Marine Pollution Bulletin 150 (2020) 110602

Contents lists available at ScienceDirect

Marine Pollution Bulletin

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

The development of a national approach to monitoring estuarine health based on multivariate analysis

D.E. Clark[a,b,*], J.E. Hewitt[c], C.A. Pilditch[b], J.I. Ellis[d]

[a] Cawthron Institute, Private Bag 2, Nelson, 7042, New Zealand
[b] University of Waikato, Gate 1, Knighton Rd, Hamilton, 3240, New Zealand
[c] National Institute of Water and Atmospheric Research, PO Box 1115, Hillcrest, Hamilton, 3216, New Zealand
[d] University of Waikato, New Zealand

ARTICLE INFO

Keywords:
Biotic index
Benthic communities
Ecological health
Environmental status
Sedimentation
Contaminants

ABSTRACT

New Zealand has a complex coastal environment spanning a large latitudinal gradient and three water masses. Here we assess whether multivariate analyses of benthic macrofaunal community composition can be a sensitive approach to assessing relative estuarine health across the country, negating the need for regional indices and reducing reliance on reference sites. Community data were used in separate canonical analyses of principal coordinates to create multivariate models of community responses to gradients in mud content and heavy metal contamination. Both models performed well (R^2 = 0.81, 0.71), and were unaffected by regional and estuarine typology differences. The models demonstrate a sensitive and standardized approach to assessing estuarine health that allowed separation of the two stressors. This approach could be applied to other stressors, countries or regions.

1. Introduction

Estuaries are among the most valuable of all ecosystems with regard to the services they provide to society (Costanza et al., 1997), many of which result from the high degree of connectivity with terrestrial systems and their proximity to people. However, as human populations have increased in coastal areas, so have the pressures on estuaries, which are exposed to multiple and cumulative stressors arising from adjacent catchments (e.g. increased sediment, nutrient and contaminant loads; Bricker et al., 2008; Johnston et al., 2015; Thrush et al., 2004), anthropogenic activities within the marine environment (e.g. fishing, dredging, shipping; Grosholz, 2002; Piló et al., 2019; Thrush et al., 1998), and global sources (e.g. climate change; Brierley and Kingsford, 2009). Such cumulative impacts have resulted in a loss of biodiversity and resilience, and an increased potential for tipping points to occur (Lotze et al., 2006). Thus, estuaries are not only one of the most heavily used, but also one of the most vulnerable natural systems worldwide (Agardy et al., 2005; Barbier et al., 2011; Lotze et al., 2006).

Environmental regulations increasingly require ecological assessment to quantify the impact of stressors on coastal ecosystem status and inform management decisions (e.g. the Clean Water Act or Oceans Act in USA, Australia or Canada; Water Framework Directive or Marine Strategy Framework Directive in Europe, and National Water Act in South Africa; Borja et al., 2008). For assessment methods to be useful they need to be (1) ecologically relevant, (2) feasible to implement, (3) linked to threshold or reference values so that users can assess the significance of an indicator value, (4) sensitive enough to measure status or trends that are relevant to policy decisions and reflect responses to management actions and ideally, (5) applicable over wide spatio-temporal scales (Borja and Dauer, 2008). Benthic macrofaunal communities are commonly used to assess environmental status (Borja et al., 2000; Dauer, 1993; Pearson and Rosenberg, 1978) because they respond relatively rapidly to stressors, integrate the effects of multiple stressors over time and are composed of a diverse range of species with differing functional roles, trophic levels and sensitivities. Incorporating community information into ecosystem health assessments allows organisms to ‘tell the story’, with respect to classifying sites along a continuum from degraded to non-degraded (Diaz et al., 2004).

Historically, the first approaches to extract information from macrofaunal community data included the calculation of simple metrics, such as the number of taxa or individuals and measures of community evenness and diversity (e.g. Margalef, 1958; Pielou, 1966; Shannon, 1948). These universally applicable metrics can be assessed against the Pearson and Rosenberg (1978) model of macrobenthic succession to provide an indication of environmental health, but they have limited ability to detect meaningful change because they do not differentiate

* Corresponding author. Cawthron Institute, Private Bag 2, Nelson, 7042, New Zealand.
E-mail address: dana.clark@cawthron.org.nz (D.E. Clark).

https://doi.org/10.1016/j.marpolbul.2019.110602
Received 5 June 2019; Received in revised form 16 September 2019; Accepted 16 September 2019
Available online 15 November 2019
0025-326X/

amongst different types of taxa (Ellis et al., 2015; Hewitt et al., 2005; Shade, 2016). The growing requirement for assessment of marine environmental status over the last two decades has led to a proliferation of more complex biotic indices, many of which also have foundations in the Pearson-Rosenberg model (Borja et al., 2015; Diaz et al., 2004). Many of these indicators (e.g. Borja et al., 2000; Grall and Glémarec, 1997; Simboura and Zenetos, 2002) work by assigning taxa into previously defined ecological groups, based on their response to stressors, and examining the relative proportion of these groups in the benthic community sample. This requires predefined knowledge of how a large number of species respond to stressors, and for many species the research to determine these responses has not been carried out.

Other approaches to tracking environmental health include multivariate methods, which describe assemblage patterns of the entire community (e.g. ordination-based approaches; Clarke, 1993; Flåten et al., 2007; Smith et al., 2001). Because multivariate approaches retain information on all taxa and their relative abundances, they can detect smaller changes in community structure (Attayde and Bozelli, 1998; Ellis et al., 2015; Gray et al., 1990; Hewitt et al., 2005; Warwick and Clarke, 1991). This sensitivity enables early detection of environmental deterioration, allowing management actions to be implemented before significant ecosystem damage occurs, thereby avoiding prolonged (and sometimes uncertain) recovery and/or costly remedial actions (Martinez-Crego et al., 2010). In addition, preservation of species composition information means outputs can be directly linked to changes in biodiversity and ecological functioning. This link with ecological functioning can be taken one step further by using multivariate approaches to assess changes in functional traits rather than species assemblages (e.g. Bremner et al., 2003; Hewitt et al., 2008).

Most biotic indices provide an overall measure of ecosystem health and are designed to be sensitive to a broad range of stressors. While this holistic approach can indicate the general health of a system and account for interactions amongst stressors, the inability to attribute degradation to a specific stressor makes targeted management action difficult (Martinez-Crego et al., 2010; Niemi et al., 2004). In addition, the desire to create biotic indices that track changes in ecosystem health in response to a suite of stressors has necessitated the use of expert judgement in index development. For those multi-stressor indices, taxa are often assigned to ecological groups using expert opinion due to lack of empirical knowledge of many species responses to stressors, and because accurately quantifying relationships between communities and multiple stressors is complex, given the uncertainties associated with interactions (Crain et al., 2008; Darling and Côté, 2008) and non-linear responses (deYoung et al., 2008). In contrast, stressor-specific indices can be developed from robust empirical relationships between benthic communities and the stressor of interest (e.g. Keeley et al., 2012; Robertson et al., 2016). In addition to providing managers with an objective assessment of health, these single-stressor indices diagnose the cause of degradation, enabling prioritization of mitigation measures. While multi-stressor indices have many merits, we advocate for the use of a suite of single-stressor indices, based on key pressures to the system, that allow managers to identify sources of degradation and interactions between stressors and apply appropriate action. These types of analyses (indices) would allow a weight of evidence approach (Magni et al., 2005) to the assessment of environmental status and methods to integrate the individual stressor scores into an overall score could be applied if required (e.g. Aubry and Elliott, 2006; Borja et al., 2004).

New Zealand spans 15 degrees of latitude and three water masses and, with more than 400 estuaries (Hume et al., 2016), provides an ideal place to test the robustness of biotic indices under different conditions. Here, we developed two stressor-specific Benthic Health Models (BHMs), which can be used to assess intertidal estuary health in response to increasing mud content (Mud BHM) and heavy metal contamination (Metals BHM) across New Zealand. Sedimentation and metal contamination are recognised as major threats to the health and functioning of estuaries globally and are routinely monitored, both in New Zealand and elsewhere (EU Marine Strategy Framework Directive, 2008; Hewitt et al., 2009; Hewitt et al., 2005; Hewitt et al., 2014; Lohrer et al., 2012; MacDiarmid et al., 2012; Magris and Ban, 2019; Rodil et al., 2013). In New Zealand, few estuaries have been unaffected by increased sediment inputs from land, increasing the total area of the estuary seafloor being classified as muddy sediments. A constrained multivariate ordination was used to model changes in community structure along each environmental gradient. The results of the models can be simplified into a health score, which allows estuary health to be tracked over time. In this study, we follow Hewitt et al. (2005) and define 'health' on the basis of the range of communities observed along gradients of anthropogenic impacts, rather than requiring identification of a "reference" condition or site. This definition identifies both acute effects and broader scale degradation in community structure.

