ELSEVIER

Ecological Indicators xxx (2007) xxx–xxx

ECOLOGICAL INDICATORS

This article is also available online at:
www.elsevier.com/locate/ecolind

Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices

A. Borja [a,*], D. Dauer [b], R. Díaz [c], R.J. Llansó [d], I. Muxika [a], J.G. Rodríguez [a], L. Schaffner [c]

[a] *AZTI-Tecnalia, Marine Research Division, Herrera Kaia, Portualdea s/n, 20110 Pasaia, Spain*
[b] *Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, USA*
[c] *Department of Biological Sciences, School of Marine Science, Virginia Institute of Marine Science, The College of William and Mary, Gloucester Point, VA 23062, USA*
[d] *Versar Inc., 9200 Rumsey Road, Columbia, MD 21045, USA*

Received 16 October 2006; received in revised form 15 January 2007; accepted 21 January 2007

Abstract

Legislation in US and Europe has been adopted to determine the ecological integrity of estuarine and coastal waters, including, as one of the most relevant elements, the benthic macroinvertebrate communities. It has been recommended that greater emphasis should be placed on evaluating the suitability of existing indices prior to developing new ones. This study compares two widely used measures of ecological integrity, the Benthic Index of Biotic Integrity (B-IBI) developed in USA and the European AZTI's Marine Biotic Index (AMBI) and its multivariate extension, the M-AMBI. Specific objectives were to identify the frequency, magnitude, and nature of differences in assessment of Chesapeake Bay sites as 'degraded' or 'undegraded' by the indices. A dataset of 275 subtidal samples taken in 2003 from Chesapeake Bay were used in this comparison. Linear regression of B-IBI and AMBI, accounted for 24% of the variability; however, when evaluated by salinity regimes, the explained variability increased in polyhaline (38%), high mesohaline (38%), and low mesohaline (35%) habitats, remained similar in the tidal freshwater (25%), and decreased in oligohaline areas (17%). Using the M-AMBI, the explained variability increased to 43% for linear regression, and 54% for logarithmic regression. By salinity regime, the highest explained variability was found in high mesohaline and low polyhaline areas (53–63%), while the lowest explained variability was in the oligohaline and tidal freshwater areas (6–17%). The total disagreement between methods, in terms of degraded-undegraded classifications, was 28%, with high spatial levels of agreement. Our study suggests that different methodologies in assessing benthic quality can provide similar results even though these methods have been developed within different geographical areas.

Keywords: Ecological integrity; Biotic indices; Comparison of methods; B-IBI; AMBI; M-AMBI; Chesapeake Bay

* Corresponding author. Tel.: +34 943 004800.
E-mail addresses: aborja@pas.azti.es (A. Borja), ddauer@odu.edu (D. Dauer).

1470-160X/$ – see front matter
doi:10.1016/j.ecolind.2007.05.003

1. Introduction

Assessment of the ecological integrity of benthic invertebrate communities in estuaries and coastal areas has progressed in recent years due in large part to legislation such as the 'Clean Water Act' in USA or the 'Water Framework Directive' (WFD) (Borja, 2005) and 'Marine Strategy Directive' (Borja, 2006) in Europe. Such policies, albeit broad in definition, explicitly recognize the link between fauna, flora and habitat, and require appropriate strategies for assessing the relative importance, status, or ecological integrity of water bodies. A plethora of tools and benthic indices have been developed for assessing such ecological integrity or status (see Díaz et al., 2004, for a review). The goal of all these indices is to reduce or summarize environmental conditions or quality to a number, which will form the basis for management decisions regarding environmental conditions.

The development of a benthic index should follow a logical path, similar to that of Weisberg et al. (1997): (i) defining criteria for degraded and undegraded sites based on non-biological measures such as bottom-water dissolved oxygen and sediment contaminant concentrations; (ii) identifying biological measures which respond to (and differ among) degraded and undegraded sites; (iii) adjusting these responses for habitat differences, if necessary; (iv) combining responsive measures into an index; and (v) validating the index using independent data. Indices formulated on ecological principles and properly validated will better communicate the complexity of ecological integrity. Benthic indices are especially relevant to management efforts because benthic invertebrates provide site-specific indicators of habitat conditions that integrate stress effects over time and over multiple types of stress (Gray, 1979), as highlighted by Ranasinghe et al. (2002).

Díaz et al. (2004) stated that there exists a tautological development of new indices, which appears to be endemic, self-propagating and rarely justified, and recommended that investigators place greater emphasis on evaluating the suitability of existing indices prior to developing new ones. A number of recent papers have compared different methodologies (Ranasinghe et al., 2002; Díaz et al., 2003; Reiss and Kröncke, 2005; Labrune et al., 2006; Quintino et al., 2006; Dauvin et al., 2007; Dauvin, 2007; Blanchet et al., this issue), but normally within the geographical area for which the indices were developed. There are also recent efforts to intercalibrate methodologies within the WFD, in order to obtain high levels of agreement in the final status classification (Reiss and Kröncke, 2005; Labrune et al., 2006; Borja et al., 2007). To our knowledge no comparison has been made between methods overseas. In this contribution we have selected for comparison two indices used within the geographical areas of Europe and USA. Of the indices studied by Díaz et al. (2004), 23% were European and 56% were developed for application in USA.

The Benthic Index of Biotic Integrity (B-IBI) developed in the USA by Weisberg et al. (1997) stratifies habitats based on benthic assemblage differences, identifies diagnostic metrics and thresholds based on the distribution of values at reference sites, and combines metrics into an index by a process that uses a simple scoring system that weights all measures equally. The B-IBI includes measures of species diversity, productivity, indicator species, and trophic composition. These measures vary with and are optimized for each habitat. The Shannon–Wiener index is the measure of diversity used, and both abundance and biomass are included in the productivity and indicator species measures. Similar measures have been successfully used in other benthic indices of biotic integrity in USA (e.g., Van Dolah et al., 1999; Llansó et al., 2002a,b).

