Environ Monit Assess (2009) 150:119–127
DOI 10.1007/s10661-008-0678-7

Assessing benthic community condition in Chesapeake Bay: does the use of different benthic indices matter?

Roberto J. Llansó · Jon H. Vølstad · Daniel M. Dauer · Jodi R. Dew

Received: 2 July 2007 / Accepted: 24 July 2008 / Published online: 4 December 2008

Abstract Federal and state environmental agencies conduct several programs to characterize the environmental condition of Chesapeake Bay. These programs use different benthic indices and survey designs, and have produced assessments that differ in the estimate of the extent of benthic community degradation in Chesapeake Bay. Provided that the survey designs are unbiased, differences may exist in the ability of these indices to identify environmental degradation. In this study we compared the results of three indices calculated on the same data, and the assessments of two programs: the Chesapeake Bay Program and the Mid-Atlantic Integrated Assessment (MAIA). We examined the level of agreement of index results using site-based measures of agreement, evaluated sampling designs and statistical estimation methods, and tested for significant differences in assessments. Comparison of ratings of individual sites was done within separate categories of water and sediment quality to identify which indices summarize best pollution problems in Chesapeake Bay. The use of different benthic indices by these programs produced assessments that differed significantly in the estimate of degradation. A larger fraction of poor sites was classified as good by the Environmental Monitoring and Assessment Program's Virginian Province and MAIA benthic indices compared to the Chesapeake Bay benthic index of biotic integrity, although overall classification efficiencies were similar for all indices. Differences in survey design also contributed to differences in assessments. The relative difference between the indices remained the same when they were applied to an independent dataset, suggesting that the indices can be calibrated to produce consistent results.

Keywords Benthic communities · Benthic index · Benthic index of biotic integrity · Biological indicators · Chesapeake Bay · Comparison of methods · Degradation

R. J. Llansó (✉) · J. H. Vølstad · J. R. Dew
Versar, Inc., 9200 Rumsey Rd., Columbia, MD 21045, USA
e-mail: rllanso@versar.com

D. M. Dauer
Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, USA

Present Address:
J. H. Vølstad
Institute of Marine Research, Postboks 1870 Nordnes, 5817 Bergen, Norway

Introduction

In recent years biological indices for assessing benthic condition and habitat quality have proliferated (reviewed by Diaz et al. 2004). These indices are the result of increasing concern over the long-term sustainability of aquatic ecosystems in the face of increasing pollution and habitat degradation in rivers, estuaries, and coastal zones around the world. The development of benthic indices has responded to the need of environmental monitoring and assessment programs to characterize ecological condition of coastal waters, and to national and international legislation, such as the federal Magnuson-Stevens Fishery Conservation Management Act in the U.S., which requires that 'essential fish habitat' be identified and protected (Benaka 1999), or the Water Framework Directive of the European Union, which calls for the protection and improvement of all European surface and ground waters (Borja 2005). Although different approaches have been used to develop the indices, the objective has remained the same: to infer environmental status from the assessment of benthic community condition and habitat quality.

The implementation of benthic indices is generally carried out by the agencies that develop them. Thus, because of overlapping responsibilities, there are instances where different indices apply to the same geographical region, and the question arises as to how the results of these indices compare. Although direct comparisons of the performance of benthic indices have been conducted (e.g., Ranasinghe et al. 2002; Borja et al. 2008), the outcomes of the assessments that employ these indices have seldom been compared. In the Chesapeake Bay, federal and state agencies conduct several programs to characterize the environmental condition of its tidal waters. Two of these programs use different benthic indices and survey designs, and have produced assessments that differ in the estimate of the extent and severity of benthic community degradation in Chesapeake Bay. The Chesapeake Bay Program (CBP) estimated that 51% of the Bay tidal waters was degraded in 1997, whereas the Mid-Atlantic Integrated Assessment (MAIA) estimated that 35% of the same area was degraded. This difference has been acknowledged, but the issue ignored. Provided that the survey designs are unbiased, differences may exist in the ability of these indices to identify environmental degradation.

