**Somerfield**, P. J., Clarke, K. R. (2013) Inverse analysis in non-parametric multivariate analyses: distinguishing of groups of associated species which covary coherently across samples. *J. Exp. Mar. Biol. Ecol.* **449**: 261–273. DOI:10.1016/j.jembe.2013.10.002


# Inverse analysis in non-parametric multivariate analyses: distinguishing groups of associated species which covary coherently across samples.

Paul J. Somerfield [a], K. Robert Clarke [a]

[a] Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK

Paul J. Somerfield, pjso@pml.ac.uk, +44 1752 633100, corresponding author

K. Robert Clarke, krc@pml.ac.uk

## HIGHLIGHTS

- Similarity Profiles (SIMPROF) analysis is used to examine associations among species
- Type 2 SIMPROF determines whether observed associations could have arisen by chance
- Type 3 SIMPROF with clustering identifies groups of coherently covarying species
- Component line plots – coherent curves – are used to display patterns
- The methods extend naturally to other types of multivariate data such as environmental variables

ABSTRACT

For decades multivariate analysis has been recognised as being appropriate for the analysis and description of complex ecological datasets, such as are routinely generated in studies of biota along gradients in time or space. The main focus of analyses tends to be the description and analysis of patterns among samples and groups of samples. Early applications of multivariate analyses to ecological data also recognised the importance of, and gave equal weight to, understanding how variables (species or taxa, in biotic datasets) varied among samples and groups of samples, but such analyses have inherent difficulties. Among these are the facts that species do not vary independently of each other, that responses of species to gradients may not be monotonic, that there are generally many more species than samples, that abundances vary widely within and among species, that some species are rare. Although some methods are routinely applied to explore species responses across and among samples to environmental gradients, few explicitly recognise that species do not vary independently. Within a very widely-used framework for the nonparametric multivariate analysis of ecological data we demonstrate how Similarity Profiles (SIMPROF) analysis and other approaches may be combined to analyse associations among species, and to visualize those relationships. Type 2 SIMPROF determines whether observed associations could have arisen by chance. Type 3 SIMPROF detects statistically distinct subsets of species which respond to gradients in a coherent manner. How different groups respond is visualised using component line plots (coherent curves). We illustrate the method using a range of datasets. We show how the method discriminates groups of species which respond differently to a single gradient, or respond coherently to different environmental or anthropogenic pressure gradients. We demonstrate how these approaches extend naturally to analyses of other types of multivariate data where the identification of coherent groups of variables is of interest.

KEYWORDS: Non-parametric multivariate; SIMPROF; r-mode analysis; inverse analysis; coherence; species association

## 1. Introduction

Biological assemblage data, representing the abundance or biomass of taxa in samples, presents many problems from a statistical perspective. Field et al. (1982) described a robust non-parametric multivariate strategy for the analysis of such data, which was expanded and clarified by Clarke (1993). The essence of the strategy is to display patterns among samples determined by appropriate resemblance measures (Clarke et al., 2006b) using clustering and ordination, and to analyse these patterns using a range of hypothesis tests and associated analyses, primarily based on ranked resemblances. Additional analyses are constantly added to the framework. A recent example is Similarity Profiles analysis (SIMPROF, Clarke et al., 2008) which tests for multivariate structure among groups of samples.

The major thrust over recent decades has been to group, analyse and display relationships among samples, termed Q-mode analysis, and to relate patterns among samples to similar analyses of explanatory variables. Although analyses of patterns among variables (r-mode analysis, or inverse analysis) was an important part of the original Field et al. (1982) strategy such analyses have not received equal attention over the years. Within the framework the main analyses focusing on patterns among variables (taxa) are Similarity Percentages breakdown (SIMPER, Clarke, 1993) which describes the contributions different taxa make to resemblances within and between groups of samples, and BVStep (Clarke and Warwick, 1998) which searches for subsets of variables which, in combination, reproduce patterns as described in a fixed target resemblance matrix.

Given that differences in abundance and composition of taxa drive differences among groups of samples it is, perhaps, surprising that more attention has not been paid to patterns of variation among variables. Indeed, it could be argued that the ultimate goal of analyses based on samples should be to understand which species are varying, how and why. One of the perceived problems with multivariate r-mode analysis is that species can only appear in an ordination or dendrogram in one place, which may not reflect the rich range of distributional relationships among species adequately. Other inherent statistical difficulties associated with species-analyses are that species do not vary independently of each other, that responses of species to gradients may not be monotonic, that there are generally many more species than samples, that abundances vary widely within and among species, and that many species are rare. Although some methods (e.g. canonical multivariate methods) are routinely applied to explore species responses across and among samples to environmental gradients, few explicitly recognise that species do not vary independently.

Given that individuals and species do not arrive independently in samples (Clarke et al., 2006a), and that species are not expected to vary independently of each other (one of the prime motivations

for adopting a multivariate analytical strategy in the first place), can we assess whether we have any statistical support for concluding that patterns among variables are interpretable? Secondly, and more importantly, assuming that we are convinced that patterns are interpretable, how can we determine whether groups of species are covarying coherently, changing in relative abundance across samples (and therefore presumably responding to environmental conditions) in a similar manner, and which are not?

Clarke et al. (2008) discussed the fact that SIMPROF could be used to examine relationships among variables, concluding that the appropriate pretreatment of the data and resemblance measures, and the precise definition of the hypotheses being tested, merited further study. Here we describe how SIMPROF analyses may be used to conduct r-mode analysis within the non-parametric multivariate analytical framework of Field et al. (1982) and Clarke (1993). We go on to demonstrate that many of the ideas and methods transfer readily to other contexts in which patterns of coherent variation among variables may be of interest.

## 2. Material and Methods, Theory and Calculation

### *2.1. Similarity Profiles Analysis (SIMPROF)*

SIMPROF tests for structure in multivariate data, and the method is described in detail in Clarke et al. (2008). Briefly, the test relies on the fact that if multivariate structure is manifest in a group of objects (samples or variables) some pairs of objects will be more, or less, similar to each other than would be expected if the data were essentially random and there was no structure. For analyses among samples, following an appropriate pre-treatment of the data (standardisation, transformation, etc.) a measure is used to calculate resemblances among samples. These resemblances are ranked from smallest to largest to form a Similarity Profile. If samples tend to be more, or less, similar to each other than expected by chance, this profile will be different to one generated from the same matrix following a random reassignment of each species' abundances *across samples*, independently for each species. A permutation distribution of profiles under the null hypothesis can be generated and used to define the range of values at each rank consistent with the null hypothesis. Informally speaking, if the observed profile falls outside the expected distribution generated under the null hypothesis it is clearly an unlikely event if the null is true, and therefore the null may be rejected. A formal test, at a predefined level of significance, say 1%, is achieved by computing a test statistic $\pi$ as the total area between the observed profile and the mean profile under such random permutations (Clarke et al., 2008).

*Type 1 SIMPROF*

SIMPROF has, to date, generally been used for Q-mode analysis, to analyse for structure within a group of samples. We now refer to this as Type 1 SIMPROF (Fig. 1). Biological data (abundance, biomass, % cover) are transformed, if required, and resemblances calculated using an appropriate measure such as Bray-Curtis (Clarke et al., 2006b). A test of the hypothesis

$H_0$: Samples are homogeneous, there is no multivariate structure

is constructed by comparing the observed Similarity Profile with many profiles generated by recomputing similarities having first randomly permuted species' values independently over samples. If the observed profile is sufficiently unlike the range of profiles generated under such null hypothesis conditions then it casts enough doubt on the truth of the null hypothesis for it to be rejected.

As such, Type 1 SIMPROF could simply be used to test whether there is evidence for multivariate structure within a set of samples, and could therefore be used as a 'global multivariate analysis test', a hurdle that must be crossed before any further analysis of a dataset is warranted. Commonly, however, Type 1 SIMPROF is used in conjunction with other multivariate methods which group samples, to ask whether interpretation of those groupings is warranted. For example, if the samples are subjected to hierarchical agglomerative clustering to produce a dendrogram, separate SIMPROF tests may be run for the samples clustered at each node. Failure to reject the null hypothesis at any node implies that interpretation of finer divisions further down that branch of the dendrogram is not warranted. Running such a combined clustering/SIMPROF analysis delimits groups of samples which are significantly different from each other, but internally homogeneous.

*r-mode analysis: measures of association*

Prior to conducting analyses in r-mode, to look at relationships among variables (rather than samples) in a standard species by samples matrix, elements of pre-treatment of the data are typically required and an appropriate measure of association between species must be selected (Clarke and Warwick, 2001). To account for variation in overall abundance among species and to make values comparable, each species needs to be 'relativised' or 'standardised'. Depending on the resemblance measure to be used, this may be done implicitly (by choosing a measure that incorporates standardisation) or explicitly. A common means of assessing relationships between variables is the standard product moment, or Pearson correlation coefficient. When pairwise correlations between pairs of species are calculated this automatically 'relativises' the data. In fact it

is a full normalisation, subtracting each species' mean from each count and dividing by the species' standard deviation (Clarke and Warwick, 2001). A positive correlation implies species are positively associated, and vice versa. A problem with using correlation for species data, however, is the way it treats joint absences. Two species will be positively correlated if they are both absent from a range of sites, and the value of the correlation will increase if further sites are added to the matrix from which both species are absent. Positive associations are implied by matching species' abundances across sites, although one would like to clearly identify when species have fully negative associations, in the sense of their presences being at entirely different sites. Both are achieved by using a similarity measure, such as one from the Bray-Curtis family (Clarke et al., 2006b), which treats joint absences appropriately. Prior to calculating such a measure among species, however, data need to be standardised explicitly. This is typically done using the species standardisation (for species $i$ in sample $j$):

$$y'_{ij} = 100 y_{ij} / \sum_{k=1}^{n} y_{ik}$$

thus converting values (abundances or biomasses) in samples to percentages of each species' total across all samples, which for each species now sum to 100. (Note that following such standardisation a transformation designed to reduce large disparities in values between species is no longer required. However, a transformation to reduce large disparities within species among samples may still be considered, prior to species standardisation, though this has not generally been used here.) In fact, the most appropriate measure to use for species' similarity is one that is closely related to Bray-Curtis but which includes this standardisation, rescaling each species' proportions to sum to 100 each time it is calculated. This coefficient is essentially that of Whittaker (1952) but treated as a similarity, a genuine index of association (rather than the 'disassociation' formula more commonly cited for Whittaker's index, Legendre and Legendre, 2012). The Index of Association (IA) takes the value 100 when two species have exactly the same percentage abundances across the samples (full positive association) and the value zero when they are found in completely different samples (full negative association). Taking $y_{ij}$ to be the abundance of the $i$th species ($i = 1, \ldots, p$) in the $j$th sample ($j = 1, \ldots, n$), the IA between species 1 and 2 is defined as:

$$IA = 100\left[1 - \frac{1}{2}\sum_{j=1}^{n}\left|\frac{y_{1j}}{\sum_{k=1}^{n} y_{1k}} - \frac{y_{2j}}{\sum_{k=1}^{n} y_{2k}}\right|\right]$$

or, equivalently,

$$IA = 100 \sum_{j=1}^{n} \min\left\{\frac{y_{1j}}{\sum_{k=1}^{n} y_{1k}}, \frac{y_{2j}}{\sum_{k=1}^{n} y_{2k}}\right\}.$$

It is helpful to retain the implied concept of 'negative' and 'positive' associations that derives from correlation, e.g. species which tend to co-occur less often than expected by chance, or tend to co-occur more often than expected, respectively, even when using a coefficient which takes values only over a positive range 0 to 100.