In Europe, the AZTI's Marine Biotic Index (AMBI) developed by Borja et al. (2000) is based upon the proportion of species assigned to one of five levels of sensitivity to increasing levels of disturbance, from very sensitive to opportunist species. This index has been tested under different stress sources (e.g., Borja et al., 2003; Muxika et al., 2005) and has been applied not only in Europe, but also in Asia (Cai et al., 2003), northern Africa (Bazairi et al., 2005) and South America (Muniz et al., 2005). Although AMBI presents some weaknesses in the inner part of estuaries or when the number of species is very low (see Borja and Muxika, 2005), the recent addition of a multivariate species richness and Shannon diversity component to AMBI, called multivariate AMBI (M-AMBI; Muxika et al., 2007), has allowed for a broader application within the WFD. This method has been intercalibrated with other European methods (Borja et al., 2007).

Please cite this article in press as: Borja, A. et al., Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices, Ecol. Indicat. (2007), doi:10.1016/j.ecolind.2007.05.003

The differences in approach and suites of measures included in different benthic indices leads to questions about whether the application of the various indices would yield different results (Ranasinghe et al., 2002). However, opportunities for comparison between indices are rare because it is unusual to have more than one benthic index available for any particular area. The availability of Chesapeake Bay (Fig. 1) benthic data used in the calculation of the B-IBI provided the opportunity to also apply and calculate AMBI for direct comparison of the two indices. Our specific objectives were to identify the frequency, magnitude, and nature of differences in assessment of Chesapeake Bay sites classified as 'degraded' or 'undegraded' by the B-IBI, AMBI, and M-AMBI.

2. Materials and methods

Chesapeake Bay data used in this study were obtained from the web site http://www.baybenthos.-versar.com, and corresponded to an extensive sampling survey undertaken in 2003 and covering 275 sampling locations from Virginia and Maryland (see Llansó et al., 2004, for details) (Fig. 1). The datasets provided species identifications, abundance, and biomass. The B-IBI values, together with the ecological integrity status, were obtained from Llansó et al. (2004). To apply AMBI, as most of the species in the current species-list (http://www.azti.es) are from the European biogeographical area (Borja et al., 2000) and some from South America (Muniz et al., 2005), it was necessary to assign the North American macrobenthic species to one of the five Ecological Groups (EG) defined by Borja et al. (2000) (i.e. EG I: species sensitive to disturbance; EG II: species indifferent to disturbance; EG III: species tolerant to disturbance; EG IV: second order opportunistic species; EG V: first order opportunistic species). The approach to assigning species not on the list was as follows:

(i) A number of references were first consulted. The following authors provided lists of pollution sensitive and opportunist species: Dauer, 1993; Rakocinski et al., 1997, 2000; Weisberg et al., 1997; Van Dolah et al., 1999; Llansó and Dauer, 2002; Llansó et al., 2002a,b.
(ii) When reference to sensitivity of the species was not found, but the same genus was present in the list, all species were assigned to the same group.
(iii) Occasionally, expert opinion of the American authors of this contribution was used to assign species to groups.

Species for which there was not enough information to be assigned to a group, were recorded as 'not assigned'. All the new assigned species are available in the new species-list (July, 2006) within the AMBI website (http://www.azti.es). Following species assignations, AMBI values were calculated using the formula in Borja et al. (2000), the free software in the same web-site, and the guidelines derived from Borja and Muxika (2005). Following these guidelines, stations with >20% of individuals not assigned or stations with 1–3 species per sample, were removed from the AMBI analysis alone, but included in the M-AMBI analysis. It is necessary to note that the AMBI scale is the opposite of M-AMBI and B-IBI, where low AMBI and high M-AMBI and B-IBI are

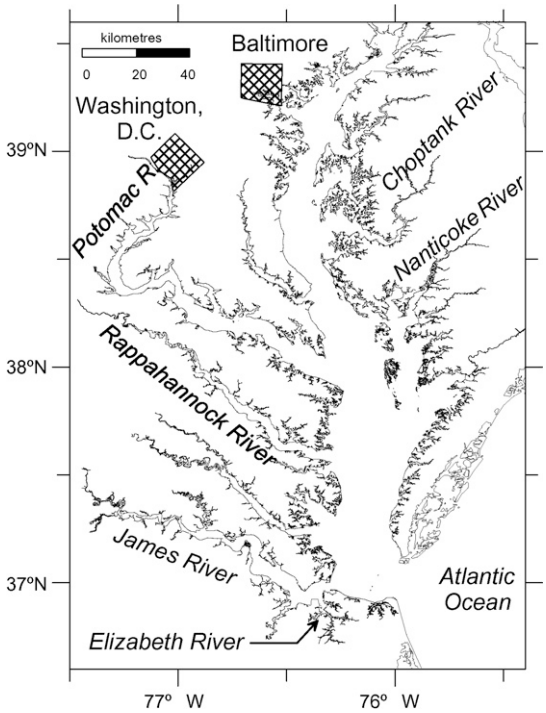

Fig. 1. Study site in Chesapeake Bay, USA. Comparison among indices was made for 275 random sites sampled August–September 2003 by the Chesapeake Bay long-term benthic monitoring and the Elizabeth River biological monitoring programs.

Please cite this article in press as: Borja, A. et al., Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices, Ecol. Indicat. (2007), doi:10.1016/j.ecolind.2007.05.003

associated with high quality environments, and high AMBI and low M-AMBI and B-IBI are associated with low quality environments.