Interagency consistency in assessment methodologies would seem to be paramount for the integration of regional monitoring efforts and the interest of federal agencies on building national assessments from state programs (USEPA 2000). Given the emphasis placed on benthic communities as indicators of environmental health (Diaz 1992), the use of various indices with potentially divergent results may limit the ability of environmental programs to provide coherent assessments of ecological condition in coastal waters. The success of integrated monitoring efforts depends on the standardization of methods and compatibility of sampling designs (Bernstein and Weisberg 2003). Thus it is important that differences in assessments such as those observed for the Chesapeake Bay be addressed to identify potential deficiencies in monitoring programs and to suggest improvements in coordination among the lead agencies.

The objectives of this study were to compare (1) three indices applicable to Chesapeake Bay: The Chesapeake Bay benthic index of biotic integrity (B-IBI), the Environmental Monitoring and Assessment Program's Virginian Province benthic index (EMAP-VP), and the Mid-Atlantic Integrated Assessment benthic index (MAIA); and (2) the assessments of the CBP and MAIA program surveys conducted in Chesapeake Bay in 1997, by evaluating results (percent of estuarine area degraded), survey design, and statistical methods. The indices were applied to a common dataset shared by both surveys and to a second independent dataset collected in the Chesapeake Bay by NOAA's National Status & Trends (NS&T) program. We were particularly interested in whether the use of different benthic indices produced differences in assessments and if survey design or estimation methods contributed to any such differences.

Methods

Indices

The B-IBI (Weisberg et al. 1997) is a multi-metric, habitat-specific benthic index of biotic integrity. It evaluates the ecological condition of a site by comparing values of key benthic community attributes to reference values expected under nondegraded conditions in similar habitat types. Various measures of benthic macroinfaunal structure and function are used as metrics: abundance, biomass, Shannon diversity, abundance and biomass of pollution-indicative and pollution-sensitive taxa, biomass and number of taxa below the sediment–water interface, and measures of trophic composition (abundance of carnivore and omnivores, abundance of deep deposit feeders, and the abundance ratio of Tanypodini to Chironomidae). The index is scaled from 1 to 5. Index values < 3 represent degraded conditions, whereas values ≥ 3 represent good benthic community condition. The B-IBI was developed specifically for the Chesapeake Bay. Further information on the B-IBI can be found in Alden et al. (2002).

The MAIA benthic index (Llansó et al. 2002) is a multi-metric index similar to the B-IBI, with the same scale and degraded/nondegraded boundary threshold. It differs from the B-IBI mainly in that it does not use biomass as a metric but includes species richness and percent dominance. Its range of application is the MAIA region (Delaware Bay to Pamlico Sound), including the Chesapeake Bay and the coastal bays of Maryland and Virginia.

The EMAP-VP benthic index (Paul et al. 2001) is a linear discriminant function based on salinity-normalized measures of benthic macroinfauna and epifauna. Three metrics are used: Gleason's diversity, tubificid oligochaete abundance, and spionid polychaete abundance. The EMAP-VP is unbounded. Sites with index values ≤ 0 represent degraded conditions, whereas values > 0 represent good benthic community condition. Its range of application is the Virginian Biogeographic Province (Cape Cod to mouth of Chesapeake Bay).

Programs

CBP is a regional partnership between the U.S. EPA's Chesapeake Bay Office, State agencies, and citizen advisory groups. Under this program, the condition of benthic macroinfaunal communities is assessed baywide every summer using the B-IBI and a probability-based sampling design (stratified simple random; Dauer and Llansó 2003). Annually, twenty-five samples are allocated randomly to each of ten strata, and estimates of degradation for each stratum and the Bay as a whole are produced. Any region within the tidal portion of the Chesapeake Bay has a known chance of being sampled.

MAIA was implemented by EPA to gather information on the extent and condition of natural resources in the mid-Atlantic region of the U.S. The estuarine component of MAIA sampled the Delaware Bay, Chesapeake Bay, the coastal bays of Maryland and Virginia, and the Albemarle-Pamlico estuarine system of North Carolina during the summers of 1997 and 1998 (Kiddon et al. 2003). The MAIA design incorporated sampling networks of existing monitoring programs. Thus many of the CBP probability-based sites for 1997 were used by MAIA. Sites were assigned to ten strata in Chesapeake Bay (different than the CBP strata), including a clustered sampling element consisting of a random selection of small estuaries (<100 km^2).

