ENVIRONMETRICS
Environmetrics 2002; **13**: 499–511 (DOI: 10.1002/env.529)

Application of two indices of benthic community condition in Chesapeake Bay

J. Ananda Ranasinghe[1,*,†], Jeffrey B. Frithsen[2], Frederick W. Kutz[3], John F. Paul[4], David E. Russell[3], Richard A. Batiuk[5], Jeffrey L. Hyland[6], John Scott[7] and Daniel M. Dauer[8]

[1]*Southern California Coastal Water Research Project, 7171 Fenwick Lane, Westminster, CA 92683-5218, U.S.A.*
[2]*National Center for Environmental Assessment, Office of Research and Development (8601), U.S. Environmental Protection Agency, 401 M Street, SW, Washington, DC 20460, U.S.A.*
[3]*Environmental Science Center, U.S. Environmental Protection Agency, 701 Mapes Road, Fort Meade, MD 20755-5350, U.S.A.*
[4]*Atlantic Ecology Division, U.S. Environmental Protection Agency, 27 Tarzwell Dr., Narragansett, RI 02882, U.S.A.*
[5]*Chesapeake Bay Program, U.S. Environmental Protection Agency, 410 Severn Avenue, Annapolis, MD 21403, U.S.A.*
[6]*NOAA, National Ocean Service, 219 Ft. Johnson Road, Charleston, SC 29412-9110, U.S.A.*
[7]*Science Applications International Corporation, 165 Dean Knauss Drive, Narragansett, RI 02882, U.S.A.*
[8]*Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, U.S.A.*

SUMMARY

The Chesapeake Bay Benthic Index of Biotic Integrity (B-IBI) and the Environmental Monitoring and Assessment Program's Virginian Province Benthic Index (EMAP-VP BI) were applied to 294 sampling events in Chesapeake Bay and the results were compared. These estuarine benthic indices are intended to identify benthic invertebrate assemblages that have been degraded by low dissolved oxygen concentrations or high concentrations of chemical contaminants. The B-IBI includes several community measures and weights them equally using a simple scoring system that compares them against values expected for undegraded sites. It includes 11 measures of species diversity, productivity, indicator species and trophic composition. The EMAP-VP BI uses discriminant function coefficients to weight contributions of species diversity and the abundances of two indicator families. The two indices agreed on degraded or undegraded classifications for benthos at 81.3% of the sites. This level of agreement is within the level of accuracy achieved during index development and, therefore, may approach the limits that can be achieved. The indices were strongly associated (Pearson's $r = 0.75$). The B-IBI was more conservative than the EMAP-VP BI, classifying 72.7% of the disagreements as degraded. The 55 sites where the indices disagreed were distributed in different habitats throughout the Bay except polyhaline sand. Many of the classification disagreements were at sites with index values close to, but on opposite sides of, the degraded–undegraded thresholds, with 49.1% of the B-IBI values within 0.5 units and 81.8% within 1.0 units; the corresponding values for sites where both indices agreed were only 23.4% and 62.7%, respectively. The pattern for the EMAP-VP BI was similar, with 61.8% and 74.6% of disagreements and only 18.8% and 38.9% of agreements within 0.5 and 1.0 units of the threshold. Although the close agreement suggests that either index is suitable for evaluating the benthic condition, the B-IBI offers some additional advantages. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: estuarine benthic invertebrates; multi-habitat biological assessment; benthic index; index of biological integrity; Chesapeake Bay

*Correspondence to: J. Ananda Ranasinghe, Southern California Coastal Water Research Project, 7171 Fenwick Lane, Westminster, CA 92683-5218, U.S.A.
†E-mail: anandar@sccwrp.org

Received 12 April 2001
Accepted 14 December 2001

1. INTRODUCTION

Assessment of the condition of benthic invertebrate communities in estuaries has progressed in recent years through the development of benthic indices. Previously, estuarine and marine benthic assessments were restricted primarily to inferring impacts from nearfield–farfield comparisons within similar habitats (e.g. Green, 1979) because of variations in benthic community composition related to salinity, sediment, or other habitat differences (e.g. Boesch, 1973, 1977). The Abundance Biomass Comparison (ABC) method (Warwick, 1986; Warwick *et al.*, 1987) is potentially habitat-independent and was used successfully in a variety of marine habitats (Austen *et al.*, 1989; Ritz *et al.*, 1989; Anderlini and Wear, 1992; Agard *et al.*, 1993) but only had limited success in estuaries (Meire and Deru, 1990; Dauer *et al.*, 1993). Recently, benthic indices applicable across habitat boundaries have been developed for estuaries and coastal areas in several geographic areas on the Atlantic (Weisberg *et al.*, 1993, 1997; Ranasinghe *et al.*, 1994; Paul *et al.*, 2001; Van Dolah *et al.*, 1999), Gulf (Engle *et al.*, 1994; Engle and Summers, 1999) and Pacific (Smith *et al.*, 2001) coasts of the United States. These indices are intended to identify degraded benthic invertebrate assemblages that are indicative of low dissolved oxygen concentrations in bottom waters or high concentrations of chemical contaminants in sediments, which are common pollution effects in estuaries.

