Ecological Indicators 85 (2018) 634–643

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Relationships between biotic indices, multiple stressors and natural variability in New Zealand estuaries

A. Berthelsen[a,*], J. Atalah[a], D. Clark[a], E. Goodwin[a], M. Patterson[b], J. Sinner[a]

[a] Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand
[b] School of People Environment and Planning, Massey University, Private Bag 11-222, Palmerston North, New Zealand

ARTICLE INFO

Keywords:
Biotic index
Macrofauna
Estuary
Ecological health
Estuary type
Natural variation

ABSTRACT

In response to the need to assess the ecological quality or health of marine benthic habitats, there has been a proliferation of biotic indices based on soft sediment macrofaunal communities. While shown to be useful in areas where they have been developed, some indices may not be readily transferrable to other regions due to differences in species ecology or composition, stressor type or magnitude, and natural variability. Using a national New Zealand dataset compiled from estuary monitoring data for 2001–2016, we used linear mixed models to determine the effect of multiple stressors (sediment mud content, metals and total phosphorus) and natural variability (associated with space, estuary type and time) on nine indices developed in New Zealand and overseas. The Richness-Integrated AZTI Marine Biotic Index (RI-AMBI), a modification of a popular overseas biotic index, had the most variation explained by stressors overall (marginal pseudo-R^2 = 0.22 compared to ≤0.15 for all other indices). This variation was primarily explained by a single stressor, sediment mud content, which is the dominant stressor in New Zealand estuaries. However, although the overall variation explained by stressors was lower for all other indices, multiple, rather than single, stressors had significant effects on some indices. For example all three stressors had a significant effect on the Traits Based Index, and the variation explained by metals was highest for this index. Relatively high amounts of natural and unexplained variation for all indices suggested that further understanding is required before operational implementation of indices at a national scale. Thus, the use of more than one index, i.e. a weight of evidence approach, is suggested to minimise uncertainty related to inaccuracy and misclassification of ecological health in New Zealand estuaries.

1. Introduction

Macrofaunal communities inhabiting marine soft sediments are often used as indicators of ecological quality or health (e.g. Borja et al., 2015). These globally common habitats are the receiving environment for many human impacts, and the sensitivity of benthic communities to human impacts has long been recognised (Pearson and Rosenberg, 1978). Many methods have been developed to assess ecological health based on analysis of multivariate community data. These include the development of biotic indices, which distil multivariate data to a univariate measure that aims to describe ecological health. Simple metrics (e.g. number of species and individuals) have been widely used. However, these have been out-performed as indicators of ecological health by biotic indices, such as those that reflect the sensitivities of different taxa to environmental gradients (Ellis et al., 2015; Simboura and Zenetos, 2002).

There has been a proliferation of biotic indices over the past 20 years and Borja and Dauer (2008) and Diaz et al. (2004) recommended that existing indices be considered before developing new ones. An important feature of an index is its capability to convey information that is meaningful for decision making, including being directly tied to management questions relating to human stressors, over a range of spatial and temporal scales (Cairns et al., 1993; Rees et al., 2008). Index responses can be tested against individual stressors (e.g. Simboura and Zenetos, 2002; Van Hoey et al., 2010). However, with increasing environmental pressures associated with both urban and rural intensification (e.g. nutrient run-off, sedimentation and metals contamination), it is preferable for indices to reflect the impacts of multiple stressors concurrently (Cairns et al., 1993; Van Hoey et al., 2010). Sensitivity to multiple stressors also increases the likelihood that an index will enable managers to identify changes in ecological health due to new and unanticipated perturbations (Cairns et al., 1993).

Most indices have been developed for northern hemisphere conditions, mostly in Europe and USA, and may not be readily transferable to other regions due to differences in species ecology or composition, stressor type or magnitude, and sources of natural variability (Gillett

* Corresponding author.
E-mail address: anna.berthelsen@cawthron.org.nz (A. Berthelsen).

https://doi.org/10.1016/j.ecolind.2017.10.060
Received 21 July 2017; Received in revised form 20 October 2017; Accepted 25 October 2017
1470-160X/

et al., 2015; Rodil et al., 2013; Van Hoey et al., 2010). Testing relationships between indices, stressors and natural variability is therefore crucial for assessing the suitability of indices in new regions (Borja et al., 2007; Diaz et al., 2004).

Some biotic indices developed overseas have been tested in New Zealand. For example, the AZTI Marine Biotic Index (AMBI, Borja et al., 2000) and the Benthic Index of Biotic Integrity (Weisberg et al., 1997) were assessed in estuaries of the Auckland region, with both performing poorly due to insensitivity to sediment and metal gradients (Rodil et al., 2013). For AMBI at least, this may be because the European ecogroups (EGs) to which taxa of differing sensitivities were assigned were based on response to enrichment (e.g. Borja et al., 2000), rather than on local stressor gradients (Rodil et al., 2013). Overseas studies have demonstrated that regionally-specific EGs can provide more accurate results due to taxon-specific responses to stressors (e.g. Gillett et al., 2015).

In New Zealand, Keeley et al. (2012a) developed regionally-specific EGs for subtidal macrofaunal taxa based on organic enrichment, and demonstrated that this increased the ability of EG-based indices to respond to organic enrichment gradients related to aquaculture farms (Keeley et al., 2012b). Local EGs for estuaries based on the response to mud have been developed by Robertson et al. (2015), as sediment mud content is a key estuarine stressor in New Zealand (Norkko et al., 2002; Robertson et al., 2016; Thrush et al., 2004). Robertson et al. (2016) subsequently developed a modified version of the AMBI (hereafter RI-AMBI) that incorporates proportional taxon richness in addition to proportional abundance of EGs used in the original AMBI. They found that using a combination of New Zealand specific and internationally defined EGs for index calculations improved the relationship with stressor gradients and furthermore that the RI-AMBI outperformed AMBI and another variation of this index, the Multivariate-AMBI (M-AMBI), which incorporates richness and diversity metrics.

As well as these EG-based approaches, an estuarine index was developed in New Zealand using a functional based approach. The Traits Based Index (TBI, Rodil et al., 2013) was developed to respond to sediment mud content and metal contamination gradients in the Auckland region, and is calculated using the richness of macrofaunal taxa in seven functional groups.

Despite the local refinement and development of biotic indices in New Zealand, there has been no comprehensive and consistent nationwide testing of the response of a range of indices (i.e. based on different approaches) to multiple stressors and natural variability. We have therefore undertaken such an assessment as a step toward selection of indices that provide accurate measures of ecological health in New Zealand estuaries and to test their comparability across wider scales. Our aim was to compile and then use a national dataset, collated from data collected using a standardised estuary monitoring protocol (Robertson et al., 2002), to assess the relationship between various biotic indices, developed both in New Zealand and overseas, multiple stressors, and natural variability (i.e. bioregion, estuary, estuary type and year).

2. Materials and methods

2.1. Macrofaunal and physico-chemical dataset

Data were obtained from intertidal estuarine surveys undertaken by New Zealand's regional government authorities during two seasonal periods (October − December and January − April) between 2001 and 2016. Surveys were conducted following a standardised estuarine monitoring protocol (Robertson et al., 2002) at unvegetated sites located at mid–low tidal height. Sites were positioned away from immediate point source discharges in order to capture overall cumulative stressor effects.

