Running head: *benthic community response to organic enrichment*

Title

Response of a marine benthic invertebrate community and biotic indices to organic enrichment from sewage disposal

Fiona E. Culhane[1,2,*], Robert A. Briers[2], Paul Tett[3], Teresa F. Fernandes[4]

[1] University of Liverpool, School of Environmental Sciences, Nicholson Building, Liverpool, L69 3GP, UK

[2] Edinburgh Napier University, School of Applied Sciences, Sighthill Court, Edinburgh, EH11 4BN, UK

[3] SAMS, Scottish Marine Institute, Oban, PA37 1QA, UK

[4] Heriot-Watt University, Institute of Life and Earth Sciences, Edinburgh, EH14 4AS, UK

*Corresponding author: Email: F.Culhane@liverpool.ac.uk Tel. +44 (0) 151 795 4646

Abstract

Nutrient enrichment is a significant cause of ecosystem change in coastal habitats worldwide. This study focuses on the change in a benthic macroinvertebrate community and environmental quality as assessed through different biotic indices following the construction of a sewage outfall pipe in the west of Scotland, from first implementation to seven years after operation of the pipe. Benthic macroinvertebrates are an important part of marine ecosystems because they mediate ecosystem processes and functions, are a key part of food webs and they provide many ecosystem services. Results indicated a clear change in benthic communities over time with an increase in species richness and changes to benthic community composition (specifically feeding type, bioturbation mode and ecological group) towards those indicative of organic enrichment. No clear spatial zonation was observed because organic carbon content increased over the entire area. According to a suite of benthic indices calculated, some negative changes were detectable following the start of sewage disposal, but largely negative community changes, and a change from 'good' to 'moderate' quality, only occurred seven years after implementation. The increase in species richness in response to increasing disturbance reduced the utility of a multi-metric index, the Infaunal Quality Index, which, instead of amplifying the signal of negative impact, dampened it. We suggest that any change in communities, regardless of direction, should be heeded, and species richness is a particularly sensitive and early warning indicator for this, but a suite of approaches is required to understand benthic community changes.

Key words
Biotic index; Multi-metric index; Benthic macroinvertebrates; Organic enrichment; Pearson-Rosenberg model; Intermediate disturbance hypothesis

INTRODUCTION

Nutrient enrichment is recognised worldwide as one of the most important causes of change in aquatic ecosystems and one of the leading sources of nutrients comes from sewage discharge, with other significant inputs coming from diffuse, agriculturally derived run-off (Diaz & Rosenberg, 2008, Smith & Schindler, 2009, WWAP, 2017). With sewage discharge can come not only nutrients, but a suite of other contaminants, such as heavy metals, changes to salinity and temperature and emerging pollutants such as microplastics (Foteinis et al., 2013, Li et al., 2018, Smith-Evans & Dawes, 1996). Around the world, sewage that is untreated or subject to minimal treatment continues to be released into coastal areas (Baum et al., 2013, WWAP, 2017). In Europe, even where there are strict regulations on the discharge of sewage, it is still an issue in many areas due to outdated sewage systems or a lag in infrastructure developments achieving the required capacity to match increasing urban populations (Kiedrzyńska et al., 2014). In this sense, coastal habitats are intentionally or unintentionally used to treat waste (Watson et al., 2016). The biota occupying benthic habitats, including the benthic macroinvertebrates, are known to be able to process waste, providing this waste treatment ecosystem service (Watson et al., 2016). However, input of waste into coastal areas can cause changes to benthic communities (e.g. Abdelrhman & Cicchetti, 2012, Bowen & Valiela, 2001, Caswell et al., 2018, Diaz & Rosenberg, 2008, Stull et al., 1986) . Eutrophication with deoxygenation due to organic enrichment, can lead to lower benthic species richness; degraded ecosystem functioning, such as bioturbation and nutrient cyclin; and decreased capacity to provide ecosystem services, including waste treatment, and climate regulation through reduced carbon sequestration (Caswell et al., 2018, Smith & Schindler, 2009, Worm et al., 2006). Changes to benthic invertebrate communities, such as loss of larger species, can have impacts on the wider ecosystem, since they are a key part of the food chain, in particular on predators with specialised feeding strategies, such as some wading birds (Bowgen et al., 2015).

Benthic macroinvertebrates are relatively sessile and long lived; they cannot avoid unfavourable conditions, and thus integrate changes of conditions over time thereby reflecting environmental conditions (Reiss & Kröncke, 2005). Where organic enrichment is most severe, anoxia (no oxygen), hypoxia (low oxygen) and hydrogen sulphite toxicity can lead to severe impacts on benthic communities, from a complete loss of fauna, to changes in organism functioning, reduced biodiversity and altered biomass (see Diaz & Rosenberg, 1995 for a review). Further away from the pollution source, a range other impacts can occur. The current understanding of macrofaunal benthic community responses to organic enrichment stems from the Pearson & Rosenberg (1978) theory. This describes a succession of macrofauna from a total lack of species at the contamination source (where pollution is severe), moving to high abundances of a few opportunistic species with little or no sediment bioturbation, and succeeding gradually with time and/or distance from the pollution source to greater species richness, larger species, lower abundances and increasingly complex sediment burrowing structures. Thus, in impacted benthic communities we can expect to see structural changes, like differences in species richness, abundance and biomass. In addition, we can expect to see functional changes, like changes in bioturbation rates, feeding types (Word, 1979) and different proportions of species with pollution tolerant or sensitive traits (Borja et al., 2000). This stereotyped response to organic enrichment (including from sewage, pulp and paper mill waste and organic dredged sediments, amongst others) allows detection of environmental impacts; this has been exploited in studies of human impacts on marine ecosystems and through the use of indices of benthic ecosystem health (e.g. Borja et al., 2011, Caswell et al., 2018, Elliott & Quintino, 2007, Muxika et al., 2005).

However, other work has shown that responses can vary from this paradigm. Species richness may show a range of responses to stress or resource availability including the humpbacked curve described by the intermediate-disturbance hypothesis, which shows an increase in diversity with increasing stress or resource availability before decreasing again as stress continues to increase (Connell, 1978, Dodson et al., 2000, Hooper et al., 2005, Huston, 2014, Mittelbach et al., 2001, Odum, 1985). This pattern has also been shown for the response to nitrogen loadings in functioning of benthic invertebrate communities, for example in bioturbation rates (Abdelrhman & Cicchetti, 2012). In some cases, nutrient enrichment can increase species richness and productivity through increased survival and recruitment, by making more resources available and mitigating the effects of other stressors such as heavy metal contaminants (Lawes et al., 2016). Other work has shown other factors influence the response of benthic communities. For example, at sites with high current speeds, high diversity and abundance can be maintained in the presence of an organic pollution source, when conditions allow for both maintained aerobic sediment conditions with an increase in food supply (Keeley et al., 2013). Current speed can influence impacts on benthic environments, either through dispersing material so it cannot accumulate; through determining the coarseness of the seabed, affecting the accumulation of finer organic material; or through determining the original benthic community which can exist there, and how sensitive it is to organic enrichment (Macleod et al., 2007, Mayor & Solan, 2011, Rees et al., 2006, Snelgrove, 1999).

In the assessment of the quality of marine benthic invertebrate communities, indices are widely used to summarise complex information into a form which is easy to interpret by a range of users. Marine benthic invertebrate community based indices include the Infaunal Trophic Index (ITI), which is based on feeding guilds (Word, 1979), and relates to evidence showing that nutrient enriched communities will functionally change by favouring surface deposit feeding over suspension feeding (Aberson et al., 2016). Other indices commonly used, amongst others, are the Azti Marine Biotic Index (AMBI) (Borja et al., 2000) and the Benthic Opportunistic Polychaete to Amphipod Ratio (BOPA) (Dauvin & Ruellet, 2007), which are based on their sensitivity or tolerance to disturbance (see Table 1 for a selection of some of the indices used). These indices summarise multivariate data into an easily understood score of environmental quality (Diaz et al., 2004). In an environmental context, indices are integral to most approaches to quality assessment that aim to safeguard ecological integrity (Borja et al., 2008, Borja et al., 2009a). In Europe, the Water Framework Directive (WFD) requires the achievement of at least good ecological status (GES) in transitional and coastal waters (EC, 2000). Indices are routinely used to assess environmental quality within the framework of this Directive (van Loon et al., 2015), therefore it is essential that these indices are effective and sensitive. There are important environmental, legislative and financial implications for policy implementation when the indices used in routine monitoring over-estimate the quality of poor areas or under-estimate quality of good areas (Quintino et al., 2006). Discrepancies and inconsistencies between indices lead to a lack of confidence in quality assessments (Quintino et al., 2006). Thus, it is recommended to use more than one index, though this can then increase the complexity of routine monitoring (Kröncke & Reiss, 2010).

A development in the use of these indices has been driven by the WFD where GES of benthic invertebrate communities is defined by the diversity, the abundance and the invertebrate taxa sensitive to disturbance (EC, 2000). This has led to the development and use of multi-metric indices, which can incorporate these different aspects of benthic ecological status explicitly. Multi-metric indices allow different aspects of biodiversity and community organisation to be integrated into a single value and should magnify a common signal in the combined metrics (Schoolmaster et al., 2012). This approach was adopted in different European countries to fulfil reporting for the WFD. For example, in the UK, routine monitoring makes use of the Infaunal Quality Index (IQI) (Phillips et al., 2014, WFD-UKTAG, 2008), which includes Simpsons Index of evenness, AMBI and species richness, while monitoring in the Netherlands uses the Benthic Ecosystem Quality Index 2 (van Loon et al., 2015). The underlying theory of different indices dictates a range of expected responses of indices to the effect of organic enrichment (see Table 1 for a description of these responses in a selection of indices). The aim of this study was to describe the response of a benthic community to known input of organic enrichment from a sewage outfall pipe over the course of several years, and to explore how a suite of indices measured this response. It was expected that, overall, species richness would decrease and that indices would detect a decrease in quality over time, as waste and contaminants accumulated, and that there would be spatial differences with distance from the source of organic enrichment.