Benthic indices are developed by (a) defining criteria for degraded and undegraded sites based on non-biological measures such as bottom-water dissolved oxygen and sediment contaminant concentrations, (b) identifying biological measures which respond to (differ among) degraded and undegraded sites, (c) adjusting these responses for habitat differences, if necessary, (d) combining responsive measures into an index, and (e) validating the index using independent data. The indices provide a valid way of simplifying and communicating complex data. They are especially relevant to management efforts because benthic invertebrates are site-specific indicators of habitat conditions that integrate stress effects over time and over multiple types of stress (e.g. Gray, 1979).

Two benthic indices have been developed for application in Chesapeake Bay; each was developed using a different approach. The Virginian Province Benthic Index (EMAP-VP BI) was developed for the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (Weisberg *et al.*, 1993; Paul *et al.*, 2001) using a discriminant analysis (DA) approach. It was developed for application in estuaries on the Atlantic coast of the U.S.A. from Cape Cod, Massachusetts to Cape Henry, Virginia, including Chesapeake Bay. The Chesapeake Bay Benthic Index of Biotic Integrity (B-IBI) was developed for the Chesapeake Bay Program (Ranasinghe *et al.*, 1994, modified by Weisberg *et al.*, 1997 and Alden *et al.*, 2002) using an Index of Biotic Integrity (IBI) approach (Karr, 1991). It was developed for application in Chesapeake Bay.

The DA approach and the IBI approach differ in at least three ways. They differ in how expectations for undegraded assemblages are modified to compensate for changes in habitat, how measures to differentiate undegraded from degraded assemblages are selected, and how selected measures are combined to produce a diagnostic, quantitative index. The DA approach (Weisberg *et al.*, 1993; Engle *et al.*, 1994; Paul *et al.*, 2001; Engle and Summers, 1999) adjusts for habitat by empirical normalization of measures that require adjustment, identifies diagnostic measures (metrics) through stepwise discriminant analysis, and combines metrics into an index by weighting them according to a canonical discriminant function. Empirical habitat normalization is usually achieved by polynomial regression. In contrast, the B-IBI approach (Weisberg *et al.*, 1997; Van Dolah *et al.*, 1999) stratifies habitats based on benthic assemblage differences, identifies diagnostic metrics and thresholds for each habitat through distribution differences, and combines metrics into an index by a process that uses a simple scoring system to weight all measures equally.

As a result of these differences in development approach, the B-IBI and the EMAP-VP BI incorporate different suites of biological measures. The B-IBI includes 11 measures of species diversity, productivity, indicator species and trophic composition; the specific measures that are included vary with and are optimized for each habitat. The Shannon–Wiener index is the measure of diversity that is used, and both abundance and biomass are included in the productivity and indicator species measures. In contrast, the EMAP-VP BI includes the same three measures in all habitats; none of them are based on biomass. It includes the Gleason index of diversity and the abundances of two families of indicator organisms: spionid polychaetes and tubificid oligochaetes.

The differences in approach and the suites of measures that are included in different benthic indices lead naturally to questions about whether their application yields different results. However, opportunities for comparison are rare because it is unusual to have more than one benthic index available for any particular area. For many estuaries, only a single benthic index is available. For many others, no benthic index has been developed yet.

The availability, in Chesapeake Bay, of two benthic indices developed using different approaches presented us with the opportunity of comparing results for the first time. We applied the DA approach based EMAP-VP BI and the IBI approach based B-IBI to the same set of samples and compared the results. Our specific objectives were to identify the frequency, magnitude and nature of differences in assessment of Chesapeake Bay sites as 'degraded' or 'undegraded' by the EMAP-VP BI and the Chesapeake B-IBI. We were also interested in identifying whether there were any reasons for selecting one index over the other for benthic assessments in Chesapeake Bay.

2. METHODS

2.1. The data

We used benthic macroinvertebrate, sediment chemistry, sediment toxicity, and water quality data collected by the U.S. EPA's Environmental Monitoring and Assessment Program (EMAP) at 294 sampling events in Chesapeake Bay from 1990 to 1993. The sites were distributed throughout the Bay using a randomly placed systematic grid (Holland, 1990; Paul *et al.*, 1992; Paul *et al.*, 1999). We only used sites that were sampled between 15 July and 30 September because both indices are applicable during this period (Weisberg *et al.*, 1997).

During each sampling event, three benthic macroinfaunal samples were collected using a 440-cm^2 Young grab, sieved through a 0.5-mm screen, and preserved in buffered, 10% formaldehyde with rose bengal. A 50-ml core from each grab was collected and stored frozen in a plastic bag for grain-size analysis. Sediment samples for chemistry and toxicity analysis were collected from additional grab samples by removing the top 2 cm of sediment with a Teflon® spatula into a clean glass jar with a Teflon® lid, and stored frozen. Dissolved oxygen and salinity were measured near the bottom using a SeaBird CTD.