Macrofaunal samples were collected using a cylindrical core, 150 mm deep, and either 130 mm (82% of samples) or 150 mm (18% of samples) in diameter, and sieved through a 0.5 mm mesh. All individuals were identified to the lowest taxonomic level practicable by experts throughout the country. Taxonomic nomenclature followed the World Register of Marine Species (WoRMS Editorial Board, 2017). Where there were taxonomic uncertainties, we aggregated to higher groups. Taxa belonging to Plantae, Vertebrata, Bryozoa, Cirripedia, Insecta, Acari, and those identified to relatively coarse taxonomic groups (e.g. Gastropoda, Polychaeta, Annelida, Bivalvia, Decapoda and Brachyura), were removed from the dataset as recommended by Borja and Muxika (2005). Macrofaunal abundance was standardised to a 130 mm diameter core (i.e. results from samples taken with 150 mm cores were scaled down). However, initial exploratory analyses indicated no significant differences in species richness between the two core diameters ($p = 0.6$, results not shown). Thus, this potential source of bias was considered negligible and sampling events using cores of either diameter (130 mm core data and scaled-down 150 mm data combined) were included in order to maximise data availability for the development of subsequent models.

Physico-chemical sediment samples were collected at each site ($n = 1 – 12$) concurrently with the macrofaunal sampling. The large range in replicate numbers was due to compositing of samples prior to laboratory analyses in some surveys, resulting in a lower number of replicate samples than originally collected, as well as differences in sampling effort in some cases. Measured variables represented common stressors in New Zealand that are natural to some extent but are exacerbated by human-induced pressures affecting estuaries, e.g. sedimentation, eutrophication and contamination (Robertson et al., 2002, Robertson et al., 2015; Thrush et al., 2004; Edgar and Barrett, 2000). The stressors included in analyses were: mud (grain size < 63 μm), nutrients (total phosphorus), and the metals copper (Cu), zinc (Zn) and lead (Pb). These stressor variables were chosen based on data availability and quality, as well as ecological relevance.

Although estuaries both overseas and in New Zealand are often limited by total nitrogen (TN) rather than total phosphorus (TP) (Howarth and Marino, 2006; Robertson et al., 2002), poor data quality for TN led us to choose TP to represent nutrients in our statistical models. TP was moderately correlated with TN values above analytical detection limit (Pearson $r = 0.68$), and more strongly correlated with measures of organic content (Ash Free Dry Weight $r = 0.71$ and Total Organic Carbon $r = 0.95$). We therefore considered TP to be a relatively good proxy for catchment-level nutrient and organic enrichment. Although there was some variation in laboratory analysis methods, particularly for grain size, no significant differences in the relationships between biotic indices and stressors were detected, providing confidence that sample processing methods were not biasing our results.

Sites were assigned to seven wider biogeographical regions based generally on the coastal physical habitats and biological communities characterised by Shears et al. (2008). Because of limited data availability, three bioregions were combined with others, resulting in four bioregions overall: 'Northern' and 'Eastern' (North Island), 'Cook Strait' (North and South Island), and 'East South' (South Island) (Fig. 1). Additionally, each site was assigned to one of two Geomorphic Classes (hereafter 'estuary type'); GC 7, tidal lagoon, or GC 8, shallow drowned valley. These estuary types were based on landscape and waterscape characteristics (e.g. geology, basin morphometry), as well as hydrodynamic features due to river and ocean forcing of the estuary (Hume et al., 2016).

2.2. Index selection and calculation

Nine biotic indices and metrics were selected for testing. Some were identified in a recent international review (Borja et al., 2015), while others were developed, tested or modified in New Zealand (Keeley

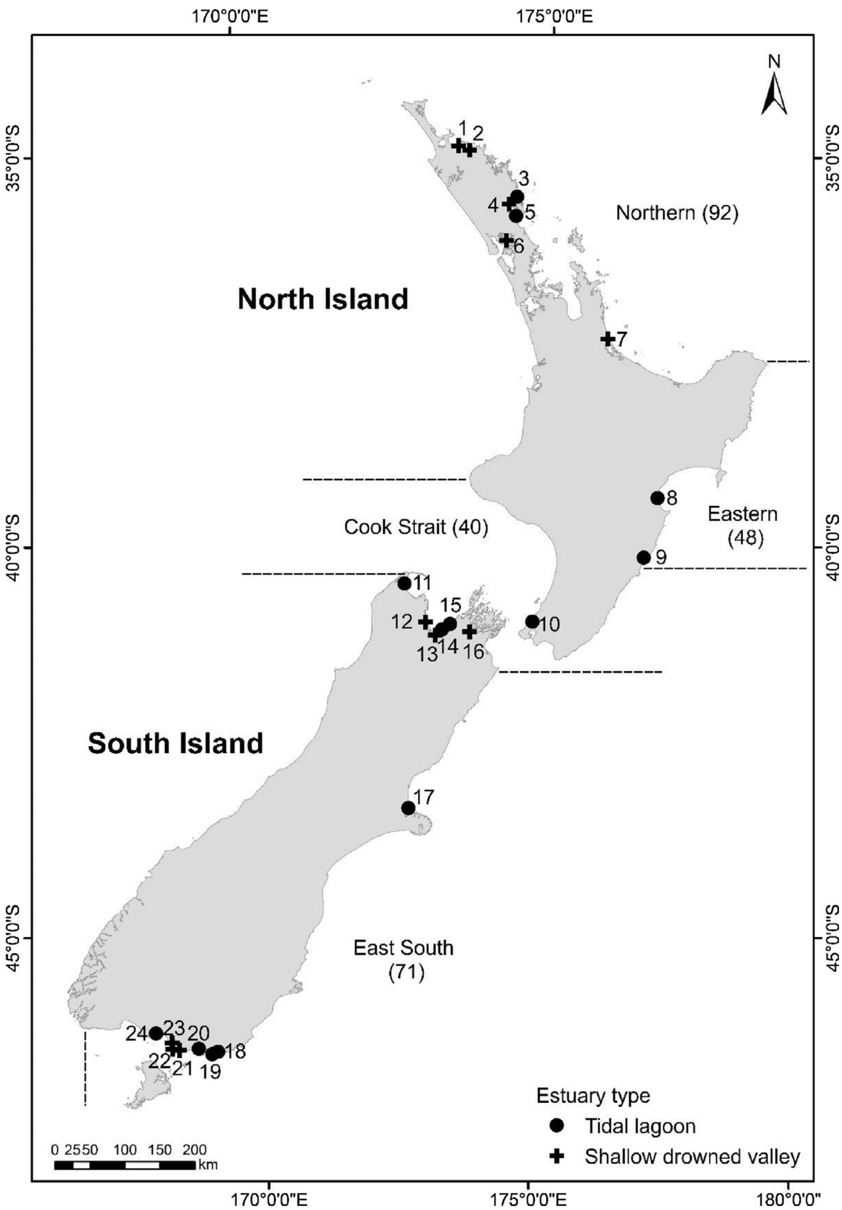

Fig. 1. Map of New Zealand showing the location and type of estuaries within each bioregion (Northern, Eastern, Cook Strait and East South) where macrofaunal and sediment samples used in this study were obtained. The number of sampling events for each bioregion is indicated in parentheses. See Appendix A in Supplementary material for estuary names and sampling details.

et al., 2012b; Robertson et al., 2016; Rodil et al., 2013). We sought indices that were previously identified as being responsive to stressors and represented a range of different approaches. The indices were: AMBI, M-AMBI, the Benthic Index (BENTIX), the Benthic Quality Index (BQI), RI-AMBI, TBI, and the Infaunal Trophic Index (ITI). In addition, simple metrics such as the log of total abundance [log(N)], and the total number of taxa (S) were included for comparison with biotic indices. For details of index and metric calculations and references see Table 1.

Indices based on EGs, i.e. AMBI, M-AMBI, RI-AMBI and BENTIX, were calculated using hybrid EG lists to reduce the number of unassigned taxa. The hybrid lists were created by supplementing regional EG classifications developed by Robertson et al. (2015) and Keeley et al. (2012a) with standard international AZTI (http://ambi.azti.es) classification. When a clash between EG assignments for a given taxon occurred, preference was given in the order of Robertson et al. (2015), then Keeley et al. (2012a), followed by AZTI classification, as preliminary analyses based on multiple regressions between indices and stressors (results not shown) indicated this provided best model fit.

