ELSEVIER

Journal of Experimental Marine Biology and Ecology 330 (2006) 55 – 80

Journal of EXPERIMENTAL MARINE BIOLOGY AND ECOLOGY

www.elsevier.com/locate/jembe

On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages

K. Robert Clarke [a,b,*], Paul J. Somerfield [a], M. Gee Chapman [b]

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

[b] *Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology Laboratories A11, University of Sydney, NSW 2006, Australia*

Received 1 June 2005; received in revised form 12 October 2005; accepted 15 December 2005

Abstract

Bray–Curtis similarity is widely employed in multivariate analysis of assemblage data, for sound biological reasons. This paper discusses two problems, however, with its practical application: its behaviour is erratic (or even undefined) for the vanishingly sparse samples that may be found as an end-point to a severe impact gradient, or a start-point in colonisation studies; and, in common with all similarity measures on species-level data, it is sensitive to inconsistency of taxonomic identification through time. It is shown that the latter problem is ameliorated by application of 'taxonomic dissimilarity' coefficients, a natural extension of the concept of taxonomic distinctness indices. Two previous suggestions for use with presence/absence data, denoted here by Γ^+ and Θ^+, are noted to be simple generalisations of the Bray–Curtis and Kulczynski measures, respectively. Also seen is their ability to permit ordinations of assemblages from wide geographic scales, with no species in common, and for which Bray–Curtis would return zero similarity for all pairs of samples.

The primary problem addressed, however, is that of denuded or entirely blank samples. Where it can be convincingly argued that impoverished samples are near-blank *from the same cause*, rather than by random occurrences from inadequate sample sizes (tow length, core diameter, transect or quadrat size etc.), a simple adjustment to the form of the Bray–Curtis coefficient can generate meaningful MDS displays which would otherwise collapse, and can improve values of the ANOSIM *R* statistic (increased separation of groups in multivariate space). It is also shown to have no effect at all on the normal functioning of a Bray–Curtis analysis when at least a modest amount of data is present for all samples.

Examination of the properties of this 'zero-adjusted' Bray–Curtis measure goes hand-in-hand with a wider discussion of the efficacy of competing similarity, distance or dissimilarity coefficients (collectively: resemblance measures) in community ecology. The inherent biological guidelines underlying the 'Bray–Curtis family' of measures (including Kulczynski, Sorenson, Ochiai and Canberra dissimilarity) are made explicit. These and other commonly employed measures (e.g. Euclidean, Manhattan, Gower and chi-squared distances) are calculated for several 'classic' data sets of impact events or gradients in space and time. Behaviour of particular coefficients is judged against the interpretability of the resulting ordination plots and an objective measure of the ability to discriminate between a priori defined hypotheses, representing impact conditions. A second-stage MDS plot of a set of resemblance coefficients, based on the respective similarities of the multivariate patterns each generates (an MDS of MDS plots, in effect), is seen to be useful in determining which coefficients are extracting essentially different information from the same assemblage matrix. This suggests a mechanism for practical classification of the plethora of resemblance measures defined in the

* Corresponding author. Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK. Tel.: +44 1752 633100; fax: +44 1752 633101.

E-mail address: krc@pml.ac.uk (K.R. Clarke).

0022-0981/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jembe.2005.12.017

literature. Similarity-based ANOSIM R statistics and Spearman ρ correlations, whose non-parametric structure make them absolutely comparable across different resemblance measures, answer questions about whether the different information extracted by some coefficients is more, or less, helpful to the final biological interpretation.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Bray–Curtis; Coefficient comparison; Dissimilarity; Second-stage MDS; Sparse assemblage; Taxonomic distinctness

1. Introduction

1.1. Guidelines underlying Bray–Curtis and related resemblance measures

A widely used strategy for multivariate analysis of community data (Clarke, 1993) computes between-sample similarity or dissimilarity coefficients prior to displaying sample patterns by non-metric multidimensional scaling (MDS), and examining differences between a priori defined groups by Analysis of Similarity (ANOSIM) tests. Community data are here regarded as any assemblage matrix, consisting of counts, biomass, % cover etc. (y_{ij}), of each of a number of taxa ($i=1, 2,\ldots,p$) in a number of samples ($j=1, 2,\ldots, n$), and we have chosen to refer collectively to dissimilarity, distance, similarity and 'nearness' coefficients as resemblance measures. Unlike some approaches, e.g. Principal Components Analysis (PCA) and Correspondence Analysis (CA), in which a particular resemblance measure is forced on the investigator (Euclidean and chi-squared distance, respectively), the above strategy caters for any coefficient, and Legendre and Legendre (1998) list about 50 possible choices! Nonetheless, the Bray–Curtis coefficient (Bray and Curtis, 1957), sometimes referred to as the Steinhaus or Odum (1950) coefficient, has been very widely employed by ecologists and environmental scientists. Bray–Curtis dissimilarity between two samples [$\equiv 100-$(Bray–Curtis similarity)] is defined as:

$$D_{12}^{\mathrm{B-C}} = 100 \cdot \frac{\sum_i |y_{i1} - y_{i2}|}{\sum_i (y_{i1} + y_{i2})} \qquad (1)$$

It, and related coefficients (Kulczynski, Canberra dissimilarity, a quantitative form of Ochiai, etc.—see later for definitions), which we coin the 'Bray–Curtis family', are implicitly motivated by the following guidelines for a dissimilarity measure.

(i) *Coincidence*: the measure takes the value zero only when the two samples are identical.
(ii) *Complementarity*: it takes its maximum value (100 or 1, the former being used here) when two samples have no species in common, irrespective of the precise abundances.
(iii) *Relative invariance*: a simple scaling change (e.g. recording biomass in g rather than mg) does not affect the relative values of a set of resemblances.
(iv) *Independence of joint absence*: exclusion or inclusion of taxa which are not present in either sample does not affect the resemblance between two samples.
(v) *Localisation*: the inclusion of an additional sample in the analysis does not affect the resemblance between existing samples.
(vi) *Dependence on totals*: the coefficient is able to combine change in relative composition with change in sample totals, where the latter carries useful information.

The universal validity of these guidelines would certainly not be accepted by all ecologists, in fact should not be accepted by any ecologist in *all* circumstances: the relevance of the guidelines will be context-dependent. What the guidelines do, however, is to make explicit the conditions under which Bray–Curtis dissimilarity, and related measures, would be expected to capture the important assemblage relationships. The extent to which, in a specific context, these guidelines should be modestly relaxed, excised, replaced or even reversed, determines the explicit or implicit coefficient choice.

While most resemblance measures satisfy the *coincidence* axiom, *complementarity* is a strong constraint, uniquely biological, which is not satisfied by many coefficients. (Indeed, some have argued against using dissimilarity measures at all, with their fixed upper limit of 100, rather than a distance measure defined to $+\infty$, see Cao et al., 1997). *Relative invariance* is clearly desirable for many types of community data (density, biomass etc.) although it is not necessarily required for coefficients based on (dimensionless) absolute count data. *Independence of joint absence* is another strong biological constraint (see below) which is failed by many measures, and while it may seem axiomatic that

the resemblance between two samples should depend only on the data for those samples, a number of coefficients do not obey *localisation* because they depend on some form of standardisation for each taxon by the total or maximum across all samples. Similarly, some measures automatically standardise only to relative composition of each species and therefore do not obey the *dependence on totals* condition. To ignore major differences in total content of each sample tends to be undesirable in environmental monitoring and some ecological work, except in cases where sample quantification (quadrat area, water volume, transect length, etc.) is poorly controlled. In more fundamental biological contexts (e.g. gut contents in diet studies), relative composition is usually preferable, but the point being made by this guideline is that the choice of whether to remove differences in sample totals should be in the hands of the experimenter, not dictated by, for example, an implicit distance measure underlying an ordination (as in CA).

Returning to (iv), *independence of joint absence* is an important and distinctive property whose strength in an ecological context is clear: species can be absent for many different reasons in different samples, and it is biologically unwise to infer that two samples are similar because neither contains a particular species. In other words, zero plays a special role in assemblage matrices — it can be misleading to treat it in the same way as any other number. To paraphrase Field et al. (1982): it is wrong to suggest that deep sea samples and brackish-water estuarine samples are similar because neither contains species found only in coastal waters. Also, extremes of spatial clustering of organisms, and low sampling volumes, often lead to replicates from entirely different groups (treatments/times/sites) having no representatives of particular species (or no species at all), and it would again be unwise to infer similarity between such samples on the strength of this. That being said, there are situations where a sample containing no organisms at all is ecologically interpretable, and dealing with such data is one of the main motivations for this paper.

1.2. Two problems with the Bray–Curtis coefficient

In spite of its general appeal, arising from the widespread validity of the above guidelines, one of the practical problems that can arise with the Bray–Curtis coefficient is its increasingly erratic behaviour as values within samples become vanishingly sparse. For example, when two samples consist only of a single individual their similarity varies between zero (the individuals are from different species) to 100 (they are from the same species). Ultimately, when both samples contain no individuals at all the Bray–Curtis coefficient is undefined, since the numerator and denominator of Eq. (1) are both zero and their ratio is indeterminate. This difficulty is an inevitable consequence of the axioms listed earlier, and the indeterminism will be shared by any resemblance measure which satisfies the *independence of joint absence* property. There are cases, however, where more knowledge about potential causes of sparse samples is available than is represented in the data matrix alone. For example, the samples could represent a series of increasingly impacted communities, from a common stressor, heading towards an end-point in which samples are entirely defaunated, or an experiment might start with complete clearance of the flora from replicate rock patches to examine the recolonisation process. In both cases, sufficient knowledge of the data structure is available to be able to say that two samples are entirely empty *from the same cause*. In these circumstances one would want to modify the dissimilarity measure so as to force it to return the value 0, rather than remain undefined. Similarly, two samples which are not empty, but nearly so, could be considered similar even if the handful of individuals they contain are from different species. For such situations, what is required (and provided later) is a modification to the definition of similarity to cope with these extreme cases, without losing the normal functioning of a coefficient such as Bray–Curtis, and its adherence to the earlier guidelines, when dealing with samples which are not denuded.

A different practical problem arises when species-level identifications are inconsistent between locations or through time. All standard similarity measures must be sensitive to such errors, if they are widespread through the species list, and occur for fixed points in time or certain spatial groups. What looks like an abrupt assemblage change at one time could simply result from improved identification skills (or taxonomic redefinition) as original taxa ‘disappear’ and are replaced by one or more ‘new’ taxa. Usually, however, this involves taxa which are closely related, e.g. a single species is now identified as two separate species within the same genus, and exploiting taxonomic relatedness of the species could bestow a degree of robustness of the analysis to such changes. The ideal, of course, is to reduce all samples to the lowest common denominator of consistent taxonomy but, where this is problematic, a presence/absence measure based on ‘taxonomic dissimilarity’, using the mean path length through the taxonomic hierarchy from a species in sample 1 to its

nearest relation in sample 2, could achieve that robustness. Instead of similarity between two samples coming only from taxa that match at the species level, contributions can now come from 'near misses' (e.g. a different species but in the same genus).

2. Methods

2.1. Definitions of taxonomic dissimilarity

Two such presence/absence based 'beta-diversity' coefficients were defined by Clarke and Warwick (1998a) and Izsak and Price (2001). These are a natural extension of the 'alpha-diversity' index of taxonomic distinctness, Δ^+ (Warwick and Clarke, 1995; Clarke and Warwick, 1998b), namely the average path length through a taxonomic (or phylogenetic/genetic) tree between every pair of species in a sample. The path length between species i and j is denoted by ω_{ij}, where for a standard Linnean classification the steps from species to genus, genus to family, etc. are regarded as equal, and the largest path length (e.g. between species in different phyla) is fixed at 100. The path length ω_{ii} between identical species is defined to be zero. In a unified notation, taxonomic dissimilarity between sample 1 (species subscripts i) and sample 2 (species subscripts j) is then formally defined in one of two ways:

$$\Gamma^+ = 100 \cdot \left(\sum\nolimits_i \min_j(\omega_{ij}) + \sum\nolimits_j \min_i(\omega_{ji}) \right) \Big/ (s_1 + s_2) \tag{2}$$

$$\Theta^+ = 100 \cdot \frac{1}{2} \left(\frac{\sum_i \min_j(\omega_{ij})}{s_1} + \frac{\sum_j \min_i(\omega_{ji})}{s_2} \right) \tag{3}$$

where s_1, s_2 are the number of observed species in samples 1 and 2. In words, Γ^+ (gamma$^+$) is the mean of all path lengths between each species in one sample and its closest relation in the other sample, whereas Θ^+ (theta$^+$) separately calculates the mean path length between all species in sample 1 and their nearest relations in sample 2, and vice versa, and then averages the two means. Clearly, if $s_1 = s_2$ then $\Gamma^+ = \Theta^+$. Note that Γ^+ is essentially the 'TD coefficient' of Izsak and Price (2001) and Θ^+ the 'optimal mapping coefficient' of Clarke and Warwick (1998a). The novel observation offered here is that it is not difficult to show algebraically that Γ^+ and Θ^+ reduce exactly to the presence/absence forms of the Bray–Curtis and Kulczynski coefficients, respectively, (see the later Eqs. (8) and (15)) when the taxonomic hierarchy is completely flattened, so that all species are in (say) the same genus. The distance to the nearest relation is then either 0 or 100 (the species is, or is not, found in the other sample). Γ^+ and Θ^+ are thus natural generalisations, to a hierarchy of relatedness, of these two measures in the 'Bray–Curtis family' (Kulczynski replaces the arithmetic mean of the sample totals in the denominator of Eq. (1) with a harmonic mean, so can be considered a variant of Bray–Curtis).

