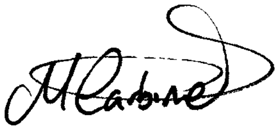
Auckland
Regional Council
TE RAUHĪTANGA TAIAO

State of Environment Indicators for Intertidal Habitats in the Auckland Region

August TR 2010/035

Auckland Regional Council
Technical Report No.035 August 2010
ISSN 1179-0504 (Print)
ISSN 1179-0512 (Online)
ISBN 978-1-877540-91-2

Reviewed by:



Name: Dr Megan Carbines
Position: Project Leader
Organisation: ARC
Date: 10.09.2010

Approved for ARC Publication by:

Name: Grant Barnes
Position: Group Manager
Organisation: ARC
Date: 10.09.2010

Recommended Citation:
van Houte-Howes, K.; Lohrer, D. (2010). State of Environment Indicators for intertidal habitats in the Auckland Region. Prepared by NIWA for Auckland Regional Council. *Auckland Regional Council Technical Report 2010/035.*

State of Environment Indicators for Intertidal Habitats in the Auckland Region

Kristel van Houte-Howes
Drew Lohrer

Prepared for
Auckland Regional Council

NIWA
PO Box 11115
Hamilton 3251

Project Number ARC10211

Contents

Reviewed by: Approved for release by:

Dr S. Thrush Dr J. Hewitt

Formatting checked

A. Bartley

1 Executive Summary

The Auckland Regional Council requires easily understandable information on the ecological integrity of invertebrate communities in estuarine and coastal areas. This information can be used for State of Environment reporting or more generally for communicating to the public about the health status of coastal habitats.

Overseas, a number of indices of ecological integrity have been developed, but not without some scientific controversy. Here we investigate two overseas indices, AMBI and B-IBI, using Auckland Regional Council data sets, to see whether these indices correlated with gradients of heavy metal concentration and sediment mud content in Auckland area estuaries. The temporal variability of each index in non-impacted locations was assessed using the ARC's Manukau Harbour monitoring data. Lastly, the performance of a new index based on New Zealand functional diversity data, called NIWACOOBII, was compared to that of AMBI and B-IBI.

Non-impacted sandflats (benthic monitoring sites) in Manukau Harbour were rated by AMBI as "unpolluted" and "slightly polluted", and index values at these sites were observed to be very stable over time. Monitoring sites in Mahurangi Harbour were also rated as unpolluted / slightly polluted, though the significant gradient in sediment mud content among sites in the Harbour was not detected by the AMBI index. AMBI scores calculated for 95 Regional Discharges Project sites demonstrated a lack of sensitivity to detect what is believed to be a reasonably strong and ecologically significant contaminant gradient. The relationship between AMBI scores and percent muddiness at the RDP sites was stronger than for metals but remained weak overall ($r^2 < 0.10$).

Similar to the AMBI, the B-IBI index classified all of the Manukau and Mahurangi Harbour sites as unpolluted or slightly polluted. Again, the correlation between B-IBI scores and muddiness was weak for the Mahurangi monitoring sites (where mud content varied between 9 and 47%) and for the 95 RDP sites (where mud content varied between 0 and 96%). The RDP metal gradient was also poorly tracked with B-IBI scores.

These results imply that we should be concerned about the ability of the overseas indices to detect change before catastrophic degradation. However, they also imply moderate to high environmental quality in our valuable harbour and estuarine ecosystems relative to at least some other locations.

The NIWACOOBII index was developed based on the richness of species in 7 functional groupings, with changes in index values reflecting potential shifts in ecological resilience. These functional groups were selected for consistency of response to increased muddiness and metals in the Mahurangi and RDP datasets. Accordingly, the NIWACOOBII was more effective than AMBI or B-IBI at tracking gradients of muddiness (both datasets) and heavy metals (RDP dataset). Although the index needs to be further refined and tested on independent data, in conjunction with the Benthic Health Model, it offers a useful way of assessing some of the elements of ecosystem health in our harbours and estuaries.

2 Introduction

Benthic indices that assess the ecological integrity of invertebrate communities in estuarine and coastal areas are being developed in many different countries (Borja et al. 2008, Weisberg et al. 1997, Bremner et al. 2006, de Juan et al. 2009, Rosenberg et al. 2004, Llansó et al. 2002, Borja & Muxika 2005) in response to legislation such as the "Clean Water Act" in the USA or the Water Framework Directive (WFD) and Marine Strategy Directive (MSD) in Europe. This has stimulated the development of an increasing number of tools for assessing ecological integrity or status (Borja et al. 2008).

The development of a benthic index usually involves:

i) defining criteria for degraded and un-degraded 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 un-degraded 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.

The ARC, which requires information on the ecological integrity of invertebrate communities in estuarine and coastal areas for State of Environment reporting, expressed interest in trialling indices that have proven useful overseas. The ARC also sought information on the efficacy of indices being developed here in New Zealand (supported by NIWA's Coasts & Oceans OBI). Thus, indices from Europe, USA and New Zealand were applied to existing ARC data sets and assessed in terms of ease of calculation, variability among years and seasons, and sensitivity to contaminant gradients. In this report, we compare and contrast some of these indices and discuss their suitability for detecting two environmental stressors (mud and heavy metals) that are recognised as major threats to the health and functioning of Auckland area estuaries.

2.1 Indices used in the European Union and USA.

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 opportunistic species. This index has been used in Europe primarily, but has also been applied in Asia, northern Africa and South America (Borja et al. 2008).

Although AMBI was designed to assess effects of organic over-enrichment, it has subsequently been used to account for the effects of different types of stressors (e.g., Borja et al. 2003; Muxika et al. 2005). Another positive feature of the index is its free availability on the internet (www.azti.es) and the production of easily understandable plots by the software in a standard format. However, this index is generally designed to compare marine communities of the same basic type, and thus habitat variation in the area of interest reduces the utility of AMBI.

The Benthic Index of Biotic Integrity (B-IBI) was developed in the USA by Weisberg et al. (1997). It 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. Measured input parameters to the B-IBI include species diversity, productivity, indicator species and trophic composition. An advantage of B-IBI over AMBI is that it accounts for habitat variation by using reference sites. (Note, reference sites are also a key element of the European Water Directive; patterns of change at monitored sites are rated relative to those at reference sites). Like AMBI, the B-IBI can be used to assess different types of stressors. However, information on the "pollution sensitivity" of all the species (which likely varies with pollution type) is critical to model performance.

A review of the international literature revealed several other indices in addition to AMBI and B-IBI (Table 1). However, these other indices tended to be modifications or local applications of the original indices, AMBI and B-IBI. For example, the index called BENTHIX (Simboura & Zenetos 2002) is essentially identical to AMBI, except that it uses two sensitivity groupings for species instead of five groupings like AMBI. A multivariate extension to the AMBI has also been developed and given a new name, M-AMBI (Borja et al. 2004, Bald et al. 2005, Muxika et al. 2007). The most similar index to the B-IBI is called MAIA, which was developed for the Chesapeake Bay region of the USA (Llansó et al. 2002, Llansó & Dauer 2002). Like B-IBI, MAIA is based on a scoring system approach. One of earliest indices developed (the abundance-biomass comparison, or ABC; Warwick 1986, Clark 1990) was also considered. However, this index and any others that required measurements of macrofaunal biomass were not pursued further, as biomass data is both difficult and time consuming to collect and is currently not recommended for Auckland Regional Council monitoring. We therefore focused primarily on the two best known and most often applied indices, AMBI and B-IBI.

Table 1:

A listing of some of the available indices of biological integrity used in Europe and the USA, along with short descriptions and the primary references where index development was discussed.

Index Name	Locale	Input and Calculation Details	References
AMBI (AZTI's Marine Biotic Index)	Europe	Species are apportioned into 5 pollution sensitivity groups, from very sensitive to very hardy. Software for calculations freely available at http://www.azti.es, or in latest version of PRIMER	Borja et al. 2008 Borja et al. 2000 Borja & Muxika 2005 www.azti.es – free software
B-IBI (Benthic Index of Biotic Integrity)	USA	Attributes of benthic community structure and function (abundance, biomass, Shannon diversity etc.), are scored according to thresholds established from reference data.	Weisberg et al. 1997 http://sci.odu.edu/chesapeakebay/data/benthic/BIBIcalc.pdf Borja et al. 2008
BENTHIX	Europe	Same as AMBI, but based on proportions of species in only two sensitivity groups: 1 = sensitive, 2= tolerant. Available in PRIMER.	Simboura & Zenetos 2002 Borja et al. 2000
M-AMBI (Multivariate-AMBI)	Europe	A multivariate extension of AMBI (see above) that compares monitoring results with reference conditions. Software for calculations freely available at http://www.azti.es	Borja et al. 2004 Bald et al. 2005 Muxika et al. 2007
MAIA (Mid-Atlantic Integrated Assessment)	USA	An integrated average score of a combination of metrics (abundance, diversity, species and trophic composition, percent abundance of pollution sensitive and insensitive species/taxa) that performed best according to several criteria.	Llansó et al. 2002 Llansó & Dauer 2002

Index Name	Locale	Input and Calculation Details	References
BTA (Biological Traits Analysis)	Europe	Species are assigned to functional or traits groups and changes in functional composition relative to known stressors are investigated using fuzzy correspondence analysis (FCA), co-inertia analysis (CoI) and/or non-metric MDS	Bremner et al. 2006 de Juan et al. 2009
BQI (Benthic Quality Index)	Europe	Tolerance values were determined for benthic species. Based on a combination of species tolerance values, abundance and diversity a benthic quality index (BQI) was calculated. MDS used to analyse how different BQI were distributed.	Rosenberg et al. 2004
RBI (Relative Benthic Index)	Europe	Based on responses of marine benthic communities to anthropogenic and natural disturbances, using 6 categories (total number of species, number of crustacean species, number of mollusc species, number of crustacean individuals, and the presence or absence of positive and negative indicator species); developed for particular areas by selecting different indicator species.	Anderson et al. 1998 Anderson et al. 2001
ABC (Abundance-Biomass-Comparison)	Europe	K-dominance curves for species abundances and biomasses are plotted on the same graph. Position of one curve relative to the other identifies three sediment conditions: undisturbed, moderately disturbed and grossly disturbed. *W* statistic produced.	Warwick 1986 Clark 1990

2.2 The Benthic Health Model (BHM) and a newer functional index, NIWACOOBII

In 2001, the ARC commissioned a study to develop criteria related to urban stormwater impacts in estuaries based on sediment heavy metal contaminant data coupled with macrofaunal community composition data (Anderson et al. 2002). The goal was to use the criteria to classify the health of sites and to provide a means by which health status could be tracked through time for management purposes. While complicated statistical techniques were used (ordination), easily interpretable plots were produced showing shifts in benthic community composition across a gradient in sediment heavy metal concentration. Further refinements to the Benthic Health Model (BHM) were presented by Hewitt et al. (2005) and testing continues to show the value of this approach (Anderson et al. 2006, Hewitt et al. 2009, Hewitt & Ellis *Draft Report*). However, it is not advisable to apply the BHM to areas outside of the Waitemata-Manukau region, for which it was developed, until its broader applicability is demonstrated. Additionally, the model is explicitly focused on stormwater/heavy metal contaminants despite a recognition that many other potential stressors (and stressor interactions) can influence macrobenthic community structure. Finally, although the BHM demonstrates clear and statistically significant shifts in community composition in conjunction with relatively modest increases in sediment metal contamination (sensitivity is a key strength of the model), the BHM does not indicate which types of species are being affected or whether individual species abundances are increasing or decreasing. Some of this information can be ascertained with more detailed examination of the data, but the development of other complementary indicators of ecosystem health and integrity directly related to ecosystem function would be of benefit. In this report, we provide details of a functional traits index, called NIWACOOBII, in which we examine increases and deceases in the number of individuals and taxa in 7 functional trait groups in response to environmental stress gradients.

3 Methods

3.1 AZTI's Marine Biotic Index (AMBI)

The AMBI index was designed to rank the ecological quality of European coastal areas, analysing the response of soft bottom communities to natural and human-mediated changes in water and sediment quality (Borja et al. 2000). This index is based on the degree of sensitivity/tolerance of different types of species to an environmental stress gradient (e.g., increasing organic enrichment). The index recognises five distinct ecological groups whose abundances are supposed to vary predictably according to stress levels (Borja et al. 2000, Borja, 2005):

Group I. Species very sensitive to organic enrichment and present under unpolluted conditions (initial state). They include the specialist carnivores and some deposit feeding tubicolous polychaetes.

Group II. Species indifferent to enrichment, always present in low densities with non-significant variations with time. These include suspension feeders and generalist (less selective) carnivores and scavengers.

Group III. Species tolerant to excess organic matter enrichment. These species may occur under normal conditions, but their populations are stimulated by organic enrichment. They are surface deposit-feeding species, such as tubicolous spionids.

Group IV. Second-order opportunistic species. Mainly small sized polychaetes: subsurface deposit feeders, such as cirratulids. More abundant in polluted conditions than Group III, less common in polluted conditions than group V.

Group V. First order opportunistic species. These are deposit feeders that proliferate in reduced (low oxygen) sediments and tend to be absent from unpolluted sites that lack organic enrichment.

The index values calculated by the AMBI model vary depending on the relative abundances of individuals in each of the above-listed ecological groups. The AMBI calculation also involves a weighting system (0, 1.5, 3, 4.5, 6, see formula below) that gives Group V species the highest weight (Borja et al. 2000, Borja, 2005).

The AMBI is simply a sum of the weighted relative abundances of individuals in the difference ecological groups, i.e.,

$$[(0\times\%GroupI)+(1.5\times\%GroupII)+(3\times\%GroupIII)+(4.5\times\%GroupIV)+(6\times\%GroupV)]/100$$

With pollution-tolerant Group V species having the highest associated weight, the highest AMBI scores reflect the most polluted sites. The output produced using the AMBI equation can be reported in a discrete categorical format called the Biotic Index (BI) or in a continuous format called the Biotic Coefficient (BC). Obviously, the BI and BC values correspond to one another; BI 0 denotes BC values between 0 and 0.2, BI 1 is for BC values of 0.2 to 1.2, BI 3 is for BC values of 1.2 to 3.3, etc. (see Borja et al.

2000, for details). Guidelines for interpreting AMBI outputs are given by Borja et al. (2000): BI scores of 0 – 1 reflect unpolluted sites, scores of 2 – 4 indicate slightly or moderately polluted sites, and scores of 5 – 7 correspond to heavily or extremely polluted sites.

3.1.1 Assigning Species to AMBI's Five Ecological Groups

The data we examined with AMBI were obtained from the ARC's long term monitoring programmes in Manukau and Mahurangi Harbours and from 95 Regional Discharges Project (RDP) sites in the Waitemata and Manukau Harbours. The data provided species identifications and average abundances at each site for all monitored species in each estuary. To enable us to apply AMBI in a New Zealand context and because most of the AMBI species listed are from European and South American biogeographical areas (http://www.azti.es) it was necessary to assign the New Zealand macrobenthic species to one of the five ecological groups.