The NS&T program is conducted by NOAA to determine the spatial extent and severity of chemical contamination and associated adverse biological effects in coastal bays and estuaries throughout the U.S. It sampled the Chesapeake Bay during the summers of 1998 (northern portion of the Bay), 1999 (middle portion), and 2001 (southern portion) (Hartwell and Hameedi 2007). Generally 3 sites were randomly allocated to each of 65 strata. The focus of the survey was the larger open expanses of the Bay assessed by the Sediment Quality Triad approach (Chapman 1990). Estimates of benthic community condition for the tidal waters of the Chesapeake Bay were not produced under this program. Thus the NS&T survey was not compared to the CBP or MAIA program surveys in this study, but the data collected under the NS&T were used to assess how well the indices compared when applied to an independent data set.

Data analysis

Benthic macroinvertebrate (0.05–mm sieve), sediment chemistry, sediment toxicity, and water quality data were obtained for each of the three programs and surveys shown in Table 1. The

Table 1 Environmental monitoring programs and benthic indices examined in this study

Program	Survey	Number of sites	Sampled by	Design	Benthic index
Chesapeake Bay Program (CBP)	Aug–Sep 1997	250	CBP	Stratified random sampling, 10 strata, all tidal waters (Dauer and Llansó 2003)	B-IBI
Mid-Atlantic Integrated Assessment (MAIA)	July–Sep 1997–1998	250 118 9	CBP EPA-ORD EPA-AED	Stratified and clustered random sampling, 10 strata, all tidal waters (Kiddon et al. 2003)	EMAP-VP, MAIA[a]
National Status & Trends (NS&T)	Aug–Sep 1998–1999 2001	208	NOAA	Stratified random sampling, 65 strata, open waters of Bay and tributaries (Hartwell and Hameedi 2007)	B-IBI

[a]Index developed for the MAIA Program but not used by this program in the 1997–1998 assessment of the Chesapeake Bay

B-IBI, EMAP-VP, and MAIA indices were computed for each of the three surveys using the methods described in Llansó (2002), Paul et al. (2001), and Llansó et al. (2002), respectively. The EMAP-VP index values computed were compared to the index values provided in the MAIA program website (http://www.epa.gov/maia/). This was done to identify and resolve discrepancies, and to understand specific usage of data, so that the same methods could be applied to all datasets. Special attention was paid to the species standardization protocols specified by each index, including which species to include or remove (e.g., epifauna) from the analysis.

Table 2 Site classifications during the development of the three indices differed in their criteria. Therefore, one set of classification criteria common to all indices was adopted for this study

Criterion	This study	
	Degraded	Nondegraded
Dissolved oxygen (ppm)	≤ 2.0	> 3.0
Effects range median (ERM) exceedances	Any	None
Effects range low (ERL) exceedances	> 10	< 3
Toxicity	Toxic[a]	Not toxic
ERM quotient	High or very high[b]	Medium or low

Degraded Meet any of the criteria, *Nondegraded* meet all criteria

[a]Toxic, <80% survival and significant difference in *Ampelisca* or Microtox tests

[b]High or very high, >0.044 mean quotient (Hyland et al. 2003)

The three indices were compared on the Chesapeake Bay MAIA and NS&T survey datasets (CBP data were part of the MAIA dataset) using the following approaches. First, the level of agreement of index results was examined using linear regression and site-based measures of agreement (Kappa analysis, Agresti 1990) of sites classified as degraded or nondegraded by the indices. Because biological indices generally show greater variability in low salinity environments, the spatial distribution of disagreements was also studied by salinity zone. Second, comparison of ratings was done for sites classified by dissolved oxygen, contaminant, and sediment toxicity criteria (Table 2) to identify which indices best summarized nutrient enrichment and sediment contaminant problems in Chesapeake Bay. Finally, estimates of degradation for CBP and MAIA program strata and the Chesapeake Bay as a whole were calculated using each of the three indices and the statistical estimation methods used by each survey, and tested for significance (confidence interval overlap test, Schenker and Gentleman 2001). The NS&T survey did not sample the entire Bay, thus estimates of degradation were not computed for this survey.