*Type 2 SIMPROF*

Type 2 SIMPROF (Fig. 1) addresses the fact that species are not expected to vary independently of each other (independence implying species will appear positively or negatively associated only by chance), and thereby assesses whether we have any statistical support for concluding that patterns among variables are interpretable. Using an appropriate coefficient, such as the Index of Association, associations among all species in a matrix are quantified and a Similarity Profile constructed. Randomising species' relative proportions separately across samples breaks down any true associations, generating a range of associations commensurate with a null hypothesis:

$H_0$: Species are not associated with each other.

Rejecting the null hypothesis in a Type 2 SIMPROF test allows us to conclude that there are genuine associations, above and beyond those that could arise by chance. As such it may again be thought of as a global test for multivariate structure, this time assessed directly from the viewpoint of the species themselves, rather than the effect any species interdependencies have on the sample structure (as seen in a Type 1 SIMPROF test). If the test fails to reject the null hypothesis then the investigator has no statistical support for examining particular species' associations further. In fact the interest is often not so much on the rejection of the null, but the nature of the departure of the observed profile from the simulated ones. Genuine negative associations are implied if there is an excess of similarities with values lower than expected under the null hypothesis, as evidenced in the Similarity Profiles analysis. Conversely, excess similarities higher than expected imply that at least some species are genuinely positively associated with each other.

*Type 3 SIMPROF*

Having explored the associations among all species and rejected a null hypothesis of 'no associations among species', a more interesting and useful question to address is: which species are associated? The null hypothesis here is that there are subsets of species which, internally, have exactly the same degree of association and therefore form coherent groups, so we need to construct a test that will determine whether this null hypothesis can be rejected with a predefined degree of confidence. For a given subset of species, Type 3 SIMPROF addresses the null hypothesis:

$H_0$: Species are coherently associated

and, as with Type 1 SIMPROF, such Type 3 analyses would normally be run in a sequence, having identified hierarchical subsets of species to test using some form of cluster analysis.

In common with many multivariate analyses examining patterns of variation among species, prior to such an analysis sequence, relatively rare species, whose occurrences across samples are essentially sporadic, need to be removed. Two species which only occur as single individuals in the same sample will have an association of 100, but if they are in different samples this will be 0. A species which only occurs as one or two individuals in a single sample can contain no useful information about associations (positive or negative) with other species. Thus rare species introduce noise into an analysis which may mask any genuine patterns. This is in marked contrast to analyses based on resemblances among samples, in which rare species play little part and there is no need to exclude them. In a Type 1 SIMPROF analysis those species contributing to that structure are those determining patterns in similarities among samples. Although transformations may be used to alter the weighting of abundant and rare species in an analysis, even with the strongest transformation (presence/absence) a single occurrence of a species in one sample is unlikely to alter a pattern of similarities among samples in any meaningful way. Prior to calculating similarities among species however, it is almost always desirable to remove rare species and there are a number of ways this might be carried out. The most often used is to remove species which contribute less than a set percentage of total abundance in any one sample. An alternative is to adjust this percentage until a preselected number of species are retained. A useful alternative could be to use frequencies of occurrence of species, retaining species that occur in more than a set number of samples.

Having omitted rare species, the Type 3 SIMPROF test again (as with Type 2) operates on interspecies associations, this time calculated from a matrix of standardised abundances of a subset of species (Fig. 1). The appropriate permutations are not, now, of species' relative proportions separately across samples. Instead, these relative proportions are permuted across species within

samples. If exchanging the relative proportions of species' abundances within samples significantly alters the similarity profile among variables, then the null hypothesis will be rejected. If a group of species have the same mutual associations, then swapping the proportional abundances among them, within samples, will not significantly alter those associations. For example, if a group of species has high abundance for sample 1, mid-range abundance for sample 2, absent for sample 3, present in small numbers for sample 4 etc., then randomly permuting values across species within those samples will not change that pattern of abundances and thus not lead to a different set of associations for those species.

Type 3 SIMPROF is therefore of most use in association with a multivariate technique that groups variables, such as hierarchical agglomerative clustering, or binary divisive clustering, based on species associations. Starting at the 'top' of such a dendrogram, at each division a SIMPROF test is run to assess whether there is evidence for rejecting the hypothesis that all species below that point in the tree have exactly the same association with each other, i.e. they form a coherent group. In other words, Type 3 SIMPROF is used to determine how far down the dendrogram an investigator has statistical support for interpreting divisions. If the test fails to reject the null hypothesis of coherent associations among species at a node, differences in distributions across samples of species below that node should not be considered meaningful.

Of course, there is nothing specific in the methods described here which makes them of relevance only to variation in measures of species' abundance. The key, as in most analyses, is to choose appropriate measures of resemblance (Clarke et al., 2006b) and to pre-treat the data accordingly. For example, to examine whether sets of environmental variables, measured on different scales, vary coherently across a set of samples, an appropriate measure of association would now be the Pearson correlation coefficient. The correlation calculation incorporates normalisation of variables to a common scale but, to allow the meaningful permutation of values across variables necessary for Type 3 SIMPROF, the variables in the dataset still need to be pretreated with a normalisation step before entry to the analysis.

*2.2. Analytical software*

All the analyses were conducted using PRIMER v6 (Clarke and Gorley, 2006) and an α development version of PRIMER 7 (PRIMER-E, Plymouth, UK). All analyses and plots are possible using the current v6 with the exception of line plots, which could be plotted using a range of software, although here PRIMER 7 was used. Note, however, that the default SIMPROF permutation procedure in v6 is not designed to carry out the novel Type 3 SIMPROF analysis and will permute the data matrix incorrectly. To obtain the correct permutations for this analysis requires a temporary

switching of the definition of 'samples' and 'variables'. Note also that v6 SIMPROF plots resemblances from largest to smallest, so the profile plots presented here have been reversed, and in v6 Whittaker's Index of Association is calculated as a distance rather than a similarity.

*2.3. Example Datasets*

To demonstrate the new insights that these methods can bring to analyses we use as examples a series of datasets, for the majority of which multivariate patterns among samples have been widely seen and discussed, particularly in the context of non-parametric multivariate analyses (Clarke and Warwick, 2001). Note that where species names appear in what follows the names used are those from the original studies, although where considered relevant updated names are also provided. The following datasets are used in this study:

*Exe estuary*

Warwick (1971) describes assemblage data on 140 species of free-living nematode at 19 intertidal sites in the Exe estuary, UK. Analysed here are abundances averaged over 6 sampling occasions in one year. This dataset is the one analysed by Field et al. (1982), in which study various aspects of r-mode analysis were discussed and exemplified.

*Bay of Morlaix*

A time-series of benthic macrofaunal abundances in samples, each comprising 10 pooled grab samples (1 $m^2$) of sediment, collected at station 'Pierre Noire' in the Bay of Morlaix. Sampling occurred on 21 occasions between April 1977 and February 1982. The time-series spans the period of the wreck of the 'Amoco-Cadiz' oil tanker in March 1978 (Dauvin, 1984). The sampling site was some 40 km from the initial tanker disaster but substantial coastal oil slicks reached the vicinity.

*Loch Linnhe*

A time-series of benthic macrofaunal samples from a station in a western Scottish sealoch, each comprising abundances in a number of pooled grab samples. Samples were collected annually from 1963 to 1973, covering the period of commissioning of a wood-pulp mill (Pearson, 1975). The later years show increasing pollution effects on the macrofauna, except that in 1973 a recovery was noted following a decrease in pollution loading (Pearson, 1975; Warwick, 1986).

*Clyde sewage sludge disposal site*

Pearson and Blackstock (1984) reported the results of biotic and abiotic sampling across the sewage-sludge disposal site at Garroch Head, Firth of Clyde, Scotland. Here, abiotic data from 1983 from 12 sites along an E-W transect, are analysed. Recorded for each site was a suite of (mainly) contaminant variables, the metals Cu, Mn, Co, Ni, Zn, Cd, Pb, Cr (all in ppm), % Carbon, % Nitrogen, and the depth of the water column. Stations in the middle of the transect show clear signs of gross pollution (Pearson, 1987; Warwick et al., 1987).

*Bremerhaven workshop*

A workshop on biological effects techniques sponsored by the International Council for the Exploration of the Sea (ICES) and the Intergovernmental Oceanographic Commission (IOC) was held at Bremerhaven, Germany, in 1990 (Stebbing et al., 1992). The main objective was to test a wide range of biological-effects techniques on contaminant gradients under the conditions in which they might be used in a monitoring programme. Data used here are a mix of 11 biochemical and histological/histochemical variables, measuring induction of detoxification mechanisms (e.g. EROD), lysosomal stability (neutral-red retention and acridine orange assays), lipid structural features (vacuole size), etc., made on dab (*Limanda limanda*) from 5 sites along a putative pollution gradient from the mouth of the Elbe to the Dogger Bank, North Sea.

## 3. Results and specific discussion

### *3.1. Exe nematodes*

In a worked example Field et al. (1982) analysed the Exe estuary dataset using both Q- and r-mode analyses. A cluster analysis and non-metric multidimensional scaling (MDS) ordination from this Q-mode analysis based on fourth-root transformed abundances identified, informally, a series of sample groups (1A, 1B, 2-4) linked at an arbitrary level of similarity in the associated cluster analysis. A reanalysis of the data using Type 1 SIMPROF (Fig. 2A) reproduces their ordination, but also shows how the apparent groupings have a degree of statistical support.

In their r-mode analysis (re-analysed here as Fig. 2B), it was clear that groupings among species determined by r-mode clustering (using Bray-Curtis on standardised abundances of a reduced set of species) could be informally related to grouping among samples determined in the Q-mode analysis. This reflected the fact that several assemblages found under differing environmental conditions within the estuary were essentially distinct. Before interpreting such a plot, however, it may be instructive to run a Type 2 SIMPROF test, if only to establish beyond reasonable doubt that there is genuine structure in the data to interpret. Figures 3A and 3B show results of Type 2 SIMPROF tests

run on the Exe data using two alternative measures of association. Note that while there is no requirement to run Type 2 SIMPROF on the reduced set of species, it generally makes sense to do so. The main interest is in how the observed profile deviates from profiles generated by permutation. Leaving all the rare species in the analysis has the effect of adding large numbers of joint absences, and a smaller number of perfect associations, compressing the area of interest in the profiles by extending the tails without affecting the test to any great extent. As mentioned above, in reality rare species are uninformative with respect to studies of association, so it makes sense to omit them, as is done here.

The Type 2 SIMPROF test for the Exe nematodes, based on interspecies correlations (Fig. 3Aa), is highly significant ($\pi = 0.043$, $p < 0.001$) indicating that there are genuine associations to interpret. It is clear from the plot that there are more positive correlations than would be expected if the null hypothesis were true. It is unclear, however, to what extent such positive associations reflect matching joint absences (rather than presences) across sites. Less easy to see is that there are also more negative correlations than would be expected, the observed profile lying below the 99% limit of profiles expected under null hypothesis conditions for all correlations with values <-0.04 or so. The long tail of weak negative correlations is generated by the normalisation of species (to a zero mean and unit variance across sites) inherent in calculating Pearson correlation, making absences take different values for each species in the calculation of each pairwise correlation. The comparable Type 2 SIMPROF test using the more appropriate Index of Association (Fig. 3B) is, as is to be expected, also highly significant ($\pi = 3.3$, $p < 0.001$) indicating that some species are more or less associated with each other than expected. Arguably, the profile is clearer, indicating both excess positive associations (observed profile above the upper 99% limit) and excess negative associations (observed profile below the lower 99% limit). The long tail of zero values indicates associations between species that do not occur in the same samples (again strengthening the case for omitting rare species from such analyses). It should be remembered that the communities under different conditions in parts of the Exe are more or less distinct (Fig. 2), so the excess of observed negative associations reflect the high species turnover across the spatial layout. A type 2 SIMPROF test for a different dataset, from the Bay of Morlaix time-series (Fig. 3C), is also highly significant ($\pi = 5.3$, $p < 0.001$), but presents a different pattern of associations. Here there is an excess of positive associations but no evidence of negative associations, reflecting a low species turnover across time-points and an absence of species that tend to occur in the same place but only at different times.