The M-AMBI was calculated by factor analysis (FA) of AMBI, species richness (as number of *taxa*) and Shannon–Wiener diversity index values (for details, see Borja et al., 2004; Bald et al., 2005; Muxika et al., 2007). This method compares monitoring results with reference conditions by salinity stretch (see below), in order to derive an Ecological Quality Ratio (EQR), as specified by the WFD. The EQR (or M-AMBI value) expresses the relationship between observed values and reference condition values. At 'high' status, the reference condition may be regarded as an optimum where the EQR approaches one. At 'bad' status, the EQR approaches zero.

The M-AMBI analysis uses the Euclidean metric distance between each of the locations and the reference locations, together with the distance between the 'high' status and 'bad' status reference condition (see Muxika et al. (2007), for terminology and details). The distance between the 'high' and 'bad' reference condition is the maximum Euclidean distance, which is set to one. All other stations are located between the reference conditions and have M-AMBI values ranging from 0 to 1 (in some cases it is possible to have vectorial values above the 'high: 1' or under the 'bad: 0' reference conditions (see Borja et al., 2004; Bald et al., 2005; Muxika et al., 2007); in such cases, the locations were be assigned to 'high' or 'bad' status, respectively). The calculation of these values was made by using the software provided in http://www.azti.es, together with that for AMBI calculation.

As the main objective of this study is to compare US and European approaches, we defined 'bad' status as the lowest possible richness and diversity value (0) and the highest AMBI value (6) exclusive of azoic sediments (which normally get a value of 7). Taking into account that high reference conditions are not defined for the area, we used the highest richness and diversity values found in each salinity stretch, and increased them by 10–15%. Using this method we obtain the reference conditions necessary to derive M-AMBI, and they can be considered as quality objectives for the future in an undegraded situation (Table 1).

Table 1
Highest richness (number of taxa) and diversity and lowest AMBI values measured at each of five salinity zones, together with the proposed high reference conditions (bad reference conditions occurred in all zones and were 0 for diversity and richness and 6 for AMBI)

	AMBI	Species richness	Shannon diversity
Measured			
Polyhaline	1.90	43	4.01
High mesohaline	1.43	28	3.84
Low mesohaline	1.96	21	3.09
Oligohaline	1.84	16	3.19
Tidal freshwater	2.61	17	3.13
Proposed			
Polyhaline	1.7	50	4.5
High mesohaline	1.2	30	4.0
Low mesohaline	1.9	25	3.5
Oligohaline	1.7	20	3.5
Tidal freshwater	2.2	20	3.5

Agreement between B-IBI, AMBI, and M-AMBI was assessed following four approaches: (i) linear regression between indices, including all data; (ii) Kappa analysis to compare multiple condition categories; (iii) match/mismatch between indices; and (iv) correlation analyses for each salinity stretch, to compare indices and environmental parameters.

For Kappa analysis, sites were classified as 'meets goal', 'marginally degraded', 'degraded', and 'severely degraded' using the B-IBI terminology, and as 'high', 'good', 'moderate', 'poor', and 'bad' status following the AMBI and M-AMBI terminology. Using these categories, the proportion of sites where the indices disagreed were quantified, and a weighted Kappa analysis (Cohen, 1960; Fleiss and Cohen, 1973) was used to compare the multiple condition categories, with more weight given to categories further apart from the degraded-undegraded threshold. Kappa values shows the next levels of agreement: (i) Null < 0.05; (ii) Very low: 0.05–0.2; (iii) Low: 0.2–0.4; (iv) Moderate: 0.4–0.55; (v) Good: 0.55–0.7; (vi) Very Good: 0.7–0.85; (vii) Almost perfect: 0.85–0.99; and (viii) Perfect: 1 (Monserud and Leemans, 1992).

Match/mismatch between indices was computed by using 'meets goal' in B-IBI and grouping 'high' and 'good' in AMBI and M-AMBI, in both cases as undegraded, and the remainder of categories as degraded. The spatial distribution of disagreements

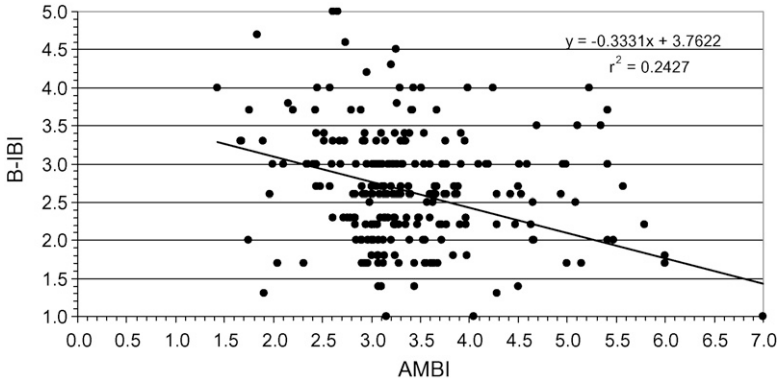

Fig. 2. Linear regression between AMBI and B-IBI for the entire dataset.

and their distribution across habitats were also studied to identify whether disagreements were more likely in any particular geographic area or any particular habitat. To further describe relationships between the indices, and the biological and environmental variables, a Principal Component Analysis (PCA) was conducted with the entire dataset: depth, salinity, silt-clay content, TOC, bottom dissolved oxygen concentration, taxa richness, diversity, and percentage of the five EG groups; and values of B-IBI, AMBI, and M-AMBI.

3. Results

Of the 217 taxa identified in the 2003 Chesapeake Bay dataset, 184 (85%) were not initially listed in the AMBI list or assigned to ecological groups. After assignment, 46 taxa (21% of the total) were in EG I, 50 (23%) in EG II, 41 (19%) in EG III, 39 (18%) in EG IV, and 12 (6%) in EG V, while 29 taxa (13%) remained unassigned.