The BHM approach has been successfully applied at estuary- (Ellis et al., 2015) and regional-scales (Hewitt et al., 2005), however, a national model that is able to detect changes across regional species pools or estuarine types has not been tested to date. National models would provide a standardised assessment method to enable the health of an estuary to be placed in a wider context and reduce the costs required to develop separate estuary-scale or regional-scale models. In addition to being sensitive to changes in ecosystem health, biotic indices need to be unaffected by different species pools (Berthelsen et al., 2018a; Gillett et al., 2015; Keeley et al., 2012) and natural environmental contexts (Barbone et al., 2012; Berthelsen et al., 2018a). These requirements are particularly important when developing a national index for a country such as New Zealand, with a strong latitudinal gradient and estuaries open to three different water masses. To this end, we developed national BHM models and tested their ability to discriminate between effects caused by the two stressors despite differences in regional species pools and estuarine physical type (i.e. tidal lagoons and shallow river valleys).

2. Methods

2.1. Macrofaunal and physico-chemical dataset

Data were obtained from surveys undertaken between 2002 and 2017, by New Zealand's regional government authorities for the purposes of estuarine monitoring (815 sampling events across 70 estuaries). Where information was available for multiple years and seasons, only one sampling event was used, with preference given to data collected between 2010 and 2014 (66% of sites) and spring/summer (October to March; 72% of sites), the years and months when most data was collected, in order to reduce potential between-year and between-season variability. Counts of larval planktonic groups (e.g. megalope, larvae and eggs) and juvenile taxa were removed from the dataset before model development, which limits the effect of recruitment pulses on the models. The 192 sites, from 34 estuaries, spanned 12 degrees of latitude and encompassed two dominant estuary types and a range of bioregions (Fig. 1). Surveys were carried out according to a standardised protocol (Robertson et al., 2002), with samples collected from sites located at mid-to-low tidal height away from point-source discharges. Some variations in salinity and exposure were expected to be present across site locations. However, sites suspected to be significantly influenced by freshwater, based on their location or the presence of high abundances of insects, were removed from the dataset.

Macrofaunal samples (n = 3 to 15 replicates per site) were collected using a 13 cm diameter core extending 15 cm into the sediment and sieved using a 500 μm mesh. Experts identified organisms to the lowest practicable resolution. Taxonomic nomenclature followed the World Register of Marine Species (WoRMS Editorial Board, 2017) and where differences in taxonomic resolution arose, we aggregated to higher taxonomic groups. Some taxa were removed from the dataset before the analysis (refer to Appendix A in the Supplementary Material for

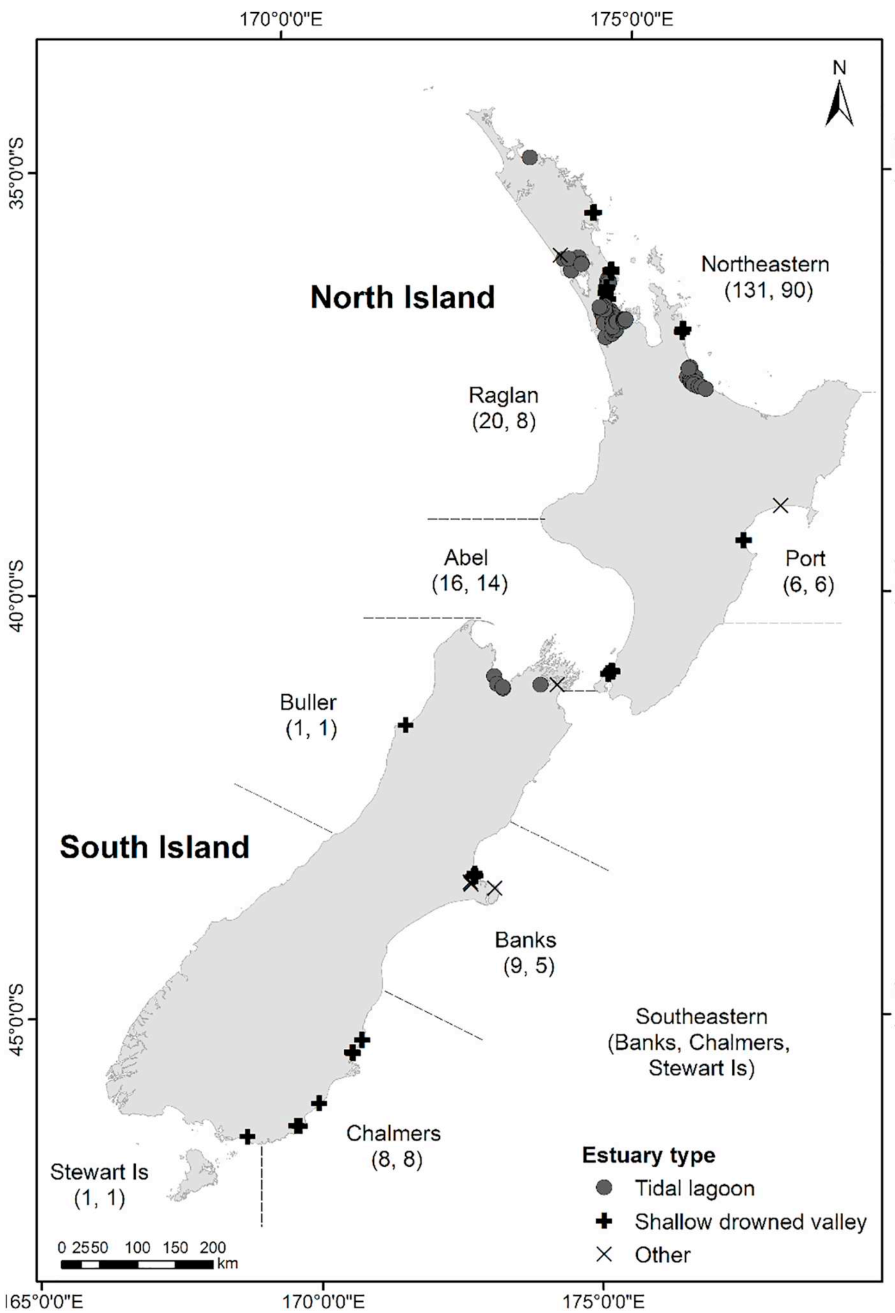

Fig. 1. Map of New Zealand showing the location and estuary type (Hume et al., 2016) for the sites used to construct the Mud and Metals Benthic Health Models (BHMs). The number of sites for each bioregion (Northeastern, Port, Raglan, Abel, Buller, Banks, Chalmers, Stewart Island), as defined by Shears et al. (2008), is indicated in parentheses for both the Mud BHM (first number) and Metals BHMs (second number).

justification). Taxonomic resolution was the same for both models and the final datasets had 125 (Mud BHM) and 109 (Metals BHM) taxa, with 80% of taxa identified to family level or lower (refer to Appendix A in Supplementary Material for a complete list of taxa used in the models).

Sediment samples (n = 1 to 12 replicates per site) were collected to a depth of 2 cm concurrent with macrofaunal samples. Samples were analysed for mud content (grain size $< 63\,\mu m$) using either wet sieving or laser diffraction analysis. To increase comparability between different sediment grain size analyses, we converted sediment mud proportions to a percentage of the < 2 mm sediment fraction (e.g. percentage of $< 63\,\mu m$ out of the < 2 mm sediment fraction) because the maximum grain size analysed differed between analysis methods (e.g. Malvern Mastersizer laser only analyses grains < 2 mm, while all grain sizes are generally analysed during wet sieving). Exploratory analysis on final models showed no pattern associated with differing sediment grain size analysis methods. At most sites (133 out of 192 from 29 of the 34 estuaries), sediment samples were also analysed for total concentrations (mg kg^{-1}) of copper (Cu), lead (Pb) and zinc (Zn), which are the key heavy metals of concern in New Zealand (ARC, 2004). Despite slight variations in metal analysis methods between sites, results from the different analytical methods were assumed to be comparable by Berthelsen et al. (2018b). In general, the methods followed the US EPA 200.2 protocol of strong acid (nitric/hydrochloric) digestion followed by Inductively Couple Plasma Mass Spectrometry (US EPA, 1994).

