EPA Public Access

Author manuscript

Ecol Indic. Author manuscript; available in PMC 2019 June 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Ecol Indic. 2018 June 1; 89: 818–827. doi:10.1016/j.ecolind.2017.08.067.

Adaptation and application of multivariate AMBI (M-AMBI) in US coastal waters

Marguerite C. Pelletier¹, David J. Gillett², Anna Hamilton³, Treda Grayson⁴, Virginia Hansen⁵, Erik W. Leppo³, Stephan B. Weisberg², and Angel Borja⁶

¹US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Laboratory, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI, USA

²Southern California Coastal Water Research Project, 3535 Harbor Blvd., Suite 110, Costa Mesa, CA, USA

³TetraTech, 400 Red Brook Blvd., Owings Mills, MD 21117, USA

⁴United States Environmental Protection Agency, Office of International and Tribal Affairs, American Indian Environmental Office, 1200 Pennsylvania Avenue, N. W., Washington DC, USA

⁵US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Laboratory, Gulf Ecology Division, One Sabine Island Drive, Gulf Breeze, FL 32561, USA

⁶AZTI Tecnalia Marine Research Division, Herrera Kaia Portualdea n/s, 20110 Pasaia, Spain

Abstract

The multivariate AMBI (M-AMBI) is an extension of the AZTI Marine Biotic Index (AMBI) that has been used extensively in Europe, but not in the United States. In a previous study, we adapted AMBI for use in US coastal waters (US AMBI), but saw biases in salinity and score distribution when compared to locally calibrated indices. In this study we modified M-AMBI for US waters and compared its performance to that of US AMBI. Index performance was evaluated in three ways: 1) concordance with local indices presently being used as management tools in three geographic regions of US coastal waters, 2) classification accuracy for sites defined *a priori* as good or bad and 3) insensitivity to natural environmental gradients. US M-AMBI was highly correlated with all three local indices and removed the compression in response seen in moderately disturbed sites with US AMBI. US M-AMBI and US AMBI did a similar job correctly classifying sites as good or bad in local validation datasets (83 to 100% accuracy vs. 84 to 95%, respectively). US M-AMBI also removed the salinity bias of US AMBI so that lower salinity sites were not more likely to be incorrectly classified as impaired. The US M-AMBI appears to be an acceptable index for comparing condition across broad-scales such as estuarine and coastal waters surveyed by the US EPA's National Coastal Condition Assessment, and may be applicable to areas of the US coast that do not have a locally derived benthic index.

Keywords

benthic index; estuary; invertebrates; coastal assessment

1. Introduction

Macrobenthic invertebrate communities are a central part of estuarine and coastal condition assessment programs (Diaz et al., 2004; O'Brien et al., 2016). The interpretation of benthic community composition, particularly for a management audience, is typically achieved using indices that distill complex species composition data into easily communicated condition scores (Pinto et al., 2009). The AZTI-Marine Biotic Index (AMBI; Borja et al., 2000), an abundance-weighted, tolerance value index that assesses habitat condition based upon the relative abundance of taxa in different tolerance value groups, is one of the most frequently used indices in Europe (Borja et al., 2015).

This index is popular because it responds to human pressures (Borja et al., 2003; Muxika et al., 2005), does not require extensive calibration and validation datasets, and uses a generalized conceptual reference definition (sensu Stoddard et al., 2006), which includes indicators commonly used by experts when assessing the status (Borja et al., 2014). However, in a pan-European study, Grémare et al. (2009) showed some weaknesses in its way of assessing sensitivity/tolerance levels (i.e. existence of a single sensitivity/tolerance list) and recommended clarification of the sensitivity/tolerance levels for individual species. Reflecting this, shortly after the publication of AMBI (Borja et al., 2000) several authors published variants of AMBI (BENTIX (Simboura and Zenetos, 2002), and MEDOCC (Pinedo and Jordana, 2007)) to address discrepancies in the assignment of tolerance groups and differences in the disturbance gradient compared to the theoretical model it was based on.

In the years after AMBI and other variants have been published, several authors have illustrated that AMBI performance can be improved when using tolerance values tailored to the local setting (Rodil et al., 2013; Gillett et al., 2015; Robertson et al., 2016). In addition, AMBI performance is less robust when there are few individuals and species present in the sample, as would be expected in the low salinity portions of an estuary (Borja and Muxika, 2005). To address this problem, Muxika et al. (2007) combined AMBI scores with habitat measures of species richness and diversity to producing multivariate AMBI (M-AMBI).

In a previous study, Gillett and colleagues (2015) modified and expanded the ecological group (EG) classifications to create an integrated list of benthic species found along the west, gulf and east coasts of the US. Using this new and expanded list to calculate AMBI for US waters (US AMBI) improved performance of this index. US AMBI was able to differentiate between *a priori* good and bad sites from three different areas of the country, and was correlated with the local indices from these areas. However, it tended to compress scores towards moderate condition. In addition, the index was correlated with grain size and salinity. The correlation with salinity resulted in a misclassification of reference sites as degraded in oligohaline and tidal freshwater habitats (Gillett et al., 2015).

In this study, we addressed these issues by adapting the M-AMBI framework of Muxika et al. (2007) for the conterminous US coast. We then examined whether the US M-AMBI improved upon the AMBI using the same data sets previously used by Gillett et al. (2015) to evaluate US AMBI.