In the laboratory, macroinvertebrates were identified to the lowest practical taxonomic level and counted. Biomass was determined as shell-free dry weight after drying at 60°C for 48 h; before measuring biomass, bivalves longer than 2 cm were shucked and smaller shells were removed by acidification in 10% HCl. Grain-size analysis measured the proportion of sands in the sediment as the fraction of dry-weight retained on a 63 μ sieve, while silt and clay content were determined by pipette analysis (Folk, 1974).

Sediment chemistry samples were analyzed for 24 polycyclic aromatic hydrocarbons, 18 congeners of polychlorinated biphenyls, DDTs, 11 chlorinated pesticides, tributyl tins, and 15 metals. Total organic carbon was determined using a CO_2 analyzer after acidifying the sediment with H_3PO_4. Sediment toxicity was measured using a ten-day acute, static, non-renewal *Ampelisca abdita* test following ASTM (1991) protocols. For each toxicity test, 200 ml of sediment sample was placed in a 1-litre glass test chamber and covered with 600 ml of seawater; five replicate test chambers with 20 organisms in each replicate were used for each sample. Additional detail on the sampling and laboratory methods employed is available in Paul *et al.* (1999).

2.2. The indices

For each of the 294 sampling events, we calculated values for the EMAP-Estuaries Virginian Province Benthic Index (EMAP-VP BI: Paul *et al.*, 2001) and the Chesapeake Bay B-IBI (B-IBI: Ranasinghe *et al.*, 1994b, modified by Weisberg *et al.*, 2002 and Alden *et al.*, 2002) using the field and laboratory data.

2.2.1. EMAP-VP BI. The EMAP-VP BI includes three measures: the Gleason Diversity Index (Gleason, 1922) and the abundances of tubificid oligochaetes and spionid polychaetes which are two families of annelid worms; some species in each family are known to be pollution indicative (e.g. Weisberg *et al.*, 1997). The Gleason Diversity Index and tubificid abundance are adjusted for salinity effects.

The EMAP-VP BI is calculated as:

$$\{1.389(\mathrm{PG} - 51.5)/28.4\} - \{0.651(\mathrm{NT} - 28.2)/119.5\} - \{0.375(\mathrm{SA} - 20.0)/45.4\}$$

where PG, the percent expected Gleason Index is given by

$$100^{*}\,\mathrm{G}/\{4.283 - (0.498^{*}\,\mathrm{Sal}) + (0.0542^{*}\,\mathrm{Sal}^2) - (0.00103^{*}\,\mathrm{Sal}^3)\}$$

NT, the salinity normalized tubificid abundance is given by

$$\text{the mean tubificid abundance per sample } - 500^{*}\,\mathrm{e}^{-15^{*}\mathrm{Sal}}$$

G is the Gleason (1922) Diversity Index for all three samples (number of species in all three samples divided by the natural logarithm of the total number of organisms in all three samples), Sal is bottom water salinity in parts per thousand, and SA is the mean abundance of spionids per sample.

The EMAP-VP BI is unbounded. Values below zero represent degraded benthic communities and values above zero represent undegraded conditions.

2.2.2. B-IBI. The B-IBI (Ranasinghe *et al.*, 1994 modified by Weisberg *et al.*, 1997 and Alden *et al.*, 2002) includes community-level attributes from several ecological classes. The attributes are the Shannon–Weaver Diversity Index, total abundance, total biomass, abundance and biomass of pollution-indicative taxa, abundance and biomass of pollution-sensitive taxa, abundance of carnivores and omnivores, abundance of deep deposit feeders, abundance weighted pollution tolerance score, and the percentage of chironomid larvae belonging to the Tanypodinae, a pollution sensitive tribe of chironomids. The B-IBI is calculated by scoring each of several attributes as either 5, 3 or 1 depending on whether the value of the attribute in a sample approximates, deviates slightly from, or deviates strongly from values at undegraded sites in similar habitats; the scores are averaged across attributes. Scoring thresholds were developed for each of seven salinity and sediment grain size benthic habitats in Chesapeake Bay. The attributes that are scored and the scoring thresholds vary from habitat to habitat.

The B-IBI is scaled from 1 to 5. Sites with values of 3 or more are considered to represent benthic communities approximating undegraded conditions. Values less than 3 represent degraded communities.

2.3. Data analysis

We compared the indices in four ways. First, we measured the level of agreement between the indices by comparing site classifications and correlation analysis. We evaluated the extent and severity of site classification differences in terms of 'degraded' and 'undegraded' designations for benthic communities as the proportion of sampling events for which the indices disagreed and the deviations of disagreeing index values from the degraded–undegraded threshold. We also measured the strength of associations among the indices using the non-parametric Spearman and parametric Pearson correlation coefficients. Second, we examined the spatial distribution of disagreements and their distribution across habitats to identify whether disagreements were more likely in any particular geographic area or any particular habitat. Third, we evaluated index disagreements relative to three measures of stress: dissolved oxygen concentrations in bottom waters, sediment contaminant concentrations and sediment toxicity. Finally, we evaluated the stability of site classifications and index agreement at sites where data from multiple sampling events were available.