The following steps to assign species not on the AMBI list were as follows;

(1) References were consulted providing lists of pollution sensitive and tolerant species (Hewitt et al. 2009; Gibbs & Hewitt, 2004).

(2) When a New Zealand species was not found on the AMBI list, but the same genus was, the species was assigned to the same group.

(3) When the genus of a New Zealand species was not found on the AMBI list, but a member of the same family was, the species was assigned to the same group.

(4) When neither the genus or family of a New Zealand species was found on the AMBI list, but a species in the same super-family was, the species was assigned to the same group.

(5) A few species remained unassigned (such as the crabs *Austrohelice crassa* and *Heteroplax hirtipes*, and the surface deposit feeding horn snail *Zeacumantus lutulentus*).

Although the method used to categorise species into ecological groups suffered from some taxonomic vagueness, we were able to match most of our species to an AMBI equivalent to create a species list acceptable to the AMBI model. Following species assignments, AMBI values were calculated using the free software available on the AZTI's webpage (www.azti.es).

3.2 Benthic Index of Biotic Integrity (B-IBI)

The B-IBI index was developed to assess benthic community health and environmental quality at sites in Chesapeake Bay (east coast, USA). The B-IBI evaluates the ecological condition of a sample by comparing values of benthic community attributes to reference values expected under non-degraded conditions in similar habitat types. The B-IBI is calculated by comparing the value of a metric (related to benthic community structure and function, e.g., species diversity, productivity, species composition, and trophic composition), from a sample of unknown quality to thresholds established from reference data distributions. These thresholds called "restoration goals" were established as the 5th or 95th and 50th (median) percentile values of reference sites for each metric-habitat combination. Each metric is scored on a 5, 3 or 1 scale, depending on whether its value at a site approximates, deviates slightly from, or deviates greatly from conditions at reference sites. These scores are then averaged to form the index.

Samples with index values of 3.0 or more are considered to have good benthic condition indicative of good habitat quality (Weisberg, 1997; Llanso & Dauer, 2002).

3.2.1 Scoring of Metrics

The B-IBI is designed to account for variability in benthic communities according to habitat. Habitat factors considered to be most important in affecting the index are salinity and sediment type. Samples are therefore assigned to a salinity class ranging from tidal freshwater (0 ppt) to polyhaline (≥18.0 ppt). Our data only related to the polyhaline (PO) salinity class. Within the polyhaline class, samples have to be further assigned into two sediment classes according to the silt-clay content of the sample: "mud" has a silt-clay (<63 μm) content by weight of >40%, whereas "sand" has a silt-clay content of 0 – 40%.

Metrics recommended for use with PO sand habitats were Shannon-Weiner species diversity index, total species abundance, total species biomass, % abundance of pollution sensitive taxa, % biomass of pollution indicative taxa, % abundance of carnivores and omnivores & % abundance of deep-deposit feeders. Metrics for use by PO mud habitats were Shannon-Weiner species diversity index, total species abundance, total species biomass, % biomass of pollution indicative taxa, % biomass of pollution sensitive taxa, % abundance of carnivores and omnivores.

Because of limitations in the data and information available, we used the abundance based metrics where species-specific biomass was unavailable, as recommended by Llanso & Dauer (2002). Similarly, information about pollution sensitive taxa and pollution indicative taxa was limited, so these were combined into a single metric called % abundance of pollution sensitive taxa. Sensitivity scores were based on sensitivity information already available from assessments made using the AMBI model to assign species to ecological groups. Abundance of carnivores, omnivores and deposit feeders was assessed using existing knowledge of feeding behaviour. The final indices calculated for both sand and mud habitats were: Shannon-Weiner species diversity index, total species abundance, % abundance of pollution sensitive taxa, % abundance of carnivores and omnivores and % abundance of deep-deposit feeders.

3.2.2 Selection of Reference Sites

Reference sites are those that show no chemical contaminant impact (Weisberg et al. 1997). In order to test the B-IBI model, reference data distributions were calculated from mud and sand habitats. The sites used for the analysis were selected on the following basis using contaminant and percent silt-clay content data available from 95 RDP sites.

Sites were initially sorted according to the two applicable categories, mud (>40%) and sand (0 – 40%), which allocated approximately half of the sites to each category. Sites were then further sorted on the basis of increasing values along a pollution gradient (defined by heavy metal concentrations). The contaminant levels were taken from the Benthic Health Model of Anderson et al. (2006) and correspond to the CAP model principal component axis 1 from whole sediment samples (<500 μm = PC1.500). In the Benthic Health Model, five groups along a gradient from non-contaminated to contaminated were identified. The non-contaminated sites (Group 1) corresponded to PC1.500 values less than -1.9. Ideally only these sites would be used as our reference sites. However, when this rule was applied to the selected muddy sites from our data set, none of the sites met the criterion, and only a small number of sandy sites did. To circumvent this issue, the non-contaminated site limit was instead set to a much higher level of PC1.500 = 0 to include an adequate number of both muddy and sandy sites, which, according to the Benthic Health Model, included all sites from Groups 1 and 2 and a few from Group 3. Refer to Anderson et al. (2006) for more details. This left us with two sets of reference sites, one for muddy habitats and one for sandy habitats. The index was then validated by examining its response to a new set of sites with known environmental stress, which were RDP sites with a PC1.500 value greater than zero (0). The index was further validated by examining its response to the Manukau and Mahurangi datasets.

3.2.3 Selection of Data Sets for Index Testing

As mentioned above, existing data from Manukau, Mahurangi and Waitemata Harbours were used for index testing. All of the data were collected in association with ARC funded programmes. The reasons for selecting particular data for use in our analyses depended on the particular questions being addressed.

To assess the sensitivity of the indices in detecting a pollution gradient, the RDP data set was ideal. There were 95 sites sampled in the RDP programme, with mean abundances of various macrofaunal taxa and sediment metal concentrations (copper, lead and zinc concentrations) measured at each site. The positions of the sites were specifically selected to encompass a gradient of storm water contaminants and subsequent analyses of sediment metal concentrations confirmed the gradient. Furthermore, the RDP data were the basis of the Benthic Health Model. While individual concentrations of contaminants were available both from the <63um and the <500um fraction of the sediment, data from the <500um fraction was used (following Anderson et al. 2006). Individual metal values (i.e., zinc, lead, and copper separately) were used as was the combined metal metric given by PC1.500 axis values.

To asses the versatility of the indices (i.e., their ability to detect other pollutant gradients besides metals), data from Mahurangi Harbour was used. Increasing muddiness has been identified as a stressor in Mahurangi Harbour (Halliday & Cummings 2009). Five intertidal sites in the Harbour are regularly monitored (CB, HL, JB, MH, TK) and these sites vary in mud content from low (JB <10%) to high (HL

>40%). Three October sampling times were used, Oct-1994, Oct-1995 and Oct-2005. We examined years from early on in the Mahurangi monitoring programme (1994, 1995) relative to a later year (2005) due to indications of muddy terrigenous inputs during the intervening time period. We wanted to see if the indices might reflect changes associated with such trends.

To assess the natural variability of index values at uncontaminated sites, we used the data from six sites in the Manukau Harbour that have been monitored for over 20 years. Invertebrate data from six sandy sediment sites collected in 1989 – 1992 (Site EB, KP, PS), 1989-1995 (Site CH) and 1989-1997 (Site AA, CB) were used. As many benthic macrofauna exhibit seasonal cycles in abundance (Hewitt et al. 1994, Cummings et al. 2001), only data from October was used, which avoided the major recruitment peaks of most of the dominant species.

3.3 Functional Index (NIWACOOBII)

An analysis of the abundance of individuals and taxa in 29 distinct functional groupings (Table 2) was performed on two data sets encompassing stress gradients: Mahurangi (muddiness) and RDP (metal contamination). First, a master list of intertidal soft-sediment species that have been found in and around North Island was compiled. Each species was then assigned to the functional groupings defined in Table 2 using in-house knowledge and the best available information from the literature. The functional groupings were based on macrofaunal attributes that included feeding behaviours, positions in the sediment column, degrees of motility, types of topographic features created (tubes/pits/mounds), body sizes, body shapes, and so on. In some cases, "fuzzy" coding was used when the role of a species did not fall distinctly into one category or another. For example, organisms can be coded as both "Top" (found in the upper 0-2 cm of the sediment column) and "Deep" (found in the 2-10 cm sediment horizon) by assigning each code a value of 0.5.

Only a small subset of the species on the master list is going to be present at any particular site and time. For Mahurangi, numbers of individuals and taxa in each of the 29 functional groups ($Ninds_{group,}$ $Ntaxa_{group}$) were calculated at five sites on three occasions each. The five sites were situated along a sedimentation gradient from 1 (least muddy) to 5 (most muddy), with JB =1, MH = 2.5, TK = 2.5, CB = 4, and HL = 5. All macrofauna present in the samples (i.e., not just the routinely monitored taxa) were analysed on each of the three occasions (Oct 1994, Oct 1995, Oct 2005). For the RDP data, $Ninds_{group}$ and $Ntaxa_{group}$ were calculated at all 95 sites (generally different sampling occasions for each). Metal contaminant values for each site were taken directly from PC1.500 of the Benthic Health Model (Anderson et al. 2006).

For each data set, correlations between stress level (either mud or metals) and functional composition ($Ninds_{group}$ and $Ntaxa_{group}$) were calculated, thus there were 116 correlations performed (29 functional groups x 2 stress types x 2 response variables). We tabulated results on the number of correlations that were positive versus negative (i.e., increasing versus decreasing $Ninds_{group}$ and $Ntaxa_{group}$ with increasing stress) and also examined the strength and significance of correlation coefficients (Pearson's *r* values). Based on the results (see Results section), one functional grouping from each of the 7 functional categories was used to construct the index. Index values were standardized to fall between 0.0 and 1.0, with maximum index values (1.0) determined from maxima in the RDP and Mahurangi data sets.

Table 2:

Listing of the 29 functional groupings used in the NIWACOOBII analysis. The asterisks next to Body size and Degree of motility indicate that no fuzzy coding was used because the corresponding functional groupings (middle column) were mutually exclusive.

Functional Category	Functional Groupings	Code
Body shape/type	Calcium-shelled	Calcium
	Globular-shaped (length ≈ width)	Globular
	Worm-shaped (length >>> width)	Worm
Body size *	Large	Large
	Medium	Medium
	Small	Small
Degree of motility *	Freely motile on or in sediment	Free
	Limited movement, usually in sediment	Limited
	Sedentary / movement in a fixed tube	Sedentar
	Semi-pelagic	Spel
Direction of sediment particle movement	Depth to depth	DD
	Depth to surface	DS
	Surface to depth	SD
	Surface to surface	SS
Feeding behaviour	Deposit feeder	Dep
	Grazer	Grazer
	Predator	Pred
	Scavenger	Scav
	Suspension feeder	Sus
Living position	Attached	Attached
	Deeper than 2 cm	Deep
	Surface epifauna	Epif
	Top 2 cm	Top
Sediment topography feature created	Permanent burrow	Burr
	Erect structure / tube	Erect
	Simple hole or pit	Hole
	Mound	Mound
	Trample marks	Trample
	Trough	Trough

3.3.1 Analysis

To understand the ability of each index to track known environmental gradients, we regressed index scores versus rankings of muddiness and heavy metal contamination at Mahurangi Harbour and RDP sites. The percent variability explained by each index was indicated by r^2 values from least squares regression fits. We also compared

AMBI, B-IBI and NIWACOOBII values with Benthic Health Model outputs. Scatterplots were presented to illustrate the distribution of the data and the level of agreement between B-IBI, AMBI, NIWACOOBII, PC1.500, muddiness and BHM values.

4 Results

4.1 AZTI's Marine Biotic Index (AMBI)

This index is based on the degree of sensitivity/tolerance of different species in a community to an environmental stressor. Using the AMBI model, datasets from Mahurangi, Manukau and the RDP project produced BC scores in the range of 0.3 to 4.3, which corresponded to BI ranks ranging from 1 (unpolluted) to 4 (moderately polluted). The AMBI index classified 47% of the RDP sites correctly when compared to contaminant levels from PC1.500 values (<0). Figure 1 shows AMBI results for four haphazardly selected RDP sites (Site 1 = Ann's Creek 2002, Site 2 = Ann's Creek 2005, Site 3 = Auckland Airport, Site 4 = Awatea Rd). The AMBI scores at these four sites were 4, 3, 1 and 2. These sites had been previously categorized by the Benthic Health Model as having ranks of 5, 4, 1 and 4 (Anderson et al. 2006), which is reasonably similar to the trend of the AMBI output. Figure 1 also shows the proportions of individuals at each site that were present in AMBI's ecological groupings I-V.

In contrast to a few RDP sites that received AMBI ratings of BI 3 or 4, all of the monitoring sites in Manukau and Mahurangi were classified as BI 1 (unpolluted) or BI 2 (slightly polluted). The relative stability of AMBI values at the Manukau monitoring sites is shown in Figures 2-7 (October data at CH 1989-95, AA 1989-97, CB 1989-97, EB 1989-92, KP 1989-92, PS 1989-92). Site AA was the most stable site for the longest period, and had AMBI classifications of BI 1 (unpolluted) in each of the 8 years analysed. Site CH was the least stable of the Manukau sites across years, particularly with respect to the changing proportions of individuals in the five ecological groupings.

The relationship between sediment contaminants and AMBI BC scores across the 95 RDP sites is shown in Figure 8. The linear least squares regression fit of BC scores versus PC1.500 values was poor and explained <4% of the variability (r^2 = 0.0377). The correlation between BC scores and sediment mud content at these sites was also low (Fig. 9), but was somewhat better (r^2 = 0.1180) than the correlation with PC1.500. The trend was for increasing AMBI BC scores with increasing mud content, suggesting declining numbers of sensitive species and increasingly impacted sites.

Relationships between AMBI scores and individual metals (Cu, Zn, Pb 500) at the RDP sites are also shown (Figs. 10 & 11). Movements from sites scored BI 1 to sites scored BI 3 (i.e., from unpolluted to moderately polluted sites) did not coincide with significant increases in sediment metal concentrations (Fig. 10). However, there were weak positive trends for copper and zinc (Cu r^2 = 0.0601, Zn r^2 = 0.0356; Fig. 11).

The relationship between BC scores and muddiness at the Mahurangi sites was weak (Fig. 12, r^2 = .034). However, as expected along a gradient of increased stress, the trend was for increasing AMBI values with increased muddiness, both overall and for the individual years of 1994, 1995 and 2005 (Fig. 12, top). Trends across years (within sites) were more variable and revealed no significant harbour-wide trend of increased sedimentation impacts (Fig.12, bottom). In fact, AMBI scores suggested somewhat improved conditions at MH, HL and TK in 2005 relative to 1994-1995 (Fig.12, bottom).