Table 1 A selection of indices used in the assessment of marine environmental health assessment and calculated in this study with their expected responses to organic enrichment

INDEX	FORMULA	Expected Response to Organic Enrichment
Species richness (S)	Number of species	Although varied responses of species richness to stress have been found, in general, it is expected to decrease with increasing organic enrichment according to the Pearson-Rosenberg model (1978) (but see text).
Abundance (N)	Number of individuals	Overall abundance (made up of opportunistic species) is expected to increase with increasing organic enrichment, except in cases of severe enrichment where it is expected to decrease (Pearson & Rosenberg, 1978).
Pielou's evenness index (J') (Magurran, 2004)	$J' = H'/log(s)$ *…where H' is the Shannon Weiner Index,* $H' = -\sum_i pi(lnpi)$ *...where pi is the proportion of individuals of species i in the total abundance*	Evenness is expected to decrease with increasing enrichment as the proportion of a few tolerant species increases while the proportion of other species decrease.
A/S (Quintino et al., 2006)	Ratio of abundance to species richness (A/S)	A/S has an inverse relationship with quality, and therefore is expected to increase with increasing enrichment as the proportion of a few tolerant species increases while the proportion of other species decrease
Taxonomic Distinctness (TD, Δ*) (Clarke and Warwick in Magurran, 2004)[a]	$\Delta^* = [\sum\sum_{i<j} \omega_{ij} x_i x_j] / [\sum\sum_{i<j} x_i x_j]$ *…where* ω_{ij} *is the taxonomic distances through the classification tree between every pair of species (the first from species i and the second from species j), and the double summation ranges over all pairs i and j of these species (i < j)*	Taxonomic distinctness is expected to decrease with increasing organic enrichment as the community becomes more homogenous
Azti Marine Biotic Index (AMBI) (Borja et al., 2000)	Biotic coefficient (AMBI) = [(0 x %GI) + (1.5 x %GII) + (3 x %GIII) + (4.5 x %GIII) + (6 x %GV)]/100 *…where GI-GV refer to groups of species with different tolerances to organic pollution. GI are the most pollution sensitive and GV the most tolerant*	AMBI has an inverse relationship with quality and is expected to increase with increasing organic enrichment as the proportions of sensitive species to tolerant species change.

INDEX	FORMULA	Expected Response to Organic Enrichment
Infaunal Quality Index (IQI) (Version 4)(Phillips et al., 2014)[b]	$IQI = \frac{\left(\left(0.38 \times \left(\frac{1-AMBI/7}{(1-AMBI/7)_{max}}\right)\right) + \left(0.08 \times \left(\frac{1-\lambda'}{1-\lambda'_{max}}\right)\right)\right) + \left(0.54 \times \left(\frac{S^{0.1}}{S^{0.1}_{max}}\right) - 0.4\right)}{0.6}$	IQI is expected to decrease as organic enrichment increases as the proportion of pollution tolerant species increase, the evenness decreases and species richness decreases.
Infaunal Trophic Index (ITI) (Word, 1979)	$ITI = 100 - \left[33 - 1/3\left(\frac{0n_1 + 1n_2 + 2n_3 + 3n_4}{n_1 + n_2 + n_3 + n_4}\right)\right]$ *...where n_i is the number of individuals in Functional Feeding Group i* *Groups: 1 = suspension feeders, 2 = surface detritus feeders, 3 = surface deposit feeders, 4 = sub-surface deposit feeders*	ITI is expected to decrease with increasing organic enrichment as the proportion of functional feeding groups changes from being dominated by mainly sub-surface deposit feeders in degraded conditions to being dominated by mainly suspension feeders in reference conditions.
Benthic Opportunistic Polychaete Amphipod Index (BOPA) (Dauvin & Ruellet, 2007)	$BOPA = \log_{10}\left[\left(\frac{f_P}{(f_A + 1)}\right) + 1\right]$ *…where f_P is the frequency of opportunistic polychaetes and f_A is the frequency of amphipods*	BOPA has an inverse relationship with quality and is expected to increase with increasing organic enrichment as the proportions of sensitive amphipods to opportunistic polychaetes changes.
Benthic Quality Index (BQI) (Rosenberg et al., 2004)	$BQI = \left(\sum_{i=0}^{n}\left(\frac{A_i}{totA} \times ES50_{0.05i}\right)\right) \times log_{10}(S + 1)$ *...where Ai is the abundance of species i* *ES50 $_{0.05}$ is the ES50 (expected number of species in 50 individuals) at 5% of the population of species i*	BQI is expected to decrease as organic enrichment increases, as the proportion of pollution tolerant species increases and species richness decreases.

[a] For Taxonomic Distinctness, classification tree is constructed using phylogenetic distance

[b] For IQI, 1-λ' is the Simpson's Index of evenness. The max values are the expected maximum or reference values for a given habitat, in this study: fine sands or muds where 1-AMBI/7 $_{max}$ = 0.96; 1-λ' $_{max}$ = 0.97; S $_{max}$ = 68.

METHODS

Study site

Ironotter Point is located at Greenock in the west of Scotland (55°5833' N, 4°4840' W) (Figure 1 (a) and (b)) (O'Reilly et al., 1997, SEPA, 1996). Data used were collected by the Scottish Environment Protection Agency (SEPA), to assess the impact of a sewage outfall pipe. A 14km sewer to be discharged through a sea outfall 1.2km offshore at a depth of 25m, was commissioned for the point in 1991. This discharged waste from a primary sewage treatment plant. A baseline subtidal benthic survey was carried out in 1989, one year after the pipe was laid, but before it became operational, and surveys after the pipe became operational were carried out in 1992, 1995 and 1998 (O'Reilly et al., 1997, SEPA, 1996). Initially the pipe received waste from a population of around 20,000 people and this was phased up to around 88,000.

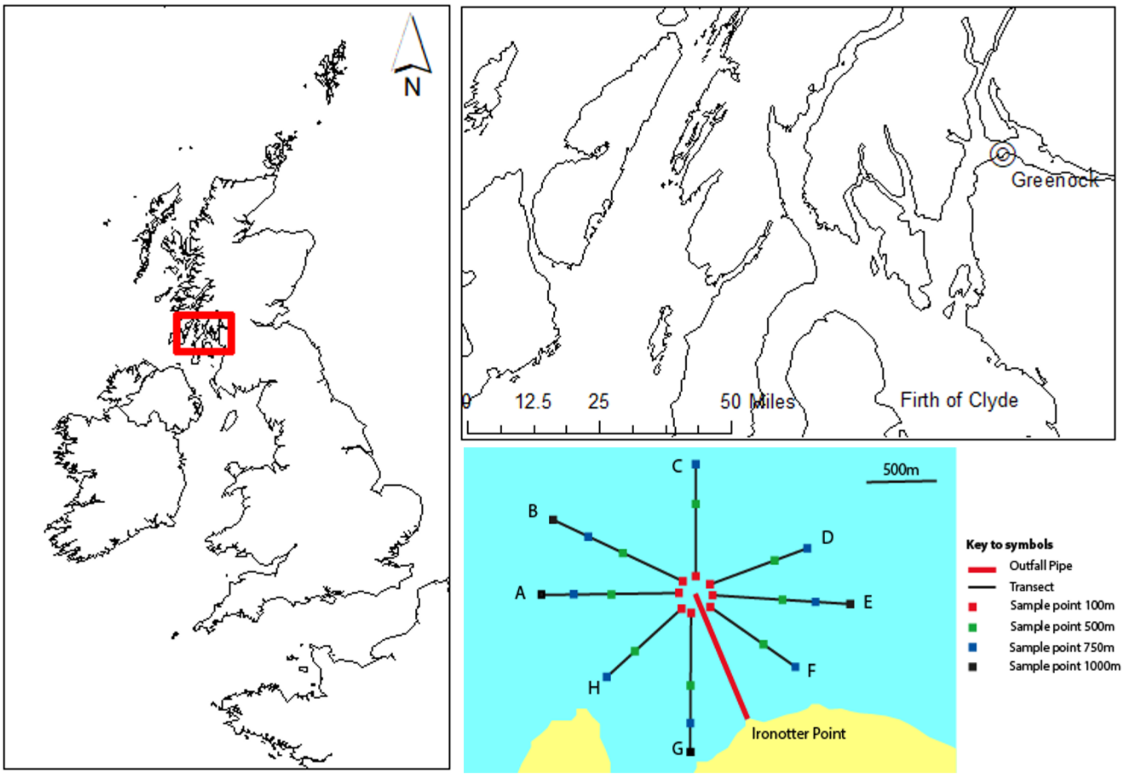

Fig. 1 Study area in Scotland, UK (a) and the location of the sea sewage outfall at Ironotter Point in the Firth of Clyde, off Greenock (b). Sampling stations were located along transects, radiating out from a distance of 100-1000m from the outfall point (c) ((a) and (b) were produced using ArcGIS version 10.6 (http://www.arcgis.com/features) and (c) was adapted from SEPA summary report (SEPA, 1996) and O'Reilly et al. 1997);, see Table S1 for full details.

Samples (0.1 m^2 Day Grab) of benthic macroinvertebrates were collected along a series of transects (Figure 1 (c)). Eight transects, which radiated out from the discharge point were used to sample, with two or three replicate samples taken at each sampling station. Stations were located 100, 500, 750 and 1000m from the discharge and were of similar depths (Figure 1 (c); Table S1, supporting information). Samples were sieved (500 µm) and identified to species level, where possible. The team of taxonomists did not change over the course of the surveys and had a common leader throughout, therefore bias in species identification by different taxonomists is not assumed to be a factor in this study. Due to the duration covered by the dataset, species names were checked and harmonised for synonyms. Organic carbon content (loss on ignition, Byers et al. (1978)) was determined from one sample of the top 5cm of sediment taken at each location. Data available were from 1989, 1992, 1995 and 1998 for macroinvertebrates and 1989, 1992 and 1995 for organic carbon. There was a change in sampling regime in 1998, due to a change in priorities and distribution of funding at this time, resulting in a minimal survey being carried out. For this year, only two transects were sampled, and organic carbon content was not measured (see Table S1).