To continue with the Exe nematode example, the main question of interest is how species are varying across samples or, more specifically, which species are varying coherently across samples. Agglomerative clustering of the interspecies association matrix (Fig. 4) divides species into

hierarchical groups. Field et al. (1982) used a slice through this dendrogram to divide species into groups corresponding, approximately and informally, to the main groups of samples (Fig. 2B). A sequence of Type 3 SIMPROF tests on groups of species joined at each node in the dendrogram reveals further structure, with each species falling into one of 8 groups (A-H) which are significantly different to each other, and within each of which a hypothesis of coherent variation across samples cannot be rejected. Plotting these Type 3 SIMPROF groups on the r-mode ordination (Fig. 2B) shows the relationship between them and the previous groupings. Group 1A corresponds to SIMPROF group A, 1B to SIMPROF group B, 2 to a combination of two distinct SIMPROF groups (C and E) and a single species (*Theristus normandicus*, SIMPROF group D), 3 to SIMPROF group H and 4 to SIMPROF groups F and G. Note that the Type 3 SIMPROF tests were run, here, using a significance level of 5%. Reducing the significance level makes the test more conservative, i.e. less likely to reject the null hypothesis, and doing so groups *Theristus normandicus* with the species in group E but otherwise makes no difference to the overall groupings. Note also, that as each test is based on a subset of possible permutations of the data it is possible for the results to differ marginally if repeated. Binomial calculations show that tests which give an average p value in the region of 5% are likely to return an actual value of p somewhere between 3.5% and 6.5% from 1000 permutations, and between 4.6% and 5.4% from 10,000 permutations (this is a general result, applying to all permutation tests). When a species falls close to the boundary of a predefined significance level, as happens here for *Microlaimus honestus* on a 5% level test (potentially separating it from the other species in group E), more permutations would determine on which side of the boundary it falls, but this is to place too much emphasis on the chosen significance level. However, it is worth repeating the tests, for different nominal significance levels, to appreciate how relatively stable the groupings of species tend to be in a SIMPROF test, almost whatever the chosen significance level. In this case, the only difference that repeats at a range of significance levels make at all is in the grouping of *Theristus normandicus* and *Microlaimus honestus* in relation to group E.

Determining which species are varying coherently does not, however, describe how they are varying. An obvious way to visualise how species are varying coherently is to use component line plots of their percentage abundances in samples (noting that the line element is purely to aid visualisation of pattern). To produce such plots, however, it is helpful to have a sensible ordering of samples. For the Exe study the Q-mode analysis was used to provide an ordering of the samples (sites). Percentage abundances of species in groups derived from the Type 3 SIMPROF analysis were plotted using this ordering (Fig. 5) and these clearly show how groups of species vary coherently across sites in different ways. The strong species-turnover across sites inferred earlier is confirmed. Each group of species tends to occur only at a limited subset of sites, although some species are

more site-specific than others. What is remarkable is that the clear multivariate structure of the sites described by the Q-mode analysis (Fig. 2A) results from different groups of species appearing and disappearing along one or more putative gradients. With recourse to additional information, of course, it is then possible to examine these patterns in some depth. For example, Clarke et al. (2008) described a constrained divisive clustering (LINKTREE) of the Q-mode sites matrix derived from the nematode abundances, using thresholds derived from measured environmental variables as constraints (incidentally, also using Type 1 SIMPROF analysis to provide a stopping rule for the algorithm). This showed that different environmental variables, or combinations of environmental variables, explained the separation of each of the biotic clusters. For example, the first major division separated sites with higher organic content and low sulphide (1-11) from sites with lower organic content and higher sulphide (12-19). The coherence plots confirm that no groups of species span this boundary to any extent, and it is easy to see which species are associated with these different conditions. At a finer level, low interstitial salinity and higher organic matter content separated sites 1-4 from sites 7-9. It may be concluded, therefore, that species within Type 3 SIMPROF group A (e.g. *Axonolaimus spinosus* and *Anoplostoma viviparum*) have similar responses to these variables, occurring where salinities are low and organic content high, whereas species in Type 3 SIMPROF group B (e.g. *Axonolaimus paraspinosus* and *Viscosia viscosa*) occur where salinity is higher and organic content lower.

Even among the 5 groups initially discriminated (1A, 1B, 2-4) a SIMPER analysis (Clarke, 1993) would require examination of tables presenting average contributions of individual species, to 5 groups of within-group similarities and 10 (=(N.N-1)/2) groups of pair-wise between-group dissimilarities, in order to develop an understanding of how different species contribute to resemblances among groups of samples. The combination of type 3 SIMPROF tests and coherent curves as a method for analysing and visualising species' variation across samples, and therefore changes in community structure, provides an intuitively interpretable alternative.

### *3.2. Morlaix macrofauna*

Although the Exe nematodes provide a good example with which to describe the methods, the fact is that the main overall patterns are rather stark differences between groups of similar sites with, as has been shown, correspondingly distinct groups of associated species. How does Type 3 SIMPROF deal with a more complex situation? For the Morlaix macrofauna a dendrogram produced by hierarchical agglomerative clustering, coupled with sequential Type 3 SIMPROF tests (Fig. 6), indicates significant structure. There are 11 coherent species groups (A-K) containing two or more species. Note that, in common with Type 1 SIMPROF tests, it is not possible to discriminate profiles

when there are only two samples (or here, variables) to compare (Clarke et al., 2008), so in situations where a pair of species form an isolated group of two variables in the dendrogram, such as *Bathyporeia nana* and *B. elegans*, the test cannot reject the null hypothesis of coherence at any level of association. That being said, these species are clearly shown by the sequence of SIMPROF tests to be varying differently to all other species, and grouping them is justified. Four taxa (*Spiophanes bombyx*, *Nucula turgida*, *Goniada maculata*, unidentified Nemertes) varied across samples in such a fashion that they were significantly different to all others, although they all fall into a supercluster with Type 3 SIMPROF groups B, C and K.

Again, having identified which taxa are varying coherently across samples, the next step is to examine the 'coherent curves' (Fig. 7). The sample ordering is a natural one in this case, as the dataset represents a time-series. The relative biological complexity of the situation, compared to the Exe, is striking. Different curves clearly represent a great variety of patterns in abundance across samples. Analysis of community variation among the 21 sampling times, from 4th-root transformed abundances of all 251 species (Fig. 7, top left), shows a marked community change following the oil spill and a possible partial recovery, together with a clear seasonal cycle. Species in group A were present in high abundance prior to the spill, but almost disappeared in the year following it, slowly recovering (with a clear seasonal cycle) over the next three years without reaching pre-spill abundance. These include amphipods in the genus *Ampelisca*, which on the basis of analyses of this dataset as well as others are considered to be classic indicators of the impact of oil spills on soft-bottom macrobenthic communities (Gesteira and Dauvin, 2000). Here, however, a range of Group A species in other taxonomic groups such as polychaetes *Polydora antennata* (now *Pseudopolydora antennata*, Spionidae), *Notomastus latericeus* (Capitellidae), *Ampharete acutifrons* (Ampharetidae), *Exogone hebes* (Syllidae) and *Phyllodoce groenlandica* (Phyllodocidae), and the tanaid *Apseudes latreillei* (now *Apseudopsis latreillii*, Apseudidae), are shown to have responded to the spill in a similar and indistinguishable fashion. While species in group A all but disappeared following the spill, those in group B declined to a lesser extent and recovered (with seasonal cycles) over the following three years. In addition to a decline in abundance immediately after the spill, species in group C apparently failed to recruit in 1980, whereas group D is characterised by seasonal absences and a particularly strong recruitment in 1978, immediately after the spill. Type 3 SIMPROF groups E-J generally contain taxa that were not present in high abundance prior to the spill, but which increase in abundance at some time after it, often with a subsequent decline in abundance at a later stage. The groups differ from each other in terms of the timing of the increase (generally getting later E-J), its duration, and the strength of the seasonal cycle overlain on the pattern. Group K, on the other hand, contains taxa that tend to be present at all sampling times and which do not

undergo strongly synchronised seasonal cycles. Interestingly, this group contains *Hyalinoecia bilineata* (now *Aponuphis bilineata*, Onuphidae), one of the species considered characteristic of the benthic community at the site (Dauvin, 1984), whereas the main characterising species, *Abra alba*, is among a group of species (group F) that increased strongly in abundance a year after the spill, before falling back towards pre-spill levels of abundance. Thus these analyses allow us to visualise long-term trends and shorter-term cycles and present a clear, objective and novel view of coherent patterns among species.

*3.3. Linnhe macrofauna*

The Linnhe macrofaunal dataset (Fig. 8), in contrast, represents a time-series in which samples were collected at one time in each of a series of years. As such, analysis of the time-series can convey no information about intra-annual community variability. Species are associated only if their abundances covary coherently across years. Although Type 3 SIMPROF analyses generally utilise a grouping structure applied to the variables, typically using clustering, there is no requirement to present this analysis every time such an analysis is conducted. Presentation of the coherent curve plots may suffice, though some way of linking the identities of the taxa involved to the groups of curves is appropriate. Coherent curves derived from Type 3 SIMPROF analysis of the Linnhe time series (Fig. 8) show species falling into six objectively defined groups (A-F). The pulp mill was commissioned in 1966 and discharges increased to high levels in 1971-2, before decreasing in 1973 (Pearson, 1975). Taxa in group A tended to decline sharply in abundance in 1966, although continued to occur in smaller numbers in following years. Taxa in group B tended to occur in low numbers prior to 1966, peak in abundance in 1966, and then decline and disappear in later years. Taxa in group C tended to appear in large numbers in 1967, following the commencement of discharge, but then to fall back to low numbers, while those in group D, while not showing a sharp peak in abundance, tended to occur in most years but with relatively low abundance in 1970-2, the years of peak discharge. Group E species increased in abundance in 1970 before falling back to low abundance in 1971, whereas species in group F dominated the assemblage in 1971-2. Although the plots tend to accentuate peaks in abundance they also reveal subtle patterns. Species from groups A and D were present in 1963, at the beginning of the series, and present in comparable abundance at the end of the series. Taxa in groups E and F, characteristic of high organic enrichment in 1970-2, occurred in low and variable numbers prior to the commencement of discharge in 1966.

*3.4. Clyde sediment chemistry*

While the methods described here, particularly Type 3 SIMPROF, are potentially of great value in understanding species' variation (and covariation) across spatial and temporal layouts, as mentioned above there is no reason why the methods cannot be applied to different types of data. The abiotic data from the Clyde, representing a gradient of contamination associated with sewage-sludge disposal at site 6 (Fig. 9A), are measured on different scales. Sediment concentrations of metals are recorded in ppm, depth in metres, and carbon and nitrogen as percentages. To reduce the effects of outliers the concentration and percentage data warrant a log transformation (Clarke and Ainsworth, 1993) and, to place all variables on a common scale, normalisation (dividing each value of each variable by the standard deviation of that variable, having subtracted the mean value) is required. An appropriate measure of resemblance among variables is now the Pearson correlation coefficient. This is in contrast to the situation where the normalisation step (inherent in the calculation of the coefficient), applied to species data rich in joint absences, led to a range of negative correlation values (see above). Now negative correlations are of potential importance, to distinguish variables which increase towards the disposal site from those that decrease. In fact, clustering of variables (Fig. 9B) with Type 3 SIMPROF distinguishes only two groups of coherent variables. The main group (Fig. 9C), consisting of a range of metals along with % C and % N, are high at the site of disposal and decrease towards the ends of the transect. A smaller group (Fig. 9D), consisting of Mn, Co, Ni and depth, shows no response to the disposal site, instead generally decreasing along the transect, albeit with lowest values at site 9. In fact the line plots for this latter group make it abundantly clear why site 9 is widely separated in the ordination plot (Fig 9A) from sites at a similar distance to the disposal centre on the other arm of the transect (sites 3 and 4).