The likely merit of this generalisation in terms of added robustness to certain types of taxonomic errors has already been noted, but an additional advantage is that two samples with no species in common, and thus with Bray–Curtis dissimilarity of 100%, can now take a range of dissimilarities <100. If two samples tend to have species in similar genera or families to each other then taxonomic dissimilarity is low, whilst if they do not share many evolutionary branches the dissimilarity will remain large. It is now therefore possible to generate meaningful MDS plots and cluster analyses of assemblages from samples over a wide geographic scale, perhaps with no species in common at all; an example is given later. (This is not the place for a discussion on the shortcomings of Linnean taxonomies in representing phylogenies, but note that nothing in the formulation of Γ^+, Θ^+ or Δ^+ limits their use to taxonomic rather than phylogenetic or genetic trees—or even functional hierarchies!)

2.2. Definition of zero-adjusted Bray–Curtis

Turning now to the primary problem addressed by this paper, of denuded samples, the requirement is to modify the behaviour of the Bray–Curtis coefficient so that it is less erratic for samples with few individuals, and is defined for samples with complete absences. The solution is ad hoc, but simple, and rather analogous to that for use of the $\log(1+x)$ transformation in univariate statistics: $\log(x)$ is undefined as x tends to zero, hence one can 'feather in' its behaviour for small x by adding a constant before taking the log. Here, the solution is also to add a constant (to the denominator of the Bray–Curtis coefficient) though a more helpful way of thinking of this is as adding a 'dummy species' to the original abundance matrix, with value 1 for all samples. The effect of this extra species, on the dissimilarity between, say, samples 1 and 2 (Eq. (1)), is clearly to make no difference to the numerator (adding a term of $|1-1|$) and to increase the denominator by 2

(adding a term of (1+1)). The *zero-adjusted Bray–Curtis* coefficient is therefore:

$$D_{12}^{\mathrm{B-Cadj}} = 100 \cdot \frac{\sum_i |y_{i1} - y_{i2}|}{2 + \sum_i (y_{i1} + y_{i2})}. \quad (4)$$

The dummy species gives the desired value of D for two samples which are empty: they now have a single 'species' in common, with the same abundance, which gives a dissimilarity of zero. In fact, since for assemblage data all the y's are content variables which can never be negative, the denominator of Eq. (4) is always strictly positive, so D is always defined. Furthermore, for two samples which are very sparse, the dummy species gives them a certain amount of similarity; they share a (dummy) species, even if they have no real species in common. This is what was required: the dissimilarity between two samples tends smoothly to zero as the samples become vanishingly sparse.

There is a good reason for defining the dummy species value to be 1. Prior to calculating inter-sample resemblances, transformations are commonly employed (Clarke and Green, 1988) to downweight the contributions of the dominant species in relation to the less common and rarer ones. For a dummy species value of 1 and a transformation chosen from the usual sequence of power transforms of increasing severity (none, square root, fourth root, reduction to presence/absence) it makes no difference to the $D^{\mathrm{B-Cadj}}$ coefficient in Eq. (4) whether the transform is performed after the addition of the dummy species or before. Clearly, a power transform of the value 1 always returns the value 1, so the constant 2 is always added to the denominator of Eq. (4). In fact, the logarithmic transform can also be embedded in this sequence, if one uses the form $\log_2(1+y)$ rather than logs to the base e or base 10; again the dummy value of 1 is unchanged by this transformation because $\log_2(1+1)=1$.

When working with abundances which are integer counts, a dummy value of 1 is also a natural choice: it is the smallest, strictly positive number that the matrix could contain. For other content variables (biomass, % cover etc.), the natural choice is similarly to take the dummy species value as the smallest non-zero number in the table (the analogy is very close with common practice in univariate statistics, for the choice of c in the $\log(c+x)$ transformation, for non-integral x). Clearly, the larger the dummy value, the greater the risk of a dampening effect on dissimilarities between samples which are not particularly sparse. A dummy value which is too small, however, risks total irrelevance. It will always force two empty samples to be 100% similar, but near-empty samples which have their content in different species will still be deemed highly dissimilar, thus failing to moderate the erratic behaviour of Bray–Curtis for near-denuded samples. A judicious compromise is required here.

2.3. Example comparison of zero-adjusted Bray–Curtis with the unadjusted measure

It is apparent from the form of Eq. (4) that only a slight difference is possible between the zero-adjusted measure, with a dummy value of 1, and the original Bray–Curtis dissimilarities, for samples of counts which contain at least a modest number of individuals. The dummy species has the same abundance in all samples and therefore its contribution to the relative dissimilarity among samples can only be marginal, and indirect (through changing the sample totals). The magnitude of the effect can be seen for a small, hypothetical data set (Table 1), representing a colonisation process for independent quadrats, examined at a series of 12

Table 1
Hypothetical data matrix of 4 species by 12 samples, with decreasingly sparse abundances, showing the difference between the usual Bray–Curtis dissimilarities, calculated for adjacent samples, contrasted with the zero-adjusted Bray–Curtis coefficient

	Samp1	Samp2	Samp3	Samp4	Samp5	Samp6	Samp7	Samp8	Samp9	Samp10	Samp11	Samp12
Species 1	0	0	1	1	0	1	0	0	10	10	0	0
Species 2	0	0	0	0	1	0	0	5	0	20	0	50
Species 3	0	0	0	0	0	1	0	5	0	20	0	50
Species 4	0	0	0	0	0	0	2	0	10	0	50	100
(Dummy species	1	1	1	1	1	1	1	1	1	1	1	1)

	1 v 2	2 v 3	3 v 4	4 v 5	5 v 6	6 v 7	7 v 8	8 v 9	9 v 10	10 v 11	11 v 12
B–C dissimilarity	?	100	0	100	100	100	100	100	71.4	100	60
Zero-adjusted B–C	0	33.3	0	50	60	66.7	85.7	93.8	69.4	98.0	59.5

The latter can be thought of as simply calculating Bray–Curtis with an added row—the displayed 'dummy' species.

times, involving only 4 species. The sparseness generally reduces from left to right, though there is other structure present. Samples 8 and 9, for example, have reasonable numbers of organisms but of completely different species. For standard Bray–Curtis dissimilarities between adjacent pairs of samples, the dissimilarity is initially undefined (between blank samples) and is then unstable (0 between samples 3 and 4, but 100 between 4 and 5, 5 and 6 etc.). This does not reflect the biological reality. If these are all sparse samples at the beginning of a colonisation from a common, abiotic starting point, they should not be defined as 100% dissimilar. The behaviour of the zero-adjusted Bray–Curtis dissimilarity looks much more appropriate, as blank samples have a dissimilarity of 0, highly sparse samples tend to have low dissimilarity to each other even when their species do not match, and the adjustment makes very little difference when samples contain reasonable numbers of individuals (e.g. dissimilarities between 10 and 11, and between 11 and 12).

This preferable behaviour is also reflected in non-metric multidimensional scaling (MDS) ordination plots (Fig. 1) based on inter-sample resemblances. Note that in this paper we follow the terminology of Clarke (1993), and Clarke and Warwick (2001), in referring to non-metric multidimensional scaling as MDS, accepting that other forms of multidimensional scaling also exist (such as classical scaling, i.e. metric MDS). For the standard Bray–Curtis plot (Fig. 1a) samples 1 and 2 need to be omitted: their dissimilarity to everything else is 100, so they will collapse the non-metric MDS plot, even if this is able to tolerate an undefined dissimilarity between them (as some MDS routines are able to do). Such condensed MDS plots are a common occurrence where one or two samples have high dissimilarity to all others, and are inevitable with non-metric MDS if the dissimilarities involving such outliers are all larger than any dissimilarities among the remainder of the samples (the precise positions of the outliers are then essentially arbitrary because all the rank order relationships in the dissimilarity matrix can be satisfied by putting them anywhere 'sufficiently far away'). The display of the remaining samples, 3 to 12, fails to capture the salient features of the original matrix clearly. This is in sharp contrast with the MDS plot from the zero-adjusted Bray–Curtis dissimilarities (Fig. 1b) which retains samples 1 and 2, and clearly displays the colonisation process (right to left) without sacrificing the other sample structure orthogonal to this (e.g. the complete dissimilarity of samples 8 and 9).

The effect of the added constant on the bottom line of Eq. (4) is to make the zero-adjusted coefficient

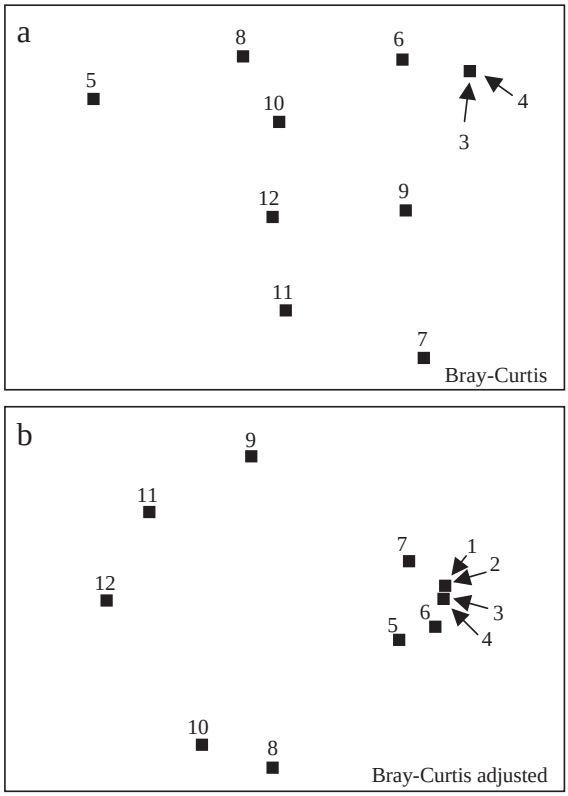

Fig. 1. MDS plots (non-metric) of the 12 hypothetical samples of Table 1, representing a colonisation process, based on: a) standard Bray–Curtis dissimilarity, omitting samples 1 and 2 (note, samples 3 and 4 are coincident); b) zero-adjusted Bray–Curtis, for all samples (note, samples 1 and 2 are coincident, as are samples 3 and 4). Stress values: a) 0.02, b) 0.08.

slightly smaller, in absolute value, than its Bray–Curtis counterpart. It is clear (Table 1) that the adjustment to Bray–Curtis only makes major changes for the sparsest of samples. Once numbers in two samples reach a reasonable level, differences in the absolute value of the dissimilarity between them become small, and their relative values (and rank order) are effectively the same. It is algebraically straightforward to quantify this effect precisely. The percentage reduction in Bray–Curtis dissimilarity D_{12}, consequent on the zero-adjustment, is $200/[T_1+T_2+2]$, where T_1 and T_2 are the totals for samples 1 and 2 ($\sum_i y_{i1}$ and $\sum_i y_{i2}$, respectively). If the total abundance of both samples, T_1+T_2, is 100 then the reduction in D is <2%. Large reductions are clearly only possible when T_1+T_2 tends to zero. As the relative values of dissimilarity tend to matter more than their absolute values one can repeat the same calculation for $(D_{12}-D_{34})$, namely the difference between the Bray–Curtis dissimilarities for samples 1 and 2 and

for samples 3 and 4. The percentage reduction in this difference, on making the zero-adjustment, is more algebraically complex, but in the special case where the totals for samples 1 and 2, T_1+T_2, and the totals for samples 3 and 4, T_3+T_4, are identical, the percentage reduction is again $200/[T_1+T_2+2]$, namely <2% if $T_1+T_2=T_3+T_4=100$.