Numerical analysis

Variation in organic carbon was analysed in relation to year, transect and distance to characterise patterns of variation before assessing the relationships between this and the values of the indices calculated. This was analysed with a factorial general linear model containing Transect, Distance from outfall and Year as fixed effects, carried out using R statistical software (R Development Core Team, 2016). An alternative model was also fitted with the inclusion of a spatial error term to account for potential spatial autocorrelation effects. There was no significant difference in the fit of the two models so the original model was kept (see Supplementary Material). Statistical significance was taken at $p<0.05$ for this and all subsequent analyses.

In order to assess biological changes in the community related to the operation of the sewage outfall pipe, the benthic community was described in a number of ways. We first analysed untransformed data of species abundance composition using SIMPER (similarity percentages), multidimensional scaling (MDS), and ANOSIM (analysis of similarity)), based on Bray-Curtis similarity, carried out using Primer 6. Community similarity was tested with respect to year of sampling and intensity of pollution, defined as discrete categories of percentage organic carbon content, increasing at regular intervals, across the range of values in the dataset. These categories were 0 – 2.5%, >2.5 – 5%, >5 – 7.5%, >7.5 – 10%, >10 – 12.5%, >12.5 – 15% organic carbon.

The total number of species found in each year and the proportion of species unique to each year are shown, though this is purely descriptive since there were differences in the number of samples and replicates taken in different years (see Table S1). The five species most responsible for generating observed patterns in each year were identified using SIMPER analysis (as above). Their ecological group (from AMBI Groups I-IV, Table 1); functional feeding group (from ITI feeding groups, Table 1); and their bioturbation mode (diffusive mixing, surface deposition, upward conveyer or downward conveyer) (classification and information from Bolam et al. (2017), MarLIN (2006), Solan et al. (2004)), were identified.

A set of ten indices was calculated for each individual sample (Table 1). Each replicate sample was treated separately in all subsequent analyses, (i.e. not pooled), to account for differences in the number of replicates taken in some cases (if replicates were pooled before calculating indices, results could be influenced by species richness, which could differ according to differences in sampling effort (Magurran, 2004)). In this way, each index was calculated on a comparable sample. Indices calculated were species richness (S), total abundance (N), measures of evenness (Pielou's evenness index and Abundance/Species richness) and taxonomic distinctness, all calculated using Primer 6 software. Five further indices calculated have associated, pre-defined quality classifications. These were AMBI, IQI, ITI, BOPA and BQI (index quality categories used are presented in Table S2, shown in supporting information). AMBI was calculated using AMBI software (http://ambi.azti.es/).

Spearman rank correlation was carried out in order to relate index results to the year of sampling, distance from the outfall and percentage organic carbon content, so as to explore any relationships between these and index values. In order to remove the effect of the confounding variable (year or organic carbon content), partial correlations were used. Only the magnitude of the Spearman rho coefficients is given and no associated p-values in order to avoid Type I errors due to multiple comparisons. The mean of each index for each year overall was also determined to explore changes in indices over time.

For the five indices (IQI, BQI, AMBI, BOPA and ITI) that have associated quality classifications, we investigated consistency between classifications to assess whether they performed differently in their ability to detect changes. Five quality categories derived from the WFD were used for IQI, BQI, AMBI and BOPA (see Table S2). In order of decreasing quality, these are 'high', 'good', moderate', 'poor' and 'bad'. ITI has four quality categories, these are 'reference', 'normal', 'changed' and ‘degraded’. For this study, these were assumed to correspond to the five WFD categories as follows: ‘reference’ = ‘high’; ‘normal’ = ‘good’; ‘changed’ = ‘moderate’ and ‘poor’; and ‘degraded’ = ‘bad’. A quality classification was determined for each individual replicate and for the mean of each station in each year. The quality classifications of the individual replicates were determined to be (1) in agreement (where all indices agree e.g. all indices assign a ‘high’ or ‘reference’ classification to a sample), (2) be ‘similar’ (where two quality classifications are given but adjacent on the scale of quality e.g. where all indices assign either a ‘good’ or a ‘moderate’ classification), or (3) ‘disagree’ (where three or more quality classifications are given, or two classifications are two levels apart on the scale of quality e.g. where one index assigns a ‘poor’ classification and another assigns a ‘good’ classification). Multidimensional scaling (MDS) based on Bray Curtis similarity was conducted on the samples in the same way as previously described. However, in this case, data were tested using ANOSIM in respect to these three levels of agreement, i.e. whether sample classifications across all years agreed, were similar or disagreed. This was related to the previous MDS plots to highlight whether the level of agreement corresponded to community compositions which were associated with a particular year or level of organic carbon content.

RESULTS

Organic carbon increased over time from an overall average of 1.2% in 1989 and 1.8% in 1992 to 8% in 1995, where the minimum found at any point in 1995 (min 3.78% in 1995) was greater than the maximum found at any point in 1989 or 1992 (max 3.73% in 1992) (Figure 2). Enrichment increased significantly with distance over time (Table 2) and this was particularly evident in 1995, with the greatest levels of organic enrichment being furthest from the pollution source (Figure 3). This suggests the effluent was well dispersed over the study area. If data had been available for 1998, the effect of distance may have been even greater but this finding of increased organic content with distance from the outfall suggests that greatest enrichment may not be at the sites closest to the outfall, as would be expected. Subsequent analyses thus focussed mainly on the effects of year of sampling and organic carbon content, and not on distance (see also supporting information, Figures S1). The silt/clay fraction at the sites increased from an average of 1.64% in 1989 (range 0.37-3.96), to 2.38% in 1992 (range 0.74-4.89). Sediment grain size properties were not available for subsequent years.

Table 2 Linear model summary of the effects of distance from sewage outfall, year of survey and transect on organic carbon content at Ironotter Point.

	Organic Carbon		
	df	**F**	**p**
Distance	8	2.56	**0.0234**
Year	1	133.25	**<0.0001**
Transect	14	0.76	0.7010
Distance*Year	1	12.83	**0.0009**
Distance*Transect	7	0.25	0.9693
Year*Transect	7	0.45	0.8628
Distance*Year*Transect	7	0.12	0.9962

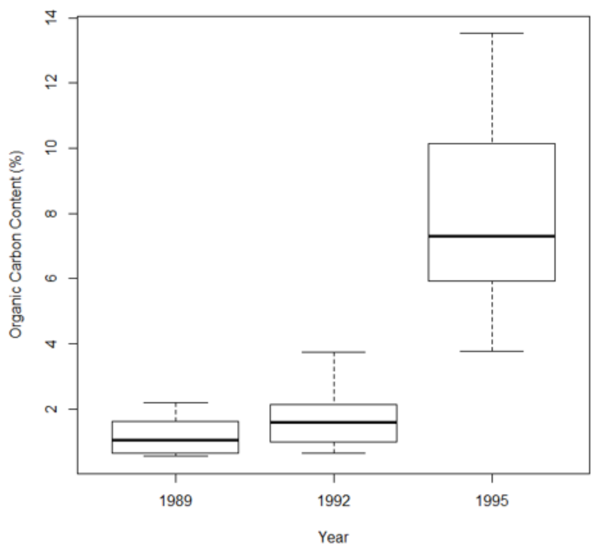

Fig. 2 Median, interquartile range and minimum and maximum values of percentage organic carbon content across all sediment samples taken at a sea sewage outfall at Ironotter Point (Firth of Clyde, UK) in each surveyed year (1989 (pre-disposal) n=22, 1992 n=22, 1995 n=28).

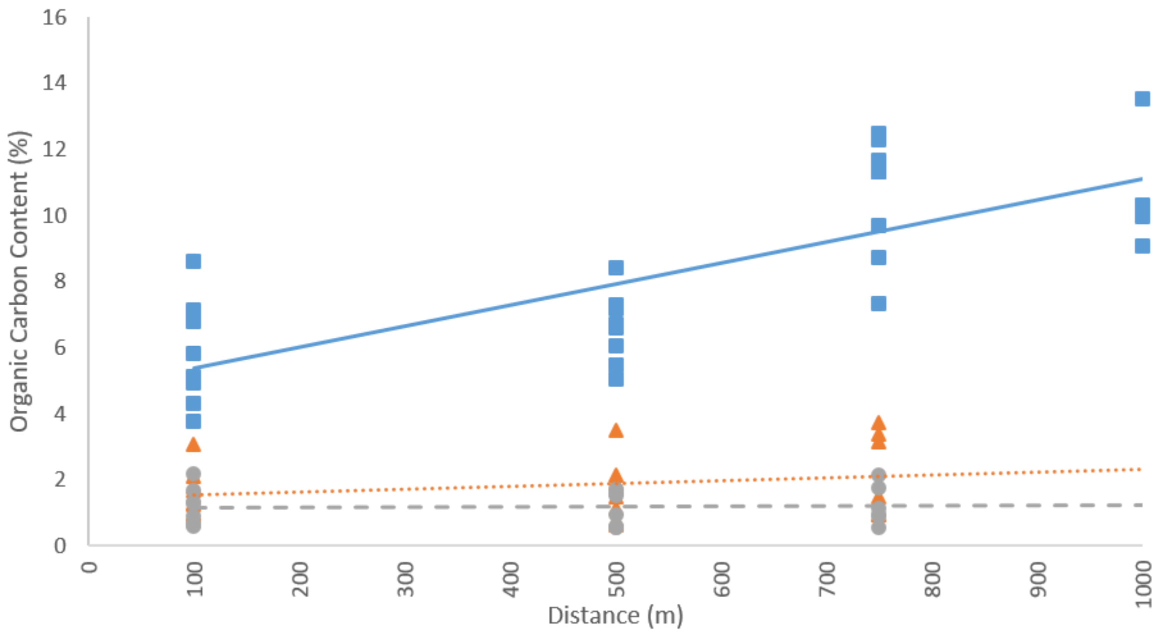

Fig. 3 Organic carbon content at each sample point in each year taken along transects located 100m to 1000m away from a sea sewage outfall at Ironotter Point (Firth of Clyde, UK) (1995 – blue square and solid line, 1992 – orange triangle and dotted line, 1989 (pre-disposal) – grey circle and dashed line)