3. RESULTS

There was a high degree of agreement between EMAP-VP BI and Chesapeake B-IBI results. Degraded and undegraded condition assessments agreed for 81.3% of the 294 sampling events (Table I). The two indices were highly associated (Figure 1) with a Pearson correlation coefficient of 0.75 and a Spearman correlation coefficient of 0.74.

Most of the sites where the indices disagreed had index values close to the degraded–undegraded threshold (Table II, Figure 1). For the B-IBI, 49.1% of the disagreements were within 0.5 index units, and 81.8% within 1.0 index units of the threshold; the corresponding values for sites where both indices agreed were only 23.4% and 62.7%, respectively (Table II). The pattern for the EMAP-VP BI was similar, with 61.8% and 74.6% of disagreements and only 18.8% and 38.9% of agreements within 0.5 and 1.0 units of the threshold, respectively (Table II).

Further evidence that index classification disagreements are rare events common only at marginal sites came from the 17 sites with multiple sampling events. Only ten of the 55 results disagreed, and seven of the disagreements were from only three sites. There were six sites with disagreements for one or more sampling events and eight sites at which index results reversed for one or both indices, i.e. sites were classified as 'undegraded' on one sampling event and 'degraded' on another. Four of the six sites with disagreements were sites at which reversals occurred. At the sites with disagreements or

Table I. Application of the EMAP-VP BI and the Chesapeake Bay B-IBI to 294 sampling events in Chesapeake Bay

		Chesapeake Bay B-IBI	
		Undegraded	Degraded
EMAP-VP BI	Undegraded	159	40
	Degraded	15	80

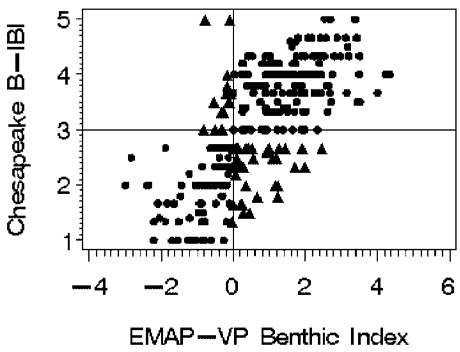

Figure 1. Association between Chesapeake Bay B-IBI and EMAP-VP BI values. Reference lines indicate degraded–undegraded thresholds. Index values for points in the lower-left and upper-right quadrants (dots) are in agreement, while points in the upper-left and lower-right quadrants (triangles) are not

reversals, complete reversals where both indices indicated identical changes in status were more common than disagreement among the indices.

In cases where the indices disagreed, the B-IBI was more conservative than the EMAP-VP BI. It classified 72.7% of the disagreements as degraded (Table III, Figure 1).

Sites where index results disagreed were distributed throughout the habitats of the Bay except polyhaline sand (Table IV). The indices agreed on results for 78.6% or more of the sites with salinity over 5 ppt, but only 62.7 and 69.7% of the sites in the less saline tidal freshwater and oligohaline habitats; 29 of the 55 disagreements (52.6%) occurred in these two habitats. Ten of the 15 sites classified as degraded by the EMAP-VP BI and undegraded by the B-IBI were in the tidal freshwater habitat (Table III).

Forty-two of the 55 site classification disagreements occurred at sites in areas that were in transition between areas with predominantly undegraded sites and areas with predominantly degraded sites (Figure 2). They were not restricted to any particular geographic area or region of the Bay.

There were indications in some cases that both indices failed to detect expected degradation effects of low dissolved oxygen or sediment chemical contamination on benthic communities. The EMAP-VP BI classified several sites with potential chemical degradation effects as undegraded in several habitats, while the Chesapeake B-IBI did so only in the tidal freshwater habitat (Table III). The EMAP-VP BI classified as undegraded 12 sites in five habitats with sediment contaminant levels exceeding Long *et al.*'s (1995) ER-M guidelines and multiple chemicals exceeding the ER-L guidelines (Table III). Long *et al.* (1995) characterize the contaminant ER-Ls as concentration levels about the lowest at which biological effects have been observed and ER-Ms as concentrations above which biological effects are likely. Toxic sediment effects, defined as *Ampelisca* survival $< 80\%$ and significantly different from controls, were observed only for three of the sites at which the indices disagreed; two of these sites were classified as undegraded by the Chesapeake B-IBI and the other as undegraded by the EMAP-VP BI.

Sixteen of the 22 sites in low mesohaline and saltier habitats considered to be 'undegraded' by the EMAP-VP BI were in areas often subject to summer hypoxia (Barth *et al.*, 1990) although only two of them had bottom dissolved oxygen measurements below 2 ppm when they were sampled. None of the 15 sites classified as 'undegraded' by the B-IBI and 'degraded' by the EMAP-VP BI had bottom dissolved oxygen below 4 ppm or was in an area often subject to summer hypoxia (Barth *et al.*, 1990).