2.2. Model development and validation

All data were averaged to the level of site to construct the models. Differing numbers of replicates can lead to bias in multivariate analyses by underestimating species richness at sites with lower numbers of replicates and thereby overestimating dissimilarity (e.g. Chao et al., 2005). Exploratory analysis showed a slight reduction in species richness at sites where only three replicates were collected (mean number of taxa per core was 15 (n = 3) vs 21–23 (n > 3)), however, these represented only 15% of samples and no patterns were observed that would indicate the number of replicates was influencing model outputs (i.e. sites were dispersed across the health score gradient). Previous studies have found that models based on all available information (i.e. a mixture of sample sizes) were most useful (Anderson et al., 2002, 2006).

Two models were developed, one based on community response to sediment mud content (Mud BHM) and the other based on response to sediment Cu, Pb and Zn concentrations (Metals BHM). Several New Zealand studies have demonstrated that mud content can be used as an indicator of stress related to sedimentation from land-based sources (Anderson, 2008; Ellis et al., 2017; Robertson et al., 2015; Thrush et al., 2003, 2005). Exploratory analyses examining the influence of other environmental variables (nitrogen, phosphorus, organic matter and salinity) showed mud and metals to be the key environmental stressors structuring benthic communities at our monitoring sites (data not shown). We used log-transformed percentage mud content as the environmental gradient for the mud model. Total extractable Cu, Pb and Zn were highly correlated (Pearson's $r = 0.85–0.91$) so a Principal Component Analysis (PCA) was used to derive a single variable (the first principal component axis; PC1) that would characterise a gradient corresponding to increases in the concentrations of all three metals. The PCA was performed on log-transformed Cu, Pb and Zn concentrations and the PC1 axis (PC1 metals) explained 92% of the variance. Log-transformations were chosen to render the data as close to normally distributed as possible for modelling and exploratory analyses indicated that the choice of transformation did not affect model outputs. Zero values were assigned to metal concentrations below analytical detection limits (22% of sites for Cu, < 2% of sites for Pb and Zn). Mud concentrations at sites within the Mud BHM ranged from 0-98% and metal concentrations at sites within the Metals BHM ranged from 0-49 mg kg^{-1} for Cu, 0–70 mg kg^{-1} for Pb and 0–288 mg kg^{-1} for Zn (untransformed). Given these values represent maximum concentrations observed across 70 estuaries over the past 15 years, we believe they cover the range of values likely to be encountered in most estuaries across New Zealand.

Canonical analysis of principal coordinates (CAP; Anderson and Robinson, 2003; Anderson and Willis, 2003) was used to derive the model relationship between macrofaunal community structure and each environmental gradient (i.e. mud and metals). CAP allows a constrained ordination to be carried out on the basis of any dissimilarity or distance measure of choice and determines the axes that best discriminates an environmental gradient. All CAP analyses were performed on square-root transformed Bray-Curtis macrofaunal community dissimilarities (Bray and Curtis, 1957) using 9999 permutations, with separate CAP models constructed for mud and metals. A square-root transformation (standard down-weighting for macrofaunal count data; Clarke and Gorley, 2015) was chosen to de-emphasis the influence of dominant taxa while still allowing differences in relative abundance to influence the results, as this was considered important for determining estuary health. Leave-one-out residual sum of squares was used to decide upon an appropriate value for the number of Principal Coordinate Analysis (PCO) axes (m) and diagnostics were checked to ensure this was appropriate for each model (Anderson et al., 2008).

Model CAP scores were simplified into a five-category health score system by splitting the CAP score gradient into five evenly spaced groups, which were re-scaled from 1 (least impacted) to 6 (most impacted) for ease of interpretation. One-way PERMANOVA was used to test whether the ecological health groups corresponded with significant differences in community structure. Unrestricted permutation of raw data was used, with 9999 permutations, type III sum of squares and ecological health group as a fixed factor. As a form of model validation, changes in community structure across the five ecological health groups were characterised using SIMPER to ensure that the discriminating taxa across groups were consistent with what is known about the habitat preferences and metal tolerances of organisms. Discriminating taxa that cumulatively contributed between 70-74% to the similarity of each group were assigned to one of three categories based on literature (refer to Appendix B in the Supplementary Material for more information). For the Mud BHM, the grain-size preference categories were sandy, intermediate/unknown and muddy, with the intermediate/unknown group a placement for taxa that showed a preference for habitats with intermediate grain-size or for species that could not be assigned based on the literature. For the Metals BHM, the metal sensitivity categories were sensitive, mixed/unknown and tolerant, with the mixed/unknown group a placement for taxa that showed an inconsistent response to metal contamination or for species that could not be assigned based on the literature. All statistical analyses were carried out using the statistical software PRIMER 7 (v 7.0.13) with the PERMANOVA + add-on (Anderson et al., 2008; Clarke and Gorley, 2015).

The accuracy of each CAP model at identifying and predicting real and repeatable patterns in the data was measured by its ability to 1) correctly place validation sites onto the environmental gradient and 2) be unaffected by temporal variability that was not associated with changes in environmental drivers. The first validation is an important step because high canonical correlation does not necessarily mean good predictive power (Anderson et al., 2006). For example, high canonical correlation can be achieved by simply increasing the number of principal coordinate analysis (PCO) axes (m) to be used in the CAP analysis. Validation sites were chosen to maximize spread across the environmental gradient and included a range of estuaries and regions. All validation sites were independent sampling events, taken from a separate dataset from the one used to develop the models. Some of the locations of the validation sites were the same as some of the model sites but sampled in a different year, similar to the validation procedure used for the regional model (Anderson et al., 2006). Twenty-nine sites were used to validate the mud model and 20 were used for the metals model; equivalent to 15% of the number of model sites. Mud content at the Mud BHM validation sites ranged from 0.6 to 93% mud while maximum Cu, Pb and Zn concentrations at the Metals BHM validation sites were 43, 65 and 216 mg kg^{-1} respectively.

The BHMs were used to place each validation site onto the environmental gradient axes by calculating the Bray-Curtis dissimilarity between that site and the sites in the model. An option within the CAP procedure in PRIMER 7 allows the addition of new sites to the model without altering distances among other points because the dissimilarity between any two sites does not depend on the other sites in the model (Anderson et al., 2008). Physico-chemical values calculated using the BHMs were the predicted values along the environment gradient. Linear regression of sampled versus predicted physico-chemical values (either ln % mud or PC1 metals) were used to identify sites whose predicted values deviated most from their observed values and in which direction. A 1:1 line (i.e. with slope (b) = 1 and intercept (a) = 0) was drawn to help interpret the positions of the points. If prediction is exact, the points would lie precisely on this line. The slope of the linear relationship, b, and the strength of the relationship (coefficient of determination, R^2), between the predicted and observed values was also used to determine validation success. Models were considered good if b and R^2 were close to 1.

We also tested whether natural temporal variability in community composition across years resulted in a site sampled at a different time, but with similar mud or metal concentrations, having markedly different CAP scores (designated as being greater than the range of values for a single group). Nine sites (4–6 sampling events per site) were used to test the Mud BHM and seven sites (2–3 sampling events per site) were used to test the Metals BHM.