Results

Linear regressions of pairs of index values accounted for 30–44% of the variability in the MAIA dataset, and 36–55% of the variability in the NS&T dataset (Fig. 1). There was a high

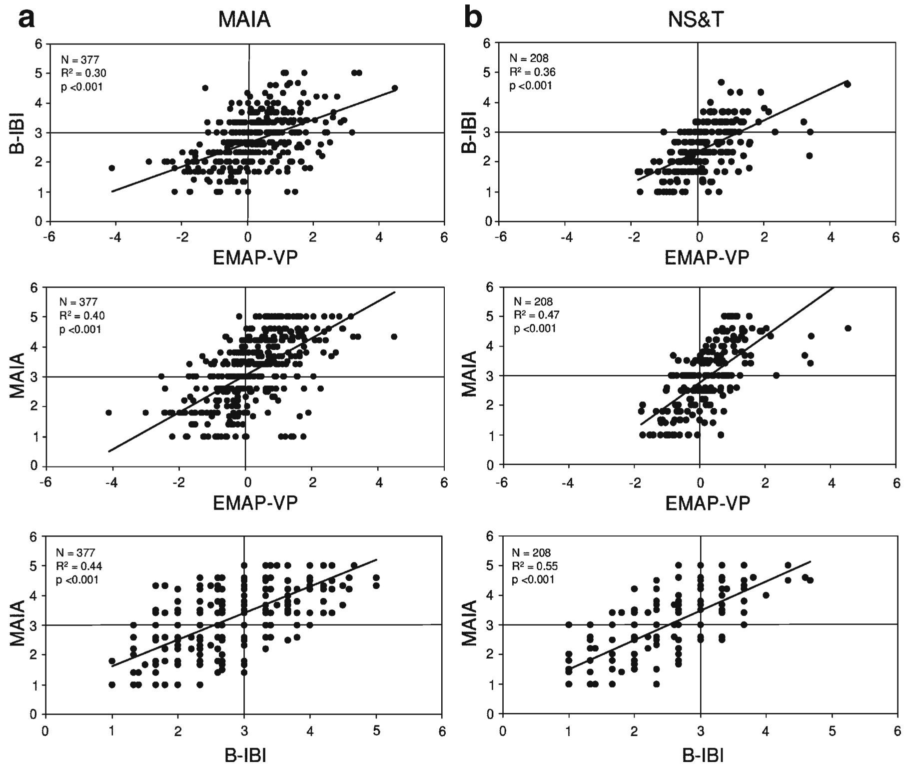

Fig. 1 Linear regressions between the values of three indices calculated on the Chesapeake Bay Mid-Atlantic Integrated Assessment (MAIA) (**a**) and National Status & Trends (NS&T) (**b**) survey datasets. Reference lines indicate degraded/nondegraded thresholds

degree of dispersion over the entire range of the indices, but the linear relationships were statistically significant. The association between the MAIA and B-IBI was better than between either of these two indices and the EMAP-VP. For sites classified as degraded or nondegraded by the indices, agreement between indices was 69–77% for the MAIA dataset, and 66–74% for the NS&T dataset (Fig. 2), with kappa indicating moderate agreement (kappa = 0.4–0.5). Agreement was only 58% and 53% for the two datasets, respectively, when all 3 indices were considered together. Diversity measures explained a large portion of the variability within indices (Fig. 3), and thus probably accounted for most of the agreement between indices. Disagreement between indices was similar across salinity zones (Table 3), suggesting that the source of disagreement was unrelated to the generally higher variability of biological indices in low salinity environments. There was no one salinity habitat that consistently had a majority or minority of disagreements; the variability observed was most likely due to the variable number of sites in each habitat.

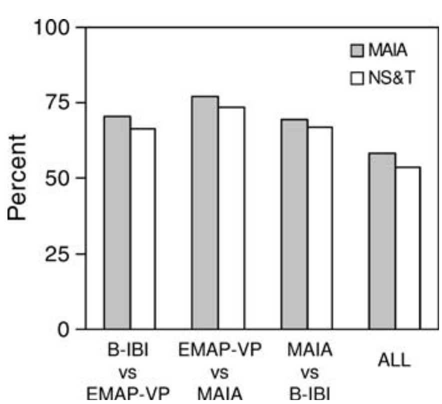

Fig. 2 Percent agreement for sites classified as degraded or nondegraded by three indices. $N = 377$ and 208, for Mid-Atlantic Integrated Assessment (MAIA) and National Status & Trends (NS&T) survey datasets, respectively