We also tested for correlations between aspects of the biotic assemblage and the index scores (Figs. 13 & 14). Average species abundance, diversity and % sensitive species for all RDP data were compared across sites with the same BI scores. The RDP data was used because this dataset provided the best range of BI scores (1-4) for evaluation. Average species abundance and % sensitive species decreased from BI 1 to 2 (Fig. 13). However, species abundance increased in BI 3, while the number of sensitive species decreased. Conversely, species diversity increased from BI 1 to 2 and then decreased (BI 3) (Fig. 13).

BC scores mainly reflected the percentage of sensitive species in the RDP data set; they were less related to macrofaunal diversity or total abundance (Fig. 14). When BC scores were regressed against the abundance of sensitive species, the explained variability was 90% ($r^2 = 0.9038$, Fig. 14, bottom). Similarly, the abundance of sensitive species explained variability in the data from Mahurangi and the Manukau by 90% and 89% respectively. These high correlations reflect the manner in which the AMBI index is constructed and how it relies on information on species sensitivities to organic enrichment in order to produce results. There was essentially no relationship between BC scores and the abundance of deep deposit feeders or carnivores and omnivores at the RDP sites ($r^2 < 0.02$). However, in Mahurangi and Manukau, the percent variability explained for these groups was higher (deep deposit feeders, 10-28%, carnivores and omnivores up to 34%).

Figure 1:

AMBI output at four RDP sites. Left vertical axis gives the percentages of species in ecological groups I-V (see legend on Figure). Right vertical axis gives BI scores on a scale from 0 (unpolluted) to 7 (extremely polluted). Sites 1 to 4 are Ann's Creek 2002, Ann's Creek 2005, Auckland Airport and Awatea Rd, with BHM rankings of 5, 4, 1, and 4 respectively.

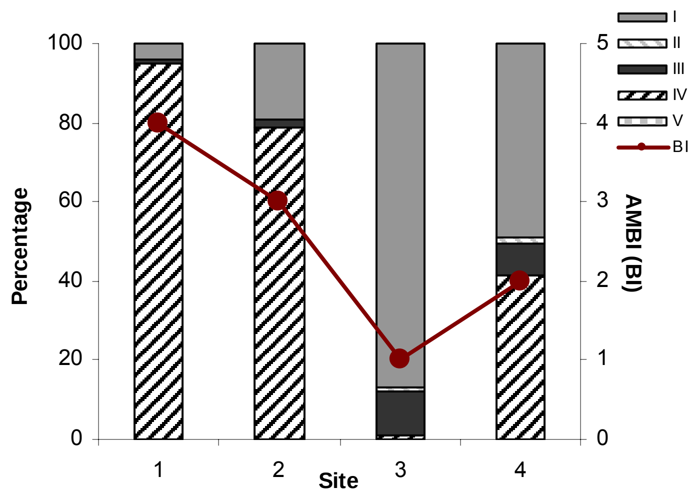


Figure 2:

MBI output at Site CH in Manukau Harbour, 7 consecutive years of October data. See Fig. 1 legend for plot details. The percentage of group I species (i.e., those sensitive to organic enrichment) is lower at this site than at other Manukau monitoring sites (see subsequent figures). During 1989-1995, discharge from a sewage treatment oxidation pond influenced CH.

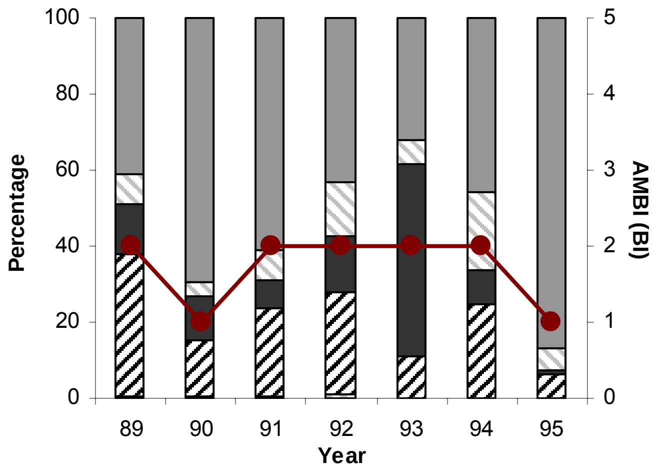


Figure 3:

AMBI output at Site AA in Manukau Harbour, 9 consecutive years of October data. See Fig. 1 legend for plot details. Site AA was scored BI 1 (unpolluted) and dominated by ecological group I in all 9 years. We know from the monitoring data and other research that AA is the most stable of the Manukau monitoring sites (Thrush et al. 2008a, Hewitt & Thrush 2009).

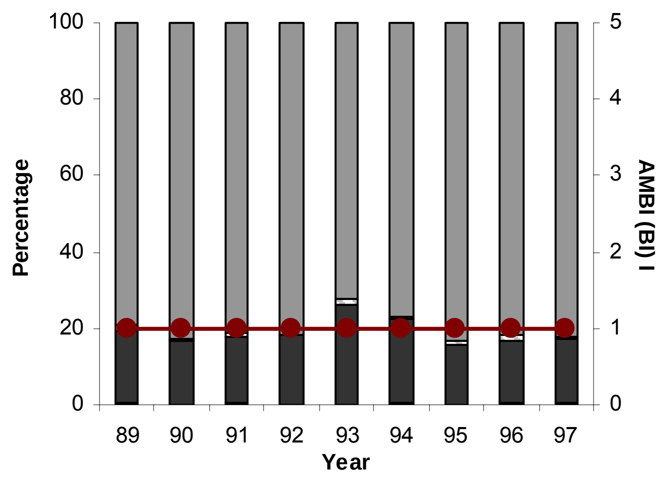


Figure 4:

AMBI output at Site CB in Manukau Harbour, 9 consecutive years of October data. See Fig. 1 legend for plot details. The BI score at CB in 1995 was 2 (slightly polluted). BI scores in all other years were 1 (unpolluted).

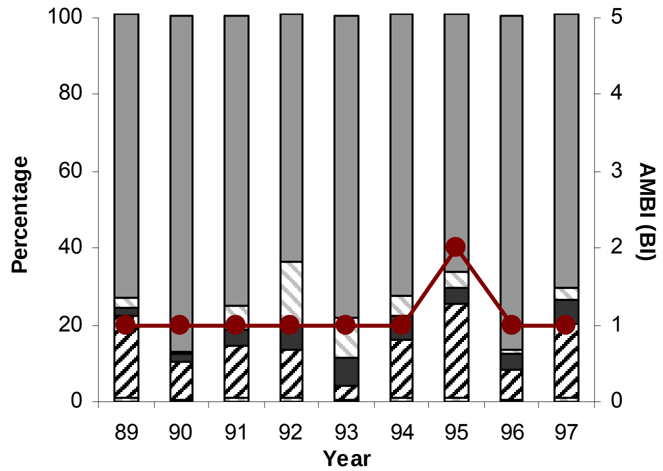


Figure 5:

AMBI output at Site EB in Manukau Harbour, 4 consecutive years of October data. See Fig. 1 legend for plot details. The BI score at EB was stable from 1989-1992 (2, slightly polluted).

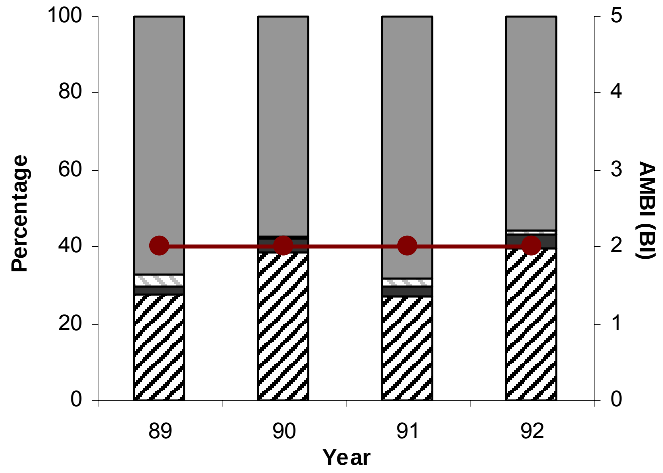


Figure 6:

AMBI output at Site KP in Manukau Harbour, 4 consecutive years of October data. See Fig. 1 legend for plot details. The BI score at KP was stable from 1989-1992 (2, slightly polluted). Sites KP, EB and PS all had stable BI scores of 2 (slightly polluted).

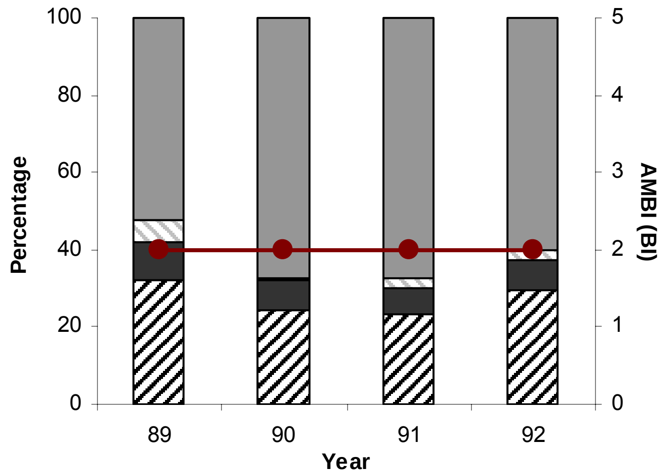


Figure 7.:

AMBI output at Site PS in Manukau Harbour, 4 consecutive years of October data. See Fig. 1 legend for plot details. The BI score at PS was stable from 1989-1992 (2, slightly polluted), which was the same as at EB and KP. However, unlike EB and KP, the percentage of the most sensitive group I species (gray) was >60% at PS in all 4 years.

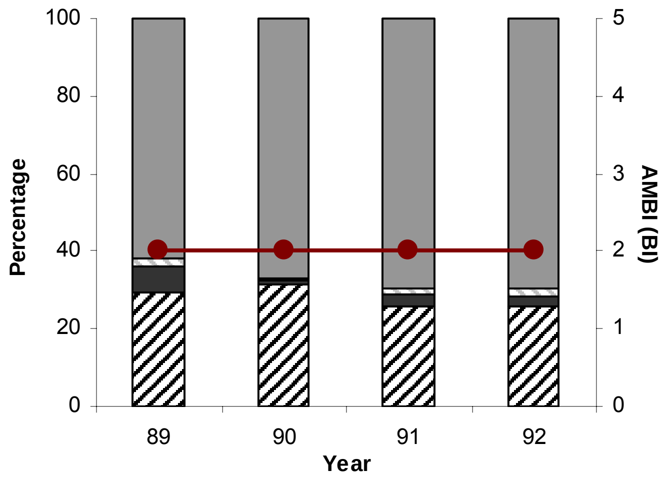


Figure 8:

Relationship between AMBI output (BC scores) and PC1.500 (a metric of heavy metal contamination; Anderson et al. 2006) across 95 RDP sites. The BC scores of the AMBI should have increased (rated sites as increasingly polluted) with increasing PC1.500 values. The low slope and high variability around the least squares regression fit (r2 = 0.038) shows that the AMBI did not track the heavy metal contamination gradient well.

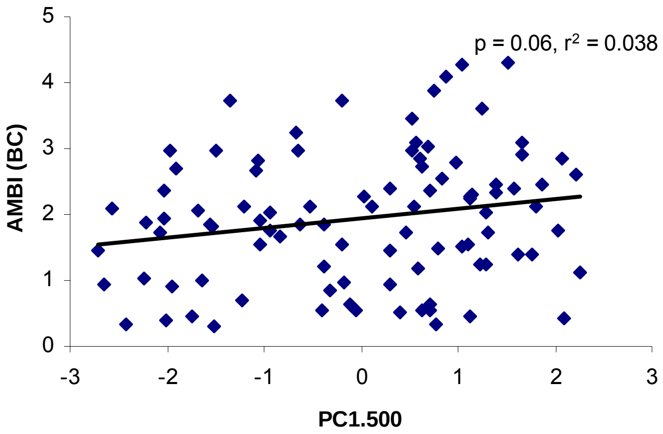


Figure 9:

Relationship between AMBI output (BC scores) and sediment muddiness (percent silt+clay) across 95 RDP sites. The BC scores of the AMBI should have increased (rated sites as increasingly polluted) with increasing muddiness. The low slope and high variability around the least squares regression fit (r2 = 0.118) shows that the AMBI did not track the muddiness gradient well.

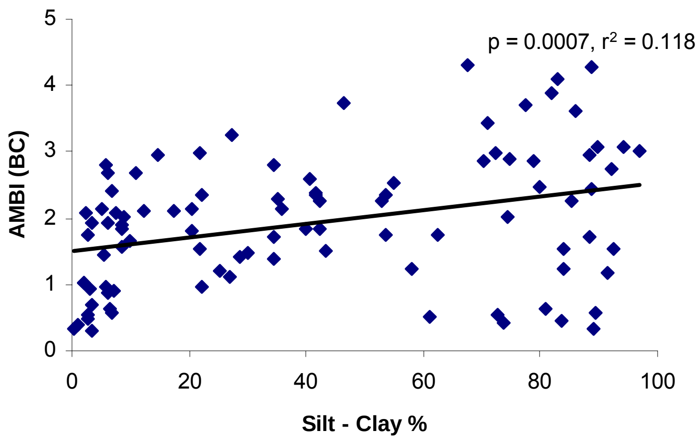


Figure 10:

Average heavy metal concentrations (± 1 standard error; Cu, Zn, Pb on particles <500 µm) in sediments from RDP sites that have the same AMBI BI score. If the AMBI was successfully tracking pollution by heavy metals, we would have expected significant differences in metal contamination at sites with BI scores of 1 (unpolluted), 2 and 3 (slightly or moderately polluted).