Co-variance between mud and metals can make it difficult to separate stressor effects. The potential for interactions between the Mud

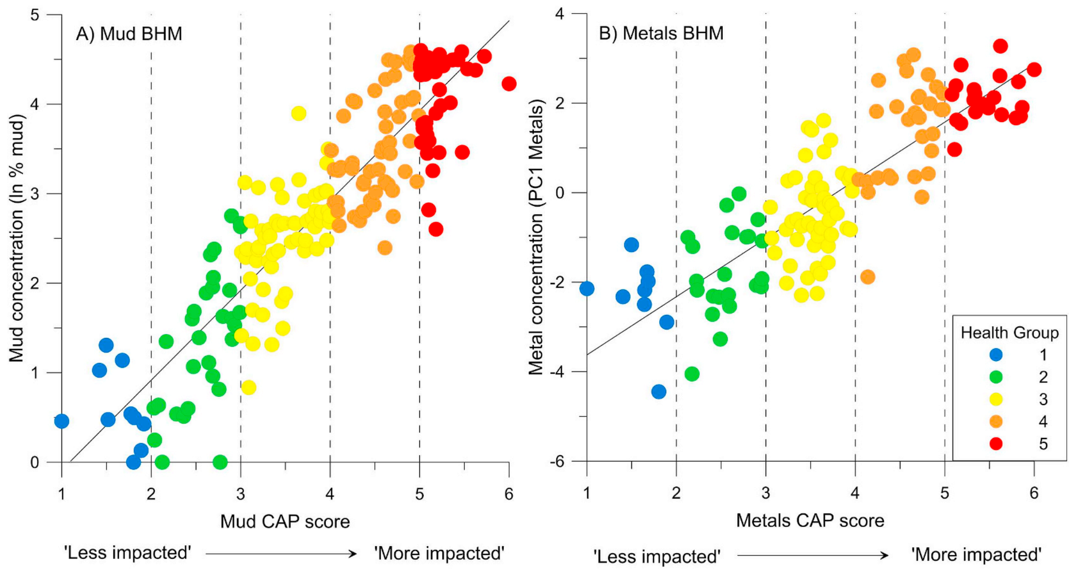

Fig. 2. Benthic Health Models (BHMs) developed using canonical analysis of principal coordinates (CAP) constrained by either A) mud (ln % mud) or B) metals (first axis of principal component analysis based on log transformed copper, lead and zinc). Grey dashed lines and symbol colours demarcate the ecological health categories for each model. A linear regression has been fitted for each of the models; Mud BHM y = 1.0038x – 1.0911, R^2 = 0.81, Metals BHM y = 1.3002x – 4.9258, R^2 = 0.71. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and Metals BHMs was examined in two ways. First the Pearson's correlation coefficient between the Mud and Metals CAP scores was calculated to examine the potential for interaction between the two models, with correlation coefficients $r \geq |0.95|$ representing a strong interaction (Anderson et al., 2008). Secondly, because all the sites in the Metals BHM were also included in the Mud BHM, the variance in macrofaunal structure explained by each independent variable (mud and metals) could be partitioned. Following the methods of Anderson et al. (2008) and Borcard et al. (1992), sequential multiple linear regressions were conducted using the DistLM routine in PRIMER on the macrofaunal abundance (square root transformed) Bray-Curtis dissimilarities to partition the variance explained by mud and metals and identify the mixed effect.

2.3. Testing the model across different regions, estuary types and scales

In order to apply the BHM approach on a national scale, it is important that the models produce consistent results across different environmental contexts and species pools. To test whether the models were affected by such differences, sites were grouped by estuary type based on Hume et al.'s (2016) classification of New Zealand hydrosystems and region based on Shears et al.'s (2008) biogeographic classification scheme (Fig. 1). Due to limited data availability for the Metals BHM, three bioregions (Banks, Chalmers and Stewart Island) were combined into a single group (Southeastern) for this model, and groups with less than five sites (Fig. 1) were removed from the analysis of both the Metals and Mud BHMs. This resulted in two levels for the 'estuary type' factor for both models (tidal lagoons and shallow drowned valleys), six levels for the 'region' factor for the Mud BHM (Abel, Banks, Chalmers, Portland, Raglan and Northeastern) and five levels for the 'region' factor for the Metals BHM (Abel, Southeastern, Portland, Raglan and Northeastern). After initial data exploration following the protocol of Zuur et al. (2010), analysis of covariance (ANCOVA) using type III sum of squares was used to test if the relationship between the model CAP scores and the environmental gradient (either mud or metals) varied with region or estuary type using Statistical Analysis Software (SAS).

To understand how the national outputs relate to assessments carried out at finer scales of resolution, national BHM CAP scores were compared to those generated from separate BHMs developed using data from one estuary (Tauranga Harbour; Ellis et al., 2015) or one region (Auckland; Hewitt et al., 2005) using Spearman's rank correlations. Eighteen sites in the national BHM were also in the single estuary BHM and 44 (Mud BHM) and 43 (Metals BHM) sites in the national BHM were also in the single region BHM.

3. Results

3.1. Model performance and validation

The CAP analyses underlying the Mud and Metals BHMs performed well (Fig. 2). The CAP model (m = 29) based on mud content resulted in a canonical correlation of 0.90 (R^2 = 0.81), with the permutation test indicating that correlation between the CAP scores and the mud gradient was significantly different from zero ($p < 0.0001$). CAP analysis based on metals (m = 20) also showed a strong (canonical correlation = 0.84, R^2 = 0.71) and significant ($p < 0.0001$) relationship between benthic macrofaunal communities and sediment metal concentrations.

Sites were split into five ecological health groups, based on model CAP scores, and information on stressor values observed at sites within each group is provided in Appendix B of the Supplementary Material. For both models, PERMANOVA indicated a significant difference in community structure across the five groups ($p < 0.0001$). Pairwise comparisons showed these differences were significant across all groups ($p < 0.04$), apart from Group 1 and 2 for the Metals BHM, which was not significant ($p = 0.065$). SIMPER analysis showed that community dissimilarity was 84% and 78% between Groups 1 and 5 for the Mud and Metals BHMs, respectively.

As another form of model validation, taxa characterising each ecological health group were identified using SIMPER and compared with known information related to grain-size preferences or levels of metal contamination, from previous studies (refer Appendix B), to determine if the BHMs placed taxa in the expected ecological health groups. Unsurprisingly, taxa driving differences between Mud BHM groups have differing grain-size preferences, with most of the taxa characterising Group 1 preferring sand (e.g. the shellfish *Austrovenus stutchburyi* and *Paphies australis*, the gastropod *Notoacmea*, the polychaete *Aonides* and phoxocephalid amphipods) and many of the taxa characterising Group 5 preferring mud (e.g. the crabs *Austrohelice*, *Hemigrapsus* and *Hemiplax*,

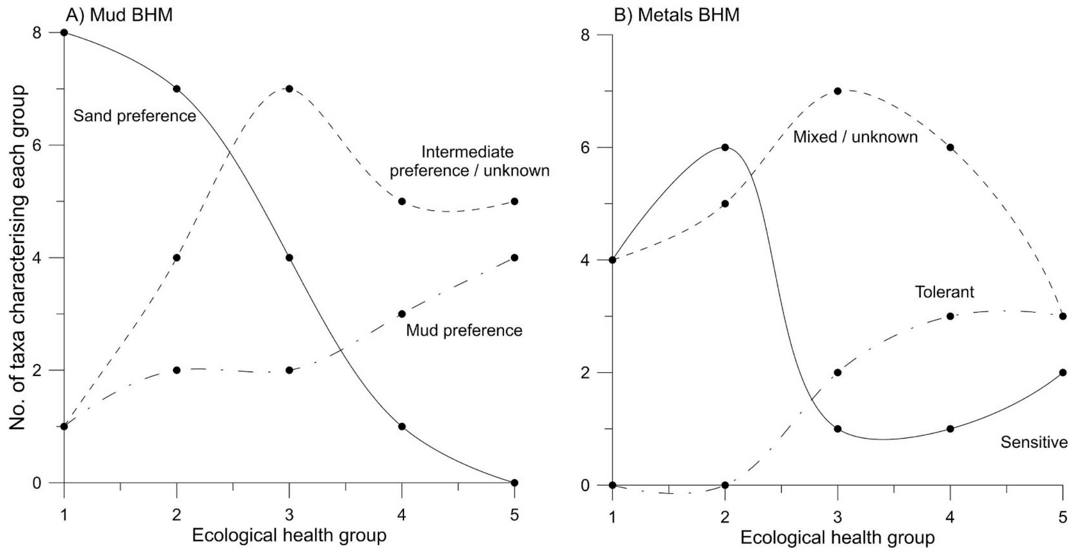

Fig. 3. A) Number of taxa characterising Mud Benthic Health Model (BHM) ecological health groups, grouped by grain-size preference (sand, intermediate/unknown, mud). B) Number of taxa characterising Metals Benthic Health Model (BHM) ecological health groups, grouped by metal contamination sensitivity (sensitive, mixed response/unknown, tolerant). Taxa characterising each ecological health group were identified using SIMPER (taxa that cumulatively contributed between 70-74% to the similarity of each group) and grain-size preferences and metal contamination sensitivities were assigned based on literature. Refer to Appendix B in the Supplementary material for further details.