A classical multivariate approach to analysing this dataset (Clarke and Warwick, 2001), and variable sets with similar properties, would be to use correlation-based PCA (Chatfield and Collins, 1980; Pearson, 1901). The MDS (Fig. 9A) will be indistinguishable from a PCA ordination in this case, because it is based on the same distance measure, Euclidean distance, and 2-d is all that is necessary to display the pattern in these 12 points over 11-d space (PCs 1 and 2 account for 88% of the total variance). The high internal coherence of the two sets of variables (revealed here by the Type 3 SIMPROF analysis) is also clear, in this case, from the definition of the first two principal components. PC1 picks out roughly equally weighted combinations (coefficients >0.3) of the first variable set with near zero coefficients for the second set, and vice-versa (coefficients >0.4) for the second principal component. This point is made rather more simply and directly by the two groups of coherent curves.

### *3.5. Bremerhaven workshop biomarkers*

Analysis of the biomarker dataset (Fig. 10) is rather less straightforward. Firstly, each of the variables reflects, in some way, the relative health of the fish at the point of sampling. The measurements are not, however, all expected to vary in the same way. For example, high values of some (e.g. oxyradicals or N-ras) indicate poor health, while low values of others (e.g. lysosomal neutral-red retention assay or tubulin) also indicate poor health. It is, therefore, no longer desirable to discriminate between variables which are highly negatively correlated with each other, so the measurement scale for some variables was reversed so that lower values always indicate poorer health. Secondly, the scales on which measurements are recorded vary among variables, ranging from presence/absence (e.g. N-ras) through various ordered categorical scales (e.g. Lys NRR, Ubiquitin, Tubulin) to continuous scales (e.g. EROD, Lys AO, Lipid vac). In itself this is not a problem. Following appropriate data treatments, such as transformation of individual variables, data can be normalised and variables correlated with each other as before. The correlation matrix can then be clustered and hierarchical divisions tested using Type 3 SIMPROF. The issue is then how best to visualise the resulting patterns. The Type 3 SIMPROF tests are based on the full set of replicates, because the interest is in determining whether certain biomarkers covary across individual samples. When it comes to constructing coherent curves, however, the ordering of samples of interest is not that of individual samples, but the fact that groups of (10) replicate samples were taken at 5 sites along a gradient. Also, owing to the various types of scale on which measurements were made, line plots based on replicate data (e.g. those for variables which are binary or with few categories) are unlikely to be very revealing. The answer, of course, is to do the tests on the replicate data but to plot the average (normalised) value for each variable at each station along the gradient to visualise the coherent curves (Fig. 10). This is somewhat analogous to the commonly encountered situation in which a single variable is tested by ANOVA using the full set of replicates, but display and interpretation of the results uses plots of means, not replicates.

The combined clustering and Type 3 SIMPROF tests identified three significantly different groups of variables and a further single variable (N-ras) which was significantly different from the others. The resulting coherent curves (Fig. 10) show a group of variables (Group A) that indicate poor fish health at Stations S3 (close to the mouth of the Elbe) and S9 (in the central North Sea), a further group (B) indicating improving health from S6 to S9, the largest group (C) indicating improving health from S3 to S7, and a single variable (N-ras) showing a decrease in health at S3 compared to the other stations. These Type 3 SIMPROF tests clearly discriminate biomarker groups giving significantly different responses (with internal consistency) and, presumably, these indicate different mechanisms and pathways of impact. Group A includes EROD, a form of cytochrome P450 involved in detoxification of carcinogens, ubiquitin, required for protein turnover (breakdown via lysosomes

and proteasomes) and cathepsin D, a protein-degrading enzyme in lysosomes involved in protein turnover. This group suggests a functional link relating to protein turnover within cells. If the putative pollution gradient (from S3 to S9) is driving the patterns in this group then the response is non-monotonic, perhaps representing increased turnover at intermediate levels of pollution. Group B, endoplasmic reticulum and tubulin, suggests a functional link relating to internal packaging and transport of proteins. Group C includes indicators of lysosomal stability (acridine orange and neutral red retention assays), pinocytosis by which material is ingested and transferred to lysosomes for digestion, oxyradicals which is a measure of superoxide radical production, and lipid vacuoles which is a measure of degenerative fatty change within cells caused by exposure to toxins. The functional link here is possibly the role of lysosomal autophagy in cellular defence against oxidative stress (Moore et al., 2006). N-ras, varying separately from the other groups, is an oncoprotein that is involved in cell growth and linked to the development of some cancers. It is recognised that a major challenge facing ecotoxicologists is to integrate individual biomarker responses into a set of tools and indices capable of indicating and monitoring the degradation of health of particular types of sentinel organisms (Moore et al., 2013). Different groups of biomarkers are indicative of alternate causes, consequences and degrees of degradation (Moore et al., 2006), but there is a tendency to analyse responses separately, ignoring the fact that multiple biomarkers can often be derived from the same samples (Clarke, 1999). This example illustrates an alternative approach which may be of considerable benefit in developing a holistic understanding of biomarker responses, and to deciding which to include in monitoring programmes.

## 4. General discussion

The fact that patterns in community structure are not readily apparent from inspection of large species-by-samples arrays motivates multivariate statistical analysis that centres on reducing the complexity of such matrices. This simplification is usually achieved by some graphical representation of biological relationships among samples, such as clustering or ordination, and statistical testing to identify and characterise changes in community structure in time and space and relate these to changing environmental or experimental conditions (Clarke and Warwick, 2001). In this paper we explain, and give examples of, different ways in which Similarity Profiles analysis can be used to understand relationships among variables in multivariate datasets, extending the previous use of the methodology for identifying sample patterns (Clarke et al., 2008). These tests then lead seamlessly to new graphical representations of the data, support interpretation of changes in communities across samples, and allow these to be related to spatio-temporal changes in other variables.

There has been great debate among ecologists over decades concerning efforts to elucidate assembly rules, and ways in which interactions between species influence patterns of co-occurrence (Weiher and Keddy, 1999). Much of this work has focused on the influence of negative species interactions, detecting patterns of negative covariation among species using null model analyses of presence/absence matrices (Ulrich and Gotelli, 2010), and a large part of this debate has focused on the appropriateness, or otherwise, of different null models for testing competing hypotheses and on statistical aspects of various testing structures. It is increasingly being realised that analyses based on abundances may be more ecologically relevant (Ulrich and Gotelli, 2010). Type 2 SIMPROF may have a contribution to make in this area. It specifically addresses the question of whether there is any statistical support for interpreting associations among species at all, based on randomisations of real data. It may be used with abundance or presence/absence data, and by examining the actual profiles in relation to the simulated ones it may be determined whether there are more negative (or positive) associations than would be expected if the null hypothesis were true. Interestingly, apart from 'long baseline' situations where there is strong turnover (either spatially or temporally) in species composition overall, in our experience there is rarely strong evidence of excess negative associations among species, especially when an appropriate measure of association, such as the Index of Association, is used. There is an obvious opportunity here to apply Type 2 (and Type 3) SIMPROF more widely in this context.

Type 3 SIMPROF, with the accompanying coherent curve plots, has obvious relevance to other methods which aim to display variables across samples as lines or profiles, such as Parallel Coordinates Analysis (Wegman, 1990) or 'replicate line plots' (Cook et al., 2007). In fact, the coherent curves in this paper are a form of Parallel Coordinate plot, in that all variables are converted to a common scale (either by standardisation or normalisation) and are plotted against implied parallel coordinates representing different samples. In Parallel Coordinate plots efforts to identify similar samples, or similarly behaving variables, are generally approached by 'brushing' or colouring lines. Type 3 SIMPROF offers an alternative, powerful and statistically robust way of grouping variables in such analyses. There is also growing interest in time-series clustering (Liao, 2005), in diverse applications including commercial energy consumption (Košmelj and Batagelj, 1990), earthquakes and mining explosions (Shumway 2003), FMRI brain activity mapping (Wismüller et al., 2002) and many others (Liao, 2005). While the focus has been on metrics and clustering methods (Liao, 2005), the combination of methods presented here represents a novel approach, in that they can determine whether there is statistical support for examining relationships among variables in the first place (Type 2 SIMPROF) and, if so, identify and display how coherent groups of variables vary through time (Type 3 SIMPROF).

Bearing in mind that these are statistical tests, one should lose sight of neither the specific hypotheses being tested nor the power of the tests to reject the null hypothesis. For example, the samples in the Morlaix dataset are generally very representative of the biological conditions, being an amalgam of 10 separate 0.1 $m^2$ Smith-McIntyre grab samples. Thus variation in community structure from sample to sample is unlikely to be over-influenced by background spatial heterogeneity. As such we can interpret differences in peaks of abundance in terms of, for example, seasonal cycles. If an investigator were specifically interested in the seasonal cycles of species, one approach would be to average the abundances of each species within seasons before repeating the full analysis. The analysis would therefore address coherent seasonal cycles ignoring inter-annual variation, grouping species together if they tend to peak in abundance at the same time each year. The power of the tests, however, is a function of the number of observations (in this context samples available for each variable), with more observations implying greater power to detect differences (Clarke et al., 2008). Thus with very long time series (or large numbers of observations) the hypothesis of coherent association is very likely to be rejected even if differences are, in fact, small. The opposite is also true, so in a situation where the time series is reduced to, say, four seasonal samples (representing average abundances in winter, spring, summer, autumn), the analysis will likely have less power to discriminate groups of variables with distinctive seasonal patterns than if monthly times series are used. Thus, as with all testing structures, it is advisable to consider the balance between power (the number of observations), the reliability of the signal carried by each observation, the significance level of the tests, and precisely what it is that the results imply. For example, if one has a multi-decadal time-series of high-frequency observations (say daily or weekly) and one is interested in long-term trends, it may make sense to average samples within years to maximise the relevant signal.

As with hierarchical Type 1 SIMPROF tests for multivariate structure among samples (Clarke et al., 2008), and as exemplified in the Exe nematode example above, the results of Type 3 SIMPROF analyses are generally robust to the choice of p value. The initial grouping of variables is based on a separate analysis, made using a series of appropriate steps including pretreatment of the data and choice of a sensible measure of association, and it is only after the samples are grouped that SIMPROF tests the significance (or otherwise) of those groupings. Provided enough permutations are used to be able to determine critical regions for p values as small as 0.1% (at least 9999 is recommended in that case), choosing between p = 5%, 1% or 0.1% will often make little difference to the resulting sets of coherent groups identified. Typically, a couple of the groups might be split into two groups (perhaps one group and a singleton) using less stringent values of p, e.g. 5%, than with more demanding p values such as 0.1%. Bearing in mind that many such SIMPROF tests are

often carried out in a Type 3 (and Type 1) analysis, a more stringent p value might be preferred but it should also be remembered that ideally, as with all tests, there needs to be a balance between the choice of significance level (Type I error) and the power of the test (the complement of the Type II error). Although there is no formal framework for addressing statistical power in this context (Somerfield et al., 2002), if the computation of variable associations is based on many samples (higher power to detect differences) then a more stringent significance level might be suggested. But it is important to appreciate that precise tinkering with the p values for a suite of SIMPROF tests is likely to make only marginal differences to the definition of coherent groups. Pragmatically, therefore, it may be appropriate to recalculate the outcome with, say, the three p values of 5%, 1% and 0.1%, and to comment on the slight variations in the plots that the alternative p values would induce, where these seem sufficiently worthwhile reporting. In the Clyde environmental variables analysis (Fig 9) for example, p values of 1%, 0.1% (or even vastly smaller) will all identify exactly the same two groups, whereas p = 5% separates out the Ni variable as a singleton group distinguishable from the remainder of the second set.