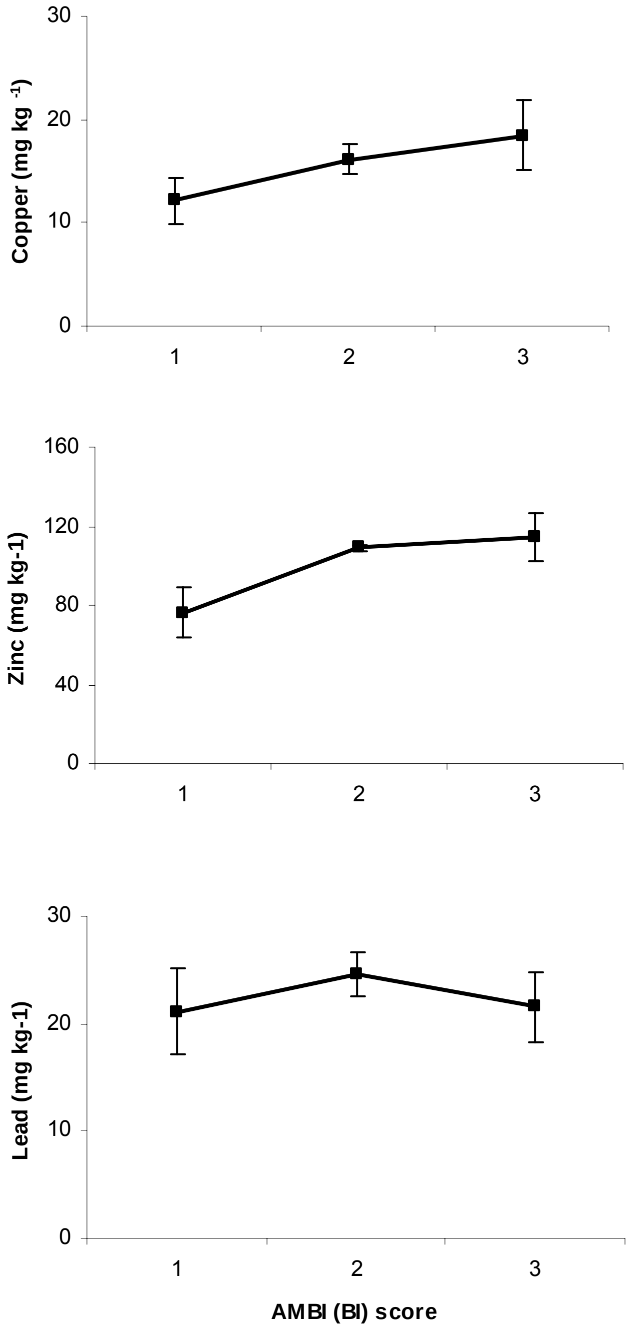


Figure 11:

Relationship between AMBI output (BC scores) and individual metal species (Cu, Zn, Pb on particles <500 µm) across 95 RDP sites. The scattering of points and the low r2 values suggest that AMBI scores do not reflect sediment metal contamination levels very well.

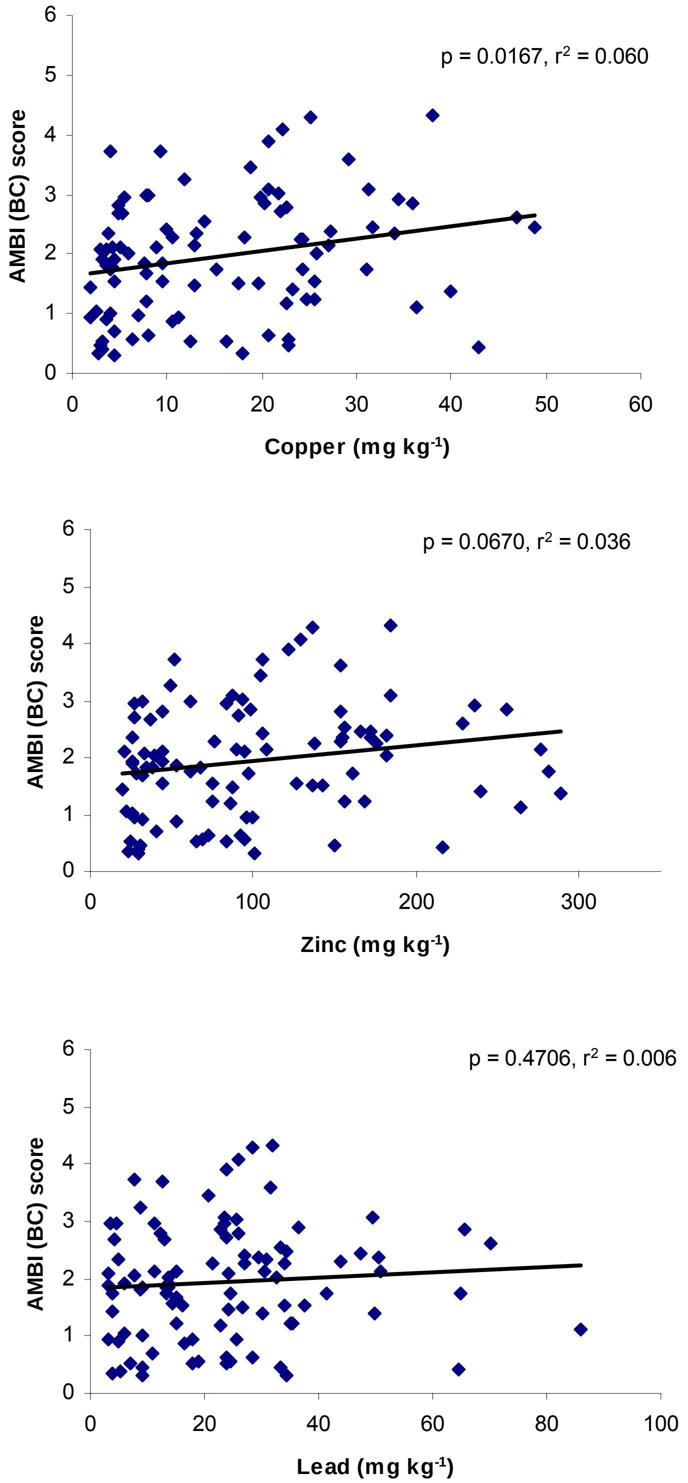
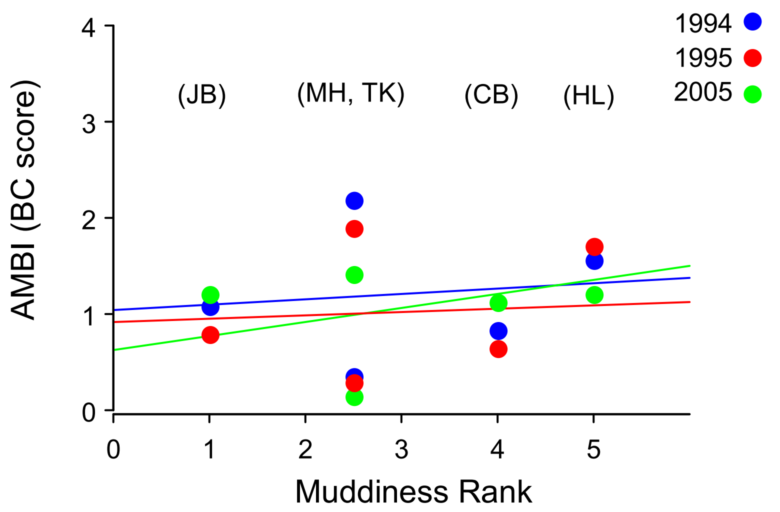


Figure 12:

AMBI (BC) scores at Mahurangi sites based on October data at JB, MH, TK, CB and HL from 1994, 1995 and 2005. We expected AMBI scores to increase with increasing muddiness (a type of pollutant), but they did not correlate with muddiness rank (top panel: p values from 0.3 to 0.8, r2 values from 0.01 to 0.10). Between 1994/1995 and 2005, a time period in which muddiness in the Harbour may have increased, there were no consistent changes in AMBI scores (bottom panel: 3 sites went down, 2 sites went up).



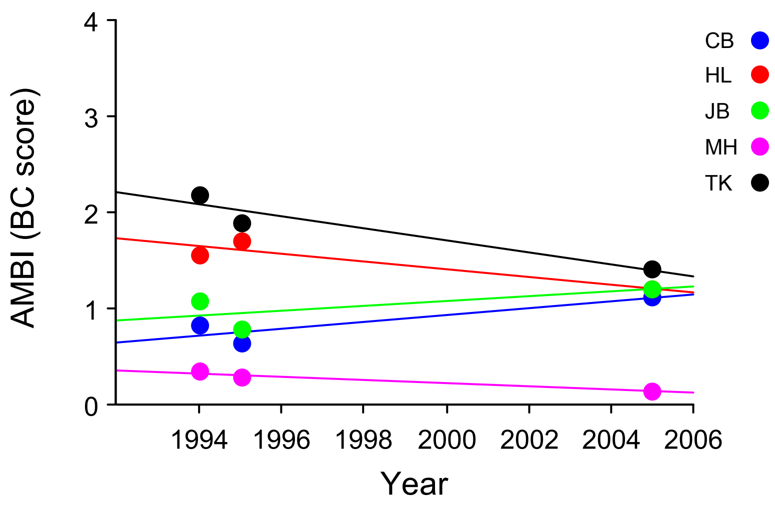


Figure 13:

Average (± 1 standard error) abundance (top), % sensitive species (middle) and Shannon-Wiener H' diversity (bottom) at RDP sites having the same AMBI BI scores. The Pearson & Rosenberg paradigm predicts that all measures would decline with declining habitat quality (indicated by a shift in BI scores of 1 to 3).

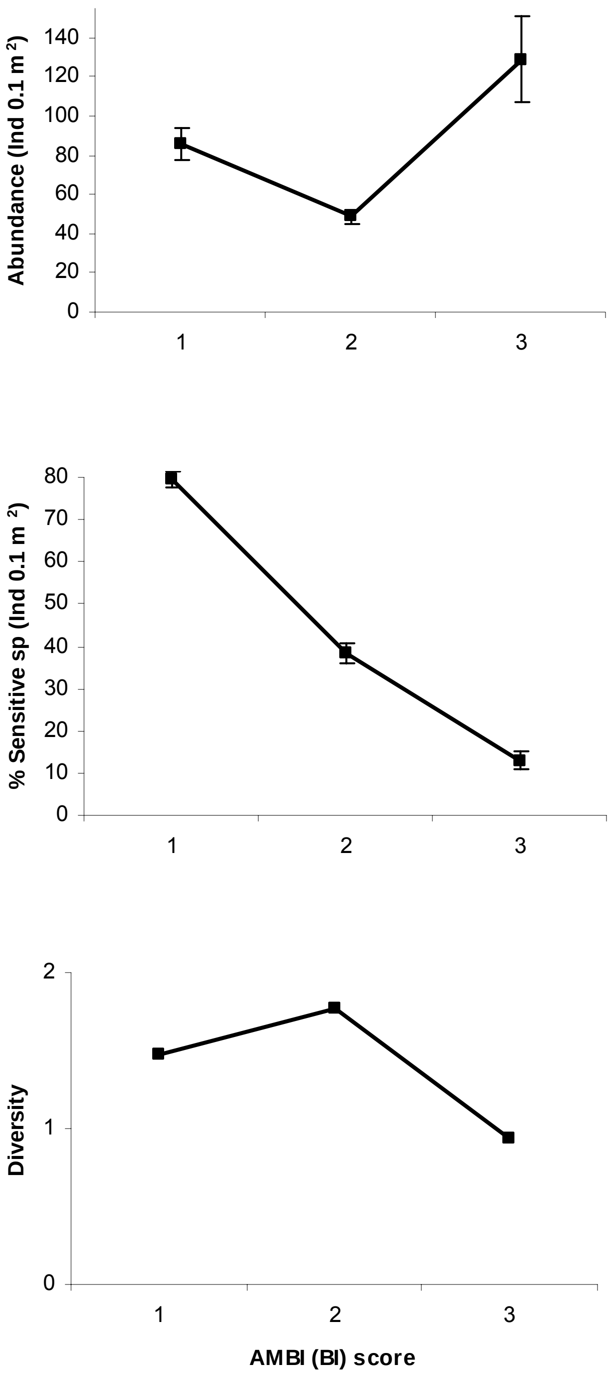
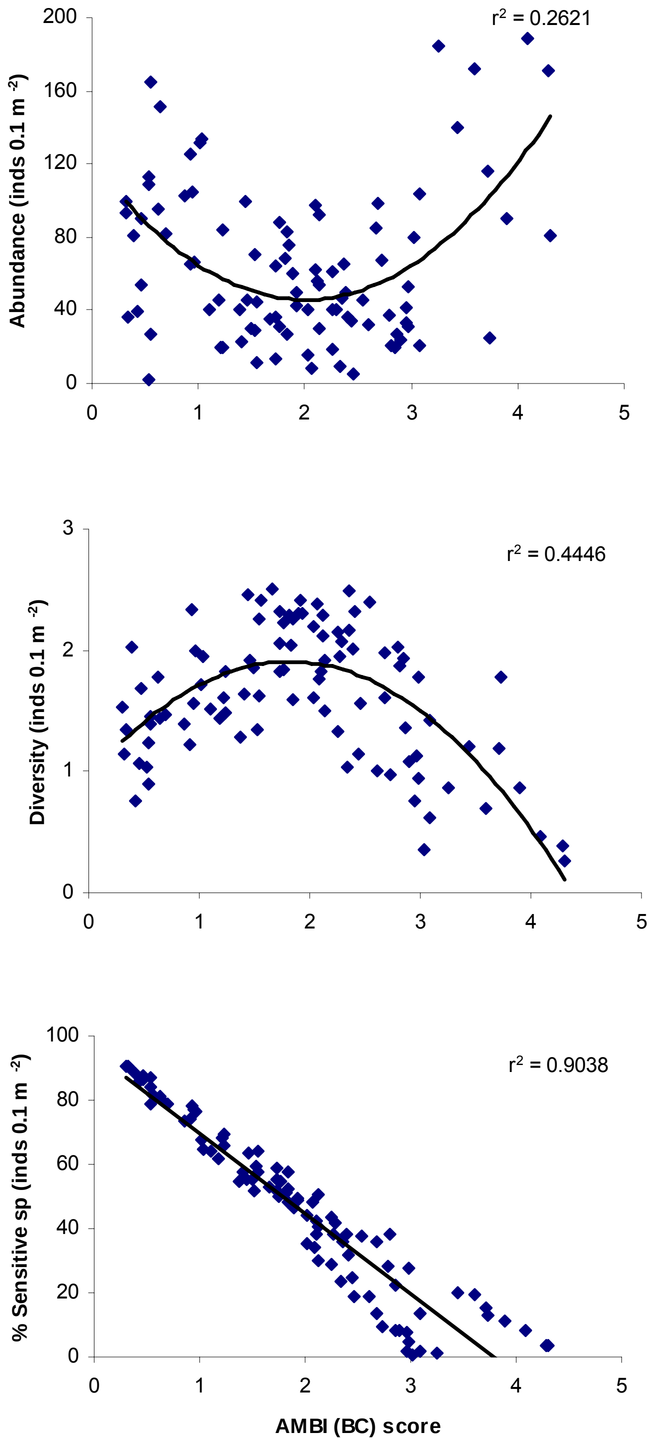


Figure 14:

Unlike BI scores of AMBI (which are categories), the BC scores of AMBI are continuous. BC scores were plotted against abundance (top), Shannon-Wiener *H′* diversity (middle) and % sensitive species (bottom). The BC scores were most strongly related to % sensitive species.



4.2 Benthic Index of Biotic Integrity (B-IBI)

The B-IBI evaluates the ecological condition of a sample by comparing values of benthic community attributes to reference values expected under non-degraded conditions in similar habitat types. A total of forty-three sites met the criteria as reference sites from the RDP dataset. Thirty-seven of these came from sand habitats and six from mud habitats. The number of reference sites for mud habitats was low because concentrations of heavy metals tend to be highest at the muddiest sites (Fig. 15). However, we were able to proceed with B-IBI calculations.

Five metrics were used to calculate the B-IBI scores (Table 3). Table 4 lists the selected metrics and their thresholds. Note that sites with average index values of 3.0 or more are considered to have good benthic condition indicative of good habitat quality. This is opposite to AMBI where low BI and BC scores indicate good quality habitat. The B-IBI index classified 58% of the RDP sites correctly, relative to contaminant levels indicated by PC1.500 values (<0). Manukau and Mahurangi sites all classified as unpolluted or slightly polluted (B-IBI >3).