Capitella polychaetes, oligochaetes and corophid amphipods; Fig. 3A). Similarly, taxa driving differences across Metals BHM groups have differing sensitivities to copper, lead and zinc (Fig. 3B). Many of the taxa characterising Metals BHM Group 1 have been found to be sensitive to metals (e.g. the shellfish *A. stutchburyi, P. australis* and *Macomona liliana*, orbinid and *Prionospio aucklandica* polychaetes, cumaceans and amphipods) while taxa more tolerant of metals (e.g. nereid and *Cossura* polychaetes, the crabs *Austrohelice*, *Hemigrapsus* and *Hemiplax* and the bivalve *Arthritica*) only begin to characterise benthic community structure in Group 3 and higher.

Both the Mud and Metals BHMs were good at predicting the position of validation sites along the environmental gradients ($R^2 = 0.90$ and 0.82, respectively), with the slope of the line close to one for both models (Fig. 4). For the temporal validation, which aimed to show that there would be no change in CAP scores if the stressor values did not change, most sites stayed within the range of an ecological health group (i.e. CAP scores within a range of 1.0), indicating that CAP scores were relatively stable and that the ecological health group boundaries are suitable (Appendix C).

Moderate correlation was observed between the CAP scores from the two models ($r = 0.76$) suggesting there is potential for interaction between the two models. However, the relationship between the two models was variable (Fig. 5) and DistLM showed that of the 13% variation in macrofaunal structure collectively explained by mud and metals, only 4.4% was shared between the two variables leaving 8.6% of variation that was independently explained by either mud or metals on their own. Furthermore, species shifts associated with changes in mud were not consistently the same as species shifts associated with changes in metals (Fig. 3). The models had reduced ability to discriminate between stressors at the higher end of the range; sites with

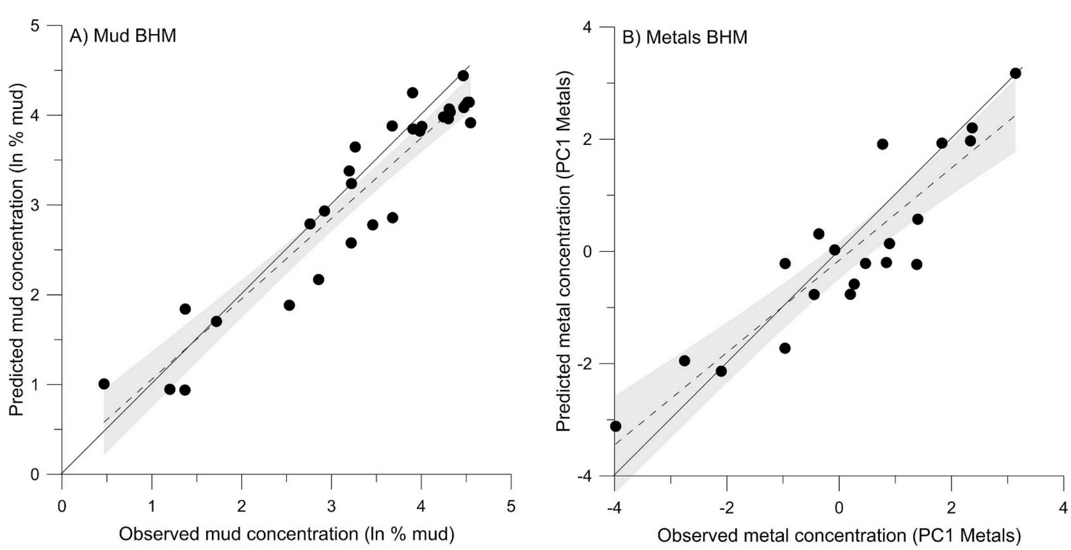

Fig. 4. Validation of Benthic Health Models (BHMs) comparing A) observed mud (ln % mud) and B) metal (first axis of principal component analysis based on log transformed copper, lead and zinc) concentrations with concentrations predicted by the BHMs on the basis of benthic macroinvertebrate community composition. The dashed line is the linear regression line (with 95% confidence interval indicated by grey shading) and the solid line has a slope of 1 and an intercept of zero (i.e. 1:1 line) and indicates where all points would lie if model predictions were perfect. Mud BHM $y = 0.8966x + 0.1614$, $R^2 = 0.90$. Metals BHM $y = 0.82x - 0.16$, $R^2 = 0.82$.

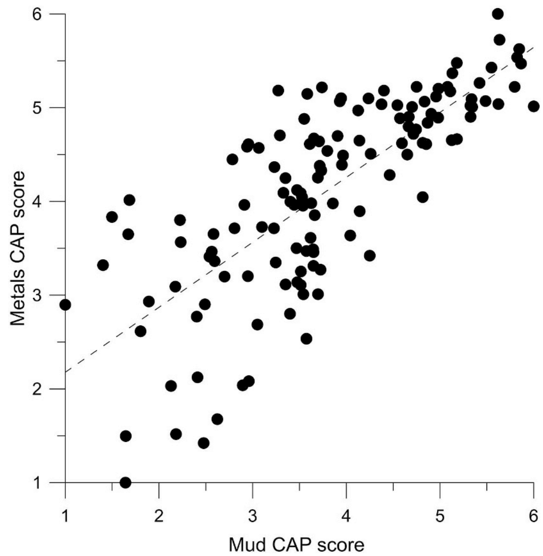

Fig. 5. Relationship between the Mud and Metals Benthic Health Model (BHM) canonical analysis of principal coordinates (CAP) scores. A linear regression (dashed line) has been fitted ($y = 0.8376x + 0.5805$, $R^2 = 0.58$).

high metal concentrations always had high mud content but this was not always the case the other way around.

3.2. *Effect of region, estuary type and scale on model results*

The results of the linear regressions indicated that both the Mud and Metals BHMs could be applied across all regions and estuary types tested. The relationship between the CAP model scores and the environmental gradients did not differ across regions for either model (mud*region $p = 0.802$; metals*region $p = 0.103$). Similarly, there was no significant interaction between estuary type and environmental gradient for either model (mud*estuary type $p = 0.647$; metals*estuary type $p = 0.067$).

Spearman's rank correlations between the national BHM CAP scores and the single region and single estuary BHM CAP scores showed that the national BHMs ranked sites in a similar way as the regional models (strong and moderate correlations; Mud BHM $r = 0.98$, Metals BHM $r = 0.76$) but the correlations with the single estuary models were not as high (moderate correlations; Mud BHM $r = 0.68$, Metals BHM $r = 0.42$). Refer to Appendix D of the Supplementary Material for more details.

4. Discussion

In this study, we successfully developed two models that track the health of estuarine benthic communities in response to two key coastal stressors; terrestrial sedimentation and heavy metal contamination. This approach to estuary health assessment has been previously applied on a regional- (Hewitt et al., 2005) and estuary-scale (Ellis et al., 2015) and here we have developed models that can be used at a national level. With the plethora of biotic indices available for monitoring (refer Borja et al., 2015; Diaz et al., 2004 for reviews), and the range of agencies responsible for coastal management, achieving consistent assessment across countries or continents can be challenging (Borja et al., 2009). Like many other countries, New Zealand does not have a standardised approach making it difficult to compare health across estuaries and set national standards. Additionally, many of the biotic indices developed overseas are not readily transferable to New Zealand due to differences in species ecology and composition, stressor type or magnitude and estuary geomorphology (Berthelsen et al., 2018a; Rodil et al., 2013). The transferability of a biotic index developed in one region to another part of the world will consistently be affected by these differences, although the development of regionally specific eco-groups may improve the performance of some indices (Gillett et al., 2015). The results of our study show that our BHMs are suitable for tracking the effects of increasing mud content and metal contamination on benthic community health in estuaries across New Zealand. The models can be applied in two widespread estuary types and across most regions. Thus, we have demonstrated the utility of the BHMs as a sensitive and standardised approach to national estuary health monitoring.