Based on a common set of classification criteria (see Table 2), agreement between indices was acceptable as to the percentage of sites correctly classified by the three indices as degraded or nondegraded. The indices correctly classified

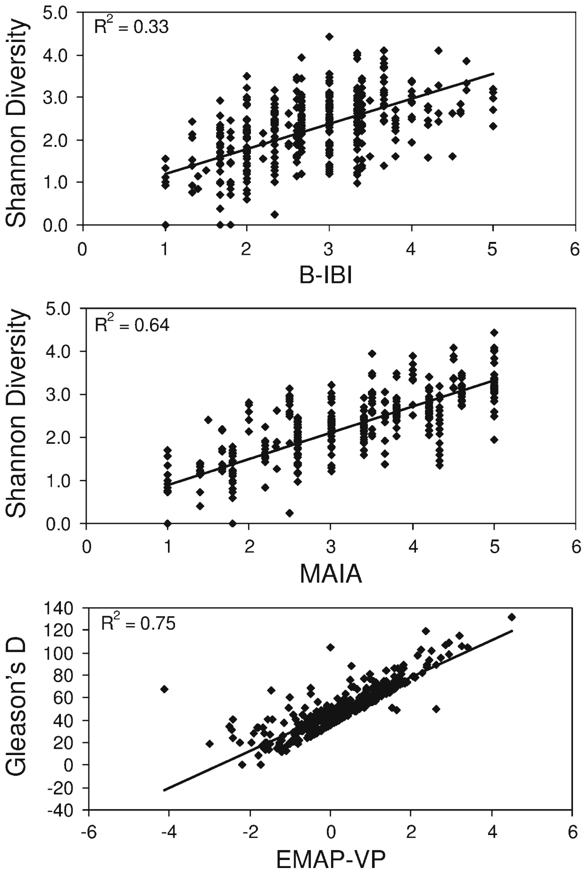

Fig. 3 Linear regression of diversity measures and index values (Mid-Atlantic Integrated Assessment survey). Gleason's D = Salinity-normalized Gleason's D

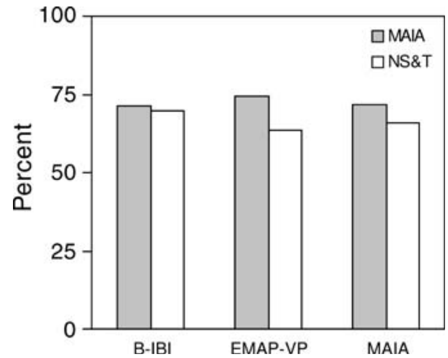

Fig. 4 Percent correct classification of sites by three indices based on dissolved oxygen, contaminant, and toxicity criteria. $N = 212$ and 176, for Mid-Atlantic Integrated Assessment (MAIA) and National Status & Trends (NS&T) survey datasets, respectively

71–75% of the sites in the MAIA dataset, and 64–70% of the sites in the NS&T dataset (Fig. 4). Dissolved oxygen was not available for NS&T; thus it was obtained by interpolating Bay Program bottom dissolved oxygen values for each of the NS&T sites (USEPA 2003, pp. 185–186). This may have contributed to the slightly lower classification efficiencies of NS&T sites. The B-IBI had more balanced type I and type II errors, while the EMAP-VP and MAIA indices underestimated degradation (Fig. 5a). This pattern was consistent in the NS&T dataset (Fig. 5b).

The use of different benthic indices by CBP and MAIA produced assessments that differed significantly in the estimate of degradation (percent of estuarine area degraded). Using the CBP survey design, the estimate of degradation was significantly higher for the B-IBI than for the EMAP-VP and MAIA indices (Fig. 6a). Using the MAIA survey design, the estimate was still higher for the B-IBI but it was not statistically significant (Fig. 6b). However, biases were identified in the MAIA program estimates due to its method of post-stratifying the CBP data. CBP sites within the MAIA strata did not have equal selection

