

# A method of linking multivariate community structure to environmental variables

K. R. Clarke, M. Ainsworth

Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, United Kingdom

**ABSTRACT:** The method of choice for multivariate representation of community structure is often non-metric multi-dimensional scaling (MDS). This has great flexibility in accommodating biologically relevant (i.e. non correlation-based) definitions of similarity in species composition of 2 samples, and in preserving the rank-order relations amongst those similarities in the placing of samples in an ordination. Correlation-based techniques (such as Canonical Correlation) are then inappropriate in linking the observed biotic structure to measured environmental variables; a more natural approach is simply to compare separate sample ordinations from biotic and abiotic variables and choose that subset of environmental variables which provides a good match between the 2 configurations. In fact, the fundamental constructs here are not the ordination plots but the (rank) similarity matrices which underlie them: a suitable measure of agreement between 2 such matrices is therefore proposed and used to define an optimal subset of environmental variables which 'best explains' the biotic structure. This simple technique is illustrated with 3 data sets, from studies of macrobenthic, meiobenthic and diatom communities in estuarine and coastal waters.

## INTRODUCTION

An increasingly accepted paradigm for the analysis of marine community data is that outlined by Field et al. (1982). An array of species abundance or biomass data for a number of samples taken at different sites and/or times is (a) transformed, to achieve a desired balance of the contributions of common and rarer species; (b) converted to a triangular matrix of similarity between every pair of samples (using a similarity or dissimilarity coefficient which is not a function of 'joint absences', such as that of Bray & Curtis 1957); (c) the similarities then permit a low-dimensional display of biotic relationships among the samples by non-metric Multi-Dimensional Scaling (MDS, Kruskal & Wish 1978); (d) the patterns in the biotic analysis are interpreted in terms of the environmental data, e.g. by superimposing the values for each physical or chemical variable in turn on the MDS ordination.

The merits of MDS techniques for ordination are now well-established in ecological applications, and comparative studies have consistently demonstrated their reliability (e.g. Kenkel & Orloci 1986). The wider Field et al. (1982) strategy has been much quoted (about 80 'non-self' citations are listed in the Science Citation

Index), and Clarke (1993) reviews some practical experiences with this approach in the intervening decade; inevitably, some of the procedures have proved of less utility than others and certain classes of problem were not comprehensively addressed in the original work. The current paper concentrates on one of these – the linking of environmental data to biotic patterns.

The asymmetric approach of the above paradigm, allowing the biota to 'tell their own story' before deducing links to specific environmental variables, has been adopted in a number of practical studies (e.g. Gray et al. 1988, 1990, Warwick et al. 1990, Warwick & Clarke 1993). The graphical techniques for elucidating these links (e.g. as seen in Fig. 9 of Field et al. 1982) have tended to be rather inadequate however, in that only one environmental variable at a time is related to the biotic ordination. Thus, in a manner consistent with the above strategy, it has not been readily possible to answer questions such as: (i) How well is the community structure 'explained' by the full set of environmental variables measured? (ii) Which variables are redundant in the sense of failing to strengthen the 'explanation' of biotic patterns once certain other variables are taken into account?

The intention of this paper is to propose a simple method for addressing such questions and to examine its performance for several literature data sets.

## METHOD

### Concept

Our approach relies on the premise that pairs of samples which are rather similar in terms of a suite of physico-chemical variables would be expected to have rather similar species composition, provided the relevant variables determining community structure (and *only* these variables) have been included in the analysis. Thus an ordination of these abiotic variables, representing the mutual environmental similarities among samples, should closely resemble the ordination of samples based on the biota. Selecting different combinations of the full environmental variable set should allow us to determine an 'optimal' match of the separate biotic and abiotic ordinations. The omission of a key determinant will degrade the match, as will the *inclusion* of environmental variables that differ markedly between the samples but have no effect on community composition.

Two related advantages are conferred by separating the construction of environmental and biotic ordinations, only linking the two by a 'pattern matching' stage.

(1) *Separate definitions of among-sample similarity can be used for each.* As noted earlier, species data requires a rather careful formulation of similarity, with appropriate trade-offs between the contributions of common and rarer species, and independence of the chosen coefficient from joint absences (as Field et al. 1982 put it: 'estuarine and abyssal samples are [no more] similar because both lack outer shelf species'). A typical species abundance matrix is dominated by zero counts and this makes correlation (or Euclidean distance) a poor basis for assessing similarity. By contrast, physico-chemical data matrices can usually be handled by conventional multivariate statistics. The variables can often be transformed to approximate normality – sometimes with different transformations needed for separate classes of variables – and correlation-based definitions of similarity are then to be preferred.

(2) *Few constraints are implied on the nature of the link between biota and environment.* It is clear that the original premise of this section could be met under a typical mix of conditions in which the numbers (or biomass) of some species are linearly related to an environmental gradient, other species are non-linearly but still monotonically related to the gradient, and the

remainder are strongly non-monotonic (unimodal) over different ranges of one or several abiotic variables. In contrast, 'classical' statistical methods such as Canonical Correlation (e.g. Mardia et al. 1979) would compute correlations, and thus assume linear relationships, *among* and *between* the biotic and abiotic variables. This is quite unrealistic.

How then should the 'pattern matching' be undertaken? One possibility is to use a Procrustes analysis (Gower 1971) on the separate ordinations for biota and environmental variables. This is a technique for optimal matching of 2 different configurations of points with the same set of labels, using only rotation, reversal and shrinking of one plot in relation to the other, i.e. preserving the *relative* position of the points within each plot. Their goodness-of-fit is then assessed by some criterion involving 'squared distance apart' of the 2 configurations. However, this has the disadvantage that it operates on ordinations in a specified number of dimensions. Thus the particular combination of environmental variables giving an optimal match of biotic and abiotic patterns may vary, depending on whether we view the sample relationships in 2- or 3-dimensional ordinations. This is undesirable. The complex data typically encountered scarcely ever admits of a *perfect* representation (zero stress) in a specified number of dimensions, so choice of dimensionality for convenient viewing of the patterns is inevitably somewhat arbitrary. A more objective criterion would be based on the triangular similarity (or dissimilarity) matrices underlying both ordinations. The definition of similarity of 2 samples will have been carefully selected, in an appropriately different way for biotic and abiotic variables, and it can therefore be argued that the similarity matrices are the fundamental constructs, representing all that is known about sample relationships.

### Coefficient

How then should one define a coefficient of agreement between the biotic and abiotic (dis)similarity matrices? An obvious candidate is a *rank* correlation computed between all the elements of the 2 triangular matrices. A rank coefficient is an appealing choice because of the differing units and modes of construction of the biotic and abiotic matrices (e.g. Bray-Curtis dissimilarities and Euclidean distances respectively); their common denominator is the relative ordering of dissimilarities within each matrix. The argument for using only these ranks is further strengthened by the knowledge that the displays of sample relationships, through accurate (low stress) MDS ordinations, are effectively functions only of the *rank* (dis)similarities.

We have studied the performance in this context of standard rank correlation coefficients, such as Spearman's  $\rho_s$  or Kendall's  $\tau$  (Kendall 1970), for several literature data sets, including the illustrations of the following section. Spearman's coefficient is defined as

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

where  $\{r_i; i = 1, \dots, N\}$  = ranks of all the sample similarities calculated using the biotic data and  $\{s_i; i = 1, \dots, N\}$  = ranks of sample similarities defined from abiotic data. This correlation is computed by first 'un-peeling' the 2 triangular similarity (or dissimilarity) matrices in the same way, to give 2 vectors of length  $N = n(n - 1)/2$ , where  $n$  = number of samples. Each vector is then replaced by its ranks (the integers 1 to  $N$ , in some order) and Eq. 1 applied.

Some points of clarification are needed here. Firstly, similarities and dissimilarities are always defined to be exactly complementary (whatever the coefficient used), in the sense that the rank order of similarities is the inverse of the rank order of dissimilarities. Secondly, high similarities correspond to low *rank similarities*, following the usual convention that the highest similarity is given a rank of 1, the next highest a rank of 2, and so on. Thirdly, note that the rank similarities  $\{r_i\}$  (or  $\{s_i\}$ ) are not a set of mutually independent variables, since they are based on a large number ( $N$ ) of strongly inter-dependent similarity calculations. Standard statistical tests and confidence intervals for  $\rho_s$  or  $\tau$ , derived for the case where the  $\{r_i\}$  and  $\{s_i\}$  are ranks of independent and identically distributed observations, are therefore totally invalid.