2.4. Computations

All resemblance calculations and ordination plots in this paper were carried out with the PRIMER package, version 6 (Clarke and Gorley, 2006); plots were additionally annotated with presentation graphics software.

3. Results

3.1. Taxonomic dissimilarity analyses

Returning to the taxonomic dissimilarity measure Γ^+, two examples are now given of its practical usefulness. Fig. 2 displays the MDS ordination plots from soft-sediment benthic macrofaunal assemblages in 20 sediment cores taken at different distances (in a cross-shaped design) from a centre of oil-drilling activity, in the Valhall oilfield of the Norwegian sector of the North Sea (data discussed by Olsgard et al., 1997). Symbols of increasing size represent distances of samples from the oilfield centre, 500 m to 6 km. Displayed on each plot are community data from two years, 1988 and 1991, both taken some time after routine operation of the field commenced. Fig. 2a uses standard Bray–Curtis on presence/absence data (also known as Sørenson, see the later Eq. (8)), whereas 2b calculates the taxonomic dissimilarity Γ^+(Eq. (2)); plots for Θ^+ are similar. Whilst there is a clear suggestion of changing assemblage structure with distance from the oilfield, the dominant feature in Fig. 2a is of a time change between 1988 and 1991. This is purely artefactual, however, there being almost a doubling in the number of different species identified between the 1988 and 1991 lists; this does not represent a recovery process but a known increase in expertise in identification skills over that period. The use of taxonomic dissimilarity in Fig. 2b mitigates this artefact to a large degree (though it cannot remove it altogether), since much of the increase in the species list is almost certainly due to 'splitting' of formerly identified single species into two or more from within the same genus. The comparative pattern of community differences as the oilfield is approached in the two years is now more evident, and shows relatively little time change, both in the differences between distance groups and the overall scale of assemblage change with distance.

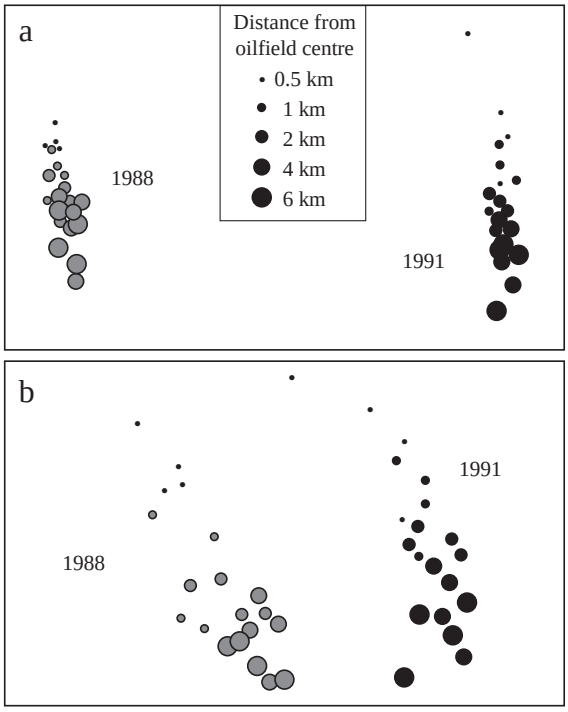

Fig. 2. MDS ordination for two years of macrobenthic assemblages from 5 distances along 4 perpendicular transects away from the Valhall oilfield, based only on species lists and a) standard Bray–Curtis, b) taxonomic dissimilarity Γ^+. Stress values: a) 0.05, b) 0.11.

Fig. 3 represents a more artificially constructed dataset designed to illustrate the behaviour of taxonomic dissimilarity in a case where a Bray–Curtis analysis cannot be carried out. Real data of species lists, of all fish recorded by FishBase (www.fishbase.org) as present in a cross-section of world island groups (see Fig. 3), are pruned to remove any species which occur at more than one of these locations. By definition, the resulting (pseudo-)endemic species matrix has Bray–Curtis dissimilarity of 100 for all pairs of islands, and an ordination is impossible. Taxonomic dissimilarities are perfectly well-defined, however, and an interpretable MDS plot results, displaying some natural juxtapositions of islands on this wide biogeographic scale (Fig. 3).

3.2. Zero-adjusted Bray–Curtis analyses

The remainder of this paper is centred around its main topic of adjustment to the quantitative Bray–Curtis coefficient (Eq. (1)) to deal with denuded samples (Eq.

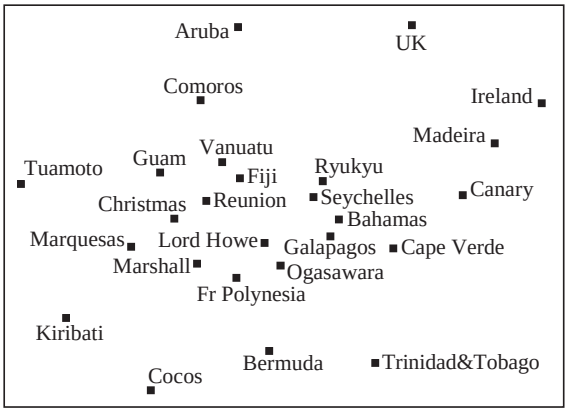

Fig. 3. MDS ordination of 26 island groups, based on taxonomic dissimilarity Γ^+ from lists of 'endemic' fish species for each location (endemic only in the sense of no species being found in more than one of these island groups). Stress value: 0.15.

(4)). A series of real data sets are used, the first three studies containing vanishingly sparse data for some sites or times, and they consist of a spatial defaunation gradient, a study of a naturally sparse assemblage and a time series of impact and recovery. Two further studies demonstrate that the adjustment in no way disrupts the normal behaviour of Bray–Curtis analyses in impact gradient studies which are either species rich throughout the gradient, or which are species-poor close to the main impact but still with high numbers of individuals of opportunist species. To place the performance of the adjusted coefficient in context, further standard resemblance measures are gradually introduced, related to the guidelines of the Introduction, and their practical performance examined.

Example 1. Boulby macrofauna, England

These data, from samples taken in September 1992, were collected as part of a long-term monitoring programme, designed to detect effects of mining effluent discharged into coastal waters from a potash mine at Boulby, East Cleveland (Brown and Shillabeer, 1997; Craig et al., 1993). Dense, saline effluent containing non-toxic clays is discharged through outfalls and the effect on the benthic community is primarily physical smothering. These data illustrate the behaviour of the Bray–Curtis adjustment in a realistic context, where the nature of the impact is to defaunate areas of the seabed.

Five sites are considered, with three replicates at three of the sites, giving a matrix of 49 species by 11 samples. Sites 33, 231 and 147 represent high levels of impact resulting in highly sparse assemblages. Sites 56 and 83 are relatively much less impacted (site 83 having a larger number of individuals, and a somewhat different species mix, than site 56). Fig. 4 contrasts the MDS ordinations based on square-root transformed data, using the Bray–Curtis measure (Fig. 4a) and its zero-adjusted form (Fig. 4b). For standard Bray–Curtis, the erratic behaviour for the highly sparse sites 231, 33 and 147 dominates the plot; they are not represented as similar to each other, with 147 and 231 appearing on 'opposite sides' of the two relatively less-impacted sites, 56 and 83. The real distinction between these latter two sites is completely lost. For the zero-adjusted form, however, the distinction between the impacted (right) and less-impacted sites (left) is the dominant feature, and the distinction between sites 56 and 83 (in relation to their replicate variability) is maintained.

Example 2. Boulder-field fauna, Australia

These data describe mobile macrofauna on the undersurfaces of similar-sized boulders, sampled from a boulder field in NSW, Australia, with 25 replicate

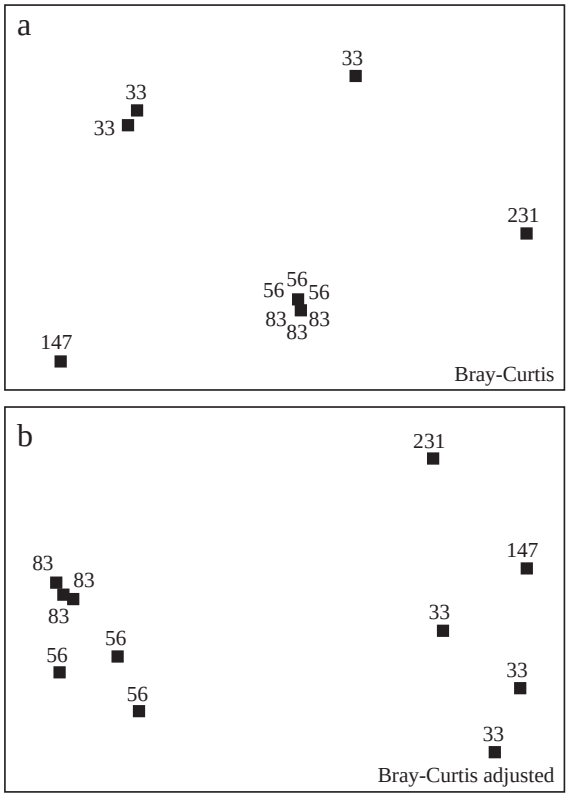

Fig. 4. Soft-sediment macrofauna from the Boulby sand patch. MDS plots contrasting: a) Bray–Curtis, with b) its zero-adjusted form, for 11 samples (by 49 species) from five sites. There are 3 replicates at each of the relatively unimpacted sites 56 and 83, a further 3 replicates at the impacted site 33, but only one at each of the impacted sites 147 and 231. Data were root-transformed before dissimilarities were calculated. Stress values: a) 0.01, b) 0.04.

boulders being examined at each of 4 sites. A total of 39 separate taxa were identified, but counts of any one taxon were small throughout and no transformation of the data was applied prior to the following analyses. The data are sparse, as is natural for this fauna (Chapman, 2002), particularly for sites 3 and 4, and six of the boulders (three from each of sites 3 and 4) had no attached organisms at all.

For a standard Bray–Curtis analysis empty samples must be removed altogether. The resulting MDS ordination (Fig. 5a) collapses into one main group and a single outlier from site 3, which is almost empty, containing only two individuals (in different species). If this sample, too, is removed, a better MDS plot results (Fig. 5b). In some cases (arguably here also) it is necessary to go through several such removals of denuded outliers before a stable structure emerges, based on samples with reasonable total counts.

For the zero-adjusted Bray–Curtis measure (Fig. 5c) it is not necessary to remove the single outlier from site 3, or even the six blank samples from sites 3 and 4, as they now have low dissimilarity with other impoverished samples. This MDS plot also appears to show a better separation of the sites than Fig. 5b. This can be more formally assessed by the ANOSIM R statistic (Clarke and Green, 1988; Clarke, 1993). This statistic is the difference between the average rank dissimilarities among groups and within groups, divided by a constant that ensures that R is always ≤ 1. $R=0$ when there are no group differences and $R=1$ when all samples in different groups are more dissimilar to each other than any samples in the same group. The usefulness of this rank-based formulation here is that the absolute values of R are comparable, as measures of group separation, from analyses based on entirely different dissimilarity coefficients.

A 1-way ANOSIM permutation test using the standard Bray–Curtis coefficient (having removed the all-blank samples) shows that sites are significantly different (first column of Table 2). All the pairwise comparisons given in the remaining columns of Table 2 are also clearly significant, although it might be considered that the test for sites 3 and 4 is borderline. Significance is of less interest here than the absolute values of R, which show that in every case (bar one, with no change) the zero-adjusted coefficient (without removals) improves the separation of the sites compared to the standard coefficient.