Table II. Percentage of samples within 0.5 and 1.0 index units from the degraded–undegraded threshold for index agreements and disagreements between the Chesapeake B-IBI and the EMAP-VP BI

		B-IBI						EMAP-VP BI			
Habitat	Salinity	Disagreements			Agreements			Disagreements		Agreements	
	(ppt)	*n*	Within 0.5 (%)	Within 1.0 (%)	*n*	Within 0.5 (%)	Within 1.0 (%)	Within 0.5 (%)	Within 1.0 (%)	Within 0.5 (%)	Within 1.0 (%)
All		55	49.1	81.8	239	23.4	62.7	61.8	74.6	18.8	38.9
Freshwater	< 0.5	19	63.2	79.0	32	43.8	78.1	52.6	73.7	28.1	46.9
Oligohaline	0.5–5	10	30.0	100.0	23	60.9	87.0	70.0	80.0	26.1	56.5
Low mesohaline	5–12	5	20.0	60.0	44	20.5	54.6	40.0	80.0	20.5	40.9
High mesohaline sand	12–18	7	57.1	85.7	31	12.9	61.3	42.9	42.9	12.9	25.8
High mesohaline mud	12–18	8	37.7	62.5	48	14.6	50.0	87.5	87.5	16.7	43.8
Polyhaline sand	> 18	0	—	—	39	12.8	28.2	—	—	5.1	15.4
Polyhaline mud	> 18	6	66.7	100.0	22	13.6	68.1	83.3	83.3	31.8	54.6

Table III. Index disagreements where sediment chemical concentrations exceeded Long *et al.* sediment quality guidelines. Long *et al.* (1995) characterize the ER-L as the lowest concentration level at which biological effects have been observed and the ER-M as the concentration at which biological effects are likely

Habitat	Disagreements (n)	Number with 1 or more chemicals> ER-M	Number with 7 or more chemicals> ER-L
A: Classified undegraded by the EMAP-VP BI			
All	40	3	11
Freshwater	9	1	1
Oligohaline	9	2	6
Low mesohaline	4	0	2
High mesohaline Sand	7	0	1
High mesohaline Mud	7	0	1
Polyhaline sand	0	0	0
Polyhaline mud	4	0	0
B: Classified undegraded by the B-BI			
All	15	4	6
Freshwater	10	4	6
Oligohaline	1	0	0
Low mesohaline	1	0	0
High mesohaline sand	0	0	0
High mesohaline mud	1	0	0
Polyhaline sand	0	0	0
Polyhaline mud	2	0	0

Table IV. Distribution of disagreements by habitat

Habitat	Salinity (ppt)	Sampling events (n)	Agreement (%)	Disagreements (n)
Freshwater	< 0.5	51	62.7	19
Oligohaline	0.5–5	33	69.7	10
Low mesohaline	5–12	49	89.8	5
High mesohaline sand	12–18	38	81.6	7
High mesohaline mud	12–18	56	85.7	8
Polyhaline sand	> 18	39	100.0	0
Polyhaline mud	> 18	28	78.6	6

4. DISCUSSION

The two estuarine benthic indices agreed at 81.3% of the sites where they were applied. Forty-two of the 55 sites at which they disagreed were of marginal habitat quality and were situated in transitional or intermediate areas (Figure 2). In general, disagreeing index values were within ranges of uncertainty for the indices and many of the disagreements were values close to, but on either side of, the degraded–undegraded threshold (Figure 2, Table IV). At sites with multiple sampling events, reversals by both indices were more common than disagreements among the indices. Six sites where index disagreements occurred were sampled on multiple occasions; four of these sites also changed in status with respect to degraded or undegraded classification by one or both indices among sampling events.

The observed 81.3% level of agreement is within the level of accuracy achieved during benthic index development and, therefore, may approach the limits of agreement that can be achieved. Classification accuracy of 80–100% was achieved in all habitats with more than ten validation samples during development of the B-IBI (Weisberg *et al.*, 1997) and accuracy of 74–84% was achieved for the southeastern estuaries B-IBI in all habitats with more than ten samples (Van Dolah *et al.*, 1999). The EMAP-VP BI achieved classification accuracy of 81–88% (Paul *et al.*, 2001).

One of the reasons for the high level of agreement among the indices may be their use of similar criteria to define degraded and undegraded sites during development. The definitions for both indices were based on dissolved oxygen concentrations in bottom water, sediment chemical contaminant concentrations and sediment toxicity, which are among the most common anthropogenic stress effects in estuaries. Although the measures were the same for both indices, the values used in the definitions differed slightly.

The high level of agreement was somewhat surprising because of differences in the geographic distribution of the sites used to develop the two indices. In keeping with its applicability throughout EMAP's Virginian Province (Paul *et al.*, 1999) the EMAP-VP BI used sites from Cape Cod, Massachusetts to Cape Henry, Virginia, while the Chesapeake B-IBI only used sites in Chesapeake Bay. We expected to observe differences due to the geographic generality of the EMAP-VP BI and the geographic specificity of the Chesapeake B-IBI. Our results indicate that any differences of this type are small, if any.

