Spatial and temporal dynamics in macrobenthos during recovery from salmon farm induced organic enrichment: When is recovery complete?

Nigel B. Keeley [a,b,*], Catriona K. Macleod [b], Grant A. Hopkins [a], Barrie M. Forrest [a]

[a] Cawthron Institute, Nelson 7010, New Zealand
[b] Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, Taroona, Tasmania, Australia

ARTICLE INFO

ABSTRACT

This study documents eight years of benthic recovery at a highly impacted salmon farm. Substantial recovery occurred in the first 2 years, and was assessed to be complete after ∼5 years. However, minor differences were still evident, along with some on-going benthic instability, attributable to medium-scale spatial movements and successional patterns of macrobenthos. Quantifying the endpoint of 'recovery' proved challenging due to: lack of a widely accepted definition, inherent variability in recovering sediments, differing trajectories of impact and reference sites, and statistical challenges. More complex biotic indices and metrics incorporating multiple variables were the most robust indicators. Statistical tests for 'parallelism' in the trajectories of Cage and Reference sites proved useful, but results were contingent upon how the method was applied, and should therefore be used in conjunction with data-visualisation methods. The study highlights the importance of a predetermined recovery endpoint, and using multiple indicators and a weight-of-evidence assessment approach.

1. Introduction

In marine benthic systems, impacts associated with organic enrichment are common and widespread, due to the prevalence of diffuse (e.g. land runoff, Diaz and Rosenberg, 2011) and point source (e.g. outfalls, Cardell et al., 1999; Taylor et al., 1998) discharges of anthropogenic wastes. Two considerations that are critical to evaluating the degree of impact on the environment are spatial scale and 'reversibility' of effects. Strong gradients of ecological succession are common, and the fundamental biological and chemical changes are generally well described (Gray et al., 1979; Kalantzi and Karakassis, 2006; Pearson and Rosenberg, 1978). However, there is less certainty associated with delineating the outer extent of enrichment effects, and the point in time at which a given location can be considered to have recovered from adverse effects; in part this uncertainty reflects natural variability (in both time and space) in environmental conditions (e.g. Hewit and Thrush, 2007; Hewitt et al., 1997; Thrush, 1991) and often a lack of understanding around what constitutes 'natural' conditions.

Finfish aquaculture is a significant point source of organic matter (via waste feed and fish faeces) to the marine environment. The primary discharges of waste feed and faeces typically result in highly enriched conditions in the immediate vicinity of the culture

site (Brooks et al., 2002; Karakassis et al., 2000). In extreme cases, conditions immediately beneath the stocked cages can become anoxic, and virtually azoic (no animal life present), representing 'worst-case' conditions in terms of the duration of the pathway to recovery from impact (Keeley et al., 2012a; Pearson and Rosenberg, 1978). Such situations provide a good case study for understanding benthic enrichment and recovery processes (Keeley et al., 2012a). Additionally, the practice of site fallowing (temporarily retiring a site), that is often used for management purposes, provides a commercial incentive to better understand recovery, as the relative time-scales and processes of recovery and re-impact influence fallowing efficacy.

A wide range of farming conditions can be encountered in finfish aquaculture (i.e. differing farm type, farming intensity and age), which can occur across a range of environments. This situation means there will be a variation in the severity of impact at the start of fallowing (e.g. Borja et al., 2009), and variation in the capacity of a given site to recover from adverse effects. For example, it is generally accepted that high energy sites will recovery faster than low energy sites (Borja et al., 2010). Thus, it is not surprising that estimates of benthic recovery times vary greatly, ranging from weeks (Ritz et al., 1989) to >11 years (Wan Hussin et al., 2012). Several studies, especially those undertaken around smaller fish farms, have suggested that complete recovery (biological and chemical) can occur within 6 months of fallowing (Brooks et al., 2003), and in some cases within periods as short as 7–14 weeks (Brooks et al., 2003, cited in Brooks et al., 2004; Ritz

* Corresponding author at: Cawthron Institute, Nelson 7010, New Zealand. Tel.: +64 3 5393257; fax: +64 3 5469464.
   E-mail address: nigel.keeley@cawthron.org.nz (N.B. Keeley).

et al., 1989). The general consensus from studies conducted over the medium-term (i.e. up to 3 years), is that marked improvement occurs in the first 6–12 months, but that recovery generally remained incomplete (Karakassis et al., 1999; Lin and Bailey-Brock, 2008; Macleod et al., 2008; Pereira et al., 2004; Villnas et al., 2011). Long-term (i.e. >3 years) studies of recovery are scarce; one that was conducted over 7 years estimated full chemical remediation would take 5.3 years and that biological remediation may take much longer (Brooks et al., 2004).

While the spread of these estimates will in part be attributable to the levels of impact at the point of fallowing and varying underlying environmental conditions, there are also multiple definitions of 'recovery' that are likely to contribute to the variances. Brooks et al. (2003) distinguished biological and chemical remediation; highlighting characteristically different recovery pathways, and providing specific criteria for recovery in each case. Other studies have emphasised differences between species-based, community recovery, and 'functional recovery' (Macleod et al., 2008); i.e. the point at which ecosystem function is re-established, but not necessarily with the same communities that were present pre-impact. It is generally assumed, that once functional recovery is achieved, an 'equilibrium state' will ensue (Macleod et al., 2008; Young et al., 2001). The concept of 'sustainable ecological succession', indicated by consistent presence and abundances of a limited number of species, has also been proposed as a measure of recovery (Ellis, 2003).

The difficulties associated with determining the point of recovery are further exacerbated by problems that arise when attempting to evaluate the question statistically. Many impact studies lack an appropriately defined assessment of pre-impact conditions, against which recovery can be quantitatively compared (Verdonschot et al., 2013). Consequently, recovery is assessed by comparison of conditions against selected spatial reference sites, that may in fact be naturally different, and the opportunity to evaluate the degree of change at a particular site is lost. Another problem with using spatial comparison as the reference point for recovery is that it may not always be appropriate to assume a strict equilibrium (or a single 'stable state') in biological systems (Beisner et al., 2003; Parker and Wiens, 2005). There may instead be a 'dynamic equilibrium' or shifting baseline (Macleod et al., 2008; Parker and Wiens, 2005; Verdonschot et al., 2013) and/or several possible alternative stable states (Beisner et al., 2003). Hence recovery should be assessed against a backdrop of both temporal and spatial variation.

Conventional beyond-BACI designs (e.g., Underwood, 1991, 1992) are considered to be one of the best approaches for monitoring recovery (Verdonschot et al., 2013). However, they tend to be resource intensive, requiring both multiple reference sites, and multiple randomly timed samplings within each specified time window. Few multi-year monitoring programs are initiated with this level of sampling effort in place, and maintaining such a design over a long time-frame is uncommon as the cost can be prohibitive. In addition, although beyond-BACI designs clearly partition the multiple sources of variation, the design is premised upon there being two fixed periods, 'before' and 'after' (e.g., Aguado-Giménez et al., 2012), whereas in most long-term datasets time is often a continuous variable that may have a non-linear response. Therefore, with a beyond-BACI approach it can be difficult to directly address the questions "was recovery complete?" and if so, "when did it occur?".

Recovery can be conceptually defined as occurring when the impacted resource reaches the level at which it would have been, had it not been impacted in the first place. At that point, the influence of impact-related factors will have diminished to a situation where levels of the resource vary temporally in a natural way (Parker and Wiens, 2005 and U.S. Code of Federal Regulation, 2001). The concept of 'varying temporally in a natural' way implies

an assumption of 'parallelism', whereby impact and reference sites will begin to respond similarly; for example, to wider oceanographic processes. This is useful statistically, and methods (based on the BACI approach) have been developed accordingly, which were used to assess recovery from the *Exxon Valdez* oil spill (Skalski et al., 2001). These methods appear to have broader applications, which we explore in this paper along with a variety of other indicators and approaches for evaluating the remediation process and exploring the concept of recovery 'end points'. Our analysis is based on a 10 year dataset that provides a baseline characterization of a highly impacted seabed beneath and adjacent to a salmon farm, which was followed by eight years of annual monitoring of the spatial and temporal patterns of recovery.

2. Methods

2.1. Study sites and sampling procedures

This study was conducted at a commercial Chinook salmon (*Oncorhynchus tshawytscha*) farm site located in the outer reaches of the Marlborough Sounds, New Zealand (Fig. 1). The farm was situated in a sheltered embayment over muddy-sand sediments (average mud content = 78–84%), in water depths ranging between 28 and 35 m, with relatively low current speeds (mid-water mean current speed ≈ 3 cm s^{-1}).

The farm was fallowed in 2001 after approximately seven years of consistent and relatively intensive use (average feed usage of ~180 mt month^{-1}). Benthic sampling was undertaken in the Austral spring (October/November) as follows: two years prior to fallowing in 1999 (T-2), immediately after the farm was fallowed in 2001 (T0), followed by annual monitoring until 2009 (T8), with the exception of 2008 when no sampling was undertaken. Seabed samples were collected beneath the site previously occupied by the cages ('Cage stations'), at 25 m intervals along a north-western transect ('Gradient' stations) running away from the farm, and at fixed Reference stations (Fig. 1). Not all sampling stations were sampled in every year; most notably, two reference sites were
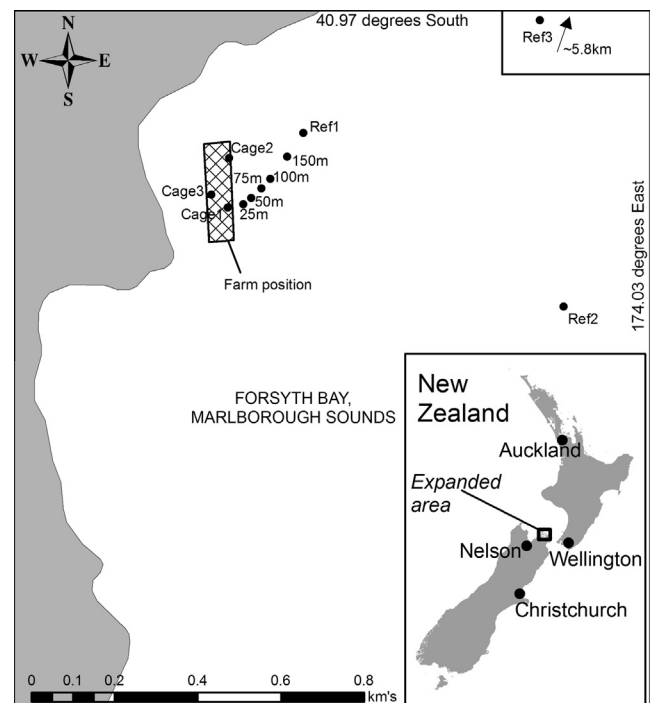

Fig. 1. Location of study site and sampling stations in relation to the farm.

added at ~400 m and ~5800 m distance from the former cage site in 2003 and 2009 respectively (sampling events denoted by ' × ' on subplots in Fig. 2).