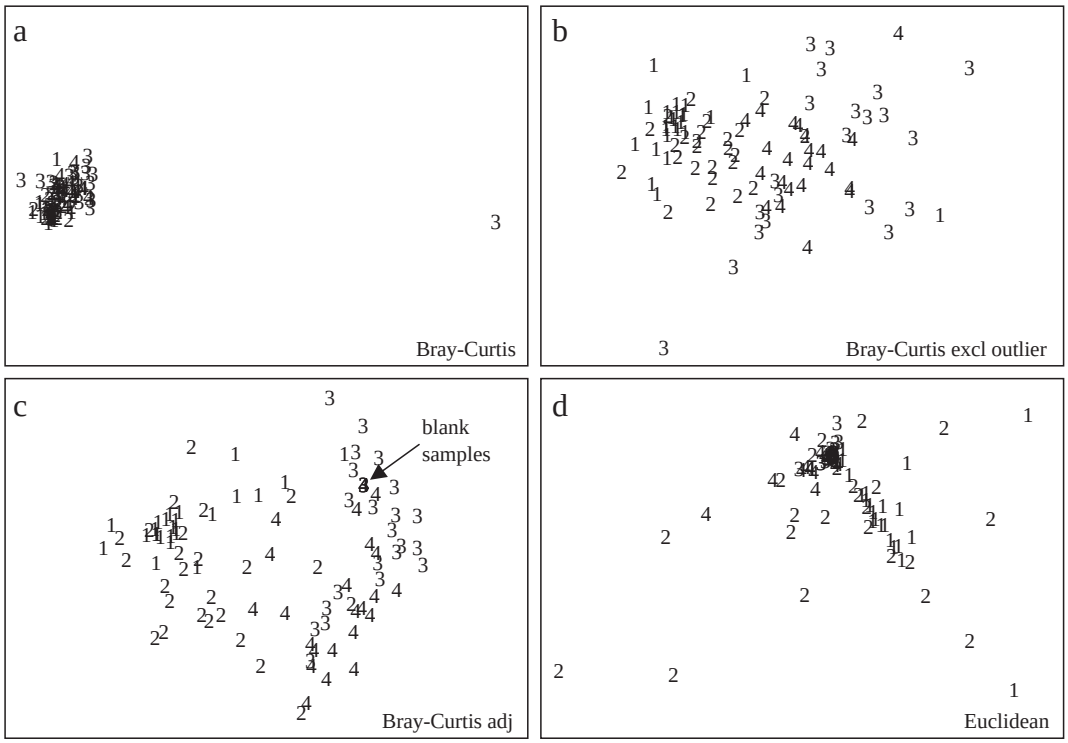

Fig. 5. Mobile macrofauna from a boulder-field, NSW, Australia. MDS plots for untransformed counts from 25 replicate boulders at 4 sites, contrasting: a) Bray–Curtis, excluding 6 all-blank samples; b) Bray–Curtis, excluding the further outlying sample from site 3; c) zero-adjusted Bray–Curtis, using all 100 samples; d) Euclidean distance, also using all 100 samples. Stress values: a) 0.10, b) 0.12, c) 0.17, d) 0.08.

Table 2
Boulder-field mobile macrofauna, at 4 sites in NSW, Australia: ANOSIM *R* statistic for testing for differences between all sites (first column) and pairwise comparisons of sites (subsequent columns), with significance level from 999 permutations given in brackets (this can never be smaller than 0.1%)

	Global *R*	*R* for 1 v 2	*R* for 1 v 3	*R* for 1 v 4	*R* for 2 v 3	*R* for 2 v 4	*R* for 3 v 4
Bray–Curtis	0.41 (0.1%)	0.12 (0.3%)	0.59 (0.1%)	0.75 (0.1%)	0.45 (0.1%)	0.46 (0.1%)	0.06 (2.6%)
Zero-adj B–C	**0.51 (0.1%)**	**0.12 (0.3%)**	**0.84 (0.1%)**	**0.82 (0.1%)**	**0.69 (0.1%)**	**0.52 (0.1%)**	**0.12 (0.5%)**
Kulczynski	0.39 (0.1%)	0.11 (0.4%)	0.59 (0.1%)	0.75 (0.1%)	0.42 (0.1%)	0.40 (0.1%)	0.05 (4.4%)
Euclidean	0.22 (0.1%)	0.08 (0.8%)	0.44 (0.1%)	0.41 (0.1%)	0.24 (0.1%)	0.17 (0.1%)	0.04 (4.7%)

Rows are four different dissimilarity measures; the best site separations (largest *R*) are consistently achieved by the zero-adjusted Bray–Curtis coefficient, shown in bold.

Some of the improvements are quite dramatic, for example the 1 v 3 and 2 v 3 comparisons. Of course, there is a relationship between power and the number of replicates used for a test (Somerfield et al., 2002), so minor improvements in significance levels could result from the increased number of replicates used in the zero-adjusted case, but the improvements here are in the values of *R*, which are not, in general, related to the number of replicates. Instead, *R* reflects the true size of location shifts in communities between the sites. The *R* values in Table 2, and the comparison of Fig. 5c with b, both demonstrate the improvement in behaviour of the modified coefficient.

To widen the comparison, Kulczynski's (1928) measure is added to Table 2. This obeys all the guidelines in the Introduction and can thus be considered to be in the 'Bray–Curtis family'. It was identified by Faith et al. (1987) as their best-performing coefficient, marginally preferable to Bray–Curtis, in a comparative study which attempted to recreate a simulated 2-d ecological gradient on a rectangular grid. The coefficient is defined as:

$$D_{12}^{\mathrm{Kul}} = 100\left(1 - \frac{\sum_i \min\{y_{i1}, y_{i2}\}}{\mathrm{HM}\left(\sum_i y_{i1}, \sum_i y_{i2}\right)}\right) \tag{5}$$

where $\mathrm{HM}(T_1, T_2) = 2/[(1/T_1) + (1/T_2)]$ is the harmonic mean of T_1 and T_2. In fact, although not immediately obvious, Kulczynski is algebraically the same as the Bray–Curtis coefficient (Eq. (1)) except that the simple arithmetic mean $(T_1 + T_2)/2$ in the denominator of Eq. (1) is replaced by a harmonic mean in Eq. (5). If the totals for samples 1 and 2 are the same, the two coefficients are therefore identical, and in practice one would expect them to give different analysis patterns only if the sample totals (post-transformation) were rather variable, with some of them relatively close to zero. As a variation on Bray–Curtis, it too is undefined for all-blank samples and the 6 such replicates have been removed for the calculations. The resulting MDS plot is not shown but is very similar to Fig. 5b upon removal of the same site 3 outlier, and the ANOSIM tests return *R* values (Table 2) which are very close to those for standard Bray–Curtis, and inferior to those for the Bray–Curtis adjustment. The Kulczynski coefficient, sometimes advocated as an alternative to Bray–Curtis, is therefore seen to have equally erratic behaviour for sparse samples.

The use of the zero-adjusted form in Eq. (4) implies, of course, that some of the 'Bray–Curtis family' guidelines of the Introduction are contravened. As noted previously, no coefficient which strictly obeys the *independence of joint absence* property can return a defined value for the dissimilarity between two all-blank samples. If this guideline is to be broken, why not, it might be asked, move to another commonly used coefficient that does not obey *independence of joint absence*? An example would be Euclidean distance:

$$D_{12}^{\mathrm{Euc}} = \sqrt{\sum_i (y_{i1} - y_{i2})^2} \tag{6}$$

which is the distance measure implicit in many commonly applied multivariate methods, such as Principal Components Analysis (PCA, Krzanowski and Marriott, 1994). In fact, there are good reasons for not doing so in assemblage studies involving many species, most of which are absent in some of the samples. If patterns of species presence/absence are then at least as important to interpretation as quantification of species which *are* present, then Euclidean and other simple distance measures, such as Manhattan, Minkowski etc., invariably perform badly (zero plays no special role and is treated in the same way as any other number). Most of the guidelines of the Introduction are not marginally contravened, as is the case for the Bray–Curtis adjustment, but are not fulfilled at all. The point is that these are not a motivating set of guidelines for simple distance-based measures: Euclidean distance breaks *complementarity* and *independence of joint*

absence in a major way, with important practical consequences. The issue is returned to later, but note for now that an MDS based on Euclidean distance shows a radically different picture of these data (Fig. 5d). It is dominated almost completely by the widely differing sample totals, rather than any presence or absence structure. It is also much less successful in separating the groups. Sites 1 and (particularly) 2 have high variability, whereas replicates from sites 3 and 4 collapse on each other as a consequence of the latter's generally smaller total abundances across all species. This is in stark contrast with both kinds of Bray–Curtis plots, in which the variability is of a similar magnitude across sites. The lesser ability to separate the groups with the Euclidean distance-based analysis is indicated by much lower values for ANOSIM R (Table 2), although even inconsequential differences are still significant due to the high power resulting from 25 replicates in each site group.

Example 3. Coral communities, Indonesia

In a study from Tikus Island, Thousand Islands, Indonesia (Warwick et al., 1990), coral assemblages were quantified in terms of % cover along 10 replicate 30 m line transects in each of the years 1981, 1983, 1984, 1985, 1987 and 1988. A total of 75 different species were observed. Significant changes were expected following the 1982–3 El Niño, and Fig. 6a demonstrates this for the standard Bray–Curtis coefficient, calculated on square-root transformed data. In fact, as is shown later, the transformation makes little difference to the outcome but, under the approach of Clarke and Warwick (2001), a mild transform is called for since one species of *Heliopora* in particular accounts for up to 40% of the cover for a few of the transects, and would tend to be given a rather dominant weight in an untransformed analysis. The plot shows a strong shift in the assemblage between 1981 and 1983 with subsequent years falling intermediately, suggesting some degree of recovery from the El Niño event. Also there is apparently increased variability among replicates in 1983 compared to other years. In contrast, for the MDS plot based on the zero-adjusted Bray–Curtis coefficient (Fig. 6b) the location shift is, if anything, clearer still, and the multivariate dispersion is seen to be more equitable across the years.

As was the case for the boulder-field example above, a Euclidean distance plot (Fig. 6c), again based on root-transformed data, is less successful in separating the groups and presenting a clear signal of change and partial recovery. It is dominated almost completely by the sample totals, which are larger and more variable (in absolute terms) for replicates from 1981 than the low values of 1983, the latter points therefore collapsing onto each other.

Contrasting MDS patterns based on two further dissimilarity measures can also be seen in Fig. 6. In order to be strictly comparable, both are again calculated from root-transformed data arrays. The first uses the Kulczynski measure (Fig. 6d). Compared with the standard Bray–Curtis analysis (Fig. 6a) the Kulczynski coefficient is clearly less successful in separating the years, particularly the 1981 samples, which generally can be distinguished with the other coefficients. Fig. 6e shows the effect of a different coefficient, χ^2 distance:

$$D_{12}^{\chi^2 dist} = \sqrt{\sum_i \frac{1}{y_{i+}/\sum_i y_{i+}} \left(\frac{y_{i1}}{\sum_i y_{i1}} - \frac{y_{i2}}{\sum_i y_{i2}} \right)^2},$$

$$y_{i+} = \sum_j y_{ij}, \tag{7}$$

which is commonly encountered as it is the distance measure implicitly underlying correspondence analysis and its derivatives (Detrended Correspondence Analysis, Canonical Correspondence Analysis, etc., see Ter Braak, 1986). This coefficient is derived essentially from a (discrete) multinomial model for the entries in the species by samples matrix, and a set of guidelines motivating its use would therefore include a restriction to counts behaving as genuine frequencies (arrivals of individuals of a species into the sample being independent of each other, and of other species; irrelevance of marginal totals to interpretation etc.). These assumptions are poor for the continuous variables of percentage cover in this study, as is seen from the failure of Fig. 6e to capture much meaningful structure (and three outlying replicates from 1983 have been removed for the MDS to get even this far). In many ways it is the exact converse of the Euclidean distance ordination, with the 1981 replicates, and those from 1985 and 1987, condensing into the middle of the ordination plot, whereas the relatively denuded samples from 1983 and, to some extent, 1984 are widely scattered. An important distinction between the two measures is that whilst Euclidean distance tends to be dominated by differences in sample totals (as remarked for Fig. 6c), χ^2 distance removes the effect of differences in total cover altogether by standardising each y_{ij} value by its sample total $\sum_i y_{ij}$. Thus, χ^2 distance reverses the *dependence on totals* guideline; it also fails to fulfill *complementarity* and *localisation*. Another problem, in complete contrast to Euclidean distance, is