In itself, the above does not compromise the use of  $\rho_s$ , say, as an index of agreement of the 2 triangular matrices. One might expect, however, that it would be a less than ideal measure because of the fact that few of the equally-weighted difference terms in Eq. 1 involve 'nearby' samples. In contrast, the premise at the beginning of this section makes it clear that we are seeking a combination of environmental variables which attains a good match of the *high* similarities (low rank similarities) in the biotic and abiotic matrices. This is more a mechanistic point than an interpretational one; it is simply that the value of  $\rho_s$ , when computed *from triangular similarity matrices*, will tend to be swamped by the larger number of terms involving 'distant' pairs of samples. These have high rank values and the squaring of differences in the numerator of Eq. 1 makes the coefficient sensitive to small *relative* differences in these large ranks. For example, 15 community samples, strongly clustered into 5 groups of size 3, generate only 15 within-group

similarities but 90 between-group terms; of importance here would seem to be a search for abiotic variables which discriminate the same 5 clusters – the precise disposition of the groups in relation to each other then seems of less significance, though this feature would tend to dominate the unweighted coefficient. The point is seen again in the first example of the next section.

Down-weighting of the larger ranks in Eq. 1 can be achieved by adding a denominator term, inside the summation, which is an increasing (and symmetric) function of  $r_i$  and  $s_i$ . Amongst a number of possibilities (inevitably), the algebraically simplest choice of weighting term is  $(r_i + s_i)$ , giving coefficient

$$\rho_w = 1 - c \sum_{i=1}^N \frac{(r_i - s_i)^2}{(r_i + s_i)} \quad (2)$$

where  $c$  is chosen so that  $\rho_w$  (like  $\rho_s$ ) takes values in the range  $-1$  to  $+1$ . More (or less) severe weighting could be achieved by higher (or lower) powers of  $(r_i + s_i)$  in the denominator.

Algebraic manipulation shows that

$$c = \frac{6}{N(N - 1)} \quad (3)$$

with extreme values  $\rho = -1$  and  $+1$  corresponding respectively to the cases where the 2 sets of ranks are in complete opposition or complete agreement. The former case is unlikely to be attainable in practice because of the constraints inherent in a similarity matrix;  $\rho$  will take values around zero if the match between the 2 patterns is effectively random but  $\rho$  will typically be positive.

The algebra also yields the alternative, and entirely equivalent, formulation

$$\rho_w = \frac{1}{N - 1} (12\bar{H} - 5N - 7) \quad (4)$$

where

$$\bar{H} = \frac{1}{N} \sum_{i=1}^N \frac{2}{r_i^{-1} + s_i^{-1}} \quad (5)$$

is the average over all  $i = 1, \dots, N$  matrix elements of the harmonic mean of the ranks  $r_i$  and  $s_i$ . This suggests the use of the term *harmonic rank correlation* for the weighted Spearman coefficient  $\rho_w$ , and this alternative formulation is (arguably) helpful in visualising the behaviour of  $\rho_w$ , for example the fact that it is minimised when the ranks  $\{r_i\}$  and  $\{s_i\}$  are exact reversals of each other. Ultimately, the efficacy of this (or a more severe) down-weighting of the larger rank similarities can only be judged by its performance in real and simulated applications; this issue is returned to in the later examples.

### Procedure

The above elements can now be put together to define the procedural steps, sketched in Fig. 1.

(i) Biotic and abiotic data matrices are handled separately, initially transforming each according to the needs of the differing similarity measures (discussed further in 'Results').

(ii) The among-sample similarity matrix for the biota is constructed only once but the equivalent triangular matrix for the abiotic data is computed many times, in fact for all possible combinations of environmental variables at each 'level of complexity' of explanation (variables taken singly, 2 at a time, 3 at a time, etc.).

(iii) The rank correlation (e.g.  $\rho_w$ ) between the biotic and abiotic triangular matrices is calculated in every case. The highest few coefficients at each level of complexity are tabulated, allowing the extent of improvement or deterioration in the match to be traced, as further variables are added. The examples given later suggest that  $\rho_w$  will reach a maximum for some (proper) subset of the  $v$  abiotic variables, and the maximum  $\rho$  at each level will usually increase monotonically as the number of variables increases, up to the optimum combination, and then decrease monotonically (though with shallower slope) as more variables are added. This makes the optimum combination a very natural set of 'best explanatory variables' of the biotic pattern.

(iv) The final step is to display the biotic MDS in conjunction with ordinations of the most important environmental variable combinations; a sensible policy is often to plot the best (maximum  $\rho$ )  $k$ -variable configuration, for  $k = 2, 3, 4, \dots, v$ . This achieves 2 things: it places the absolute values of  $\rho$  in context (for example,  $\rho_w$  of 0.8 or more seems to correspond to a very close match), and it demonstrates the importance or otherwise of small improvements in  $\rho$  that may arise from increasing the complexity of the abiotic description (for example, it appears not to be appropriate to quote  $\rho_w$  to more than 2 decimal places).

It should not be forgotten that this final display step is not definitive in the selection of the 'best match'. Indeed, if a 2-dimensional MDS is a rather poor representation of the sample relationships in the biotic data (*stress* > 0.15, say) then there is no guarantee that the optimal  $\rho$  combination will 'look best' in visual comparisons of 2-dimensional ordination plots. The examples of the following section show, however, that where stress is low there is a high degree of concordance between conclusions suggested by ordinations and by formal comparison of  $\rho$  values: even modest increases in  $\rho_w$  are reflected in improvements in the placing of one or more points in the abiotic plot relative to the biotic MDS. Note that such comparisons are aided if compatible ordination techniques are used for the separate biota and environmental analyses. Thus, whilst Principal Component Analysis (PCA) is the natural means of displaying the (suitably transformed) environmental data, corresponding to dissimilarities based on Euclidean distance, non-metric MDS of this Euclidean distance matrix will give a more strictly comparable ordination to the biotic MDS (the latter from Bray-Curtis dissimilarities). The comparative plots are then based largely on the data (*rank* similarities) that are used to define the  $\rho$  coefficient. This approach has therefore been adopted for the environmental data of the later illustrations though, in these cases, there are only minor differences between MDS on Euclidean distance and the corresponding PCA; this is to be expected when the higher-dimensional structure is well-represented in 2 dimensions.

## RESULTS AND DISCUSSION

The methodology is illustrated by 3 examples of data analysis, covering different groups of marine biota (macrofauna, meiofauna and diatom communities), a spread of locations (in estuarine and coastal waters) and contrasting types of abiotic data (structuring of communities by pollution impact and by natural environmental variables).

### Exe estuary nematodes

An instructive starting point is the data set used as an example in Field et al. (1982). The nematode component of benthic meiofaunal communities at several inter-tidal locations in the Exe estuary, UK, was enumerated by Warwick (1971). Fourth-root transformed abundances of 182 nematode species in 19 samples (each a total over 6 seasonal 'replicates') were subjected by Field et al. (1982) to non-metric MDS, employing Bray-Curtis similarities. The resulting ordi-

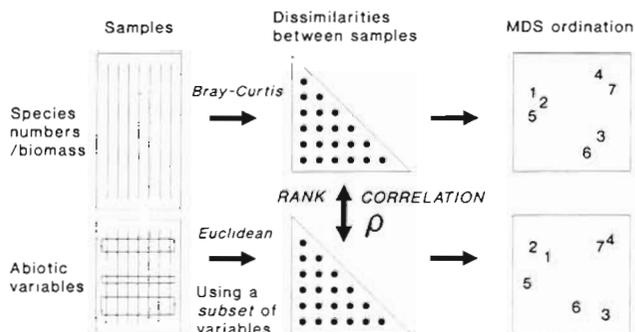
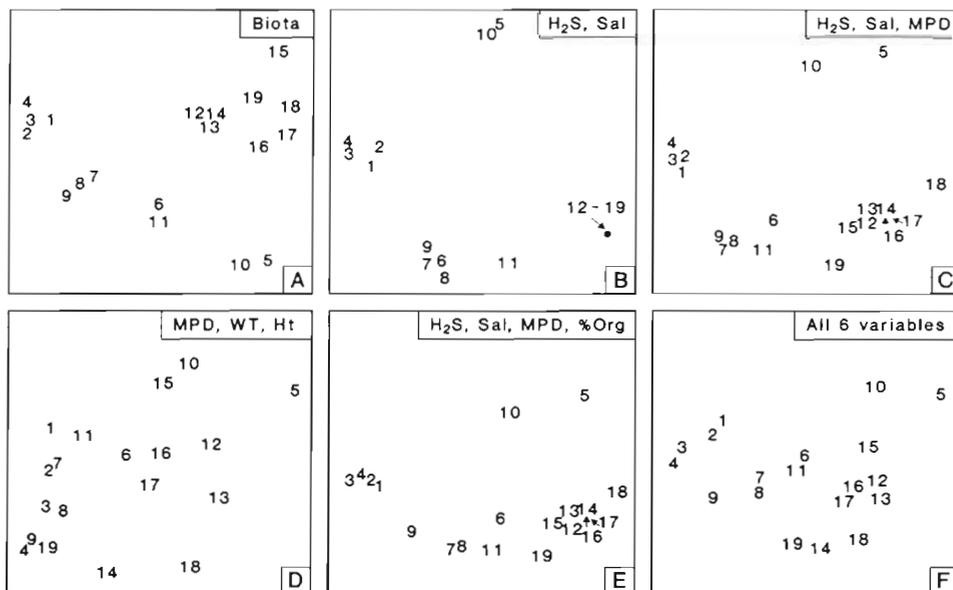


Fig. 1 Schematic diagram of the linking of biotic pattern to environmental variables: selection of abiotic variable subset maximising rank correlation ( $\rho$ ) between (dis)similarity matrices

Fig. 2. Exe estuary nematodes. Non-metric multi-dimensional scaling (MDS) ordinations of the 19 sampling sites, based on: (A) transformed abundances of 182 nematode species; (B) to (F) selected subsets of the 6 environmental variables (see text). The associated 'matching coefficients'  $\rho_w$  of environmental to biotic similarity patterns (Table 1A) are: (B) 0.76, (C) 0.80, (D) 0.40, (E) 0.79 and (F) 0.77. Stress values for the 2-dimensional ordinations, using Kruskal's stress formula 1, are: (A) 0.05; (B) 0; (C) 0.04; (D) 0.07; (E) 0.04; and (F) 0.06



nation is shown in Fig. 2A, a 2-dimensional plot providing an excellent representation (stress is very low at 0.05).