2. Materials and Methods

2.1. Approach

US AMBI was combined with two additional metrics to create US M-AMBI. The National Coastal Assessment (NCA) datasets used to develop the ecological group species list in our initial study (Gillett et al., 2015) was used to derive the High and Bad thresholds for each habitat (salinity zone) needed for the M-AMBI algorithm. These datasets were also used to select the additional metrics used in US M-AMBI. US M-AMBI was evaluated using the three regional datasets used to evaluate US AMBI in a previous study (Gillett et al., 2015). US M-AMBI was assessed for 1) concordance with local indices presently being used as management tools in three geographic regions of US coastal waters, 2) classification accuracy for sites defined *a priori* as good or bad and 3) insensitivity to natural environmental gradients.

2.2. NCA Calibration Dataset

Benthic invertebrate macrofaunal samples from 4,061 stations located in coastal waters of the conterminous US were collected during the summer months from 1999 to 2006 in the Atlantic, Gulf of Mexico, and Pacific waters of the US by the U.S. Environmental Protection Agency's National Coastal Assessment (NCA; U.S. EPA, 2016). Following local conventions, stations from the Pacific coast were sampled with a 0.1 m² grab and sieved on a 1-mm screen. In contrast, stations from the Atlantic and Gulf of Mexico coasts were sampled with a 0.04 m² grab and sieved on 0.5-mm screen. All specimens were identified to the lowest possible taxonomic level (typically species) and followed standard NCA QA/QC protocols for identification. They were further harmonized for taxonomy using the WoRMS (WoRMS Editorial Board, 2016) and ITIS (2016) databases.

All of the stations sampled for macrobenthos were also sampled for sediment chemistry (grain size, total organic carbon (TOC), heavy metals, PAHs, PCBs, etc.), sediment toxicity, and water quality (salinity, dissolved oxygen (DO), etc.) with sampling and laboratory protocols detailed in U.S. EPA (2016).

2.3. Calculation of Indices

M-AMBI is calculated by combining the AMBI score, Shannon-Weiner Diversity (H'), and species richness (S). The value of each metric for each sample is standardized and then combined via a factor analysis; factor scores are then placed along orthogonal gradients of condition created from a user-defined reference (High) and highly degraded (Bad) anchor points for each habitat. The resultant position in Euclidean space is the index score of the sample (Muxika et al., 2007). This approach allows the user to create local-specific expectations of condition and interpret benthic samples using a best attainable reference condition definition (*sensu* Stoddard et al., 2006).

In this study, habitat was defined as salinity zone, following the Venice Classification System (1958), as was done by Borja et al. (2008) to remove the salinity bias seen in US AMBI (Gillett et al., 2015). High and Bad thresholds were calculated for each of the metrics (Table 1). The Bad threshold was the worst possible value for that metric (e.g., AMBI score

of 6, diversity score of 0). The High threshold was based on the 95th percentile of the data for a metric that was higher at unimpacted sites (richness, diversity), and the 5th percentile for a metric that was higher at impacted sites (AMBI, % oligochaetes). West coast sites in both the NCA dataset and Southern California dataset (Ranasinghe et al., 2012) were sampled using a larger grab size which would be expected to inflate the species richness compared to smaller samples. Because of this, High thresholds were calculated separately for the polyhaline and euhaline habitats on the west coast. Higher species richness in lower salinity habitats on the west coast relative to similar habitats for the rest of the US was not observed, likely due to the low number of lower salinity samples from the west coast. For this reason, the lower salinity expectations were calculated for the entire US.

Although the ecological group classifications from Gillett et al. (2015) were used for this study, US AMBI was recalculated for all stations using raw rather than natural log transformed abundance, as transformation dampened the relationship between AMBI and chemical stressors, so that contaminated stations (Effects Range Median Quotient (ERMQ) >1) were classified primarily as slightly to moderately disturbed rather than moderately to highly disturbed (Figure 1). Classification accuracy was assessed for all habitats. Because the NCA datasets had no *a priori* good and bad sites, reference and impaired sites were calculated based on sediment contaminant data, amphipod toxicity, TOC, and DO concentrations (Table 2). For most habitats, US M-AMBI was calculated using US AMBI, species richness and diversity, resulting in classification accuracy greater than 70%. However, classification accuracy in the tidal freshwater habitat was only 41%, so additional metrics (dominance, % oligochaetes, and % naidid oligochaetes) were explored as a replacement for species richness. The percent oligochaetes metric was chosen to replace species richness as it resulted in the highest overall classification accuracy (73%). Naidid oligochaetes had comparable accuracy at 73%, but was not chosen as it was a subset of oligochaetes. Dominance had much lower accuracy at 57%.

Species richness, Shannon Weiner and dominance (lambda) were calculated using the DIVERSE routine in PRIMER 6.0. US M-AMBI was calculated using the M-AMBI portion of the scripts from Sigovini et al. (2013) in R version 3.1.2. The M-AMBI scores were separated into different condition classes (Table 3) based on Borja et al. (2012a).

2.4. Regional Validation Datasets

The three regional datasets used for validation in Gillett et al. (2015) were also used for this study (Table 4). These datasets were compiled from monitoring studies from the mid-Atlantic US (Delaware Bay, NJ/DE to Pamlico Sound, NC), Southeast (NC to southern FL), and Southern CA (Point Conception, CA to the US-Mexico border). All three datasets had a peer-reviewed benthic index (Table 5) that had been used for that area, and had benthic invertebrate data along with associated water quality, sediment quality and sediment toxicity data. Stations without salinity measures were not used for this study, with the exception of the Southern CA dataset where salinity was assumed to be ~30 (polyhaline) in open coastal sites. These datasets also had *a priori*-defined Good (High) and Bad sites (Hyland et al., 1999; Llanos et al., 2002; Weisberg et al., 2008) that were used to assess classification accuracy (Table 4).

2.5 Index Evaluation

First, the indices were assessed for concordance with previously published assessment indices presently being used as management tools in each of the three regions where the validation data sets were collected (Llanos et al., 2002; Smith et al., 2001, Van Dolah et al., 1999). Comparisons were made using all sites with matching local benthic index scores. Agreement with existing local benthic indices was assessed by Spearman's rank correlation using IBM SPSS, Version 20.