The general applicability of the methods introduced in this paper should not be underestimated. Although we have chosen to exemplify them in the context of species abundances or measures of variables in samples, there are many other contexts in which they may be used even within ecology. For example, one could analyse a matrix of biological traits × species to determine whether species fall into coherent trait-based groups (see Somerfield et al., 2008 for a discussion of how one might construct such a matrix and the appropriate measure(s) of resemblance to use), or a matrix of taxonomic or genetic information × individuals to ask whether there is statistical support for discriminating individuals taxonomically (Mühling et al., 2006). A matrix of gut contents in different prey could be analysed to group predators and prey into guilds (French et al., in press). It should also be remembered that although we have used hierarchical agglomerative clustering to define grouping structures which are then tested using Type 1 and Type 3 SIMPROF , there is no reason why other groupings, either intrinsic (derived from the data) or extrinsic (derived from other knowledge about species or samples), may not be used.

The methods described in this paper are additions to a robust framework for non-parametric multivariate analysis widely used in ecology and other fields. In keeping with the transparent, permutation-based philosophy of that framework, no underlying distributional assumptions are made: there are no ‘black box’ models which the user has to take on trust. In fact the final outputs (Figs 5, 7-10) are simple line plots, the like of which a practitioner may have already plotted for particular variables, across some natural ordering of the samples. The key step that this paper adds, however, is a fully defensible statistical justification for grouping these line plots into batches of

coherent curves, demonstrating the full range of significantly different response groups of those variables across sample conditions, and a sometimes striking degree of uniformity of response within each group. While Type 1 SIMPROF (as a SIMPROF test for samples) and to a lesser extent Type 2 SIMPROF (as a SIMPROF for similarities amongst variables) were discussed by Clarke et al. (2008), at the time they wrote "conditions under which it would make sense to permute entries for each sample across variables must be rather rare". In fact, as we show here, this is not the case if appropriate combinations of data pretreatment and resemblance are applied, and the hypotheses to be addressed are understood. The result, Type 3 SIMPROF, especially when combined with visualisation tools such as coherent curve plots, can bring a powerful degree of organisation to any study aiming to understand relationships among variables and samples in a multivariate context.

**Acknowledgements**

We thank Ray Gorley of Primer-E Ltd for programming computer routines used in this work. PJS acknowledges National Capability support from the UK Natural Environment Research Council. The work is a contribution to the Marine Life Support Systems research area of the Plymouth Marine Laboratory, to projects Marinexus and Charm2 funded in part by the INTERREG IVA France (Channel) – England programme (FEDER), and to the EU FP7 projects Devotes (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status, Grant Agreement number 308392) and Vectors (VECTORS of Change in Oceans and Seas Marine Life, Impact on Economic Sectors, Contract number 266445).KRC acknowledges his positions as an honorary fellow at the Plymouth Marine Laboratory and adjunct professor at Murdoch University, W Australia.

## References

Chatfield, C., Collins, A.J., 1980. Introduction to multivariate analysis. Chapman and Hall, London.

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

Clarke, K.R., 1999. Non-metric multivariate analysis in community-level ecotoxicology. Environ. Toxicol. Chem. 18, 118-127.

Clarke, K.R., Gorley, R.N., 2006. PRIMER v6: User manual/tutorial. PRIMER-E, Plymouth.

Clarke, K.R., Warwick, R.M., 1998. Quantifying structural redundancy in ecological communities. Oecologia 113, 278-289.

Clarke, K.R., Warwick, R.M., 2001. Change in marine communities: an approach to statistical analysis and interpretation, 2nd edn. PRIMER-E, Plymouth.

Clarke, K.R., Chapman, M.G., Somerfield, P.J., Needham, H.R., 2006a. Dispersion-based weighting of species counts in assemblage analyses. Mar. Ecol. Progr. Ser. 320, 11-27.

Clarke, K.R., Somerfield, P.J., Chapman, M.G., 2006b. 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., Somerfield, P.J., Gorley, R.N., 2008. Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. J. Exp. Mar. Biol. Ecol. 366, 56-69.

Cook, D., Hofmann, H., Lee, E-K., Yang, H., Nikolau, B., Wurtele, E., 2007. Exploring gene expression data, using plots. J. Data Sci. 5, 151-182.

Dauvin, J-C., 1984. Dynamique d'écosystèmes macrobenthiques des fonds sédimentaires de la baie de Morlaix et leur perturbation par les hydrocarbures de l'Amoco Cadiz. Thèse Doct. Etat, Sci. Nat., Université Pierre et Marie Curie, Paris.

Field, J.G., Clarke, K.R., Warwick, R.M., 1982. A practical strategy for analysing multispecies distribution patterns. Mar. Ecol. Progr. Ser. 8, 37-52.

French, B., Clarke, K.R., Platell, M.E., Potter, I.C., in press. An innovative statistical approach to constructing a readily comprehensible food web for a demersal fish community. Est. Coast. Shelf Sci.

Košmelj, K., Batagelj, V., 1990. Cross-sectional approach for clustering time-varying data. J. Classification 7, 99-109.

Legendre, P., Legendre, L., 2012. Numerical Ecology, 3rd Engl edn. Elsevier, Amsterdam.

Liao, T.W., 2005. Clustering of time-series data – a survey. Pattern Recognition 38, 1857-1874.

Moore, M.N., Allen, J.I., Somerfield, P.J., 2006. Autophagy: role in surviving environmental stress. Mar. Env. Res. 62, S420-S425.

Moore, M.N., Viarengo, A.G., Somerfield, P.J., Sforzini, S., 2013. Linking lysosomal biomarkers and ecotoxicological effects at higher biological levels. In: Amiard-Triquet, C., Amiard, J-C., Rainbow, P.S. (Eds), Ecological biomarkers: indicators of ecotoxicological effects. CRC Press, Boca Raton, FL, pp. 107-130.

Mühling M., Somerfield, P.J., Harris, N., Belay, A., Whitton, B., 2006. Phenotypic analysis of Arthrospira (Spirulina) strains (cyanobacteria). Phycologia 45, 148-157.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Phil. Mag. 2, 559-572.

Pearson, T. H., 1975. The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. IV. Changes in the benthic fauna attributable to organic enrichment. J. Exp. Mar. Biol. Ecol. 20, 1-41.

Pearson, T.H., 1987. The benthic ecology of an accumulating sludge disposal ground. In: Capuzzo, J., Kester, D. (Eds.), Biological Processes and Wastes in the Ocean. R.E. Kneger, Melbourne, FL, pp. 195–200.

Pearson, T.H., Blackstock, J., 1984. Garroch Head Sludge Dumping Ground Survey, Final Report. Dunstaffnage Marine Research Laboratory, Oban.

Shumway, R.H., 2003. Time-frequency clustering and discriminant analysis. Stat. Probab. Lett. 63, 307-314.

Somerfield, P.J., Clarke, K.R., Olsgard, F., 2002. A comparison of the power of categorical and correlational tests applied to community ecology data from gradient studies. J. Anim. Ecol. 71, 581-593.

Somerfield, P.J., Clarke, K.R., Warwick, R.M., Dulvy, N.K., 2008. Average functional distinctness as a measure of the composition of assemblages. ICES J. Mar. Sci. 65, 1462-1468.

Stebbing, A.R.D., Dethlefsen, V., Carr, M., (eds) 1992. Volume 91 (1992) MEPS Special, book version. Biological effects of contaminants in the North Sea. Results of the ICES/IOC Bremerhaven Workshop. Inter-Research, Amelinghausen.

Ulrich, W., Gotelli, N.J., 2010. Null model analysis of species associations using abundance data. Ecology 91, 3384-3397.

Warwick, R.M., 1971. Nematode associations in the Exe estuary. J. Mar. Biol. Ass. U.K. 51, 439-454.

Warwick, R.M., 1986. A new method for detecting pollution effects on marine macrobenthic communities. Mar. Biol. 92, 557-562.

Warwick, R.M., Pearson, T.H., Ruswahyuni, 1987. Detection of pollution effects on marine macrobenthos – further evaluation of the species abundance-biomass method. Mar. Biol. 95, 193-200.

1 Wegman, E.J., 1990. Hyperdimensional data analysis using parallel coordinates. J. Am. Stat. Soc. 85,
2 664-675.
3 Weiher, E., Keddy, P. (eds), 1999. Ecological assembly rules, perspectives, advances, retreats.
4 Cambridge University Press, Cambridge, UK.
5 Wismüller, A., Lange, O., Dersch, D.R., Leinsinger, G.L., Hahn, K., Pütz, B., Auer, D., 2002. Cluster
6 analysis of biomedical image time series. Int. J. Comput. Vision 46, 103-128.
7 Whittaker, R.H., 1952. A study of summer foliage insect communities in the Great Smoky Mountains.
8 Ecological Monographs 22, 1-44.

**Figure legends**

**Fig 1.** Schematic diagram for construction of the three Types (1 to 3) of SIMPROF test on a quantitative species by samples matrix, and the respective null hypotheses $H_0$ that they test. The diagram applies equally to other (non-commonly scaled) variables such as environmental data, replacing 'species standardisation' with 'variable normalisation', 'association' with 'correlation' etc.

**Fig 2. Exe nematodes**. A) Non-metric MDS (nMDS) plot of 19 sites from 4th-root transformed, time-averaged abundances of 140 nematode species, using Bray-Curtis similarities between sites. Continuous contours denote the groups 1 (A&B), 2-4 identified by Field et al (1982) from group-average (Q-mode) cluster analysis, corresponding to a 17.5% similarity slice through the dendrogram. Dashed contours are a 35% similarity slice (corresponding approximately to the groups established by Type 1 SIMPROF tests). B) nMDS of the 'most important' 52 species (those accounting for $\geq$5% of total abundance at one or more of the sites), defining resemblance between species by Whittaker's index of association (equivalent to Bray-Curtis on species-standardised abundances, untransformed). Contours show species groups identified by (r-mode) cluster analysis at a 5% slice through the dendrogram, which can be approximately (and informally) matched to the 4-5 sample groups, as in Field et al (1982). Species numbers, and the symbols denoting clusters established by Type 3 SIMPROF tests, are defined in Fig 4.

**Fig 3.** Similarity profiles from Type 2 SIMPROF tests for: A & B) Exe nematodes, based respectively on Pearson correlation and Whittaker's index of association, as resemblance measures calculated among the subset of 52 species shown in Fig 2B; C) Morlaix macrofauna, based on the index of association among the subset of species (again 52, coincidentally) shown in Fig 6. Continuous lines denote the observed profile, the full set of pairwise resemblances ordered from smallest to largest (y axis) plotted against their rank (x axis). Dashed lines are limits within which 99% of resemblances

would be expected to fall, for any given rank, under the null hypothesis of no association amongst species.

**Fig 4. Exe nematodes.** Dendrogram from (r-mode) group-average clustering of the 52 'most important' species, based on Whittaker's Index of Association among species, as in Fig 2B. Continuous lines indicate the 8 'coherent groups' (A - H) which were significantly differentiated by Type 3 SIMPROF tests (at the 5% level). Within each of these groups, the null hypothesis that all pairs of species have the same association to each other cannot be rejected, the subgroup structure identified by cluster analysis thus having no statistical support (dashed lines).