The species assemblage in each year was distinct as revealed by MDS, which showed a temporal trend in composition, with communities shifting along the first dimension with a strong pattern related to the year of sampling (Figure 4 (a)) (One-way ANOSIM, R=0.657, p<0.01). There were significant differences in composition between all years with the greatest differences found between 1989, before the sewage pipe was implemented, and the last survey in 1998 (ANOSIM pair wise comparison, R=0.948). Differences in composition between the years increased with time also with 1989 and 1992 being the most similar (ANOSIM pair wise comparison, R=0.551), 1992 and 1995 being more different (ANOSIM pair wise comparison, R=0.577) and 1995 and 1998 being more different again (ANOSIM pair wise comparison, R=0.718). The taxonomic composition within years varied across samples and became slightly more similar in later years (SIMPER average similarity 1989: 35%, 1992: 45%, 1995: 41%, 1998: 41%). A strong pattern related to organic matter content was also found (One-way ANOSIM, R=0.273, p<0.01) and this was related to the year of sampling (Figure 4 (b)). The greatest differences were between the communities found in samples with 'unknown' organic carbon values (from 1998) and those with the lowest carbon content (One-way ANOSIM, R=0.856 (0-1% carbon) and R=0.874 (>1-2% carbon), p<0.01). The strength of the relationship between benthic community similarity and organic carbon was weaker than found with year but organic carbon data were not available for 1998. The large change in organic matter content in 1995 (from an average of less than 2% previously to more than 8% on average), parallels the large changes in community composition found between 1992 and 1995. Although no data for 1998 were available, based on the fact that waste was being received and the capacity of the waste treatment plant was being phased up, it could be assumed that there was a corresponding further increase in organic matter and this was also paralleled by a further large change in community composition between 1995 and 1998.

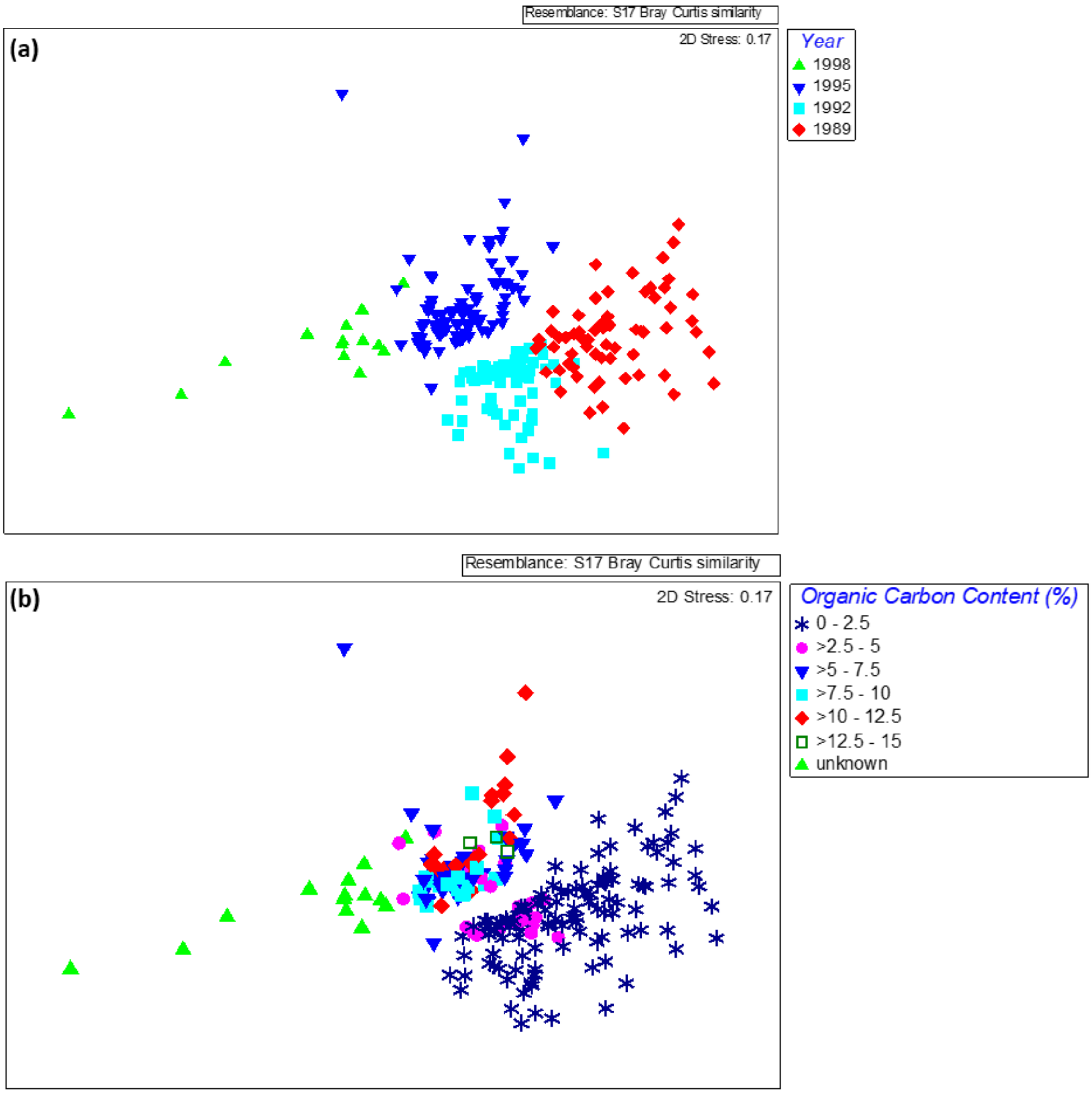

Fig. 4 MDS plots of benthic species data obtained from samples collected at Ironotter Point (Firth of Clyde, UK) according to (a) year and (b) organic carbon content in sediment. 1989 was the baseline year, before implementation of the sewage outfall pipe. Each point represents one sample with a total of 228 samples.

Across all years, 471 taxa were identified. Even when accounting for the difference in the numbers of samples taken, species richness increased over time and was greater in each year since sewage disposal began compared to before the sewage plant was operational (the least number of samples were taken in 1998 but there was still greater species richness than in either 1989 or 1992) (Table 3). 1989 had the greatest proportion of its species (19%) classified as being in AMBI Group I, i.e. the most sensitive species to organic enrichment. This compares to only 5% Group I species in 1998. The number of species that were only found in a given year also increased over time, particularly in

1995, when these made up 30% of all species (though this year also saw the most samples taken). Some of these species were in AMBI Group I and this proportion decreased over time.

Table 3 The total (pooled) number of species found across all samples taken in each year at Ironotter Point (1989 n= 64, 1992 n= 66, 1995 n= 84, 1998 n=14). The number of species in each year that were found only in that year ('unique' species). The proportion of each of these that are species classified as Ecological Group I species (i.e. AMBI Group I, see Table 1).

Year	Total Number of Species Found	Proportion of Species in Ecological Group I (%)	Number of Unique Species (Proportion of Total Found Per Year)	Proportion of Unique Species in Ecological Group I (%)
1989	190	19	29 (15%)	45
1992	208	8	20 (10%)	55
1995	350	11	105 (30%)	35
1998	232	5	40 (17%)	30

In 1989, the community composition was distinct from other years in that the species found to be driving community patterns were the polychaetes *Anobothrus gracilis, Spiophanes kroyeri* and *Mediomastus fragilis* and the bivalves *Nucula nitidosa* and *Thyasira flexuosa* (Table 4). From 1992 onwards, some second-order opportunistic species featured amongst the most important species (from AMBI group IV). In this year, polychaetes *Scalibregma inflatum, Chaetozone setosa*, *Chaetozone zetlandica* and *M. fragilis* were found, in addition to *T. flexuosa*. In 1995, *T. flexuosa, C. setosa* and *M. fragilis* were also driving community patterns, along with the Nemertean, *Tubulanus* sp. and the polychaete *Nephtys* sp. The polychaete *Ophryotrocha hartmanni,* the most dominant species in 1998, was more than five times more abundant than the most dominant species in any other year. Other species that characterised the community in this year were the polychaetes *M. fragilis*, *Prionospio fallax* and *Melinna palmata* and the bivalve *T. flexuosa*. Across years, most of the important species were from AMBI group III, indicating they are tolerant. Amongst the most important species, the most common mode of bioturbation was found to be surface deposition, where fauna deposit materials at the sediment surface, while the most common feeding types were surface deposit and detritus feeders. *S. kroyeri*, a suspension feeder, was present in every year, but only important in driving community patterns in 1989, indicating a change in functioning after implementation of the pipe.