Second, the index was assessed for classification accuracy using the *a priori* good and bad sites that were used to develop the regional indices (Table 4). These comparisons were only conducted on a subset of stations that had an *a priori* classification. Agreement between index classification and *a priori* status was quantified as percent agreement from contingency tables generated using SPSS, Version 20. Kappa, a strength of agreement index (Cohen, 1960; Agresti, 1996) was also calculated and categorized according to Landis and Koch (1977). The condition thresholds used to define good or bad for each index are detailed in Table 6.

Third, the indices were assessed for resilience to natural salinity or sediment composition gradients, by calculating Spearman's rank correlations between index score and salinity or grain size (% silts and clays) at reference sites within each dataset. The correlations were only conducted on the subset of reference sites to ensure that impacts due to natural gradients were not confounded by the presence of anthropogenic stressors. Reference sites from each dataset were selected using the criteria in Table 2.

3. Results

3.1. Overall

US M-AMBI performed as well or better than US AMBI for all three evaluations: 1) concordance with the three local indices, 2) classification accuracy based on *a priori* good or bad sites and 3) insensitivity to natural environmental gradients.

3.2. Comparison of M-AMBI to regional indices

There was good agreement between the local indices and US M-AMBI (Table 7). Correlations between US M-AMBI and the local indices were >0.70 and were highly statistically significant. Overall classification accuracy of US M-AMBI relative to the local indices ranged from 86% to 97%, and kappa ranged from moderate to almost perfect.

While the US AMBI response was significantly correlated with the local index (Gillette et al., 2015), high variability was observed (Figures 2A, 3A, 4A). This was especially evident in the mid-Atlantic and southeast data. In the mid-Atlantic, sites classified as degraded using the local index were classified by US AMBI as undisturbed through heavily disturbed. In the southeast, sites classified as having healthy benthos by the local index ranged between undisturbed and moderately disturbed by US AMBI. The compression in response cited by Gillette et al. (2015) was especially evident in the southeast where sites classified as having unhealthy benthos by the local index were not classified as being highly disturbed by US

AMBI (Figure 3A) and in southern California where most sites classified as undisturbed by the local index were classified as slightly disturbed by US AMBI (Fig 4A). US M-AMBI, which adjusted for salinity and included two additional metrics in addition to US AMBI, was calculated for these same sites and greatly improved response relative to the local indices (Figures 2B, 3B, 4B). For all three validation datasets, US M-AMBI response was less variable, and more closely related to the local index scores than was US AMBI. The US M-AMBI response extended the range of condition, despite having fewer impaired sites than unimpaired or moderately impaired sites.

3.3. Comparison to a priori Good/Bad sites from regional validation datasets

US M-AMBI did an excellent job correctly classifying the *a priori* good-bad validation sites (Table 8). Overall classification accuracy ranged from 83% in the mid-Atlantic dataset to 100% in the southern California dataset. Strength of agreement (Landis and Koch, 1977) ranged from moderate in the mid-Atlantic dataset to almost perfect in the southeast and southern California datasets. This was equivalent or better than both the US AMBI index and classification of the local index (Table 9). Overall classification accuracy of US M-AMBI was equivalent or higher than US AMBI, and higher than the local indices. However, classification accuracy for all indices was > 75%.

3.4. Removal of natural environmental gradient bias

Grain size was correlated with US AMBI but not US M-AMBI in unimpacted sites in the mid-Atlantic validation dataset (Table 10). In the southeast dataset, grain size in unimpacted sites was not significantly correlated with US AMBI but was significantly correlated with US M-AMBI (Table 10); however, the sample size was very low and the range of grain size in the unimpacted sites was extremely truncated compared to the entire dataset. A significant relationship had previously been shown between US AMBI and salinity in unimpacted sites (Gillett et al., 2015). This was confirmed in this study for two of the three validation datasets (Table 10) with US AMBI calculated using raw rather than transformed abundance. The presence of a correlation between US AMBI and salinity was problematic because low salinity reference sites tended to have higher (worse) US AMBI scores than higher salinity reference sites (Figure 5A), even in the absence of anthropogenic stressors. There was no significant correlation between US M-AMBI and salinity for all datasets in unimpacted sites (Table 10). When US M-AMBI was calculated for these same sites, the lower salinity sites were classified as being in High or Good condition (Figure 5B), removing this source of bias.

4. Discussion

The US M-AMBI assessment tool developed for US coastal waters as part of this study performed as well or better across all performance criteria than did the US AMBI tool (Gillett et al., 2015). US M-AMBI eliminated the ‘compression’ of response relative to local index response seen with US AMBI by including additional metrics responsive to environmental stressors. US M-AMBI also removed the bias seen at low salinities with US AMBI (Gillett et al., 2015). The inclusion of salinity-calibrated community response metrics

with US AMBI scores produced a better performing index that could be used in a wider variety of estuarine habitats, as was observed by Muxika et al. (2007).

Poor performance in low salinity environments is a frequent concern with the development and application of estuarine benthic indices (Borja and Muxika, 2005; Weisberg et al., 1997; Thompson et al., 2013; Neto et al., 2014). These habitats present a challenge for condition assessment because of naturally lower species richness in low salinity environments (Attrill and Rundle, 2002; Ranasinghe et al., 2012), narrowing the range of response that indices must delineate. Another challenge is that lower salinity habitats transition between the freshwater systems upstream and saline waters downstream, so resident biota are subject to natural osmotic stress and high turbidity (Diaz, 1994; Draheim, 1998; Attrill and Rundle, 2002). This estuarine quality paradox (Elliott and Quintino, 2007; Dauvin and Ruellet 2009) in which structural community measures are confounded by natural physical stresses often leads to imperfect classifications using assessment tools designed for application across the full range of estuarine salinities and sediment types. By combining species tolerance values (i.e., AMBI scores) with salinity-calibrated community response measures, assessment index response can be improved.