**Fig 5. Exe nematodes.** Groups of 'coherent curves', namely component line plots for the groups of species identified in Fig 4, showing the consistency of species responses within a group and (in this case) the high species turnover between groups. The y axes are percentages of the total abundance of each species found across the 19 sites (i.e. 'species-standardised', untransformed data); the x axis rearranges the 19 sites to preserve the groupings in the site dendrogram referred to in the legend of Fig 2A, rotated approximately to coincide with the layout of this nMDS plot. Groups D and E, though separated by the Type 3 SIMPROF test are plotted together, the single-species group D being separately identifiable by the dashed line. They have relatively high similarity (Fig 4) and the text describes how these two groups are not separated at stronger significance levels.. Species within other groups are not individually identified because of their statistically inseparable responses.

**Fig 6. Morlaix macrofauna.** Dendrogram from group-average clustering of the 52 'most important' species (those accounting for $\geq$0.5% of total abundance at one or more of the times) based on Whittaker's Index of Association calculated among (untransformed) species abundances, for soft-sediment benthos at 21 sampling times (spanning the Amoco-Cadiz oil spill). As in the legend to Fig

4, continuous lines indicate the coherent groups determined by a series of Type 3 SIMPROF tests (11 groups with multiple species, A-K, and four further single-species groups, somewhat similar to B, C and K).

**Fig 7. Morlaix macrofauna.** Top left: nMDS based on Bray-Curtis similarities among 21 sampling times, from 4th-root transformed abundances of all 251 species observed, appearing to display a marked community change following the oil spill and a possible partial recovery, together with a clear seasonal cycle. For a subset of 52 species, the other 11 graphs show line plots for the coherent species groups A-K, as identified by SIMPROF tests in Fig 6 (omitting the single-species groups). The y axes are the relative species abundances (untransformed but each standardised to a common total of 100%); x axes are the sequential time points (roughly at 3-month intervals).

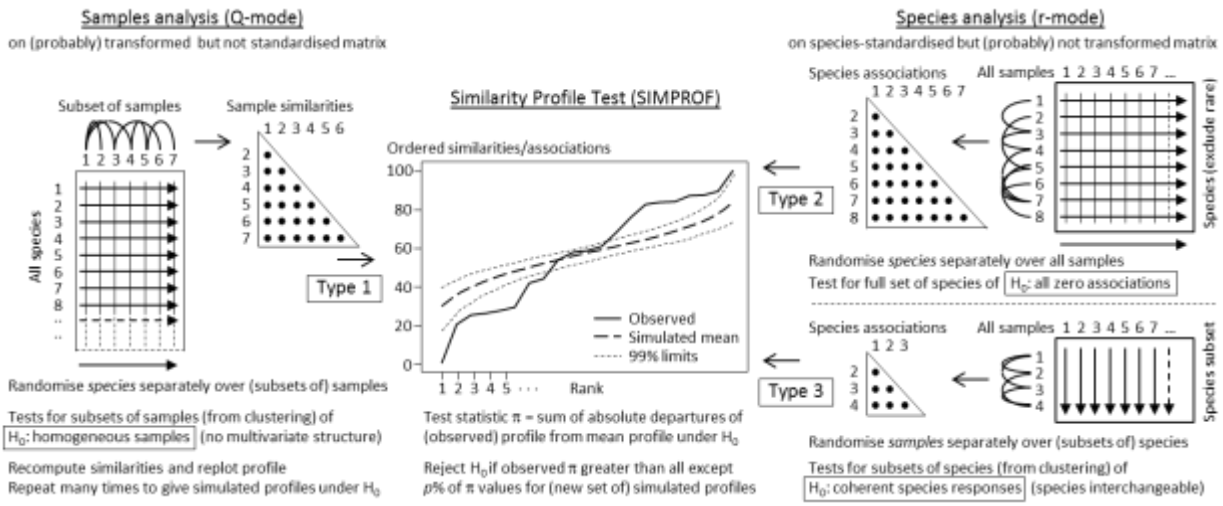
**Fig 8. Linnhe macrofauna.** Groups of 'coherent curves' from (untransformed) abundance data, species-standardised across 11 years for a subset of 51 species (those accounting for $\geq$1% of total abundance in one or more years). Type 3 SIMPROF tests at the 1% level, applied to a cluster analysis on the index of association among species, gave 6 statistically distinguishable groups (A - F) with between 6 and 13 members, and one further single-species group (*Mysella bidentata*, present only in 1973).

**Fig 9. Clyde sediment chemistry.** A) nMDS for 12 sites on a transect across a sludge dump-ground (site 6), based on samples in one period for 11 environmental variables (metal concentrations, % carbon and nitrogen, and water depth, Dp). All but water depth were subjected to a log(1+x) transformation, all normalised to a common mean (0) and variance (1), and the ordination based on Euclidean distances among the 12 sites. B) Group-average cluster analysis on Pearson correlations among the 11 (transformed) variables, the darker lines indicating the only two groups distinguished

by Type 3 SIMPROF tests (at about the 1% level or finer) on the normalised variables. C) & D) Line plots for the two coherent groups of variables identified in (B), of normalised variables plotted along the transect of sites

**Fig 10. Bremerhaven workshop biomarkers.** The 4 groups of coherent curves produced by Type 3 SIMPROF tests on 11 biomarker response variables measured on dab collected at 5 sites (S3, S5, S6, S7 and S9) in the North Sea, along a putative contaminant gradient. Cath D = cathepsin D; End Ret = endoplasmic reticulum; Lys AO = lysosomal Acridine Orange; Lys NRR = lysosomal Neutral-Red Retention; Pinocyt = pinocytosis; Oxyrad = oxyradicals; Lipid vac = lipid vacuoles. On replicate data (10 pools of material for each site), the EROD, Oxyrad and Lipid vac variables were log(1+x) transformed, all variables normalised, variables indicated by (-) were reversed in sign (low values then consistently implying impact), and Pearson correlation among variables, over the 50 samples, used to produce group-average clustering and the 4 SIMPROF groups A - D (on 1% level tests). The line plots show the averages of the normalised variables over the replicates at each site.