Based on the AMBI classification, Chesapeake Bay exhibited some degree of disturbance. Of the 263 stations used in the study (after removing those recommended in AMBI guidelines, see Section 2), 141 stations (54%) were classified as slightly disturbed by AMBI, 90 (34%) as moderately disturbed, 14 (5%) as heavily disturbed, and 18 (7%) as extremely disturbed. This last group consisted of the azoic stations.

Linear regression of AMBI and B-IBI accounted for 24% of the variability of the entire dataset, with a high degree of dispersion over the entire range of both indices (Fig. 2). When evaluated by saline regimes, the explained variability increased in the polyhaline (38%), high mesohaline (38%), and low mesohaline (35%) areas, remained similar in tidal freshwater (25%), and decreased in oligohaline areas (17%).

When M-AMBI was regressed with the B-IBI, the explained variability was 43% for linear regression, and 54% for logarithmic regression (Fig. 3). By salinity regime, the highest explained variability was found in mesohaline and polyhaline areas (53–63%), while the lowest explained variability was in the oligohaline and tidal fresh water areas (6–17%) (Fig. 3).

Direct comparison of ecological status between site classifications was not possible due to the different number of status categories between the M-AMBI (five levels) and the B-IBI (four levels). New boundaries for the M-AMBI were determined from the regression lines, taking into account the B-IBI limits (Table 2). Using this approach, 'bad' status corresponded to 'severely degraded' sites in the B-IBI, and 'high-good' status corresponded to 'meets goal'. Hence, B-IBI and M-AMBI can be compared (Table 3). The Kappa statistic showed a moderate level of agreement between site classifications ($\kappa = 0.46$). Removing tidal freshwater and oligohaline sites increased agreement slightly ($\kappa = 0.48$), but it remained moderate.

When comparing both methods in terms of degraded versus undegraded categories, using 0.58 as threshold for M-AMBI and 3.0 for B-IBI (Table 2), the disagreement for the entire dataset represents 28%

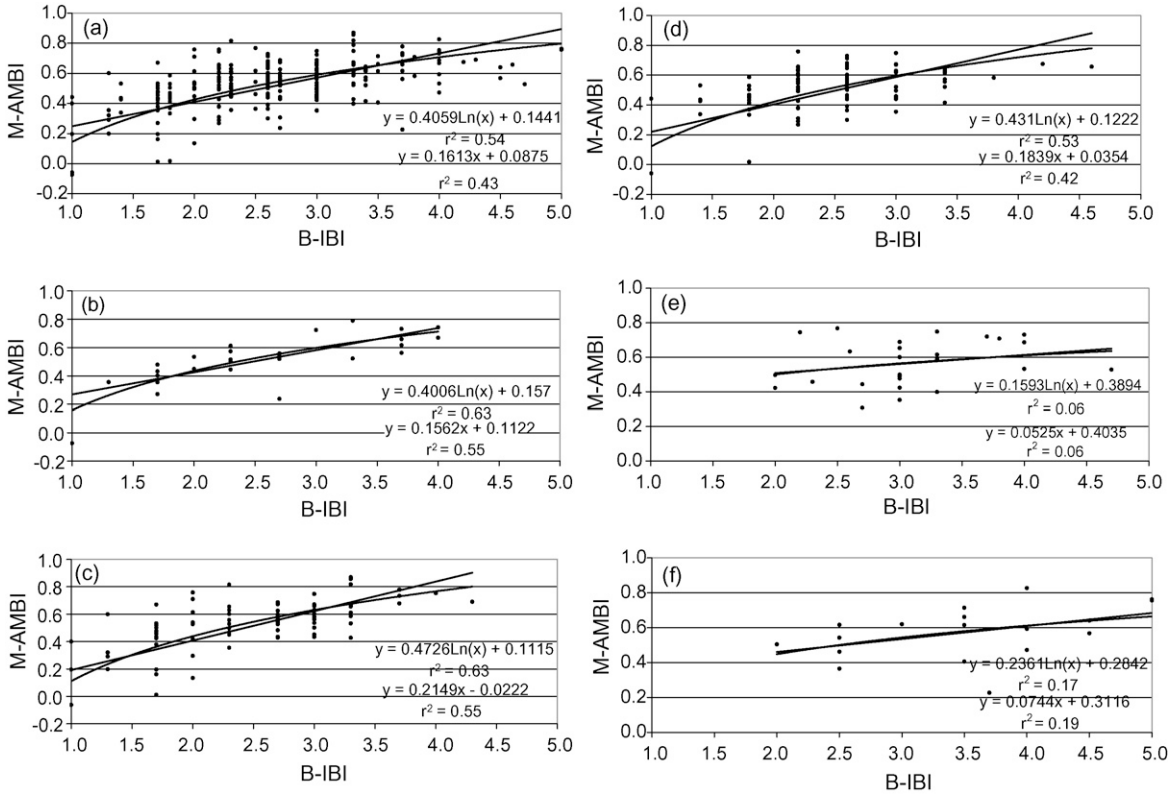

Fig. 3. Linear and logarithmic regression between M-AMBI and B-IBI for: (a) the entire dataset, (b) polyhaline (>18 psu), (c) high mesohaline (12–18 psu), (d) low mesohaline (5–12 psu), (e) oligohaline (0.5–5 psu), and (f) tidal freshwater (<0.5 psu) zones.

of the cases (15.6% and 12.7%, respectively) (Table 4), in spite of good spatial agreement (Fig. 4). When the M-AMBI indicated undegraded conditions but the B-IBI indicated degraded conditions, the highest percentages of disagreement occurred in low mesohaline (21%) and high mesohaline (16%) areas (Table 4). Conversely, when the B-IBI indicated undegraded conditions and M-AMBI indicated degraded conditions, the highest levels of disagreement were in oligohaline (31%) and tidal freshwater (22%) areas (Table 4).

