Ecological Indicators 45 (2014) 308–312

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Is there a significant relationship between the benthic status of an area, determined by two broadly-used indices, and best professional judgment?

Ángel Borja[a], Sandra Marín[b,*], Rosa Núñez[b], Iñigo Muxika[a]

[a] AZTI – Tecnalia, Marine Research Division, Herrera Kaia Portualdea s/n, 20110 Pasaia, Spain
[b] Institute of Aquaculture, Austral University of Chile, P.O. Box 1327, Puerto Montt, Chile

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 7 April 2014
Accepted 7 April 2014

Keywords:
Best professional judgment
Benthic macroinvertebrates
Human disturbance
Benthic quality assessment
Chile

ABSTRACT

Two benthic indices to assess the quality status (the AZTI's Marine Biotic Index (AMBI) and multivariate-AMBI (M-AMBI)) are being used extensively in different habitats worldwide. We try to interpret what is behind these indices making them suitable for different habitats. To demonstrate that, we used best professional judgment (BPJ), applying it to a dataset from southern Chile, to determine the criteria proposed by 12 experts in assessing the status. The experts were provided with raw species abundance data, from 12 stations within a gradient of disturbance, from unaffected to severely affected. There was a very good agreement among experts (kappa values 0.72–0.77), with highly significant ($p < 0.001$) correlation between BPJ and AMBI and M-AMBI classifications, and an agreement of 76.4% and 81.9%, respectively. When comparing BPJ in Chile with other results in Europe, USA and northern Africa, similar patterns can be identified: (i) the number of criteria identified for classification is very high (range 7–12); (ii) the experts use several criteria together in the BPJ assessment; and (iii) the rank of the most important criteria is indicator species, richness, and diversity/dominance. These criteria are included in indices such as AMBI and M-AMBI. Hence, although experts are classifying samples subjectively when applying BPJ, they are corroborated in their opinions when using such indices. This fact can explain why these indices are so widely used.

1. Introduction

In recent times, there is an increasing need to assess the status of the oceans, making sustainable human marine activities e.g. fishing, aquaculture, etc. (Halpern et al., 2012). Assessing the status requires tools allowing defining marine health in a comprehensible way (Borja et al., 2013; Tett et al., 2013). Environmental legislation worldwide has led to an over-inflation of methods to assess the status of different ecosystem components (Birk et al., 2012).

One of the most used components in assessments is soft-bottom benthic macroinvertebrates (Pearson and Rosenberg, 1978; Weisberg et al., 1997; Dauer et al., 2000; Borja et al., 2003). The benthic indices available to assess the status are numerous (Diaz et al., 2004; Marques et al., 2009; Pinto et al., 2009).

Some of the benthic indices are used in local areas, and sometimes using them over large geographic areas can be problematic, because they are usually developed within specific habitats (Borja and Dauer, 2008). There is no certainty that indices developed in different regions or habitats assess biological condition on the same scale (Teixeira et al., 2010). However, there are two indices (AZTI's Marine Biotic Index: AMBI; Borja et al. (2000); and multivariate-AMBI: M-AMBI; Muxika et al. (2007)), which are being used extensively in different habitats worldwide, from estuaries to deep oceanic waters and from tropical areas to polar ones (see Borja et al. (2012) for a summary on the use of M-AMBI). But, why AMBI and M-AMBI are so frequently used in assessing marine benthic quality worldwide? Our hypothesis is that benthic status of a given area determined by AMBI and M-AMBI is highly correlated with the benthic status assigned by professional judgment.

Interpreting what is behind an index which makes it suitable for different habitats could be complicated, yet needed, especially when anthropogenic activities are to be environmentally regulated by these indices, e.g. in Europe, these and other indices are legally bound in assessing the benthic ecological status (Borja et al., 2009a). One potential solution to interpret this is to apply best professional judgment (BPJ) to establish the criteria that experts use in assessing the status of 'blind' samples across regions. This method has been demonstrated as useful in providing a uniform scale for calibrating

* Corresponding author. Tel.: +56 65 2277119.
E-mail address: smarin@spm.uach.cl (S. Marín).

http://dx.doi.org/10.1016/j.ecolind.2014.04.015
1470-160X/

any index, in different regions (Weisberg et al., 2008; Teixeira et al., 2010; Dauvin et al., 2012).

In order to explain the broad use of AMBI and M-AMBI in assessing the status in those regions our objectives were: (i) to evaluate the level of agreement among experts using BPJ to assess the condition of benthic communities in southern Chile; (ii) to determine the criteria used by experts to assess the status; and (iii) to determine if those criteria are common across geographies in other regions of the world.

2. Methods

From a project whose objective was to monitor environmental performance of salmon culture in the Inner Sea of Chiloé, southern Chile (41°28′ S–72°56′ W; 43°71′ S–73°27′ W), 12 samples were selected to encompass a range of conditions from unimpacted to highly disturbed benthic communities. The samples were collected with 0.1 m^2 Van Veen grab and sieved through 1-mm screen. After identification and counting, samples were ordered from best to worst and selected at even intervals using AMBI (Borja et al., 2000). In the calculation we used the free software available at http://ambi.azti.es, with the species list of February 2012. Hence, three samples were classified as "unaffected", which is the best situation, and should bear a community at a least affected or unaffected (undisturbed) site by human activities; three samples were classified as "marginal deviation from unaffected", which is a community that shows some indication of stress or disturbance, but within the measurement error of unaffected condition; three samples were classified as "affected", where there is confidence that the community shows evidence of physical, chemical, or anthropogenic stress, with a clear unbalanced situation; and three were classified as "severely affected", where the magnitude of stress is high and benthic communities are very disturbed.