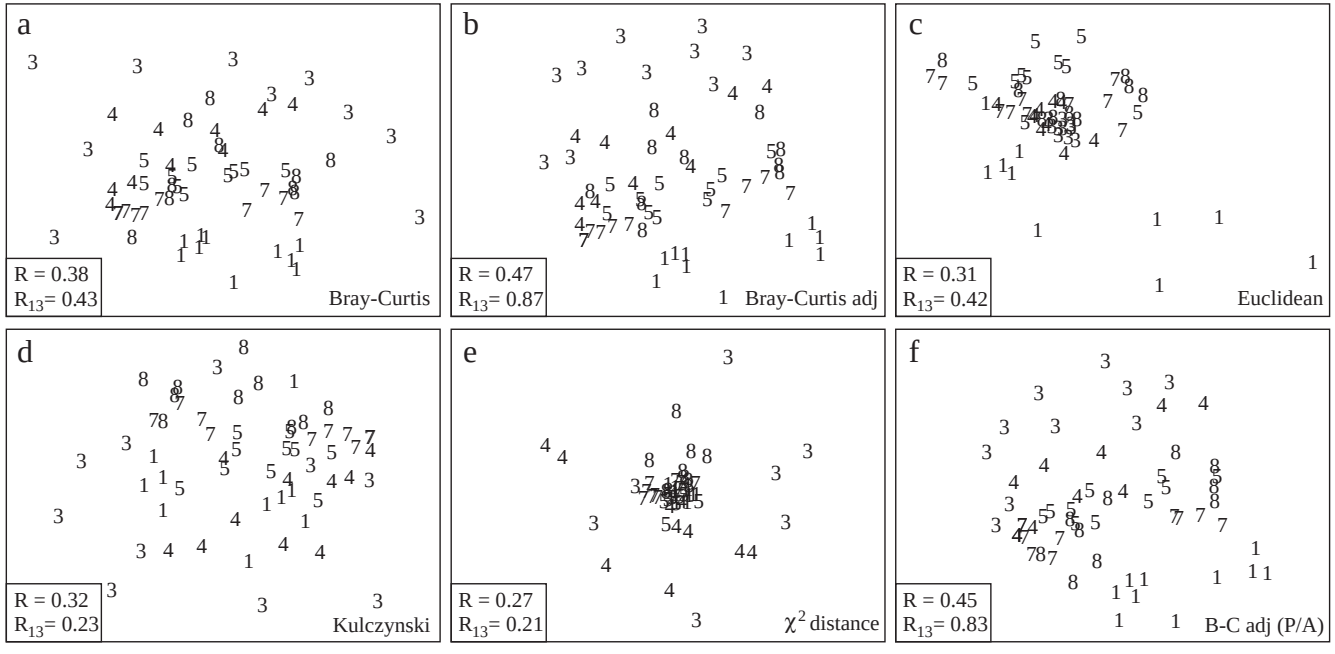

Fig. 6. Coral assemblages from Tikus Island, Thousand Islands, Indonesia. MDS plots for square-root transformed % cover, of 75 species on 10 transects, from each of 6 years (1=1981, 3=1983, 4=1984, 5=1985, 7=1987, 8=1988), with the exception of plot (*f*), which is based only on presence/absence data. Dissimilarity measures used are: a) standard Bray–Curtis, b) its zero-adjusted form, c) Euclidean distance, d) Kulczynski, e) χ^2 distance, f) zero-adjusted Bray–Curtis, based on presence/absence data, namely zero-adjusted Sørensen. *R* denotes the global ANOSIM statistic for a test of no differences between all years; R_{13} is the pairwise ANOSIM *R*, testing only for differences between 1981 and 1983. Stress values: a) 0.18, b) 0.21, c) 0.12, d) 0.20, e) 0.18, f) 0.21.

that small entries can be troublesome. Eq. (7) shows why: there is a denominator term within the outer summation that is the sum over all samples for each species, denoted y_{i+}. Rare species, accounting for very little cover across all samples, will give values close to zero for this denominator, potentially causing instability in the coefficient. Here, the problem is exacerbated by the continuous nature of the percentage cover values, recorded down to 0.1%. It is this instability to rare species and (to a much smaller extent) the enforced sample standardisation that makes the ordination for χ^2 distance totally different, and equally unhelpful, from that based on Euclidean distance. Bray–Curtis on unstandardised data, in contrast, can be thought of as intermediate in its use of sample totals, producing a mix of relative compositional information, in which presence/absence structure plays a definite role (especially under strong transformation), with a modest contribution from the total (transformed) cover values. That the latter is not dominant can be seen by an analysis using (adjusted) Bray–Curtis on sample-standardised data, giving an MDS intermediate between Fig. 6a and b (not shown).

All of the plots in Fig. 6 have high stress values, and the conclusions above about relative visual separations of year groups should be treated with caution, until examined more formally by hypothesis tests. Non-parametric ANOSIM tests are not compromised by the approximations necessary to view a 2-dimensional ordination pattern, since they take place in the full high-dimensional space of the (rank) dissimilarity matrices. The ANOSIM R statistics for the global tests of 'no difference between any years' are displayed in the bottom left corner of the plots in Fig. 6. They match the visual impressions given by the plots, in spite of the high stress values for the latter. Pairwise R values for the test of 1981 v 1983 are also given (all displayed R values are significant at least at the 0.1% level), and make the point even more clearly: the zero-adjusted Bray–Curtis dissimilarity is the most effective by far of the coefficients considered for this data, with $R_{13}=0.87$. (The adjusted Bray–Curtis analysis on standardised, then root-transformed data gives $R=0.37$, $R_{13}=0.56$, but standardisation does not seem biologically appropriate here, since loss of total coral cover could be an important consequence of the El Niño event).

From this example, a final point needs to be made about the Bray–Curtis adjustment. Since the entries are continuous values, not discrete counts, the lowest non-zero entry is not 1, and, as commented in the Methods section, other choices for the dummy species value might be considered. Experience demonstrates, however, that such choices make rather little difference, even here where the data contain several sparse samples. This is illustrated by examining the final plot in the sequence, Fig. 6f, which uses a zero-adjusted Bray–Curtis coefficient based only on the presence/absence of coral taxa (i.e. 1's and 0's in place of % cover). For presence/absence (P/A) data, Bray–Curtis reduces to the Sørensen (1948), or Dice (1945) dissimilarity:

$$D_{12}^{\mathrm{Sor}} = 100 \cdot \frac{b+c}{2a+b+c}, \tag{8}$$

where a is the number of species which are found in both samples 1 and 2; b denotes the number found in sample 1 but not sample 2; and c the number in sample 2 but not sample 1. Adding a dummy species with a value of 1 for all samples is the only natural choice here since it is a P/A matrix, and this gives the zero-adjusted Sørensen dissimilarity:

$$D_{12}^{\mathrm{Soradj}} = 100 \cdot \frac{b+c}{2+(2a+b+c)}. \tag{9}$$

The MDS plot using this coefficient (Fig. 6f) would be exactly the same, whatever quantitative value we had chosen for the dummy species, prior to the P/A transformation. Furthermore, the outcome looks remarkably similar to Fig. 6b, based on the mildly (square-root) transformed quantitative data. The ANOSIM R value, e.g. for the 1981 to 1983 difference, remains large ($R_{13}=0.83$), as is also the case for the other extreme, of no transformation (MDS not shown, $R_{13}=0.88$), demonstrating both the efficacy and relative robustness of a Bray–Curtis analysis in this case, when some form of adjustment is made for near-blank samples.

Example 4. Oilfield macrofauna, Norway

Fortunately, the alarmingly disparate plots of Fig. 6, when radically different dissimilarity measures are employed, are not typical of all data sets. Sometimes conclusions are robust to a wide choice of coefficients. In particular, the effect of the zero-adjustment to Bray–Curtis should be entirely negligible in cases where there are no genuinely impoverished samples. In a commonly cited study of an oilfield impact gradient on soft-sediment macrobenthos (the Ekofisk oilfield, Gray et al., 1990) there are no near-empty samples, the clear impact for the 4 sites within 100 m of drilling operations being expressed as a rise in numbers of certain opportunist species, as well as a decline in numbers of more vulnerable taxa. Samples (along 5 radial transects) are grouped into distance classes, A: >3.5 km; B: 3.5–1 km;

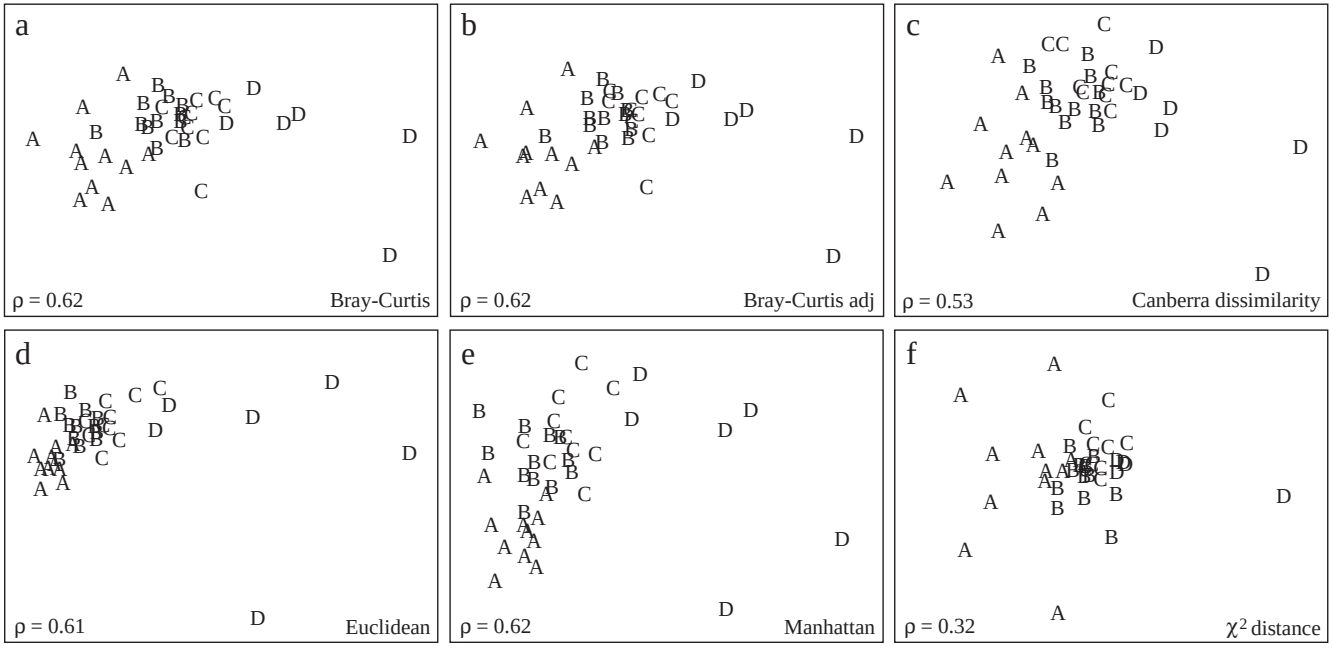

Fig. 7. Macrobenthic assemblages from the Ekofisk oilfield, Norwegian sector of the N Sea. MDS plots for square-root transformed species counts, of 174 species from 39 sites on 5 transects outwards from the oilfield centre. The sites were categorised, a priori, into four distance groups (D: <250 m from the centre, C: 250 m–1 km, B: 1– 3.5 km, A: >3.5 km). The ordinations used coefficients: a) Bray–Curtis; b) zero-adjusted Bray–Curtis; c) Canberra dissimilarity; d) Euclidean distance; e) Manhattan distance; f) χ^2 distance. Also given is the rank correlation, ρ, of the faunal dissimilarities to a model distance matrix reflecting the ordered group structure A to D: $\rho=1$ implies perfect serial change of communities with distance from the centre; $\rho=0$ implies no serial change. Stress values: a) 0.12, b) 0.11, c) 0.18, d) 0.06, e) 0.11, f) 0.18.