M-AMBI was calculated using the method outlined in Sigovini et al. (2013). BQI was calculated using taxon sensitivity values $ES50_{0.05}$ as per Leonardsson et al. (2009) derived for 44 taxa using our dataset. For ITI, taxa were assigned to functional feeding groups based on published information and expert advice from local taxonomists. For TBI, taxa were matched to traits according to Rodil et al. (2013), although in a small number of cases an exact match was not possible and matching was instead based on similar taxa. Prior to TBI calculation, all meiofauna (i.e. Ostracoda, Nematoda, and Copepoda) were removed from the dataset. All biotic indices were calculated from all replicate data that met operational limits, except for TBI, which was calculated from average values per sampling event. Operational limits for AMBI, M-AMBI, RI-AMBI and BENTIX for each replicate were: assignment of EGs to taxa > 20%, number of individuals > 3, and number of species > 3 (Borja and Muxika, 2005).

Table 1
Equations used to derive biotic indices. EGI, EGII, EGIII, EGIV and EGV are ecological groups. N = number of individuals, S = number of taxa, H = Shannon-Weiner Diversity Index.

Biotic Index	Equation	Equation sources
AZTI Marine Biotic Index (AMBI)	$[(0 \times \%\ N_{EGI}) + (1.5 \times \%\ N_{EGII}) + (3 \times \%\ N_{EGIII}) + (4.5 \times \%\ N_{EGIV}) + (6 \times \%\ N_{EGV})]/100$	(Borja et al., 2000)
Richness-integrated AZTI Marine Biotic Index (RI-AMBI)	$[0 \times (\%N_{EGI} + \%\ S_{EGI}) + 1.5 \times (\%N_{EGII} + \%\ S_{EGII}) + 3 \times (\%\ N_{EGIII} + \%\ S_{EGIII}) + 4.5 \times (\%N_{EGIV} + \%\ S_{EGIV}) + 6 \times (\%\ N_{EGV} + \%\ S_{EGV})]/200$	(Robertson et al., 2016)
Multivariate AZTI Marine Biotic Index (M-AMBI)	$S/S_{max} + H/H_{max} + (1 - AMBI/6)/3$	(Muxika et al., 2007; Sigovini et al., 2013)
Benthic Index (BENTIX)	$[6 \times \%\ N_{EGI} + N_{EGII} + 2 \times (\%\ N_{EGIII} + N_{EGIV} + N_{EGV})]/100$	(Simboura and Zenetos, 2002)
Infaunal Trophic Index (ITI)	$100 - 3.33 \times [(0 \times N_1 + 1 \times N_2 + 2 \times N_3 + 3 \times N_4)/(N_1 + N_2 + N_3 + N_4)]$ where N_1, N_2, N_3 and N_4 are the number of individuals in suspension detritus, interface detritus, surface deposit and subsurface deposit feeding groups, respectively.	(Word, 1979)
Benthic Quality Index (BQI)	$\Sigma\ (N_i/N_{class} \times ES50_{0.05i}) \times \log_{10}(S + 1) \times (N_{total}/(N_{total} + 5)$ where $ES50_{0.05i}$ is the taxon sensitivity value, calculated as the 5th percentile of the rarefied species richness (ES50) for each sample calculated according to Hurlbert (1971), N_i is the number of individuals in taxon i, N_{class} is the total number of individuals of taxa having a sensitivity value, S is total number of taxa and N_{total} is the total abundance of individuals in the sample. A half saturation constant of 5 was used to reduce the index value when total abundance is low ($N < 20$).	(Rosenberg et al., 2004; Leonardsson et al., 2009)
Traits Based Index (TBI)	$1 - (\Sigma max - \Sigma actual)/\Sigma max$ where $\Sigma actual = S_{Top} + S_{Erect} + S_{SS} + S_{Medium} + S_{Sus} + S_{Sedentary} + S_{Worm}$. $\Sigma max = S_{TopMax} + S_{ErectMax} + S_{SSMax} + S_{SedentaryMax} + S_{SusMax} + S_{MediumMax} + S_{WormMax}$. where Max is the maximum expected value for a given number of replicates based on rarefaction of empirical data (see Table A1 in Appendix A of Rodil et al., 2013). Trait definitions (e.g. Top, Erect, SS, Medium, Sus, Sedentary and Worm), justification for the use of these particular traits, and full details on index calculation are described in Rodil et al. (2013). Σmax = 133.6, 175.4, 204.6, 212.5, 226.4 for 3, 6, 9, 10 and 12 replicates, respectively. Σmax values were extrapolated when the number of replicates differed from those listed.	(Rodil et al., 2013)
Log Abundance [log(N)]	log(N)	
Number of species (S)	S	

2.3. Statistical analyses

Linear mixed models were used to determine the amount of variation for each biotic index accounted for by sediment mud content, total phosphorus (TP), and metals contamination. Natural variability associated with space, estuary type, and time was also accounted for in the models. Macrofaunal and stressor data were averaged by sampling event (i.e. estuary-site-year-month). Zero values were assigned to stressor values below analytical detection limits. After initial data exploration and quality assurance following a standard protocol (Zuur et al., 2010), the dataset comprised 251 sampling events (n = 251) from 143 sites located in 24 estuaries (Fig. 1, Appendix A in Supplementary material). Mud and TP were square-root transformed.

Principal component analyses (PCA) were performed on the log-transformed normalised metal concentrations (i.e. Zn, Cu and Pb), and the first axis (metals PC1, 92.3% of the total variance, hereafter "metals") was used to characterise the overall range in metal contamination. All response (index) and predictor (stressor) variables were centered and scaled before the analyses (by subtracting the overall mean from each observation and dividing the result by the overall standard deviation), to allow direct comparison of regression coefficients and inference about the relative sizes of effects among stressors and indices. The inverse of AMBI and RI-AMBI values were used in models to allow direct comparison with other indices, i.e. a higher index value suggests better benthic condition.

Indices were generally normally distributed, thus the models were fitted with Gaussian errors. However, some indices had mild degrees of skewness and were transformed to improve normality, including square-root transformation of TBI and S, and reverse square-root transformation of AMBI and BENTIX (i. e. $\sqrt{max(x) - x}$). Associations between pairs of indices were tested using Pearson correlations (df = 249 and $p < 0.001$ for all results). Collinearity among predictor variables was checked using a variance inflation factor (VIF < 3, Zuur et al., 2010). All models were fitted with mud, metals, and TP as fixed effects. To avoid over-parameterisation of the models, stressor interactions were not included.

Because data exploration showed the relationship between indices and stressors varied between estuary types, the models were fitted with uncorrelated random intercepts and slopes for each estuary type. Additionally, to account for natural variability in index responses to stressors, the effects of estuary (nested in bioregion), bioregion, and sampling year were included as random effects in the models (random intercepts only). Our interest in random effects lies in the variation among them rather than the specific effects of each variable. They allow variation among levels around the intercept (intercept only) and/or the slope (random slopes) of each model that is quantified as the standard deviation (SD). Because the number of replicates differed between sampling events, the models were weighted by the square-root of the number of replicates (range 1 – 15).

The use of linear mixed models allowed the separation of the variability of a response variable (biotic index) into two components: the fixed effects of stressors and the random effects of natural variability. Stressor (fixed) effects are ideally maximised in a desirable index to respond to stressors, while natural variability (random effects) is ideally minimised so that an index is comparable over a range of spatio-temporal scales. To quantify these two components, marginal pseudo-R^2 (accounting for fixed effects) and conditional pseudo-R^2 (accounting for fixed and random effects) were calculated for each index (Nakagawa and Schielzeth, 2013).

