Coastal Research Program), and Elizabeth Ebersole (Maryland Department of Natural Resources) for critically reviewing the manuscript. Thuzar Myint managed the database and conducted the analysis of data. Virginia Engle (EPA Gulf Ecology Division) and an anonymous reviewer provided valuable peer-review comments. This manuscript has been subjected to the EPA's review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.

Literature Cited

Adams, D. A., J. S. O'Connor, and S. B. Weisberg. 1998. Sediment quality of the NY/NJ Harbor system. Final report. An investigation under the Regional Environmental Monitoring and Assessment Program (REMAP). EPA/902/R-98/001. U.S. Environmental Protection Agency—Region 2, Division of Environmental Science and Assessment, Edison, New Jersey.

Alden, III, R. W., 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.

American Society for Testing Materials. 1993. Guide for conducting 10-day static sediment toxicity tests with marine and estuarine infaunal amphipods. ASTM E-1367-92. American Society for Testing Materials, Philadelphia, Pennsylvania.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15:185–211.

Bradley, M. P. and R. B. Landy. 2000. The mid-Atlantic integrated assessment (MAIA). *Environmental Monitoring and Assessment* 63:1–13.

Bricker, S. B., C. G. Clement, D. E. Pirhalla, S. P. Orlando, and D. R. G. Farrow. 1999. National Estuarine Eutrophication Assessment: Effects of Nutrient Enrichment in the Nation's Estuaries. National Oceanic and Atmospheric Administration, National Ocean Service, Silver Spring, Maryland.

Bulich, A. A., M. W. Greene, and D. L. Isenberg. 1981. Reliability of the bacterial luminescence assay for determining toxicity of pure compounds and complex effluent, p. 338–347. *In* D. R. Branson and K. L. Dickson (eds.). Aquatic Toxicology and Hazard Assessment, Proceedings of the Fourth Annual Symposium, ASTM STP 737. American Society for Testing and Materials, Philadelphia, Pennsylvania.

Chaillou, J. C., S. B. Weisberg, F. W. Kutz, T. E. DeMoss, L. Mangiaracina, R. Magnien, R. Eskin, J. Maxted, K. Price, and J. K. Summers. 1996. Assessment of the ecological condition of the Delaware and Maryland coastal bays. EPA/620/R-96/004. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

Culliton, T. J., M. A. Warren, T. R. Goodspeed, D. G. Remer, C. M. Blackwell, and J. J. McDonough, III. 1990. 50 years of population change along the nation's coast. Second report of a Coastal Trend Series. National Oceanic and Atmospheric Administration, National Ocean Service, Rockville, Maryland.

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

Dauer, D. M., M. F. Lane, H. G. Marshall, and K. E. Carpenter. 1998. Status and trends in water quality and living resources in the Virginia Chesapeake Bay: 1985–1997. Prepared for the Virginia Department of Environmental Quality, Richmond, Virginia. Old Dominion University, Norfolk, Virginia.

EATON, L. 2001. Development and validation of biocriteria using benthic macroinvertebrates for North Carolina estuarine waters. *Marine Pollution Bulletin* 42:23–30.

ENGLE, V. D. AND J. K. SUMMERS. 1999. Refinement, validation, and application of a benthic condition index for Northern Gulf of Mexico estuaries. *Estuaries* 22:624–635.

ENGLE, V. D., J. K. SUMMERS, AND G. G. GASTON. 1994. A benthic index of environmental condition of Gulf of Mexico estuaries. *Estuaries* 17:372–384.

HARDING, L. W. 1994. Long-term trends in the distribution of phytoplankton in Chesapeake Bay: Roles of light, nutrients and stream flow. *Marine Ecology Progress Series* 104:267–291.

HILSENHOFF, W. L. 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomologist* 20:31–39.

HYLAND, J. L., L. BALTHIS, C. T. HACKNEY, G. MCRAE, A. H. RINGWOOD, T. R. SNOOTS, R. F. VAN DOLAH, AND T. L. WADE. 1998. Environmental quality of estuaries of the Carolinian Province: 1995. Annual statistical summary for the 1995 EMAP-Estuaries Demonstration Project in the Carolinian Province. NOAA Technical Memorandum NOS ORCA 123. National Oceanic and Atmospheric Administration, National Ocean Service, Office of Ocean Resources Conservation and Assessment, Silver Spring, Maryland.

HYLAND, J. L., W. L BALTHIS, C. T. HACKNEY, AND M. POSEY. 2000. Sediment quality of North Carolina estuaries: An integrative assessment of sediment contamination, toxicity, and condition of benthic fauna. *Journal of Aquatic Ecosystem Stress and Recovery* 8:107–124.

KARR, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21–27.

KARR, J. R., K. D. FAUSCH, P. L. ANGERMEIER, P. R. YANT, AND I. J. SCHLOSSER. 1986. Assessing biological integrity in running waters: A method and its rationale. Special Publication 5. Illinois Natural History Survey, Urbana, Illinois.