C: 1–0.25 km; D:<0.25 km. Fig. 7 shows MDS plots based on the following coefficients: standard Bray–Curtis (Eq. (1)); its zero-adjusted form (Eq. (4)); the Stephenson et al. (1972) form of Canberra dissimilarity, defined as

$$D_{12}^{\mathrm{Can}} = \frac{100}{p_{12}} \sum_i \frac{|y_{i1} - y_{i2}|}{(y_{i1} + y_{i2})}, \tag{10}$$

where joint absences are excluded and p_{12} is the number of species present in either samples 1 or 2 (or both); Euclidean distance (Eq. (6)); Manhattan distance, defined as

$$D_{12}^{\mathrm{Man}} = \sum_i |y_{i1} - y_{i2}|; \tag{11}$$

and χ^2 distance (Eq. (7)). The MDS for Kulczinski (Eq. (5)) is not shown because it is totally indistinguishable from the standard Bray–Curtis plot of Fig. 7a. The Canberra dissimilarity is included here as a further example of a fairly widely used 'biological' coefficient, which obeys all the guidelines of the Introduction and can thus be considered another member of the Bray–Curtis family. Manhattan distance is more akin to the Euclidean distance measure (contravening the same guidelines), though with the potential advantage of being less prone to distortion by outliers since it operates with absolute differences not squared differences. Analyses for all coefficients use square-root transformed species abundances.

In quantifying the success of different coefficients in delineating the impact gradient, the ANOSIM R statistic is less appropriate than in the previous example. It tests only for unordered differences between groups, whereas the optimum model to test here is of serial change: groups A and D are further apart than groups A and C, or B and D, which in turn are further apart than groups A and B, B and C, or C and D. Such serial change is better expressed in a 'seriation with replication' statistic ρ, a Spearman rank correlation between the biotic dissimilarities and a model distance matrix reflecting the ordering of the distance groups away from the oilfield (Somerfield et al., 2002). A permutation procedure is used to test the significance of ρ, essentially a non-parametric Mantel (1967)-type test, and again because of the non-parametric formulation, values of ρ are absolutely comparable between the different coefficient choices. (Note that the choice of ANOSIM R or seriation ρ parallels that in univariate statistics, between 1-way ANOVA and linear regression with replication.)

The MDS ordinations (Fig. 7), and the accompanying ρ statistics, show firstly, and reassuringly, that the dummy species adjustment makes no difference whatsoever to the Bray–Curtis analysis. Secondly, the other coefficients also give rise to a broadly similar picture, with a clear pattern in most cases of assemblage change moving out from the oilfield centre. The Canberra dissimilarity is nearly as successful as Bray–Curtis in distinguishing the distance groups, though less satisfyingly linear (Fig. 7c). Its greater stress can be attributed to the equal weighting given to rarer and common species by the move inside the summation of the denominator term $(y_{i1}+y_{i2})$. Species i values of 0 and 1, for the two samples, now give exactly the same dissimilarity contribution as species i values of 0 and 100, so that rare species which, almost by definition, have a low 'signal-to-noise ratio' can be given too much importance. Although the Euclidean-distance plot (Fig. 7d) has a tendency to be dominated by the larger variation among replicates close to the oilfield centre, it too captures the key distinction between sites as far out as 3.5 km and those beyond — the division between B and A labels. Manhattan distance gives a configuration somewhat intermediate between Euclidean distance and Canberra dissimilarity (Fig. 7e). The only disappointing plot, showing a much less clear delineation of the gradient, is that for χ^2 distance (Fig. 7f; it was also necessary in this case to remove one of the D samples, closest to the oilfield centre, to avoid a collapsed MDS plot).

The reason that most of these coefficients show similar patterns in this example is the short baseline of community change, in terms of turnover of species. The species found at opposite ends of the gradient are not entirely different, and the abundance matrix cannot be made to 'block diagonalise' (i.e. with non-zero entries all close to the diagonal) when samples are ordered by increasing distance from the oilfield. In fact, on a casual visual inspection of the original data matrix, it is hard to see much structure at all. Many species occur across all samples, and the assemblage gradient is characterised as much by changes in density of ubiquitous species as by species deletions and accessions (though both types of change are present). Guidelines such as *complementarity*, *independence of joint absences*, and *dependence on totals* become less biologically important over a short baseline of change.

Example 5. Sludge disposal site macrofauna, Scotland

In complete contrast, the final example considers a long baseline of community change, in another classic impact gradient study. The data are abundances of 84 benthic macrofauna species in 12 samples (labelled 1 to 12) taken along a single E–W transect at Garroch Head, Firth of Clyde, Scotland (Pearson and Blackstock,

1984). Site 6 is at the centre of a sewage-sludge disposal site, and sites 1 and 12 are furthest from the centre. In a meta-analysis by Warwick and Clarke (1993), on studies of sub-tidal soft-sediment macrobenthos subject to disturbance by organic enrichment, this was considered to encompass the widest impact range, and species turnover from one end of the gradient to the other is complete. MDS plots based on 4th-root transformed counts in the 12 samples, for Bray–Curtis and its zero-adjusted form (Fig. 8a, b), are again indistinguishable. Although the central site is heavily organically enriched (and metal contaminated) it is not defaunated, as there are large numbers of a few species of opportunists (e.g. *Capitella capitata*). There is no necessity for a 'near zero' adjustment to the behaviour of Bray–Curtis because no samples are sparse, so the inclusion of the dummy species is simply irrelevant, and is seen not to affect the standard Bray–Curtis plot in any way.

The remaining plots in Fig. 8 make another interesting point. The main feature of Fig. 8a (and b) is the clear progression towards the disposal-site (samples 1 to 6), and the retrogression back to the starting point (samples 7 to 12). The latter is perhaps along a slightly different trajectory but, nonetheless, samples at opposite ends of the transect (1 and 12) have very similar biological communities, and the mid-point sample (6) has a very different community.

There is something of a curve in this gradient, as is often found with assemblages driven by strong single gradients (commonly referred to as the 'arch' or 'horseshoe' effect), but the pattern is unmistakable. Contrast this with the MDS based on Euclidean distance (Fig. 8c). The arch effect now appears so pronounced that the mid-point, sample 6, is located close to the end-point samples, 1 and 12. The stress is relatively low here, so this is not an artefact of representing multivariate relationships in low-dimensional space, it is a genuine reflection of the high similarity that the Euclidean coefficient gives to sites 1 and 6, and sites 6 and 12. This makes no sort of biological sense, as sites 6 and 12 have no species in common at all. What is in common between them is that samples are sparser, both in terms of total numbers and species richness, at the ends of the transect and in the middle of the disposal-site, than at sites adjacent to the disposal centre. This is, arguably, the intermediate disturbance hypothesis (Connell, 1978) in action, and its effect here is to confound an analysis based on Euclidean distance.

The primary reason for Euclidean distance failing to capture the biological reality in this case is that it flouts *complementarity* in a major way. This is best illustrated by what is known as Orloci's paradox (Orloci, 1978). In Table 3, consider the top-left array of counts of three species 1, 2, 3 in three samples A, B, C. Biologically, A must be considered more similar to C than either of them is to B (they have no species in common with B). Yet, Euclidean distance (Eq. (6)) clearly reverses this order, ascribing greater similarity to A and B (distance $\sqrt{75}$) than A and C (distance $\sqrt{200}$). Bray–Curtis, on the other hand, returns the more natural order, biologically speaking, of 100% dissimilarity for A and B (and B and C) but only 50% for A and C. Zero-adjusted Bray–Curtis, in its efforts to define dissimilarity between two all-blank samples, must contravene the *complementarity* guideline, returning 88% for the dissimilarity between A and B, and 95% for B and C. However, A and C still have a dissimilarity of 48%, retaining essentially the same relationships as for standard Bray–Curtis and only departing mildly from *complementarity*.

Inevitably, it is possible to find situations where the adjusted Bray–Curtis measure fails *complementarity* in the same way as Euclidean distance does (two samples with no species in common becoming more similar than two samples that share a common species). The top-right array in Table 3, for example, gives adjusted Bray–Curtis dissimilarities of 50% for A and B but 88% for A and C. However, this can only happen for low counts, as in samples A and B, and it is at this point that the *complementarity* guideline begins to look less appealing. Multiply the counts in the right-hand array by 15, without altering the dummy species value of 1, and the rank order of the adjusted coefficients ($\times 15 D^{\mathrm{B-Cadj}}$ column in Table 3) is restored to that of standard Bray–Curtis. This indicates that the zero-adjusted Bray–Curtis measure must also disobey *relative invariance*; if it did not, for low abundance samples, then its purpose would not be achievable. When samples A and B only contain one individual each, we might be happy to interpret them as from similarly impoverished communities, displaying a natural degree of sampling fluctuation, and thus having greater similarity than A and C. Replace each individual by 15, for all three samples, and we might justifiably take the opposite view: A and B should be less similar than A and C. This is precisely the distinction achieved by the adjusted Bray–Curtis coefficient: denuded samples can be deemed rather similar even when they have no species in common, samples with more data cannot.

The message of this digression, for the real data matrix from the Clyde samples, is that the adjusted Bray–Curtis coefficient is only likely to fall outside the 'envelope' of the Bray–Curtis family, defined by the guidelines of the Introduction, in extreme cases with many denuded samples (such as the boulder-field data

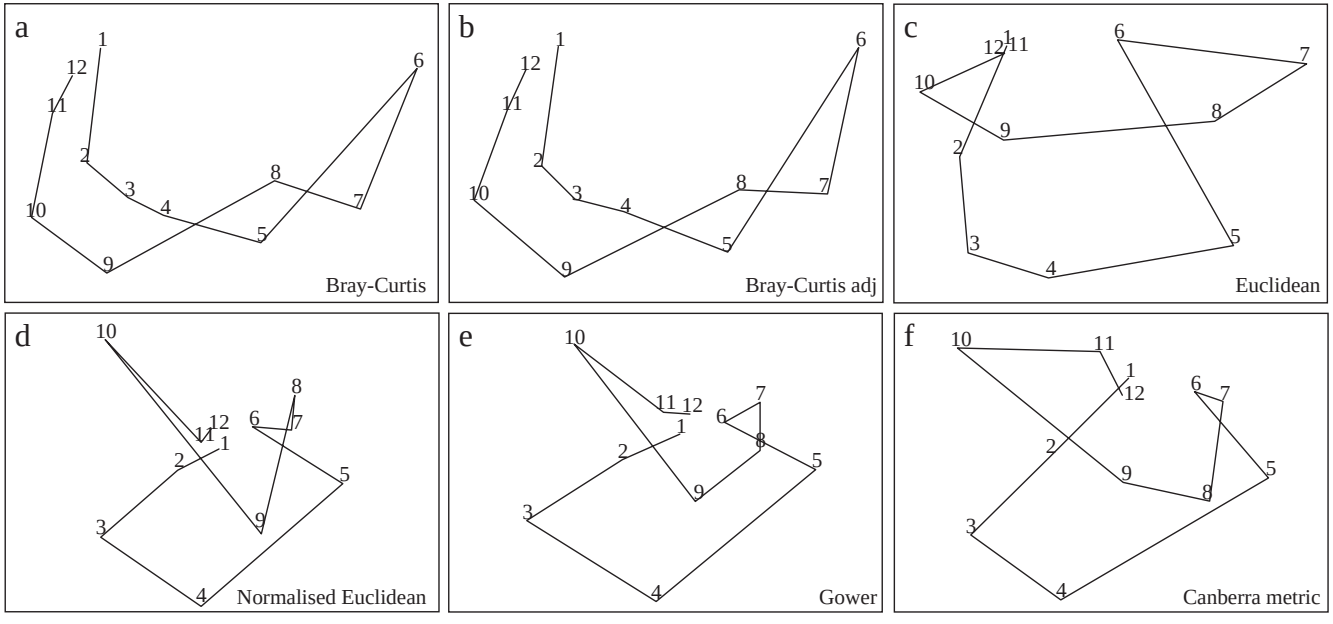

Fig. 8. Macrofauna from the Garroch Head disposal site, Firth of Clyde, Scotland. MDS plots for 4th-root transformed species counts, of 84 species from 12 sites on a single E–W transect (1–12) crossing the centre of the disposal ground (site 6). The plots used: a) Bray–Curtis dissimilarity; b) zero-adjusted Bray–Curtis; c) Euclidean distance; d) normalised Euclidean; e) Gower's coefficient; f) Canberra metric (*not* the Canberra dissimilarity, note). Stress values: a) 0.07, b) 0.07, c) 0.06, d) 0.11, e) 0.1, f) 0.13.