Purposely, no model selection for fixed effects was conducted as the interest in the analyses was the comparison of the relative sizes of the effects of each of the three selected stressors. Selection of each model's optimal random structure was based on data exploration, AIC criteria and residual inspection. Because it is challenging to calculate degrees of freedom to obtain *p*-values in complex unbalanced mixed model designs (Bolker et al., 2009), we opted not to report *p*-values, but to interpret statistical significance of regression coefficients based on the overlap of their 95% confidence intervals (CI) with zero. Models were validated by plotting residuals versus fitted values, versus each covariate in the model. They were also validated by assessing residuals for temporal and spatial dependency. Linear mixed models were fitted using the lme4 library (Bates et al., 2014) of the software R (R Core Team, 2014).

3. Results

3.1. Data summary

The dataset contained 125 macrofaunal taxa, of which 27% were identified to species, 24% to genus, 31% to family, and 18% to order or above. At the level of bioregion, the most commonly occurring taxa were amphipods, polychaetes (Capitellids e.g. *Heteromastus filiformis*, Spionids e.g. *Prionospio* sp. and *Scolecolepides* sp., and Nereids e.g. *Nicon aestuariensis*), and bivalves (*Austrovenus stutchburyi*, *Arthritica bifurca* and *Macomona liliana*). The proportion of taxa assigned to EGs based on Robertson et al. (2015), Keeley et al. (2012a), and AZTI were 52%, 5%, and 25% respectively, with the rest (18%) unable to be assigned.

Summary values for each of the stressors and biotic indices are displayed in Table 2. Maximum zinc concentration (231 mg/kg) was higher than the national low threshold guidelines (ANZECC ISQG Low = 200 mg/kg, ANZECC, 2000) developed specifically for Australasia, based on the "effects range-low" principle (e.g. Long and Morgan, 1990) using the lower 10 percentile of biological effects data. Maximum concentrations of copper (38 mg/kg) and lead (40 mg/kg) were below the national low threshold guidelines (65 mg/kg and 50 mg/kg, respectively). Sediment mud content ranged from 0 to 82% and TP concentrations ranged from those indicative of unenriched to highly enriched sites (53 – 1413 mg/kg) (Robertson et al., 2002). All indices covered a range of values indicative of those from 'bad' to 'poor' and 'moderate' to 'high' ecological health based on previously determined health thresholds (Keeley et al., 2012b; Robertson et al., 2016; Rodil et al., 2013). EG-based indices AMBI, RI-AMBI and BENTIX were highly correlated with each other (r ≥ 0.74), as were M-AMBI, TBI, BQI and S (r ≥ 0.83). Moderate correlations occurred between log(N) and BQI (r = 0.60), log(N) and S (r = 0.59), and between M-AMBI and RI-AMBI (r = 0.53).

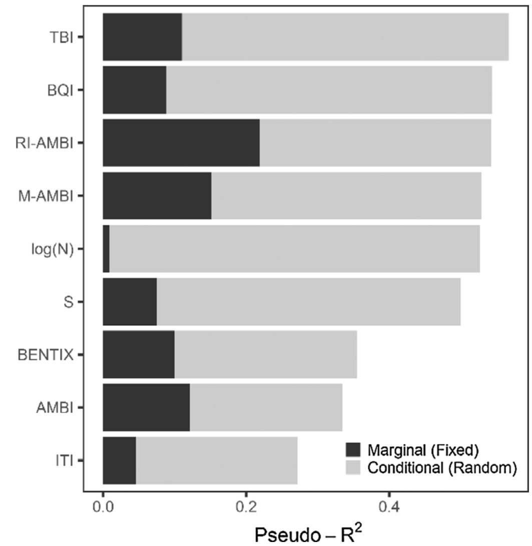

Fig. 2. Marginal (stressor, fixed effects) and conditional (natural variability, random effects only) pseudo-R^2 values obtained from linear mixed models of each biotic index. Stressors represent fixed effects of mud, total phosphorus (TP) and metals; natural variability represents random effects of estuary nested within bioregion, bioregion, year and estuary type.

3.2. Biotic indices and stressor relationships in relation to natural variability

Most indices (TBI, BQI, RI-AMBI, M-AMBI, log(N) and S) had a similar amount of variation explained by the stressors (fixed effects) and natural variability (random effects) combined, with overall pseudo-R^2 values ranging between 0.50 and 0.57 (Fig. 2). Overall pseudo-R^2 values were smaller (≤0.36) for BENTIX, AMBI and ITI.

The index with the largest amount of variation accounted for by the stressors (mud, metals and TP) was RI-AMBI (marginal pseudo-R^2 = 0.22), followed by M-AMBI, AMBI and TBI with marginal pseudo-R^2 = 0.15, 0.12 and 0.11, respectively (Fig. 2). However, for all indices, natural variability was larger (combined random effects of bioregion, estuary nested within bioregion, estuary type and year) than the variability explained by multiple stressor effects. Conditional pseudo-R^2 values (random effects only) ranged between 0.21 and 0.52 compared to marginal pseudo-R^2 values (fixed effects) that ranged between 0.01 and 0.22. Overall natural variability was smallest for AMBI (conditional pseudo-R^2 random effects only = 0.21), followed by ITI, BENTIX and RI-AMBI (conditional pseudo-R^2 random effects only = 0.23, 0.26 and 0.32, respectively), compared to other indices (≥0.38).

Table 2
Summary of values for (a) all physico-chemical sediment variables and (b) biotic indices in the dataset. Values below analytical detection limits (ADL) are shown.

Variables	Minimum	q25	Median	q75	Maximum	Mean	SD
a) Stressors							
Mud (%)	0.00	4.14	12.45	26.30	81.89	18.04	17.78
Total Phosphorus (mg/kg)	53.00	229.50	337.00	446.50	1413.00	357.18	182.30
Copper (mg/kg)	< ADL	< ADL	5.13	8.70	38.00	6.26	6.39
Zinc (mg/kg)	< ADL	18.50	35.00	51.50	231.00	40.85	33.99
Lead (mg/kg)	< ADL	2.06	4.47	7.63	40.47	5.92	6.10
b) Biotic indices							
AMBI	0.11	1.50	1.78	2.30	3.97	1.85	0.63
BENTIX	2.12	4.01	4.87	5.41	6.00	4.63	0.96
BQI	1.32	3.57	4.81	6.18	8.65	4.87	1.66
ITI	5.32	24.84	33.34	39.66	71.29	32.47	12.40
log(N)	2.08	3.53	4.33	4.96	6.64	4.28	1.01
M-AMBI	0.27	0.46	0.53	0.63	0.80	0.54	0.11
RI-AMBI	0.83	1.59	1.84	2.25	3.51	1.93	0.49
S	4.00	7.26	10.30	14.27	24.33	10.94	4.63
TBI	0.06	0.18	0.26	0.37	0.64	0.28	0.12