Twelve benthic experts from Chile were provided species-abundance data from those 12 sampling locations and asked to determine the condition of the benthos at each location, using their own expertise and criteria. The experts neither know the level of disturbance nor the location in Chile. Of the 12 benthic ecologists, 10 were from academic institutions (universities and research institutes), and 2 from private consulting firms. Benthic experts came from the north part of Chile (30°24′ S–71°19′ W) to the very south, the Magellan region (53°36′ S–70°53′ W) and their experience in benthic monitoring ranged from 6 to 30 years.

The experts were asked to rank the relative condition of the sites from 1 (best) to 4 (worst) within each sample. As commented above, the four condition categories or quality classes should be related to: (1) "unaffected"; (2) "marginal deviation from unaffected"; (3) "affected"; and (4) "severely affected". The experts would designate these conditions using the criteria to classify them as liberally as each expert desired, without recommending them on the use of any index or method, only based on their experience, as done in other research elsewhere (Teixeira et al., 2010).

However, the experts were also asked to identify the criteria they used to evaluate the benthos and rate their importance as follows: (1) very important; (2) important, but secondary; (3) marginally important; and (4) useful, but only to interpret other factors. Criteria that were not used by an expert were assigned a rank of 5 for the purpose of calculating an average importance of that attribute among the experts.

In addition to AMBI values, M-AMBI values were calculated, using the same software. As the sampling area has reference stations (unaffected), the lowest AMBI value and the highest richness and Shannon's diversity values were used as reference conditions in the M-AMBI calculation.

Table 1
Condition categories assigned by the 12 benthic experts to each of the 12 samples. Key to condition categories: 1 – "unaffected"; 2 – "marginal deviation from unaffected"; 3 – "affected"; 4 – "severely affected".

Experts	Samples											
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Exp 1	1	2	2	4	3	3	2	4	2	1	4	1
Exp 2	3	1	2	4	3	1	2	4	1	2	4	3
Exp 3	3	1	2	4	3	1	3	4	1	2	4	2
Exp 4	3	2	2	4	2	1	3	4	2	2	4	3
Exp 5	2	2	1	4	2	1	3	4	2	2	4	1
Exp 6	2	3	2	4	2	3	2	3	3	2	3	2
Exp 7	2	3	1	4	3	1	2	4	3	2	4	3
Exp 8	2	1	1	4	2	1	2	4	1	1	4	1
Exp 9	2	1	1	4	2	1	2	4	2	3	4	2
Exp 10	3	2	1	4	3	1	3	4	2	2	4	3
Exp 11	2	1	1	3	2	1	2	4	1	2	4	2
Exp 12	2	1	1	4	2	1	3	4	1	3	4	2

Although AMBI and M-AMBI have five assessment categories, when assessing the ecological status in Europe (after Muxika et al., 2007), we have used here the four abovementioned categories in order to compare results with those obtained by Weisberg et al. (2008), in California; Teixeira et al. (2010), in Europe and US; and Dauvin et al. (2012), in Europe and northern Africa. To make them comparable, the moderate and poor categories in AMBI and M-AMBI, as defined in Muxika et al. (2007), were merged into "affected".

The level of agreement on condition categories assigned by the experts and those calculated using AMBI and M-AMBI was evaluated using Kappa analysis (Cohen, 1960; Landis and Koch, 1977) by establishing the level of agreement using the equivalence table of Monserud and Leemans (1992), from null to perfect. Fleiss–Cohen weights were applied (Fleiss and Cohen, 1973) because misclassifications between distant categories (e.g. between "unaffected" and "affected", or "unaffected" and "severely affected") are more important than misclassifications between closer categories (e.g., between "unaffected" and "marginal deviation from unaffected", or "affected" and "severely affected").

The level of agreement in ranking sites among all the experts and AMBI and M-AMBI values was evaluated using regression analysis.

3. Results

There was substantial agreement in condition categories assigned by the experts (Tables 1 and 2). This is especially remarkable for the worst quality locations (S4, S8 and S11), in which 11 out of the 12 experts agreed on sample condition category. For those locations with the best condition (S6, S9 and S10), there is a high agreement in S6 (10 out of the 12 experts). In contrast, for intermediate quality stations, there was some dispersion in the categories assigned by experts.

There also was some deal of consensus in ranking of samples (Table 2) among the experts. Hence, samples unaffected or with marginal deviation from unaffected (as classified by AMBI) present the lowest mean condition categories values, samples affected show intermediate values, and those severely affected present the highest values. Something similar can be shown when compared with M-AMBI values.

The regression between mean condition categories values assigned by expert judgment and both AMBI and M-AMBI values were highly significant ($p < 0.001$) (Figs. 1 and 2).

Kappa analyses indicated a very good agreement among experts in their four condition category assignments, when compared with AMBI (kappa value of 0.72) and M-AMBI (kappa value of 0.77) (Table 3). However, some mismatch can be detected in this table when compared at the level of good ('unaffected'/'marginal