At each station, sediment samples were collected using a van Veen grab (0.1 m^2), with water depth (Depth, m) and distance from farm (Distance, m) recorded. All samples were collected in triplicate (i.e. $n = 3$), except in 2006 ($n = 2$) and 2009 ($n = 5$). Sediments were retained from each sample for the determination of grain size

distribution (dried and analysed gravimetrically for size class fractions from silt–clay through to gravel), organic matter content (%OM measured as% ash free dry weight; Luczak et al., 1997) and macrofaunal community composition. Sediment grain size and%OM measures were determined from sub-samples collected using a 5.5 cm diameter Perspex core, with the surface 30 mm kept for analysis. Macrofauna sub-samples were collected using a 130 mm diameter (0.0132 m^2) core (100 mm sediment depth).
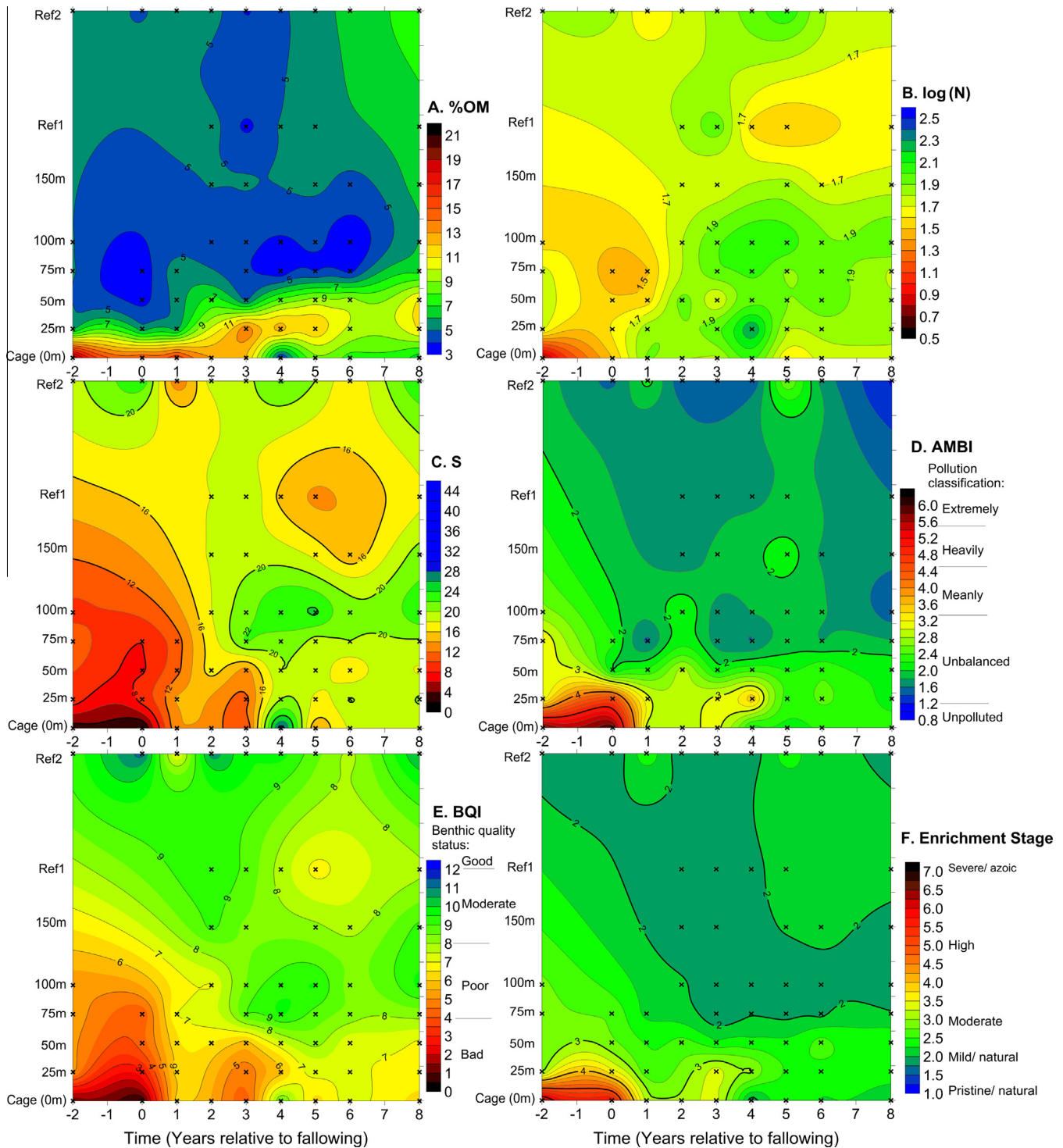

Fig. 2. Changes in%OM and log(N + 1), S, AMBI, BQI and ES with space and time at for the study site. Time is given in years relative to year of fallowing (=T0) and space represents distance (m) from the cages. AMBI index – high values indicate a more polluted status (Borja et al., 2000); BQI index – high values indicate high benthic quality (Rosenberg, 2008); ES – lower values reflect lower levels of enrichment (Keeley et al., 2012a,b). Crosses denote sampling events.

Macrofauna were sorted and enumerated to the lowest practicable level and their abundances recorded. Macrofauna count data were used to calculate total abundance (N), number of taxa (S), Pielou's evenness (J'), Shannon diversity (H') and the AZTI's Marine Biotic Index (AMBI, Borja et al., 2000), Benthic Quality Index (BQI, Rosenberg et al., 2004) and Multivariate-AMBI (M-AMBI, Muxika et al., 2007). Qualitative assessments of sediment odour, Beggiatoa (bacterial mat) coverage and sediment out-gassing using pre-specified categories (as described in Keeley et al., 2012b) were also made at each station.

2.2. Data analyses

2.2.1. Indicators of recovery

Pearson correlation coefficients between variables were determined from replicate-level data, with a $\log_{10}(x + 1)$ transformation applied to N. As several of the indicators were highly correlated, a subset of variables was selected for further analysis. The selection was based on their: potential to provide complementary information, weak correlation with other variables, representation of different components of the benthos, and/or different levels of analytical complexity (e.g. biotic indices). The chosen indicators were: %OM, $\log_{10}(N + 1)$, S, AMBI, and BQI, as well as overall enrichment stage (ES). ES is a numerical derivative of all of the physico-chemical and biological variables combined, and provides a robust indication of overall impact status (Keeley et al., 2012a). Although BQI and AMBI were highly correlated, both were retained because they have previously been found to be particularly good indicators of enrichment (Keeley et al., 2012a and references therein), and are computationally quite different and utilise different pollution tolerance classification schemes (Borja et al., 2000; Rosenberg et al., 2004). Furthermore, there are defined values for BQI and AMBI which specifically relate to levels of benthic quality and pollution status (Borja et al., 2000; Rosenberg et al., 2004); these classifications are shown in Fig. 2.

Spatial and temporal relationships for individual parameters were interpolated into a grid using the Kriging method in Surfer 9, and displayed in 2-dimensional plots. The x-axis represents 'Time' in years relative to fallowing (T0) and the y-axis represents 'Space' as distance from cages (m).

2.2.2. Multivariate analysis

Macrofaunal community data were analysed using PRIMER 6 (Clarke, 2006). Data were square-root transformed to reduce the influence of the highly abundant taxa and then averaged at the station-level. Multi-dimensional scaling (MDS) was used to display the Bray–Curtis similarities (zero adjusted with dummy variable due to some samples containing very few individuals, Clarke et al., 2006) between Cage, 25 m, 75 m and Reference stations over time (all years: T-2 to T8). One-way, single factor ANOSIM (Clarke et al., 2006) was also conducted on replicate-level data to obtain a pair-wise assessment of statistical differences among sites at each point in time (i.e. for each of T-2 to T8).

Species succession was described by using the SIMPER procedure to identify those taxa which contributed most to Bray–Curtis similarities within sample groups for each survey; groups consisted of sample replicates pooled within Cage or References stations. Bubble plots were generated to display temporal patterns in specific taxa using the ggplot() function in the ggplot2 library in R, where the x-axis represents time (in years) and the y-axis represents individual species. The colour gradient of the symbols indicates the average (%) contribution to the groups' similarity (from SIMPER analysis), and bubble size indicates abundance (square-root transformed) at the given time. Taxa shown are restricted to those which contributed to the top 90% of the similarity. The y-axis (i.e. taxa) was sorted (from top to bottom) according to the sequential contribution of those taxa to the groups' similarity, starting with T0 and progressing to T8. This creates a gradient of species succession; with those species that played an important role early in the recovery process placed at the top of the plot and those that contributed in the later stages toward the bottom. Taxa-specific 'Eco-Groups' (EG) are displayed where available, which are established classifications for species sensitivity to organic enrichment that are used to calculate AMBI; these groups range from 'I' (very sensitive) to 'V' (first order opportunistic taxa) (Refs. Borja et al., 2000; Keeley et al., 2012b and http://ambi.azti.es/).

Macrofauna data were also analysed according to functional feeding groups ('FFG'; deposit feeders, filter feeders, suspension feeders, scavengers, carnivores, omnivores and grazers), and higher level taxonomic groups ('HLT'; e.g. higher level Order, Class or Phyla), with the exception of polychaetes, which were separated into two groups based on mobility: sedentary and errant.