Table 3
Hypothetical data on abundances of 3 species (1, 2, 3) in two sets of 3 samples (A, B, C) with, below, values of Euclidean distance (D^{Euc}), Bray–Curtis dissimilarity ($D^{\mathrm{B-C}}$), and zero-adjusted Bray–Curtis dissimilarity ($D^{\mathrm{B-Cadj}}$) between pairs of samples in each array

	A	B	C		A	B	C
Sp 1	0	5	0	Sp 1	0	1	0
Sp 2	5	0	15	Sp 2	1	0	15
Sp 3	5	0	15	Sp 3	0	0	15
	D^{Euc}	$D^{\mathrm{B-C}}$	$D^{\mathrm{B-Cadj}}$		$D^{\mathrm{B-Cadj}}$	$\times 15D^{\mathrm{B-Cadj}}$	$D^{\mathrm{B-C}}$
A v B	8.7	100	88.2	A v B	50	93.8	100
A v C	14.1	50	47.6	A v C	87.9	93.1	93.5
B v C	21.8	100	94.6	B v C	93.9	99.6	100

Also given for the right-hand set is the adjusted Bray–Curtis dissimilarity when the counts in the array above have been multiplied by a factor of 15 ($\times 15D^{\mathrm{B-Cadj}}$).

of Example 2). Where even modest amounts of data are available for most samples, its behaviour will track other Bray–Curtis type coefficients very closely. Euclidean distance, on the other hand, is always outside this 'envelope', obeying entirely different criteria in which *complementarity* is deemed disadvantageous (e.g. explicitly in Cao et al., 1997; implicitly in Warton and Hudson, 2004). It is, however, undeniable that in the context of a longish baseline of species turnover, the failure to obey *complementarity* leads to the conflict seen in Fig. 8a (or b) and c, and the misleading representation, in the Euclidean distance plot, of the monotonic gradient of assemblage change as the centre of the disposal site is approached.

Euclidean distance is, as previously mentioned, the implicit dissimilarity measure underlying Principal Components Analysis and a PCA of this data looks very like Fig. 8c. The practical difference between a non-metric MDS and a PCA ordination greatly diminishes when both use effectively the same dissimilarity measure, especially (as here) when there are a small number of samples which fit relatively comfortably into low-dimensional space (stress = 0.06 for Fig. 8c). This is covariance-based PCA, so called, because Euclidean distances are calculated on the (transformed) common scale of measurement for all variables (= species), namely the dimensionless scale of counts. More common is correlation-based PCA, in which the variables are first normalised prior to distance calculation (but after transformation, if any). The equivalent MDS ordination would therefore use normalised Euclidean distance as its dissimilarity coefficient (Fig. 8d). Normalisation involves subtracting the mean and dividing by the standard deviation of each species, taken across all samples. This contravenes the *localisation* guideline, since the introduction of further samples would then potentially change the mean and standard deviation used in this normalising step. This is not a minor violation if the additional samples come from an environment where a particular species is much more abundant than elsewhere in the matrix. Fig. 8d, however, demonstrates that the flouting of yet another guideline is a price not worth paying in this instance, as the problems inherent in Fig. 8c are not corrected by normalisation, but are exacerbated. The mid-transect sites 7 and 8 are now drawn close to the transect end-points, 1, 11 and 12, and the pattern of monotonic response to the gradient is almost entirely lost. Normalising is commonplace with a Euclidean distance analysis, but this is usually forced by the variables being on different measurement scales, so there is little alternative to giving them potentially equal weight. After normalisation the units are the common, dimensionless, units of standard deviation about the mean. Normalisation is not required for assemblage matrices because they are generally on a common scale across species (counts, biomass, area cover etc.), and it is usually counter-productive because it gives exactly the same weight to rare as common species and thus increases the ratio of 'noise' to 'signal'.

The final two plots in Fig. 8 show the effects of two alternative ways, sometimes advocated, of giving species potentially equal weight. The simple form of Gower's coefficient (Gower, 1971):

$$D_{12}^{\mathrm{Gow}} = \frac{1}{p}\sum_i \frac{|y_{i1} - y_{i2}|}{R_i},$$

$$R_i = \max_j(y_{ij}) - \min_j(y_{ij}), \tag{12}$$

divides each value y_{ij} by the full range of values for that species across all samples, before calculating absolute distances. In practice, this corresponds to a species standardisation of the matrix by the species maxima (since the minima are typically zero) before application of Manhattan distance. In addition to Gower's

coefficient, another recommendation by Gower and Legendre (1986) for use in ordination is the original form of the Canberra metric (Lance and Williams, 1967):

$$D_{12}^{\mathrm{CanMet}} = \sum_i \frac{|y_{i1} - y_{i2}|}{(y_{i1} + y_{i2})} \tag{13}$$

which differs from Canberra dissimilarity (Eq. (10)), importantly, in respect of the latter's division by the variable factor p_{12}, the number of non-jointly absent species in samples 1 and 2. The standardisation inherent in the Canberra metric is able to satisfy the *localisation* guideline (which Gower does not) but neither coefficient obeys *independence of joint absence* or *complementarity*. As a consequence, they too fail in this case to capture the biological reality of the impact gradient for the Clyde samples (Fig. 8e,f).

4. Discussion

4.1. Previous recommendations on coefficient choice

Several authors have compared sets of resemblance measures, in order to choose those with 'optimal' performance. Mostly, optimality has been judged as success in analysing artificially constructed test data sets. For example, Kulczynski's coefficient (Eq. (5)) was identified by Faith et al. (1987) as the best performer, marginally preferable to Bray–Curtis, in reconstructing a simulated 2-dimensional ecological gradient, based on a model of homoscedastic Gaussian curves for the species counts. Hajdu (1981) constructed 'ordered comparison case series' (OCCAS) which were simple, two species, test cases, involving linear increases or decreases in species responses. Gower and Legendre (1986) used Hadju's OCCAS to demonstrate that the best coefficients for linear ordination were Gower's coefficient (Eq. (12)) and the Canberra metric (Eq. (13)). Another commonly invoked criterion is that of using distance measures which possess metric behaviour, defined as the distance between samples 1 and 3 always being greater than, or equal to, the sum of the distances between samples 1 and 2, and samples 2 and 3 (the 'triangle inequality'). That is, the measure behaves like a genuine distance in some high-dimensional space.

None of the above artificial test cases, criteria or recommendations seem particularly convincing in the practical cases studied here. Firstly, the Kulczynski coefficient is seen to obey the same set of motivating guidelines as Bray–Curtis and very often gives an indistinguishable analysis, the two coefficients being close to monotonicity (i.e. placing dissimilarities in the same order), as Legendre and Legendre (1998) point out. Its functional form (Eq. (5)) shows that it can only return a different order if some samples are highly impoverished relative to others, because only then can there be a substantial difference between the arithmetic and harmonic averages of the two sample totals in the denominators of these coefficients. In Bray–Curtis, which uses an arithmetic mean, there will only be instability if both totals are close to zero. With Kulczynski, the harmonic mean in the denominator will be near zero if either sample total is near zero, which implies that Kulczynski will have a slightly greater tendency to instability than Bray–Curtis. There is some evidence for this in the Tikus corals example (Fig. 6a and d), which does have substantial variation in the degree of sparseness of samples.

Secondly, the recommendation given by Legendre and Legendre (1998), to use either the original forms of the Gower coefficient (Eq. (12)) or the Canberra metric (Eq. (13)) for linear ordination, is seen to be sub-optimal for the practical contexts considered here. Both measures contravene most of the initial guidelines and fail to capture the strongly 'linear' impact gradient with distance from the disposal centre, in the Clyde study. 'Linear' is an over-used word in statistics generally. The 'linear ordination' of Gower and Legendre (1986) refers to a strict linear relationship between the original measurement scale (species counts) and the final ordination. This is an unrealistic model for the effect of the impact gradient here, since this has more to do with patterns of species turnover (presence/absence structure) than linear trends in abundance over a short baseline of change.

Thirdly, the requirement for coefficients to behave as strict metrics is not a guideline that it would be helpful to add to the set in the Introduction. It arises as a requirement in Principal Co-ordinates Analysis, if negative eigenvalues are to be avoided for some of the axes (PCoA, Gower, 1966). This is not a constraint on PCoA/metric scaling in modern practice, however, since several authors have used semi-metrics (such as Bray–Curtis) effectively, not only to display low-d PCoA ordinations but also to construct ANOVA-type generalisations in the mix of real and imaginary high-d space that semi-metric resemblance matrices lead to (see the major contribution by McArdle and Anderson, 2001). A simpler observation is that metric property is clearly irrelevant to the widely employed non-parametric strategy of MDS ordination and ANOSIM permutation tests, because the rank order of dissimilarities (which is all ANOSIM depends on) can be identical for two

resemblance measures, one of which is metric and the other not. An example is Euclidean distance (a metric, naturally) and the square of Euclidean distance:

$$D_{12}^{\mathrm{Euc}^2} = \sum_i (y_{i1} - y_{i2})^2 \tag{14}$$

which is not a metric, but clearly gives resemblances in the same rank order as Euclidean distance itself.

A plethora of resemblance measures have been defined or applied in the ecological literature (Legendre and Legendre, 1998, provide a long, but not exhaustive list), in part because it is easy to invent and motivate new ones. For example, the Kulczynski dissimilarity (Eq. (5)) applies to fully quantitative data but, for the special case of presence/absence (P/A) lists only, it reduces to:

$$D_{12}^{\mathrm{KulPA}} = 100 \cdot \left(1 - \frac{a}{\mathrm{HM}(a+b, a+c)}\right). \tag{15}$$

This uses the notation of Eq. (8), which gives the analogue of the Bray–Curtis coefficient in the P/A case (Sorensen), and the definition of the harmonic mean HM from Eq. (5), the quantitative Kulczynski coefficient. Another P/A coefficient of the same type is due to Ochiai (1957):

$$D_{12}^{\mathrm{Och}} = 100 \cdot \left(1 - \frac{a}{\sqrt{(a+b)(a+c)}}\right), \tag{16}$$

and this will give analyses which are, in some sense, intermediate between those of Sorensen and Kulczynski P/A. This is because it replaces the arithmetic mean of the sample totals, in the Sorensen coefficient, with a geometric mean, whereas Kulczynski (Eq. (15)) uses the harmonic mean in the same place, as previously noted. A geometric mean is intermediate between an arithmetic and a harmonic mean. There does, however, appear to be no reference in the literature to a quantitative form of Ochiai, which is naturally defined as:

$$D_{12}^{\mathrm{B-C(GM)}} = 100 \left(1 - \frac{\sum_i \min\{y_{i1}, y_{i2}\}}{\sqrt{\sum_i y_{i1} \sum_i y_{i2}}}\right) \tag{17}$$

i.e. a 'geometric Bray–Curtis', with a geometric mean divisor replacing the arithmetic mean of the sample totals. This satisfies all the guidelines of the Introduction and is therefore another member of the Bray–Curtis family: its behaviour will be intermediate between that of the quantitative Bray–Curtis and Kulczynski coefficients.

The ease with which new coefficients can be written down highlights the necessity of a disciplined approach to their assessment. We have so far suggested two ways of predicting the likely distinctiveness and behaviour of any newly defined coefficient, namely, referring it to the guidelines of the Introduction, and deducing its performance from the similarity of its functional form to an existing coefficient, whose behaviour in particular contexts is better understood. There is a limit, though, to how far algebraic inspection can be used to predict the likely performance of a coefficient. What is needed here is an effective means of assessing whether any newly defined measure contributes essentially different information than any of the existing canon, and which, if any, of the known coefficients it most closely resembles.

4.2. Comparing coefficients with 2nd-stage MDS plots

A simple means of comparing a number of different coefficients, for any one study, is to derive MDS ordination plots of the among-sample resemblances, as in Figs. 5–8, and to compare the resulting patterns. This, however, is at best a subjective means of comparison, as any ordination is only a low-dimensional (usually 2-dimensional) representation of the true high-dimensional relationships amongst the samples. Faced with the problem of comparing large numbers of ordination plots derived from one set of abundance data, Somerfield and Clarke (1995) introduced the concept of 'second-stage' MDS. This has since been used effectively (e.g. Kendall and Widdicombe, 1999; Olsgard et al., 1997, 1998; Olsgard and Somerfield, 2000) to understand the consequences for the resulting ordinations of changing the prior transformation applied to the raw data matrix, and/or identifying the taxa to coarser taxonomic levels than species (i.e. pooling some variables). It has also been used to identify which faunal groups best reflect the multivariate pattern seen in an environmental variable matrix, defining an impact gradient for example. Here, we use the same idea to inter-relate competing definitions of resemblance.