The correlation between sediment contaminant levels at the RDP sites (PC1.500) and the B-IBI scores is shown in Figure 16. The linear least squares fit explained very little variation (r^2 = 0.0211). The correlation between B-IBI scores and the RDP mud gradient was somewhat stronger (Fig. 17), but the percent variability explained remained relatively low (r^2 = 0.1046). Nevertheless, the change in B-IBI scores with increased muddiness did seem to indicate declining numbers of sensitive species.

Relationships between B-IBI scores and individual metals (Cu, Zn, Pb 500), as opposed to the combined PC1.500 metric, are shown in Figures 18 and 19. There were no strong relationships observed.

We also determined whether or not aspects of the biotic assemblage were correlated with B-IBI scores (Figs. 20 & 21). Average species abundance, Shannon-Wiener *H′* diversity and % sensitive species for all RDP data were compared across sites with the same B-IBI scores. The RDP data was again used because this dataset provided the best range of B-IBI scores (1-5) for evaluation. Average species abundance decreased and % sensitive species and species diversity increased with an increasing B-IBI score (>3) (Fig 20).

Relationships between B-IBI scores versus species diversity and the abundance of sensitive species were moderate (r^2 = 0.5385 and r^2 = 0.3369 respectively). The highest degree of dispersion over the entire range of variables was for species abundance (r^2 = 0.0918) (Fig. 21). Conversely, B-IBI scores explained just 5% and 4% of the variability in sensitive species abundance at Mahurangi and the Manukau.

Like the AMBI, the correlation between B-IBI scores and muddiness at the Mahurangi sites was weak (r^2 = 0.093). B-IBI scores decreased with increased muddiness across the sites in 1995 and 2005, but the scores increased along the mud gradient in 1994 (Fig. 22, top). At TK, MH and CB, there were trends of increased B-IBI scores (suggesting improved conditions) between the years of 1994/1995 and 2005 (Fig. 22, bottom).

Figure 15:

Relationship between sediment heavy metal contaminants (PC1.500 values from Anderson et al. 2006) and sediment muddiness (percent silt+clay) at the 95 RDP sites. The muddiest sites tended to be the most contaminated with heavy metals, which is not surprising as metals are known to bind to silt and clay particles.

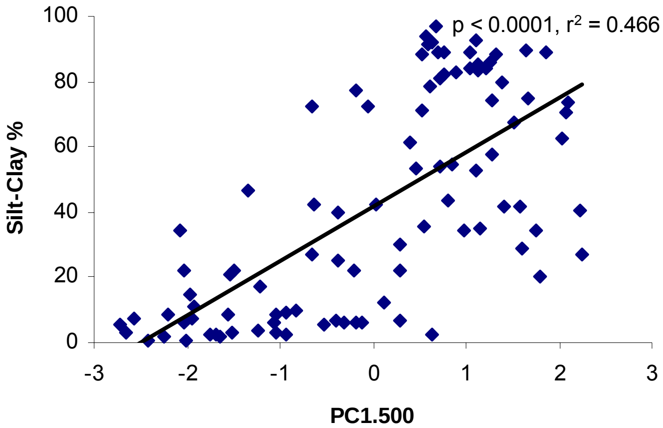


Table 3:

Mean values, divided by habitat (sand/mud) and estuary, for the different components comprising the B-IBI. Of the RDP sites, reference sites are denoted REF, whilst contaminated sites are denoted CONT. The REF and CONT division was based on Benthic Health Model output (Anderson et al. 2006).

Estuary	RDP				Mahurangi		Manukau
Habitat	Sand REF	Mud REF	Sand CONT	Mud CONT	Sand	Mud	Sand
Abundance	73.11	49.88	55.02	60.83	54.87	49.06	78.83
Species Diversity (*H'*)	1.87	1.659	1.74	1.35	1.82	1.76	2.20
Pollution sens. sp (%)	55.08	39.07	51.82	37.953	73.48	59.73	70.89
Carni/omnivores (%)	13.31	21.92	33.56	34.223	8.73	7.26	14.28
Deep dep. feeders (%)	56.88	55.58	47.57	42.73	73.16	62.69	58.25

Table 4:

Thresholds used to score components of the B-IBI in sand versus mud habitats.

	Scoring criteria		
Sand Habitat	5	3	1
Average abundance	>41-94	14-41 or >94-154	<14 or >154
Species Diversity (*H'*)	>2	0.9-2.0	<0.9
% Abundance of pollution sensitive spp.	>55	4-55	<4
% Abundance of carnivores/omnivores	>12	3-12	<3
% Abundance of deep deposit feeders	>64	14-64	<14
Mud Habitat			
Average abundance	>26-70	8-26 or >70-106	<8 or >106
Species diversity	>1.8	1-1.8	<1
% Abundance of pollution sensitive sp	>38	13-38	<13
% Abundance of carnivores/omnivores	>13	8-13	<8
% Abundance of deep deposit feeders	>68	11-68	<11

Figure 16:

Relationship between B-IBI output and PC1.500 (a metric of heavy metal contamination; Anderson et al. 2006) across 95 RDP sites. The B-IBI scores should have decreased (rated sites as increasingly polluted) with increasing PC1.500 values. The low slope and high variability around the least squares regression fit (r2 = 0.0394) shows that the B-IBI did not track the heavy metal contamination gradient well.

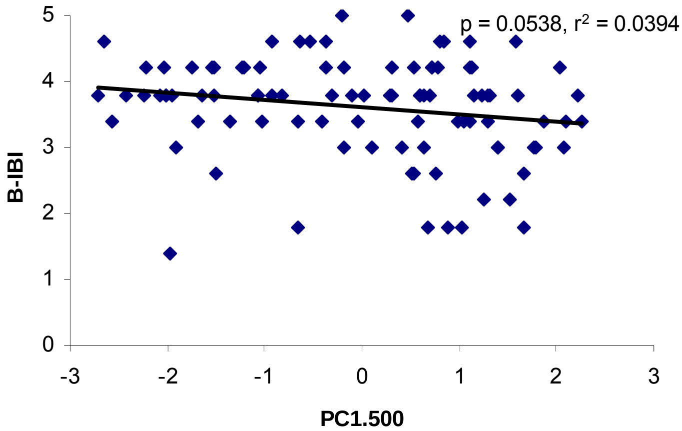


Figure 17:

Relationship between B-IBI output and sediment muddiness (percent silt + clay) across 95 RDP sites. The B-IBI scores should have decreased (rated sites as increasingly polluted) with increasing muddiness. The low slope and high variability around the least squares regression fit (r2 = 0.1046) shows that the B-IBI did not track the mud gradient well.

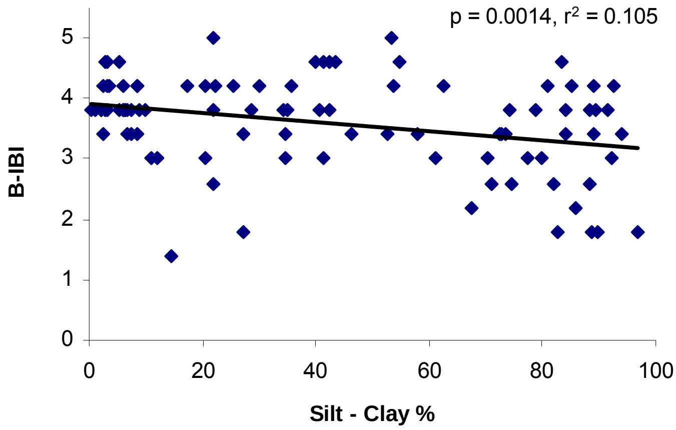


Figure 18:

Average heavy metal concentrations (± 1 standard error; Cu, Zn, Pb on particles <500 μm) in sediments from RDP sites relative to B-IBI category. If the B-IBI was successfully tracking pollution by heavy metals, we would have expected significant differences in metal contamination across the four B-IBI categories (higher scores should have had lower metal concentrations).

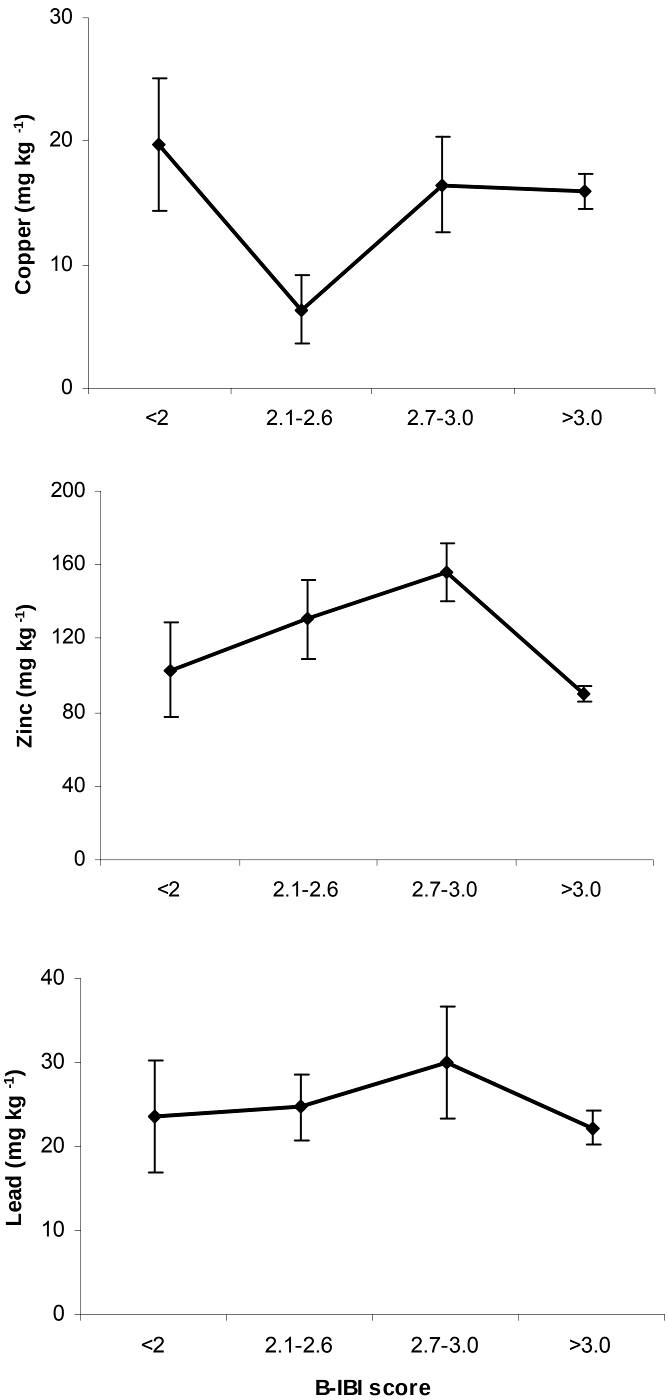
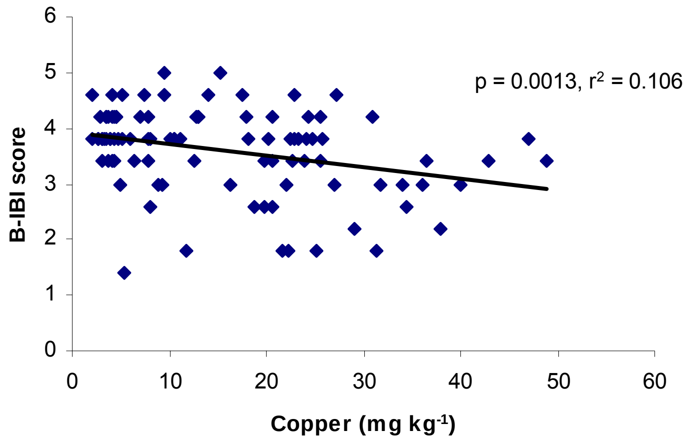
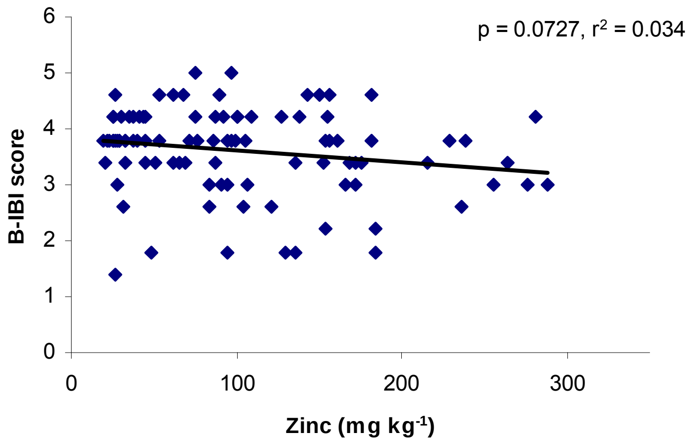


Figure 19:

Relationship between B-IBI output and individual metal species (Cu, Zn, Pb on particles <500 µm) across the 95 RDP sites. The scattering of points and the low r2 values suggest that B-IBI scores do not reflect sediment metal contamination levels very well.





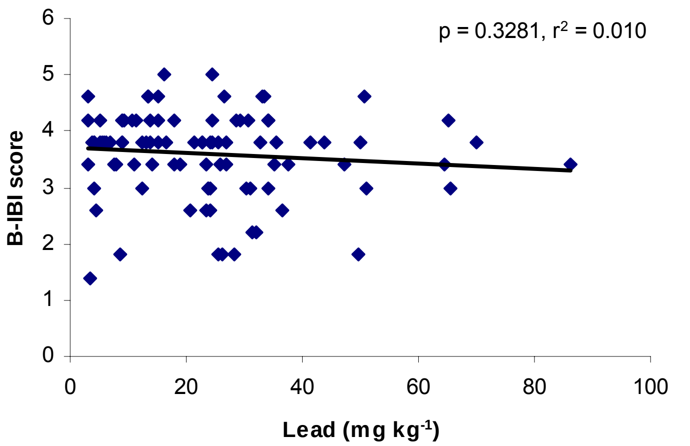


Figure 20:

Average (± 1 standard error) abundance (top), % sensitive species (middle) and Shannon-Wiener H' diversity (bottom) at RDP sites divided by B-IBI category. The Pearson & Rosenberg paradigm predicts that all measures would increase with increasing habitat quality (indicated by increased B-IBI scores).