2.2.3. Assessment of recovery

Recovery level was determined by comparing the environmental parameters from the Cage and Reference stations using six different approaches, with varying levels of complexity (Table 1). Methods 1 and 2 involved plotting and visually comparing point in time estimates against background conditions. The mean values (with standard error) for all stations and times were overlaid with point-in-time 95% confidence intervals for Reference stations as well as the natural range of conditions that was encountered over

Table 1
Definitions of the six different approaches used to evaluate recovery.

Method	Definition
1. Long-term background range	The point in time that the mean value for the Cage stations first falls within the overall background range of conditions at the Reference stations and remains there for the remainder of the study
2. Point-in-time background range	The point in time that the standard error bars for the Cage stations first overlap with the point-in-time percentiles for the Reference stations and remains there for two or more consecutive years
3. Point-in-time ANOVA/PERMANOVA	The first point at which the Cage stations are considered statistically comparable ($P < 0.05$) to the Reference stations using nested ANOVA or PERMANOVA (Factors: Treatment, Station(Treatment))
4. Parallelism- forward stepping	The mid-point of the first time window for which a non-significant ($P > 0.05$) interaction term is obtained, moving forward in time (from T0 to T8). Conducted for different length time windows
5. Parallelism- backward stepping	The mid-point of the last time window for which a non-significant ($P > 0.05$) interaction term is obtained, moving backward in time (from T8 to T0). Conducted for different length time windows
6. Biological and chemical remediation (from Brooks et al., 2003)	Chemical: "the reduction of accumulated organic matter with a concomitant decrease in free sediment sulphide and an increase in sediment redox potential under and adjacent to salmon farms to levels at which more than half the reference area taxa can recruit and survive" Biological: "the restructuring of the infaunal community to include those taxa whose individual abundance equals or exceeds 1% of the total abundance at a local reference station. Recruitment of rare species representing < 1% of the reference abundance is not considered necessary for complete biological remediation"

the course of the study, based on the 5th and 95th percentiles for all Reference station data.

Method 3 involved simple point-in-time statistical comparisons using nested models to assess differences between the Cage and Reference stations (factor: 'Treatment'), where Station was a random factor nested within Treatment. Equivalent models were constructed for univariate and multivariate analyses, the former using the 'aov()' function in R (R Team, 2011) and the latter using the ANOSIM procedure.

Methods 4 and 5 used a test for 'parallelism' (forward and backward stepping) following the methods of Wiens and Parker (1995) and more recently Skalski et al. (2001); also known as a level-by-time interaction. Parallelism assumes that after impact (in this case organic enrichment), control and impact profiles converge over time and eventually track (or parallel) each other as impacted sites begin to respond solely to the same regional climatic changes or oceanographic conditions as the reference sites. Hence, parallelism between mean profiles for (in this case) the impacted and Reference stations provides inferential evidence of recovery (Skalski, 1995; Skalski et al., 2001). Population or community level differences between control and impacted sites are not considered in assessing recovery in this manner, only the relative patterns of the temporal trends (Skalski et al., 2001).

Parallelism analysis requires data to be analysed on a scale where natural differences between sites and temporal effects have an additive effect on population levels (Skalski et al., 2001). Where the strength of the response varies greatly, such as in population data, a transformation will be required to reduce any differences in amplitude; this also emphasises the trends such that tests for parallelism can be applied. Animal abundance (N) was the only variable log-transformed; all other variables responded normally and/or the scale for change was constrained. A nested linear mixed effects model permitting random slope and intercept was constructed using the lmer() function in the nlme library in R (Zuur et al., 2009), where year ('Ye') was treated as a continuous variable, treatment ('Tr', Cages versus Refs) as a fixed factor, and station ('St') as a random factor. As such, the 'Ye × St (Tr)' interaction became the test for parallelism. The test was applied to a reduced time series, for example three consecutive surveys out of the 10 year dataset, starting with the first or last year sampled. If the interaction term was non-significant then the window was moved forward (if forward stepping), or back (if backward stepping) one year, and the test repeated (Table 1). The resulting P-value was overlaid on scatterplots of the environmental variables, using horizontal bars to display the windows over which the tests apply, and to assist with visualisation of the results.

A comparable multivariate model was also constructed using Permutational analysis of variance (PERMANOVA + for PRIMER, Anderson et al., 2008), this approach tested for recovery in square-root transformed macrofauna composition data (MCD), FFG, HLT count data, and also the collective influence of all the univariate environmental variables combined ('All Vars'). All Vars analysis was undertaken using Euclidean distances, with data first normalised ((x – SD)/SD) to account for differing scales and arbitrary origins (Clarke, 2006). Differences between Cage and Reference stations through time were tested using a three-factor nested repeated measures design: factor 1 = Year ('Ye', 8 levels, fixed), factor 2 = Treatment ('Tr', 2 levels, fixed), factor 3 = Station(Treatment) ('St(Tr)', 4 levels, random). Significant terms were further investigated where required using a posteriori pairwise comparisons with 9999 permutations. Type I SS (sequential) were used, as some of the subsets were unbalanced. Further analysis of the components of variation for terms of interest was undertaken by calculating the distance among centroids in PERMANOVA and plotting the resulting matrix using principle coordinates (PCO, Anderson et al., 2008).

3. Results

3.1. Spatial and temporal patterns in indicator variables

Prior to fallowing (i.e. T-2 to T0), the seabed beneath the cages was highly impacted (Fig. 2). Organic matter content was markedly elevated (%OM 15–20%, Fig. 2a) and the macrofaunal community was impoverished with few enrichment tolerant taxa remaining (predominantly *Capitella capitata*) (Fig. 2b and c). The BQI was low (0.5–1.5), and the AMBI was relatively high (3.4–5.8), indicative of "bad benthic quality" and "poor ecological" quality statuses respectively (Fig. 2d and e). Total abundance (N) was reduced with an average of 5 individuals at T-2 and 31–40 individuals at T0, compared with a range of 40–120 individuals at the Reference sites (Fig. 2b). Average ES at T0 was 5.6 indicating highly enriched overall conditions (Keeley et al., 2012a). Seabed impacts were highly localised, with a marked reduction in%OM to near-background levels (4–7%) within ∼25 m of the cages (Fig. 2a). Macrofauna composition improved markedly within the first 25 m, continuing to improve with increasing distance from the Cage stations. Nonetheless, an effect was still clearly evident (low S, H′ and BQI) ⩾100 m from the cages.

One year after fallowing,%OM at the fallowed Cage stations was approximately 20% lower (%OM = 12%) than that observed at T-2. Over the same time period, there was considerable improvement in the biological indicators: S increased from approximately 3–16 taxa per core, and the AMBI and BQI biotic indices indicated an improvement to an 'unbalanced' or 'meanly' polluted state and to poor benthic quality, respectively (Fig. 2d and e). According to the main biological indicators (i.e. log(N), S, AMBI, BQI) recovery over the next two years (T2 and T3) was negligible, although%OM continued to decrease to around 10%. In the fourth year of recovery, most of the indicators (%OM, S, AMBI, M-AMBI) showed further substantial improvement at the fallowed Cage stations, achieving levels comparable to those found at the Reference stations. An exception was the BQI, which although improved (to BQI 8), was not yet comparable to background levels (BQI 9–10). At T5 a slight deterioration was evident, particularly in S and BQI. However, between surveys T6 and T8, all results (except BQI) suggested that conditions were similar to the Reference stations. BQI scores continued to indicate an impacted state; an average of ∼7 at the Fallowed-Cage stations compared with 8–11 at the Reference stations (Fig. 2e).

Unexpectedly, there was an apparent increase in enrichment at the Gradient stations (especially 25 m) midway through the study, mainly associated with the distribution of%OM. Initially the peak in%OM was at the Cage stations, but after 3 years (T3) this peak had shifted outwards to the 25 m station, where it remained for the following 4 years (Fig. 2). An increase in%OM was also evident at the 50 m station at T5, but to a lesser degree. The biology appeared to follow a similar temporal and spatial response pattern; at T4 there was a peak in N at the 25 m station and a general increase in N and S at the 75 m and 100 m stations. Notably, S also peaked temporarily at both the 100 m and Fallowed-Cage stations at T4. The biotic indices responded similarly with an initial increase in AMBI, decreased M-AMBI and BQI at ∼50 m from the cages. AMBI and M-AMBI improved to levels comparable to the Reference stations at T4, but once again BQI remained elevated, suggesting that macrofaunal composition was still impacted.

3.2. Patterns in macrofaunal composition

Multivariate analysis of the macrofaunal count data for all stations and times suggests a progressive convergence with time (T0–T8, moving from right to left on Fig. 3), as the fallowed Cage and near-cage (25 m) samples became increasingly similar to the
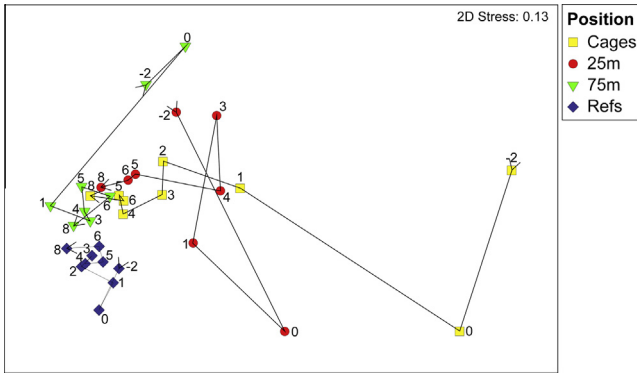
Fig. 3. MDS ordination of time-series data (T-2 to T8, T0 = year of fallowing) for Cage, 25 m, 75 m and Reference stations, based on Bray–Curtis similarities of station-averaged, square-root transformed macrofauna count data.

Reference samples. Differences between years at the Reference stations were comparatively small. Recovery in the macrofauna composition at the Cage stations was greatest in the first year after fallowing (T0–T1 R Statistic = 0.49, $P < 0.01$, Fig. 3). The magnitude of recovery diminished in each subsequent year until T4, after which changes in community composition were relatively minor.