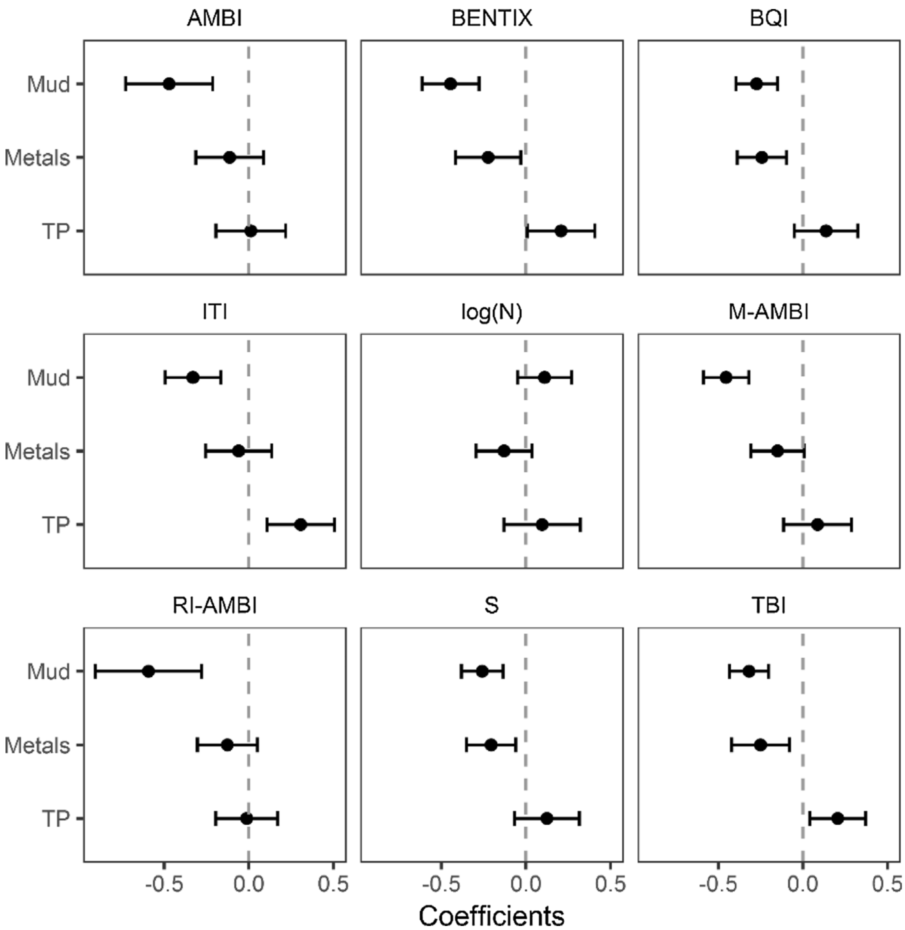

Fig. 3. Regression coefficients (± 95% CI) of fixed effects obtained from linear mixed models for each biotic index in response to mud, total phosphorus (TP) and metals (n = 251). Models were fitted with estuary nested within bioregion, bioregion, year and estuary type, as random effects.

Of the stressors, mud typically had the largest negative effect on index values, the only exception being log(N) (Fig. 3, Appendix B in Supplementary material). Mud had a significant effect on all indices (except log(N)), with the largest effect for RI-AMBI (regression coefficient = -0.59 ± 0.28 95% CI) and AMBI (-0.47 ± 0.21). Metals had a significant effect on TBI (-0.25 ± 0.08), BQI (-0.24 ± 0.10), BENTIX (-0.22 ± 0.03) and S (-0.21 ± 0.06). TP had a significant effect on ITI (0.31 ± 0.51), TBI (0.21 ± 0.37) and BENTIX (0.21 ± 0.41). TBI and BENTIX were the only indices for which all three stressors had a significant effect. With the exception of mud for log(N), mud and metals had negative effects on all indices, corresponding to poorer ecological health. On the other hand, the effect of TP on the indices was generally either negligible or positive, indicative of better ecological health.

RI-AMBI and AMBI had the smallest estuary-to-estuary variability within bioregions (SD ≤ 0.40), compared to other indices (SD 0.41 – 0.55, Fig. 4, Appendix B in Supplementary material). Larger-scale variation among bioregions was negligible for AMBI, RI-AMBI, BENTIX and ITI, and was larger for all other indices (SD 0.31 – 0.80). Inter-annual temporal variation was small in comparison to spatial variability associated with bioregion and estuary and was smallest for BENTIX and M-AMBI (SD = 0.00 and 0.08, respectively) compared to all other indices (SD 0.13 − 0.25). Intercept variation around estuary type was small or near negligible for most indices, although larger for log(N), BENTIX and ITI (SD 0.18 – 0.38). For most indices there was small variability in the slopes of the relationships between indices and stressors in relation to estuary type. However, there was considerable variation in the slopes of the relationship with mud for AMBI and RI-AMBI (SD = 0.14 and 0.20, respectively) in relation to estuary type. Similarly, slope variability for the relationship between TP and log(N), S, BQI and M-AMBI was larger (SD 0.08 – 0.11) than for other indices.

4. Discussion

4.1. Index and stressor relationships

To provide an overall measure of ecological health that is useful for managers, an index should be sensitive to human-related stressors and applicable across a wide range of contexts. Of the indices we assessed, RI-AMBI had the most variation accounted for by stressors, i.e. it was the most sensitive to stressors overall in our models.

Fine sediments within coastal environments are increased by human impacts and are recognized internationally as a major ecological threat (Airoldi, 2003, Gray 1997, GESAMP, 1994). In New Zealand estuaries, mud (i.e. fine sediment) is a dominant stressor on benthic communities and has been exacerbated by anthropogenic activities such as changes in land-use and coastal development (Robertson et al., 2015).

In our study, mud was the only stressor to have a significant effect on the RI-AMBI. This was not surprising as RI-AMBI, along with the other EG-based indices, was more likely to respond primarily to mud due to our choice of EG list (i.e. preference for NZ-specific EGs based on mud tolerance). A relatively strong relationship (simple linear regression $R^2 > 0.5$) between RI-AMBI and mud has already been demonstrated by Robertson et al. (2016).

However, benthic macrofauna within estuaries are often subject to multiple stressors relating to a variety of human pressures. Indices reflecting this may often be more useful than those sensitive to only one or highly correlated stressors. For TBI, there was a significant effect of all three stressors, as was also the case for BENTIX. TBI offers a different measure of ecological health based on functional redundancy and ecological resilience (Rodil et al., 2013), and in this respect is complementary to all other indices in our study.

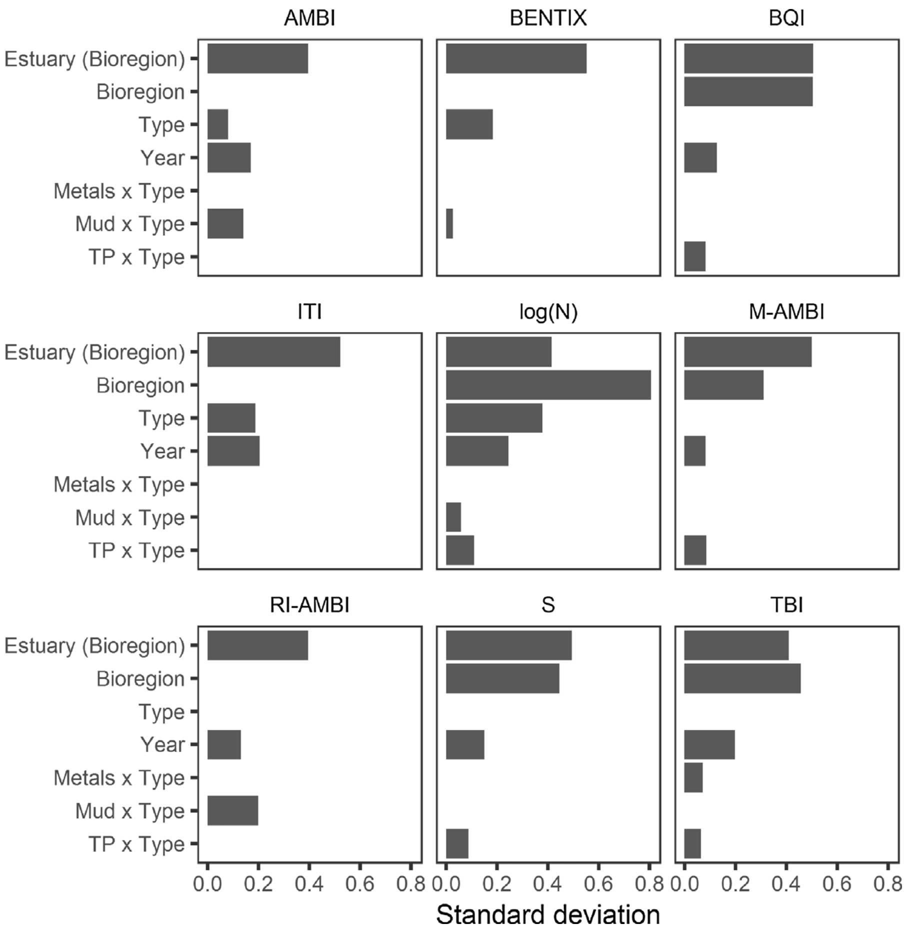

Fig. 4. Standard deviations for random effects obtained from linear mixed models of each biotic index to account for natural variability in relation to: estuary nested within bioregion, bioregion, year and estuary type, as random intercepts; and estuary type and mud x type, metals x type and total phosphorus (TP) x type, as random slopes.