The correlation coefficients between indices and environmental parameters by salinity were significant at the 0.1% level for polyhaline, high mesohaline, and low mesohaline areas, but not significant for oligohaline and tidal freshwater areas (Table 5). Significant

Table 2
Equivalence between M-AMBI and B-IBI boundaries used to define ecological status and degradation

Degradation	M-AMBI				B-IBI	
	Status	(1)	(2)	(3)	Status	Boundaries
Undegraded	High	>0.83	>0.85			
Undegraded	Good	0.62–0.83	0.55–0.85	>0.58	Meets goal	≥3.0
Degraded	Moderate	0.41–0.62	0.39–0.55	0.5–0.58	Marginal	2.7–2.9
Degraded	Poor	0.20–0.41	0.20–0.39	0.4–0.5	Degraded	2.0–2.6
Degraded	Bad	<0.20	<0.20	<0.4	Severely degraded	1.0–2.0

M-AMBI boundaries: (1) defined in Borja et al. (2004); (2) intercalibrated in Borja et al. (2007); (3) defined in this exercise, in relation to B-IBI.

Please cite this article in press as: Borja, A. et al., Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices, Ecol. Indicat. (2007), doi:10.1016/j.ecolind.2007.05.003

Table 3
Number of sites classified as meets goal, marginal, degraded, and very degraded by the B-IBI and the M-AMBI

	M-AMBI				Total
	Meets goal	Marginal	Degraded	Severely degraded	
B-IBI					
Meets Goal	63 (44)	14 (11)	17 (11)	4 (1)	98 (67)
Marginal	5 (5)	8 (8)	4 (3)	2 (1)	19 (17)
Degraded	31 (26)	17 (16)	18 (16)	10 (9)	76 (67)
Very degraded	7 (7)	10 (10)	27 (24)	38 (38)	82 (79)
Total	106 (82)	49 (45)	66 (54)	54 (49)	275 (230)

Numbers in parentheses exclude tidal freshwater and oligohaline sites.

Table 4
Number (*N*) and percentage (%) of disagreements between M-AMBI and B-IBI, for sites classified as degraded and undegraded, for the entire dataset and by salinity stretch

Salinity stretch	M-AMBI (undegraded) and B-IBI (degraded)		M-AMBI (degraded) and B-IBI (undegraded)	
	N	%	*N*	%
Polyhaline (30)	2	6.7	2	6.7
High mesohaline (99)	16	16.2	9	9.1
Low mesohaline (102)	21	20.6	12	11.8
Oligohaline (26)	3	11.5	8	30.8
Tidal freshwater (18)	1	5.6	4	22.2
Entire dataset (275)	43	15.6	35	12.7

Number of sites in parenthesis.

correlations between indices and individual parameters were variable. The only consistent correlations were in the high mesohaline area where all indices were associated with water depth, bottom dissolved oxygen concentration, and total organic carbon (TOC). Also, note that all indices showed decreasing benthic quality as dissolved oxygen decreased in mesohaline areas, and this relationship was highly significant (Table 5). Indices and structural parameters (richness and diversity) are highly correlated, except in oligohaline and tidal freshwater for AMBI and B-IBI (Table 5).

In PCA, 50% of the variability was explained by the first two components. Component 1 was related to the benthic indices (M-AMBI and AMBI) and diversity; component 2 was related to physical parameters (salinity, depth and dissolved oxygen); and component 3 was related to EG III (composed by species tolerant to organic enrichment and typical of estuaries) (Fig. 5). AMBI was negatively related to richness and percentage of EG IV; B-IBI was positively related to dissolved oxygen and negatively related to depth; and M-AMBI was positively related to diversity (Fig. 5).

4. Discussion

The level of agreement in determining benthic integrity or ecological status between the B-IBI and AMBI was moderate and probably due to different boundary and threshold settings, which makes comparisons between indices difficult. In fact, when we readjusted boundaries in the M-AMBI, the level of agreement increased dramatically. The same situation was found in Europe when intercalibrating different methods within the WFD, where it was necessary to adjust boundaries for each methodology, in order to achieve better agreement between methods (Borja et al., 2007).

When the comparison was made in terms of degraded-undegraded locations, and the M-AMBI used as alternative to the AMBI, the two-benthic indices agreed over 72% of the sites. This agreement is slightly

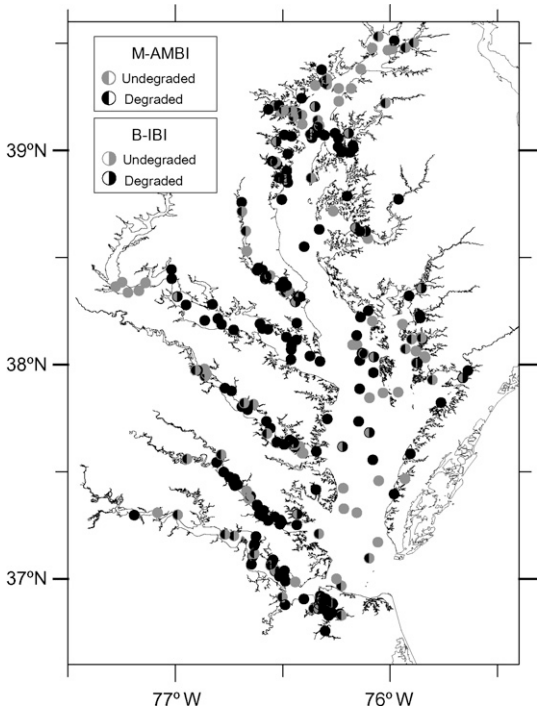

Fig. 4. Comparison between the results obtained with M-AMBI and B-IBI in terms of degraded (moderate/poor/bad or marginal/degraded/severely degraded) and undegraded (high/good or meets goal).