Considerable recovery was also evident at the 25 m station between T0 and T1 (Fig. 3). However, the changes in the community observed at T3 and T4 did not follow the expected recovery trajectory, as the community became more dissimilar to the Reference stations than that observed at T1. Substantial recovery occurred between T4 and T5, with the 25 m stations becoming comparable to Reference stations (T5, 25 m – Reference, R Statistic = 0.43, $P = 0.13$). Macrofaunal composition at the 75 m station was initially significantly different from both the Cage and Reference stations (i.e. at T-2, R Statistics > 0.91, $P = 0.1$), but comparable to the 25 m station (R Statistic = 0.14, $P = 0.8$). There was substantial recovery at the 75 m station in the first year (T0–T1), but minimal change thereafter (Fig. 3), as the community was comparable to the Reference stations from T1 onwards (R Statistic < 0.6, $P > 0.1$).

The difference between T0 and T1 at the fallowed Cage stations was mostly due to a shift from a community dominated by *Capitella capitata* (a first-order opportunist, Eco-Group V) to one dominated by Eco-Group IV (second-order opportunists: dorvilleid polychaetes, nematodes) and enrichment tolerant taxa (*Theora lubrica*, amphipods, and *Arthritica bifurca*; EcoGroup III, Fig. 4). Abundances of *C. capitata* decreased to the point of not being a significant component of the assemblage at T2 and then disappeared altogether at T4. Nematodes decreased in abundance at
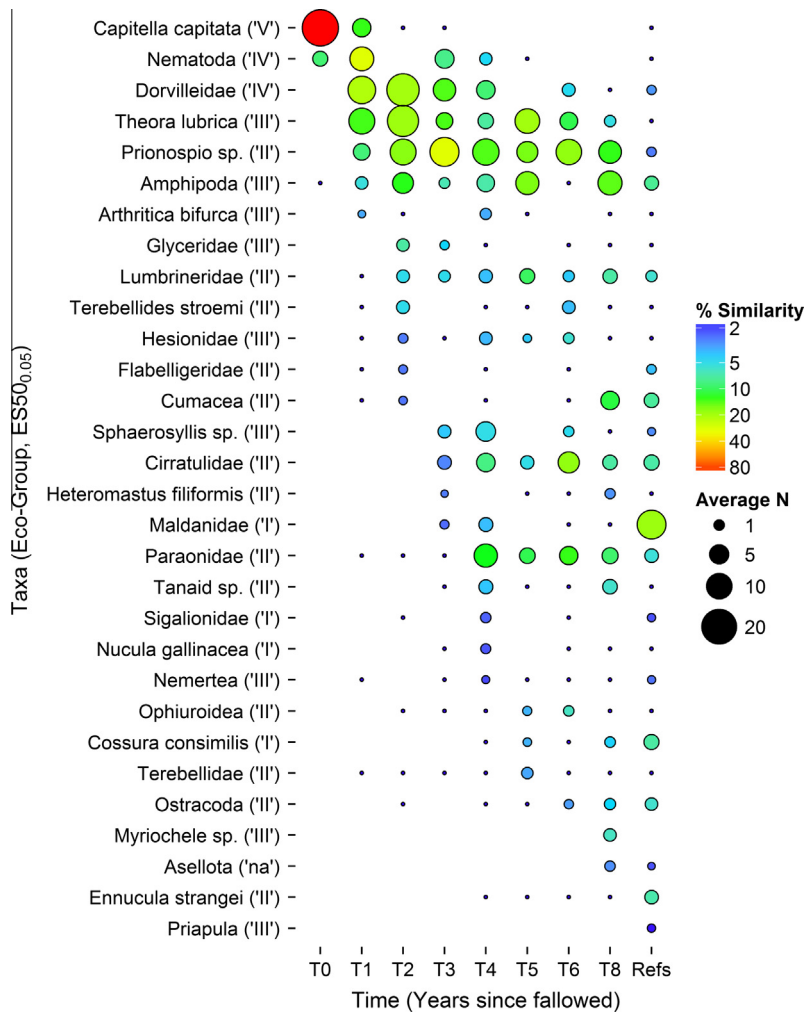

Fig. 4. Bubble plot summarising SIMPER analysis results of macrofauna count data at former Cage stations (T0–T8) and Refs at T8. Species shown are those that contributed most to the top 90% of the groups' similarity and are sorted (from top to bottom) according to their relative contribution (indicated by colour gradient). Bubble size indicates (square-root of) average total abundance (N). Bracketed values indicate previously established Eco-Group (Borja et al., 2000) values for each taxon. Small blue dots indicate taxa that were present, but did not contribute significantly to the similarity.

T2, but increased at T3, before again declining markedly and disappearing at T6. Second-order opportunists and enrichment tolerant taxa (*T. lubrica*, dorvilleid polychaetes, *Prionospio* sp. and amphipods) all continued to increase in abundance at T2 and T3. *T. lubrica* and *Prionospio* sp. remained dominant taxa through to the conclusion of the study (T8), whereas the importance of dorvilleid polychaetes diminished at T5 (Fig. 4). Several Eco-Group II and III taxa, including polychaetes (belonging to the Families Glyceridae, Lumbrineridae, Hesionidae, Flabelligeridae and Trichobranchidae) and cumaceans, were important at T2. The numerical importance of most of these taxa was short lived, with the exception of lumbrinerid polychaetes, which were important contributors to the similarity of the assemblages throughout.

While the abundance of many early colonizers decreased at T3, cirratulid polychaetes, *Sphaerosyllis* sp., maldanid polychaetes and *Heteromastus filliformis* became important contributors for the first time (Fig. 4). Cirratulids, *Sphaerosyllis* sp. and maldanids (an Eco-Group I taxon) continued to increase in abundance at T4, at which point paraonid polychaetes, *Tanaid* sp., sigalionid polychaetes and *Nucula gallinacea* (an EcoGroup I bivalve) became notable components of the assemblage for the first time in the recovery phase,

with cirratulids and paraonids being dominant taxa for the remainder of the study. *Cossura consimilis*, brittle stars (Ophiuroidea) and terebellid polychaetes all featured prominently late in the recovery process (T5 and T6), and are taxa which are either considered sensitive, or indifferent to enrichment. *Myriochele* sp. (an EcoGroup III taxon) and isopods (Asselota) became prevalent for the first time at T8.

A relatively small shift in background or 'natural' conditions was observed in the macrofauna count data between T0 and T8 (Fig. 3), principally due (in reducing order of importance) to reduced abundances of: ophiuroids, *Nucula nitidula*, *T. lubrica*, Spionidae, *Cadulus teliger*, *Echinocardium cordatum* and *Neilo australis*, and increased abundances of paraonids and *C. consimilis*.

Multivariate analysis of FFG data showed substantial recovery in the first year, followed by a high degree of temporal (inter-annual) variability, due to fluctuations in the relative abundances of scavengers, omnivores and carnivores (Fig. 5a). The temporal changes in HLT groups were largest in the first 2 years, as the community shifted from being highly dominated by sedentary polychaetes and nematodes, to one with a more balanced assemblage of errant polychaetes, crustaceans, bivalves and amphipods (Fig. 5b). At the
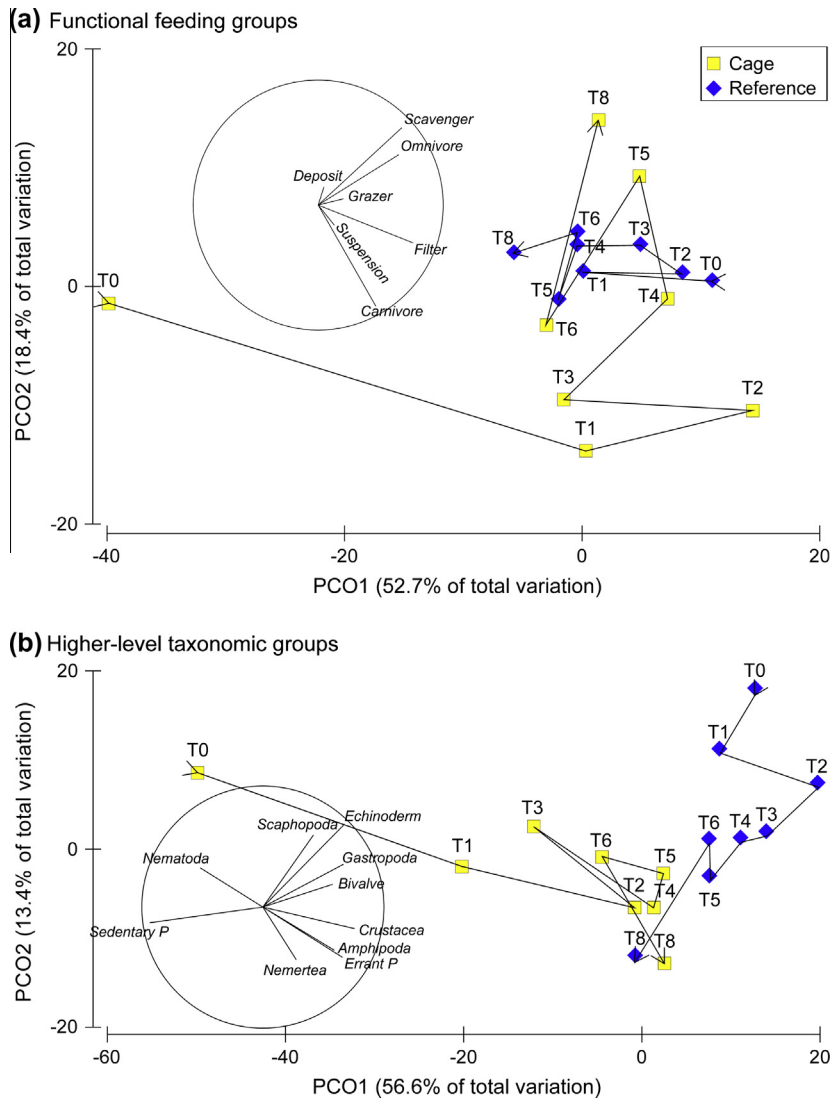

Fig. 5. PCO ordination of Distance among Centroids for Cage and Reference stations from T0 to T8, based on Bray-Curtis similarities (+d) of square-root transformed macrofauna count data grouped according to (a) functional feeding groups and (b) higher-level taxonomic groups. Overlaid with Pearson correlation vectors to indicate main drivers of differences in 2-dimensional space.

conclusion of the study (T8), the HLT composition of macrofauna at the fallowed Cage and Reference stations was very similar. Although the composition of the Reference station taxa also changed through time, the direction was different to that of the main recovery pathway, and was mainly associated with a reduction in numbers of echinoderms, gastropods, scaphopods and bivalves between T0 and T8 (Fig. 5b). At a taxa level, the differences at T8 were mostly due to higher abundances of *Myriochele* sp. *Tanaid* sp., *T. lubrica*, *Prionospio aucklandica* and *H. filiformis*, and lower numbers of *Ennucula strangei*, maldanid and flabelligerid polychaetes and Priapulid worms at the fallowed Cage stations (Fig. 4). Of the 30 dominant taxa identified by the SIMPER analysis (Fig. 4), only six were not shared in common between the Cages and reference stations at T8; interestingly, two of these were first order opportunists that were present only at the References stations.