In its original formulation, the European M-AMBI used scaled expectations of species diversity and species richness to overcome poor index performance in low salinity habitats (Muxika et al., 2007). Other authors tried also to overcome these problems, by modifying the method, the reference conditions or the condition class boundaries, both in Europe (Teixeira et al., 2009; Sigovini et al., 2013; van Loon et al., 2015) and in other continents (Borja and Tunberg, 2011; Forde et al., 2013; Cai et al., 2014; Feebarani et al., 2016). In our index development process, we considered additional macrobenthic community metrics that would have a wider range in response and potentially provide better discrimination in low salinity habitats than species richness or diversity. Based upon overall classification accuracy, the relative abundance of oligochaetes was determined to be a better complementary metric in tidal freshwater habitats. Using oligochaetes to partially evaluate the condition of these types of estuarine habitats fits well with our understanding of tidal freshwater and oligohaline estuarine ecology in which oligochaetes are much more diverse and the dominant component of the macrobenthic community (Diaz, 1994; Gillett et al., 2007; Gillett and Schaffner, 2009). This additional metric is similar to those used by experts when classifying samples subjectively (Weisberg et al., 2008; Teixeira et al., 2010).

One motivation for this study was to develop a macrobenthic-based assessment index that would be comparable across the entire conterminous US coast; however, conducting bioassessments at that scale is challenging. A good bioassessment index needs to be responsive to a variety of stressors (e.g., legacy and emerging toxic chemicals, habitat alteration, hypoxia) while being insensitive to natural environmental gradients (salinity, depth). For continental-scale application, a good index needs to perform in a consistent fashion across different habitats (e.g., cold upwelling influenced waters, embayments, or lagoons) (Grémare et al., 2009), especially when the index needs to be applied across environmental gradients in estuaries. One approach to address this issue would be to develop expectations for every region or potential habitat, as large-scale variability in benthic response can decrease index performance (Engle and Summers, 1999). The approach taken

in this study was to try to minimize the variance in benthic response due to natural environmental gradients while also minimizing the number of separate habitats examined, as setting reference condition for a large number of habitats can be problematic (Borja et al., 2012b). The final bioassessment index should serve to inform management and regulatory goals for a waterbody (i.e., beneficial uses or ecosystem services).

Although the focus of this study was to develop an index that could be used to compare estuaries across the US, the good correspondence with the local indices for all three validation datasets suggests that this index might be applicable areas of the United States without a local index. In these areas, care should be taken to ensure that sites are comparable to those used in this study. If not (e.g., offshore sites in deep water), additional habitats and habitat expectations may need to be defined to accurately characterize sites using US M-AMBI. The hybrid EG list may need to be supplemented in these local areas to ensure that enough local species are characterized. This may be achieved by applying genus level characterization to local species without an EG classification, or may require assembling a local benthic workgroup to determine ecological group assignments. US M-AMBI output should be assessed to determine whether it adequately characterizes local conditions. This may be done by comparing the index to physiochemical condition of sites, best professional judgement of sites (Weisberg et al. 2008), or multivariate assessments of the benthos. Finally, condition class boundaries might be altered to better refine application of the index for local application.

Although setting of condition classes has been a major issue for European waters, this has not been of major concern in US waters. Both the mid-Atlantic and southeast indices are indices of biotic integrity where the index thresholds are dependent on the underlying metrics, which are in turn dependent on an underlying, and set, reference distribution. The southern California index is a species tolerance index tolerance based an underlying pollution gradient, and the condition classes are in turn based on the percent deviation from reference conditions. These thresholds have shifted slightly since the initial paper by Smith et al. (2001) due to addition of data rather than a deliberate adjustment of thresholds. Based on the results of the present study, use of the condition classes of Borja et al (2012a) for application of US M-AMBI in a national assessment to assess condition and trends appear to be adequate at this time. Future work with US M-AMBI will examine the pressures, stressors and natural structuring and modifying factors impacting benthic condition as assessed by US M-AMBI. It will also be important to update the EG classifications as new species are added to the national species list.

Although there are a number of well-calibrated, commonly used macrobenthic indices that reflect local conditions well in various regions of the US (Diaz et al., 2004; Pinto et al., 2009; O'Brien et al., 2016), none can be applied across all of the nation's coasts. Most of these indices have a local specificity produced by responsiveness to local stressor exposures and calibration over the habitat gradients of their specific waterbody, but as such do not apply to continental scales because they are tied to local reference conditions. Development of indices that are comparable across broad-scales are needed for environmental assessment programs such as the US EPA's National Coastal Condition Assessment (NCCA), which supports the US Clean Water Act, and the European Water Framework Directive (WFD)

monitoring (Birk et al., 2012). The US M-AMBI developed here appears to meet the need to assess US coastal waters. As an index designed for continental-scale applicability, US M-AMBI comes close to providing the precision and accuracy of a locally-developed index, while also providing managers a tool to interpret local conditions in a national context.

Acknowledgments

We would like to thank the AMBI benthic workgroup members for helpful suggestions and John Kiddon, Sandi Robinson, Hugh Sullivan and two anonymous reviewers for their technical reviews. This manuscript has been reviewed by U.S. EPA's National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, Rhode Island and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is ORD Tracking# ORD-021143.