Table 4 The five species identified using SIMPER analysis as being primarily responsible for observed patterns across years at Ironotter Point with total abundance, average abundance, contribution (%), and cumulative total of contributions (%), ecological group (i.e. AMBI group, see Table 1), bioturbation mode (MarLIN, 2006, Solan et al., 2004), and functional feeding type (ITI group, see Table 1)

Species	Total abundance	Average Abundance	Contribution (Cumulative)	Ecological group	Bioturbation mode	Functional feeding type
1989(n=64)						
Anobothrus gracilis	746	11.66	11.77 (11.77)	III	Surface deposition	Surface detritus feeder
Nucula nitidosa	573	8.95	8.53 (20.30)	I	Diffusive mixing and Surface deposition	Surface deposit feeder
Thyasira flexuosa	550	8.59	7.70 (28.00)	III	Surface deposition	Surface deposit feeder
Spiophanes kroyeri	506	7.91	7.18 (35.18)	III	Surface deposition	Suspension feeder
Mediomastus fragilis	438	6.84	5.51 (40.69)	III	Upward conveyer	Surface deposit feeder
1992 (n=66)						
Scalibregma inflatum	1,658	25.12	12.23 (12.23)	III	Diffusive mixing and Downward conveyer	Surface detritus feeder
Chaetozone setosa	1,701	25.77	12.05 (24.28)	IV	Surface deposition and Downward conveyer	Surface detritus feeder
Thyasira flexuosa	1,220	18.48	7.56 (31.84)	III	Surface deposition	Surface deposit feeder
Chaetozone zetlandica	881	13.35	5.87 (37.71)	IV	Surface deposition and Downward conveyer	Surface detritus feeder
Mediomastus fragilis	1,348	20.84	5.54 (43.25)	III	Upward conveyer	Surface deposit feeder
1995 (n=84)						
Tubulanus sp	2,156	25.67	7.96 (7.96)	II	Diffusive mixing and Surface deposition	Surface deposit feeder
Thyasira flexuosa	2,536	30.19	7.55 (15.51)	III	Surface deposition	Surface deposit feeder
Chaetozone setosa	2,026	24.12	7.48 (23.00)	IV	Surface deposition and Downward conveyer	Surface detritus feeder
Mediomastus fragilis	1,836	21.86	6.79 (29.78)	III	Upward conveyer	Surface deposit feeder
Nephtys sp.	1335	15.89	4.84 (34.62)	II	Diffusive mixing	Surface deposit feeder
1998 (n=14)						
Mediomastus fragilis	3,346	239.00	17.16 (17.16)	III	Upward conveyer	Surface deposit feeder
Ophryotrocha hartmanni	13,975	998.21	13.64 (30.80)	IV	Diffusive mixing and Surface deposition	Sub-surface deposit feeder
Prionospio fallax	2,006	143.29	12.24 (43.04)	IV	Surface deposition	Surface detritus feeder
Thyasira flexuosa	1,058	75.57	53.35 (53.35)	III	Surface deposition	Surface deposit feeder
Melinna palmata	771	55.07	6.26 (59.61)	III	Surface deposition	Surface deposit feeder

Spearman correlations of indices S, N, A/S, ITI, BOPA and BQI showed strong relationships with organic content and year (Table 5), reflecting the MDS which showed a strong effect of both organic content and year on the benthic community composition. However, not all indices reflected the expected decrease in environmental quality with time and with organic enrichment. Only J', ITI, A/S, BOPA and AMBI indicated a decrease in quality with year and with organic enrichment. Only ITI showed decreasing quality when the effect of year was removed from the effect of organic matter (partial correlation analysis). Several indices showed a decrease in quality when the effect of organic matter was removed from the effect of year. This could indicate that some of the change over time was not due to the increase in organic matter, however no organic matter content data were available for the last survey which may have influenced these results by reducing the strength of the correlation between the indices and organic carbon. There may also not have been a linear relationship with organic carbon content. However, most indices found an improvement in quality with year and organic matter content. The strongest correlations found were between species richness and abundance with year and organic matter (Spearman rank correlation, r between 64-79%, Table 5). Weak correlations with distance (Spearman rank correlation, r between +/-20% for all indices, Table 5) could be due to the distance effect not being present, or being weak in the early years while becoming more pronounced later. Distance from the outfall did not appear to influence change in the community until the final year (for which no organic carbon data are available) and this was reflected by the quality classifications of indices (Figure S1). Thus, those samples closest to the outfall did not have worse quality than those further away but there were similar quality classifications over the whole area.

Table 5 Correlation between indices and environmental variables at Ironotter Point. Spearman rank correlations with percentage correlation, *r*. Partial correlation carried out to remove effect of confounding variable 'year' from effect of 'organic carbon content and vice versa. Darker shades indicate a stronger relationship. Organic carbon content (%) data was not available for last year of sampling, 1998.

Shade	% Correlation
	≥80
	≥70 <80
	≥50 <70
	≥30 <50
	<30

	Organic Carbon Content	Year	Distance	Organic Carbon Content (year removed)	Year (organic carbon content removed)
S	64.5	79.0	1.0	-5.6	58.4
N	66.2	79.1	-4.7	11.9	52.4
J'	-30.7	-50.6	4.4	7.1	-34.8
A/S	51.3	61.8	-8.0	10.3	35.1
TD (Δ*)	34.1	22.7	32.0	37.1	-15.2
AMBI	6.6	35.9	-18.8	-37.1	48.0
IQI	30.0	13.1	20.8	24.6	-6.2
ITI	-48.2	-57.3	-0.6	-30.8	-17.6
BOPA	16.4	48.7	-5.9	-34.5	54.9
BQI	54.8	34.4	12.1	34.8	-8.6

Note: A/S, BOPA and AMBI index values have inverse relationships with quality

IQI, the multimetric Infaunal Quality Index used for WFD classification in the UK, showed a positive relationship with organic carbon content (Table 5). This outcome for IQI is influenced by the components of this index, which include species richness and AMBI (Table 1). Over time (Figure 5(f)), IQI slightly decreases in the first year after implementation of the sewage outfall, then increases from 1992 to 1995 and goes from good quality to high quality. Subsequently, a decrease in the average IQI index value is indicated from 1995 to 1998. This resulted in a weak overall correlation with year (Spearman rank correlation, r=13%, Table 5), despite the change in community reflected by the ANOSIM (ANOSIM pair wise comparison between 1989 and 1998, R=0.948). AMBI did not change much over time until the final year (Figure 5(e)), while species richness increased over time (Figure 5(a)). In the final year of sampling, AMBI and IQI indicated the change in the community as a negative one. However, while AMBI changed from an overall good classification in 1989-1995 to moderate in 1998, IQI remained at a good classification in the last year of sampling, disguising the relatively slight but unacceptable degradation in quality found by AMBI and other indices, including

BOPA (Figure 5(g)) and ITI (Figure 5(h)), but reflecting the continued increase in species richness over time. ITI and the measures of evenness, A/S and J' (Figure 5(c), (d)), were the only indices to find a consistent deterioration in quality over time. These responses from the indices do not parallel the large change in community composition and the increase in organic carbon content found between 1992 and 1995, and the increasing importance of second order opportunistic species in the communities from 1992 onwards, indicating organic enrichment.

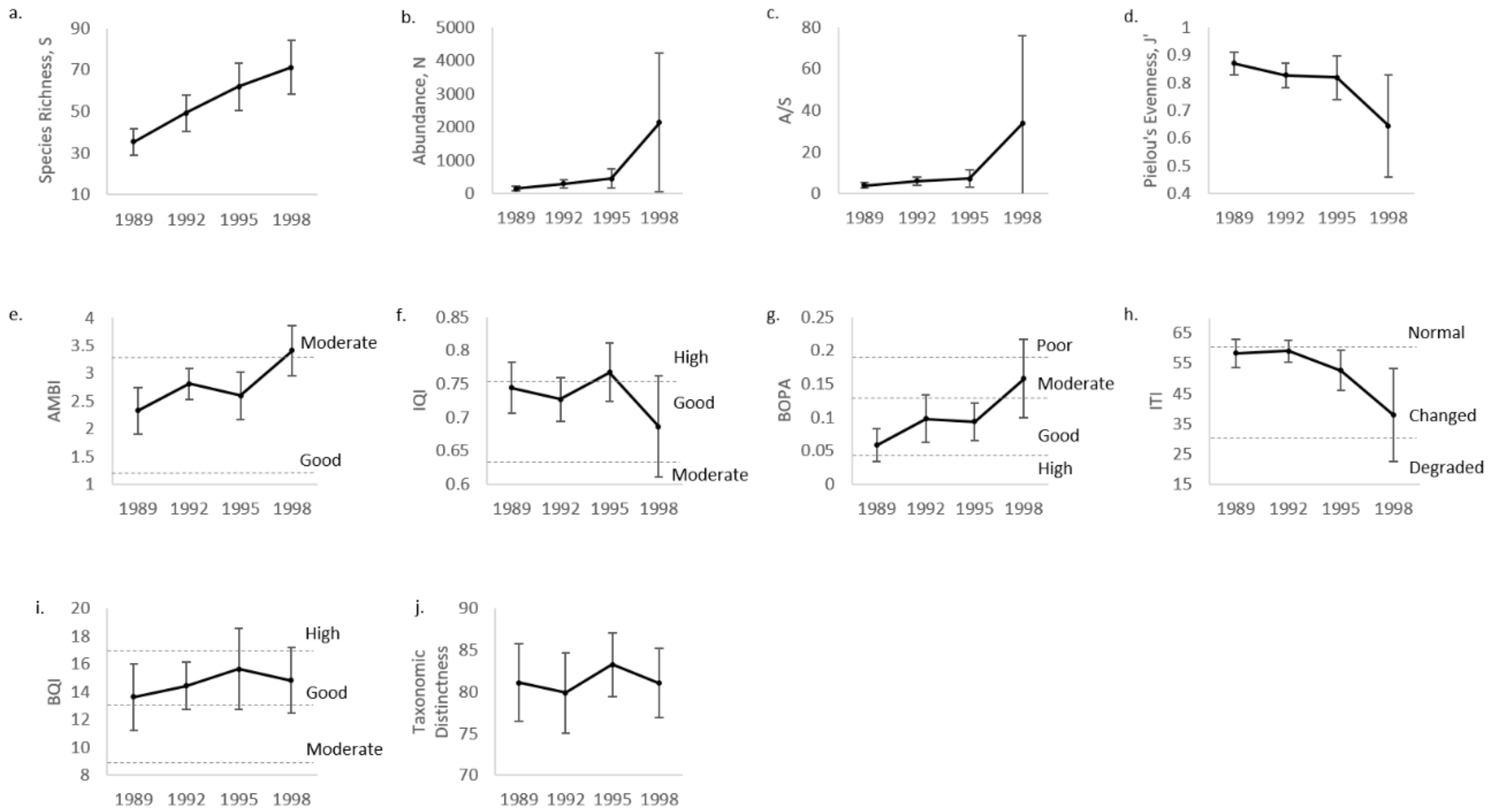

Fig. 5 Change in macrobenthic community indices over time at Ironotter Point (Firth of Clyde, UK) (average across all sites with standard deviation, 1989 n= 64, 1992 n= 66, 1995 n= 84, 1998 n=14). 1989 was the baseline year, before implementation of the sewage outfall pipe. Note that AMBI, A/S and BOPA have inverse relationships with quality. For details of indices, see Table 1. Quality classification threshold information can be found in Table S2.