3.3. Assessment against recovery criteria

The variables that most consistently depicted recovery were S, BQI, ES and *All Vars* (all variables assessed using multivariate analyses, Table 2). Recovery according to these variables was generally shown to have been achieved after 4 years; however, the range was greater for S and ES (2–6 years) and for *All Vars* (2–4.5 years). AMBI and MCD also indicated time lengths between 4 and 5.5 years, but also that recovery had not been achieved in some instances (i.e.>5.5 years). The most commonly occurring time estimate (indicated by the mode) across all variables was 4–5.5 years. Note that because the midpoint of the time window was used as the 'recovery point' for the parallelism method, the maximum recovery timeline was effectively 5.5, 6 and 6.5 years for the 3, 4 and 5 year windows, respectively.

In terms of method-specific differences, the background range methods and the point in time ANOVA/ PERMANOVA tests were reasonably consistent in indicating that recovery had occurred after 4 years (Table 2). The most common recovery time estimates for the forward stepping parallelism method was 2, 4.5 and 5.5 years for the 3-, 4- and 5-year windows respectively. It is apparent that the 2 year estimate from the 3-year time windows indicated early parallelism, yet recovery had clearly not occurred (Figs. 6–8). Despite the temporal responses being similar over this period, the indicator levels remained substantially different, and the subsequent time window (T2–T5) indicated further significant differences between Reference and fallowed Cage stations. When the slightly longer 4-year window was used to test for parallelism, the recovery that had been apparent during T1–T3 using the 3-year window was no longer evident (Table 2, Figs. 6–8).

The most common time estimates using the backward stepping method was >5.5 years, suggesting recovery had not occurred during the study period. However, periods of parallelism were evident earlier in the time-line, highlighting a possible issue with the backward stepping approach. In the case of the AMBI and BQI indices, this was due to a small divergence and reduced sample variability, increasing statistical power. This divergence was mainly due to improving conditions at the control stations (decreasing AMBI and increasing BQI), while the conditions at the fallowed Cage stations remained relatively static (Fig. 7b and 8a). Similarly, although significant differences were still sometimes evident in the last time window for%OM, levels of%OM were lower at the Cage stations than at the Reference stations, consistent with a less enriched state at the former farm site.

In several instances, the main term (either 'Tr', in the case of the point-in-time analyses, or the 'Ye × Tr' interaction for parallelism) was not significant early on, but the term involving Station nested within Treatment (i.e. 'St(Tr)', or 'Ye × St(Tr)') was highly significant (e.g. Appendix Figs. S1 and 2, Tables S1 and 3). This was usually due to a large amount of variation in the factor St(Tr) masking differences in the main term. Further analysis of the components of variation for that term showed that the Cage Stations still changed substantially more than the reference Stations, but that the rate at which the Cage stations were changing was often different (e.g. Appendix S1–S3). Unfortunately, recovery was greater at one station and as a consequence the main interaction term was not significant. In these instances discretion was used to select the point most indicative of recovery.

Total abundance (log transformed) generally indicated recovery early on in the study (mode = 2 years) suggesting that the fallowed Cage stations were comparable to the Reference stations at T0, despite obvious differences in many other variables (Table 2). FFG also appears to suggest a relatively rapid recovery (only 1 year), but was inconsistent with assessments for later years suggesting recovery was not complete at the conclusion of the study.

In terms of biological remediation according to the Brooks et al. (2003) criterion (Method 6 in Table 1), a total of 19 dominant Reference station taxa (defined as those whose individual abundance equals or exceeds 1% of the total abundance) were identified. These were amphipods, polychaete worms (*Cossura consimilis*, *Prionospio* sp., *Sphaerosyllis* sp., dorvilleids, cirratulids, hesionids, lumbrinerids, maldanids, paraonids, sigalionids,), nematodes, cumaceans, ophiuroids, ostracods, priapulid worms and small bivalves (*Theora lubrica*, *Nucula gallinacean* and *Ennucula strangei*). At T-2 and T0, only 2.6% and 4.4% of these taxa (respectively) were present at the Cage stations. Substantial recolonization occurred at T1 and T2, with 26% and 41%, respectively, of these taxa present. At T4

Table 2

Summary of recovery estimates for indicator variables and multivariate analyses based on five different methods. **'Ye × Tr' non-significant at an earlier time, but 'Ye × St(Tr)' remained significant due to large between-cage Station differences (*Appendix* Tables S1, S3). Temporal difference at fallowed-Cage stations still considerably larger than for Reference stations until specified time (see Appendix).

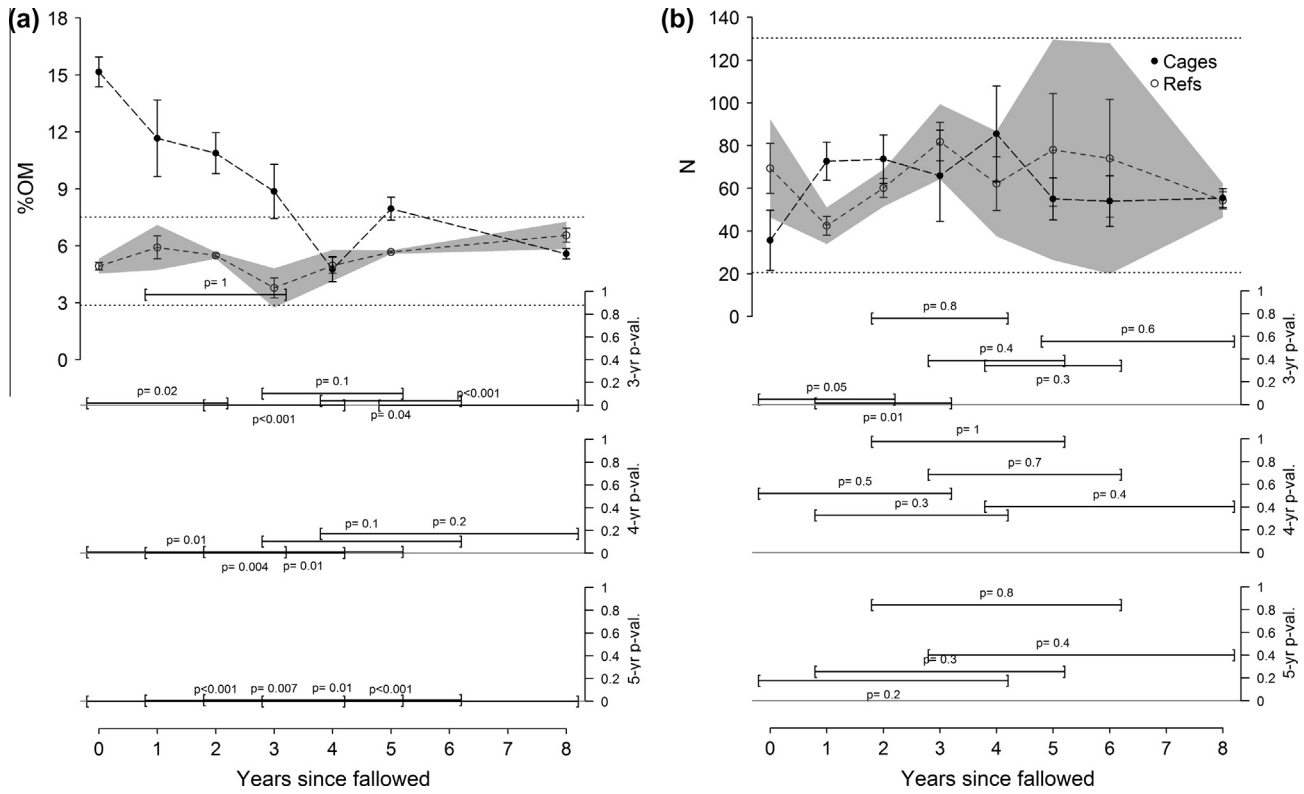| Method	Univariate tests						Multivariate tests				MODE
	%OM	log(N)	S	AMBI	BQI	ES	All Vars	MCD	FFG	HLT	
1. Background range	6–8	0	1	4	4	4	–	–	–	–	4
2. Point-in-time background	6–8	2	4	>8	4	4	–	–	–	–	4
3. Point-in-time ANOVA/PERMANOVA	4	0	4	5	4	4	4*	4*	1	4*	4
4. Parallelism – forward stepping											
3-year windows	2	3	2	5	2	2	3*	5*	1	1	2
4-year windows	4.5	1.5	3.5	>6	>6	6	4.5*	4.5*	1	4.5*	4.5
5-year windows	>5.5	2	4	5.5	5.5	>5.5	4*	>5.5	4*	>5.5	5.5
5. Parallelism – backward stepping											
3-year windows	>6.5	2	4	>6.5	>6.5	5	4*	5*	>6.5	5*	>6.5
4-year windows	4.5	1.5	>6	>6	>6	6	3*	4.5*	>6	4.5*	>6
5-year windows	>5.5	2	4	5.5	5.5	>5.5	2*	>5.5	>5.5	>5.5	>5.5
MODE	4.5	2	4	5–6	4	4	4	4–5.5	1	4.5– > 5.5	