In addition to being sensitive enough to detect ecologically meaningful changes, indices must also be robust across the ecological and environmental contexts over which they will be applied (Borja and Dauer, 2008). We tested this by examining the response of the BHMs across different regions and estuary types. The BHMs responded to mud and metals in the same manner across all regions and estuary types tested, indicating these models were robust and suitable for application in many estuaries across New Zealand. The lack of regional differentiation suggests that local environmental drivers (e.g. anthropogenic activities, sediment grain-size, hydrodynamics) may be more important in structuring communities than regional species pools, which are driven by factors such as species dispersal and biogeographic history (Ricklefs, 1987). This finding is supported by other studies, which have found macrobenthic biodiversity to be influenced more by local conditions than regional ones (de Juan and Hewitt, 2011; Edgar et al., 1999). However, regional variations in benthic community structure may have also been concealed by the level of taxonomic resolution required to develop a national-scale model. The BHM approach requires a common pool of taxa and higher levels of taxonomic resolution (e.g. family vs genus/species) are often required to aggregate infrequent species into common groups or correct for inconsistencies in taxonomic resolution across source data. While reducing the number of taxonomic units can help the model perform well across a range of regions, it may obscure species-specific responses to stress, decreasing model sensitivity overall. Taxonomic resolution in our dataset was primarily constrained by inconsistencies across sites and better taxonomic standardisation could have enabled more robust models, across a wider range of regions and estuary types, to be developed.

When attempting to apply biotic indices on a nationwide scale, it is important to understand how outputs relate to assessments carried out at finer scales of resolution, as these may provide a more precise estimate of environmental status for managing specific locations and their problems. Our study showed that the national BHMs ranked the health of sites in a similar manner to models developed using regional data but may not have been as sensitive as models developed using data from a single estuary. As mentioned earlier, this decrease in sensitivity may have arisen from aggregation of taxa to higher levels of taxonomic resolution, potentially obscuring species-specific responses to stress. Additionally, the smaller stressor gradient in the single estuary model may allow it to discriminate over smaller changes in health. We tested this by creating a new national model that was restricted to the same stressor range as the single estuary model and observed an improvement in the correlation between the model health score rankings (refer to Appendix D of Supplementary Material for further details). Reduced power caused by having fewer data points for comparison may also contribute to inconsistency between model health score rankings, and this was supported by a slight decrease in concordance between the regional and national models when comparing fewer sites (refer Appendix D).

Even though the single estuary model may provide a more sensitive measure of estuary health, having a national-scale model delivers clear advantages. As BHM outputs are on a relative scale, a national-scale model enables the health of the estuary to be placed in a national context and provides consistency across the country. Having a national model also reduces the substantial costs that would be required to develop separate estuary-scale or even regional-scale models, making it possible for managers to utilize this assessment tool to evaluate any estuary for which they have appropriate macrofaunal data.

The outputs of the BHMs can be simplified into a five-category health score system, which allows managers to easily track the relative health of sites through time or identify thresholds for undesirable conditions, which may trigger management action (Rees et al., 2008). Monitoring directional/trend targets is a robust and reliable method and is largely independent of the concept of reference conditions because it only requires relative assessments of ecological quality status (Borja et al., 2012). It can indicate how a site is changing in response to an increasing pressure, even if the site was already impacted when monitoring began. The BHM ecological health groups provide an indication of the health of a site in the context of New Zealand, however, managers need to consider more than just the relative health category when setting management targets as the category boundaries do not necessary reflect ecological thresholds. Establishing type-specific reference conditions could help to define appropriate thresholds in different settings (e.g. upper or lower reaches of estuaries) and there are a range of methods available to estimate these (Barbone et al., 2012; Borja et al., 2012; EU Water Framework Directive, 2000; Stoddard et al., 2006). However, reference conditions can be difficult to define in estuaries due to their high natural variability and the scarcity of locations remaining in an undisturbed state (Barbone et al., 2012; Berthelsen et al., 2018a; Chainho et al., 2007). Further research is required to understand where community thresholds lie along different environmental gradients and in different contexts, which could inform management goals or adjustment of group boundaries in the future.

Studies have suggested estuarine sediments with less than 10–30% mud support more diverse, abundant and/or resilient benthic communities (Ellis et al., 2017; Robertson et al., 2015, 2016; Rodil et al., 2013). The boundary between Mud BHM Group 3 and 4 occurs around 18% mud and transitions to Group 5 around 50% mud. Therefore, depending on management goals, aiming for Mud BHM health scores in groups less than 4 may be appropriate. However, when interpreting Mud BHM health scores, it must be acknowledged that hydrodynamic controls on sedimentation rates may naturally result in upper reaches of estuaries being muddier than outer reaches, dependent on estuary type and the magnitude of sediment inputs. The risk of natural processes affecting the use of the Mud BHM can be alleviated in three ways. First, adjustment of thresholds or reference conditions, which consider these natural variations, can be used when setting management targets (Chainho et al., 2007). Second, sites can be selected to represent both inner and outer areas of estuaries. Third, rather than relying on one-off assessments of health, we recommend examining Mud BHM health scores over time and acting if a site is progressively decreasing in 'health' with respect to sedimentation.

Guidelines regarding acceptable levels of metal loading in coastal sediments vary (refer Burton, 2002 for a review), but many sediment quality guidelines set two threshold values, one below which effects rarely occur (threshold effects e.g. TEL, ERL, SQGV) and one above which effects are likely to occur (midrange/extreme effects e.g. PEL, ERM, SQG-High; Long et al., 1995; MacDonald et al., 1996; Simpson et al., 2013). Most threshold effect values fall within Group 4 or 5 of the Metals BHM while almost all midrange or extreme values are beyond those measured in our nationwide study (refer to Appendix E of Supplementary Material for more details). However, as observed in other studies (Hewitt et al., 2009; Tremblay et al., 2017), both these lower and upper thresholds may be too high to protect benthic communities, given we observed significant changes in community structure at lower metal concentrations. Many of these guidelines are developed from single-species, laboratory dose-response experiments with mortality as an endpoint (Calow, 1998), which do not accurately represent the complexities of coastal systems. Indeed, guidelines derived from field-based species sensitivity distributions (Bjørgesæter and Gray, 2008; Hewitt et al., 2009; Kwok et al., 2008) tend to be lower than other guidelines outlined in Appendix E, corresponding to Metals BHM Group 3 and 4.

Although single-stressor models have advantages in terms of providing objective measures of health and diagnosing the cause of degradation, interactions between stressors can confound outputs and any strongly co-varying environmental variables should be examined to ensure the model can discriminate between them. A moderate correlation was observed between the Mud and Metals BHMs, reflective of the fact that metals commonly bind to fine sediments and/or organic matter (Power and Chapman, 1992). However, consistent with previous studies (Ellis et al., 2015; Hewitt and Ellis, 2010; Thrush et al., 2008), we found the collinearity between mud content and metal concentrations was not sufficient to prevent partitioning out individual effects of these stressors on macrofaunal communities. Only 4.4% of the explained macrofaunal community variation was shared by mud and metals, suggesting that both variables are important in structuring benthic communities, with neither being a replacement for the other. The differences in taxa driving changes across the two models (Fig. 3) also supports this conclusion. However, the Metals BHM may have reduced ability to discriminate between mud and metals effects in Group 5, so we suggest the use of bivariate plots of Mud and Metals CAP scores when assessing site changes (Hewitt and Ellis, 2010). If sites are moving along only one of the two axes, effects can be attributed to that stressor, but if sites are moving in both directions, a close inspection of which species are responding to the changes may be required to ascertain the environmental driver.

Multivariate approaches to assessing health have been found to be more sensitive than univariate methods because they preserve information on all taxa and their relative abundances (Attayde and Bozelli, 1998; Ellis et al., 2015; Gray, 2000; Hewitt et al., 2005; Warwick and Clarke, 1991). However, it is precisely for this reason that the BHMs are constrained to being applied under the same conditions as the data used to develop them. Differences in species composition restrict the application of these models to intertidal portions of estuaries within New Zealand, although this does not preclude the development of new models for other environments or regions of the world. The outputs of the models appear robust across most regions and for the two estuary types tested (which represent more than half of the estuaries in New Zealand; Hume et al., 2016), but further research is needed to determine their suitability for assessing health in other estuary types. Although the incorporation of data collected across multiple months likely reduces the influence of seasonal fluctuations in species composition on model results, it is recommended that data for new sites is collected at similar seasonal time periods (in this case October to March, i.e. the time period for which most model data was collected). Our models capture the range of mud and metal concentrations likely to be encountered in most New Zealand estuaries, however, if metal values increase significantly, new sites would need to be added to the model to extend its range, affecting comparison with earlier health model scores. The BHMs provided good indicators of benthic community health at a national level in response to mud and metals, however, we advocate the use of multiple indicators to gain a more complete understanding of overall health, particularly those that represent responses to other stressors (e.g. nutrients) or the condition of other taxonomic groups (e.g. plankton, fish).