Five indices, which have an associated quality classification (IQI, BQI, AMBI, BOPA and ITI) were used to assess the quality for each sample (e.g. ‘good’, ‘moderate’, ‘poor’) and the level of agreement between these indices was determined (see Table S3 in supporting information and Figure 6). At Ironotter, most index quality classifications indicated a decrease in quality in 1998 but no decrease in quality was evident before this, apart from ITI which showed a greater frequency of ‘changed’ classification from 1995 onwards when sediment organic matter substantially increased (Table S3). According to IQI and BQI, quality appeared to increase after the sea pipe was put in place and decrease again in 1998. This does not reflect the change in community composition and organic enrichment found over time, which indicated a decrease in quality, as described above. Looking at all samples, a total of 7% of index classifications ‘agreed’, 44% showed a ‘similar’ classification (no more than one quality classification higher or lower) and 48% ‘disagreed’ (three quality classifications given for the same sample point) and there were significant differences in average similarity of the communities at sites where indices ‘disagreed’ versus those where they were ‘similar’ (One-way ANOSIM pairwise comparison, $R=0.109$, $p<0.01$). All the samples where the indices agreed were found in 1989 and 1992 and most of the similar classifications were also found in these two years (Figures 4(a)). Most of the disagreement between indices occurred in the years 1995 and 1998, after organic matter content had increased considerably. The ITI finding largely ‘changed’ quality and the IQI finding largely ‘high’ quality was responsible for the majority of disagreement between indices (Table S3). Looking at distribution of quality assignment by indices overall showed IQI and BQI indicated better quality (80% of classifications were good or high) than BOPA and AMBI (around 40% good and no high classifications), while ITI assigned the lowest quality classifications (only changed or degraded classifications) (Figure 6).

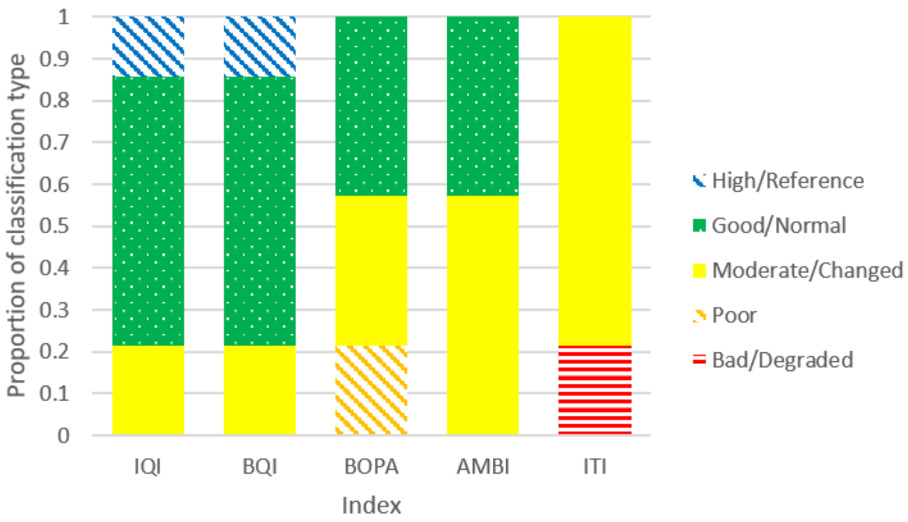

Fig. 6 Differences in quality classifications in macrobenthic community indices found between five indices for the same array of samples taken over four years between 1989-1998 (n=228) at Ironotter Point (Firth of Clyde, UK).

DISCUSSION

The Pearson-Rosenberg paradigm (1978) predicts that when moving away from a source of organic pollution in space and/or time species richness will gradually increase until stable communities are reached, while abundance will increase quickly at first, before then decreasing. The current study of Ironotter Point included a baseline survey before implementation of a sewage sea outfall, followed by three surveys taken one, four and seven years after the pipe was in use. Due to the known input of organic matter and subsequent organic enrichment at this site, the expected response over time was a decrease in species richness and an increase in the abundance of opportunistic species at the pollution sources, as waste output increased or accumulated over time at the site. While abundance of opportunistic species did increase, the species more sensitive to organic pollution were maintained, resulting in an overall increase in species richness over the course of the study. This response was in line with studies that have found nutrient enrichment can lead to increased species richness, greater availability of resources, greater productivity and rates of functioning and increased survival and recruitment in benthic communities (e.g. Abdelrhman & Cicchetti, 2012, Borja et al., 2009b, Krumhansl et al., 2015, Lawes et al., 2016). Species richness can respond to disturbance by initially increasing, corresponding to the intermediate disturbance hypothesis (Connell, 1978, Dodson et al., 2000, Hooper et al., 2005, Huston, 2014, Mittelbach et al., 2001, Odum, 1985).

Studying this site over a longer time period may have shown a subsequent decrease in species richness, as this is the expected response to organic enrichment (Pearson & Rosenberg, 1978).

A previous study showed that sites with higher currents can support high diversity alongside high abundances of opportunistic species in the presence of disturbance (Keeley et al., 2013). In the first year after operation of the pipe at Ironotter, organic matter levels were only slightly elevated but had substantially increased in 1995. The carbon content found in 1995 (on average 8% and a range of 3.78-13.52%) is comparable to that found in impacted sewage disposal sites elsewhere (e.g. at the Garroch Head sludge disposal central point values ranged from 6-15% (Caswell et al., 2018) and at the Nervion estuary the means at the most impacted sites were around 10%, and at intermediate sites 6-7% (Borja et al., 2006)). Nevertheless, in 1995, most indices did not detect impacts on the community due to the enrichment. Given that increased organic carbon was observed over the entire area, even up to 1000m from the outfall, this suggests the site was dispersive and may have a relatively high flow. However, at the same time, the build-up of organic matter found in 1995, was more comparable to low flow sites in Keeley et al. (2013), where higher levels of biodiversity were not supported. The levels of species richness supported at this site (on average 61.9 in 1995 and 71.1 in 1998) are much greater than those reporting sewage impacts elsewhere with similar levels of organic enrichment (e.g. 6.6-24.7 species at the most impacted stations at Garroch Head (Caswell et al., 2018)). The species richness here is similar to that found at disposal sites in Liverpool Bay, where the elevated species richness at disposal sites compared to reference sites was attributed to natural site differences (Whomersley et al., 2007). However, that study had much lower levels of organic carbon than found in this study (ranging from 1.92-3.04%) and did not have pre-disposal data available to know if sites were naturally different.

Even with an increase in species richness, other indices which take species identity and the proportion of opportunistic species into account, would still be expected to indicate a decrease in quality, as found in other studies (e.g. Borja et al., 2009b). At Ironotter, ITI detected a slight decreasing trend with organic enrichment and showed a decreasing trend with year from 1992 onwards, reflecting a change in the functional feeding group composition and a relative decrease in suspension feeders, like *Spiophanes kroyeri*. AMBI indicated a decrease in quality over time but, in contrast to what has been found elsewhere (Borja et al., 2006), showed no strong trend with increasing organic matter content. The decrease in quality detected did not manifest in a change in quality classification until the final year of sampling. From 1992 to 1995, AMBI showed no change because disturbance sensitive taxa were maintained in the community, even in the presence of organic enrichment. BOPA performed similarly to AMBI. However, during this time, the community was changing, and the differences were primarily driven by species that have been found to be indicative of enrichment elsewhere. For example, in 1989, important species in the community included *S. kroyeri* and *Nucula nitidosa*, species that are associated with undisturbed sites (e.g. Caswell et al., 2018, Reiss & Kröncke, 2005). Other species found in this year included *Mediomastus fragilis* and *Thyasira flexuosa,* both of which have been associated with intermediate levels of organic enrichment (Caswell et al., 2018, Pearson & Rosenberg, 1978, Rees et al., 2006). These then increased in abundance in 1992, along with further organic enrichment indicator species *Scalibregma inflatum*, *Chaetozone setosa*, and *Chaetozone zetlandica* (Pearson & Rosenberg, 1978, Rees et al., 2006). A similar pattern of increasing abundance of these enrichment indicator species continued until, in 1998, the community was dominated by opportunistic species such as *Ophryotrocha hartmanni* (Cardell et al., 1999), showing a clear indication of an enriched community. The multimetric index, IQI, detected an increase with organic matter and a small decrease in quality with year when the effect of organic matter was removed. This linear trend hid what was initially a slight increase in quality with year before a slight decrease again. This was reflected in the quality classifications, which mostly changed from good to high, to good again in the final year. This also reflects the influence of species richness on this index. Species richness and AMBI overall demonstrated opposing trends at this site and this resulted in the outcome that the negative trend indicated by AMBI was dampened so that IQI showed little change in quality over time. Out of the five indices that have quality classifications, apart from ITI, negative impacts implying a degraded state were not detected by the indices until around seven years after the pipe was installed. This may be a suggestion of the resilience of the benthic community since higher diversity is often correlated to greater resilience or stability (e.g. Worm et al., 2006). Other studies have found a lag in response of benthic communities, where they remain relatively stable before a critical threshold of oxygen saturation is reached and the community collapses (Diaz & Rosenberg, 1995, Josefson & Widbom, 1988). In this study, the communities did not collapse, but species richness increased alongside clear indications of changes in the community, ultimately towards a degraded state. The ITI results showed that functional characteristics, in terms of the feeding modes present within the community, had changed by 1995. Thus, it would appear resilience and stability were decreasing over time and with increasing enrichment.