A suite of 6 environmental variables was recorded for each of the 19 sites (again averaged over the 6 sampling occasions, where appropriate): median particle diameter of the sediment (MPD), its organic content (% Org), depth of the blackened H<sub>2</sub>S layer (H<sub>2</sub>S), interstitial salinity (Sal), depth of the water table (WT) and height up the shore (Ht), the latter being appropriately scaled to tidal positions.

Field et al. (1982) superimposed these variables one at a time on the faunal MDS, as symbols whose sizes reflect the magnitude of the environmental variable at each site. Their Fig. 9 shows consistent patterns in some variables but, to return to the questions posed in the 'Introduction', are there changes in these variables *in combination* which could be sufficient to account for division of the 19 sites into the 5 groups observed in Fig. 2A? (These 5 groups are also clearly delineated in the corresponding cluster analysis, Fig. 3 of Field et al. 1982). A corollary would ask if any of the abiotic variables were irrelevant to the biotic description, or positively prejudicial to a good explanation because they vary at odds with the faunal pattern. Answering these questions by the procedure in the 'Method' section leads to the results of Table 1 and the plots of Fig. 2B to F. The latter are 2-dimensional MDS ordinations of several abiotic variable subsets, based on Euclidean distance matrices (or they can be regarded as PCA plots of environmental variables, since the difference is small in this case).

Table 1A shows the combinations of variables which give rise to the largest rank correlations ( $\rho_w$ ) between the biotic and abiotic sample similarities, as the size of

Table 1. Exe estuary nematodes. Combinations of environmental variables, taken *k* at a time, yielding the 'best matches' of biotic and abiotic similarity matrices for each *k*, as measured by (A) harmonic rank correlation  $\rho_w$  (weighted Spearman), and (B) standard Spearman coefficient  $\rho_s$ . Bold type indicates the combination with maximum  $\rho$  overall. Sediment variables are H<sub>2</sub>S: depth of H<sub>2</sub>S layer; Sal: interstitial salinity; MPD: median particle diameter; %Org: % organic content; Ht: height up the shore; WT: depth of water table

A: Harmonic rank correlation	
<i>k</i>	Best variable combinations ( $\rho_w$ )
1	H <sub>2</sub> S %Org Sal ... (.62) (.54) (.53)
2	H <sub>2</sub> S, Sal H <sub>2</sub> S, MPD H <sub>2</sub> S, %Org Sal, %Org ... (.76) (.67) (.65) (.61)
3	<b>H<sub>2</sub>S, Sal, MPD</b> H <sub>2</sub> S, Sal, %Org H <sub>2</sub> S, Sal, WT ... (.80) (.75) (.73)
4	H <sub>2</sub> S, Sal, MPD, %Org H <sub>2</sub> S, Sal, MPD, Ht ... (.79) (.78)
5	H <sub>2</sub> S, Sal, MPD, %Org, Ht ... (.79)
6	H <sub>2</sub> S, Sal, MPD, %Org, Ht, WT (.77)
B: Spearman rank correlation	
<i>k</i>	Best variable combinations ( $\rho_s$ )
1	H <sub>2</sub> S %Org Sal ... (.71) (.64) (.58)
2	H <sub>2</sub> S, Sal H <sub>2</sub> S, %Org H <sub>2</sub> S, MPD Sal, %Org ... (.80) (.76) (.73) (.68)
3	H <sub>2</sub> S, Sal, MPD H <sub>2</sub> S, Sal, %Org H <sub>2</sub> S, %Org, Ht ... (.82) (.80) (.77)
4	<b>H<sub>2</sub>S, Sal, MPD, %Org</b> H <sub>2</sub> S, MPD, %Org, Ht ... (.83) (.80)
5	H <sub>2</sub> S, Sal, MPD, %Org, Ht ... (.82)
6	H <sub>2</sub> S, Sal, MPD, %Org, Ht, WT (.79)

the subset of environmental variables increases. Thus, *on its own*, the variable which best groups the sites, in a manner consistent with the faunal pattern, is the depth of the H<sub>2</sub>S layer ( $\rho_w = 0.62$ ); next best is the organic content of the sediment ( $\rho_w = 0.54$ ), etc. Of course, since the faunal ordination is not essentially 1-dimensional (Fig. 2A), one would not expect a single environmental variable to provide a very successful match, though knowledge of the H<sub>2</sub>S variable alone does distinguish points to the left and right of Fig. 2A (Sites 1 to 4 and 6 to 9 have lower values than for Sites 5, 10 and 12 to 19, with Site 11 intermediate).

The best 2-variable combination also involves depth of the H<sub>2</sub>S layer but adds the interstitial salinity. The correlation ( $\rho_w = 0.76$ ) is markedly better than for any other 2-variable subset. Fig. 2B displays the corresponding ordination; there are only 2 variables involved so, in effect, the figure is a simple scatter plot of H<sub>2</sub>S against salinity, rotated to aid visual comparison with Fig. 2A. The interstitial salinity distinguishes Sites 1 to 4 from 6 to 9 and Sites 5 and 10 from 12 to 19.

The best 3-variable combination retains the above 2 variables and adds the median particle diameter of the sediment; the matching coefficient rises to  $\rho_w = 0.80$  and, in fact, this is the maximum value of  $\rho_w$  for any size of subset. The ordination of the 3 variables is shown in Fig. 2C; it can be seen that the inclusion of median particle diameter draws Sites 6 and 11 closer and adds some meaningful differentiation of the 12 to 19 group (e.g. note the proximate positions of Sites 12 to 14). In fact, Fig. 2C now provides a surprisingly good match to Fig. 2A, with all the main conclusions about site relationships being apparent from both plots. The only obvious discrepancy is in the location of the (5, 10) cluster; in Fig. 2C this has changed places with the large 12 to 19 group. This is of no real significance: the faunal data clearly indicate that there are 5 distinct clusters, with the (5, 10) group rather unlike anything else but closer to (6, 11) and 12 to 19 than the other groups; similar conclusions are apparent from the arrangement of sites in the environmental variable plot, Fig. 2C. The point to bear in mind is that the ordinations are only a visual aid. The matching is carried out on the relative similarities (i.e. which samples are 'close' to which other samples), and these are *not* constrained to a 2-dimensional layout. In fact, this example provides a good illustration of why a Procrustes approach to matching the faunal and environmental ordinations could be unsatisfactory. Fig. 2A and C could not be made to match closely by rotation, reflection or scaling operations; yet, as we have remarked, they tell a surprisingly similar story.

At this point, it might not be unreasonable to query whether a roughly similar division of the sites would arise from any combination of (say) 3 environmental

variables; perhaps the latter are so confounded with each other that they *all* tell essentially the same story (irrespective of the range of values observed for  $\rho_w$ !). This can be simply refuted by plotting some combinations of environmental variables giving rise to lower matching coefficients. Fig. 2D shows the ordination of the 3 variables: median particle diameter, depth of the water table and height up the shore (the corresponding  $\rho_w$  is 0.40). Whilst there are still some points of similarity, arising from the inclusion of median particle diameter, the match with Fig. 2A has largely broken down.

More subtly, note what happens when further variables are added to a description which already appears to provide a good match. Table 1A shows that the best 4-variable combination simply adds knowledge of the % organic content to the best 3-variable combination, but this contributes nothing further to the explanation. In fact,  $\rho_w$  declines slightly and this is reflected in the corresponding ordination Fig. 2E, with a parting of Site 9 from Sites 7 and 8 which is not observed in the faunal plot Fig. 2A. Adding the depth of the water table and height up the shore, the match is seen to deteriorate further, as indicated by the declining value of  $\rho$  and as observed in Fig. 2F: Site 14 is separated from Sites 12 and 13 and the spacing between Sites 1 to 4 increases; neither of these changes mirrors the pattern in Fig. 2A.