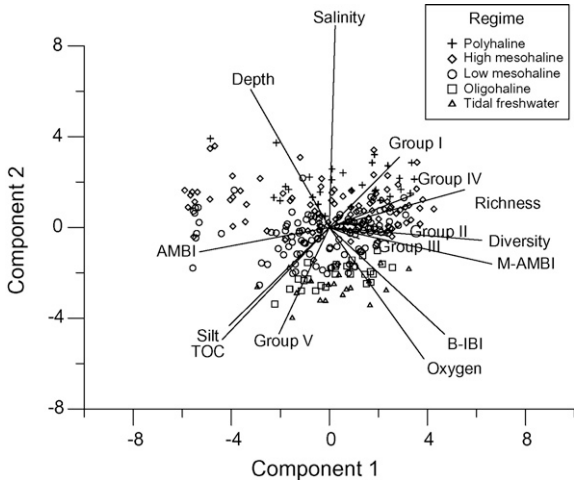

Fig. 5. Principal Component Analysis for the entire dataset. Groups I–V, represent the percentage of each ecological group, as described in Section 2.

under the range of agreement (80–87%) found in Europe when intercalibrating different methods for the WFD (Borja et al., 2007). The differences in agreement when comparing USA and European methodologies probably results from a scaling problem of one or both indices across habitats, rather than to underlying index performance. Where the indices disagreed, some values were within ranges of uncertainty for the indices and many of the disagreements were close to, but on either side of, the degraded–undegraded threshold (e.g. 14 of the 35 values where the indices disagreed were close to the 0.58 boundary).

One of the reasons for this level of agreement between different indices may be their use of similar criteria to define degraded and undegraded sites during development (Ranasinghe et al., 2002). The definitions for B-IBI were based on dissolved oxygen concentrations in bottom water, sediment chemical contaminant concentrations, sediment toxicity, and organic carbon content, which are among the most common anthropogenic stressors in estuaries. In the case of M-AMBI, the basis is the AMBI, which has been tested against different anthropogenic impacts, which includes anoxia and hypoxia, sediment toxicity (metals, PAH), and others (Borja et al., 2000, 2003, 2006, 2007; Muxika et al., 2005). In this study the highest values of AMBI were associated with the most degraded sites, which had high TOC and low dissolved oxygen concentrations, at some of the salinity stretches (Table 5). High values of B-IBI and M-AMBI are usually associated with high levels of dissolved oxygen, richness, and diversity, normally representative of healthy environments (Weisberg et al., 1997; Borja et al., 2007).

Alden et al. (2002), studying the relative discriminatory power of individual metrics within the B-IBI, determined that the most important metrics were pollution-indicative taxa abundance, pollution-sensitive taxa abundance, and diversity (especially in mesohaline and polyhaline habitats). This pattern is consistent with the results of this study, and could be the reason of the level of agreement between the B-IBI and the M-AMBI, because M-AMBI includes proportions of indicative and sensitive taxa (AMBI) together with diversity and richness.

When comparing B-IBI and the U.S. Environmental Monitoring and Assessment Program's Virginian Province Benthic Index (EMAP-VP BI), Ranasinghe et al. (2002) found a high level of

Table 5
Correlation coefficients between indices, environmental and structural parameters, by salinity zone

	Polyhaline (30)	High mesohaline (99)	Low mesohaline (102)	Oligohaline (26)	Tidal freshwater (18)
AMBI–M-AMBI	**−0.887**	**−0.897**	**−0.884**	−0.513	−0.472
AMBI–B-IBI	**−0.615**	**−0.617**	**−0.591**	−0.413	−0.500
M-AMBI–B-IBI	**0.743**	**0.744**	**0.651**	0.255	0.441
AMBI–Depth	0.562	0.517	0.233	−0.522	−0.469
AMBI–Salinity	−0.217	−0.058	0.130	−0.060	−0.541
AMBI–Oxygen	**−0.591**	**−0.462**	**−0.404**	0.047	−0.345
AMBI–TOC	−0.108	**0.482**	**0.388**	0.113	−0.300
AMBI–Silt/Clay	−0.016	0.281	**0.369**	−0.025	−0.132
AMBI–Richness	**−0.634**	**−0.714**	**−0.655**	−0.084	−0.021
AMBI–Diversity	**−0.615**	**−0.753**	**−0.759**	−0.154	−0.151
M-AMBI–Depth	−0.316	**−0.570**	**−0.449**	0.294	−0.188
M-AMBI–Salinity	0.265	0.110	−0.244	0.232	0.161
M-AMBI–Oxygen	0.388	**0.547**	**0.524**	−0.142	0.611
M-AMBI–TOC	0.037	**−0.594**	**−0.499**	−0.353	−0.301
M-AMBI–Silt/Clay	−0.024	**−0.445**	**−0.483**	−0.337	−0.292
M-AMBI–Richness	**0.836**	**0.922**	**0.881**	**0.837**	**0.850**
M-AMBI–Diversity	**0.831**	**0.931**	**0.918**	**0.863**	**0.884**
B-IBI–Depth	−0.116	**−0.557**	**−0.370**	0.260	0.226
B-IBI–Salinity	**0.571**	−0.105	−0.290	−0.347	0.165
B-IBI–Oxygen	0.328	**0.624**	**0.359**	0.217	0.064
B-IBI–TOC	−0.218	**−0.381**	−0.219	−0.442	0.012
B-IBI–Silt/Clay	−0.242	−0.239	−0.240	−0.204	0.116
B-IBI–Richness	**0.754**	**0.661**	**0.517**	−0.084	0.240
B-IBI–Diversity	**0.533**	**0.773**	**0.657**	0.278	0.304