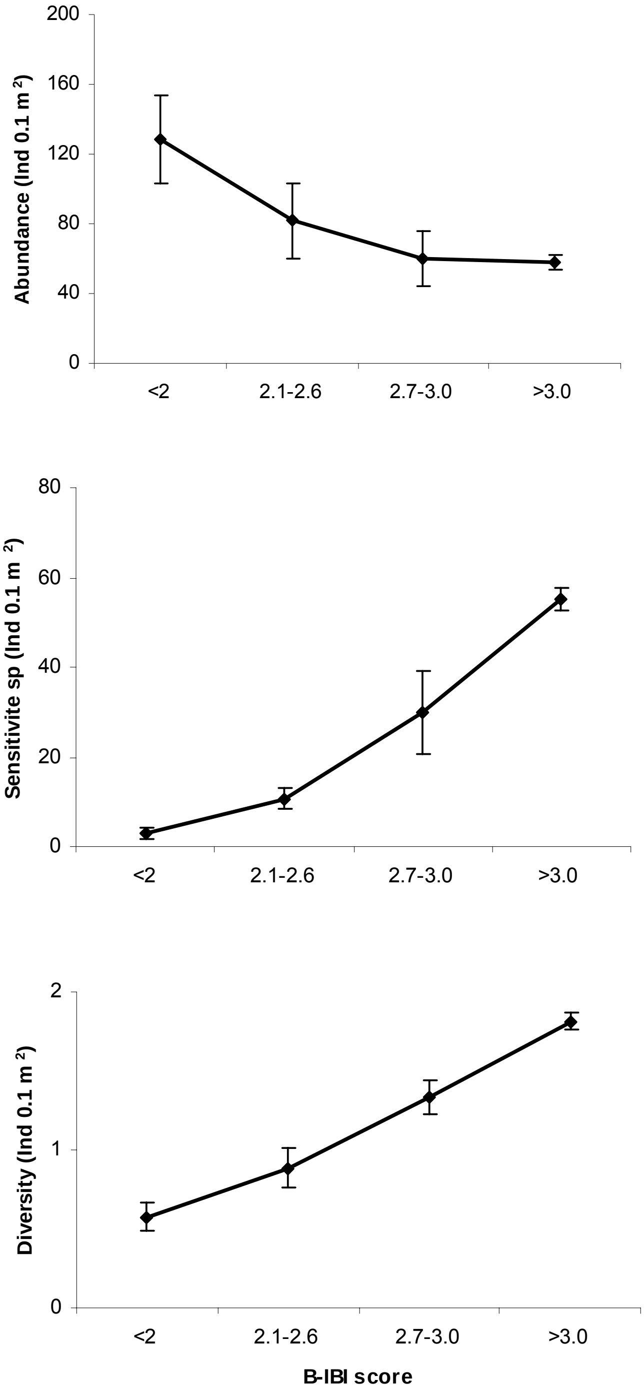


Figure 21:

B-IBI scores were plotted against abundance (top), % sensitive species (middle) and Shannon-Wiener H' diversity (bottom). The B-IBI scores were most strongly correlated with H' diversity (r2 = 0.538), but there was a reasonably good correlation with % sensitive species also.

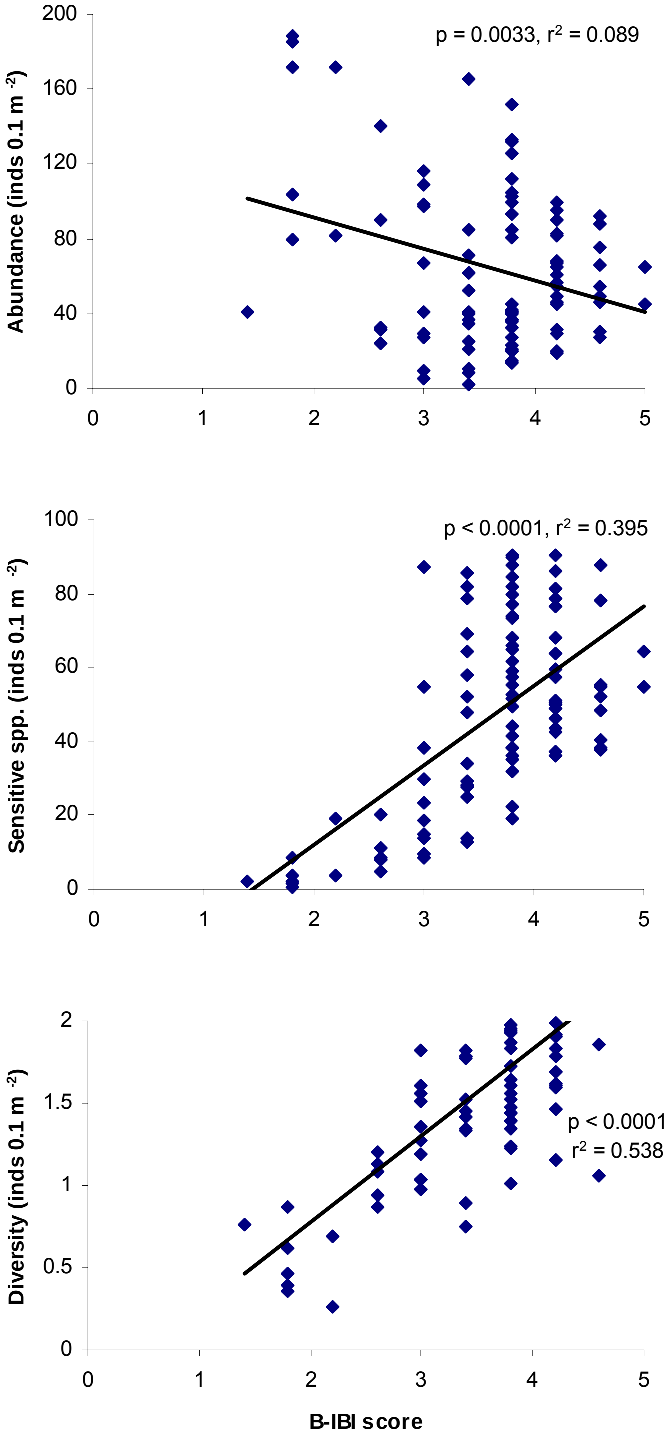
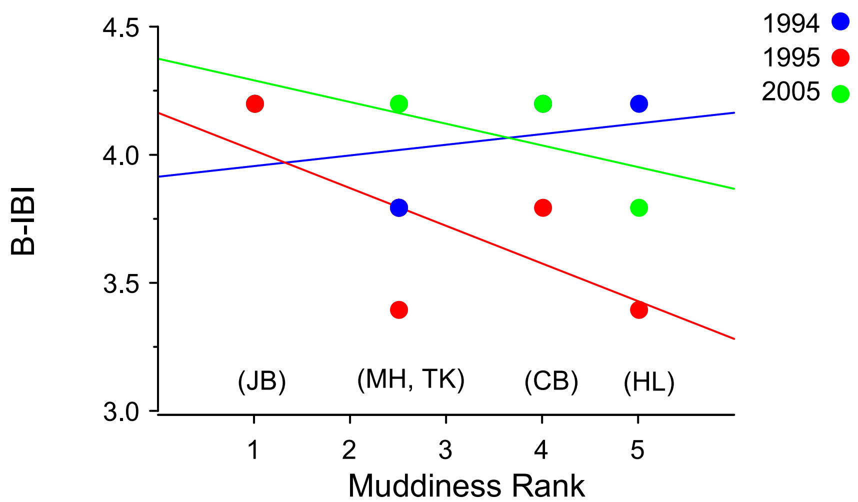
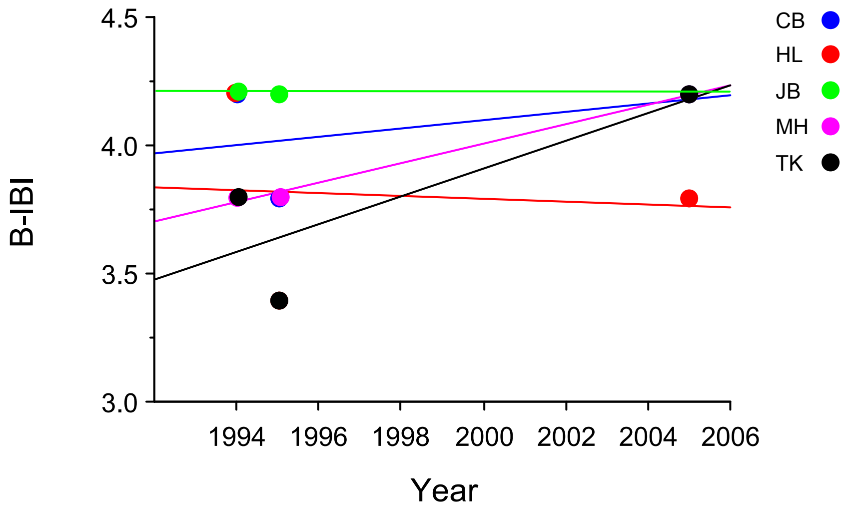


Figure 22:

B-IBI scores at Mahurangi sites based on October data at JB, MH, TK, CB and HL from 1994, 1995 and 2005. We expected B-IBI scores to decrease with increasing muddiness (a type of pollutant), but they did not correlate well with muddiness rank overall (top panel: p values from 0.17 to 0.63, r2 values from 0.08 to 0.53). Between 1994/1995 and 2005, a time period in which muddiness in the Harbour may have increased, there were no consistent changes in B-IBI scores (bottom panel: 3 sites went up, 2 sites stayed essentially the same).





4.3 Comparison between AMBI, B-IBI and PC1.500 values

Based on the AMBI and B-IBI classifications, RDP sites exhibited various levels of disturbance, ranging from severely degraded to meeting restoration goals. When comparing both methods in terms of degraded versus non-degraded status, 26 sites were classified as meeting restoration goals (Table 5). In comparison, the Benthic Health Model classified only 11 sites as non-contaminated (Group 1), corresponding to a PC1.500 value < -1.9[1]. This difference is because the new but significantly higher threshold level of PC1.500 < 0 that we used to identify degraded and non-degraded sites for the B-IBI index included 43 sites that came mainly from BHM Groups 1 and 2 with a few from Group 3.

Variation between the AMBI and B-IBI classifications increased among sites that were classed as degraded, severely degraded or marginal in either index. For example, 60 sites graded by the AMBI index as marginal were graded by the B-IBI index as severely degraded (4), degraded (3), marginal (8) and meets goals (45). Linear regression of AMBI and B-IBI indicated poor concordance of these indices (accounting for 32% of the variability in the dataset), with a higher degree of dispersion for sites in the high quality low AMBI and high B-IBI range (Fig. 23).

Table 5:

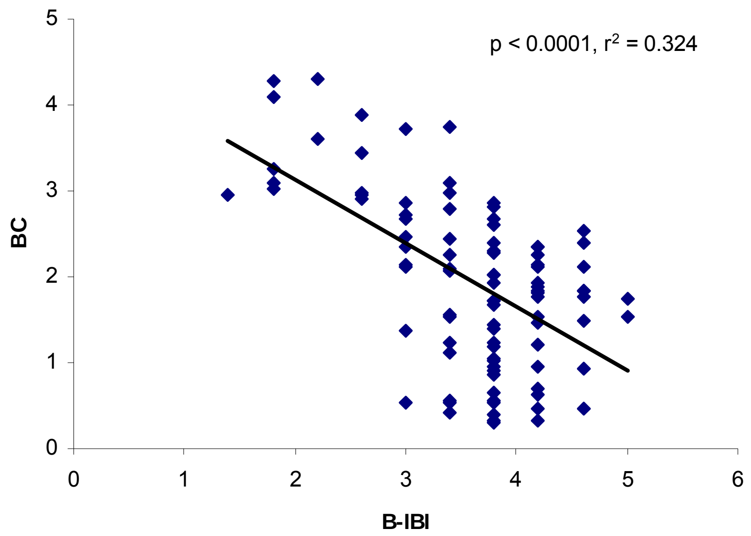
Number of RDP sites classified as Severely Degraded, Degraded, Marginal, and Meets Goals by the two overseas indices tested during this investigation, B-IBI and the AMBI. The range of values for each classification category is given (B-IBI in first two columns, AMBI in first two rows. Data cells give the number of RDP sites classified by B-IBI, followed by the number classified by AMBI (in parentheses).

		(AMBI)			
B-IBI		Severely Degraded	Degraded	Marginal	Meets Goals
	Range	5.0 - 6.0	3.3- ≤6.0	1.2- ≤3.3	0.0- ≤1.2
Severely Degraded	<2	6 (0)	2 (8)	4 (60)	0 (27)
Degraded	2.1-2.6	7 (0)	4 (8)	3 (60)	0 (27)
Marginal	2.7-3.0	10 (0)	1 (8)	8 (60)	1 (27)
Meets Goals	>3.0	72 (0)	1 (8)	45 (60)	26 (27)

[1] Criteria for "meeting restoration goals" have not been defined in New Zealand. Instead, the BHM and related ranking systems (Environmental Response Criteria, TEL, ISQG) tend to focus on "healthy" vs "polluted". However, it could be suggested that BHM categories of 1 and 2 are more likely to be meeting restoration goals.

Figure 23:

Relationship between index scores for the two overseas indices tested during this investigation, AMBI and B-IBI. AMBI (BC) scores are plotted on the y-axis, while B-IBI is plotted on the x-axis. The two indices tracked each other reasonably well (r2 = 0.3237).



4.4 Functional Index (NIWACOOBII)

An analysis of the abundance of individuals and taxa in 29 distinct functional groupings was performed on two data sets encompassing stress gradients: Mahurangi (muddiness) and RDP (metal contamination).The analysis of the 29 NIWACOOBII functional groupings revealed consistent responses to increasing stress levels for both mud and metals. Generally, the number of taxa present per group, $Ntaxa_{group}$, decreased with increasing stress levels (27 negative correlations with mud, 27 negative correlations with metals, no correlations were significantly positive) (Table 6 and 7).

Correlations between $Ninds_{group}$ and stress levels were more variable, with both significant positive and significant negative correlations recorded. Sixteen of the 29 correlations with PC1.500 were negative (55%), although only 8 significant negative correlations and 5 significant positive correlations were identified. Twelve of the 29 groups were negatively correlated with increased mud content at the Mahurangi sites (3 significant negative correlations, 0 significant positive correlations). Although there were fewer significant correlations with mud than for metals due to the smaller number of data points analysed (i.e., 15 for Mahurangi, cf 95 for RDP), the correlation coefficients for the mud gradient at Mahurangi tended to be stronger (Table 6 and 7).

Table 6:

Correlations between two environmental stressors (mud and metals) and the number of individuals and taxa (Ninds_{group} and Ntaxa_{group}, respectively) in 29 functional trait groups. Refer to Table 2 for descriptions of the 29 functional trait groups listed in the first column. The numbers in the table are Pearson's correlation coefficients; * denotes significance at $p<0.05$, whereas ~* indicates $0.05<p<0.10$.