Fig. 6. Comparison of temporal profiles of Cage and Reference stations for average: (a)%OM and (b) N (note, analyses conducted on log transformed data), with corresponding *p*-values indicating results of test for parallelism (i.e. Ye × Tr term) for 3, 4 and 5 year time windows (indicated by horizontal bars on bottom three *y*-axes). *Y*-axes correspond to specified *p*-value. Vertical error bars on main plots represent SE. Grey shaded area represents point-in-time 95% CI for Reference stations, and horizontal dashed lines indicate the range of background conditions over the entire study (5th and 95th percentiles for Reference Stations).
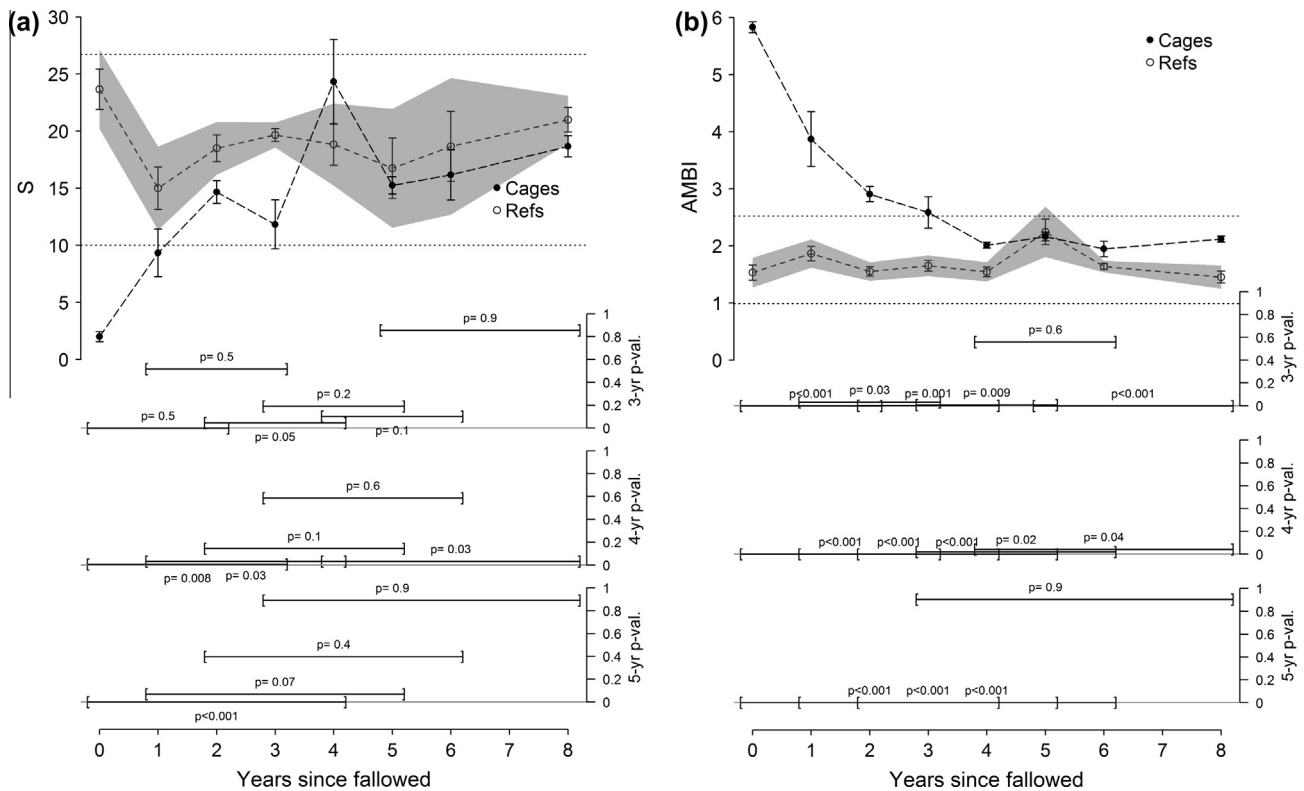

Fig. 7. As in Fig. 6, but for (a) S and × (b) AMBI.
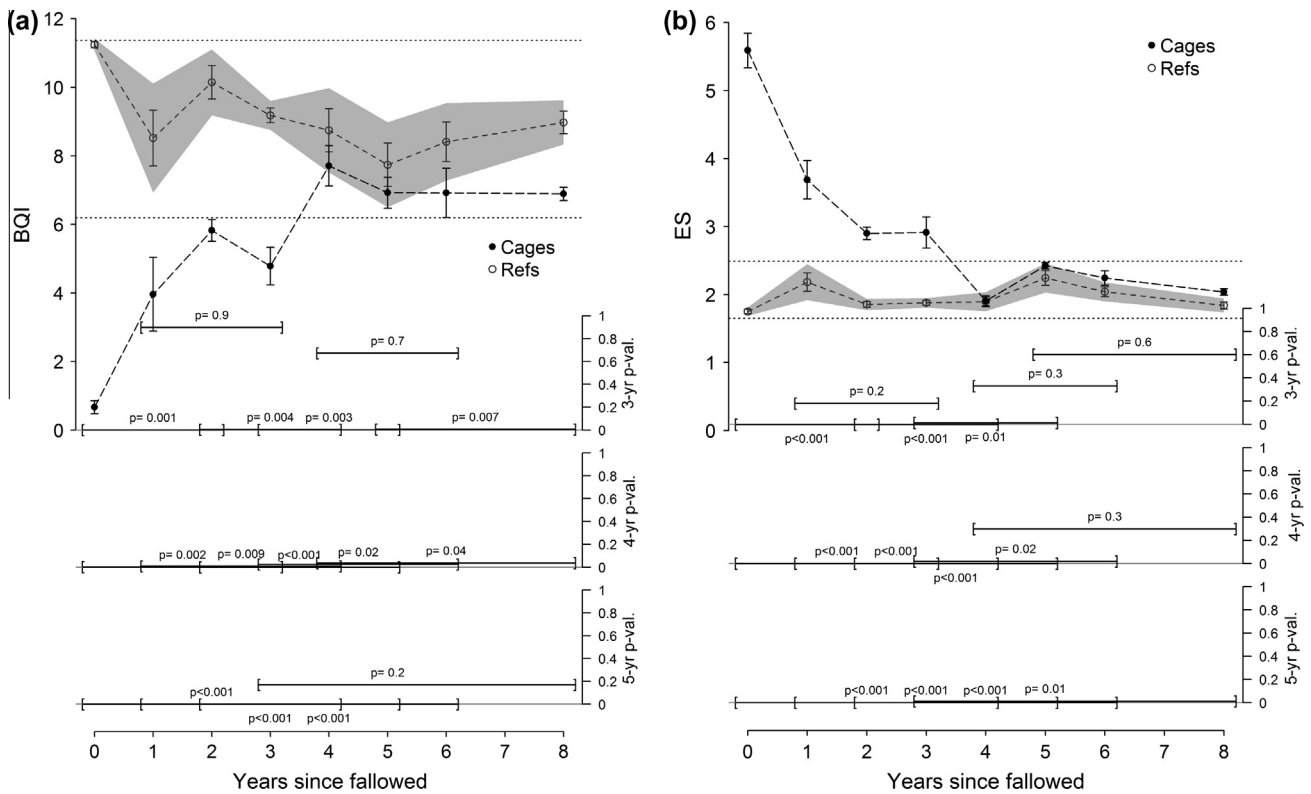
Fig. 8. As in Fig. 6, but for (a) BQI and (b) ES.

the level of recolonization peaked, with 69% of the dominant Reference station taxa being present at the Cage stations. This declined back to 48% at T5 and remained around 50% (±5%) through until the conclusion of the study. If Cage and Reference stations are compared based on the presence-absence of all shared taxa (i.e. irrespective of the 1% threshold), a similar picture of remediation emerges, only with a slightly higher proportion of common taxa. This comparison shows that at T1, ~35% of the taxa were common between Cage and Reference stations, and by T2, ~57% were shared, which met that aspect of the chemical remediation criteria in Method 6 of Table 1.

4. Discussion

4.1. Recovery timeframes

The 10 year duration of this study allowed for an evaluation of relatively long-term (i.e. 8 years) recovery processes and interactions, which is not possible with shorter-term assessments where recovery has often not been realised (Karakassis et al., 1999; Lin and Bailey-Brock, 2008; Macleod et al., 2008; Villnas et al., 2011). Substantial improvement in seabed health beneath the former salmon farm site was observed in the first 2 years, followed by more gradual and variable-dependent improvements over the following 2–3 years. A weight-of-evidence assessment suggested that 'recovery' had ostensibly been achieved after 4–5.5 years. At this point sediment conditions were, in many respects, comparable to the Reference stations however, significant differences were still evident in some environmental indicators. This is consistent with previous estimates for recovery of macroinvertebrates from organic enrichment (i.e. from 2 to ~7 years), but short when compared to other impact sources and longer-lived, slower-turnover, biological components (Borja et al., 2010). The fact that recovery occurred in years rather than decades from what was a highly impacted state in a low flow area is significant, and reinforces

previous assertions that salmon farming is unlikely to have long-term adverse benthic impacts (e.g. Lu and Wu, 1998).

Chemical remediation, as defined by Brooks et al. (2003), requires significant improvements in %OM and sediment chemistry (sulphides and redox) such that more than half the reference taxa can recruit and survive, and this consistently occurs earlier in the process than biological remediation. Although no chemical data were available in our study, %OM had clearly declined at the former cage site, and the associated biological criterion was achieved after two years. This is a relatively long period compared with other fish farm related studies (Brooks et al., 2003; Ritz et al., 1989), which reported similar remediation within a few weeks to 6 months, but was much shorter than the 5.4 years estimated for a biologically atypical site where the macrofauna was dominated by bivalves (Brooks et al., 2004). Bivalves are typically suspension or deposit feeders and generally considered sensitive to enrichment (Borja et al., 2000; Pearson and Rosenberg, 1978). However, it is important to note that this atypical community also lacked opportunistic polychaetes, which are critical to the recovery process (Macleod et al., 2007).