References

- Agresti, A. An introduction to categorical data analysis. John Wiley and Sons; New York: 1996.
- Attrill MJ, Rundle SD. Ecotone or ecocline: ecological boundaries in estuaries. *Estuar. Coast. Shelf Sci.* 2002; 55:929–936.
- Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecol. Indic.* 2012; 18:31–41.
- Borja A, Franco J, Pérez V. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Mar. Pollut. Bull.* 2000; 40:1100–1114.
- Borja A, Muxika I. Guidelines for the use of AMBI (AZTI's marine biotic index) in the assessment of the benthic ecological quality. *Mar. Pollut. Bull.* 2005; 50:787–789. [PubMed: 15950990]
- Borja A, Dauer D, Díaz R, Llansó RJ, Muxika I, Rodríguez JG, Schaffner L. Assessing estuarine benthic quality conditions in Chesapeake Bay: a comparison of three indices. *Ecol. Indic.* 2008; 8:395–403.
- Borja A, Tunberg BG. Assessing benthic health in stressed subtropical estuaries, eastern Florida, USA using AMBI and M-AMBI. *Ecol. Indic.* 2011; 11:295–303.
- Borja A, Mader J, Muxika I. Instructions for the use of the AMBI software (Version 5.0). *Revista de Investigacion Marina, AZTI-Technalia.* 2012a; 19:71–82.
- Borja A, Dauer DM, Gremare A. The importance of setting targets and reference condition in assessing marine ecosystem quality. *Ecol. Indic.* 2012b; 12:1–7.
- Borja Á, Marín SL, Muxika I, Pino L, Rodríguez JG. Is there a possibility of ranking benthic quality assessment indices to select the most responsive to different human pressures? *Mar. Pollut. Bull.* 2015; 97:85–94. [PubMed: 26099789]
- Cai W, Borja Á, Liu L, Meng W, Muxika I, Rodríguez JG. Assessing benthic health under multiple human pressures in Bohai Bay (China), using density and biomass in calculating AMBI and M-AMBI. *Mar. Ecol.* 2014; 35:180–192.
- Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 1960; 20:37–46.
- Dauvin JC, Ruellet T. The estuarine quality paradox: is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems? *Mar. Pollut. Bull.* 2009; 59:38–47. [PubMed: 19084874]
- Diaz RJ. The response of tidal freshwater macrobenthos to sediment disturbance. *Hydrobiologia.* 1994; 278:201–212.
- Diaz RJ, Solan M, Valente RM. A review of approaches for classifying benthic habitats and evaluating habitat quality. *J. Environ. Manage.* 2004; 73:165–181. [PubMed: 15474734]
- Draheim, RC. Masters Thesis. The College of William and Mary; Gloucester Point, Virginia, USA: 1998. Tidal Freshwater and Oligohaline Benthos: Evaluating the Development of a Benthic Index of Biological Integrity for Chesapeake Bay.

- Elliott M, Quintino V. The estuarine quality paradox: environmental homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. *Mar. Pollut. Bull.* 2007; 54:640–645. [PubMed: 17418874]
- Engle VD, Summers JK. Refinement, validation, and application of a benthic condition index for Gulf of Mexico estuaries. *Estuaries.* 1999; 22:624–635.
- Forde J, Shin PK, Somerfield PJ, Kennedy RM. M-AMBI derived from taxonomic levels higher than species allows Ecological Status assessments of benthic habitats in new geographical areas. *Ecol. Indic.* 2013; 34:411–419.
- Gillett DJ, Holland AF, Sanger DM. On the ecology of oligochaetes: monthly variation of community composition and environmental characteristics in two South Carolina tidal creeks. *Estuar. Coast.* 2007; 30:238–252.
- Gillett DJ, Schaffner LC. Benthos of the York River. *J. Coast. Res.* 2009; (Special Paper 57):80–98.
- Gillett DJ, Weisberg SB, Grayson T, Hamilton A, Hansen V, Leppo EW, Pelletier MC, Borja A, Cadien D, Dauer D, Diaz R, Dutch M, Hyland JL, Kellogg M, Larsen PF, Levinton JS, Llansó R, Lovell LL, Montagna PA, Pasko D, Phillips CA, Rakocinski C, Ranasinghe JA, Sanger DM, Teixeira H, VanDolah RF, Velarde RG, Welch KI. Effect of ecological group classification schemes on performance of the AMBI benthic index in US coastal waters. *Ecol Indic.* 2015; 50:99–107.
- Grémare A, Labruno C, Vanden Berghe E, Amouroux JM, Bachelet G, Zettler ML, Vanaverbeke J, Fleischer D, Bigot L, Maire O, Deflandre B, Craeymeersch J, Degraer S, Dounas C, Duineveld G, Heip C, Herrmann M, Hummel H, Karakassis I, Kedra M, Kendall M, Kingston P, Laudien J, Occhipinti-Ambrogi A, Rachor E, Sarda R, Speybroeck J, Van Hoey G, Vincx M, Whomersley P, Willems W, Włodarska-Kowalczyk M, Zenetos A. Comparison of the performances of two biotic indices based on the MacroBen database. *Mar. Ecol. Prog. Ser.* 2009; 382:297–311.
- Hyland JL, Van Dolah RF, Snoots TR. Predicting stress in benthic communities of southeastern U.S. estuaries in relation to chemical contamination of sediments. *Environ. Toxicol. Chem.* 1999; 18:2557–2564.
- Hyland J, Balthis L, Karakassis I, Magni P, Shine J, Vestergaard O, Warwick R. Organic carbon of sediments as an indicator of stress in the marine benthos. *Mar. Ecol. Prog. Ser.* 2005; 295:91–103.
- Integrated Taxonomic Information System. [accessed 1.05.16] 2016. <https://www.itis.gov/>
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
- Llansó RJ, Scott LC, Hyland JL, Dauer DM, Russell DE, Kutz FW. An estuarine benthic index of biotic integrity for the mid-Atlantic region of the United States. II. Index development. *Estuaries.* 2002; 25:1231–1242.
- Long, ER., Morgan, LG. NOS OMA 52. NOAA Technical Memorandum. NOAA NOS; Seattle, WA: 1990. The Potential for Biological Effects of Sediment Sorbed Contaminants Tested in the National Status and Trends program.
- Long ER, MacDonald DD, Smith SL, Calder FD. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ. Manage.* 1995; 19:81–97.
- Long ER, Field LJ, MacDonald DD. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ. Toxicol. Chem.* 1998; 17:714–727.
- Muxika I, Borja Á, Bald J. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. *Mar. Poll. Bull.* 2007; 55:16–29.
- Neto JM, Feio MJ, Teixeira H, Patrício J, Serra SRQ, Franco JN, Calapez AR, Constantino E. Transitional and freshwater bioassessments: one site, two perspectives? *Mar. Pollut. Bull.* 2014; 78:153–164. [PubMed: 24246653]
- O'Brien A, Townsend K, Hale R, Sharley D, Pettigrove V. How is ecosystem health defined and measured? A critical review of freshwater and estuarine studies. *Ecol. Indic.* 2016; 69:722–729.
- Pelletier MC, Campbell DE, Ho KT, Burgess RM, Audette CT, Detenbeck NE. Can sediment total organic carbon (TOC) and grain size be used to diagnose organic enrichment in estuaries? *Environ. Toxicol. Chem.* 2011; 30:538–547. [PubMed: 21298700]