There is an interesting contrast here with an unrelated, but in some ways analogous statistical technique, that of multiple regression (e.g. Draper & Smith 1981). Layouts such as Table 1 have the appearance of lattice structures from linear models, where the value in brackets would represent the residual sum of squares from fitting a subset model. There, the addition of further explanatory variables guarantees an 'improved fit' of the data points to the model, in the sense of always reducing the residual sum of squares. By contrast, here, the inclusion of environmental variables which have no effect on community structure is not just neutral in its effect on the matching coefficient but is virtually guaranteed to *decrease*  $\rho$ . For example, Sites 1 to 4 cover the full range of possible sampling positions in the intertidal zone which, of course, explains why the corresponding points are 'pushed apart' in Fig. 2F. This is not matched in Fig. 2A and height up the shore is clearly seen not to be a forcing variable of the *overall* community pattern in this data set.

Note the satisfying consistency of the 'lattice' structure in Table 1A: the best combination at one level is always a subset of the best combination on the line below. The analysis has not been constrained to achieve this; *all* combinations have been evaluated and simply ranked. Furthermore, the introduction of some variables (e.g. depth of the water table and

height up the shore) virtually never seems to result in increasing  $\rho$ , wherever this occurs in the lattice. There is no guarantee of such a simple structure and straightforward interpretation in all cases, just as there is no guarantee of simplicity in interpreting multiple regression results, though it is interesting to note that the other 2 illustrations of this paper also show a simple hierarchical pattern for the best variable combinations at each level of complexity.

There remains one final point to make about this example. The analysis has been repeated using the simple Spearman rank correlation coefficient  $\rho_s$  instead of the weighted coefficient  $\rho_w$ ; the results are in Table 1B. (It has also been repeated using Kendall's  $\tau$  but with conclusions which are virtually identical to those using  $\rho_s$ .) The absolute values of  $\rho_s$  are seen to be generally higher than those of  $\rho_w$ , and this is true of other examples also. In itself, this fact is of no significance (just as the values of Kendall's  $\tau$  tend on average to be lower than Spearman's  $\rho_s$ ; Kendall 1970, section 1.21). What matters is whether the *rankings* of  $\rho$  values for the different variable combinations are much the same for  $\rho_s$  and  $\rho_w$ ; if so, the weighting has achieved nothing. In fact, the pattern of best variable combinations at each level is rather similar in Table 1A and B, but there *is* a difference in the ranking of  $\rho$  values, with  $\rho_s$  suggesting (albeit marginally) that a 4-variable set provides a better match than the best 3-variable set. There seems little evidence to support such a conclusion from the ordinations (contrast Fig. 2C and E with A) and, indeed, some suggestion to the contrary in the separation of Site 9 from Sites 7 and 8, noted previously. The weighting is clearly having some effect and, for the present examples at least, one that appears to improve concordance of the formal results with simple visual comparisons of the respective (low stress) ordinations. This bears out the reasoning in the previous section that  $\rho_s$  may be over-dominated by the larger dissimilarities, and the use of some form of weighting, as in  $\rho_w$ , would seem advisable – at least for sample sizes of similar order to the current examples ( $n \approx 20$ ).

### Messolongi lagoon diatoms

Danielidis (1991) describes a study of diatom communities in the Messolongi, Aitoliko and Kleissova lagoons in western central Greece. Abundance data for 193 diatom taxa were collected from 17 sampling sites (Fig. 3) at 2 to 3 month intervals; the analysis here uses aggregated data over the 4 sampling times: June, August, October 1984 and January 1985. A suite of environmental variables for the 17 sites was also collected synchronously; these too have been averaged over the same sampling occasions. They consist

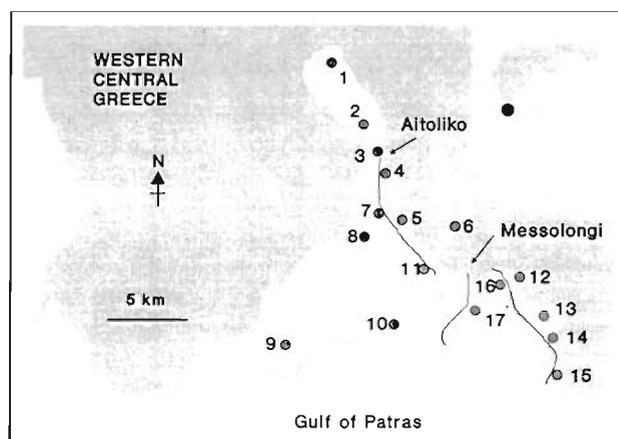


Fig. 3. The 17 sampling sites, for diatom communities and associated environmental variables, in the Messolongi, Aitoliko and Kleissova lagoons, western central Greece

of levels of inorganic nitrogen, nitrate, nitrite, ammonia, silicate, phosphate, salinity, temperature, pH and dissolved oxygen.

Table 2 presents the results of linking biotic and environmental patterns, based on the harmonic rank correlation  $\rho_w$  between respective similarity matrices. For each row of the table (complexity of environmental description) the best variable combinations are listed in rank order, including (at least) all  $\rho$  values greater than that for the best variable combination in the row above. A simple hierarchical structure is again apparent for the optimal combination at each level, with the overall maximum  $\rho$  occurring for the 5-variable subset: inorganic nitrogen, phosphate, salinity, silicate and dissolved oxygen.

Fig. 4 shows the sample configurations from MDS analyses based on the biota (4th-root transformed abundances and Bray-Curtis dissimilarities) and on the best 2-, 5- and 10-variable combinations of environmental data (Euclidean dissimilarities). The concordance is striking, with the main outline of the community pattern (Fig. 4A) already apparent from information on only inorganic nitrogen and phosphate (Fig. 4B), and the optimum environmental combination (Fig. 4C) providing a near-perfect match. Additional variables contribute nothing further; including them all (Fig. 4D) is seen to degrade the match by, for example, poorer placing of Site 8 (separated from Site 16) and Site 4 (now 'above' Site 15, separated from Sites 1 and 3).

One could argue here that some degree of matching should be anticipated, in that the environmental variables are not entirely abiotic; for example, measured silicate levels presumably reflect, in part, silica content of the diatom frustules. Nonetheless, the main conclusion cannot be gainsaid: the overall community struc-

Table 2. Messolongi lagoon diatoms. Combinations of variables, *k* at time, giving the largest rank correlations  $\rho_w$  between biotic and environmental similarity matrices; bold type indicates the best combination overall. In-N: inorganic nitrogen; Sal: salinity; O<sub>2</sub>: dissolved oxygen; Temp: temperature; other abbreviations are standard. All variables except Sal and O<sub>2</sub> are log-transformed

<i>k</i>	Best variable combinations ( $\rho_w$ )					
1	In-N (.75)	NO <sub>3</sub> (.64)	Sal (.61)	SiO <sub>2</sub> (.38)	...	
2	In-N, PO <sub>4</sub> (.79)	In-N, NO <sub>3</sub> (.74)	PO <sub>4</sub> , Sal (.74)	PO <sub>4</sub> , NO <sub>3</sub> (.74)	...	
3	In-N, PO <sub>4</sub> , Sal (.80)	In-N, PO <sub>4</sub> , NO <sub>3</sub> (.80)	In-N, PO <sub>4</sub> , SiO <sub>2</sub> (.79)	In-N, Sal, SiO <sub>2</sub> (.78)	...	
4	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> (.83)	PO <sub>4</sub> , Sal, SiO <sub>2</sub> , NO <sub>3</sub> (.81)	In-N, PO <sub>4</sub> , O <sub>2</sub> , NO <sub>3</sub> (.80)	In-N, PO <sub>4</sub> , Sal, O <sub>2</sub> (.79)	...	
5	<b>In-N, PO<sub>4</sub>, Sal, SiO<sub>2</sub>, O<sub>2</sub></b> <b>(.86)</b>	PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> (.83)	In-N, PO <sub>4</sub> , SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> (.83)	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , NO <sub>3</sub> (.82)	...	
6	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> (.86)	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , pH (.84)	...	...	...	
7	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> , pH (.85)	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> , NH <sub>3</sub> (.84)	...	...	...	
8	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> , pH, NH <sub>3</sub> (.84)	...	...	...	...	
9	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> , pH, NH <sub>3</sub> , NO <sub>2</sub> (.82)	...	...	...	...	
10	In-N, PO <sub>4</sub> , Sal, SiO <sub>2</sub> , O <sub>2</sub> , NO <sub>3</sub> , pH, NH <sub>3</sub> , NO <sub>2</sub> , Temp (.79)	...	...	...	...	

ture from a large abundance matrix, heavily transformed so that many species contribute to the MDS pattern, is replicated almost perfectly through knowledge of only 5 environmental variables. This result has obvious implications for prediction as well as explana-

tion. That the environmental and biotic variables be, in some sense, 'independently measured' is not a prerequisite of the method. We are not attempting at this stage to erect a formal statistical framework which would permit hypothesis testing, construction of confidence intervals for  $\rho_w$ , etc. Indeed, it is clear that  $\rho_w$  could be exploited as a simple measure of agreement between the similarity structure underlying any 2 ordinations with the same label sets, however these are obtained.