One definition of biological remediation requires the complete re-establishment of the dominant taxa (>1% by abundance) at reference sites (Brooks et al., 2004). In the present study only 68% of the dominant Reference taxa had re-established after four years, and the level subsequently reduced with only ~50% re-establishment at the end of the study. Hence, alternate definitions of recovery proposed by Brooks et al. (2003, 2004) can be applied at differing timescales, and comparisons with the current data clearly shows that estimates of the timeframe for recovery not only vary markedly based on these definitions, but can also differ according to the local ecology.

4.2. Indicators of recovery

In terms of notable responses of individual taxa, the approximate point of 'recovery' (i.e. ~5 years) identified here coincided

with large reductions in the abundance of nematodes and dorvilleid polychaetes. These enrichment tolerant taxa are prevalent under moderate-to-high levels of enrichment (ES 3– 5, Keeley et al., 2012b) and therefore their substantive temporal decline during recovery may be a useful indicator of biological remediation. At the same time, several other taxa became established as important components of the macrofauna (ophiuroids, the polychaete *Cossura consimilus*, members of the polychaete family Terebellidae, ostracods and the bivalve *Ennucula strangei*). Four of these are listed on the AMBI database (http://ambi.azti.es/) as EG I or II taxa and hence appear to be good 'universal' indicators of unimpacted conditions (Borja et al., 2000) and recovery. There were also other taxa that were central components of the macrofauna in the initial phases, but then remained dominant throughout the recovery process and hence would not be useful indicators of recovery stage (e.g. the bivalve mollusc *Theora lubrica*, the polychaete *Prionospio* sp. and members of the polychaete family Lumbrineridae).

Unsurprisingly, the analyses highlighted the important role of *Capitella capitata* during moderate to severe enrichment (ES4 to 6) in the early stages of recovery (i.e. the first two years). However, it was interesting that despite being the dominant species in these early stages, there was not a strong peak in total abundance (N) as has been so frequently reported in association with benthic impacts (Hale and Heltshe, 2008; Pearson and Rosenberg, 1978; Rosenberg et al., 2004). Instead, N at the fallowed Cage stations remained statistically similar to the Reference stations throughout the study. Distinct differences in the proliferation of opportunists during recovery have been observed elsewhere, but the reasons for this remain unclear (Brooks et al., 2004). In the current study the lack of a strong peak in N is possibly an artefact of sampling timing and frequency, which was annual, whereas the 'peak of opportunists' can occur over the first few months of recovery (e.g. Pereira et al., 2004). This situation undermines the utility of N as an indicator of ecological succession during long-term recovery studies.

Consequently, although ecological differences are critical to defining change in response to organic enrichment, it is risky to depend upon a few specific indicator species or simplistic measures of ecological condition, particularly in the later stages of recovery. Beyond 3 years, impacts were less obvious and were mainly evident as compositional differences in the macrofauna and as a result, variables such as S tended to suggest recovery earlier on. The variables that were most consistent in their estimate of recovery tended to the more complex biotic indices, which take into account some aspect of the species identity or functional role (BQI and AMBI) and ES (which integrates across all variables – biotic and abiotic, Keeley et al., 2012a). The multivariate approach, which also integrated all variables, was similarly robust. This reinforces previous findings that Eco-Group based indices, and complex metrics like BQI and ES, are most useful for discerning enrichment effects associated with finfish aquaculture (Borja et al., 2009; Keeley et al., 2012a) and medium-term recovery states (Borja et al., 2006). The simplistic individual variables tended to be more susceptible to variable-specific and non-intuitive responses.

The detailed assessment of recovery here is based primarily on the state of the seabed at the fallowed Cage stations and it is conceivable that some of the variability observed may be an artefact of sampling and spatial variability rather than temporal changes *per se*. Field observations during the last survey indicated some residual small-scale patchiness at the Cage stations, with small pockets of blackened (anoxic) sediments amongst natural sediments. Although this may have contributed to individual sample variability, on most occasions triplicate samples were collected from triplicate Cage stations, which would have accounted for the small-scale patchiness, and provided a reasonable estimate of overall condition. However, the spatial analyses did suggest a larger-scale patchiness and apparent 'shifting' of the enrichment peak that may explain some of the inter-annual variability. The peak in %OM occurred outside of the historical Cage footprint (25–50 m away) after 3–4 years, where there was a corresponding biological response, resulting in a higher overall enrichment stage (ES). Interestingly, at the same time the total number of taxa increased at both the Cage station, and further away at the 100 m position. This observation is somewhat perplexing, as there was no known new source of external organic matter at the Gradient stations, and considerable care was taken to accurately relocate sampling stations. One possible explanation is that the predominant current flow in the direction of the Gradient stations gradually transported organic material in that direction, the result being a slowly migrating enrichment peak. Simultaneously, the elevated densities of opportunists, especially at the 25 m station, may have themselves contributed to the apparent increase in organic content of the sediment samples. Regardless of the cause, these reasonably large-scale spatial and temporal patterns in the benthos could be responsible for both the variability observed in some indicators at the Cage stations (e.g. functional feeding groups), and potentially some of the significant Ye × Tr interactions that occurred later in the study.

4.3. End-points and methods for assessing recovery

Compositional disparity between fallowed Cage and Reference station communities can be interpreted as a failure of the simplest criteria for recovery. However, this assumes both a steady-state and spatial equilibrium, and the related successional theory that there is only one 'climax' state, to which impacted communities will return; which is generally considered to be an overly simplistic view (Beisner et al., 2003; McCook, 1994; Parker and Wiens, 2005; Young et al., 2001). In this study, failure to converge on a similar endpoint was most evident in the AMBI, BQI and the multivariate analyses of the macrofauna assemblage. Such compositional differences in the latter stages of recovery are often attributable to the absence of late successional 'equilibrium' (or climax) species that tend to have slower re-colonisation rates (Whitlach et al., 2001), or may be excluded by early colonisers (Connel and Slatyer, 1977). However, in this case the differences observed at the conclusion of the study were mainly due to differences in the relative abundances of similar taxa, most of which had comparable ecological functions. Furthermore, the assumption of steady state equilibrium was clearly not valid in this instance as there was a shift in reference conditions over the course of the study, which is evident in both the macrofauna count data and the high-level taxonomic grouping analysis. Therefore, the concepts of a dynamic equilibrium (Skalski et al., 2001) and alternative stable states (Beisner et al., 2003; Borja et al., 2010) appear to be applicable in this particular assessment of recovery.

The point at which communities have a broadly equivalent faunal composition with similar functional roles (e.g. bioturbation, feeding and reproduction strategies) has been proposed as a useful reference point for recovery (Macleod et al., 2008). The critical aspect in this context being that the biological and ecological characteristics achieve a state from which the system has 'the capacity' to fully recover (Macleod et al., 2008). Multivariate analysis and tests for parallelism of the functional feeding groups in this study revealed some on-going differences between Cage and Reference stations. Closer analysis of the data revealed that although the functional composition of the macrofauna was reasonably similar to the reference communities after two years, there were significant interactions observed in the parallelism tests due to alternations in the relative dominance of two or three of the main feeding groups. Such compositional 'instability' can occur post-disturbance if the habitat is modified such that it favours

recolonisation by species other than those which previously existed (Connel and Slatyer, 1977). Thus ecosystem function appears to have been restored relatively quickly, but on-going compositional instability suggested a stable state had not yet been achieved (Connel and Slatyer, 1977); a situation that can be indicative of impacted sediments (Karakassis et al., 1999; Mendez and Linke-Gamenick, 2001).

Parallelism is one approach that can be used to assess recovery over the longer-term, and is particularly appropriate where communities may be subject to natural changes. The basic premise of the test for parallelism is that impact and control sites will begin to track, or "parallel", each other when the influence of the impact is no longer important and they are both responding, solely to wider environmental stimuli (Parker and Wiens, 2005; Skalski et al., 2001). Most of the individual indicator variables met this criterion over the 4 to 5.5 year range. However, several of the biotic indices failed the test when the last survey was included in the assessment window - particularly when the backward stepping method was applied. In the context of the greater recovery trajectory, this divergence in the last time window was usually relatively small, due to 'deterioration' at the Reference stations (as opposed to on-going improvements at the Cage Stations), and statistical significance reflected the fact that variability among the Cage samples was generally reduced in the final year. Hence, despite statistical significance, the patterns were not always ecologically significant or consistent with incomplete recovery. This highlights a weakness that exists in both the forward and backward stepping approaches. When the backward stepping approach is applied, parallelism is assumed to have never occurred, when in fact a period of similar responses may have occurred earlier in the study. Conversely, with the forward stepping procedure parallelism can be concluded prematurely.

Window size is an important consideration when applying the parallelism test as it represents a compromise between power (longer time windows have higher degrees of freedom for the error term) and temporal resolution (longer windows being less sensitive to localised deviations from parallelism) (Skalski et al., 2001). In this study, using longer (5 year) time windows increased the likelihood that the test would indicate that parallelism had not been achieved at the conclusion of the study (i.e. when including years 6 and 8). Whereas, the shortest 3-year window identified parallelism at an early stage for most variables, between years 1 and 3, but this was rejected as evidence of recovery because subsequent windows did not demonstrate parallelism and the full range of temporal plots indicated that such a finding was premature. Thus, longer time windows are less prone to falsely identifying parallelism. However, a disadvantage to using longer time windows is that it is more difficult to nominate a single year as being the point of 'recovery' and requires monitoring to be conducted for a longer period. The recommended solution is to combine parallelism and visual assessment methods by identifying all stages of parallelism and then selecting the first period during which all points within the window remain within the range of natural background variability. This approach is most reliable when applied to more complex biotic indices and metrics that unify multiple variables.

Clearly, there are numerous unresolved challenges to reliable determination of recovery. A fundamental issue is the lack of clarity and consensus around what constitutes 'recovery', and hence this subject requires further attention by scientists and environment managers. It is evident that recovery can be considered in terms of benthic chemistry, and both structural and functional attributes of the benthos, each of which will provide a different impression of the rate of recovery, and have a different interpretation in the context of baseline or reference site conditions. Even if a definition or end-point for recovery is agreed upon, determining whether recovery has been achieved is inherently complex and requires some degree of expert judgement. Therefore, for the foreseeable future, multiple indicators of seabed recovery should be monitored and a weight-of-evidence approach applied.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.marpolbul.2013.12.008.