- Pinedo, S., Jordana, E. Spain (Catalonia and Balearic Islands). In: Carletti, A., Heiskanen, A-S., editors. Water Framework Directive Intercalibration Technical Report Part 3: Coastal and Transitional waters. JRC Scientific and Technical Reports. JRC, ies; 2007. p. 62-70.
- Pinto R, Patrício J, Baeta A, Fath BD, Neto JM, Marques JC. Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecol. Indic.* 2009; 9:1–25.
- Ranasinghe JA, Welch KI, Slattery PN, Montagne DE, Huff DD, Lee H II, Hyland JL, Thompson B, Weisberg SB, Oakden JM, Cadien DB, Velarde RG. Habitat-related benthic macrofaunal assemblages of bays and estuaries of the western United States. *Integr. Environ. Assess. Manag.* 2012; 8:638–48. [PubMed: 22987518]
- Robertson BP, Savage C, Gardner JP, Gardner A, Robertson BM, Stevens LM. Optimising a widely-used coastal health index through quantitative ecological group classifications and associated thresholds. *Ecol. Indic.* 2016; 69:595–605.
- Rodil IF, Lohrer AM, Hewitt JE, Townsend M, Thrush SF, Carabine M. Tracking environmental stress gradients using three biotic integrity indices: advantages of a locally-developed traits-based approach. *Ecol. Indic.* 2013; 34:560–570.
- Simboura N, Zenetos A. Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new biotic index. *Mediterr. Mar. Sci.* 2002; 3:77–111.
- Sigovini M, Keppel E, Tagliapietra D. M-AMBI revisited: looking inside a widely-used benthic index. *Hydrobiologia.* 2013; 717:41–51.
- Smith RW, Bergen M, Weisberg SB, Cadien DB, Dalkey A, Montagne DE, Stull JK, Velarde RG. Benthic response index for assessing infaunal communities on the southern California mainland shelf. *Ecol. Appl.* 2001; 11:1073–1087.
- Smith, RW., Ranasinghe, JA., Weisberg, SB., Montagne, DE., Cadien, DB., Mikel, TK., Velarde, RG., Dalkey, A. Extending the southern California Benthic Response Index to assess benthic condition in bays. Technical Report No 410. Southern California Coastal Water Research Project; Westminster, CA: 2003.
- Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH. Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecol. Appl.* 2006; 16:1267–1276. [PubMed: 16937796]
- Teixeira H, Neto JM, Patrício J, Veríssimo H, Pinto R, Salas F, Marques JC. Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool. *Mar. Poll. Bull.* 2009; 58:1477–1486.
- Teixeira H, Borja A, Weisberg SB, Ranasinghe JA, Cadien DB, Dauer DM, Dauvin J, Degraer S, Diaz RJ, Grémare A, Karakassis I, Llansó RJ, Lovell LL, Marques JC, Montagne D, Occhipinti-Ambroggi A, Rosenberg R, Sardá R, Schaffner LC, Velarde RG. Assessing coastal benthic macrofauna community condition using best professional judgement – Developing consensus across North America and Europe. *Mar. Pollut. Bull.* 2010; 60:589–600. [PubMed: 19969316]
- Teixeira H, Weisberg SB, Borja A, Ranasinghe JA, Cadien CA, Velarde RG, Lovell LL, Pasko D, Phillips CA, Montagne DE, Ritter KJ, Salas F, Marques JC. Calibration and validation of the AZTI's Marine Biotic index (AMBI) for Southern California marine bays. *Ecol. Indic.* 2012; 12:84–95.
- Thompson B, Ranasinghe JA, Lowe S, Melwani A, Weisberg SB. Benthic macrofaunal assemblages of the San Francisco Estuary and Delta. *Environ. Monit. Assess.* 2013; 185:2281–2295. [PubMed: 22684808]
- Tweedley JR, Warwick RM, Clarke KR, Potter IC. Family-level AMBI is valid for use in the north-eastern Atlantic but not for assessing the health of microtidal Australian estuaries. *Estuar. Coast. Shelf Sci.* 2014; 141:85–96.
- U.S. Environmental Protection Agency. National Coastal Condition Assessment 2010. EPA/841/R-15/006. Office of Water and Office of Research and Development; Washington, DC; 2015.
- U.S. Environmental Protection Agency. [accessed 1.05.16] Environmental Monitoring & Assessment Program. 2016. <https://archive.epa.gov/emap/archive-emap/web/html/index-37.html>
- Van Dolah RF, Hyland JL, Holland AF, Rosen JS, Snoots TR. A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. *Mar. Environ. Res.* 1999; 48:269–283.