It is possible that the patterns observed were due to other external factors unrelated to the organic enrichment, or where there are other factors which maintain levels of species richness in the presence of organic enrichment, such as the presence of differing flow regimes (e.g. Keeley et al., 2013), or where responses are highly site specific (Villnäs et al., 2011). However, we did not have an extended time series of data from before the baseline study or an undisturbed reference site against which to make a comparison. Partial correlations indicated the effect of year was stronger than the organic carbon effect, suggesting other factors could also be involved. These other factors may be related to the input of sewage into the system but not captured by measuring organic carbon content. For example, in the Nervion estuary, Borja et al. (2006) found AMBI to be more highly correlated with oxygen saturation than with organic matter, although there was still much stronger correlations with organic matter than were found in this study. No organic carbon data were available for the last year which may also obscure these results. The waste discharged from the sewage treatment plant at this time was likely to have other contaminants in addition to organic matter and, although not consistently measured, measurements taken in 1995 showed elevated levels of heavy metals (Table S4). Other studies have shown that the presence of nutrients can mitigate the effects of heavy metals on some benthic invertebrates by increasing recruitment and decreasing mortality (Lawes et al., 2016). However, it is likely that these highly persistent contaminants will build up in the system over the years, and not be removed in the same way that nutrients are, eventually leading to deteriorating conditions (Watson et al., 2016).

Different types of linear and nonlinear responses to environmental gradients were found at this site and responses could be confounded with several factors, measured and unmeasured, making interpretation of the index responses difficult. As well as a greater amount of information benefitting interpretation of index responses, other methods may be more suitable for measuring the response of indices to environmental and temporal gradients in order to detect nonlinear trends and to account for confounding factors. Long term studies also may be responding to natural succession events in the community (e.g. Clare et al., 2017, Rees et al., 2006, Stull et al., 1986). However, the clear shift in species composition according to ANOSIM suggested the community was showing a larger and more rapid response than would be expected under normal conditions (Clarke & Warwick, 2001) and this was most likely to be due to the input of organic material. Species richness doubled from pre-disposal to the last year of sampling, but this study shows any change, up or down, in species richness should be used as an early indication of change in the system. Echoing the recommendations of others (e.g. Kröncke & Reiss, 2010, Pinto et al., 2009, Villnäs et al., 2012), this should be considered alongside other methods such as multivariate analysis to look at species composition, measures of evenness, ITI, AMBI, environmental variables and other indices to interpret the change in the system. The complexity of potential responses (e.g. Borja et al., 2006, Caswell et al., 2018, Rees et al., 2006, Whomersley et al., 2007) shows the importance of not relying on a single index for quality classification. Structural properties of ecosystems can respond in variable ways to disturbance while functional properties may indicate, more reliably, the direction of changes in quality (Feld & Hering, 2007, Paul, 1997, Villnäs et al., 2012) or the cause of change (Culhane et al., 2014). We found those indices that take species richness into account (e.g. IQI, BQI) find a greater proportion of high/reference conditions, and those that are based only on species identity and include functional traits (e.g. BOPA, AMBI, ITI), find greater proportions of moderate, poor or bad ecosystem state.

Studies have found that, generally different benthic index results correlate positively with each other, but do not necessarily arrive at the same quality classification, suggesting a need for calibration (Blanchet et al., 2008, Labrune et al., 2006, Quintino et al., 2006, Zettler et al., 2007). In this study, the agreement between the five index classifications mainly showed agreement between indices, or similar trends in the baseline year and first year after installation. However, in 1995 most indices were found to disagree and in 1998 most indices again showed similar status. It may be that a moderate level of disturbance increases disagreement between indices and causes indices to act unpredictably. Caswell et al. (2018) found greater coefficients of variation of species richness at intermediate sites than at reference sites, highlighting greater patchiness and variability at these sites. This patchiness could also influence index results at intermediate levels of disturbance. Since indices perform less well in distinguishing intermediate disturbance, this could also be important for the detection of small changes in quality and early warning signals. AMBI detected little change in quality (index value), and no change in quality classification until the last year. Other studies have found AMBI to be unsuitable for detecting small changes over time (Kröncke & Reiss, 2010) and differences between 'good' and 'moderate' qualities can go undetected (Puente & Diaz, 2008).

Results in this study showed that most indices did not respond in the expected way to the introduction of a disturbance. This was largely due to the increase in species richness across all stations over time that we observed. Indices should make the interpretation of complex environmental data simpler and ideally reflect anthropogenic disturbance in a reliable and consistent way (Karr, 1999). During periods of moderate change, indices may be particularly unpredictable in how they classify quality. Species richness did not change in the expected direction but did change quickly and considerably in magnitude, and therefore could be considered an early warning indicator of disturbance, but further exploration and assessment would be required to interpret this change, including multivariate analysis and interpretation of environmental variables.

The use of multi-metric indices such as IQI is largely policy driven (Phillips et al., 2014, van Loon et al., 2015). Multi-metric indices should magnify a common signal in the combined metrics (Schoolmaster et al., 2012). But if those signals are opposing due to the unpredictable response of communities to change, such as in this study, this signal can be dampened. IQI still indicated 'good' quality in the final year of sampling, whereas AMBI indicated a decrease in quality to 'moderate' status. Indices summarise a large amount of information into a single value, resulting in something which, in principle, is easier to understand but is also possibly prone to misinterpretation due to loss of information (Rees et al., 2006). It may be more beneficial to use the separate component parts of multi-metric indices rather than the combined form, thus taking a more cautious and informative approach to the assessment of benthic health.

As human populations increase, the problem of waste treatment and sewage discharge into coastal areas also increases. Benthic communities here exhibited a degree of resilience to degradation in response to the input of sewage, with high species diversity maintained alongside high abundance of opportunistic species. Yet, the communities showed compositional and functional changes indicative of organic enrichment before being classified as degraded in 1998. There was a shift between 1995 and 1998, where the community switched from initially coping with the level of waste to showing clear indications of unacceptable change. We found some of the indices used, such as IQI, were less sensitive to small shifts, while species richness changed quickly. At the global extent that sewage is discharged around the world, there are significant implications for the structure and functioning of coastal ecosystems and their capacity to continue to supply the waste treatment ecosystem service. There is a need to establish the relative utility of different assessment approaches for 'early warning' signals. In particular, we need to establish when is the best time to intervene, rather than wait until major changes have occurred, and hence potential impacts are greater and more difficult to rectify.

ACKNOWLEDGEMENTS

This work was supported by funding from the EU FP6 project SPICOSA, although the work reported here was not part of the requirements or deliverables of SPICOSA. We thank Myles O'Reilly and the Scottish Environment Protection Agency (SEPA) for providing data and background information about the study site and sampling protocol. We thank colleagues from the University of Liverpool's Ecology and Marine Biology Department and anonymous reviewers for comments on a previous version of this manuscript.

SUPPORTING INFORMATION

Supporting information contains additional information on sampling and data collection, as well as additional supporting results.

Abdelrhman M.A. and Cicchetti G. (2012) Relationships between Nutrient Enrichment and Benthic Function: Local Effects and Spatial Patterns. *Estuaries and Coasts*, **35**(1), 47-59.
Aberson M.J.R., Bolam S.G. and Hughes R.G. (2016) The effect of sewage pollution on the feeding behaviour and diet of Hediste (Nereis diversicolor (O.F. Müller, 1776)) in three estuaries in south-east England, with implications for saltmarsh erosion. *Marine Pollution Bulletin*, **105**(1), 150-160.
Baum R., Luh J. and Bartram J. (2013) Sanitation: A Global Estimate of Sewerage Connections without Treatment and the Resulting Impact on MDG Progress. *Environmental Science & Technology*, **47**(4), 1994-2000.
Blanchet H., et al. (2008) Use of biotic indices in semi-enclosed coastal ecosystems and transitional waters habitats--Implications for the implementation of the European Water Framework Directive. *Ecological Indicators*, **8**(4), 360-372.
Bolam S.G., et al. (2017) Differences in biological traits composition of benthic assemblages between unimpacted habitats. *Marine Environmental Research*, **126**, 1-13.
Borja A., et al. (2011) Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. *Marine Pollution Bulletin*, **62**(3), 499-513.
Borja A., et al. (2008) Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Marine Pollution Bulletin*, **56**(9), 1519-1537.
Borja A., Franco J. and Perez V. (2000) A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Marine Pollution Bulletin*, **40**, 1100-1114.
Borja A., Muxika I. and Franco J. (2006) Long-term recovery of soft-bottom benthos following urban and industrial sewage treatment in the NerviÃƒÂ³n estuary (southern Bay of Biscay). *Marine Ecology Progress Series*, **313**, 43-55.
Borja A., Ranasinghe A. and Weisberg S.B. (2009a) Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. *Marine Pollution Bulletin*, **59**(1), 1-4.
Borja Á., et al. (2009b) Assessing the suitability of a range of benthic indices in the evaluation of environmental impact of fin and shellfish aquaculture located in sites across Europe. *Aquaculture*, **293**(3), 231-240.
Bowen J.L. and Valiela I. (2001) The ecological effects of urbanization of coastal watersheds: historical increases in nitrogen loads and eutrophication of Waquoit Bay estuaries. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**(8), 1489-1500.
Bowgen K.M., Stillman R.A. and Herbert R.J.H. (2015) Predicting the effect of invertebrate regime shifts on wading birds: Insights from Poole Harbour, UK. *Biological Conservation*, **186**, 60-68.
Byers S.C., Mills E.L. and Stewart P.L. (1978) A comparison of methods for determining organic carbon in marine sediments, with suggestions for a standard method. *Hydrobiologia*, **58**(1), 43-47.
Cardell M.J., Sardà R. and Romero J. (1999) Spatial changes in sublittoral soft-bottom polychaete assemblages due to river inputs and sewage discharges. *Acta Oecologica*, **20**(4), 343-351.
Caswell B.A., Paine M. and Frid C.L.J. (2018) Seafloor ecological functioning over two decades of organic enrichment. *Marine Pollution Bulletin*, **136**, 212-229.
Clare D.S., Spencer M., Robinson L.A. and Frid C.L.J. (2017) Explaining ecological shifts: the roles of temperature and primary production in the long-term dynamics of benthic faunal composition. *Oikos*, **126**(8), 1123-1133.