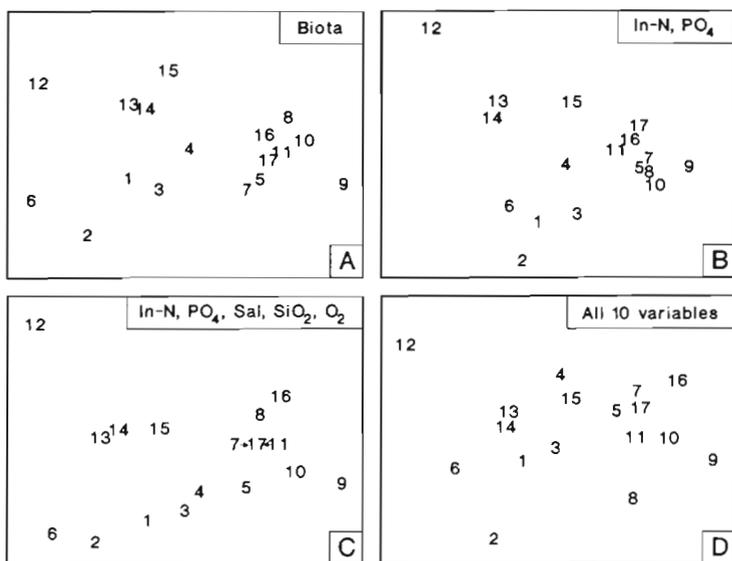


Fig. 4. Messolongi lagoon diatoms. MDS plots of the 17 sampling sites, based on: (A) transformed abundances of 193 diatom taxa; (B) to (D) the 'best' 2-, 5- and 10-variable combinations of transformed environmental variables (see text). Matching coefficients  $\rho_w$  (Table 2) are: (B) 0.79, (C) 0.86 and (D) 0.79. Stress values are: (A) 0.08; (B) 0; (C) 0.08; and (D) 0.08

Another general point of practical importance is also exemplified by the current data. Some of the environmental variables were in fact transformed prior to computation of the Euclidean dissimilarity matrix. The upper right-hand half of Fig. 5 shows why this was necessary. It is a 'draftsman plot', namely, an arrangement in a triangular array of simple scatter plots between pairs of environmental variables. The points in each plot are, of course, the values of the variables at the 17 sites, in this case untransformed. Such plots can be a helpful way of examining observations *en masse*, when assessing the likely validity of assumptions that will need to be made about the data. Here, choice of transformation is important in the definition of Euclidean dissimilarity between pairs of

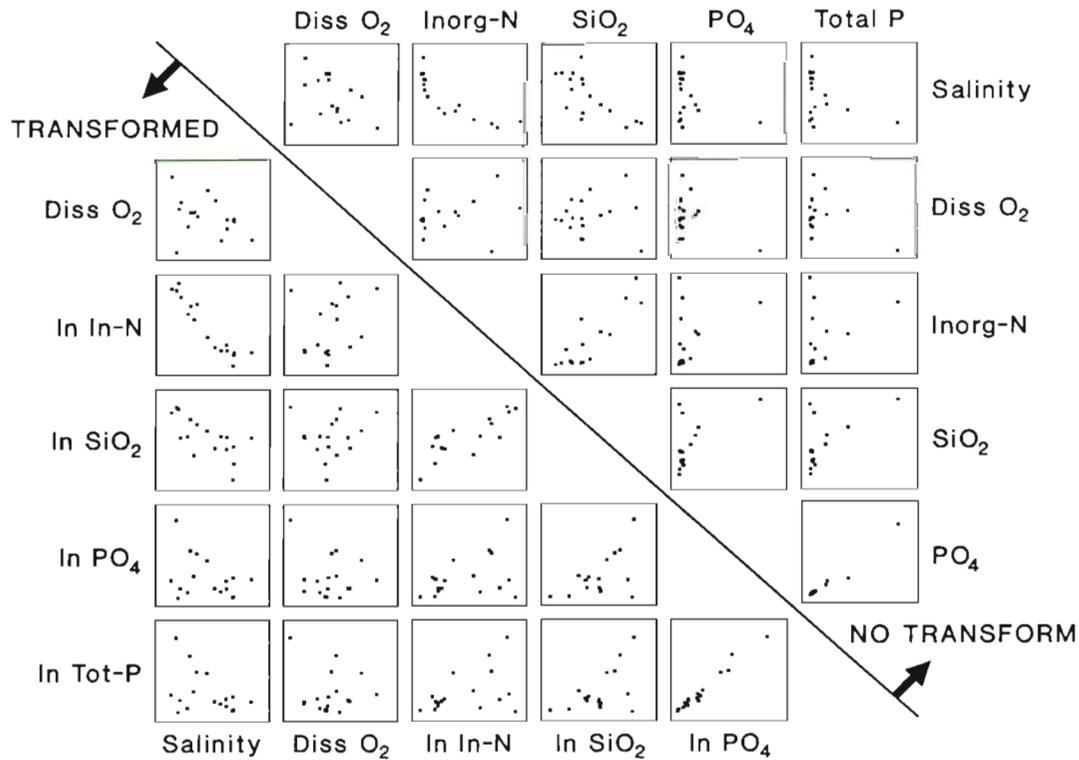


Fig. 5. Messolongi lagoons. 'Draftsman' plots (pairwise scatter plots) of selected environmental variables at the 17 sampling sites. Upper triangle: original values. Lower triangle: concentrations of inorganic nitrogen, silicate, phosphate and total phosphorus are log-transformed (note  $\ln = \log_e$ ); salinity and dissolved oxygen remain untransformed

samples, just as it is in defining Bray-Curtis dissimilarity for the species data. The use of Euclidean distance in the (normalised) environmental variable space is most effective if the data are approximately multivariate-normally distributed; that is, pairwise relationships (if present) are *linear* and the data are not markedly skewed on any of the variable axes.

The upper triangle in Fig. 5 indicates a marked right-skewness in most of the original variables, the exceptions being salinity and dissolved oxygen, and a noticeable curvilinearity of the relation between inorganic nitrogen and salinity. (Note that, for clarity of presentation, Fig. 5 includes only those variables deemed important by the matching process; however it is necessary initially to compute the full draftsman plot, for all variables). The lower triangle in Fig. 5 depicts the same draftsman plot after suitable transformation of most variables; this is seen to have largely removed the widespread skewness and also the curvilinearity in the inorganic nitrogen versus salinity plot. Note that it is not usually appropriate to tinker with minor modifications to transformations for individual variables. All that is necessary is a broad-brush approach in which related groups of variables, expected to behave in similar fashion (in terms of variance-mean relationships for example), are subjected to the *same* trans-

formation. For concentration variables this will usually involve reduction in right skewness, so the choice is often simply between log (or 4th-root, having similar effect), square root and no transformation at all. Here, all environmental variables were log transformed, with the exceptions of salinity and dissolved oxygen, and one can have confidence from the resulting draftsman plot that a dissimilarity matrix calculated from Euclidean distance (on subsequently normalised variables) is a fair summary of the sample relationships.

The draftsman plot also serves another useful 'pre-processing' function. Note that Fig. 5 includes a variable not found in the analysis of Table 2, total phosphorus; there were in fact 11 environmental variables originally recorded. Total phosphorus was omitted because of the very high degree of correlation it shows with phosphate in the (transformed) draftsman plot, Fig. 5. Whilst there would have been no technical difficulty in retaining both variables in the matching routine, there is little point in doing so. It doubles the number of combinations to search and for every combination involving phosphate, the parallel combination – in which phosphate is replaced by total phosphorus – would give a virtually identical  $\rho$  value. For example, the variable subset: inorganic nitrogen, total phosphorus, salinity, silicate and dissolved oxygen yields a

$\rho_w$  of 0.86 (cf. Table 2) and the MDS of these 5 variables is indistinguishable from Fig. 4C. Clearly, no useful information could result from retaining both variables; no amount of statistical sophistication can ever disentangle effects of variables which are almost perfectly confounded (Clarke & Green 1988)! The suggested rule-of-thumb here is that variable sets whose mutual correlation coefficients (after suitable transformation) average more than about 0.95 should be pruned to a single representative. There may occasionally be good reason for retaining one variable in preference to others in the subset – perhaps it is causally implicated, from external evidence – but whichever measurement is chosen the interpretation must recognise, of course, that this variable may stand only as a proxy for other variables. Furthermore, the same is true of *any* environmental measurement identified as contributing to the optimal match; it may stand only as a proxy for another, *unrecorded* variable.

It will be clear by now that the steps in the current example have been presented out of sequence. In fact it is desirable that a largely *prescriptive* approach is taken to the analysis, in the following order: a draftsman plot of all variables is used to select transformations and then to remove highly confounded variables, before any matching operations take place. Different transformations *are* likely to affect the conclusions – if sample similarities are represented in a different way there will almost inevitably be changes in the results of the matching process – but it seems unwise to search through several transformations in an attempt to maximise the degree of concordance between environmental and biotic patterns. The risk of 'selection bias' is then greatly magnified. Our overall strategy is predicated on the assumption that suitable transformations and similarity measures have been chosen to allow valid and meaningful biological representation of the samples by the respective similarity matrices; only then do the comparisons begin.