The TBI also had the strongest relationship with metals, not surprising as it was developed to respond to metal contamination (Rodil et al., 2013). Metal concentrations in New Zealand estuaries are generally low, although specific areas can exceed guideline levels either naturally or due to anthropogenic impacts (Robertson et al., 2002). Deliberate placing of sites away from contamination sources may have limited maximum contamination concentrations in our study, with only zinc exhibiting values above New Zealand guidelines. Despite the low metal concentrations, we still detected a significant effect of metals on some indices. It has been found that ecological effects occur at metal concentrations below New Zealand guidelines (Hewitt et al., 2009; Rodil et al., 2013; Tremblay et al., 2017), which are higher than international guidelines (e.g. Long and Morgan, 1990) based on equivalent principles. Ideally, response to metals would be evaluated experimentally (e.g. sediment spiking) and/or by applying indices to locations with a greater range of metals contamination.

Total phosphorus had a significant effect on ITI, TBI and BENTIX. In contrast to mud and metals, increasing total phosphorus concentrations, where they had an effect, were associated with better ecological health. At low levels, nutrient (nitrogen and phosphorus) enrichment in estuaries can have a positive effect on benthic communities due to increased food availability through higher primary productivity, although beyond a critical point eutrophication starts to introduce negative effects (Cloern, 2001; McGlathery et al., 2007).

4.2. Natural variability

For RI-AMBI, AMBI, ITI and BENTIX, natural variability was smaller than for the other indices, with smaller spatial scales (i.e. estuary-to-estuary within each bioregion) accounting for the largest amount of this variation. For indices with larger amounts of natural variability (S, TBI, M-AMBI, log(N), BQI), larger-scale spatial variation (between bioregions) was also relatively important. Of several indices that use taxon counts in their calculations, TBI, M-AMBI and BQI were all highly correlated with the simple metric S, and it is possible that the relatively high bioregional variability for S contributed to the underlying cause of this variability for these indices. Additionally, for several indices responses to stressors varied in relation to estuary type. Although our analyses did not include some bioregions and estuary types due to data availability, our results suggest that interpretation of indices across estuary types and/or bioregions must be done with caution until further research, and possibly index threshold calibration, is conducted (see Section 4.4). Large spatial variability at various scales is inherent to estuarine macrofaunal communities globally (Edgar and Barrett, 2002; Ysebaert and Herman, 2002) and in New Zealand (Anderson et al., 2007). However, Robertson et al. (2015) found no significant spatial effect at the level of region on estuarine macrofaunal assemblages across New Zealand. Regions in their study were based on boundaries associated with governance of regional authorities which, compared to our study, were on a larger spatial scale than estuary, but generally a smaller scale than bioregion.

Although the effect of estuary type varied across indices, for all indices except ITI, it affected index vs stressor relationships to some extent by shifting the intercept or the slope of the relationship (see Fig. 4). The main difference between tidal lagoon and shallow drowned valley estuary types is that the greater depth and planform complexity of shallow drowned valley estuaries mean they are not as well flushed (Hume et al., 2016), which may have led to differing biological communities. Estuary type has been demonstrated to influence community metrics (e.g. Barbone et al., 2012). In New Zealand, relationships between macrobenthic community differences and higher-level variables such as estuary type have been identified (Ellis et al., 2006). RI-AMBI has previously been recommended for use in all shallow, intertidal dominated New Zealand estuaries (Robertson et al., 2016), suggesting that this index is comparable across tidal lagoon and shallow drowned valley estuaries. However, the amount of variation explained by mud in conjunction with estuary type suggests that RI-AMBI may not be comparable across these two estuary types.

The reported pattern of smaller inter-annual compared to overall spatial variability in index values is consistent with previous studies on soft sediment macrofauna (Anderson et al., 2007; Edgar and Barrett, 2002; Ysebaert and Herman, 2002). Smaller temporal scale (i.e. seasonal) effects were excluded from our analyses based on initial data exploration and data availability. However, seasonal effects on biotic indices have been demonstrated overseas (Magni et al., 2006; Reiss and Kröncke, 2005) and on macrofaunal communities in New Zealand estuaries, e.g. peaks in taxon abundance driven by recruitment (Greenfield et al., 2013; Hewitt and Thrush, 2007).

4.3. Unexplained variation

Unexplained variation can limit the usefulness of an index as a measure of ecological health, because it leaves doubt concerning whether changes in the index have been caused by anthropogenic or natural forces. A relatively high amount of unexplained variation was detected for all indices and was largest for BENTIX, AMBI and ITI, and unmeasured variables within the environment are a possible source of this. Sources of variability that could influence estuarine benthic community composition not explored in our study include fetch (i.e. wave and wind exposure, Hewitt et al., 2016), salinity (Gillett et al., 2015; Zettler et al., 2007), climate (Hewitt et al., 2016), invasive species (Zaiko and Daunys, 2015), fishery harvesting, biotic interactions, and other anthropogenic contaminants, e.g. pesticides. Stressors in the fixed part of our models were chosen from existing survey data primarily based on data availability and quality. Strong relationships have previously been demonstrated between some of our chosen stressors with others (e.g. mud and organic enrichment, Robertson et al., 2016). If data had been available, the addition of more stressors may have reduced the amount of unexplained variation within our models. Multiple stressors can also interact, resulting in a variety of outcomes e.g. synergistic and antagonistic interactions (Côté et al., 2016; Crain et al., 2008). The inclusion of stressor interactions in our models could potentially have reduced the amount of unexplained variability, however these were omitted to prevent model over-parameterisation.

4.4. Index thresholds and reference conditions

For biotic indices to be useful to managers, ecological health goals need to be set and thresholds identified to trigger management actions (Chainho et al., 2007; Rees et al., 2008). Most indices that are widely used, such as AMBI, have thresholds that are developed, tested or calibrated for the regions they are used in. Thresholds have been previously recommended for RI-AMBI for use in New Zealand estuaries on a national scale, and also within the Auckland region for TBI (Robertson et al., 2016; Rodil et al., 2013). However our results suggest that threshold values may need to be calibrated to ensure that differences in natural variability are accounted for, and the index values are comparable on a national scale. Approaches such as signal detection theory could be used to test the performance (sensitivity and specificity) of biotic indices in New Zealand and to set thresholds in a standardised way (Chuševė et al., 2016).

One of the fundamental principles for the derivation of thresholds and ecological health classification is the identification of type-specific reference conditions for minimally impacted or pristine sites (Bailey et al., 2004; Pollard and Huxham, 1998). The quantification of deviation from reference conditions requires characterisation of the relationship between a given stressor and its effect. However, reference conditions can be difficult to define in estuaries due to their high natural variability (Chainho et al., 2007; Sigovini et al., 2013) and the common absence of unimpacted areas (Borja et al., 2004). There are limited data on reference conditions for New Zealand estuaries. Considering the extent of bioregional and estuary type variation reported here for some of the tested indices, the identification of reference conditions could assist in the derivation of ecological health thresholds for different regions and estuary types.