Clarke K.R. and Warwick R.M. (2001) Change in marine communities: an approach to statistical analysis and interpretation. Primer-E, Lutton, United Kingdom.
Connell J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science*, **199**, 1302-1310.
Culhane F.E., et al. (2014) Development of an operational EU policy-based marine ecosystem (services) assessment framework. Deliverable 9: Report to the European Environment Agency from the University of Liverpool. . *University of Liverpool*, 432.
Dauvin J.C. and Ruellet T. (2007) Polychaete/amphipod ratio revisited. *Marine Pollution Bulletin*, **55**(1-6), 215-224.
Diaz R.J. and Rosenberg R. (1995) Marine benthic hypoxia: A review of its ecological effects and the behavioural response of benthic macrofauna. *Oceanography and Marine Biology*, **33**, 245-303.
Diaz R.J. and Rosenberg R. (2008) Spreading Dead Zones and Consequences for Marine Ecosystems. *Science*, **321**(5891), 926-929.
Diaz R.J., Solan M. and Valente R.M. (2004) A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management*, **73**, 165-181.
Dodson S.I., Arnott S.E. and Cottingham K.L. (2000) The Relationship in Lake Communities between Primary Productivity and Species Richness. *Ecology*, **81**(10), 2662-2679.
EC (2000) Establishing a framework for Community action in the field of water policy. Directive 2000/60/EC of the European Parliament and of the Council. . *Official Journal of the European Communities*, **L327**, 1-72.
Elliott M. and Quintino V. (2007) The Estuarine Quality Paradox, Environmental Homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. *Marine Pollution Bulletin*, **54**(6), 640-645.
Feld C.K. and Hering D. (2007) Community structure or function: effects of environmental stress on benthic macroinvertebrates at different spatial scales. *Freshwater Biology*, **52**(7), 1380-1399.
Foteinis S., Kallithrakas-Kontos N.G. and Synolakis C. (2013) Heavy Metal Distribution in Opportunistic Beach Nourishment: A Case Study in Greece. *The Scientific World Journal*, **2013**, 472149.
Hooper D.U., et al. (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, **75**(1), 3-35.
Huston M.A. (2014) Disturbance, productivity, and species diversity: empiricism vs. logic in ecological theory. *Ecology*, **95**(9), 2382-2396.
Josefson A.B. and Widbom B. (1988) Differential response of benthic macrofauna and meiofauna to hypoxia in the Gullmar Fjord basin. *Marine Biology*, **100**(1), 31-40.
Karr J.R. (1999) Defining and measuring river health. *Freshwater Biology*, **41**, 221-234.
Keeley N.B., Forrest B.M. and Macleod C.K. (2013) Novel observations of benthic enrichment in contrasting flow regimes with implications for marine farm monitoring and management. *Marine Pollution Bulletin*, **66**(1), 105-116.
Kiedrzyńska E., et al. (2014) Point sources of nutrient pollution in the lowland river catchment in the context of the Baltic Sea eutrophication. *Ecological Engineering*, **70**, 337-348.
Kröncke I. and Reiss H. (2010) Influence of macrofauna long-term natural variability on benthic indices used in ecological quality assessment. *Marine Pollution Bulletin*, **60**(1), 58-68.
Krumhansl K.A., et al. (2015) Assessment of Arctic Community Wastewater Impacts on Marine Benthic Invertebrates. *Environmental Science & Technology*, **49**(2), 760-766.

Labrune C., et al. (2006) Characterisation of the ecological quality of the coastal Gulf of Lions (NW Mediterranean). A comparative approach based on three biotic indices. *Marine Pollution Bulletin*, **52**, 34-47.
Lawes J.C., Clark G.F. and Johnston E.L. (2016) Contaminant cocktails: Interactive effects of fertiliser and copper paint on marine invertebrate recruitment and mortality. *Marine Pollution Bulletin*, **102**(1), 148-159.
Li X., et al. (2018) Microplastics in sewage sludge from the wastewater treatment plants in China. *Water Research*, **142**, 75-85.
Macleod C.K., Moltschaniwskyj N.A., Crawford C.M. and Forbes S.E. (2007) Biological recovery from organic enrichment
some systems cope better than others. *Marine Ecology Progress Series*, **342**, 41-53.
Magurran A.E. (2004) *Measuring Biological Diversity*, Oxford: Blackwell Publishing.
MarLIN (2006) BIOTIC - Biological Traits Information Catalogue. Marine Life Information Network, Plymouth: Marine Biological Association of the United Kingom. Available from <www.marlin.ac.uk/biotic>.
Mayor D.J. and Solan M. (2011) Complex interactions mediate the effects of fish farming on benthic chemistry within a region of Scotland. *Environmental Research*, **111**(5), 635-642.
Mittelbach G.G., et al. (2001) What Is the Observed Relationship between Species Richness and Productivity? *Ecology*, **82**(9), 2381-2396.
Muxika I., Borja A. and Bonne W. (2005) The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. *Ecological Indicators*, **5**, 19-31.
O'Reilly M.G., Boyle J. and Miller B. (1997) The impact of a new long sea outfall on the sublittoral benthos and sediments of the lower Clyde Estuary. . *Coastal Zone Topics*, **3**(The Estuaries of Central Scotland), 129-139.
Odum E.P. (1985) Trends expected in stressed ecosystems. *BioScience*, **35**, 419-422.
Paul M.J. (1997) Back to Odum: Using ecosystem functional measures in stream ecosystem management. In Hatcher K.J. (ed) *Proceedings of the 1997 Georgia Water Resources Conference, held March 20-22, 1997, at the University of Georgia.* Athens, Georgia: Institute of Ecology, The University of Georgia.
Pearson T.H. and Rosenberg R. (1978) Changes in fauna and sediment structure along a gradient of organic enrichment.
Phillips G.R., Anwar A., Brooks L., Martina L.J., Miles A.C. and Prior A. (2014) Infaunal quality index: Water Framework Directive classification scheme for marine benthic invertebrates. *Environment Agency*.
Pinto R., Patrício J., Baeta A., Fath B.D., Neto J.M. and Marques J.C. (2009) Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecological Indicators*, **9**(1), 1-25.
Puente A. and Diaz R.J. (2008) Is it possible to assess the ecological status of highly stressed natural estuarine environments using macroinvertebrates indices? *Marine Pollution Bulletin*, **56**(11), 1880-1889.
Quintino V., Elliott M. and Rodrigues A.M. (2006) The derivation, performance and role of univariate and multivariate indicators of benthic change: Case studies at differing spatial scales. *Journal of Experimental Marine Biology and Ecology*, **330**(1), 368-382.
R Development Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria. https://www.R-project.org/.
Rees H.L., et al. (2006) Benthic responses to organic enrichment and climatic events in the western North Sea. *Journal of the Marine Biological Association of the United Kingdom*, **86**(1), 1-18.

Reiss H. and Kröncke I. (2005) Seasonal variability of benthic indices: An approach to test the applicability of different indices for ecosystem quality assessment. *Marine Pollution Bulletin*, **50**(12), 1490-1499.
Rosenberg R., Blomqvist M., C Nilsson H., Cederwall H. and Dimming A. (2004) Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin*, **49**(9), 728-739.
Schoolmaster D.R., Grace J.B. and William Schweiger E. (2012) A general theory of multimetric indices and their properties. *Methods in Ecology and Evolution*, **3**(4), 773-781.
SEPA (1996) Biological quality at Ironrotter Point, Greenock - 1989-1998. Scottish Environment Protection Agency.
Smith-Evans M. and Dawes A. (1996) Early experiences in monitoring the effects of Hong Kong's new generation of sewage outfalls on the marine environment. *Marine Pollution Bulletin*, **33**(7), 317-327.
Smith V.H. and Schindler D.W. (2009) Eutrophication science: where do we go from here? *Trends in Ecology & Evolution*, **24**(4), 201-207.
Snelgrove P.V.R. (1999) Getting to the Bottom of Marine Biodiversity: Sedimentary Habitats: Ocean bottoms are the most widespread habitat on Earth and support high biodiversity and key ecosystem services. *BioScience*, **49**(2), 129-138.
Solan M., Cardinale B.J., Downing A.L., Engelhardt K.A.M., Ruesink J.L. and Srivastava D.S. (2004) Extinction and Ecosystem Function in the Marine Benthos. *Science*, **306**(5699), 1177-1180.
Stull J.K., Haydock C.I., Smith R.W. and Montagne D.E. (1986) Long-term changes in the benthic community on the coastal shelf of Palos Verdes, Southern California. *Marine Biology*, **91**(4), 539-511.
van Loon W.M.G.M., et al. (2015) Application of the Benthic Ecosystem Quality Index 2 to benthos in Dutch transitional and coastal waters. *Journal of Sea Research*, **103**, 1-13.
Villnäs A., Norkko J., Lukkari K., Hewitt J. and Norkko A. (2012) Consequences of Increasing Hypoxic Disturbance on Benthic Communities and Ecosystem Functioning. *PLOS ONE*, **7**(10), e44920.
Villnäs A., Perus J. and Bonsdorff E. (2011) Structural and functional shifts in zoobenthos induced by organic enrichment — Implications for community recovery potential. *Journal of Sea Research*, **65**(1), 8-18.
Watson S.C.L., et al. (2016) A conceptual framework for assessing the ecosystem service of waste remediation: In the marine environment. *Ecosystem Services*, **20**, 69-81.
WFD-UKTAG W.F.D.-U.K.T.A.G. (2008) UKTAG coastal water assessment method benthic invertebrate fauna. Invertebrates in soft sediments (Infaunal quality Index (IQI)). 17.
Whomersley P., Schratzberger M., Huxham M., Bates H. and Rees H. (2007) The use of time-series data in the assessment of macrobenthic community change after the cessation of sewage-sludge disposal in Liverpool Bay (UK). *Marine Pollution Bulletin*, **54**(1), 32-41.
Word J.Q. (1979) The Infaunal Trophic Index. 19-39.
Worm B., et al. (2006) Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science*, **314**(5800), 787-790.
WWAP (2017) United Nations Worl Water Assessment Programme. The United Nations World Water Development Report 2017. Wastewater: The Untapped Resource. *UNESCO*.
Zettler M.L., Schiedek D. and Bobertz B. (2007) Benthic biodiversity indices versus salinity gradient in the southern Baltic Sea. *Marine Pollution Bulletin*, **55**, 258-270.