1

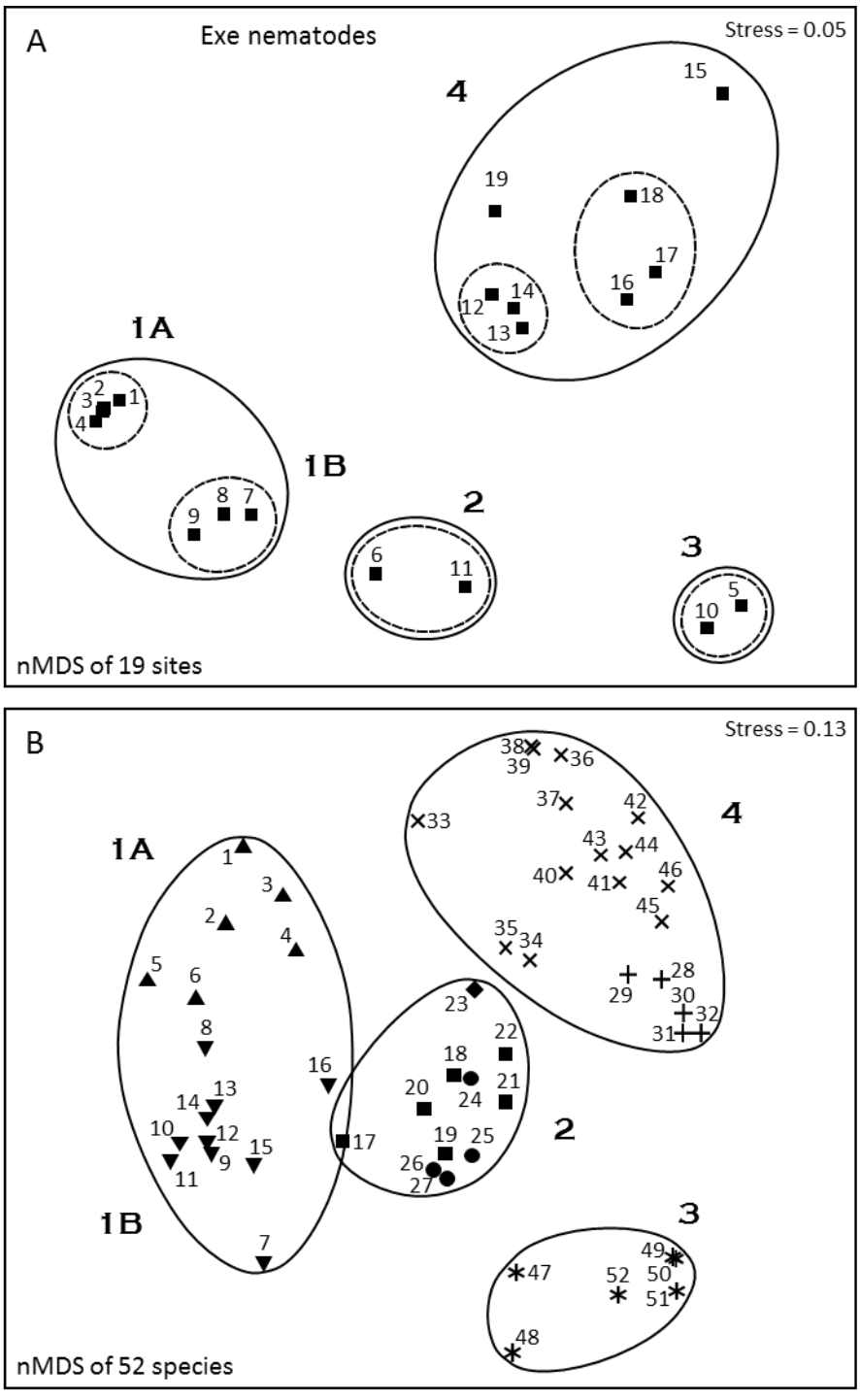



2

3 Somerfield and Clarke 2013 Figure 1

4

1

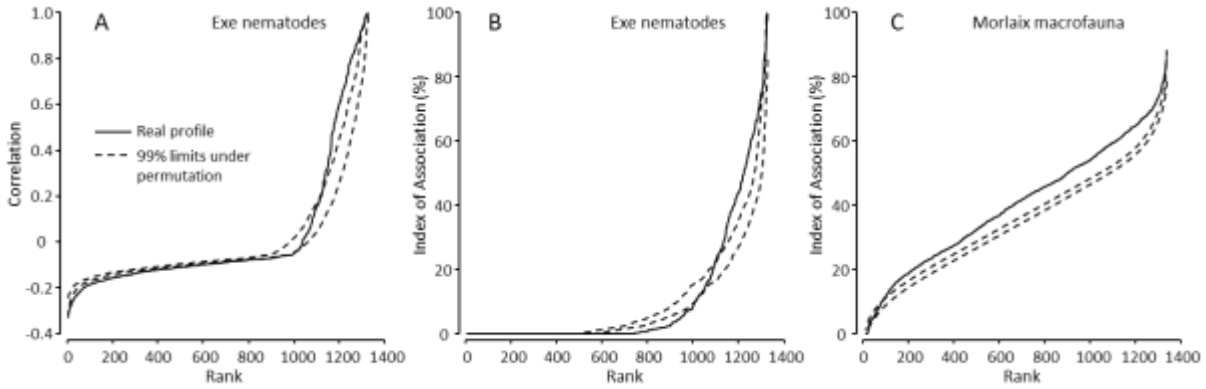



2

3 Somerfield and Clarke 2013 Figure 2

4

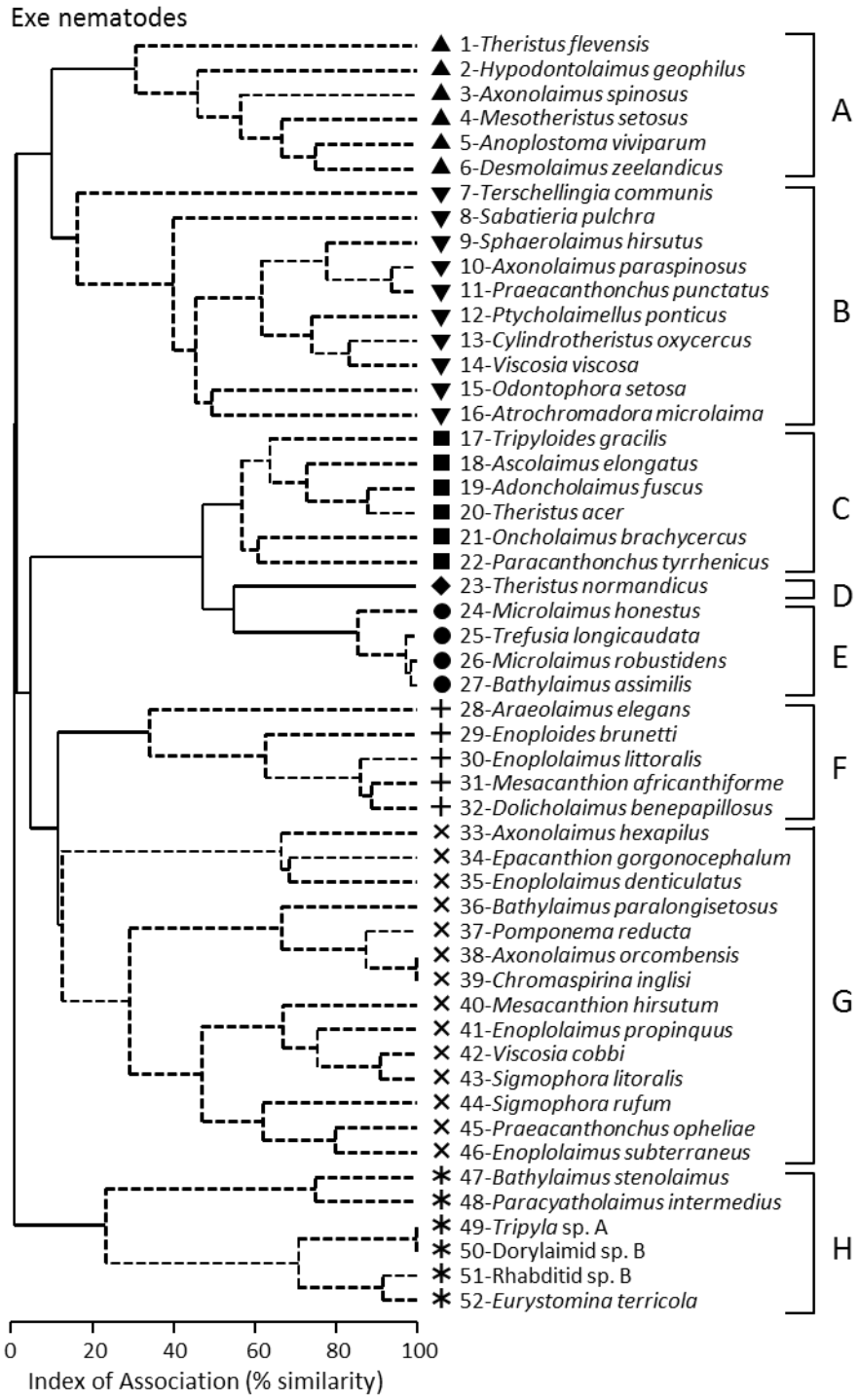
1




2

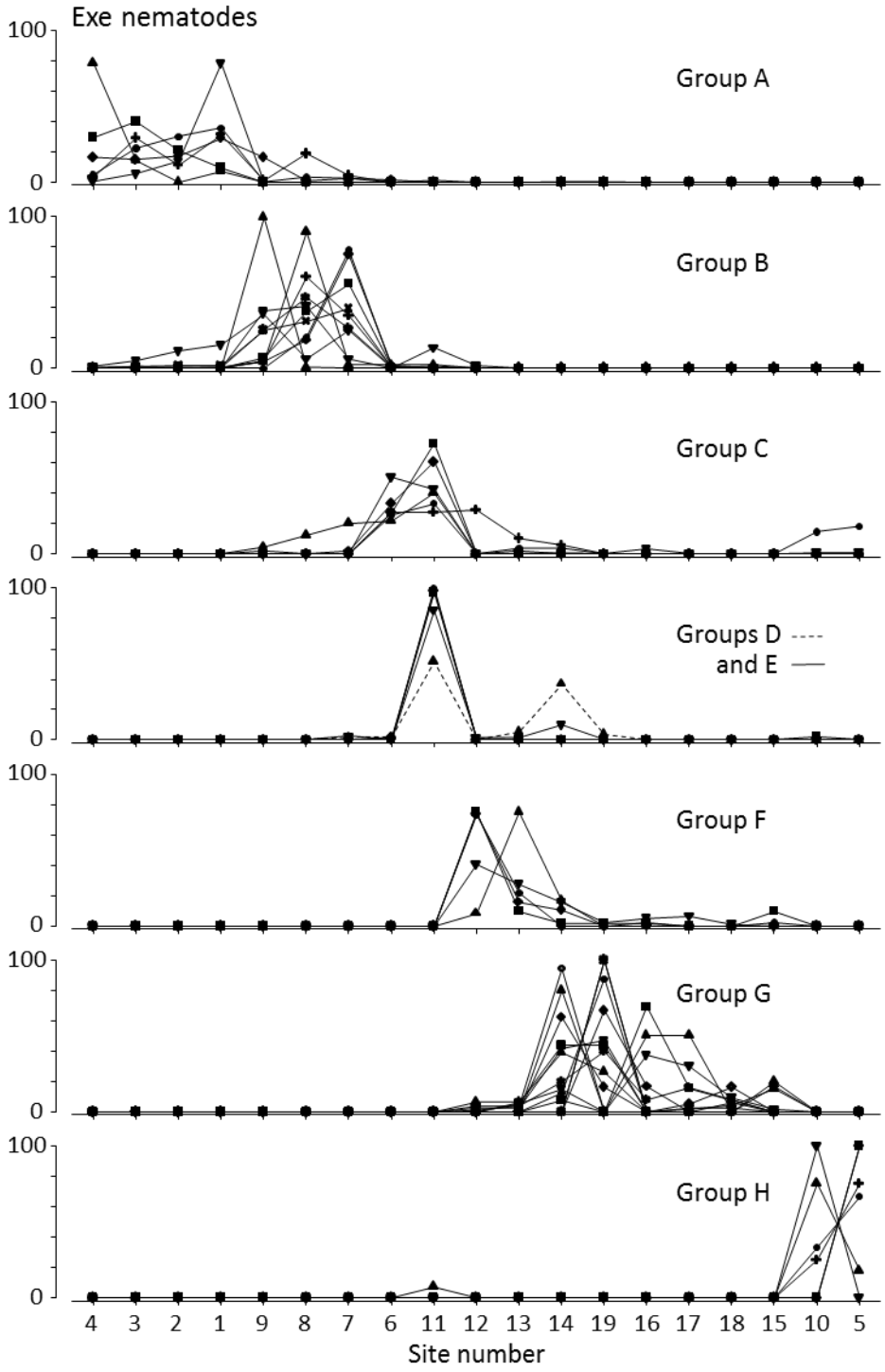
3 Somerfield and Clarke 2013 Figure 3

4




Somerfield and Clarke 2013 Figure 4

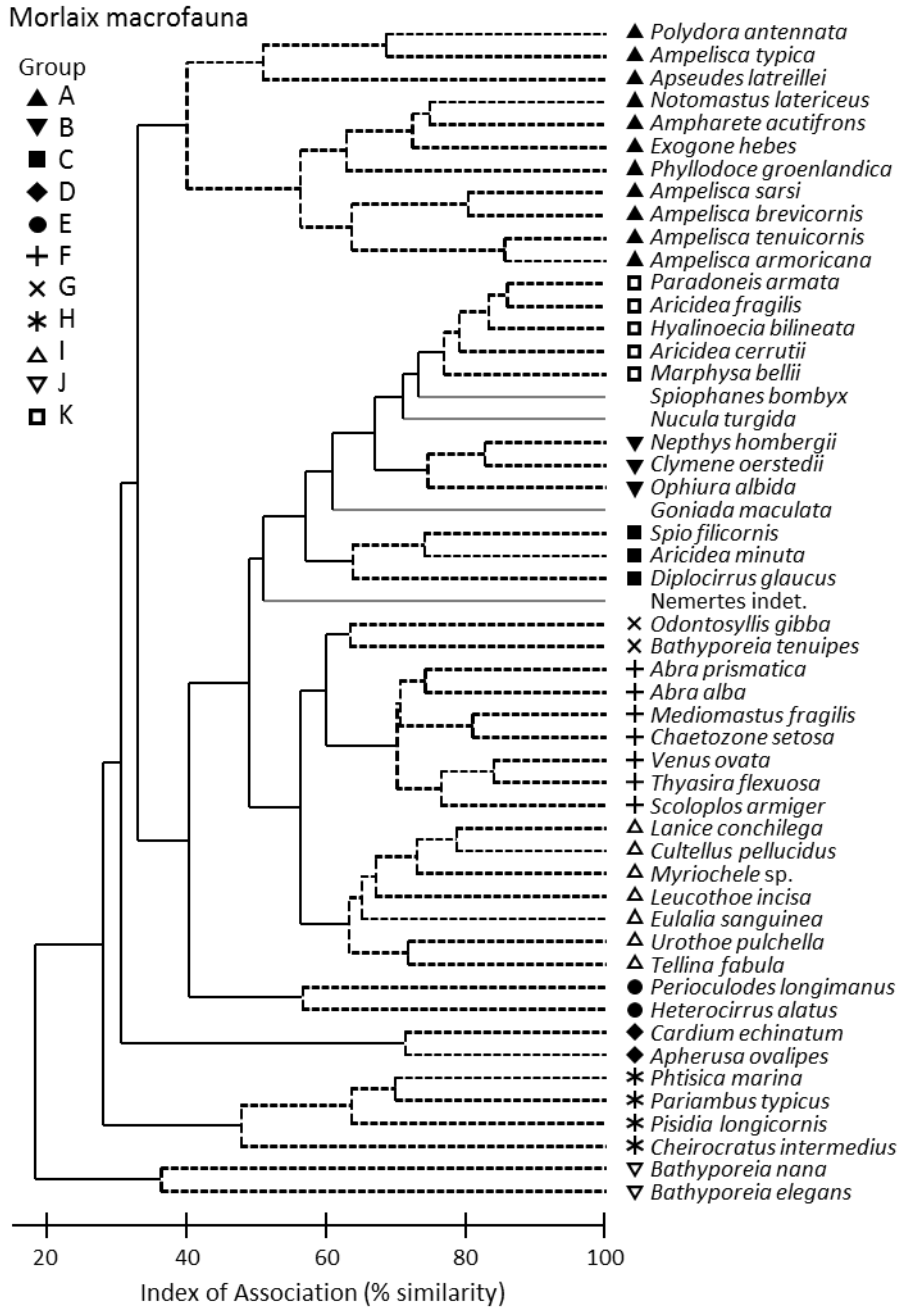
1




2