There are subtle indications that the Chesapeake B-IBI, which is more sensitive than the EMAP-VP BI and more likely to classify sites as degraded, provides the more accurate picture, especially in saline habitats. Sixteen of the 22 sites classified as degraded by the B-IBI but undegraded by the EMAP-VP BI were in areas prone to degradation (Figure 2) or had dissolved oxygen or chemical contaminant concentrations at levels of potential concern (Table III). These concerns do not exist for the five saline sites classified as undegraded by the B-IBI (Table III). Nevertheless, there is no incontrovertible evidence based on dissolved oxygen, sediment contaminant or sediment toxicity data for concluding that any of these sites should be classified in one category or the other.

The Chesapeake B-IBI has some advantages. First, benthic ecologists, managers and the public easily understand the B-IBI. The concept of scoring biological properties relative to values expected under undegraded reference conditions in the same habitat is more intuitive than weighting values using discriminant function coefficients. Second, the B-IBI incorporates information from more ecological categories, leading to increased confidence in application results, while the ecological justification for the EMAP-VP BI weighting scheme and abundance metrics is less obvious. Due to the inherent variability of biological data, the certainty of ecological conclusions increases as the number of metrics redundantly confirming status increases. The B-IBI incorporates information about diversity, abundance, biomass, proportions of pollution-indicative and pollution-sensitive species, trophic structure, and depth distribution in the sediments, weighting each type of information equally. In contrast, the EMAP-VP BI effectively consists of diversity and tubificid abundance measures in freshwaters, and diversity and spionid abundance measures in saline habitats. The diversity measure contribution is weighted about twice as important as the tubificid abundance contribution and four times that of spionid abundance with statistical justification based on discriminant function