- van Loon WMGM, Boon AR, Gittenberger A, Walvoort DJJ, Lavaleye M, Duineveld GCA, Verschoor AJ. Application of the Benthic Ecosystem Quality Index 2 to benthos in Dutch transitional and coastal waters. *J. Sea Res.* 2015; 103:1–13.
- Venice System. Symposium on the classification of brackish waters. Venice, April 8–14, 1958. *Archives for Oceanography and Limnology.* 1958; 11(Suppl):1–248.
- Weisberg SB, Ranasinghe JA, Schaffner LC, Diaz RJ, Dauer DM, Frithsen JB. An estuarine index of biological integrity (B-IBI) for Chesapeake Bay. *Estuaries.* 1997; 20:149–158.
- Weisberg SB, Thompson B, Ranasinghe JA, Montagne DE, Cadien DB, Dauer DM, Diener D, Oliver J, Reish DJ, Velarde RG, Word JQ. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. *Ecol. Indic.* 2008; 8:389–394.
- WoRMS Editorial Board. [accessed 1.05.16] World Register of Marine Species. 2016. Available from <http://www.marinespecies.org> at VLIZ

Figure 1. Comparison of US AMBI calculated using raw abundance (closed circles) or the natural log of abundance (open circles) in comparison to the Effects Range Median Quotient (ERMQ; Long et al., 1998). ERMQ is a composite measure of sediment contamination.

Figure 2.
Comparison of Mid-Atlantic benthic index to A) US AMBI B) US M-AMBI. Note that for the Mid-Atlantic benthic index, low scores indicate more impaired condition where high scores indicate less impaired condition.

Figure 3. Comparison of Southeast benthic index to A) US AMBI B) US M-AMBI. Note that for the Southeast benthic index, low scores indicate more impaired condition where high scores indicate less impaired condition.

Figure 4.

Comparison of Southern California benthic index to A) US AMBI B) US M-AMBI. Note that for the Southern California benthic index high scores indicate more impaired condition where low scores indicate less impaired condition. This is opposite of the other two regional indices

Figure 5.

Reference sites from the Mid-Atlantic validation data used to show relationship between the benthic index and salinity A) Bias in US AMBI due to salinity B) No bias in US M-AMBI due to salinity

Table 1

Reference (High) and highly degraded (Bad) anchor points for each habitat used to calculate M-AMBI scores

Salinity Bin	Region	Scale	AMBI	Species Richness	Diversity (H')	% Oligochaetes
All	NE, SE, Gulf, West	Bad	6	0	0	100
Tidal Freshwater	NE, SE, Gulf, West	High	0.15		1.93	0.00
Oligohaline	NE, SE, Gulf, West	High	0.53	16.0	2.12	
Mesohaline	NE, SE, Gulf, West	High	0.85	26.0	2.48	
Polyhaline	NE, SE, Gulf	High	0.72	44.0	2.96	
Polyhaline	West	High	0.18	76.8	3.30	
Euhaline	NE, SE, Gulf	High	0.56	61.0	3.29	
Euhaline	West	High	0.66	92.0	3.62	
Hyperhaline	NE, SE, Gulf, West	High	0.32	55.0	3.45	

Table 2

Environmental thresholds used to select reference sites to assess sensitivity to salinity and grain size and classification accuracy for each habitat in the NCA Calibration dataset

Stressor	National Threshold	
	Reference	Degraded
Sediment Chemistry ^a	0 ERL Exceedances	1 ERM Exceedances
Amphipod Toxicity ^b	90 % Survival	< 75 % Survival
TOC ^c	< 2 %	> 3.5 %
DO ^d	> 5 mg/L	< 2 mg/L
Criteria	Site must meet all thresholds	Site must meet at least 2 thresholds

^aLong and Morgan (1990), Long et al. (1995); ERL (effect range low) is the level below which adverse effects are not likely to be observed, ERM (effect range median) is the level at which adverse effects are likely to be seen.

^bUS EPA (2015).

^cHyland et al. (2005), Pelletier et al. (2011).

^dPelletier et al. (2011). For the Southern CA validation data set, missing DO values were assumed to be 5 mg/L since many of the stations were open ocean sites assumed to be well oxygenated.

Table 3

US M-AMBI Condition Classes

US M-AMBI Condition	US M-AMBI
Bad	< 0.20
Poor	0.20 – 0.39
Moderate	0.39 – 0.53
Good	0.53 – 0.77
High	> 0.77

Table 4

Information on the validation datasets used in this study.

Validation Dataset	Environmental Factors	Number of <i>a priori</i> Sites	<i>a priori</i> Good and Bad Site Determination	Reference
Mid-Atlantic (N = 484)	salinity	0 – 35	Sites were classified based on dissolved oxygen concentrations, sediment contamination, and amphipod or Microtox toxicity	Llanos et al. 2002
	% silt-clay	0 – 99.2		
	depth (m)	0.1 – 34		
Southeast (N = 57)	salinity	0.1 – 36.5	Sites were classified based on Effects Range Median Concentrations (ERMQ) based on 8 metals, 13 PAHs, total PCBs and 2 pesticides.	Hyland et al. 1999
	% silt-clay	0.3 – 98.9		
	depth (m)	0.3 – 15		
Southern California (N = 545)	salinity	27.2 – 39.4	Sites were classified based on best professional judgement of macrobenthic community condition	Weisberg et al. 2008
	% silt-clay	1 – 100		
	depth (m)	0.4 – 30		

Table 5

Validation data sets benthic indices and cut-offs used for this study.