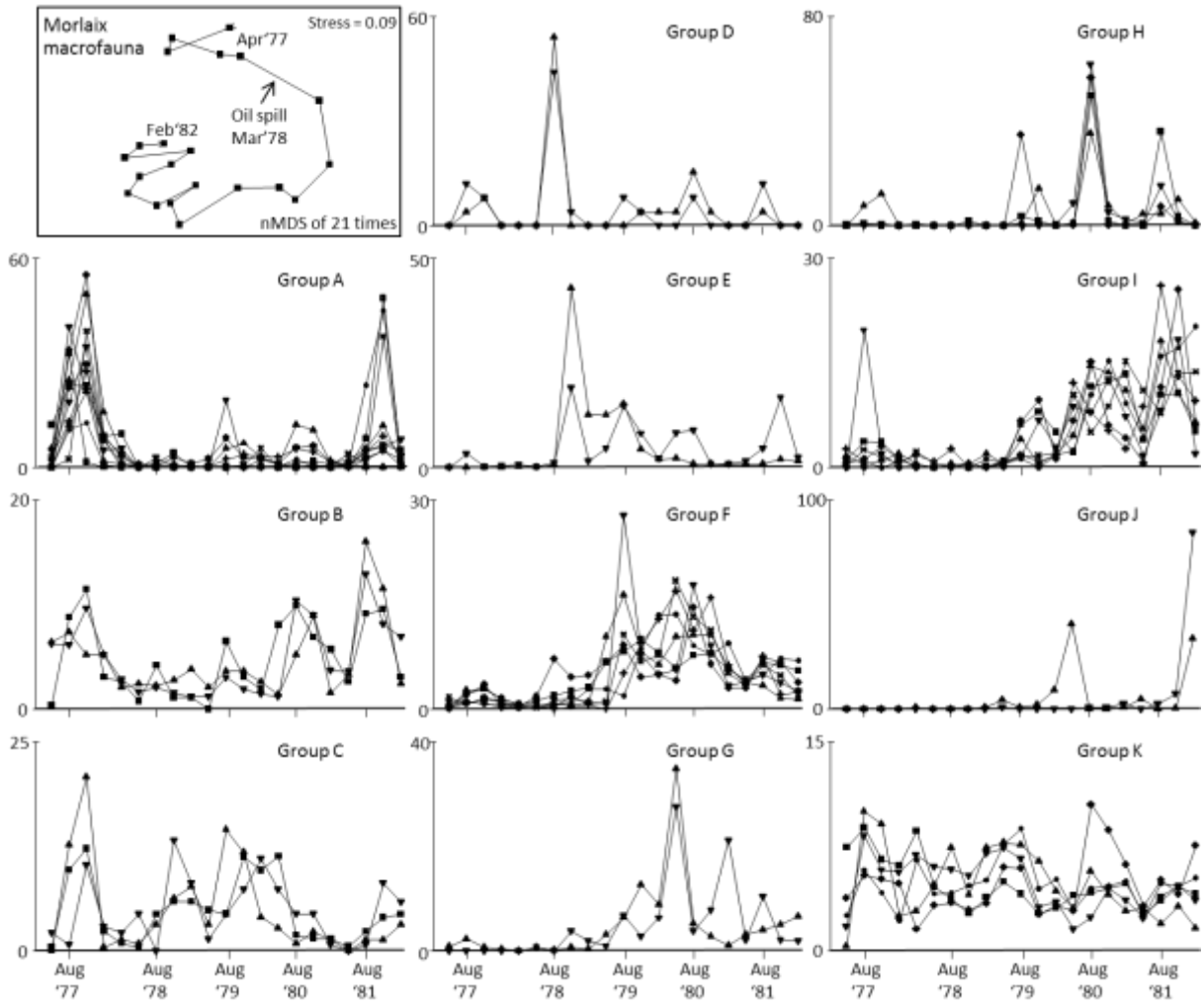
3 Somerfield and Clarke 2013 Figure 5

4




Somerfield and Clarke 2013 Figure 6

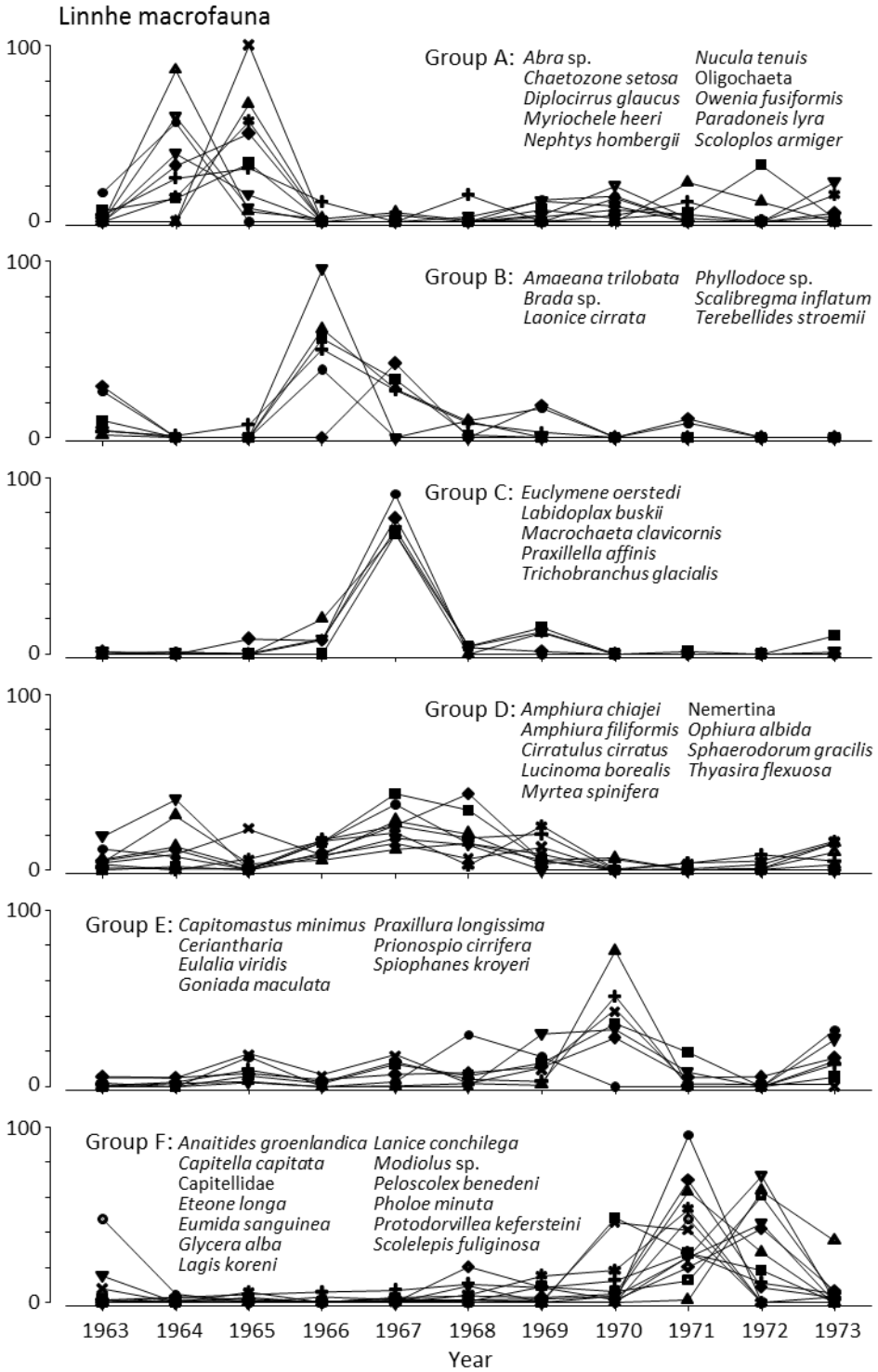
1




2

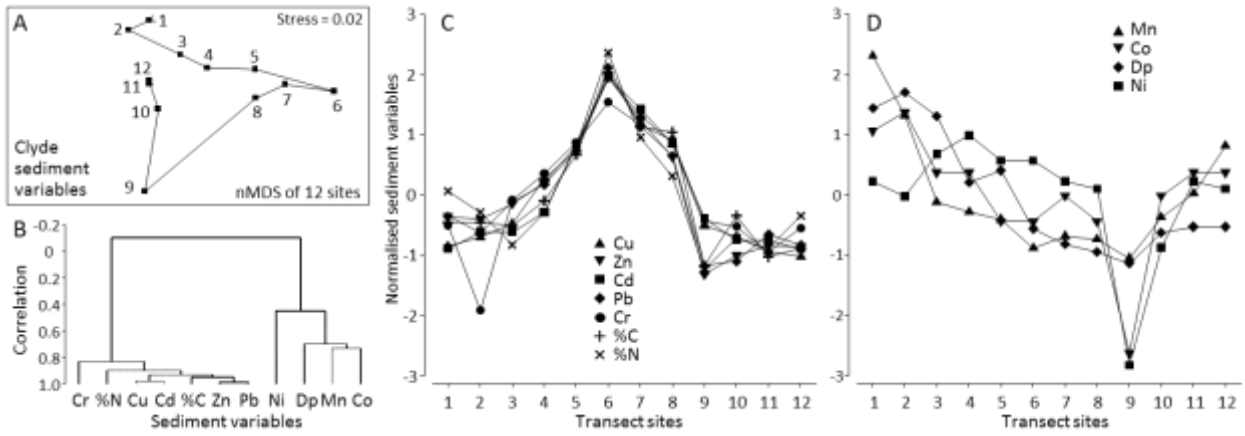
3 Somerfield and Clarke 2013 Figure 7

4




Somerfield and Clarke 2013 Figure 8

1



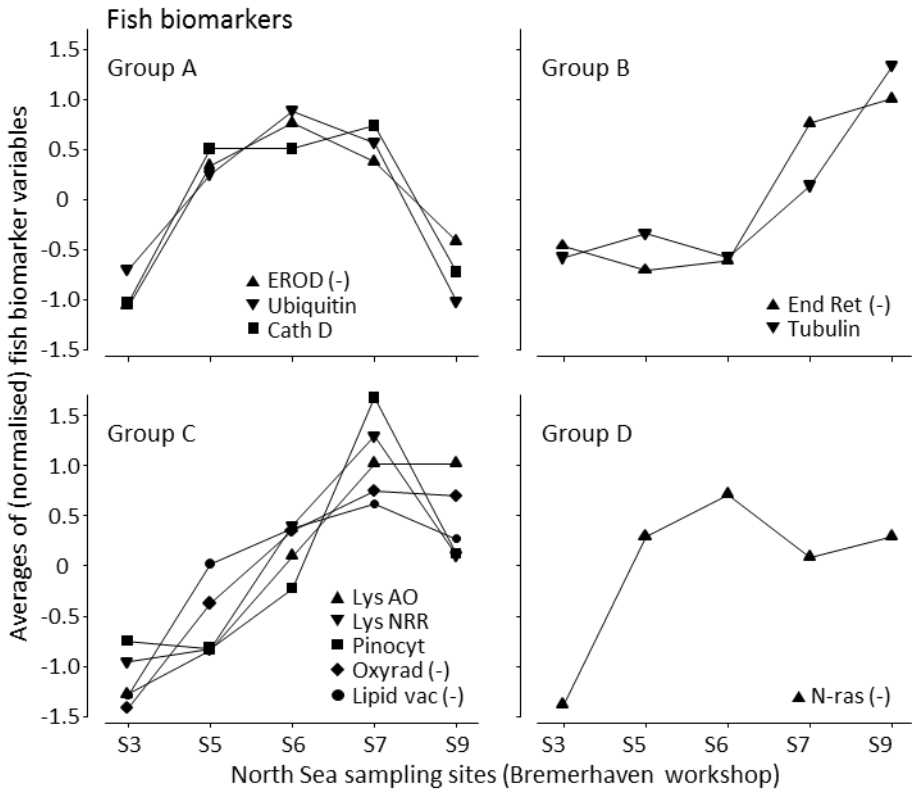

2

3 Somerfield and Clarke 2013 Figure 9

4

5

1




2

3 Somerfield and Clarke 2013 Figure 10