Bold and underlined numbers are significant at $p < 0.001$; underlined numbers are significant at $p < 0.01$. Number of sites in parentheses.

agreement between indices, despite differences in the geographic distribution of the sites used to develop the two indices. These authors found that the B-IBI was more sensitive than the EMAP-VP BI in classifying sites as degraded, and provided the more accurate picture, especially in saline habitats. This pattern is similar to that found in this study, in which the B-IBI shows the greatest number of disagreements with M-AMBI results in low salinity locations. In the case of AMBI, the same problems in low salinity habitats have been described (Borja and Muxika, 2005); however, M-AMBI may be more accurate in those habitats (Muxika et al., 2007). Probably, this reflects the natural physical stress within these areas, rendering both indices less useful in these habitats or being more difficult to assess the quality status under natural stress.

There are advantages in using B-IBI and M-AMBI: (i) benthic ecologists, managers, and stakeholders easily understand both types of indices, due to their intuitivity, as mentioned by Ranasinghe et al. (2002) and Borja and Muxika (2005); (ii) both indices incorporate information based on well-known ecological theories, such as Pearson and Rosenberg's (1978) paradigm, which lead to increased confidence in the results; and (iii) both indices incorporate information about diversity and proportions of pollution-indicative (opportunistic) and pollution-sensitive species abundance, while M-AMBI also incorporates richness, and the B-IBI a variety of other metrics.

The macrofaunal-based indices compared in this study are derived from biomass and/or abundance data, which are measures of biotic integrity in the sense of Karr et al. (1986), and as highlighted by Díaz et al. (2003). These indices are measures of community structure, and emphasize species identity, richness, and diversity, which are thought to be intrinsically important features of the benthos. Hence, when community structure indices are high, it is assumed that benthic habitat quality is also high (Díaz et al., 2003). The indices we studied were responsive to

organic enrichment gradients, as described by the model of Pearson and Rosenberg (1978) and to hypoxia and other stressors that reduce species diversity and abundance, as discussed in Borja et al. (2000, 2003, 2006, 2007), Díaz et al. (2003) and Muxika et al. (2005).

A principal factor stressing the macrobenthos and degrading benthic habitat quality in the Chesapeake Bay is low dissolved oxygen, which is spatially extensive and strongly correlated with benthic community condition, explaining 42% of the variation in the B-IBI (Dauer et al., 2000). High levels of sediment contamination are spatially limited to a few locations including Baltimore Harbor and the Southern Branch of the Elizabeth River, and explained about 10% of the variation in the B-IBI. After removing the effects of low dissolved oxygen, the residual variation in benthic community condition was weakly correlated with surrogates of eutrophication, such as water column concentrations of total nitrogen, total phosphorus, and chlorophyll *a* (Dauer et al., 2000).

5. Conclusions

Our study suggests that applying different methodologies to assessing benthic quality can provide similar results even the methods have been developed within different geographical areas. When the same ecological basis is used in indices such as B-IBI, AMBI and M-AMBI the results could produce a level of agreement, in assessing ecological integrity, close to that found in Europe when intercalibrating WFD methodologies. Much of the mismatch between indices was related to spatial variability in community structure measures and the type of habitat. However, indices that integrate structural and functional aspects of benthos, such as B-IBI and M-AMBI, hold promise as measures of benthic habitat quality because of their ability to integrate physical habitat structure with benthic communities. However, more comparison studies should be undertaken in order to improve the assessment of ecological integrity.

Acknowledgments

J.G. Rodríguez was supported by the programme Torres Quevedo (Ministerio de Educación y Ciencia and Fondo Social Europeo) during the final writing of the manuscript. The comments from two reviewers, Jean Claude Dauvin (Station Marine de Wimereux, Université des Sciences et Technologies de Lille, France) and Steve Weisberg (Southern California Coastal Water Research Project, Westminster, CA, USA), have considerably improved the initial draft. This paper is contribution number 370 from AZTI-Tecnalia (Marine Research Unit).

References

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

Bald, J., Borja, A., Muxika, I., Franco, J., Valencia, V., 2005. Assessing reference conditions and physico-chemical status according to the European Water Framework Directive: a case-study from the Basque Country (Northern Spain). Mar. Pollut. Bull. 50, 1508–1522.

Bazairi, H., Bayed, A., Hily, C., 2005. Structure et bioévaluation de l'état écologique des communautés benthiques d'un écosystème lagunaire de la côte atlantique marocaine. Comptes Rendus Biologies 328, 977–990.

Blanchet, H., Lavesque, N., Ruellet, T., Dauvin, J.C., Sauriau, P.G., Desroy, N., Desclaux, C., Leconte, M., Bachelet, G., Janson, A.L., Bessineton, C., Duhamel, S., Jourde, J., Mayot, S., Simon, S., De Montaudouin, X., this issue. Use of Biotic Indices in semi-enclosed coastal ecosystems and transitional waters habitats—implications for the implementation of the European Water Framework Directive. Ecol. Indic., doi:10.1016/j.ecolind.2007.04.003.

Borja, A., 2005. The European Water Framework Directive: a challenge for nearshore, coastal and continental shelf research. Cont. Shelf Res. 25 (14), 1768–1783.

Borja, A., 2006. The new European Marine Strategy Directive: difficulties, opportunities, and challenges. Mar. Pollut. Bull. 52, 239–242.

Borja, A., Muxika, I., 2005. Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. Mar. Pollut. Bull. 50, 787–789.

Borja, A., Franco, J., Pérez, V., 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Pollut. Bull. 40, 1100–1114.

Borja, A., Muxika, I., Franco, J., 2003. The application of a Marine Biotic Index to different impact sources affecting soft-bottom benthic communities along European coasts. Mar. Pollut. Bull. 46, 835–845.