LENAT, D. R. 1993. A biotic index for the southeastern United States: Derivation and list of tolerance values, with criteria for assigning water-quality ratings. *Journal of the North American Benthological Society* 12:279–290.

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.

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.

MALONE, T. C. 1987. Seasonal oxygen depletion and phytoplankton production in Chesapeake Bay: Preliminary results of 1985–86 field studies, p. 54–60. *In* G. B. Mackiernan (ed.). Dissolved Oxygen in the Chesapeake Bay: Processes and Effects. Maryland Sea Grant, College Park, Maryland.

MALONE, T. C., D. J. CONLEY, T. R. FISHER, P. M. GILBERT, L. W. HARDING, AND K. G. SELLNER. 1996. Scales of nutrient-limited phytoplankton productivity in Chesapeake Bay. *Estuaries* 19: 371–385.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA). 1994. Assessment of Chemical Contaminants in the Chesapeake and Delaware Bays. NOAA, Coastal Monitoring and Bioeffects Assessment Division, National Status and Trends Program, Silver Spring, Maryland.

ODUM, E. P. 1985. Trends expected in stressed ecosystems. *Bioscience* 35:419–422.

PAERL, H. W., J. L. PINCKNEY, J. M. FEAR, AND B. L. PEIERLS. 1998. Ecosystem responses to internal and watershed organic matter loading: Consequences for hypoxia in the eutrophying Neuse River estuary, North Carolina, USA. *Marine Ecology Progress Series* 166:17–25.

PAUL, J. F., J. H. GENTILE, K. J. SCOTT, S. C. SCHIMMEL, D. E. CAMPBELL, AND R. W. LATIMER. 1999. EMAP-Virginian Province four-year assessment report (1990–93). EPA 600/R-99/004. U.S. Environmental Protection Agency, Atlantic Ecology Division, Narragansett, Rhode Island.

PAUL, J. F., K. J. SCOTT, D. E. CAMPBELL, J. H. GENTILE, C. S. STROBEL, R. M. VALENTE, S. B. WEISBERG, A. F. HOLLAND, AND J. A. RANASINGHE. 2001. Developing and applying a benthic index of estuarine condition for the Virginian Biogeographic Province. *Ecological Indicators* 1:83–99.

PEARSON, T. H. AND R. ROSENBERG. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology Annual Review* 16:229–311.

PENNOCK, J. R., J. H. SHARP, AND W. W. SCHROEDER. 1994. What controls the expression of estuarine eutrophication? Case studies of nutrient enrichment in the Delaware Bay and Mobile Bay estuaries, USA, p. 139–146. *In* K. R. Dyer and R. J. Orth (eds.). Changes in Fluxes in Estuaries: Implications from Science to Management. Olsen and Olsen, Fredensborg, Denmark.

RANASINGHE, J. A., L. C. SCOTT, AND F. S. KELLEY. 1999. Chesapeake Bay water quality monitoring program, long-term benthic monitoring and assessment component. Level 1 Comprehensive report, July 1984–December 1998. Prepared for Maryland Department of Natural Resources, Tidewater Ecosystem Assessments. Versar, Inc., Columbia, Maryland.

RINGWOOD, A. H., M. E. DELORENZO, P. E. ROSS, AND A. F. HOLLAND. 1997. Interpretation of Microtox solid-phase toxicity tests: The effects of sediment composition. *Environmental Toxicology and Chemistry* 16:1135–1140.

STROBEL, C. J. 1998. Environmental Monitoring and Assessment Program, Mid-Atlantic Integrated Assessment, Estuaries Component. Field Operations and Safety Manual. U.S. Environmental Protection Agency, Atlantic Ecology Division, Narragansett, Rhode Island.

TUTTLE, J. H., R. B. JONAS, AND T. C. MALONE. 1987. Origin, development and significance of Chesapeake Bay anoxia, p. 442–472. *In* S. K. Majumdar, L. W. Hall, Jr., and H. M. Austin (eds.). Contaminant Problems and Management of Living Chesapeake Bay Resources. Pennsylvania Academy of Science, Philadelphia, Pennsylvania.

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.

WARWICK, R. M. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology* 92:557–562.

WEISBERG, S. B., J. B. FRITHSEN, A. F. HOLLAND, J. F. PAUL, K. J. SCOTT, K. J. SUMMERS, H. T. WILSON, R. M. VALENTE, D. G. HEIMBUCH, J. GERRITSEN, S. C. SCHIMMEL, AND R. W. LATIMER. 1993. EMAP-Estuaries Virginian Province 1990 Demonstration Project Report. EPA/620/R-93/006. U.S. Environmental Protection Agency, Narragansett, Rhode Island.

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.

WILSON, J. G. AND D. W. JEFFREY. 1994. Benthic biological pollution indices in estuaries, p. 311–327. *In* J. M. Krammer (ed.). Biomonitoring of Coastal Waters and Estuaries. CRC Press, Boca Raton, Florida.

Received for consideration, September 26, 2001
Accepted for publication, May 17, 2002