Table 3 Percent disagreements by salinity zone

Survey	Indices	Salinity zone (psu)				
		Polyhaline (>18)	High mesohaline (12–18)	Low mesohaline (5–12)	Oligohaline (0.5–5)	Tidal fresh (<0.5)
MAIA	B-IBI vs EMAP-VP	22.9	34.1	23.9	33.3	33.3
	MAIA vs EMAP-VP	31.4	23.8	16.3	16.7	28.6
	B-IBI vs MAIA	31.4	36.0	20.7	30.0	33.3
NS&T	B-IBI vs EMAP-VP	31.0	37.5	36.8	29.4	N/A
	MAIA vs EMAP-VP	34.0	18.1	26.3	17.6	N/A
	B-IBI vs MAIA	37.0	30.6	31.6	23.5	N/A

N/A Not applicable (not surveyed)

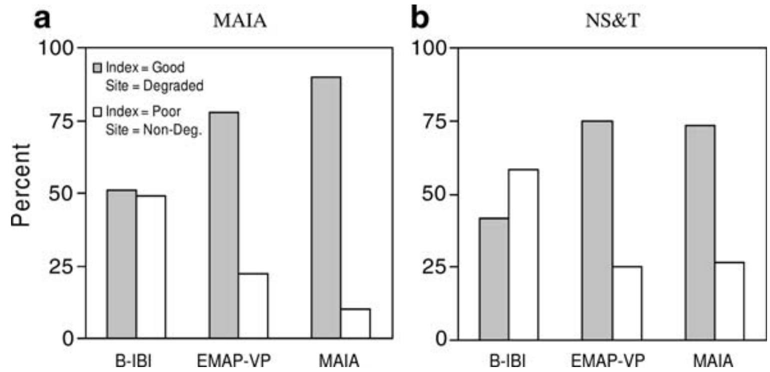

Fig. 5 Percent of Type I and Type II errors (sites misclassified) by three indices, for Mid-Atlantic Integrated Assessment (MAIA) (**a**) and National Status & Trends (NS&T) (**b**) survey datasets

probabilities, as assumed by MAIA. This assumption resulted in a marked decrease in the estimate of degradation for some strata, such as the Maryland Mid-Bay Mainstem, which contributed to the overall lower estimate of degradation for Chesapeake Bay obtained when using the MAIA survey design and the B-IBI. For example, using the stratum boundaries defined by MAIA for the Maryland Mid-Bay Mainstem, 36.1% of the estuarine area of this stratum was considered degraded by the B-IBI when equal selection probabilities were assumed, whereas 47.5% of the same area was considered degraded when area weights were applied according to the CBP survey design. The CBP design divides the Maryland Mid-Bay Mainstem into two strata, the Mid-Bay Mainstem proper and the Upper Bay Mainstem (Dauer and Llansó 2003). The Mid-Bay Mainstem is usually in worse condition than the Upper Bay Mainstem because it has a higher incidence of low dissolved oxygen events (Dauer and Llansó 2003; Llansó et al. 2008), which is one reason why these two strata are considered separately by CBP. In addition, the deep (>12 m) trough of the Mid-Bay Mainstem is azoic (Llansó et al. 2003). Weighting the deep trough separately, as CBP does, further increases the degraded area of the Maryland Mid-Bay Mainstem to 58.8%.

Discussion

Our study shows that the differences in assessments of degradation for the Chesapeake Bay were the result of the indices. Using the CBP survey design, 51% of the tidal waters of the Bay was considered degraded by the B-IBI, versus 35% by the EMAP-VP benthic index and

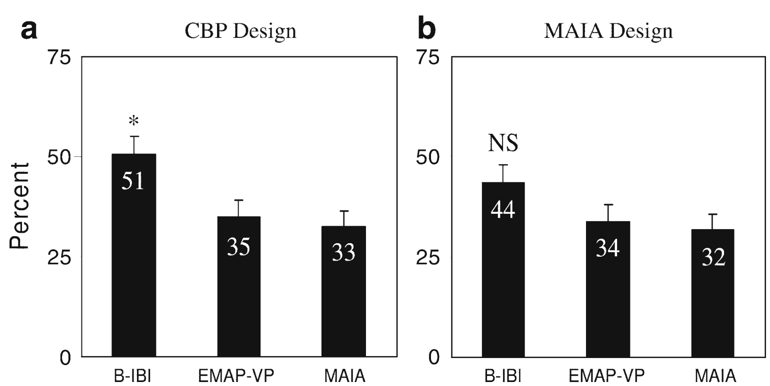

Fig. 6 Percent of Chesapeake Bay (+SE) with degraded benthic condition as measured by three indices, for Chesapeake Bay Program (CBP) (**a**) and Mid-Atlantic Integrated Assessment (MAIA) (**b**) survey designs. $N =$ 250 and 244, respectively. * Significant difference by confidence interval overlap test, *NS* not significant