4.5. Further considerations

Biotic index performance in New Zealand may require future reassessment if stressors increase beyond their current magnitude (Borja and Dauer, 2008); e.g. eutrophication is the dominant pressure in many countries more heavily populated than New Zealand (Rodil et al., 2013). Future reassessment may also be required if new stressors, other than those tested in the current study, become prominent. However, RI-AMBI, and other EG-based indices, could be adjusted to respond to other stressors through the development of new EGs for estuarine taxa.

Multivariate indices have been demonstrated to outperform simple metrics (e.g. S and N) for measuring stressor gradients in New Zealand estuaries (Ellis et al., 2015; Hewitt et al., 2005). However, as multivariate approaches have not been tested in New Zealand against biotic indices such as those in the current study, future work to develop and compare these on a national scale would determine their effectiveness.

None of the biotic indices tested in our study met all of the criteria for an ideal index. We therefore recommend the use of more than one index, i.e. a weight of evidence approach, to reduce uncertainty related to inaccuracy and misclassification of ecological health from the use of a single index (e.g. Borja and Muxika, 2005; Cairns et al., 1993; Rees et al., 2008).

Acknowledgments

We thank Massey University and the Ministry of Business, Innovation and Employment (MBIE) for funding this work (contract MAUX1502). We also acknowledge the support of New Zealand regional authorities that provided data and permission to use it: Northland Regional Council, Auckland Regional Council, Waikato Regional Council, Bay of Plenty Regional Council, Hawkes Bay Regional Council, Greater Wellington Regional Council, Marlborough District Council, Nelson City Council, Tasman District Council, West Coast Regional Council, Environment Canterbury, Christchurch City Council, Otago Regional Council and Environment Southland. Drew Lohrer from the National Institute of Water and Atmospheric Research (NIWA) calculated Traits Based Index (TBI) values. Emma Newcombe (Cawthron Institute) produced an initial literature review. Ben Robertson, Drew Lohrer (NIWA), Judi Hewitt (NIWA) and Barrie Forrest (Cawthron Institute) provided helpful comments on draft versions of the manuscript. We also thank the anonymous reviewers who provided helpful feedback on the manuscript draft.

Appendices A and B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2017.10.060.

References

ANZECC, 2000. Sediment quality guidelines. Australian and New Zealand Guidelines for Fresh and Marine Water Quality 2000 (Volume 2: Aquatic ecosystems – rationale and background information). Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Artarmon.

Airoldi, l., 2003. The effects of sedimentat on rocky coast assemblages. Oceanogr. Marine Biolol.: An Annual Rev. 41, 161–236.

Anderson, M., Pawley, M., Ford, R., Williams, C., 2007. In: Temporal Variation in Benthic Estuarine Assemblages of the Auckland Region. Prepared by UniServices for Auckland Regional Council. Auckland Regional Council Technical Publication