CRediT authorship contribution statement

D.E. Clark: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. **J.E. Hewitt:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Supervision. **C.A. Pilditch:** Conceptualization, Writing - review & editing, Supervision. **J.I. Ellis:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Supervision.

Declaration of competing interest

None.

Acknowledgements

We thank Massey University and the New Zealand Ministry of Business, Innovation and Employment (MBIE) for funding this work through the Oranga Taiao Oranga Tangata research programme (contract MAUX1502, led by Murray Patterson). Additional funding was provided by the Cawthron Institute's Internal Investment Fund and the National Institute of Water and Atmospheric Research's Strategic Science Investment Fund. We also acknowledge the support of New Zealand regional authorities that provided data and permission to use it: Auckland Council, Bay of Plenty Regional Council, Environment Canterbury, Environment Southland, Greater Wellington Regional Council, Hawkes Bay Regional Council, Nelson City Council, Northland Regional Council, Otago Regional Council, Tasman District Council, Waikato Regional Council and West Coast Regional Council. Anna Berthelsen, Javier Atalah and Eric Goodwin helped to compile the dataset and provided advice on statistical methods. Gretchen Rasch provided constructive comments on a draft version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2019.110602.

References

Agardy, T., Alder, J., Dayton, P., Curran, S., Kitchingman, A., Wilson, M., Catenazzi, A., Restrepo, J., Birkeland, C., Blaber, S., Saifullah, S., Branch, G., Boersma, D., Nixon, S., Dugan, P., Davidson, N., Vorosmarty, C., 2005. Chapter 19: coastal systems. In: Hassan, R., Scholes, R., Ash, N. (Eds.), Ecosystems and Human Well-Being: Current State and Trends: Findings of the Condition and Trends Working Group. Millennium Ecosystem Assessment.

Anderson, M., Hewitt, J., Thrush, S., 2002. The Development of Criteria for Assessing Community Health of Urban Intertidal Flats. pp. 54 NIWA report prepared for Auckland Regional Council Technical Report No. 184.

Anderson, M.J., 2008. Animal-sediment relationships re-visited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. J. Exp. Mar. Biol. Ecol. 366, 16–27.

Anderson, M.J., Gorley, R.N., Clarke, K.R., 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-E, Plymouth, UK, pp. 214.

Anderson, M.J., Hewitt, J.E., Ford, R.B., Thrush, S.F., 2006. Regional Models of Benthic Ecosystem Health: Predicting Gradients from Biological Data, vol. 317. Prepared by Auckland UniServices Limited for Auckland Regional Council. Auckland Regional Council Technical Publication, pp. 77 (plus appendices).

Anderson, M.J., Robinson, J., 2003. Generalized discriminant analysis based on distances. Aust. N. Z. J. Stat. 45, 301–318.

Anderson, M.J., Willis, T.J., 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology 84, 511–525.

ARC, 2004. Blueprint for Monitoring Urban Recieving Environments TP168. Auckland Regional Council, pp. 66.

Attayde, J.L., Bozelli, R.L., 1998. Assessing the indicator properties of zooplankton assemblages to disturbance gradients by canonical correspondence analysis. Can. J. Fish. Aquat. Sci. 55, 1789–1797.

Aubry, A., Elliott, M., 2006. The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: application to the Humber Estuary, UK. Mar. Pollut. Bull. 53, 175–185.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R., 2011. The value of estuarine and coastal ecosystem services. Ecol. Monogr. 81, 169–193.

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. Indicat. 12, 105–122.

Berthelsen, A., Atalah, J., Clark, D., Goodwin, E., Patterson, M., Sinner, J., 2018a. Relationships between biotic indices, multiple stressor gradients and natural variability in New Zealand estuaries. Ecol. Indicat. 85, 634–643.

Berthelsen, A., Clark, D., Goodwin, E., Atalah, J., Patterson, M., 2018b. National Estuary Dataset: User Manual. OTOT Research Report No. 5. Cawthron Report No. 3152. Massey University, Palmerston North, pp. 22 (plus appendices).

Bjørgesæter, A., Gray, J.S., 2008. Setting sediment quality guidelines: a simple yet effective method. Mar. Pollut. Bull. 57, 221–235.

Borcard, D.P., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology 73, 1045–1055.

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., Hutchings, P., Jia, X., Kenchington, R., Marques, J.C., Zhu, C., 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Mar. Pollut. Bull. 56, 1519–1537.

Borja, A., Dauer, D.M., 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. Ecol. Indicat. 8, 331–337.

Borja, Á., Dauer, D.M., Grémare, A., 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecol. Indicat. 12, 1–7.

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, Á., Franco, J., Valencia, V., Bald, J., Muxika, I., Jesús Belzunce, M.a., 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, 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.

Borja, A., Miles, A., Occhipinti-Ambrogi, A., Berg, T.J.H., 2009. Current status of macroinvertebrate methods used for assessing the quality of European marine waters: implementing the Water Framework Directive. Hydrobiologia 633, 181–196.

Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecol. Monogr. 27, 325–349.

Bremner, J., Rogers, S.I., Frid, C.L.J., 2003. Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. Mar. Ecol. Prog. Ser. 254 11-2003.

Bricker, S.B., Longstaff, B., Dennison, W.C., Jones, A., Boicourt, K., Wicks, C., Woerner, J., 2008. Effects of nutrient enrichment in the nation's estuaries: a decade of change. Harmful Algae 8, 21–32.

Brierley, A.S., Kingsford, M.J., 2009. Impacts of climate change on marine organisms and ecosystems. Curr. Biol. 19, R602–R614.

Burton, J., Allen, G., 2002. Sediment quality criteria in use around the world. Limnology 3, 65–76.

Calow, P.P., 1998. Handbook of Environmental Risk Assessment and Management. (Blackwell).

Chainho, P., Costa, J.L., Chaves, M.L., Dauer, D.M., Costa, M.J., 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.

Chao, A., Chazdon, R.,L., Colwell, R.K., Shen, T., 2005. A statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett. 8, 148–159.

Clarke, K.N., Gorley, R.N., 2015. PRIMER V7: User Manual/Tutorial. PRIMER-E, Plymouth, pp. 296.

Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. Aust. J. Ecol. 18, 117–143.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van der Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260.

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.

Darling, E.S., Côté, I.M., 2008. Quantifying the evidence for ecological synergies. Ecol. Lett. 11, 1278–1286.

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

de Juan, S., Hewitt, J., 2011. Relative importance of local biotic and environmental factors versus regional factors in driving macrobenthic species richness in intertidal areas. Mar. Ecol. Prog. Ser. 423, 117–129.

deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R.I., Scheffer, M., Werner, F., 2008. Regime shifts in marine ecosystems: detection, prediction and management. Trends Ecol. Evol. 23, 402–409.

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

Edgar, G.J., Barrett, N.S., Last, P.R., 1999. The Distribution of Macroinvertebrates and Fishes in Tasmanian Estuaries, vol. 26. pp. 1169–1189.

Ellis, J.I., Clark, D., Atalah, J., Jiang, W., Taiapa, C., Patterson, M., Sinner, J., Hewitt, J., 2017. Multiple stressor effects on marine infauna: responses of estuarine taxa and functional traits to sedimentation, nutrient and metal loading. Sci. Rep. 7.

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.

EU Marine Strategy Framework Directive, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive) European Parliament. Council of the European Union.

EU Water Framework Directive, 2000. Directive 2000/60/EC of the European Parliament and of the Council Establishing a Framework for the Community Action in the Field of Water Policy.

Flåten, G.R., Botnen, H., Grung, B., Kvalheim, O.M., 2007. Quantifying disturbances in benthic communities—comparison of the community disturbance index (CDI) to other multivariate methods. Ecol. Indicat. 7, 254–276.

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. Indicat. 50, 99–107.