33% by the MAIA benthic index. The magnitude of this difference decreased when the MAIA survey design was applied, so that the difference between indices was no longer statistically significant. Biases due to post-stratification of data from outside the originating program contributed to these differences, pointing to communication issues among the lead agencies. In a previous evaluation of the B-IBI and the EMAP-VP benthic index, Ranasinghe et al. (2002) concluded that there was no empirical evidence for selecting one index over the other. This conclusion was based on a detailed examination of the site-specific level of agreement between the indices, but the results of the assessments that employ these indices were not considered. In unbiased sampling designs we find that the indices produce different estimates of benthic condition.

Although the classification efficiencies between the three indices were comparable for sites classified by dissolved oxygen, sediment contaminant concentrations, and toxicity (71–75%), our study indicates three further concerns in comparing the indices applicable to Chesapeake Bay. First, site-specific agreement between the indices was only moderate, whether comparing index values or degraded/nondegraded determinations. Second, Type I and Type II errors were unbalanced for two of the indices (EMAP-VP and MAIA). In contrast, for the B-IBI the error types were more balanced. Third, a single metric, diversity, explained a large portion of the variance in the EMAP-VP and the MAIA indices, 75% and 64% respectively. In contrast, for the B-IBI, the species diversity metric accounted for a much lower portion of the variance, 33%. The deficiencies identified by these concerns may lead to a lack of appropriate restorative actions, particularly when benthic community condition is at intermediate levels.

The first two concerns show that site-specific condition determination is index-dependent. A larger fraction of poor sites was classified as good by the EMAP-VP and MAIA indices than by the B-IBI. The EMAP-VP and MAIA indices tended to underestimate degradation, and this accounted for a large portion of the difference between program estimates of benthic condition. It could be argued that the Chesapeake Bay is at the southern limit of the area targeted by the EMAP-VP, and that an index that is applicable to broad geographical regions would not be expected to perform as well as in more central locations of its range of applicability. However, the EMAP-VP is nonetheless being applied to the Chesapeake Bay to make inferences about the environmental quality of the estuary and to enable management decisions.

The third concern is most troublesome in that a single metric, species diversity, accounts for a large portion of the variation in two of the indices, and in a linear manner. Benthic community condition paradigms, e.g. for organic enrichment (Pearson and Rosenberg 1978) and for sediment contamination (Rakocinski et al. 2000), are non-monotonic, and specific metrics (SAB curves of species, abundance of individuals, and biomass) respond differently to stress. Benthic community indices for which the final condition characterization is driven by a single or very few metrics are not likely to be effective at intermediate levels of degradation. At these levels, subtle alterations and possible indirect effects on community condition are not likely to be a linear function of a single community metric.

Indeed the greatest challenge in applying estuarine and coastal indices of biological condition is to accurately estimate condition above or below a threshold for restorative action. Our study clearly demonstrates that choice among the existing benthic indices does matter. Rather than continue ignoring these issues, an attempt should be made to unify methodologies with the objective of producing meaningful characterizations of environmental status that reduce the risk of misinterpretation. The relative difference between the indices in this study remained the same when they were applied to a second independent dataset, suggesting that the indices can be calibrated to produce consistent results. Regional intercalibration exercises, similar to those of Borja et al. (2007), would foster communication among lead agencies and provide opportunities for the development and evaluation of new and existing methods.

Acknowledgements We are grateful to Harry Buffum (CSC) and John Kiddon (U.S. EPA, Narragansett, RI) for their assistance with the MAIA program data and computer code. The comments of Don Strebel and Mark Southerland (Versar, Inc.), Angel Borja (AZTI-Tecnalia, Spain), and two anonymous reviewers helped improve the paper.

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Alden, R. A. III, Dauer, D. M., Ranasinghe, J. A., Scott, L. C., & Llansó, R. J. (2002). Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics, 13,* 473–498.

Benaka, L. (Ed.) (1999). *Fish habitat: Essential fish habitat and rehabilitation.* Bethesda: American Fisheries Society.