#### Firth of Clyde macrofauna

A final data set is used to exemplify the recommended prescriptive sequence and to illustrate its application to a pollution study in which the environmental measurements are contaminant levels. Fig. 6 shows a transect of 12 sampling sites across the sewage sludge disposal ground in the Firth of Clyde (Garroch Head), sampled in 1983 for abundance and biomass of soft-sediment macrobenthic species (Pearson 1987). Biomass readings of the 84 observed species are used here, for the reasons discussed in Warwick & Clarke (1993): when biomass information is available it can be argued to be a more biologically-

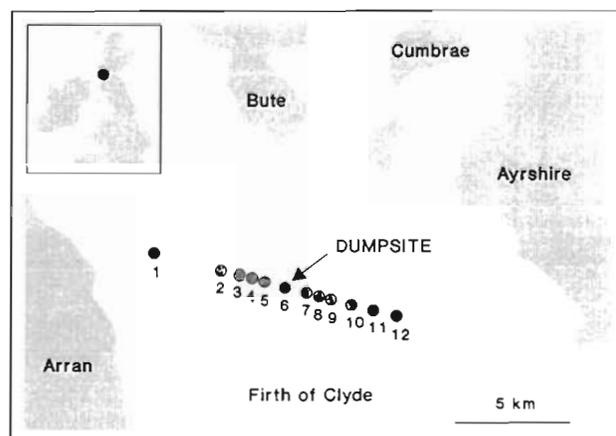


Fig. 6. The 12 sampling sites, for macrofaunal communities and sediment contaminant variables, on a transect across the sewage-sludge dumping ground in the Firth of Clyde, western Scotland, UK

relevant measure of how a community is structured. The environmental variables recorded from the 12 sites consist of the depth of the water column and sediment measures of organic enrichment (% carbon and % nitrogen) and heavy metal concentrations (Cu, Mn, Co, Ni, Zn, Cd, Pb and Cr; T. H. Pearson & J. Blackstock, Scottish Marine Biological Association unpubl. report, 1984).

On the basis of a draftsman plot of all 11 abiotic variables, a log-transformation was seen to be appropriate for the concentration data, that is, all variables except the water depth. The resulting draftsman plot for the transformed data is given in Fig. 7. There is clearly a high degree of *collinearity* in 1 subset of the variables; note the plots of Pb against Zn, %C against Zn, Pb against Cu, etc. As pointed out in the previous example, nothing is to be gained by retaining all 4 of these variables in the analysis; they exhibit the same pattern across the transect and it could never be possible to differentiate their effects. The suggested rule-of-thumb, that subsets with mutual correlation averaging more than about 0.95 should be replaced by a single representative, leads to the reduced set of 8 variables: %C, %N, Cd, Ni, Mn, Cr, Co and water depth.

The species biomass data is subjected to 4th-root transformation and a Bray-Curtis dissimilarity matrix computed. Table 3 then shows the results of matching this to the Euclidean dissimilarities from the transformed environmental variables. The results are again straightforward to interpret, with a simple hierarchical structure for the optimum combination at each level of complexity and an overall maximum  $\rho_w$  of 0.79 for just 3 environmental measurements: % carbon, % nitrogen and cadmium concentrations.

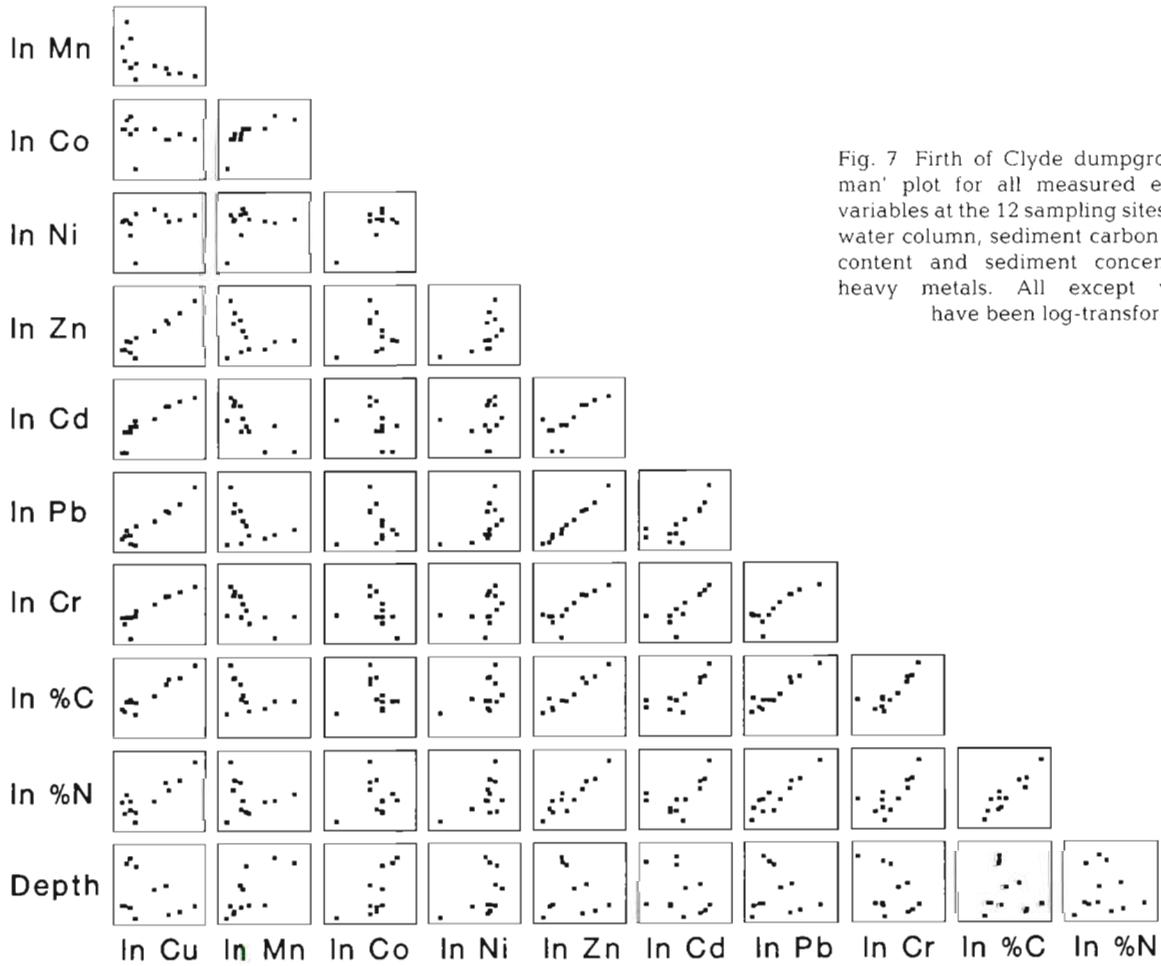


Fig. 7 Firth of Clyde dumpground. 'Draftsman' plot for all measured environmental variables at the 12 sampling sites: depth of the water column, sediment carbon and nitrogen content and sediment concentration of 8 heavy metals. All except water depth have been log-transformed

Some of the corresponding MDS plots are given in Fig. 8. The ordination for the biota (Fig. 8A) shows the clear trend of change in community structure passing from Site 1 to the sludge dumping area (centred at Site 6), with a steady reversal of the trend as one moves on to Site 12; the communities at the extremities of the transect are very similar. This major axis of change is largely replicated by the information on sediment organic content (Fig. 8B,  $\rho_w = 0.75$ ). Note, however, that there is additional structure in the biotic plot, for example the 'vertical' separation of Site 9, which is not well represented by the %C and %N information alone. Knowledge of Cd together with %C is seen to provide most of the missing structure (Fig. 8C,  $\rho_w = 0.78$ ), though the retention of %N produces a marginal improvement, for example moving Site 10 away from Site 2, and Site 8 away from Sites 7 and 5 (Fig. 8D,  $\rho_w = 0.79$ ). Of course, the 2 organic variables are closely related (though not entirely collinear), as is clear from the largely 1-dimensional structure of Fig. 8B. The substitution of other metal concentrations for Cd, or their addition to it, results in deterioration of the match.