Number 348.
Bailey, R.C., Norris, R.H., Reynoldson, T.B., 2004. Defining the reference condition, bioassessment of freshwater ecosystems: Using the reference condition approach. Springer, US, Boston, MA, pp. 27–62.
Barbone, E., Rosati, I., Reizopoulou, S., Basset, A., 2012. Linking classification boundaries to sources of natural variability in transitional waters: a case study of benthic macroinvertebrates. Ecol. Indic. 12, 105–122.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using Lme4. (arXiv preprint arXiv:1406.5823).
Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24, 127–135.
Borja, A., Dauer, D.M., 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. Ecol. Indic. 8, 331–337.
Borja, A., Muxika, I., 2005. Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. Mar. Pollut. Bull. 50, 787–789.
Borja, A., Franco, J., Perez, V., 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Pollut. Bull. 40, 1100–1114.
Borja, A., Franco, J., Valencia, V., Bald, J., Muxika, I., Belzunce, M.J., Solaun, O., 2004. Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. Mar. Pollut. Bull. 48, 209–218.
Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 42–52.
Borja, Á., Marín, S.L., Muxika, I., Pino, L., Rodríguez, J.G., 2015. Is there a possibility of ranking benthic quality assessment indices to select the most responsive to different human pressures? Mar. Pollut. Bull. 97, 85–94.
Côté, I.M., Darling, E.S., Brown, C.J., 2016. Interactions among ecosystem stressors and their importance in conservation. Proc. R. Soc. B: Biol. Sci. 283.
Cairns, J., McCormick, P.V., Niederlehner, B., 1993. A proposed framework for developing indicators of ecosystem health. Hydrobiologia 263, 1–44.
Chainho, P., Costa, J., Chaves, M., Dauer, D., Costa, M., 2007. Influence of seasonal variability in benthic invertebrate community structure on the use of biotic indices to assess the ecological status of a Portuguese estuary. Mar. Pollut. Bull. 54, 1586–1597.
Chuševė, R., Nygård, H., Vaičiūtė, D., Daunys, D., Zaiko, A., 2016. Application of signal detection theory approach for setting thresholds in benthic quality assessments. Ecol. Indic. 60, 420–427.
Cloern, J.E., 2001. Our evolving conceptual model of the coastal eutrophication problem. Mar. Ecol. Prog. Ser. 210, 223–253.
Crain, C.M., Kroeker, K., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecol. Lett. 11, 1304–1315.
Diaz, R.J., Solan, M., Valente, R.M., 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. J. Environ. Manage. 73, 165–181.
Edgar, G.J., Barrett, N.S., 2000. Effects of catchment activities on macrofaunal assemblages in Tasmanian estuaries. Estuarine Coast. Shelf Sci. 50, 639–654.
Edgar, G.J., Barrett, N.S., 2002. Benthic macrofauna in Tasmanian estuaries: scales of distribution and relationships with environmental variables. J. Exp. Mar. Biol. Ecol. 270, 1–24.
Ellis, J., Ysebaert, T., Hume, T., Norkko, A., Bult, T., Herman, P., Thrush, S., Oldman, J., 2006. Predicting macrofaunal species distributions in estuarine gradients using logistic regression and classification systems. Mar. Ecol. Prog. Ser. 316, 69–83.
Ellis, J.I., Hewitt, J.E., Clark, D., Taiapa, C., Patterson, M., Sinner, J., Hardy, D., Thrush, S.F., 2015. Assessing ecological community health in coastal estuarine systems impacted by multiple stressors. J. Exp. Mar. Biol. Ecol. 473, 176–187.
Gillett, D.J., Weisberg, S.B., Grayson, T., Hamilton, A., Hansen, V., Leppo, E.W., Pelletier, M.C., Borja, A., Cadien, D., Dauer, D., Diaz, R., Dutch, M., Hyland, J.L., Kellogg, M., Larsen, P.F., Levinton, J.S., Llansó, R., Lovell, L.L., Montagna, P.A., Pasko, D., Phillips, C.A., Rakocinski, C., Ranasinghe, J.A., Sanger, D.M., Teixeira, H., Dolah, R.F.V., Velarde, R.G., Welch, K.I., 2015. Effect of ecological group classification schemes on performance of the AMBI benthic index in US coastal waters. Ecol. Indic. 50, 99–107.
GESAMP, 1994. Anthropogenic influences on sediment discharge to the coastal zone and environmental consequences. GESAMP Reports and Studies. 52.
Gray, J.S., 1997. Marine biodiversity: patterns, threats and conservation needs. Biodiv. Conserv. 6, 153–175.
Greenfield, B.L., Hewitt, J.E., Hailes, S.F., 2013. In: Manukau Harbour Ecological Monitoring Programme: Report on data collected up until February 2013. Prepared by NIWA for Auckland Council. Auckland Council Technical Report TR2013/027. (p. 39 plus appendices).
Hewitt, J.E., Thrush, S.F., 2007. Effective long-term ecological monitoring using spatially and temporally nested Sampling. Environ. Monit. Assess. 133, 295–307.
Hewitt, J., Anderson, M., Thrush, S., 2005. Assessing and monitoring ecological community health in marine systems. Ecol. Appl. 15, 942–953.
Hewitt, J.E., Anderson, M.J., Hickey, C.W., Kelly, S., Thrush, S.F., 2009. Enhancing the ecological significance of sediment contamination guidelines through integration with community analysis. Environ. Sci. Technol. 43, 2118–2123.
Hewitt, J.E., Ellis, J.I., Thrush, S.F., 2016. Multiple stressors, nonlinear effects and the implications of climate change impacts on marine coastal ecosystems. Glob. Change Biol. 22, 2665–2675.
Howarth, R.W., Marino, R., 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: evolving views over three decades. Limnol. Oceanogr. 51, 364–376.
Hume, T., Gerbeaux, P., Hart, D., Kettles, H., Neale, D., 2016. In: A classification of New Zealand's coastal hydrosystems. Report prepared for Ministry for the Environment. pp. 120.
Hurlbert, S.H., 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52, 577–586.
Keeley, N.B., Macleod, C.K., Forrest, B.M., 2012a. Combining best professional judgement and quantile regression splines to improve characterisation of macrofaunal responses to enrichment. Ecol. Indic. 12, 154–166.
Keeley, N.B., Forrest, B.M., Crawford, C., Macleod, C.K., 2012b. Exploiting salmon farm benthic enrichment gradients to evaluate the regional performance of biotic indices and environmental indicators. Ecol. Indic. 23, 453–466.
Leonardsson, K., Blomqvist, M., Rosenberg, R., 2009. Theoretical and practical aspects on benthic quality assessment according to the EU-Water Framework Directive-examples from Swedish waters. Mar. Pollut. Bull. 58, 1286–1296.
Long, E.R., Morgan, L.G., 1990. The potential for biological effects of sediments-sorbed contaminants tested in the National Status and Trends Program. NOAA Technical Memorandum NOS OMA. pp. 52.
Magni, P., Como, S., Montani, S., Tsutsumi, H., 2006. Interlinked temporal changes in environmental conditions, chemical characteristics of sediments and macrofaunal assemblages in an estuarine intertidal sandflat (Seto Inland Sea, Japan). Mar. Biol. 149, 1185–1197.
McGlathery, K.J., Sundbäck, K., Anderson, I.C., 2007. Eutrophication in shallow coastal bays and lagoons: the role of plants in the coastal filter. Mar. Ecol. Prog. Ser. 348, 1–18.
Muxika, I., Borja, A., Bald, J., 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 16–29.
Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4, 133–142.
Norkko, A., Thrush, S.F., Hewitt, J.E., Cummings, V.J., Norkko, J., Ellis, J.I., Funnell, G.A., Schultz, D., MacDonald, I., 2002. Smothering of estuarine sandflats by terrigenous clay: the role of wind-wave disturbance and bioturbation in site-dependent macrofaunal recovery. Mar. Ecol. Prog. Ser. 234, 23–42.
Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution in the marine environment. Oceanogr. Marine Biol.: An Annual Rev. 16, 229–311.
Pollard, P., Huxham, M., 1998. The European Water Framework Directive: a new era in the management of aquatic ecosystem health? Aquat. Conserv. Mar. Freshwater Ecosyst. 8, 773–792.
R Core Team, 2014. A Language and Environment of Statistical Computing. R foundation for statistical computing, Vienna, Austria.
Rees, H.L., Hyland, J.L., Hylland, K., Mercer Clarke, C.S.L., Roff, J.C., Ware, S., 2008. Environmental indicators: utility in meeting regulatory needs. An overview. ICES J. Mar. Sci. 65, 1381–1386.
Reiss, H., Kröncke, I., 2005. Seasonal variability of benthic indices: an approach to test the applicability of different indices for ecosystem quality assessment. Mar. Pollut. Bull. 50, 1490–1499.
Robertson, B., Gillespie, P., Asher, R., Frisk, S., Keeley, N., Hopkins, G., Thompson, S., Tuckey, B., 2002. Estuarine evironmental assessment and monitoring: A national protocol. Part A – Development of the monitoring protocol for New Zealand Estuaries: Introduction, rationale and methodology. (Prepared for Supporting Councils and the Ministry for the Environment, Sustainable Management Fund Contract No. 5096, p. 93).
Robertson, B.P., Gardner, J.P.A., Savage, C., 2015. Macrobenthic-mud relations strengthen the foundation for benthic index development: A case study from shallow, temperate New Zealand estuaries. Ecol. Indic. 58, 161–174.
Robertson, B.P., Savage, C., Gardner, J.P.A., Robertson, B.M., Stevens, L.M., 2016. Optimising a widely-used coastal health index through quantitative ecological group classifications and associated thresholds. Ecol. Indic. 69, 595–605.
Rodil, I.F., Lohrer, A.M., Hewitt, J.E., Townsend, M., Thrush, S.F., Carbines, M., 2013. Tracking environmental stress gradients using three biotic integrity indices: advantages of a locally-developed traits-based approach. Ecol. Indic. 34, 560–570.
Rosenberg, R., Blomqvist, 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. Mar. Pollut. Bull. 49, 728–739.
Shears, N.T., Smith, F., Babcock, R.C., Duffy, C.A.J., Villouta, E., 2008. Evaluation of biogeographic classification schemes for conservation planning: application to New Zealand's coastal marine environment. Conserv. Biol. 22, 467–481.
Sigovini, M., Keppel, E., Tagliapietra, D., 2013. M-AMBI revisited: looking inside a widely-used benthic index. Hydrobiologia 717, 41–50.
Simboura, N., Zenetos, A., 2002. Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new biotic index. Mediterranean Marine Sci. 3, 77–111.
Thrush, S.F., Hewitt, J.E., Cummings, V.J., Ellis, J.I., Hatton, C., Lohrer, A., Norkko, A., 2004. Muddy Waters: elevating sediment input to coastal and estuarine habitats. Front Ecol. Environ. 2, 299–306.
Tremblay, L.A., Clark, D., Sinner, J., Ellis, J.I., 2017. Integration of community structure data reveals observable effects below sediment guideline thresholds in a large estuary. Environ. Sci.: Processes Impacts 19, 1134–1141.
Van Hoey, G., Borja, A., Birchenough, S., Buhl-Mortensen, L., Degraer, S., Fleischer, D., Kerckhof, F., Magni, P., Muxika, I., Reiss, H., 2010. The use of benthic indicators in europe: from the water framework directive to the marine strategy framework directive. Mar. Pollut. Bull. 60, 2187–2196.
Weisberg, S.B., Ranasinghe, J.A., Dauer, D.M., Schaffner, L.C., Diaz, R.J., Frithsen, J.B., 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. Estuaries Coasts 20, 149–158.
WoRMS Editorial Board, 2017. World register of Marine Species. Available from http://

www.marinespecies.org at VLIZ. Accessed 2017–06-07.
Word, J.Q., 1979. Classification of benthic invertebrates into infaunal trophic index feeding groups. Coastal Water Research Project Biennial Report 1980. pp. 103–121.
Ysebaert, T., Herman, P.M., 2002. Spatial and temporal variation in benthic macrofauna and relationships with environmental variables in an estuarine, intertidal soft-sediment environment. Mar. Ecol. Prog. Ser. 244, 105–124.
Zaiko, A., Daunys, D., 2015. Invasive ecosystem engineers and biotic indices: giving a wrong impression of water quality improvement? Ecol. Indic. 52, 292–299.
Zettler, M.L., Schiedek, D., Bobertz, B., 2007. Benthic biodiversity indices versus salinity gradient in the southern Baltic Sea. Mar. Pollut. Bull. 55, 258–270.
Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1, 3–14.