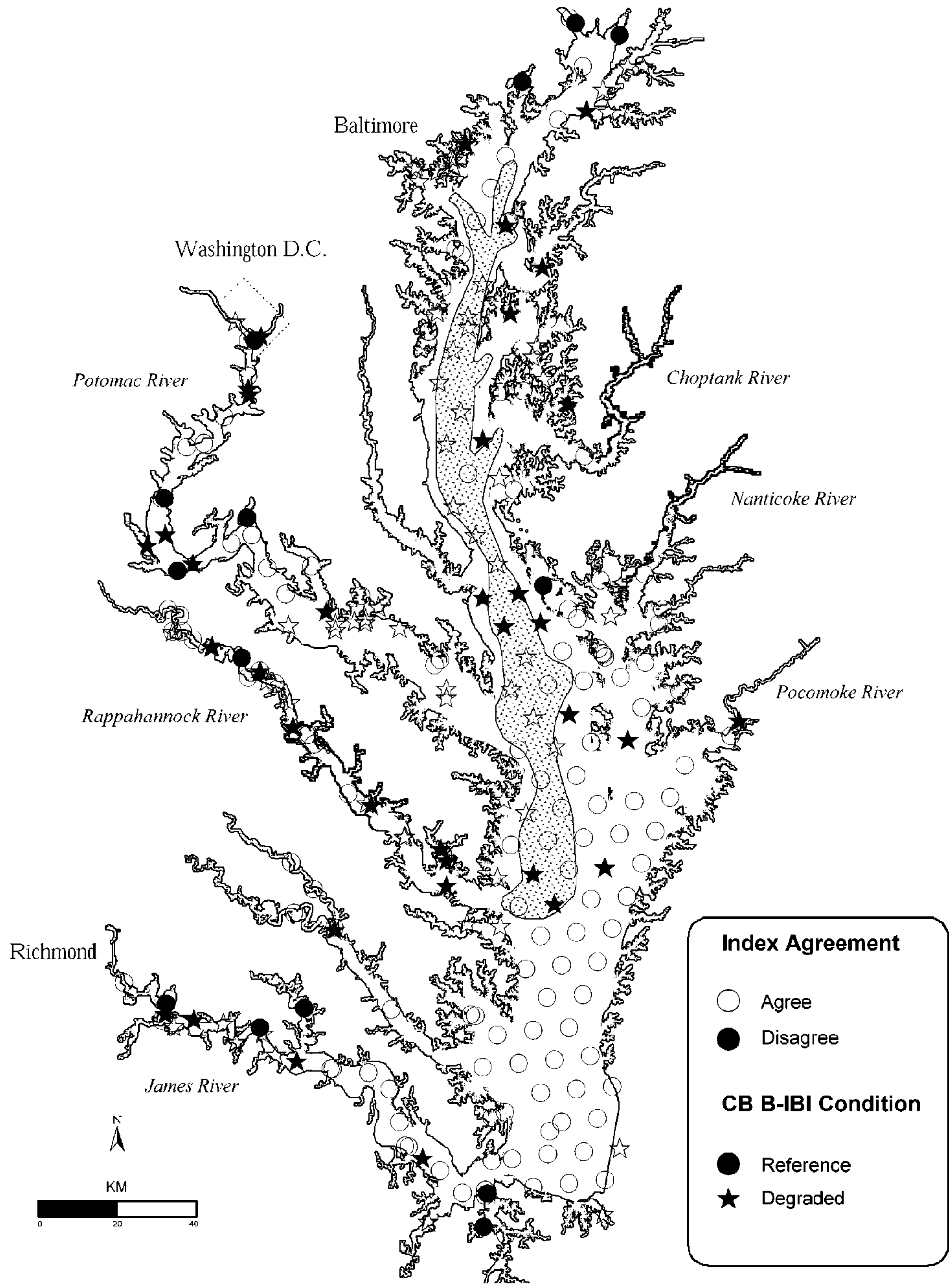

Figure 2. Locations of EMAP sampling events in Chesapeake Bay from 1990–1993. Open symbols indicate agreement between the Chesapeake Bay B-IBI and EMAP-VP BI classifications; filled symbols indicate disagreement. Circles indicate B-IBI identified undegraded sites, while stars indicate degraded sites. The stippling indicates areas regularly subject to summer hypoxia (Barth *et al.*, 1990)

coefficients (Paul *et al.*, 2001), but no ecological justification. Not all members of these two families are pollution indicative; for example, *Spiophanes bombyx* is a spionid found almost exclusively at reference sites on the east coast of the United States (Ranasinghe, unpublished data). Third, the B-IBI is flexible and suitable for application to historic data. It was developed incorporating flexibility in sampling gear, and benthic data type (abundance, biomass or depth distribution beneath the sediments). The EMAP-VP BI, on the other hand, is gear-specific to the Young Grab and does not include biomass or depth distribution information. Since almost all Chesapeake Bay data prior to 1990 were collected with sampling devices other than a Young Grab, the B-IBI is the measure of choice for assessing temporal trends or for comparisons with data collected prior to 1990. The Young Grab has also been called a Young-modified Van Veen Grab (e.g. Holland, 1990; Weisberg *et al.*, 1993; Paul *et al.*, 1999, 2001).

Using the B-IBI is also consistent with the precautionary principle because it classifies 72.7% of the disagreement sites as degraded. The precautionary principle states that where uncertainty about environmental impact exists, decisions should favor long-term sustainability of the environment (Gray, 1990; Gray and Brewers, 1996). The precautionary principle favors increasing the power of monitoring programs by paying more attention to Type II errors (Peterman and M'Gonigle, 1992; Buhl-Mortensen, 1996). In environmental terms, most monitoring programs emphasize Type I errors, reducing the error of declaring an environmental impact when there is none. Often these programs do not concern themselves with Type II errors of not declaring an environmental impact when one has occurred. The precautionary principle dictates that Type II errors are more serious to environmental management than Type I errors (Underwood, 1997).

The additional features of the B-IBI are primarily about convenience of application and conformity with conventional dogma and paradigms. Based on the close agreement between the results of the two indices, there is no empirical evidence for selecting one index or the other.

5. DISCLAIMER

Although the research described in this article has been funded wholly by the United States Environmental Protection Agency through contract number 68-W5-0054 with Science Applications International Corporation under subcontract to Versar, Inc., it has not been subjected to Agency review. Therefore, it does not necessarily reflect the views of the Agency.

ACKNOWLEDGEMENTS

This work was performed at Versar, Inc. under subcontract to Science Applications International Corporation and funded by the United States Environmental Protection Agency through contract number 68-W5-0054. We are also grateful to Allison Brindley who created the map and Thuzar Myint and Lisa Scott who compiled the data. The comments of Dr. A. K. Singh and two anonymous reviewers improved the manuscript considerably.

REFERENCES

Agard JBR, Gobin L, Warwick RM. 1993. Analysis of marine macrobenthic community structure in relation to pollution, natural oil seepage and seasonal disturbance in a tropical environment (Trinidad, West Indies). *Marine Ecology Progress Series* **92**: 233–243.

Alden RW, III, Dauer DM, Ranasinghe JA, Scott LC, Llansó RJ. 2002. Statistical verification of the Chesapeake Bay Benthic Index of Biotic Integrity. *Environmetrics* **13**: 473–498.

Anderlini VC, Wear RG. 1992. The effect of sewage and natural seasonal disturbances on benthic macrofaunal communities in Fitzroy Bay, Wellington, New Zealand. *Marine Pollution Bulletin* **24**: 21–26.

ASTM (American Society of Testing and Materials). 1991. Standard guide for conducting 10-day static sediment toxicity tests with marine and estuarine amphipods. Annual Book of ASTM Standards Volume **11.04**: 1052–1075.

Austen MC, Warwick RM, Rosado MC. 1989. Meiobenthic and macrobenthic community structure along a putative pollution gradient in southern Portugal. *Marine Pollution Bulletin* **20**: 398–405.

Barth E, Batiuk R, Brown MJ, Burch M, Domotor D, Fisher N, Hoffman F, Hollowell J, Hurley L, Jordan S, Krantz G, Magnien R, Michael B, Mihursky J, Panday N, Uphoff J, Walton G. 1990. The State of the Chesapeake Bay: third Biennial Monitoring Report—1989. Chesapeake Bay Program: Annapolis, MD.