Table 3. Firth of Clyde macrofauna. Combinations of variables,  $k$  at a time, giving the largest rank correlations  $\rho_w$  between biotic and abiotic similarity matrices; bold type indicates the best combination overall. Dep: depth of water column; other variables measure organic content (C, N) and heavy metal contamination (Cd, Ni, Mn, Cr, Co) of the sediments. All variables except water depth are log-transformed

$k$	Best variable combinations ( $\rho_w$ )			
1	C	N	Cd	...
	(.70)	(.62)	(.58)	
2	C, Cd	C, N	Cd, N	C, Co ...
	(.78)	(.75)	(.74)	(.63)
3	<b>C, Cd, N</b>	C, Cd, Co	C, N, Mn	...
	<b>(.79)</b>	(.72)	(.71)	
4	C, Cd, N, Ni	C, Cd, N, Co	...	
	(.77)	(.77)		
5	C, Cd, N, Ni, Mn	C, Cd, N, Ni, Co	...	
	(.71)	(.71)		
6	C, Cd, N, Ni, Mn, Cr	C, Cd, N, Mn, Cr, Co	...	
	(.68)	(.68)		
7	C, Cd, N, Ni, Mn, Cr, Co	...		
	(.63)			
8	C, Cd, N, Ni, Mn, Cr, Co, Dep	...		
	(.58)			

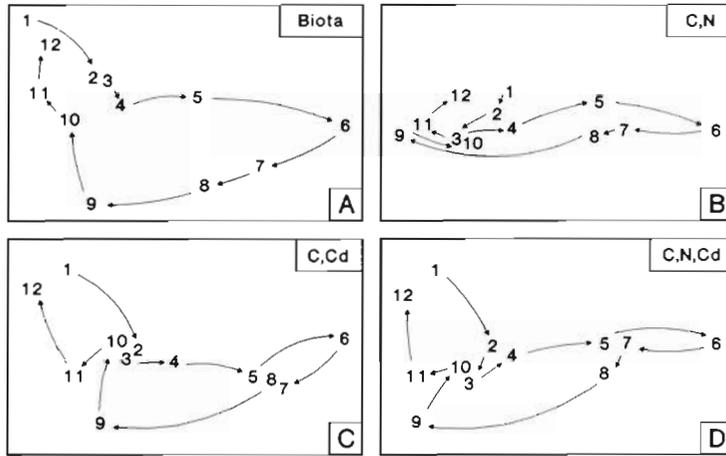


Fig. 8. Firth of Clyde macrofauna. MDS plots of the sampling sites across the sludge dumpground (centred at Site 6), based on: (A) transformed biomass of 84 macrofaunal species; (B) to (D) 3 combinations of carbon, nitrogen and cadmium levels (transformed) in the sediments. The matching coefficients  $\rho_w$  (Table 3) are: (B) 0.75; (C) 0.78; and (D) 0.79. Stress values are: (A) 0.04; (B) 0; (C) 0.01; and (D) 0.01

The most interesting conclusion of this analysis is, perhaps, that a much-studied data set (e.g. Warwick et al. 1987, Warwick 1988), hitherto regarded largely as an example of a simple organic enrichment gradient, has additional biological structure which is amenable to 'explanation' by this methodology. Of course, as noted for the previous example, there can be no guarantee that Cd concentration (for example) is directly *causal* in the shaping of the community structure; it may simply be collinear with unmeasured characteristics. Observational field studies always carry this caveat, causality being approached only through field intervention experiments and laboratory studies (also with their attendant difficulties of interpretation, e.g. Underwood & Peterson 1988).

## CONCLUSIONS

### Computation

The calculations necessary for the methodology of this paper require adequate computing facilities but are not generally prohibitive. Naturally, the computing burden increases sharply with the number of environmental variables,  $v$ . The number of combinations for which  $\rho_w$  needs to be evaluated in an exhaustive search is

$$\sum_{k=1}^v \frac{v!}{k!(v-k)!} \equiv 2^v - 1 \quad (6)$$

that is, 1023 combinations for the most extensive of the 3 examples considered here, the Messolongi diatom

data (Table 2). An exhaustive search may not always be necessary: (a) as explained earlier, the set of variables will often be reduced by excluding those which are collinear with other measures and can contribute nothing to the interpretation; (b) there may be situations in which specific variables are causally implicated, on external evidence, and it is desired to perform the matching process with these variables always included in the abiotic combination – there are no new issues of principle or practical difficulty in achieving this. (Again note the analogy to a multiple regression in which certain explanatory variables are always fitted before the contributions of others are assessed.)

Computation time for each value of  $\rho_w$  also increases at least in proportion to the square of the number of samples,  $n$ . Nonetheless, the full computation for the Messolongi data takes only about 15 min on an 80286 IBM PC (with standard maths co-processor).

The requisite software, termed the BIO-ENV program, is coded in Fortran and links to the in-house PRIMER package (Plymouth Routines In Multivariate Ecological Research), developed mainly by M. R. Carr, to run on IBM PC-compatibles at the Plymouth Marine Laboratory. This package was used to perform the MDS analyses of this paper, essentially by the Kruskal non-metric algorithm (Kruskal & Wish 1978).

### Further questions

The BIO-ENV procedure proposed in this paper should be regarded as in an embryonic state and there is considerable scope for further theoretical and practical investigation of its properties. A number of questions immediately arise.

(1) *Is there a testing framework which would assess the 'statistical significance' of the optimal combination of environmental variables? And to what extent does the repeated computation of  $\rho_w$  encourage 'selection bias' towards a more complex explanation than the data truly support?* The short answer to the first question has to be 'no'. It is unclear how a comprehensive testing framework could be constructed on the basis of the very weak assumptions that we have postulated about the nature of sample relationships in the data and the form of the link between biota and environment. One simple test is possible however: that of whether a specific environmental combination is 'better than random' in its match to the biotic pattern, that is, a test of the null hypothesis that  $\rho_w$  is effectively zero. As noted earlier, it is *not* valid to examine this by

standard tables or large sample normality approximations, even if the unweighted Spearman  $\rho_s$  coefficient is used. The standard tests are based on 'independent and identically distributed observations' underlying each set of ranks; the elements of a dissimilarity matrix are certainly not that! Instead, one can use a randomisation (permutation) test in which the sample labels in the environmental dissimilarity matrix, say, are randomly reassigned and the match  $\rho_w$  to the biotic matrix recomputed. This operation is repeated on a large number of occasions, generating the spread of small positive (and negative)  $\rho_w$  values that could be expected by chance if there is truly no relation between biotic and abiotic patterns. The 'significance' of the genuine  $\rho_w$  is then determined by where it falls in relation to the upper tail of this simulated probability distribution.

One should not forget the 'selection bias' effects, on the overall significance level, of the large number of individual tests implicit in comparing the biotic pattern with the *optimal* environmental combination from an exhaustive search. A crude compensation for this would be to demand a very high individual significance level, for example, to reject the null hypothesis only if the observed  $\rho_w$  is greater than that for any of the random permutations in  $20(2^v - 1)$  simulations. This would guarantee an 'experiment-wide' significance level of  $p < 0.05$ , although very conservatively since the  $(2^v - 1)$  potential tests are strongly interdependent. This aspect could be improved upon but is of little importance since the test itself is of limited applicability. Situations in which it is useful to test the hypothesis that there is *no* relationship between biotic and environmental patterns are presumably rather rare. Practical interest will more likely centre around whether addition of an explanatory variable 'significantly increases' an already non-zero value of  $\rho_w$ . A formal statistical framework for such tests is, perhaps, too much to expect on the present assumptions (or lack of them).

This is not as serious as it may at first appear, as evidenced by the analogy with multiple regression. It has long been recognised that the significance testing framework in, for example, stepwise multiple regression (Draper & Smith 1981) is not without logical inconsistencies, and its most defensible role is as an *exploratory* tool. As such, the main function of significance testing is then to provide a sensible 'stopping rule', since adding further explanatory variables – even entirely random ones – always produces some improvement in the residual sum of squares. By contrast, the BIO-ENV procedure has a natural stopping rule in the propensity of  $\rho_w$  to *decrease* with inclusion of unimportant variables. It seems intuitively clear that the addition of a totally irrelevant environmental

variable, whilst it may occasionally accommodate some minor feature of the biotic ordination purely by chance, is certain to conflict strongly with many of the main features of the plot and, overall, decrease the value of  $\rho_w$ . Thus the requirement to search through a large number of variable combinations, in determining the optimal match, is less likely to raise the spectre of selection bias than might at first appear. Some practical evidence for this can be adduced from the coherence of the lattice structures in the earlier examples (more comprehensive versions of Tables 1 to 3, not shown): what appear to be irrelevant variables consistently degrade the match, virtually wherever they are included. There is clearly scope, however, for addressing this issue through simulation studies.

Another central feature of the proposed methodology is the simple visual assessment that can be made, from the ordinations, of the significance (= importance) of an improvement in  $\rho_w$ . It is well-known, of course, that what a scientist or layman understands by the word 'significance' does not always equate with 'statistical significance': differences that are established beyond reasonable statistical doubt can nonetheless be inconsequential. Thus, a marginal adjustment in the location of one sample in an environmental plot may have no implications for the overall conclusions of the study, and the question of whether this adjustment leads to a 'statistically significant' improvement in the match may be academic. *Above all, the BIO-ENV procedure should be thought of as exploratory.* In fact, just as it is desirable in multiple regression that a subset model be assessed on how well it explains a further, independent data set, so the BIO-ENV procedure would benefit from a 2-stage approach. Only those environmental variables identified as contributing to the optimal match with the biotic data would be analysed in a repeat study; variables making a marginal, spurious contribution to the original match (if any) would be almost certain not to feature in the second analysis.