Bernstein, B. B., & Weisberg, S. B. (2003). Southern California's marine monitoring system ten years after the national research council evaluation. *Environmental Monitoring and Assessment, 81,* 3–14.

Borja, A. (2005). The European water framework directive: A challenge for nearshore, coastal and continental shelf research. *Continental Shelf Research, 25,* 1768–1783.

Borja, A., Josefson, A. B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., et al. (2007). An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European water framework directive. *Marine Pollution Bulletin, 55,* 42–52.

Borja, A., Dauer, D. M., Diaz, R. J., Llansó, R. J., Muxika, I., Rodríguez, J. G., et al. (2008). Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices. *Ecological Indicators, 8,* 395–403.

Chapman, P. M. (1990). The sediment quality triad approach to determining pollution-induced degradation. *The Science of the Total Environment, 97/98,* 815–825.

Dauer, D. M., & Llansó, R. J. (2003). Spatial scales and probability based sampling in determining levels of benthic community degradation in the Chesapeake Bay. *Environmental Monitoring and Assessment, 81,* 175–186.

Diaz, R. J. (1992). Ecosystem assessment using estuarine and marine benthic community structure. In A. Burton (Ed.), *Contaminated sediment toxicity assessment* (pp. 67–85). Boca Raton, FL: Lewis.

Diaz, R. J., Solan, M., & Valente, R. M. (2004). A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management, 73,* 165–181.

Hartwell, S. I., & Hameedi, J. (2007). *Magnitude and extent of contaminated sediment and toxicity in Chesapeake Bay.* NOAA Technical Memorandum NOS NCCOS 47. Silver Spring, VA: NOAA National Ocean Service.

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

Kiddon, J. A., Paul, J. F., Buffum, H. W., Strobel, C. S., Hale, S. S., Cobb, D., et al. (2003). Ecological condition of US Mid-Atlantic estuaries, 1997–1998. *Marine Pollution Bulletin, 46,* 1224–1244.

Llansó, R. J. (2002). *Methods for calculating the Chesapeake Bay benthic index of biotic integrity.* Retrieved 10 July 2008, from http://www.baybenthos.versar.com/docs/ChesBaybibi.pdf.

Llansó, R. J., Scott, L. C., Hyland, J. L., Dauer, D. M., Russell, D. E., & Kutz, F. W. (2002). An estuarine benthic index of biotic integrity for the mid-Atlantic region of the United States. II. Index development. *Estuaries, 25,* 1231–1242.

Llansó, R. J., Dauer, D. M., Vølstad, J. H., & Scott, L. C. (2003). Application of the benthic index of biotic integrity to environmental monitoring in Chesapeake Bay. *Environmental Monitoring and Assessment, 81,* 163–174.

Llansó, R. J., Dauer, D. M., & Vølstad, J. H. (2008). Assessing ecological integrity for impaired waters decisions in Chesapeake Bay, USA. *Marine Pollution Bulletin.* doi:10.1016/j.marpolbul.2008.11.011.

Paul, J. F., Scott, K. J., Campbell, D. E., Gentile, J. H., Strobel, C. S., Valente, R. M., et al. (2001). Developing and applying a benthic index of estuarine condition for the Virginian Biogeographic Province. *Ecological Indicators, 1,* 83–99.

Pearson, T. H., & Rosenberg, R. (1978). Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology: An Annual Review, 16,* 229–311.

Rakocinski, C. F., Brown, S. S., Gaston, G. R., Heard, R. W., Walker, W. W., & Summers, J. K. (2000). Species-abundance-biomass responses by estuarine macrobenthos to sediment contamination. *Journal of Aquatic Ecosystem Stress and Recovery, 7,* 201–214.

Ranasinghe, J. A., Frithsen, J. B., Kutz, F. W., Paul, J. F., Russell, D. E., Batiuk, R. A., et al. (2002). Application of two indices of benthic community condition in Chesapeake Bay. *Environmetrics, 13,* 499–511.

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55,* 182–186.

USEPA (2000). *Clean water action plan: Coastal research and monitoring strategy.* Washington, DC: U.S. Environmental Protection Agency, Coastal Research and Monitoring Strategy Workgroup.

USEPA (2003). *Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries.* EPA 903-R-03-002. Washington, DC: Region III Chesapeake Bay Program Office.

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