Boesch DF. 1973. Classification and community structure of macrobenthos in the Hampton Roads area, Virginia. *Marine Biology* **21**: 226–244.

Boesch DF. 1977. A new look at the zonation of benthos along the estuarine gradient. In *Ecology of Marine Benthos*, Coull BC (ed.). University of South Carolina Press: Columbia, SC; 245–266.

Buhl-Mortensen L. 1996. Type-II statistical errors in environmental science and the precautionary principle. *Marine Pollution Bulletin* **32**: 528–531.

Dauer DM, Luckenbach MW, Rodi AJ, Jr. 1993. Abundance biomass comparison (ABC method): effects of an estuarine gradient, anoxic/hypoxic events and contaminated sediments. *Marine Biology* **116**: 507–518.

Engle VD, Summers JK. 1999. Refinement, validation, and application of a benthic condition index for Northern Gulf of Mexico estuaries. *Estuaries* **22**(3A): 624–635.

Engle VD, Summers JK, Gaston GR. 1994. A benthic index of environmental condition of Gulf of Mexico estuaries. *Estuaries* **17**: 372–384.

Folk RL. 1974. *Petrology of Sedimentary Rocks.* Hemphill Publishing Co.: Austin, TX.

Gleason HA. 1922. On the relation between species and area. *Ecology* **3**: 158.

Gray JS. 1979. Pollution-induced changes in populations. *Transactions of the Royal Philosophical Society of London* **286**(B): 545–561.

Gray JS. 1990. Statistics and the precautionary principle. *Marine Pollution Bulletin* **21**: 174–176.

Gray JS, Bewers JM. 1996. Towards a scientific definition of the precautionary principle. *Marine Pollution Bulletin* **32**: 768–771.

Green RH. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons: New York.

Holland AF (ed.). 1990. Near Coastal Program Plan for 1990: Estuaries. EPA 600/4-90/033. U.S. Environmental Protection Agency, Environmental Research Laboratory, Office of Research and Development: Narragansett, RI.

Karr JR. 1991. Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications* **1**: 66–84.

Long ER, McDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* **19**: 81–97.

Meire PM, Dereu J. 1990. Use of the abundance/biomass comparison method for detecting environmental stress: some considerations based on intertidal macrozoobenthos and bird communities. *Journal of Applied Ecology* **27**: 210–223.

Paul JF, Scott KJ, Holland AF, Weisberg SB, Summers JK, Robertson A. 1992. The estuarine component of the U.S. EPA's Environmental Monitoring and Assessment Program. *Chemistry and Ecology* **7**: 93–116.

Paul JF, Gentile JH, Scott KJ, Schimmel SC, Campbell DE, Latimer RW. 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 JF, Scott KJ, Campbell DE, Gentile JH, Strobel CS, Valente RM, Weisberg SB, Holland AF, Ranasinghe JA. 2001. Developing and applying a benthic index of estuarine condition for the Virginian Biogeographic Province. *Ecological Indicators* **1**: 83–99.

Peterman RM, M'Gonigle M. 1992. Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin* **24**: 231–234.

Ranasinghe JA, Weisberg SB, Dauer DM, Schaffner LC, Diaz RJ, Frithsen JB. 1994b. Chesapeake Bay benthic community restoration goals. CBP/TRS 107/94. Environmental Protection Agency, Chesapeake Bay Program: Annapolis, MD.

Ritz DA, Lewis ME, Shen M. 1989. Response to organic enrichment of infaunal macrobenthic communities under salmonid seacages. *Marine Biology* **103**: 211–214.

Smith RW, Bergen M, Weisberg SB, Cadien D, Dalkey A, Montagne D, Stull JK, Velarde RG. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. *Ecological Applications* **11**: 1073–1087.

Underwood AJ. 1997. Environmental decision-making and the precautionary principle: what does this principle mean in environmental sampling practice? *Landscape and Urban Planning* **37**: 137–146.

Van Dolah RF, Hyland JL, Holland AF, Rosen JS, Snoots TR. 1999. A benthic index of biological integrity for assessing habitat in estuaries of the southeastern USA. *Marine Environmental Research* **48**: 1–15.

Warwick RM. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology* **92**: 557–562.

Warwick RM, Pearson TH, Ruswahyuni. 1987. Detection of pollution effects on marine macrobenthos: further evaluation of the species abundance/biomass method. *Marine Biology* **95**: 193–200.

Weisberg SB, Ranasinghe JA, Dauer DM, Schaffner LC, Diaz RJ, Frithsen JB. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* **20**: 149–158.

Weisberg SB, Frithsen JB, Holland AF, Paul JF, Scott KJ, Summers JK, Wilson HT, Valente RM, Heimbuch DG, Gerritsen J, Schimmel SC, Latimer RW. 1993. EMAP-Estuaries Virginian Province 1990 Demonstration Project Report. EPA 600/R-92/100. U.S. Environmental Protection Agency, Environmental Research Laboratory: Narragansett, RI.