(2) *Under precisely what assumptions is the BIO-ENV procedure valid?* The main techniques proposed for relating biotic to abiotic data are linked in a natural way to the ordination methods used for display. Thus, the assumption that sample dissimilarities are well-represented by Euclidean distance for both environmental *and* species data, and that abundances are *linearly* related to environmental gradients, leads to classical Canonical Correlation (e.g. Mardia et al. 1979) or the variant known as Redundancy Analysis (Rao 1964). Essentially, all links within and *between* biotic and abiotic variables are assumed to be adequately summarised by standard correlation coefficients, and the associated ordination technique is simply Principal Component Analysis. Similarly, the important method of Canonical Correspondence Analysis introduced by

C. J. F. ter Braak (Jongman et al. 1987) is derived from the assumption that the species abundances have unimodal responses across the measured environmental gradients, and the dual relationships of biotic and abiotic data are accurately displayed by Correspondence Analysis, with its implicit definition of ' $\chi^2$  distance' as the measure of sample dissimilarity.

Our stance has been to retain maximum freedom in choosing measures of sample similarity which are individually relevant to the biological and environmental contexts. These then lead to *separate* graphical representations, in appropriately flexible manner, by non-metric MDS ordination. (Note the contrast with the above methods, where the species-environment relationships are embedded at an earlier stage and will influence the individual representations of biotic and abiotic information.) MDS effectively relies only on the *rank* similarities, and the subsequent BIO-ENV procedure, itself only a function of these ranks, is therefore consistent with the chosen ordination technique. Selecting an environmental subset to optimise the match of biotic and abiotic patterns is an intuitively plausible and seductively simple procedure, but it is far from clear what assumptions (if any) are being made about the form of the species-environment relationships. There may be conditions under which the results are misleading; such possibilities could, in part, be examined by simulation.

(3) *What is the likely range of practical application?* Clearly the method requires that environmental data can genuinely be matched to biotic data site for site, or sample for sample. Presumably, the more closely environmental sampling is matched to the biotic material the better the chances of a good 'explanation', particularly for spatial studies. Note that the data sets of this paper are all examples of spatial, rather than temporal patterns of community structure. In fact, in 2 of the 3 cases, variations through time have been deliberately removed by averaging across seasons. Of course, the mechanics of the method would allow the linking of matched biotic and abiotic *time series* from a single site but a greater degree of caution is necessary here and there are good reasons for expecting a lower rate of success: (a) Temporal autocorrelation within each series would not be well catered for. For example, the global randomisation test for complete absence of a match would be misleading for serially-correlated data. (b) The environmental variables shaping community trends are less likely to be measurable concurrently with the biotic data. It may be necessary for the abiotic variables to incorporate lags in time (and even in space, in the context of pelagic data), and the scale of the linking procedure could then easily get out-of-hand.

Neither of these caveats is specific to the current methodology; they are a reflection of the greater diffi-

culty generally encountered in handling and interpreting time-series data.

For purely spatial studies, the BIO-ENV procedure has now been applied in a reasonably wide range of marine contexts. Apart from the 3 examples of this paper, Gee et al. (1992) employ it to assess meiofaunal community responses to differing physical factors and sediment concentrations of hydrocarbons and heavy metals, along a transect of the Elbe plume (German Bight). Agard et al. (1993) also use the program to examine the relation of tropical macrobenthic communities to a suite of physical and chemical variables recorded at 31 stations in a coastal region of Trinidad. At an earlier stage of development, some of the methods of this paper were utilised by Warwick et al. (1991) to link intertidal macrobenthic invertebrate data, for 46 sites in 6 estuaries of southwest Britain, to static and dynamic physical characteristics of the estuarine sites. In addition, Warwick & Clarke (1991) referred to this putative methodology in several contexts, including the sampling of macrobenthic communities at 39 sites around the Ekofisk oilfield in the North Sea (Gray et al. 1990). Whilst further practical application, simulation studies and comparative trials are clearly desirable, the evidence so far suggests that these ideas *can* play a useful exploratory role in identifying *potential* agents in the shaping of community structure. Perhaps the technique's greatest strength is the relative simplicity of both the concept and the resulting graphical presentations, which non-specialists appear to find very accessible.

*Acknowledgements.* We thank Melanie Austen, Martin Carr, Richard Warwick (Plymouth Marine Laboratory) and Martin Green (NIVA, Oslo) for helpful discussions, and Daniel Daniellidis (University of Athens) for permission to use data from his Ph.D. thesis for illustrative purposes. The work forms part of the Community Ecology project of the Plymouth Marine Laboratory, with the BIO-ENV software development undertaken by M.A. being funded by the U.K. Department of the Environment, as part of its coordinated programme of research on the North Sea. We also thank Charles Green for coding additional features in BIO-ENV, under funding from the U.K. National Rivers Authority.

#### LITERATURE CITED

- Agard, J. B. R., Gobin, J., Warwick, R. M. (1993). Analysis of marine macrobenthic community structure in relation to natural and man induced perturbations in a tropical environment (Trinidad, West Indies). *Mar. Ecol. Prog. Ser.* 92: 233–243
- Bray, J. R., Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27: 325–349
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* (in press)

- Clarke, K. R., Green, R. H. (1988). Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Prog. Ser.* 46: 213–226
- Danielidis, D. B. (1991). A systematic and ecological study of diatoms in the lagoons of Messolongi, Aitoliko and Kleissova (Greece). Ph.D. thesis, University of Athens
- Draper, N. R., Smith, H. (1981). *Applied regression analysis*, 2nd edn. Wiley, New York
- Field, J. G., Clarke, K. R., Warwick, R. M. (1982). A practical strategy for analysing multispecies distribution patterns. *Mar. Ecol. Prog. Ser.* 8: 37–52
- Gee, J. M., Austen, M., De Smet, G., Ferraro, T., McEvoy, A., Moore, S., Van Gausbeki, D., Vincx, M., Warwick, R. M. (1992). Soft sediment meiofauna community responses to environmental pollution gradients in the German Bight and at a drilling site off the Dutch coast. *Mar. Ecol. Prog. Ser.* 91: 289–302
- Gower, J. C. (1971). Statistical methods of comparing different multivariate analyses of the same data. In: Hodson, F. R., Kendall, D. G., Tautu, P. (eds.) *Mathematics in the archaeological and historical sciences*. Edinburgh University Press, Edinburgh, p. 138–149
- Gray, J. S., Aschan, M., Carr, M. R., Clarke, K. R., Green, R. H., Pearson, T. H., Rosenberg, R., Warwick, R. M. (1988). Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.* 46: 151–165
- 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. Prog. Ser.* 66: 285–299
- Jongman, R. H. G., Ter Braak, C. F. J., Tongeren, O. F. R. van (1987). *Data analysis in community and landscape ecology*. Pudoc, Wageningen
- Kendall, M. G. (1970). *Rank correlation methods*. Griffin, London
- Kenkel, N. C., Orloci, L. (1986). Applying metric and non-metric multidimensional scaling to some ecological studies: some new results. *Ecology* 67: 919–928
- Kruskal, J. B., Wish, M. (1978). *Multidimensional scaling*. Sage Publications, Beverly Hills
- Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate analysis*. Academic Press, London
- Pearson, T. H. (1987). The benthic biology of an accumulating sludge disposal ground. In: Capuzzo, J., Kester, D. (eds.) *Biological processes and wastes in the ocean. Oceanic processes: marine pollution*, Vol. 1. Krieger, Melbourne, FL, p. 195–200
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā A* 26: 329–358
- Underwood, A. J., Peterson, C. H. (1988). Towards an ecological framework for investigating pollution. *Mar. Ecol. Prog. Ser.* 46: 227–234
- Warwick, R. M. (1971). Nematode associations in the Exe estuary. *J. mar. biol. Ass. U.K.* 51: 439–454
- Warwick, R. M. (1988). The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar. Pollut. Bull.* 19: 259–268
- Warwick, R. M., Clarke, K. R. (1991). A comparison of some methods for analysing changes in benthic community structure. *J. mar. biol. Ass. U.K.* 71: 225–244
- Warwick, R. M., Clarke, K. R. (1993). Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. *Mar. Ecol. Prog. Ser.* 92: 221–231
- Warwick, R. M., Goss-Custard, J. D., Kirby, R., George, C. L., Pope, N. D., Rowden, A. A. (1991). Static and dynamic environmental factors determining the community structure of estuarine macrobenthos in SW Britain: why is the Severn Estuary different. *J. appl. Ecol.* 28: 329–345
- 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
- Warwick, R. M., Platt, H. M., Clarke, K. R., Agard, J., Gobin, J. (1990). Analysis of macrobenthic and meiobenthic community structure in relation to pollution and disturbance in Hamilton Harbour, Bermuda. *J. exp. mar. Biol. Ecol.* 138: 119–142

*This article was submitted to the editor*

*Manuscript first received: March 3, 1992*

*Revised version accepted: November 30, 1992*