	Correlation with MUD (Mahurangi)				Correlation with METALS (RDP sites)			
Group	Ninds_{group}		Ntaxa_{group}		Ninds_{group}		Ntaxa_{group}	
Calcium	0.32		-0.66	*	-0.44	*	-0.44	*
Globular	-0.45	~*	-0.53	*	0.10		-0.37	*
Worm	0.02		-0.70	*	0.15		-0.45	*
Large	0.19		-0.52	*	0.19	~*	-0.27	*
Medium	-0.54	*	-0.69	*	-0.18	~*	-0.43	*
Small	0.21		-0.74	*	-0.11		-0.48	*
Free	0.26		-0.65	*	-0.10		-0.44	*
Limited	0.28		-0.45	~*	-0.06		-0.32	*
Sedentar	-0.59	*	-0.81	*	-0.29	*	-0.59	*
Spel	.		.		0.01		-0.27	*
DD	0.01		-0.67	*	0.13		-0.35	*
DS	-0.01		-0.52	*	0.21	*	-0.23	*
SD	0.02		-0.40		0.11		-0.19	~*
SS	0.35		-0.66	*	-0.37	*	-0.50	*
Dep	0.27		-0.64	*	-0.15		-0.48	*
Grazer	-0.06		-0.45	~*	-0.03		-0.26	*
Pred	-0.40		-0.62	*	0.38	*	-0.37	*
Scav	-0.46	~*	-0.66	*	0.41	*	-0.37	*
Sus	-0.38		-0.78	*	-0.23	*	-0.43	*
Attached	-0.18		-0.17		-0.28	*	-0.47	*
Deep	-0.39		-0.67	*	0.15		-0.35	*
Epif	0.22		-0.69	*	-0.38	*	-0.42	*
Top	0.06		-0.72	*	-0.13		-0.48	*
Burr	0.31		-0.02		0.21	*	0.00	
Erect	-0.59	*	-0.90	*	-0.07		-0.40	*
Hole	0.01		-0.50	~*	-0.31	*	-0.35	*
Mound	0.51	~*	0.26		0.12		0.16	
Trample	-0.40		-0.43		0.37	*	0.01	
Trough	0.22		-0.17		-0.54	*	-0.51	*

Table 7:

Summary of correlation test results corresponding to Table 6.

Correlations with MUD (Mahurangi)		Correlations with METALS (RDP)	
$Ninds_{group}$	$Ntaxa_{group}$	$Ninds_{group}$	$Ntaxa_{group}$
Negative Correlations, Total n=12 (43%)	Negative Correlations, Total n=27 (96%)	Negative Correlations, Total n=16 (55%)	Negative Correlations, Total n=27 (93%)
Negative Correlations, $r < -0.23$ n=9 (32%)	Negative Correlations, $r < -0.40$ n=24 (85%)	Negative Correlations, $r < -0.23$ n=8 (28%)	Negative Correlations, $r < -0.40$ n=13 (44%)
Negative Correlations, $p < 0.05$ n=3 (10%)	Negative Correlations, $p < 0.05$ n=19 (68%)	Negative Correlations, $p < 0.05$ n=8 (28%)	Negative Correlations, $p < 0.05$ n=25 (86%)
Positive Correlations, $p < 0.05$ n= 0 (0%)	Positive Correlations, $p < 0.05$ n=0 (0%)	Positive Correlations, $p < 0.05$ n=5 (17%)	Positive Correlations, $p < 0.05$ n=0 (0%)

Based on the results of the initial analysis, a trial index using $Ntaxa_{group}$ was developed. Seven of the original 29 functional groups were retained for use in the index, with one grouping selected from each functional category (see Table 2): The seven selected groups were "Worm" (worm-shaped organisms with length much greater than width), "Medium" (organisms of intermediate body size, not large or small), "Sedentary" (organisms that do not move, or only do so within a fixed tube), "Surface-to-Surface" (organisms whose activities move sediment particles laterally across the sediment surface, as opposed to up or down) "Suspension feeders" (organisms that feed by filtering suspended particles from seawater), "Top 2 cm" (organisms that occupy the upper 2 cm of the sediment column), and "Erect" (organisms that create erect topographic features, such as tubes, that stick out of the sediment). These groups met the following criteria for inclusion:

- Consistent negative response to mud and metals ($r < 0$ for both).
- Statistically significant negative response to mud and metals ($p < 0.05$ for both).
- Strong negative response to mud and metals (average $r < -0.5$).
- Negative correlations between $Ninds_{group}$ and mud or metals (preferably both).

Index values were then calculated as follows:

1. The 7 selected $Ntaxa_{group}$ values per site were summed (i.e., $Ntaxa_{Worm}$ + $Ntaxa_{Medium}$ + $Ntaxa_{Sedentary}$ + $Ntaxa_{SS}$ + $Ntaxa_{Sus}$ + $Ntaxa_{Top}$ + $Ntaxa_{Erect}$) to produce a quantity called SUM_{actual}. These sums were calculated for all 95 RDP sites and for all 15 site/date combinations at Mahurangi.
2. A maximum expected value (i.e., a non-polluted reference value) for Manukau-Waitemata-Mahurangi intertidal sites was determined from the sums of maximum values observed across all 95 RDP and 15 Mahurangi samplings, e.g., $Ntaxa_{WormMAX}$ + $Ntaxa_{MediumMAX}$ + $Ntaxa_{SedentaryMAX}$ + $Ntaxa_{SSMAX}$ + $Ntaxa_{SusMAX}$ + $Ntaxa_{TopMAX}$ + $Ntaxa_{ErectMAX}$. The quantity was called SUM_{max}, and was constant.
3. A minimum possible value for Manukau-Waitemata-Mahurangi intertidal sites (i.e., a completely defaunated site) was set at 0.
4. The index formula was $1 - (SUM_{max} - SUM_{actual}) / SUM_{max}$, which essentially standardised the index values to fall between 0 and 1. Values near 0 would indicate highly degraded sites, and values near 1 would indicate the opposite.

The resultant index values at the sites in Mahurangi Harbour ranged between 0.30 and 0.91. Index values for RDP sites ranged between 0.14 and 0.94, with 11 values < 0.30. The lowest RDP index value came from Tauranga J (0.14), with Whau Wairau (0.19) second lowest. The highest index value of the RDP sites (0.94) came from Little Shoal Bay.

Index values correlated well with the muddiness ranks given to the sites in Mahurangi Harbour ($r^2 = 0.564$). The across-site mud gradient was detected with the NIWACOOBII in all three years (1994, 1995 and 2005; Fig. 24, top). NIWACOOBII values at MH, TK and JB increased slightly between 1994/1995 and 2005 (an indication of declining sedimentation stress over time), whilst the other two sites remained essentially unchanged (Fig. 24, bottom).

Index values also correlated reasonably well with sediment mud content (silt+clay) at the RDP sites (r^2 = 0.3313, Fig. 25). However, the index values were not as highly correlated with metals (PC1.500, r^2 = 0.2590, Fig. 26) as they were with the mud variables.

The index values did not fit neatly into the 5 categories described in the Benthic Health Model. Figure 27 shows the overlap in index values between adjacent BHM categories. The lowest index value corresponded to a RDP site with a moderately polluted classification rank in the Benthic Health Model (i.e., a "3" on the scale of 1 to 5). The second lowest index value, however, did come from a category 5 site (Whau Wairau).

Figure 24:

NIWACOOBII scores at Mahurangi sites based on October data at JB, MH, TK, CB and HL from 1994, 1995 and 2005. We expected NIWACOOBII scores to decrease with increasing muddiness (a type of pollutant), and this was indeed observed (top panel: p values from 0.03 to 0.14, r2 values from 0.56 to 0.85). Between 1994/1995 and 2005, a time period in which muddiness in the Harbour may have increased, there were no consistent changes in NIWACOOBI scores (bottom panel; 3 sites went up, 2 sites remained the same).

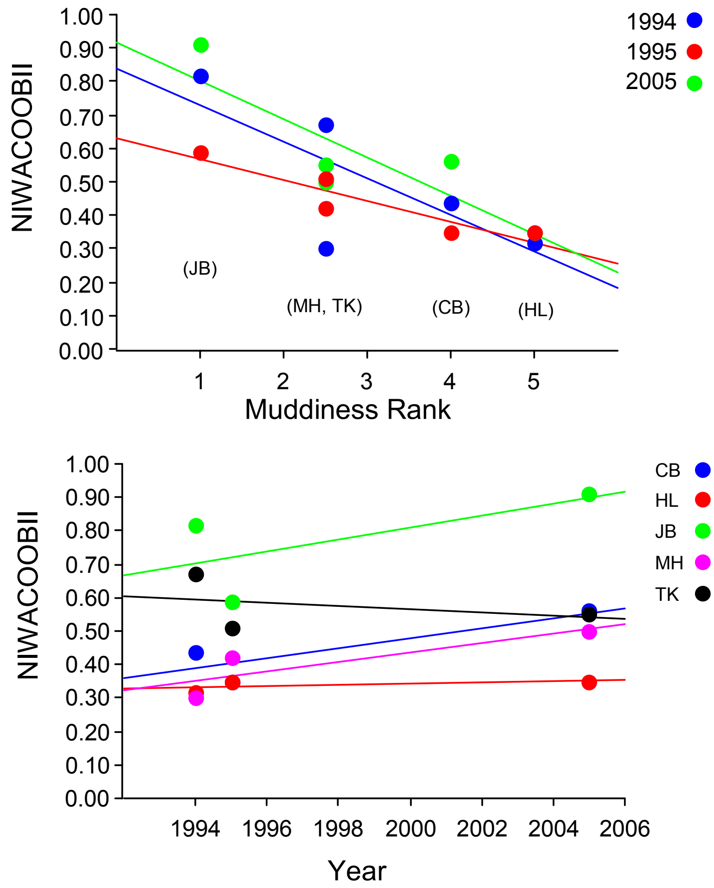


Figure 25:

Correlation between sediment mud content (x-axis, percent silt+clay) and NIWACOOBII values (low values indicating degraded sites with low functional richness). NIWACOOBII tracked the sediment mud gradient better than either AMBI or B-IBI (r2 = 0.32, cf r2 = 0.10).

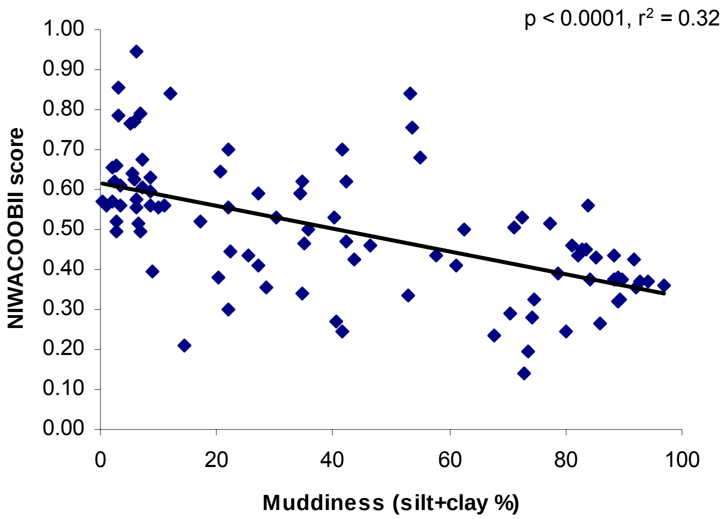


Figure 26:

Correlation between PC1.500 (x-axis, with high values indicating high sediment metal concentrations) and NIWACOOBII values (low values indicating degraded sites with low functional richness). NIWACOOBII tracked the metals gradient better than either AMBI or B-IBI (r2 = 0.25, cf r2 = 0.04).

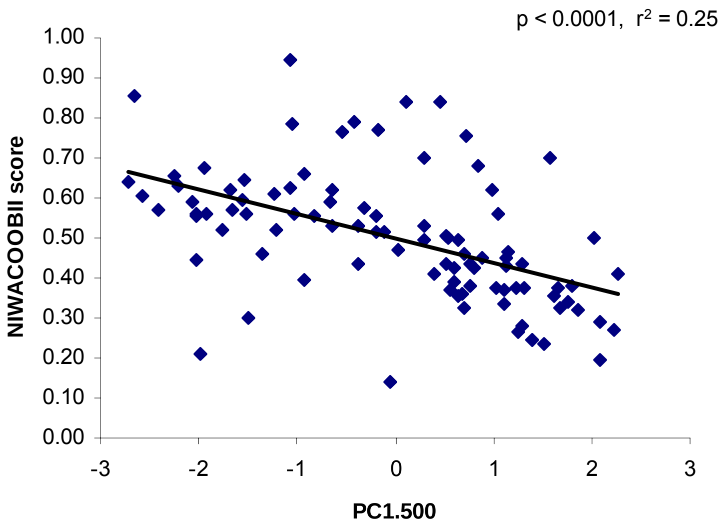
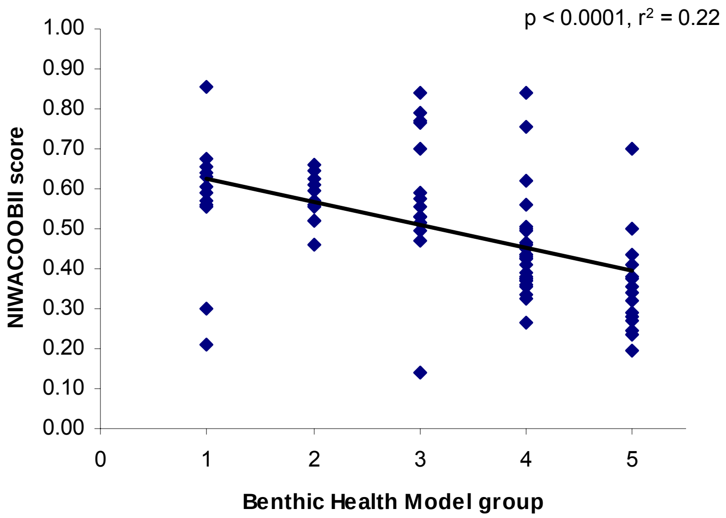


Figure 27:

Correlation between Benthic Health Model categories (with category 5 being the most degraded) and NIWACOOBII values (low values indicating degraded sites with low functional richness). There are outlying data points present in each BHM category, but the trend of declining health status indicated by the BHM is tracked by the NIWACOOBII. A correlation with actual CAP scores from the BHM is likely to be better (Hewitt & Ellis, Draft Report).



5 Discussion

The level of agreement in characterising benthic integrity or ecological status between the B-IBI, AMBI and measured estuarine contaminants (PC1.500) was moderate. A large proportion of RDP sites that we know are impacted by contaminants were classed as un-degraded by both indices. Increasing muddiness appeared to be the stressor most readily identified by the indices.

AMBI relies heavily on knowledge of species' sensitivity/tolerance to environmental stress gradients. The index is based on a well-known paradigm related to organic enrichment gradients, which suggests that benthic communities respond to improvements in habitat quality in three steps; abundances increase, species diversity increases, and species dominance shifts from pollution tolerant to pollution sensitive species (Pearson and Rosenberg 1978; Weisberg et al. 1997).