Grall, J., Glémarec, M., 1997. Using biotic indices to estimate macrobenthic community

perturbations in the Bay of Brest. Estuarine. Coastal and Shelf Science 44, 43–53.
Gray, J.S., 2000. The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. J. Exp. Mar. Biol. Ecol. 250, 23–49.
Gray, J.S., Clarke, K.R., Warwick, R.M., Hobbs, G., 1990. Etection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields. North Sea Marine Ecology Progress Series 66, 285–299.
Grosholz, E., 2002. Ecological and evolutionary consequences of coastal invasions. Trends Ecol. Evol. 17, 22–27.
Hewitt, J., 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., Anderson, M.J., Thrush, S.F., 2005. Assessing and monitoring ecological community health in marine systems. Ecol. Appl. 15, 942–953.
Hewitt, J.E., Bell, R., Costello, M.J., Cummings, V., Currie, K., Ellis, J., Francis, M., Froude, V., Gorman, R., Hall, J., Inglis, G., MacDiarmid, A., Mills, G., Pinkerton, M., Schiel, D.R., Swales, A., Law, C.S., McBride, G.B., Nodder, S., Rowden, A., Smith, M., Thompson, D., Torres, L., Tuck, I., Wing, S.R., 2014. Development of a national marine environmental monitoring programme (MEMP) for New Zealand. In: Ministry for Primary Industries New Zealand Aquatic Environment and Biodiversity Report, vol. 141. pp. 112 (plus appendices).
Hewitt, J.E., Ellis, J., 2010. Assesment of the Benthic Health Model. Auckland Regional Council TR2010/034, pp. 26 (plus appendices).
Hewitt, J.E., Thrush, S.F., Dayton, P.D., 2008. Habitat variation, species diversity and ecological functioning in a marine system. J. Exp. Mar. Biol. Ecol. 366, 116–122.
Hume, T., Gerbeaux, P., Hart, D., Kettles, H., Neale, D., 2016. A Classification of New Zealand's Coastal Hydrosystems. Prepared for Ministry for the Environment, pp. 65 (plus appendices).
Johnston, E.L., Mayer-Pinto, M., Crowe, T.P., 2015. REVIEW: chemical contaminant effects on marine ecosystem functioning. J. Appl. Ecol. 52, 140–149.
Keeley, N.B., Macleod, C.K., Forrest, B.M., 2012. Combining best professional judgement and quantile regression splines to improve characterisation of macrofaunal responses to enrichment. Ecol. Indicat. 12, 154–166.
Kwok, K.W.H., Bjorgesæter, A., Leung, K.M.Y., Lui, G.C.S., Gray, J.S., Shin, P.K.S., Lam, P.K.S., 2008. Deriving site-specific sediment quality guidelines for Hong Kong marine environments using field-based species sensitivity distributions. Environ. Toxicol. Chem. 27, 226–234.
Lohrer, A.M., Townsend, M., Rodil, I.F., Hewitt, J.E., Thrush, S.F., 2012. Detecting shifts in ecosystem functioning: the decoupling of fundamental relationships with increased pollutant stress on sandflats. Mar. Pollut. Bull. 64, 2761–2769.
Long, E.R., MacDonald, D.D., Smith, S.L., Calder, F.D., 1995. Incidence of adverse biological effects within range of chemical concentrations in marine and estuarine sediments. Environ. Manag. 19, 81–97.
Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., Jackson, J.B.C., 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. Science 312, 1806–1809.
MacDiarmid, A., McKenzie, A., Sturman, J., Beaumont, J., Mikaloff-Fletcher, S., Dunne, J., 2012. Assessment of anthropogenic threats to New Zealand marine habitats. New Zealand Aquatic Environment and Biodiversity Report No 93, 255.
MacDonald, D.D., Carr, R.S., Calder, F.D., Long, E.R., Ingersoll, C.G., 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. Ecotoxicology 5, 253–278.
Magni, P., Hyland, J., Manzella, G., Rumohr, H., Viaroli, P., Zenetos, A., 2005. In: Proceedings of the Workshop "Indicators of Stress in the Marine Benthos", vols. 8–9. UNESCO/IOC, IMC, Torregrande-Oristano, Italy, pp. 46 October 2004.
Magris, R.A., Ban, N.C., 2019. A meta-analysis reveals global patterns of sediment effects on marine biodiversity. Glob. Ecol. Biogeogr. 0.
Margalef, R., 1958. Information theory in ecology. General Systems 3, 36–71.
Martinez-Crego, B., Alcoverro, T., Romero, J., 2010. Biotic indices for assessing the status of coastal waters: a review of strengths and weaknesses. J. Environ. Monit. 12, 1013–1028.
Niemi, G., Wardrop, D., Brooks, R., Anderson, S., Brady, V., Paerl, H., Rakocinski, C., Brouwer, M., Levinson, B., McDonald, M., 2004. Rationale for a new generation of indicators for coastal waters. Environ. Health Perspect. 112, 979–986.
Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution in the marine environment. Oceanogr. Mar. Biol. Annu. Rev. 16, 229–311.
Pielou, E.C., 1966. The measurement of diversity in different types of biological collections. J. Theor. Biol. 13, 131–144.
Piló, D., Carvalho, A.N., Pereira, F., Coelho, H.E., Gaspar, M.B., 2019. Evaluation of macrobenthic community responses to dredging through a multimetric approach: effective or apparent recovery? Ecol. Indicat. 96, 656–668.
Power, E.A., Chapman, P.M., 1992. Assessing Sediment Quality. Lewis, Chelsea.
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.
Ricklefs, R.E., 1987. Community Diversity: Relative Roles of Local and Regional Processes. Science 235, 167–171.
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, vol. 5096. Prepared for Supporting Councils and the Ministry for the Environment, Sustainable Management Fund Contract, pp. 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. Indicat. 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. Indicat. 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. Indicat. 34, 560–570.
Shade, A., 2016. Diversity is the question, not the answer. ISME J. 11, 1.
Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423.
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.
Simboura, N., Zenetos, A., 2002. Benthic indicators to use in ecological quality classifications of Mediterranean soft bottom marine ecosystems, including a new biotic index. Mediterr. Mar. Sci. 3, 77–111.
Simpson, S.L., Batley, G.E., Chariton, A.A., 2013. Revision of the ANZECC/ARMCANZ Sediment Quality Guidelines CSIRO Land and Water Science Report 08/07. CSIRO Land and Water.
Smith, R.W., Bergen, M., Weisberg, S.B., Cadien, D., Dalkey, A., Montagne, D., Stull, J.K., Velarde, R.G., 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. Ecol. Appl. 11, 1073–1087.
Stoddard, J., P Larsen, D., Hawkins, C., Johnson, R., H Norris, R., 2006. Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition.
Thrush, S.F., Hewitt, J.E., Cummings, V.J., Dayton, P.K., Cryer, M., Turner, S.J., Funnell, G.A., Budd, R.G., Milburn, C.J., Wilkinson, M.R., 1998. Disturbance of the marine benthic habitat by commercial fishing: impacts at the scale of the fishery. Ecol. Appl. 8, 866–879.
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.
Thrush, S.F., Hewitt, J.E., Hickey, C.W., Kelly, S., 2008. Multiple stressor effects from species abundance distributions: interactions between urban contaminants and species habitat relationships. J. Exp. Mar. Biol. Ecol. 366, 160–168.
Thrush, S.F., Hewitt, J.E., Norkko, A., Nicholls, P.E., Funnell, G.A., Ellis, J.I., 2003. Habitat change in estuaries: predicting broad-scale responses of intertidal macrofauna to sediment mud content. Mar. Ecol. Prog. Ser. 263, 101–112.
Thrush, S.F., Lundquist, C.J., Hewitt, J.E., 2005. Spatial and temporal scales of disturbance to the seafloor: a generalized framework for active habitat management. In: Barnes, P.W., Thomas, J.P. (Eds.), Benthic Habitats and the Effects of Fishing. American Fisheries Society, Symposium Series, Bethesda, Maryland, pp. 639–649.
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.
US EPA, 1994. Methods for the Determination of Metals in Environmental Samples: EPA 200 Series, Supplement I. U. S. Environmental Protection Agency EPA- 600/R-94/111, Revision 2.8, May 1994.
Warwick, R.M., Clarke, K.R., 1991. A Comparison of some methods for analysing changes in benthic community structure. J. Mar. Biol. Assoc. U. K. 71, 225–244.
WoRMS Editorial Board, 2017. World register of marine species. Available from http://www.marinespecies.org .
Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3–14.