References

Aguado-Giménez, F., Piedecausa, M.A., Gutiérrez, J.M., García-Charton, J.A., Belmonte, A., García-García, B., 2012. Benthic recovery after fish farming cessation: a "beyond-BACI" approach. Mar. Pollut. Bull. 64, 729–738.

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

Beisner, B.E., Haydon, D.T., Cuddington, K., 2003. Alternative stable states in ecology. Front. Ecol. Environ. 1, 376–392, Front Ecol Environ 2003, 1(7), pp. 376–382.

Borja, Á., Dauer, D., Elliott, M., Simenstad, C., 2010. Medium- and long-term recovery of estuarine and coastal ecosystems: patterns, rates and restoration effectiveness. Estuaries Coasts 33, 1249–1260.

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., Muxika, I., Franco, J., 2006. Long-term recovery of soft-bottom benthos following urban and industrial sewage treatment in the Nervión estuary (southern Bay of Biscay). Mar. Ecol. Prog. Ser. 313, 43–55.

Borja, A., Rodríguez, J.G., Black, K., Bodoy, A., Emblow, A., Fernandes, T.F., Forte, J., Karakassis, I., Muxika, I., Nickell, T.D., Papageorgiou, N., Pranovi, F., Sevastou, K., Tomassetti, P., Angel, D., 2009. Assessing the suitability of a range of benthic indices in the evaluation of environmental impact of fin and shellfish aquaculture located across Europe. Aquaculture 293, 231–240.

Brooks, K.M., Mahnken, C., Nash, C., 2002. Environmental effects associated with marine netpen waste with emphasis on salmon farming in the Pacific northwest. Responsible Mar. Aquacult., 159–203.

Brooks, K.M., Stierns, A.R., Backman, C., 2004. Seven year remediation study at the Carrie Bay Atlantic Salmon (Salmo salar) farm in the Broughton Archipelago, British Columbia, Canada. Aquaculture 239, 81–123.

Brooks, K.M., Stierns, A.R., Mahnken, C.V.W., Blackburn, D.B., 2003. Chemical and biological remediation of the benthos near Atlantic salmon farms. Aquaculture 219, 355–377.

Cardell, M.J., Sarda, R., Romero, J., 1999. Spatial changes in sublittoral soft-bottom polychaete assemblages due to river inputs and sewage discharges. Acta Oecologica-Int. J. Ecol. 20, 343–351.

Clarke, K.R., Somerfield, P.J., Chapman, M.G., 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. J. Exp. Mar. Biol. Ecol. 330, 55–80.

Clarke, K.R.G., R.N., 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E Ltd, Plymouth, UK.

Connel, H.H., Slatyer, R.O., 1977. Mechanisms of succession in natural communities and their role in community stability and organization. Am. Nat. 111, 1119–1144.

Diaz, R.J., Rosenberg, R., 2011. Introduction to Environmental and Economic Consequences of Hypoxia. Int. J. Water Resour. Dev. 27, 71–82.

Ellis, D.V., 2003. The concept of "sustainable ecological succession"; and its value in assessing the recovery of sediment seabed biodiversity from environmental impact. Mar. Pollut. Bull. 46, 39–41.

Gray, J.S., Waldichuk, M., Newton, A.J., Berry, R.J., Holden, A.V., Pearson, T.H., 1979. Pollution-induced changes in populations [and discussion]. philosophical transactions of the Royal Society of London, Series B. Biol. Sci. 286, 545–561.

Hale, S.S., Heltshe, J.F., 2008. Signals from the benthos: Development and evaluation of a benthic index for the nearshore Gulf of Maine. Ecol. Indicators 8, 338–350.

Hewit, J.E., Thrush, S., 2007. Effective long-term ecological monitoring using spatially and temporally nested sampling. Environ. Monit. Assess. 133, 295–307.

Hewitt, J.E., Pridmore, R.D., Thrush, S.F., Cummings, V.J., 1997. Assessing the short-term stability of spatial patterns of macrobenthos in a dynamic estuarine system. Limnol. Oceanogr. 42, 282–288.

Kalantzi, I., Karakassis, I., 2006. Benthic impacts of fish farming: meta-analysis of community and geochemical data. Mar. Pollut. Bull. 52, 484–493.

Karakassis, I., Hatziyanmi, E., Tsapakis, M., Plaiti, W., 1999. Benthic recovery following cessation of fish farming: a series of successes and catastrophes. Mar. Ecol.-Progr. Ser. 184, 205–218.

Karakassis, I., Tsapakis, M., Hatziyanmi, E., Papadopoulou, K.N., Plaiti, W., 2000. Impact of cage farming of fish on the seabed in three Mediterranean coastal areas. ICES J. Mar. Sci. 57, 1462–1471.

Keeley, N., Forrest, B., Crawford, C., Macleod, C., 2012a. Exploiting salmon farm benthic enrichment gradients to evaluate the regional performance of biotic indices and environmental indicators. Ecol. Indicators 23, 453–466.

Keeley, N., MacLeod, C., Forrest, B., 2012b. Combining best professional judgement and quantile regression splines to improve characterisation of macrofaunal responses to enrichment. Ecol. Indicators 12, 154–166.

Lin, D.T., Bailey-Brock, J.H., 2008. Partial recovery of infaunal communities during a fallow period at an open-ocean aquaculture. Mar. Ecol. Progr. Ser. 371, 65–72.

Lu, L., Wu, R.S.S., 1998. Recolonization and succession of marine macrobenthos in organic-enriched sediment deposited from fish farms. Environ. Pollut. 101, 241–251.

Luczak, C., Janquin, M.A., Kupka, A., 1997. Simple standard procedure for the routine determination of organic matter in marine sediment. Hydrobiologia 345, 87–94.

Macleod, C.K., Moltschaniwskyj, N.A., Crawford, C.M., 2008. Ecological and functional changes associated with long-term recovery from organic enrichment. Mar. Ecol. Progr. Ser. 365, 17–24.

Macleod, C.K., Moltschaniwskyj, N.A., Crawford, C.M., Forbes, S.E., 2007. Biological recovery from organic enrichment: some systems cope better than others. Mar. Ecol. Prog. Ser. 342, 41–53.

McCook, L.J., 1994. Understanding ecological community succession: Causal models and theories, a review. Vegetatio 110, 115–147.

Mendez, N., Linke-Gamenick, I., 2001. Sediment processing in Capitella spp. (Polychaeta: Capitellidae): strain-specific diffeneces and effects of the organic toxicant fluranthene. Mar. Biol. 138, 319.

Muxika, I., Borja, A., Bald, J., 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 16–29.

Parker, K.R., Wiens, J.A., 2005. Assessing recovery following environmental accidents: environmental variation, ecological assumptions, and strategies. Ecol. Appl. 15, 2037–2051.

Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanograph. Mar. Biol. Annu. Rev. 16, 229–311.

Pereira, P.M.F., Black, K., McLusky, D.S., Nickell, T.D., 2004. Recovery of sediments after cessation of marine fish farm production. Aquaculture 1–4, 330.

Team, R., R.D.C., 2011. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritz, D.A., Lewis, M.E., Shen, M., 1989. Response to organic enrichment of infaunal macrobenthic communities under salmonid seacages. Mar. Biol. 103, 211–214.

Rosenberg, R., Blomqvist, M., C Nilsson, H., Cederwall, H., Dimming, A., 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. Mar. Pollut. Bull. 49, 728–739.

Skalski, J.R., 1995. Statistical considerations in the design and analysis of environmental damage assessment studies. J. Environ. Manage. 43, 67–85.

Skalski, J.R., Coats, D.A., Fukuyama, A.K., 2001. Criteria for oil spill recovery: a case study of the intertidal community of prince william Sound, Alaska, Following the Exxon Valdez Oil Spill. Environ. Manage. 28, 9–18.

Taylor, L.A., Chapman, P.M., Miller, R.A., Pym, R.V., 1998. The effects of untreated municipal sewage discharge to the marine environment off Victoria, British Columbia, Canada. Water Sci. Technol. 38, 285–292.

Thrush, S.F., 1991. Spatial Patterns in Soft-bottom Communities. Trends Ecol. Evol. 6, 75–79.

Underwood, A.J., 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populatons. Aus. J. Mar. Freshwater Res. 42, 569–587.

Underwood, A.J., 1992. Beyond BACI: the detection of environmental impact on populations in the real, but variable world. JEMBE 161, 145–178.

U.S. Code of Federal Regulations. 2001. Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), revised 2001. U.S. Code of Federal Regulations. Vol. 43. Sections 11:14 and 11:72.

Verdonschot, P.F.M., Spears, B.M., Feld, C.K., Brucet, S., Keizer-Vlek, H., Borja, A., Elliott, M., Kernan, M., Johnson, R.K., 2013. A comparative review of recovery processes in rivers, lakes, estuarine and coastal waters. Hydrobiologia 704, 453–474.

Villnas, A., Perus, J., Bonsdorff, E., 2011. Structural and functional shifts in zoobenthos induced by organic enrichment - Implications for community recovery potential. J. Sea Res. 65, 8–18.

Wan Hussin, W.M.R., Cooper, K.M., Froján, C.R.S.B., Defew, E.C., Paterson, D.M., 2012. Impacts of physical disturbance on the recovery of a macrofaunal community: A comparative analysis using traditional and novel approaches. Ecol. Indicators 12, 37–45.

Whitlach, R.B., Lohrer, A.M., Thrush, S.F., 2001. Scale-dependent recovery of the benthos: effects of larval and post-larval life stages. In: JY, A., SA, W., RC, A. (Eds.), Organism–sediment interactions. University of South Carolina Press, Columbia, SC, pp. 181–197.

Wiens, J.A., Parker, K.R., 1995. Analyzing the effects of accidental environmental impacts: approaches and assumptions. Ecol. Appl. 5, 1069–1083.

Young, T.P., Chase, J.M., Huddleston, R.T., 2001. Succession and assembly as conceptual bases in community ecology and ecological restoration. Ecol. Res. 19, 5–19.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. Mixed effect models and extensions in ecology with R. Springer.