Borja, A., Franco, J., Valencia, V., Bald, J., Muxika, I., Belzunce, M.J., Solaun, O., 2004. Implementation of the European Water Framework Directive from the Basque Country (northern Spain): a methodological approach. Mar. Pollut. Bull. 48 (3–4), 209–218.

Borja, A., Muxika, I., Franco, J., 2006. Long-term soft-bottom benthos recovery, following urban and industrial sewage treatment in the Nervión estuary (southern Bay of Biscay). Mar. Ecol. Progr. Ser. 313, 43–55.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 42–52.

Cai, L., Tam, N.F.Y., Wong, T.W.Y., Ma, L., Gao, Y., Wong, Y.S., 2003. Using Benthic Macrofauna to Assess Environmental Quality of Four Intertidal Mudflats in Hong Kong and Shenzhen Coast. Acta Oceanol. Sinica 22, 309–319.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Dauer, D.M., 1993. Biological criteria, environmental health and estuarine macrobenthic community structure. Mar. Pollut. Bull. 26, 249–257.

Dauer, D.M., Weisberg, S.B., Ranasinghe, J.A., 2000. Relationships between benthic community condition, water quality, sediment quality, nutrient loads, and land use patterns in Chesapeake Bay. Estuaries 23, 80–96.

Dauvin, J.C., 2007. Paradox of estuarine quality: benthic indicators and indices in estuarine environments, consensus or debate for the future. Mar. Pollut. Bull. 55, 271–281.

Dauvin, J.C., Ruellet, T., Desroy, N., Janson, A.L., 2007. The ecological quality status of the Bay of Seine and the Seine estuary: use of biotic indices. Mar. Pollut. Bull. 55, 241–257.

Díaz, R.J., Cutter, G.R., Dauer, D.M., 2003. A comparison of two methods for estimating the status of benthic habitat quality in the Virginia Chesapeake Bay. J. Exp. Mar. Biol. Ecol. 285/286, 371–381.

Díaz, R.J., Solan, M., Valente, R.M., 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. J. Environ. Manage. 73, 165–181.

Fleiss, J.L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ. Psychol. Meas. 33, 613–619.

Gray, J.S., 1979. Pollution-induced changes in populations. Trans. R. Philos. Soc. Lond. 286(B), 545–561.

Karr, J.R., Fausch, K.D., Angermeier, P.L., Yant, P.R., Schlosser, I.J., 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey Special Publication, vol. 5. Champaign, IL.

Labrune, C., Amouroux, J.M., Sarda, R., Dutrieux, E., Thorin, S., Rosenberg, R., Gremare, A., 2006. Characterization of the ecological quality of the coastal Gulf of Lions (NW Mediterranean): A comparative approach based on three biotic indices. Mar. Pollut. Bull. 52, 34–47.

Llansó, R.J., Dauer, D.M., 2002. Methods for calculating the Chesapeake Bay benthic index of biotic integrity. 1–24, http://sci.odu.edu/chesapeakebay/data/benthic/BIBIcalc.pdf.

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

Llansó, R.J., Scott, L.C., Hyland, J.L., Dauer, D.M., Russell, D.E., Kutz, F.W., 2002b. 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., Kelley, F.S., Scott, L.C., 2004. Chesapeake Bay water quality monitoring program. Long-term benthic monitoring and assessment component. Level I: comprehensive report July 1984–December 2003 (vol. 1). Prepared by Versar Inc. for Maryland Department of Natural Resources, Resource Assessment Service, Tidewater Ecosystem Assessment, Annapolis, Maryland.

Monserud, R., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. Ecol. Model. 62, 275–293.

Muniz, P., Venturini, N., Pires-Vanin, A.M.S., Tommasi, L.R., Borja, A., 2005. Testing the applicability of a Marine Biotic Index (AMBI) for assessing the ecological quality of soft-bottom benthic communities in the South America Atlantic region. Mar. Pollut. Bull. 50, 624–637.

Muxika, I., Borja, A., Bonne, W., 2005. The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Ecol. Indic. 5 (1), 19–31.

Muxika, I., Borja, Á., Bald, J., 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 16–29.

Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanogr. Mar. Biol. Annu. Rev. 16, 229–311.

Quintino, V., Elliott, M., Rodrigues, A.M., 2006. The derivation, performance and role of univariate and multivariate indicators of benthic change: case studies at differing spatial scales. J. Exp. Mar. Biol. Ecol. 330, 368–382.

Rakocinski, C.F., Brown, S.S., Gaston, G.R., Heard, R.W., Walker, W.W., Summers, J.K., 1997. Macrobenthic responses to natural and contaminant-related gradients in Northern Gulf of Mexico estuaries. Ecol. Appl. 7, 1278–1298.

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 chemical contamination. J. Aquat. Ecosyst. Stress Recov. 7, 201–214.

Ranasinghe, J.A., Frithsen, J.B., Kutz, F.W., Paul, J.F., Russell, D.E., Batiuk, R.A., Hyland, J.L., Scott, J., Dauer, D.M., 2002. Application of two indices of benthic community condition in Chesapeake Bay. Environmetrics 13, 499–511.

Reiss, H., Kröncke, I., 2005. Seasonal variability of benthic indices: an approach to test the applicability of different indices for ecosystem quality assessment. Mar. Pollut. Bull. 50, 1490–1499.

Van Dolah, R.F., Hyland, J.L., Holland, A.F., Rosen, J.S., Snoots, T.R., 1999. A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. Mar. Environ. Res. 48, 269–283.

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

Please cite this article in press as: Borja, A. et al., Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices, Ecol. Indicat. (2007), doi:10.1016/j.ecolind.2007.05.003