In our application of AMBI to New Zealand data sets, increased AMBI scores (indicative of degrading health) were associated with declines in the abundances of sensitive species and declines in species diversity. Proportionally there was an overall increase in abundance of a few tolerant species with increasing scores. This is consistent with the Pearson and Rosenberg (1978) paradigm, as decreases in sensitive species along with increases in the abundance of pollution tolerant / opportunistic species are expected with increasing stress. The progression of BI scores from un-impacted (BI = 1) to impacted (BI = 4) sites and the percentage of ecological groupings derived for some of the RDP sites (1-4) are shown in Figure 1. Although the AMBI index was able to detect some changes at these sites, scores indicating more severe levels of pollution (i.e., 5-7) were not seen at any of the New Zealand sites we examined. There are sites in New Zealand sites that are highly polluted, but these pollution hot spots represent a small proportion of our estuaries. Europe has a much larger human population densities and a longer history of pollution and broad-scale contaminant effects (with eutrophication, in particular, more common).

In European applications of AMBI, estuaries with a high BC level have been shown to have high mud and organic matter content, together with low redox potential (Borja et al. 2000). Ecotoxicological effects may play a multiple stressor role when combined with habitat change associated with mud additions and organic over enrichment (Thrush et al. 2008b). The sites in Manukau Harbour that we sample repeatedly as part of the ARC's State of Environment monitoring programme are sandy sites (low mud content) with generally low levels of storm water contamination (e.g., heavy metals). The sites were expected to have low AMBI scores and relative stability through time, and this was indeed the case at CB, EB, KP, PS and particularly AA (Fig. 3). Interestingly, Site CH was influenced by discharge from a sewage oxidation pond during the years we analysed, and AMBI scores at this site were higher and more variable at CH (Fig. 2). There was also a greater proportion of "Group III" species at this site, which perhaps illustrates how AMBI is most adept at detecting organic enrichment effects.

Although AMBI provides a good way of describing benthic macro fauna community composition, by assigning each species to an ecological grouping according to their sensitivity to a stress gradient, the index did not correlate well with the RDP metal contaminant gradient. Correlations with muddiness gradients at the RDP and Mahurangi sites were somewhat better but still weak or marginal.

The B-IBI was calculated using well known metrics of species abundance, diversity and the abundance of sensitive species, carnivores and deposit feeders. Like the AMBI, the B-IBI index was based on the paradigm of Pearson and Rosenberg (1978) and incorporated factors related to functionality. For example, it is assumed that the abundance and diversity of species living deep in the sediment will be highest at the cleanest reference sites and that the distribution of benthos among feeding guilds should be the most diverse at the cleanest reference sites (Weisberg et al. 1997).

The B-IBI values we calculated for New Zealand sites were correlated with gradients of increasing muddiness. However, like the AMBI, the B-IBI was unsuccessful at distinguishing reference sites from known degraded sites. Increasing numbers of sensitive species and diversity measures did correspond to increasing B-IBI scores, indicating the strong influence this metric had on the final B-IBI score. Better knowledge of the tolerance levels and sensitivities of species, based on life history or functionality characteristics, could improve use of the index. However, the way species respond to various stressors, like muddiness or contaminants, may be inconsistent, and this would reduce the general utility of the index to characterize sites being affected by multiple simultaneous stressors.

The problem of multiple stressors with respect to applications of the B-IBI was demonstrated by our difficulty in finding reference sites. There were plenty of clean sandy sites, but relatively few clean muddy sites. So, to apply the index in both muddy and sandy areas, the threshold values defining "clean" versus "contaminated" (with respect to heavy metals) had to be reduced quite substantially. Another constraint when using the B-IBI is that it includes several metrics that are not available in our data sets, including biomass and abundance in differing depth horizons. We substituted abundance data for biomass data. Weisberg et al. (1997) measured classification efficiency using abundance-for-biomass substitutions and found there was little reduction in efficiency. Finally, key elements of the B-IBI, including biodiversity indices such as number of taxa, Shannon-Weiner *H´* and species richness, had previously been demonstrated as not sensitive to contaminants in the Auckland Region during the development of the BHM (Anderson et al. 2002, Hewitt et al. 2005).

5.1 Comparability of AMBI and B-IBI and NIWACOOBII

Both AMBI and B-IBI used similar criteria (dissolved oxygen concentrations, sediment contaminants/toxicity, organic carbon content) and used similar and well known individual metrics (pollution sensitive taxa, abundance and diversity) to define degraded and un-degraded sites during their development, making them more likely to have a greater level of agreement (Borja et al. 2008). Studies using these indices have classified between 72-93% of sites correctly, indicating the need to combine any index with other measures of habitat quality, such as direct measures of sediment contamination and toxicity, to reduce misinterpretation of the data (van Dolah et al. 1999).

Results from our studies found BI scores ranging from 1-4, which is a much smaller range than that experienced in overseas estuaries (typical range 1-7) where periodic hypoxia either seasonally or for repeated brief periods (days or weeks) generates mass mortality or elimination of the benthic fauna (Borja et al. 2000). High BI scores (5-7) tended to correlate to increasing percentages of mud and organic matter, together with decreasing redox potential. Similarly, Borja et al. (2008) found that the principal stressor in Chesapeake Bay was low dissolved oxygen, explaining 42% of the variability in the B-IBI as well as organic enrichment. In comparison, estuaries in New Zealand are usually not subjected to extreme hypoxic events, and the index correlated best to increasing muddiness. Overseas indices may lack the sensitivity required to detect slight to moderate changes, thereby making them less useful/applicable in a New Zealand context.

Despite the current lack of consistency between the indices tested in this study, all of them provide ways of interpreting benthic data across habitats and they all seek to provide uniform scales for comparing the quality of benthic assemblages (Weisberg et al. 1997). This information can contribute to identifying areas most in need of management. That being said, the AMBI and B-IBI, like all of the others listed in Table 1, have a degree of subjectivity buried into their interpretation, and while the results may be simple to present, they are not any less complicated to truly understand or explain to a lay audience than the ordinations that underpin the Benthic Health Model.

The NIWACOOBII analysis found $Ntaxa_{group}$ to be a particularly useful variable for tracking increasing stress levels. We documented reductions in the number of species per group (negative correlations with increasing stress levels) for 93 and 96% of the functional groups tested for responses to mud and metals. The preponderance of negative trends makes intuitive sense, given that stressors/pollutants are thought to have a negative impact on biota. The consistency of trends for the two different stressor types suggests applicability to both highly urbanized areas (where heavy metal effects may dominate) and rural development areas (where increasing muddiness may be occurring). The use of reasonably broad functional groups also provides a degree of inter-region consistency, in that sub-regions with differing species lists are nevertheless all likely to have representatives in the same functional groups.

A reduction in $Ntaxa_{group}$ can be interpreted as a loss of functional redundancy. Habitats with high functional redundancy (i.e., many species present in each group) will tend to have higher inherent resistance and resilience in the face of environmental changes, as the higher numbers of species per functional group provide "insurance" for stochastic or stress-induced losses of particular species. Therefore, the NIWACOOBII analysis is not only reasonably sensitive and consistent across stressor types and regions; it is meaningful with regards to maintaining ecosystem multi-functionality.

The trial index that we created was far better than the AMBI or B-IBI in terms of significant correlation with mud and metal gradients in the Mahurangi and Waitemata-Manukau systems. However, the assessment of the NIWACOOBII was slightly biased, in that it was developed using Mahurangi and Waitemata-Manukau data; assessments on independent data sets are now required. Along with independent testing, further refinements and improvements are likely to made, which may result in a New Zealand specific index that is relevant to the key stressors in our systems.

In conclusion, we do not recommend adoption of an overseas index for ARC's State of Environment reporting at present, given that indices such as AMBI and B-IBI did not track the effects of mud and heavy metals in our estuaries particularly well. The allure of any index is in its ability to present complex information in a simple way. However, it appears that overseas indices would have to be drastically modified if they are to work well in New Zealand, which would introduce new problems in terms of simplicity and comparability to values from other nations.

With further testing using independently collected data, the NIWACOOBII may prove useful for understanding and communicating the limits of resilience in our coastal and estuarine ecosystems (Thrush et al. 2009). In particular, this index can be used in conjunction with the Benthic Health Model to interpret the significance of observed changes. The BHM is probably the more sensitive tool for documenting shifts in macrobenthic community structure along stress gradients. However, the NIWACOOBI provides resource managers with information in the tangible and meaningful currency of functional redundancy. An indication of functional redundancy in a community can be used to understand the resistance of communities to environmental change and their recovery potentials. Thus, together with the Benthic Health Model, the NIWACOOBII may provide the ARC with easily understandable information on the ecological integrity of invertebrate communities in estuarine and coastal areas which can be used for State of Environment reporting or more generally for communicating to the public about the health status of coastal habitats.

6 Acknowledgements

We thank the ARC for funding and for providing the data sets for index testing. We are grateful for the online resources provided freely at www.azti.es. NIWA's work on resilience and ecosystem functioning, which underpinned NIWACOOBII development, was funded by FRST (C01X0501). M. Carbines, J. Walker, and C. Hatton provided helpful comments on an earlier draft of this report.

7 References

Anderson, B.S.; Hunt, J.W.; Phillips, B.M.; Fairey, R.; Roberts, C.A.; Oakden, J.M.; Puckett, H.M.; Stephenson, M.; Tjeerdema, R.S.; Long, E.R.; Wilson, C.J.; Lyons, J.M. (1998). Chemistry, toxicity and benthic community conditions in selected sediments of the Los Angeles Region. Final Report. State Water Resources Control Board, Sacramento, CA, USA.

Anderson, B.S.; Hunt, J.W.; Phillips, B.M.; Fairey, R.; Roberts, C.A.; Oakden, J.M.; Puckett, H.M.; Stephenson, M.; Tjeerdema, R.S.; Long, E.R.; Wilson, C.J.; Lyons, J.M. (2001). Sediment quality in Los Angeles Harbor, USA: A Triad assessment. *Environmental Toxicology and Chemistry 20*: 359–370.

Anderson, M.J.; Hewitt, J.; Thrush, S. (2002). The development of criteria for assessing community health of intertidal flats. Prepared by NIWA (HAM2002-048) for Auckland Regional Council. *Auckland Regional Council Technical Publication 184.*

Anderson, M.J.; Hewitt, J.E.; Ford, R.B.; Thrush, S.F. (2006). Regional models of benthic ecosystem health: predicting pollution gradients from biological data. Prepared by Auckland UniServices Ltd for Auckland Regional Council. *Auckland Regional Council Technical Publication 317*.

Bald, J.; Borja, A. et al. (2005). Assessing reference conditions and physico-chemical status according to the European Water Framework Directive: A case-study from the Basque Country (Northern Spain). *Marine Pollution Bulletin 50(12)*: 1508–1522.

Borja, A.; Franco, J. et al. (2000). A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. *Marine Pollution Bulletin 40(12)*: 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. *Marine Pollution Bulletin 46(7)*: 835–845.

Borja, A.; Muxika, I. (2005). Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. *Marine Pollution Bulletin 50(7)*: 787–789.

Borja, A.; Dauer, D.M. et al. (2008.) Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices. *Ecological Indicators 8(4)*: 395–403.

Bremner, J.; Rogers, S.I.; Frid, C.L.J. (2006). Methods for describing ecological functioning of marine benthic assemblages using biological traits analysis (BTA). *Ecological Indicators 6(3)*: 609–622.

Clarke, K.R. (1990). Comparison of dominance curves. *Journal of Experimental Marine Biology and Ecology 138*: 143–157.

Cummings, V.J.; Funnell, G.F.; Schultz, D.; Thrush, S.F.; Berkenbusch, K.; Nicholls, P. (2001). Mahurangi estuary ecological monitoring programme - report on data collected from July 1994 to January 2001. Prepared by NIWA (ARC01207) for Auckland Regional Council. *Auckland Regional Council Technical Publication 175.*

de Juan, S.; Demestre, M.; Thrush, S.F. (2009). Defining ecological indicators of trawling disturbance when everywhere that can be fished is fished: A Mediterranean case study. *Marine Policy 33(3)*: 472–478.

Gibbs, M.; Hewitt, J. (May 2004). Effects of sedimentation on macrofaunal communities: a synthesis of research studies for ARC. Prepared by NIWA (HAM2004-060) for Auckland Regional Council. *Auckland Regional Council Technical Publication 264.*

Halliday, J.; Cummings, V.J. (2009). Mahurangi Estuary Ecological Monitoring Programme, Report on data collected from July 1994 to January 2009. *Auckland Regional Council Technical Publication No.*

Hewitt, J.; Anderson, M.; Hickey, C.; Kelly, S.; Thrush, S.F. (2009). Enhancing the Ecological Significance of Sediment Contamination Guidelines through Integration with Community Analysis. *Environmental Science & Technology 43(6)*: 2118–2123.

Hewitt, J.E.; Anderson, M.J.; Thrush, S.F. (2005). Assessing and monitoring ecological community health in marine systems. *Ecological Applications 15*: 942–953.

Hewitt, J.E.; Ellis, J. (2010-*DRAFT*). Assessment of the Benthic Health Model. *Auckland Regional Council Technical Publication 2010/034.*

Hewitt, J.E.; Thrush, S.F. (2009). Do species abundances' become more spatially variable with stress? *The Open Ecology Journal 2*: 37–46.

Hewitt, J.E.; Thrush, S.F.; Pridmore, R.D.; Cummings, V.J. (1994). Ecological monitoring programme for Manukau Harbour: Analysis and interpretation of data collected October 1987 – February 1993. Prepared by NIWA for Auckland Regional Council. *Auckland Regional Council Technical Publication 44.*

Llansó, R.; Scott, L.; Dauer, D.; Hyland, J.; Russell, D. (2002). An estuarine benthic index of biotic integrity for the Mid-Atlantic region of the United States. I. Classification of assemblages and habitat definition. *Estuaries and Coasts 25(6)*: 1219–1230.

Llansó, R.J.; Dauer, D.M. (2002). Methods for calculating the Chesapeake Bay benthic index of biotic integrity. www.baybenthos.versar.com

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. *Marine Pollution Bulletin 55(1-6)*: 16–29.

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

Rosenberg, R.; Blomquist, M.; Nilsson, H.C.; Cederwall, H.; Dimming, A. (2004). Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin 49(9-10)*: 728–739.

Simboura, N.; Zenetos, A. (2002). Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new biotic index. *Medit. Mar. Sci. 3(2)*: 77–111.

Thrush, S.F.; Coco, G.; Hewitt, J.E. (2008a). Complex positive connections between functional groups are revealed by neural network analysis of ecological time-series. *American Naturalist 171*: 669–677.

Thrush, S.F.; Hewitt, J.E.; Hickey, C.W.; Kelly, S. (2008b). Multiple stressor effects identified from species abundance distributions: Interactions between urban contaminants and species habitat relationships. *Journal of Experimental Marine Biology and Ecology 366*: 160–168.

Thrush, S.F.; Hewitt, J.E.; Dayton, P.K.; Coco, G.; Lohrer, A.M.; Norkko, A.; Norkko, J.; Chiantore, M. (2009). Forecasting the limits of resilience: integrating empirical research with theory. Proceedings of the Royal Society B-Biological Sciences 276: 3209–217.

Warwick, R.M. (1986). A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology 92*: 557–562.

Weisberg, S.; Ranasinghe, J. et al. (1997). An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries and Coasts 20(1)*: 149–158.