	MAIA Benthic Index	MAIA Condition	
Mid-Atlantic	< 3	Degraded	Llanos et al. (2002)
	3	Non-Degraded	
	Southeast B-IBI score	Southeast Condition	
Southeast	< 1.5	Unhealthy benthos	Van Dolah et al. (1999)
	1.5–3	Some stress	
	3	Healthy benthos	
	BRI Score	Southern CA Condition	
Southern California	< 39.96	Undisturbed	Smith et al. (2001)
	39.96 – 49.15	Low disturbance	Smith et al. (2003)
	49.15 – 73.27	Moderate disturbance	
	73.27	High disturbance	

MAIA = Mid-Atlantic Integrated Assessment

B-IBI = Benthic Index of Biotic Integrity

BRI = Benthic Response Index

Table 6

Thresholds used to convert index values into good or bad categories to assess classification accuracy

Benthic Index	Good index value	Good index category	Bad index value	Bad index category
US M-AMBI	> 0.53	Good Condition High Condition	< 0.39	Bad Condition Poor Condition
US AMBI	< 3.3	Undisturbed Slightly Disturbed	3.3	Moderately Disturbed Heavily Disturbed
MAIA BI	3	Non-Degraded	< 3	Degraded
B-IBI	3	Healthy Benthos	< 3	Unhealthy Benthos Some Stress
BRI	49.15	Undisturbed Low Disturbance	> 49.15	Moderate Disturbance High Disturbance

MAIA = Mid-Atlantic Integrated Assessment

B-IBI = Benthic Index of Biotic Integrity

BRI = Benthic Response Index

Table 7

Agreement between local indices and US M-AMBI using local datasets from three regions of the U.S.

Dataset	Local Index	Correlation (Spearman's ρ)	Overall Classification Accuracy (%)	Kappa
Mid-Atlantic	MAIA Benthic Index	$r = 0.715$ ($p < 0.0005$)	87.3	0.65 (substantial)
Southeast	Southeast B-IBI	$r = 0.829$ ($p < 0.0005$)	97.4	0.95 (almost perfect)
Southern California	Southern CA BRI	$r = -0.746$ ($p < 0.0005$)	86.3	0.49 (moderate)

MAIA = Mid-Atlantic Integrated Assessment

B-IBI = Benthic Index of Biotic Integrity

BRI = Benthic Response Index

Table 8
Accuracy of US M-AMBI using *a priori* Good/Bad sites for each of the validation datasets.

Dataset	US M-AMBI results relative to reference-impaired classification			
	Bad	Poor	Moderate	Good
Mid-Atlantic	Reference	0	1	5
	Stressed	1	17	7
	Classification Accuracy = 83.2% kappa = 0.57 (moderate)			43
Southeast	Undegraded	0	1	3
	Degraded	2	11	3
	Classification Accuracy = 90.3% kappa = 0.81 (almost perfect)			9
Southern California	Reference	0	0	0
	Marginal Deviation from Reference	0	0	1
	Affected	0	2	3
	Severely Affected	2	4	0
	Classification Accuracy = 100% kappa = 1.00 (almost perfect)			4

Southern California has four *a priori* categories, while the Southeast and Mid-Atlantic have only two

Table 9

Agreement between US M-AMBI, US AMBI and the local benthic index with the *a priori* Good/Bad sites for each of the validation datasets.

Dataset	Index	Overall Classification Accuracy (%)	Kappa
MAIA	US M-AMBI	83.2	0.57 (moderate)
	US AMBI	84.0	0.65 (substantial)
	MAIA BI	78.2	0.48 (moderate)
Southeast	US M-AMBI	90.3	0.81 (almost perfect)
	US AMBI	86.5	0.73 (substantial)
	B-IBI	83.9	0.68 (substantial)
Southern California	US M-AMBI	100.0	1.00 (almost perfect)
	US AMBI	95.2	0.90 (almost perfect)
	BRI	95.2	0.90 (almost perfect)

MAIA = Mid-Atlantic Integrated Assessment

B-IBI = Benthic Index of Biotic Integrity

BRI = Benthic Response Index

Table 10

Correlation between environmental variables (salinity, % silt-clay) and benthic indices. The correlations are based on National threshold reference sites only (see Table 1).

Dataset	Variable	Correlations			
		US M-AMBI	US AMBI	MAIA BI	
Mid-Atlantic	Bottom salinity	Spearman's rho	0.043	-0.613**	0.017
		p-value	0.742	<0.0005	0.933
		N	61	61	26
	% Silt-clay	Spearman's rho	0.133	0.575**	0.110
		p-value	0.347	<0.0005	0.675
		N	52	52	17
Southeast	Bottom salinity	Spearman's rho	-0.410	-0.900*	-0.671
		p-value	0.493	0.037	0.215
		N	5	5	5
	% Silt-clay	Spearman's rho	0.975**	0.500	0.783
		p-value	0.005	0.391	0.118
		N	5	5	5
Southern California	Bottom salinity	Spearman's rho	-0.008	-0.132	-0.156
		p-value	0.973	0.568	0.499
		N	21	21	21
	% Silt-clay	Spearman's rho	0.131	0.156	-0.238
		p-value	0.342	0.255	0.080
		N	55	55	55

MAIA = Mid-Atlantic Integrated Assessment

B-IBI = Benthic Index of Biotic Integrity

BRI = Benthic Response Index

* Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript