

A generalised analysis of similarities (ANOSIM) statistic for designs with ordered factors

PAUL J. SOMERFIELD,*¹  K. ROBERT CLARKE^{1,2} AND RAY N. GORLEY²

¹*Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK (Email: pjso@pml.ac.uk);*

and ²*Primer-E Ltd, c/o Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK*

Abstract In the study of multivariate data, for example of change in ecological communities, ANOSIM is a robust non-parametric hypothesis-testing framework for differences in resemblances among groups of samples. RELATE is a non-parametric Mantel test of the hypothesis of no relationship between two resemblance matrices. Details are given of the explicit link between the RELATE statistic, a Spearman rank correlation (ρ) between corresponding elements in the two resemblance matrices, and the ANOSIM statistic R , a scaled contrast between the among- and within-group ranks. It is seen that R can equivalently be defined as the slope of the linear regression of ranked resemblances from observations against ranked distances among samples, the latter from a simple model matrix assigning the values 1 and 0 to between- and within-group distances, respectively. Re-defining this model matrix to represent ordered distances among groups leads naturally to a generalised ANOSIM statistic, R^O , suitable for testing, for example, ordered factor levels in space or time, or an environmental or pollution gradient. Two variants of the generalised ANOSIM statistic are described, namely R^{Oc} where there are replicates within groups, and R^{Os} where there are only single samples (no replicates) within groups, for which an ANOSIM test was not previously available. Three marine ecological examples using ANOSIM to analyse an ordered factor in one-way designs are provided. These are: (1) changes in macrofaunal composition with increasing distance from an oil rig; (2) differences in phytal meiofaunal community composition with increasing macroalgal complexity; and (3) changes in average community composition of free-living nematodes along a long-term heavy metal gradient. Incorporating knowledge of an ordering structure is seen to provide more focussed, and thus stronger, ANOSIM tests, but inevitably risks losing power if that prior knowledge is incorrect or inappropriate.

Key words: hypothesis test, mantel test, multivariate analysis, ordered factors, seriation.

INTRODUCTION

Ecological studies that consider numerous variables (such as abundances or biomasses of different species) in a number of samples generate data matrices that are difficult to analyse using classical statistical approaches. To address the many statistical issues involved Field *et al.* (1982) described a robust non-parametric multivariate strategy for the analysis of such data. The analytical strategy was expanded and clarified by Clarke (1993) and continues to evolve (e.g. Clarke *et al.* 2014a). The essence of the strategy is, following appropriate pre-treatment of the data (Clarke *et al.* 2014b), to display patterns among samples as determined by appropriate resemblance measures (Clarke *et al.* 2006) using clustering (Clarke *et al.* 2016) and ordination, and to analyse these patterns using a range of hypothesis tests and associated analyses (e.g. Clarke *et al.* 2008), primarily based on ranked resemblances. In all that follows, inter-sample resemblances will be considered as dissimilarities for ease of discussion, but the methods are equally

applicable to other forms of resemblance such as similarities or distances.

A key formal hypothesis test within the framework is ANOSIM (Analysis of Similarities), a special form of Mantel (1967) test originally described for one-way layouts by Clarke and Green (1988). Classical one-way ANOSIM operates on an appropriate resemblance matrix calculated among samples, with a factor describing their *a priori* group structure (e.g. of different sites, times, treatments, etc.) underlying the null hypothesis to be tested, namely H_0 : ‘no differences among groups of samples’. If the null hypothesis is true, then the average rank resemblance among samples within groups is expected to be the same as the average rank resemblance among samples from different groups. The ANOSIM statistic R is defined as the scaled difference between the average between-group (\bar{r}_B) and within-group (\bar{r}_W) ranks:

$$R = \frac{(\bar{r}_B - \bar{r}_W)}{M/2} \quad (1)$$

where $M = n(n-1)/2$ and n is the total number of samples being considered. Clearly, under the null

*Corresponding author.

Accepted for publication March 2021.

hypothesis, R would be expected to take values (positive or negative) ‘close’ to zero, and increasing departure from H_0 would result in increasingly larger positive values for R . The scaling in equation (1) ensures that R falls within the range -1 to 1 , and takes the value $R = 1$ only under maximal separation of the groups, that is if all samples within groups (replicates) are less dissimilar to each other than any pair of samples from different groups. Values of R substantially less than 0 are not usually to be expected as this implies that samples within groups are generally less similar to each other than samples in different groups, a possibility only for a mislabelled or seriously inappropriate design. Note that the usual mathematical terminology for ranks assigns to the highest observation a rank value of 1 (the lowest number). This inversion of order can be very confusing so for equation (1), and throughout this paper, it is conceptually much simpler and more natural to reverse this and always define rank values as increasing with increasing dissimilarity, or model distance, between samples. Under this convention, the ranks in equation (1) are, therefore, rank dissimilarities (in early papers they were referred to as rank similarities).

If H_0 is true, then all samples effectively belong to a single group. The spread of possible values of R under the null hypothesis can be determined by randomly permuting the sample labels and recalculating R for each random reallocation, or for a random subset if there is a large number of possible permutations (Hope, 1968). The significance level of the observed value of R is then determined by comparing it to the range of values obtained under permutation, with rejection of the null hypothesis when the observed R is sufficiently large (positive) to have rarely or never occurred under permutation. This is a global test, and if it rejects the null hypothesis, pairwise tests may then be constructed by applying the same procedure to appropriate subsets of dissimilarities and re-ranking.

The non-parametric RELATE procedure tests the null hypothesis H_0 : ‘there is no relationship between two resemblance matrices’, using a matrix rank correlation and permutations (Somerfield & Gage 2000). Such a test is more obviously a type of Mantel test and may be used whether there are replicates (Somerfield *et al.* 2002) or not (Somerfield & Gage 2000). The ANOSIM test, using the R statistic, is formally equivalent to a RELATE test in which the inter-sample dissimilarities are correlated with a simple model matrix (e.g. see Legendre & Legendre 2012, pp. 608–611). Such models are distance matrices representing a specific idealised alternative to the null hypothesis, and thus guide construction of a test statistic effective in rejecting the null in favour of the (composite) alternative hypothesis of interest. Here, the alternative is H_1 : ‘there are (unspecified) differences between groups’, and the model specifies samples in the same group to be a distance 0 apart and samples from different groups a distance 1 unit apart (Fig. 1). The units of the model matrix are not important because correlation between matching elements is calculated having first ranked both matrices. Between the corresponding elements of the ranked resemblances among observed samples $\{r_i; i = 1, \dots, M\}$ and the model matrix ranks $\{s_i; i = 1, \dots, M\}$, the Spearman coefficient (e.g. Kendall, 1970) is then:

$$\rho = 1 - \frac{6}{M(M^2 - 1)} \sum_{i=1}^M (r_i - s_i)^2 \quad (2)$$

which again lies in the range $(-1, 1)$ with the extremes $\rho = -1$ and $+1$ corresponding to the cases where the two sets of ranks are in complete opposition or complete agreement, though as with negative R the former is unlikely in practice because of the constraints inherent in a dissimilarity matrix. Values of ρ around zero correspond to the absence of any match or concordance in pattern between the elements in the two matrices. The null

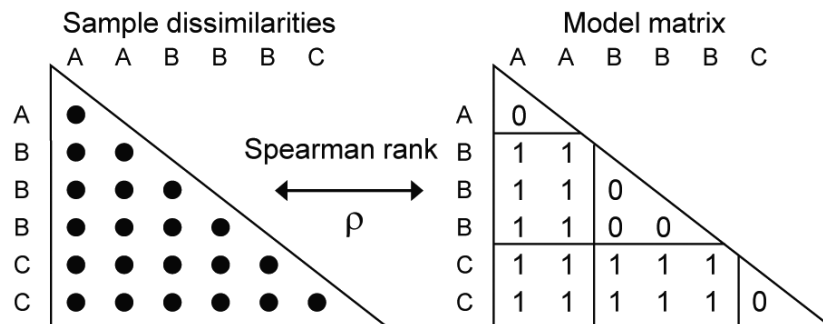


Fig. 1. Schematic diagram of a RELATE test of ‘no difference between groups’ for 3 groups (A, B, C) with 2, 3, 2 replicates showing the structure of the model matrix.

hypothesis here is H_0 : ‘no relationship between matrices’ and the spread of ρ values which are consistent with the null hypothesis can be determined by re-computing its value for random permutations of the sample labels in one of the two matrices (holding the other fixed). The significance of the test is then determined by comparing the observed ρ value with the distribution of values obtained from these random permutations.

As noted earlier, although the RELATE ρ statistic for the model matrix of Fig. 1 is not the same as an ANOSIM R statistic, the procedures (which permute the labels over samples in the same way for the two tests) produce test results which are equivalent for this simple one-way design.

Ordered models

Factors may be ordered in space or time. Examples could be that samples represent conditions along an environmental gradient, or times along a time-series, and it may be desirable to test for concordance of the observed dissimilarity structure with such hypothesised serial patterns (Sommerfield *et al.* 2002). Then, the test is not of the null:

$$H_0: A = B = C = D = \dots$$

against the general alternative

$$H_1: A, B, C, D, \dots \text{ differ (in ways unspecified)}$$

but of the same null H_0 against an ordered alternative, denoted symbolically as:

$$H_2: A < B < C < D \dots$$

Both alternative hypotheses H_1 and H_2 are composite (an infinite amalgam of simple hypotheses), as in a conventional univariate ANOVA, but the multivariate context makes precise mathematical definition cumbersome. The differing focus of the two alternatives is clear, however, a test statistic designed for alternative H_1 should have some power to detect differences between two or more of the groups in any

arrangement of those groups in multivariate space. A statistic designed for alternative H_2 , however, should be more narrowly focussed on detecting a steadily increasing separation of the successive groups. Thus A and B, or B and C, are more similar (less different) than A and C or B and D, which themselves are more similar than A and D, and so on. Hypothesis H_2 is thus an appropriate alternative for testing, say, for an inter-annual drift in an assemblage away from its initial state, or for serial change in community composition along an environmental gradient such as increasing water depth or levels of a nutrient or pollutant. A schematic of a simple model matrix, effective in constructing a form of RELATE test appropriate to the alternative hypothesis H_2 , is exemplified in Fig. 2, the ‘seriation with replication’ test (Sommerfield *et al.* 2002). Here, samples in the same group are considered to be distance zero units apart, in adjacent groups A to B and B to C one unit apart, and A to C two units apart. The RELATE test then simply correlates the observed dissimilarity ranks $\{r_i\}$ against model distance ranks $\{s_i\}$. Sommerfield *et al.* (2002) showed that if such groups of replicates really are serially ordered then this RELATE approach will have more statistical power to detect differences than an equivalent ANOSIM test for (unordered) differences among groups.

Both ANOSIM and RELATE are forms of Mantel test, and between them they allow formal hypothesis tests to support analyses of differences among samples reflecting differences among groups, of common patterns where there are no groups, of one-way and two-way designs with or without replication, and they offer alternative approaches if factors are, or are not, ordered. The purpose of this paper is to detail the difference in algebraic form and contextual motivation of these two formulations, and to demonstrate that they are best unified in a single testing framework for exploring relationships of samples to hypothesised models.

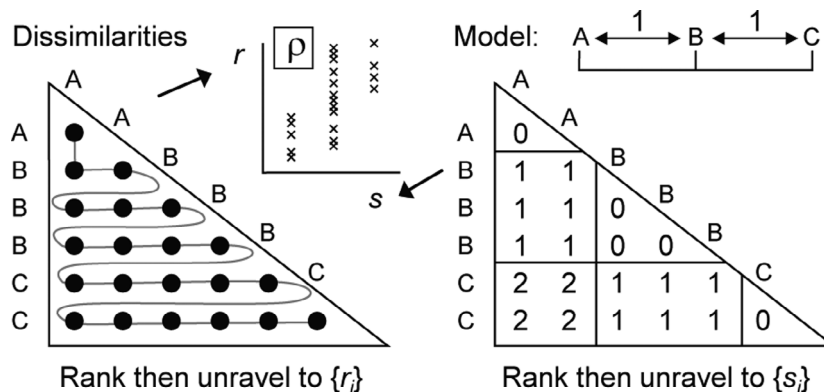


Fig. 2. Schematic diagram of a ‘seriation with replication’ RELATE test for samples in 3 ordered groups.

METHODS

Relation of standard ANOSIM R to RELATE ρ for unordered groups

For a structure in which groups A, B, C, ... are unordered, the model ranks $\{s_i\}$ are constructed from model distances of 0 between any samples in the same groups and 1 between any two samples in different groups (expanding Fig. 1 to the case of many groups). The choice of 0 or 1 is arbitrary, and could be replaced by any constants a and b such that $a < b$, since these distances are then ranked.

Defining $i = 1, \dots, w_0$ as indexing the within-group dissimilarities and $i = w_0 + 1, \dots, w_0 + w_1$ as among-group dissimilarities (where $w_0 + w_1 = M$), ranking the model distances gives the two sets of tied ranks:

$$s_i = \begin{cases} (w_0 + 1)/2 & \text{for } i = 1, \dots, w_0 \\ w_0 + (w_1 + 1)/2 & \text{for } i = w_0 + 1, \dots, w_0 + w_1 \end{cases} \quad (3)$$

the result following simply from the sum of the first n integers being $n(n + 1)/2$.

The RELATE (Spearman ρ) rank correlation is simply the Pearson correlation of the two sets of ranks $\{r_i\}$ and $\{s_i\}$, namely:

$$\rho = S_{rs} / \sqrt{S_{rr} \cdot S_{ss}} \quad (4)$$

where $S_{rs} = \sum_{i=1}^M (r_i - \bar{r})(s_i - \bar{s})$, $S_{ss} = \sum_{i=1}^M (s_i - \bar{s})^2$ and $S_{rr} = \sum_{i=1}^M (r_i - \bar{r})^2$ denote the usual sums of cross-products and sums of squares.

The key to establishing the results of this paper is to note that:

$$\rho = (S_{rs} / S_{ss}) \cdot \sqrt{S_{ss} / S_{rr}} = \beta \cdot \sqrt{S_{ss} / S_{rr}} \quad (5)$$

where β is the fitted slope of the least-squares linear regression of $\{r_i\}$ on $\{s_i\}$. A simple special case of three widely separated groups is illustrated in Fig. 3. It is a standard result for a 2-point linear regression, with mean y values of \bar{y}_1 and \bar{y}_2 at the only two x values x_1 and x_2 , that the fitted line joins the two means \bar{y}_1 and \bar{y}_2 irrespective of the numbers of observations making up each mean. Denoting, as usual, the average rank between (among) all groups by \bar{r}_B , and within groups by \bar{r}_W , using (3) the regression slope is therefore:

$$\begin{aligned} \beta &= (\bar{r}_B - \bar{r}_W) / \left[\frac{2w_0 + w_1 + 1}{2} - \frac{w_0 + 1}{2} \right] \\ &= (\bar{r}_B - \bar{r}_W) / (M/2) \end{aligned} \quad (6)$$

This is the definition of the standard ANOSIM statistic R , for the one-way unordered groups case, and motivates the generalisation of R to the ordered groups case (and any other model for the $\{s_i\}$) as the slope of the linear regression of observed ranks $\{r_i\}$ on model ranks $\{s_i\}$.

In the simple unordered case, it follows from (5) that

$$R = \rho \sqrt{S_{rr} / S_{ss}} \quad (7)$$

and since both \bar{r} and \bar{s} must always be $(M + 1)/2$ for ranks, whether there are tied ranks or not:

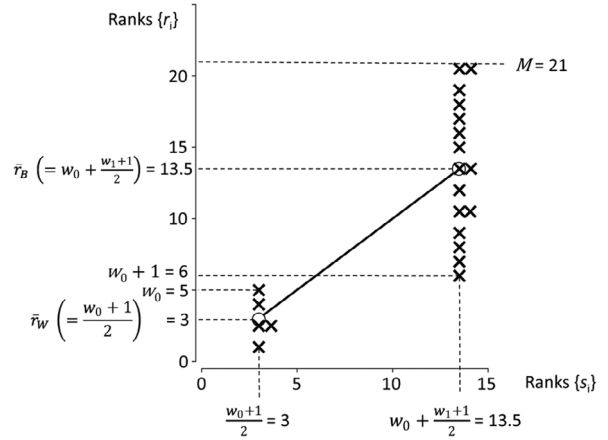


Fig. 3. Scatter plot of $\{r_i\}$ against $\{s_i\}$ for a simple case of 3 widely separated groups with 2, 3, 2 samples (so with 5 within-group dissimilarities and 17 between-group dissimilarities) with some ties in the $\{r_i\}$ (indicated by dithered symbols), showing the key elements discussed in the text. Mean ranks in $\{r_i\}$ within and between groups indicated by circles. The slope β of the linear regression is, and has to be, 1.

$$\begin{aligned} S_{ss} &= \sum (s_i - \bar{s})^2 \\ &= w_0 \left[\frac{w_0 + 1}{2} - \frac{M + 1}{2} \right]^2 + w_1 \left[\frac{2w_0 + w_1 + 1}{2} - \frac{M + 1}{2} \right]^2 \\ &= w_0 (w_1/2)^2 + w_1 (w_0/2)^2 \\ &= Mw_0w_1/4 \end{aligned} \quad (8)$$

The general expression for S_{rr} is a little more complex (not shown), but if there are no tied dissimilarity values (not unrealistic in many contexts) it simplifies, since

$$\sum r_i^2 = \sum i^2 = M(M + 1)(2M + 1)/6 \quad (9)$$

and

$$\begin{aligned} S_{rr} &= \sum r_i^2 - M\bar{r}^2 \\ &= M(M + 1) \left[\frac{2M + 1}{6} - \frac{M + 1}{4} \right] \\ &= M(M^2 - 1)/12 \end{aligned} \quad (10)$$

Hence, from (7),

$$R = \rho \sqrt{\frac{(M^2 - 1)}{3w_0w_1}} \quad (11)$$

and it is clear that R and ρ are effectively the same statistic in this simple case, differing only by a proportionality constant. An important point is that the proportionality constant is purely a function of the design (w_0 , w_1 and their sum M are not functions of the data $\{r_i\}$, only of the design). Clearly, the equations defining slopes and correlations dictate that β ($=R$) is zero if, and only if, ρ is zero, but a key issue for comparability across different tests is now to establish their respective maximum values.

Figure 3 shows equally clearly that the slope must be maximised for this 2-point regression when the lower of the two y -axis means is as small as possible and the upper mean as large as possible, that is the within-group ranks $\{r_i\}$ are all (strictly) lower than the between-group ranks, thus:

$$\bar{r}_W = (w_0 + 1)/2, \quad \bar{r}_B = w_0 + (w_1 + 1)/2 \quad (12)$$

As indicated in the figure, this is irrespective of ties in $\{r_i\}$, provided any ties are wholly inside the within-group and between-group sets, and not straddling the boundary between $i = w_0$ and $i = w_0 + 1$.

The means in (12) are then simply the two discrete values on the x -axis, $s_1 = (w_0 + 1)/2$ and $s_2 = w_0 + (w_1 + 1)/2$, so the slope is clearly maximised at $\beta (= R) = 1$, the well-known maximum from the original derivation of ANOSIM R . The same is not true of ρ . Though $\rho \leq 1$ is axiomatic for a correlation coefficient, it is evident from Fig. 3 that a perfect correlation $\rho = 1$ (i.e. all points on a scatter plot are co-linear) is never attainable – except for an unlikely case in which there are only two distinct observed dissimilarity values (e.g. zero within groups and 100 % between groups). R is, therefore, a more useful descriptive statistic than ρ for interpretation and comparison of the magnitude of between-group differences across different tests, since it always takes a maximum value of 1 when all dissimilarities between groups are greater than any dissimilarity within groups. In the latter case, ρ is seen to be less than 1, and its attained maximum varies not only with the design constants w_0 and w_1 but can also vary with the data, in cases where some observed dissimilarities are equal to each other, giving ties in the ranks $\{r_i\}$.

Maximum value for β in the general case

In the general case, the $\{s_i\}$ represent any set of ranks from modelled ‘distances’, not only the seriation model (equi-spaced samples in time or space) which is the main thrust of this paper. Other examples might include unequal spacing, circularity (e.g. seasonality) and geographical or environmental ‘distances’ among sample locations. An alternative to the (Spearman) RELATE correlation ρ between observed ranks $\{r_i\}$ and model ranks $\{s_i\}$ is, therefore, the slope (β) of the least-squares linear regression of $\{r_i\}$ on $\{s_i\}$.

To illustrate the simplest case, Fig. 4 shows the scatter plot of ranks in observed dissimilarities $\{r_i\}$ against the model ranks $\{s_i\}$ for serially ordered groups A, B and C with 2, 3 and 2 replicates per group, where the groups are maximally separated, and there are no ties in the observations. It is intuitively clear from Fig. 4 that the slope is maximised as the ranks $\{r_i\}$ on the y -axis are separated into ever-increasing values at each increasing step in s on the x -axis. But the regression slope in this case needs to be determined, under the fully general scenario with many discrete steps for model ranks $\{s_i\}$ on the x -axis, bearing in mind that there could also be ties within each of the separated sets of $\{r_i\}$ ranks on the y -axis.

In any situation representing maximum separation among groups, the $\{s_i\}$ ranks are grouped into $j = 1, 2, \dots, k$ tied sets, with the j th set (Φ_j) having t_j members, each of tied

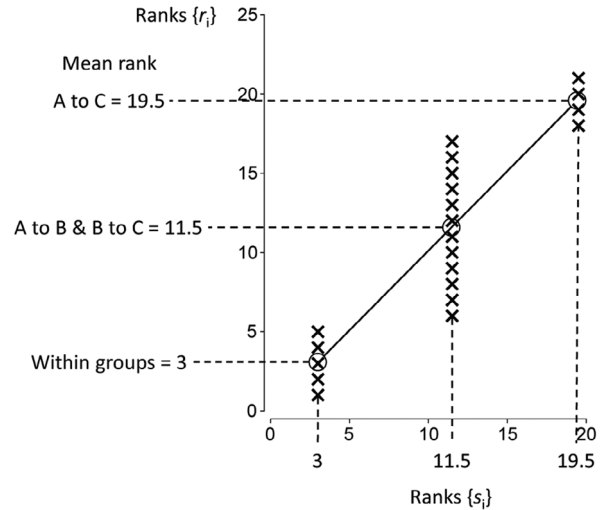


Fig. 4. Scatter plot of $\{r_i\}$ against $\{s_i\}$ for 3 groups with 2, 3, 2 replicates respectively showing the case for the most extreme ordered separation of the groups $A < B < C$. The y -axis means are denoted by circles.

value $c_j (=1, \dots, k)$. The key step to note (self-evident from Fig. 4 but carrying over to the fully general case) is that under this best possible separation, the sum of the t_j ranks from the j th step of the s -axis, namely $t_j c_j$, is always equal to the sum of the matching ranks on the r -axis. This is irrespective of whether there are also ties within the $\{r_i\}$ ranks, on condition (as before) that none of those ties is across adjacent s -axis sets (strict separation). The absence of any overlap (or equality) of values on the y -axis (for r_i) across the sets of tied ranks on the x -axis (s_i values) dictates that the respective r -axis means are also $\{c_j; j = 1, \dots, k\}$. It seems intuitively clear, therefore, that the slope of the regression line $\beta = 1$ in that case, and that β can never exceed 1.

However, the well-known ‘regression to the mean’ in standard (asymmetric) linear regression of y on x , and the unbalanced numbers of $\{r_i\}$ values at each s -step, make it necessary to prove this intuitive conclusion in the fully general case, so technically this is demonstrated as follows:

$$S_{sr} = \sum_{i=1}^M \left(r_i - \frac{M+1}{2} \right) \left(s_i - \frac{M+1}{2} \right) = \sum_{j=1}^k \left[\sum_{i \in \Phi_j} \left(r_i - \frac{M+1}{2} \right) \right] \left(c_j - \frac{M+1}{2} \right) \quad (13)$$

$$= \sum_{j=1}^k \left[t_j c_j - t_j \left(\frac{M+1}{2} \right) \right] \left(c_j - \frac{M+1}{2} \right) = \sum_{j=1}^k t_j \left[c_j - \frac{M+1}{2} \right]^2$$

$$S_{ss} = \sum_{i=1}^M \left(s_i - \frac{M+1}{2} \right)^2 = \sum_{j=1}^k t_j \left[c_j - \frac{M+1}{2} \right]^2 \quad (14)$$

so $S_{rs} = S_{ss}$ and therefore the slope $\beta = S_{rs}/S_{ss} = 1$, which clearly cannot be exceeded by any other configuration of the $\{r_i\}$ ranks. In other words, if all dissimilarities between

pairs of samples deemed further apart under the model are strictly greater than any dissimilarities between pairs of samples considered closer together by the model, then $\beta = 1$.

In contrast, Fig. 4 makes it clear that ρ does not attain its theoretical maximum value 1 even in this simple case of maximal separation, as the observational ranks are scattered about the regression line. This will almost always be the case for real replicated data sets: maximal separation of groups with $\beta = 1$ will give correlations ρ which are < 1 , making β the more useful of the two statistics for practical interpretation.

In the fully general case, the relationship between β and ρ depends on the structure of ties in the observation ranks $\{r_i\}$, as well as the model design $\{s_i\}$, though not on the interplay between the two sets of ranks – that is clear from the general formulation of equation (5). Following on from the unordered case earlier, where a simple relation was given (equation 11) for the design-specific proportionality constant between β and ρ (when there are no ties in the response ranks r), it is worth noting how that relationship changes for a three-group seriation of the type seen in Fig. 2 (but with any number of replicates within each group), as follows.

Under the usual seriation model in which A to B and B to C are considered as the same distance apart (say 1) and A to C is larger (say 2), let w_0 denote the number of within-group ranks, w_1 the number of A to B plus B to C ranks, and w_2 the number of A to C ranks, so that $M = w_0 + w_1 + w_2$. Assuming, as previously, that the observed ranks $\{r_i\}$ are not tied, the expression (derivation not shown) becomes:

$$\beta = \rho \sqrt{\frac{M^2 - 1}{3 \cdot (w_0 + w_2) \cdot (w_1 + \frac{w_0 \cdot w_2}{M})}} \quad (15)$$

in which the w_0 on the bottom line of the previous expression (equation 11) now gets replaced by $w_0 + w_2$, and the w_1 on the bottom line gets replaced by $w_1 + (w_0 \cdot w_2)/M$, which is $\geq w_1$. It is apparent that if $w_2 = 0$ this reduces to the previous formula, representing the unordered group case. Thus again, in this simple seriation case, β (a generalisation of R) and ρ are effectively the same test statistic, differing only by a proportionality constant that is solely a function of the design being analysed (note that this will not be true if there are ties in the observed ranks $\{r_i\}$).

A generalised ANOSIM statistic

The two statistics ρ and β are closely related, differing in numerical value according to a proportionality constant dictated only by the particular design being analysed (where there are no ties in the observed dissimilarities). The slope β of the linear regression of $\{r_i\}$ on $\{s_i\}$ reduces to the usual ANOSIM R statistic in the unordered case. The equations defining slopes and correlations dictate that this slope is zero if, and only if, ρ is also zero. The slope of the regression can never exceed 1 and it takes that value only under a generalisation of the (non-parametrically) most extreme multivariate separation that can be observed between groups. This was previously characterised by the mantra:

‘all dissimilarities between groups are larger than any within groups’, to which we now must add: ‘and all dissimilarities between groups which are placed further apart in the model matrix are larger than any dissimilarities between groups which the model puts closer together’. Thus, the generalised ANOSIM statistic is here defined as the slope of the linear regression of $\{r_i\}$ on $\{s_i\}$, and this definition applies whatever the form of the model matrix ranks $\{s_i\}$. Extending the nomenclature of ANOSIM, it is denoted in the particular case of ordered groups by R^O (the superscript upper-case O denoting ‘ordered’). Testing of this statistic uses the appropriate permutation distribution as before, because standard tests (or interval estimates) for the slope of the regression cannot be used owing to the high degree of internal dependency among the $\{r_i\}$ (dissimilarities are not mutually independent).

It is useful to further distinguish two cases for the ordered one-way ANOSIM statistic, namely ordered category and ordered single statistics, denoted by R^{Oc} and R^{Os} . The difference is simply that the notation R^{Oc} is used when the data have replicates within groups, so that it is influenced by both the presence of group structure and by the ordering of those groups, whereas R^{Os} refers to one-way layouts with no replicates, where the test is then entirely based on whether or not there is a serial ordering (trend) in the multivariate pattern of the ‘groups’ (i.e. single samples in this case), in the specified order. Technically, the computation is no different: both are simply the slope of the regression of the ranks $\{r_i\}$ on $\{s_i\}$, though clearly the un-replicated design requires a reasonable number of ‘groups’ (at least 5, in the one-way case) to generate sufficient permutations to have any prospect of demonstrating serial change.

Example datasets

Ekofisk

Hobbs (1987) surveyed benthic macrofauna in soft sediments around the Ekofisk oil platform in the North Sea, and Gray *et al.* (1990) analysed these data in a multivariate context. Thirty-nine sites at different distances (100 m to 8 km) and different directions away from the oil platform were sampled, to examine evidence for changes in the assemblage with decreasing distance to the oil rig. The sites were allocated (somewhat arbitrarily, but *a priori*) into 4 distance groups, A: >3.5 km from the rig (11 sites), B: 1–3.5 km (12 sites), C: 250m–1 km (10 sites) and D: <250 m (6 sites). At each site, 3 samples were taken with a Day grab (0.1 m²), sampling to a minimum depth of 10 cm. The samples were sieved on a 1 mm sieve to extract the fauna which was subsequently preserved and identified. Samples from each site were pooled prior to analysis. Bray–Curtis resemblances were calculated between sites following a square-root transformation of the pooled abundances.

As sites are grouped into distance classes at decreasing distances to the oil-field centre, an ordered one-way ANOSIM test, with sites used as replicates for the four distance groups, seems preferable here to the standard (unordered) ANOSIM. Though the null hypothesis $H_0: A = B = C = D$ is the same, the ordered hypothesis $H_1: A < B < C < D$ is

an appropriate alternative for directed community change with distance. There is no need for the test to have power to detect an (uninterpretable) alternative in which, for example the communities in D are very different from C and B but then very similar to A, so by restricting the alternative to a smaller set of possibilities, a more powerful test statistic R^O for detecting that alternative, and for appropriately measuring its magnitude, may be employed.

St Marys phytal meiofauna

Gee and Warwick (1994a,b) studied the relationship between the fractal dimension of the physical structure of species of macroalgae, and the community structure of the fauna inhabiting those algal species. Algae were sampled from a series of sites in the Isles of Scilly, UK. Organisms were subsequently extracted from the algal samples, identified and counted. Full details are to be found in the original papers (Gee & Warwick 1994a,b). Here data on the abundances of 99 taxa (species or putative species) of meiofauna from three species of macroalgae, collected from four sites on the island of St Marys, are analysed. For organisms in the meiofaunal size range, the three species of algae differ in their fractal dimension and complexity, in the order *Chondrus crispus* < *Lomentaria articulata* < *Cladophora rupestris* (Gee & Warwick 1994a,b), so as an alternative to an unordered test for differences among species of algae, a test for the specific ordering of meiofaunal communities inhabiting those algae in relation to their structural complexity could also be informative. Although all samples represent equal volumes of alga, it is likely that samples differ in terms of the amount of surface area and living space sampled, so samples were standardised (converted to percentages) prior to the analysis. Also, abundances of some organisms were orders of magnitude greater than others, so the standardised data matrix was fourth-root transformed prior to calculating resemblances with the Bray–Curtis coefficient.

Fal nematodes

Somerfield *et al.* (1994) sampled meiofaunal communities in five creeks in the Fal estuary system, Cornwall, UK. The area has a history of metal mining going back to the Bronze Age, and the accumulation of waste and spoil runoff has resulted in levels of metals in the sediments in various creeks that differ considerably. The five creeks sampled, therefore, represent different points along a long-term gradient in heavy metal contamination, in the order Restronguet, Mylor, Pill, St Just and Percuil, with the highest metal concentrations in Restronguet and these concentrations then reducing by approximately a half from each creek to the next. The data used here are free-living nematode abundances, fourth-root transformed and then averaged within each creek to give five ordered samples.

Data analyses

All the analyses were undertaken with PRIMER v7 (Clarke & Gorley 2015). Testing utilised the ANOSIM and RELATE routines. Resemblances were ordinated using non-metric multidimensional scaling (nMDS).

RESULTS

Ekofisk

Figure 5a shows the nMDS for the 39 sites, with the 4 distance groups (differing symbols) clearly showing a pattern of steady community change with decreasing distance towards the oil rig. Figure 5b plots the $39 \times 38/2 = 741$ rank dissimilarities $\{r_{ij}\}$ against the (ordered) model ranks $\{s_i\}$, the four sets of tied ranks for the latter representing (left to right): within A, B, C or D; then A to B, B to C or C to D; then A to C or B to D; and finally A to D. The fitted regression of r on s has a strong slope of $R^O = 0.656$, the ordered ANOSIM statistic, and this is larger than its value for 9999 random permutations of the group labels to the 39 samples, so $p < 0.01\%$ at least (and here it would clearly be significant at effectively any proposed significance level chosen *a priori*). The contrast is with a standard (unordered) ANOSIM test which records the lower (though still highly

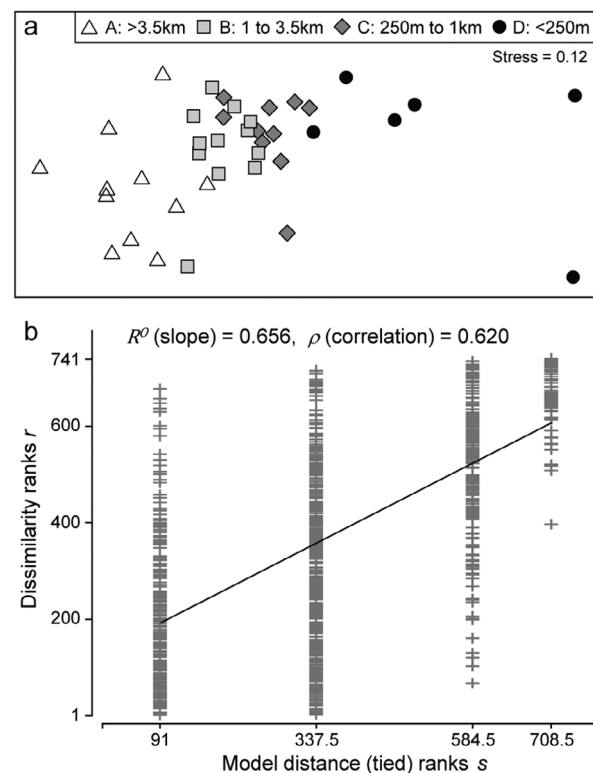


Fig. 5. Ekofisk macrobenthos. a) nMDS of the 39 sites (from square-root transformed abundances of 173 species and Bray–Curtis similarities), with the four distance groups from the oil rig indicated by differing symbols. b) Scatter plot of rank dissimilarities (r) among the 39 sites against tied ranks (s) from a serial-ordering model of groups, showing the fitted regression line with slope R^O , the ordered ANOSIM statistic.

significant) value of $R = 0.54$. Clearly, if there are only two groups, R^O and R become the same statistic, so the pairwise tests between all pairs of groups which follows this (global) ordered ANOSIM test are the same as for the usual unordered analysis.

For the four Ekofisk distance groups, the pairwise R values show the pattern expected from a gradient of change: for groups one step apart (A to B, B to C, C to D), $R = 0.56, 0.16, 0.55$; for two steps (A to C, B to D), $R = 0.76, 0.82$; and for three steps (A to D), $R = 0.93$ (all 'significant' by conventional criteria, $p < 1\%$). In this case, Fig. 5b clearly demonstrates how the (global) R^O captures both the standard ANOSIM R 's contrast of within and between-group ranks (the left-hand set of points *vs.* the right-hand three sets) and the regression relation of greater change with greater distance (the right-hand three). This will not always be the case, however, since clear group differences which are not ordered in the way postulated by the alternative hypothesis can lead to lower (and possibly non-significant) values of R^O than obtained by the standard ANOSIM R . An example of this is seen later.

St Marys phytal meiofauna

Figure 6 shows a steady pattern of meiofaunal community change among species of macroalgae consistent with the increasing fractal dimension of the algal species. The observed value of the ordered ANOSIM statistic is $R^O = 0.716$. Only one of the $12!/(4!4!4!2) = 17325$ potentially distinct permutations yields a value ($R^O = 0.724$) greater than the observed value, so $p = 0.012\%$. The pairwise tests give values of R (unordered) of 0.42 ($p = 8.6\%$) between *Chondrus*

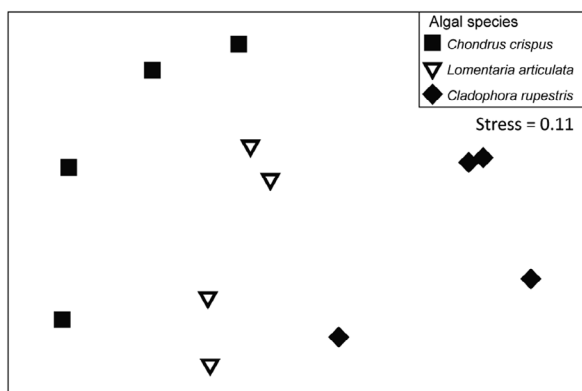


Fig. 6. St Marys phytal meiofauna. nMDS of the meiofaunal samples from each of three species of macroalgae, representing a gradient of increasing fractal dimension and complexity, from four sites on the island of St Marys in the Isles of Scilly, UK. For organisms in the meiofaunal size range *Chondrus crispus* is the least complex, and *Cladophora rupestris* is the most complex.

and *Lomentaria*, 0.76 ($p = 2.9\%$) between *Lomentaria* and *Cladophora*, and 0.91 ($p = 0.29\%$) between *Chondrus* and *Cladophora*, the species which differ most. The test for unordered differences among species of algae gives $R = 0.67$, a smaller value of R than R^O for the ordered test. It also has a p value which is somewhat less extreme, $p = 0.09\%$, since 5 out of the possible $12!/(4!4!4!3!) = 5775$ potentially distinct permutations yielded a value that was equal to or greater than the observed value of R , one of which was, naturally, for the original labelling. It is worth noting here that in addition to selecting a more appropriate and interpretable statistic in this context, an ordered ANOSIM procedure permits a stronger test (Sommerfield *et al.* 2002), as reflected in the increased number of potentially distinct permutations (17325 *cf.* 5775). That smaller numbers of potentially distinct permutations can generate less powerful tests can be seen *in extremis* in the final example, a situation in which the standard ANOSIM procedure is powerless, there being no possible permutations other than the observed configuration.

Fal nematodes

If there is no replication within groups, the standard ANOSIM R is undefined and no test is possible. Because the factor (Creek) is ordered, however, it is possible to test for a linear sequence, consistent with the community structure of the nematodes becoming increasingly dissimilar with increasing differences in the levels of metals in the sediment, by using the R^{Os} statistic. The ordered ANOSIM test gives $R^{Os} = 0.74$, which is the most extreme of all 60 possible permutations, so $p = 1.7\%$.

DISCUSSION

Defining a generalised ANOSIM R statistic for ordered levels of factors allows the unification of different hypothesis tests into a common framework. All variants of the ANOSIM statistic (R, R^{Oc}, R^{Os}) take a value centred at 0 if the appropriate null hypothesis is true, and a value of exactly 1 in the most extreme case of the alternative hypothesis, maximal separation or ordering, as previously described. Values of the statistics are interpretable as scaled measures of effect size, and are therefore comparable across different tests and even different datasets. The significance of observed values of the ANOSIM statistics can be tested using relevant sets of permutations. The tests require no distributional assumptions and are fully non-parametric.

The R^O statistic is closely linked to the matrix correlation ρ , differing in numerical value largely by a design-specific proportionality constant (entirely so if

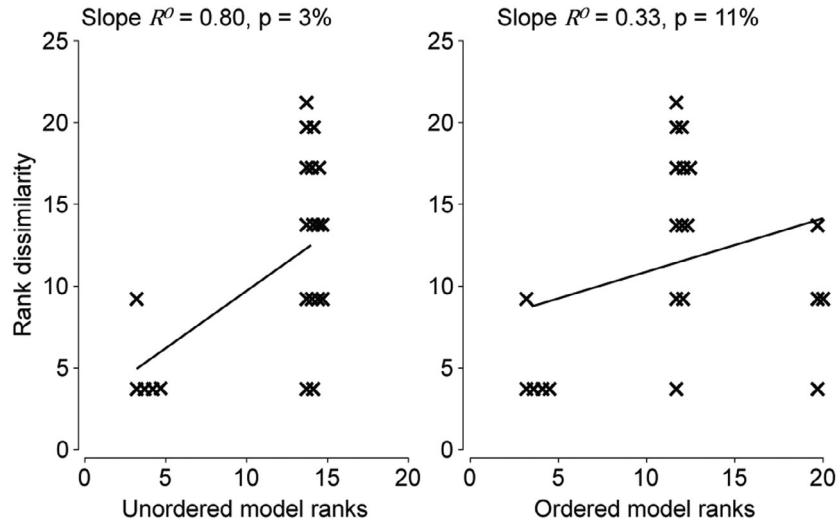


Fig. 7. Scatter plots of dissimilarity ranks $\{r_i\}$ against unordered and ordered model ranks $\{s_i\}$ for sample data in groups (A < B < C) of 2, 3, 2 replicates, as in Fig. 2, where the groups are not actually ordered according to this model but to A = C < B. Tied ranks in $\{r_i\}$ are dithered. Unordered ANOSIM indicates a clear difference among groups (slope $R^O = R = 0.80$, $p = 3\%$), while ordered ANOSIM does not ($R^O = 0.33$, $p = 11\%$).

there are no tied dissimilarities). Generally, this will mean that the numerical value of ρ will be less than that of R^O but, as with most generalisations, there will be exceptions. For ρ to equal or exceed R^O , the observational ranks $\{r_i\}$ need to be more heavily tied than are the modelled distances $\{s_i\}$. This could arise if, for example observed dissimilarities took few values (say 0 or 100) and were being correlated with a model having several levels, a rather unlikely case! One should also note the asymmetry of the R^O statistic in contrast to the symmetry of ρ . The generalised ANOSIM concept is restricted to regressing real data in the ranks $\{r_i\}$ on modelled distances in the ranks $\{s_i\}$, and it does not make sense to carry out the regression the other way around. The RELATE ρ statistic, on the other hand, is appropriate for a wider sweep of problems where the interest is in comparing the sample patterns of any two resemblance matrices. This contrast is, in part, an issue of what to do about tied ranks and identifies a context-dependent dichotomy noted early in the development of non-parametric methods (Kendall 1970). Namely, are two judges in perfect agreement only if they rank 10 candidates in exactly the same order, or does placing the candidates into the same two groups of 5 ‘acceptable’ and 5 ‘not acceptable’ count as perfect agreement? Here, ρ (the former, which does not adjust for tied ranks) will be more appropriate for some problems, and generalised R (the latter, which does, in effect, build in an adjustment for ties in the $\{s_i\}$) will be more appropriate for other problems.

Somerfield *et al.* (2002) showed that in situations where the groups of data are genuinely ordered, a test that takes this *a priori* ordering into account will

have greater power to detect differences among groups than testing against an unordered alternative. Here, the concept of ordered factors is extended into the ANOSIM framework. Not only does the use of an ordered statistic allow more information about the sample relationships to be built into its construction, it also increases the number of potential permutations under the null hypothesis. This influences aspects of the power and precision of the test in detecting ordered, as opposed to unordered, change. It is important to note, however, that in situations where differences among groups are not ordered (or ordered differently to the ordering implicit in the model being tested), then testing using the ordered statistic may be less powerful than the unordered alternative. Fig. 7 shows an example.

Finally, the focus of this paper is on developing and demonstrating a non-parametric framework for testing for differences among ordered groups. It concentrates on linear (or serial) ordering of groups in a simple one-way layout, such as might be expected through time, along an environmental gradient or in response to increased doses of an experimental treatment, as these are common situations in many ecological investigations. It should be noted, however, that the ideas and results described here carry over seamlessly to situations where the model reflects other relationships among groups of samples. A simple example could be a test for seasonality where the model matrix would need to reflect the fact that samples are not simply linearly arranged in time, but are cyclical with groups of samples close to the end of one year being close to samples at the beginning of the next, with samples at their greatest distance apart when separated by half a year. If all

dissimilarities between groups are larger than any within groups, and all dissimilarities between groups which are further apart in the (cyclical) model matrix are larger than any dissimilarities between groups which the model puts closer together, then the generalised *R* statistic will take a value of 1, just as in the simpler linear examples described in this paper.

ACKNOWLEDGEMENT

All those who contributed to collecting the data used in this study, or making it available, are gratefully acknowledged, including J.S. Gray, G. Hobbs, J.M. Gee, R.M. Warwick, A.A. Rowden and S. Widdicombe.

FUNDING

P.J.S. acknowledges funding support from the UK Natural Environment Research Council (NERC) through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science (Grant no. NE/R015953/1) and from the NERC and Department for Environment, Food and Rural Affairs, Marine Ecosystems Research Programme (Grant no. NE/L00299X/1).

CONFLICT OF INTEREST

The authors are unaware of any conflict of interest.

AUTHOR CONTRIBUTION

Paul J. Somerfield: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **K. R. Clarke:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **R. N. Gorley:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Validation (equal).

DATA AVAILABILITY STATEMENT

All data are freely available from the corresponding author.

REFERENCES

Clarke K. R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–43.

- Clarke K. R. & Gorley R. N. (2015) *PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth, UK.
- Clarke K. R., Gorley R. N., Somerfield P. J. & Warwick R. M. (2014a) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 3rd edn. PRIMER-E, Plymouth, UK.
- Clarke K. R. & Green R. H. (1988) Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Progr. Ser.* **46**, 213–26.
- 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., 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.
- Clarke K. R., Somerfield P. J. & Gorley R. N. (2016) Clustering in non-parametric multivariate analyses. *J. Exp. Mar. Biol. Ecol.* **483**, 147–55.
- Clarke K. R., Tweedley J. R. & Valesini F. (2014b) Simple shade plots aid better long-term choices of data pre-treatment in multivariate assemblage studies. *J. Mar. Biol. Assoc. U.K.* **94**, 1–16.
- 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.
- Gee J. M. & Warwick R. M. (1994a) Metazoan community structure in relation to the fractal dimensions of marine macroalgae. *Mar. Ecol. Progr. Ser.* **103**, 141–50.
- Gee J. M. & Warwick R. M. (1994b) Body-size distribution in a marine metazoan community and the fractal dimensions of macroalgae. *J. Exp. Mar. Biol. Ecol.* **178**, 247–59.
- Gray J. S., Clarke K. R., Warwick R. M. & Hobbs G. (1990) Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar. Ecol. Progr. Ser.* **66**, 285–99.
- Hobbs G. (1987) *Environmental survey of the benthic sediments of the Ekofisk and Eldfisk oilfields, July 1987*. Oil Pollution Research Unit Report No. FSC/OPRU/24/87. Oil Pollution Research Unit, Field Studies Council, Orierton Field Centre, Pembroke.
- Hope A. C. A. (1968) A simplified Monte Carlo significance test procedure. *J. R. Stat. Soc. Series B* **30**, 582–98.
- Kendall M. G. (1970) *Rank Correlation Methods*. Griffin, London.
- Legendre P. & Legendre L. (2012) *Developments in Environmental Modelling, 24. Numerical Ecology*, 3rd English edition. Elsevier, Amsterdam.
- Mantel N. (1967) The detection of disease clustering and a generalised regression approach. *Cancer Res.* **27**, 209–20.
- Somerfield P. J., Clarke K. R. & Olsford F. (2002) A comparison of the power of categorical and correlational tests applied to community ecology data from gradient studies. *J. Anim. Ecol.* **71**, 581–93.
- Somerfield P. J. & Gage J. D. (2000) Community structure of the benthos in Scottish sea-lochs. IV Multivariate spatial pattern. *Mar. Biol.* **136**, 1133–45.
- Somerfield P. J., Gee J. M. & Warwick R. M. (1994) Soft-sediment meiofaunal community structure in relation to a long-term heavy metal gradient in the Fal estuary system. *Mar. Ecol. Progr. Ser.* **105**, 79–88.