A second-stage MDS can be thought of as an ordination of ordinations, or to be more precise, an ordination of the relationships amongst the resemblance matrices that underlie each of the (first-stage) ordinations. Here, a number of among-sample resemblance matrices are derived from the same set of abundances, using different resemblance coefficients. A measure of (second-stage) resemblance is then calculated between the different (first-stage) resemblance matrices, which is then used to create the second-stage MDS ordination. The second-stage resemblance measure of choice (Somerfield and Clarke, 1995; Clarke and Warwick, 2001) is simply a Spearman rank correlation ρ, calculated pairwise between corresponding elements in

each of the resemblance matrices. The use of a rank correlation allows for the different (and not necessarily linearly related) measurement scales for differing coefficients, and its application to the elements of the resemblance matrices ensures that the inter-comparisons of coefficients are carried out in full high-dimensional space, not in the low-dimensional approximations of respective ordination plots. A large value of ρ implies that two coefficients are behaving in the same way, a small value implies that multivariate analyses from the two coefficients will differ. A negative value only happens in extreme cases, but implies that analyses differ still further, to the point of being contradictory (for example, the matrices underlying Fig. 6c and e give $\rho = -0.22$). If necessary, it would therefore be appropriate to turn Spearman ρ into a strict similarity measure by a simple location shift to ensure it is always positive (similarity $= 100(1+\rho)/2$ puts it into the range 0 to 100).

Fig. 9 gives second-stage MDS plots for the four main comparative studies presented above, for which the corresponding first-stage MDS plots are in Figs. 5–8. The latter, however, contain ordinations for only a selection of coefficients, whereas second-stage plots allow direct comparison of most of the measures in this paper. Note, though, that the taxonomic dissimilarities of Eqs. (2) and (3) were excluded because taxonomic hierarchies were not available for most of these studies, and it is important for the comparative interpretation of the second-stage plots to use the same set of coefficients throughout. (In fact, experience with other studies bears out the previous suggestion that Γ^+ and Θ^+ will track each other closely unless sample totals very widely, with some approaching zero, and they also demonstrate their genesis from the Bray–Curtis family, being generalisations of the Sørenson and Kulczynski presence/absence measures.) Any all-blank samples have also had to be deleted for Fig. 9, so that the standard coefficients are defined for all samples. The 13 coefficients that are displayed cover the full set of equations defined here, excluding: Euclidean distance squared (Eq. (14)), which (by definition) falls on top of Euclidean distance; Ochiai (Eq. (16)) which falls between Sorensen and Kulczynski P/A; and 'geometric Bray–Curtis' (quantitative Ochiai, Eq. (17)), which falls between Bray–Curtis and quantitative Kulczynski, as predicted. Added to the set is a repeat of Euclidean distance (Eq. (6)) but with a prior normalisation of the data matrix. It should be noted here also that we have not included the popular Jaccard P/A dissimilarity, defined as:

$$D_{12}^{\mathrm{Jac}} = 100 \cdot \frac{b+c}{a+b+c}, \tag{18}$$

because it is strictly monotonic with Sørensen (Eq. (8)), i.e. it places the among-sample dissimilarities in the same rank order, so the point for Jaccard in Fig. 9 would always be coincident with that for Sørensen. Note that Jaccard is another member of the 'Bray–Curtis family' because it is simply the P/A form of Canberra dissimilarity (Eq. (10)).

4.3. Interpreting the 2nd-stage plots for the 4 studies

The second-stage plots of Fig. 9 tell us a great deal about the relationships between the different coefficients, and do so accurately (stress is very low in all cases). Firstly, there is a commonality to the overall pattern, irrespective of the particular study. Thus, Euclidean distance and χ^2 distance tend to be polarised, usually reflecting the greatest difference in analysis outcomes of any two measures. They contravene different combinations of the introductory guidelines, but the key distinction is that whilst Euclidean distance is unstandardised, so that less-abundant species contribute rather little, χ^2 distance is doubly standardised, with abundances first divided by sample totals across species and then the resulting squared differences divided by species totals across samples, so that less-abundant species are upweighted, to some extent. Both of these standardisations have a major impact, as can be seen from the location on the plot of the normalised Euclidean measure, which includes a standardisation (by species standard deviation across samples). This is always well removed in Fig. 9 from the non-normalised Euclidean matrix but also from χ^2 distance. Gower and the Canberra metric standardise in different ways. Gower divides by the range of abundances for each species across all samples. Canberra is more locally standardised, with division by the sum of abundances for that species across only the two samples being compared. Nonetheless, the Gower, Canberra metric and normalised Euclidean coefficients all do have a standardisation or normalisation of species, potentially giving all species equal weight, and this clearly drives a common behaviour.

Manhattan distance is always seen to be intermediate between this group and (non-normalised) Euclidean distance. It shares an absence of standardisation with Euclidean distance but it shares a numerator of absolute differences, rather than squared differences, with Gower and the Canberra metric. The absolute differences in Manhattan should generally provide a greater robustness to outliers than the squared differences in Euclidean distance. The main, common distinction in the plots of Fig. 9, however, is between these distance coefficients

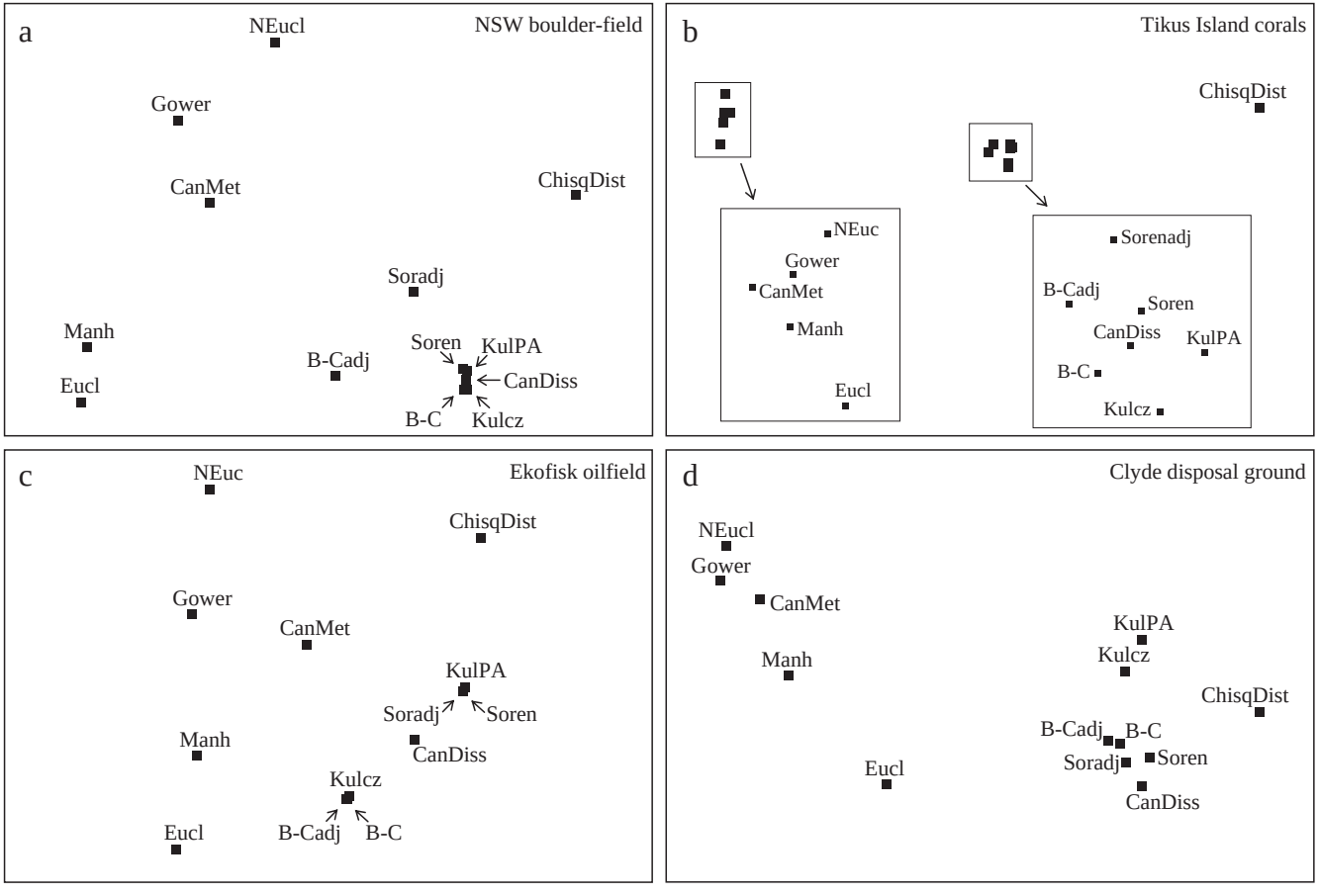

Fig. 9. Second-stage MDS (an MDS of MDS plots, in effect) for the four studies: a) NSW boulder-field macrofauna (as Fig. 5); b) Tikus Island corals (as Fig. 6); c) Ekofisk oilfield macrobenthos (as Fig. 7); d) Clyde disposal-site macrobenthos (as Fig. 8). Plot b) also shows MDS subset plots (re-runs of MDS, not just zoomed graphics) for the two collapsed groupings of coefficients. Each plot is based on Spearman rank correlations, treated as similarities, between the entries of every pair of 13 resemblance matrices, each of the latter computed among all samples in that study (excluding any all-blank samples). The 13 coefficients being compared are B–C: Bray–Curtis dissimilarity (Eq. (1)); B–Cadj: zero-adjusted Bray–Curtis (Eq. (4)); Kulcz: Kulczynski dissimilarity (Eq. (5)); Eucl: Euclidean distance (Eq. (6)); NEucl: Euclidean on normalised data; ChisqDist: χ^2 distance (Eq. (7)); Soren: Sørensen dissimilarity (P/A data, Eq. (8)); Soradj: zero-adjusted Sørensen (P/A data, Eq. (9)); CanDiss: Canberra dissimilarity (Eq. (10)); Manh: Manhattan distance (Eq. (11)); Gower: simple Gower's coefficient (Eq. (12)); CanMet: original form of Canberra metric (Eq. (13)); and KulPA: Kulczynski dissimilarity (P/A data, Eq. (15)). Stress values: a) 0.05, b) 0.01 (0 and 0.01 for subset MDS plots), c) 0.05, d) 0.01.

and the (generally tighter) group of dissimilarities termed the Bray–Curtis family: Bray–Curtis, Kulczynski, Canberra dissimilarity (and quantitative Ochiai) and their P/A forms (Sørensen, Kulczynski P/A, Jaccard and Ochiai, respectively). The key unifying element here is the way these all satisfy the introductory guidelines, in particular the *independence of joint absence* and *complementarity* criteria, at least one of which all other coefficients in Fig. 9 contravene. As previously noted, the zero-adjusted forms of Bray–Curtis and Sørensen also fall within the ambit of the Bray–Curtis family in nearly all cases, since they only mildly contravene the guidelines, and then only for extremely sparse sample sets such as the boulder-field data of Fig. 9a.

Turning to differences in the plots of Fig. 9, rather than their common features, it is clear that switching resemblance measures has greater consequences for some data matrices than others. The least polarised of the four plots is that for the Ekofisk study (Fig. 9c), in which the differing coefficients were seen, in Fig. 7, not to influence the conclusions greatly (with the possible exception of χ^2 distance). The baseline of assemblage change is short here, with effects on the communities being rather subtle, outside the four or five sites within two hundred metres of the oil-drilling activity. Thus, relative to smaller overall changes resulting from the different coefficients, the differences caused by transformation choices (prior to resemblance computation) are accentuated. Quantitative Bray–Curtis, based on mildly transformed data (square root), separates from Sørensen, which is Bray–Curtis on severely transformed data (reduced to presence/absence); 4th-root and log transformations would give points intermediate between these two. This is roughly the position occupied by the Canberra dissimilarity (Eq. (10)), which is seen to fall marginally closer to Sørensen than Bray–Curtis on most of the plots. This could have been anticipated: Canberra dissimilarity makes the important modification to the Canberra metric (Eq. (13)) of removing its dependence on joint absences, and Fig. 9 shows that this radically alters its behaviour. It now satisfies all the guidelines in the Introduction, but the local standardisation based on the sum of the two species abundances (discussed under the Canberra metric above) implies that contributions from common and rarer species are potentially given equal weight. This, of course, is what a Bray–Curtis P/A (Sørensen) analysis does, in a slightly different way. The other feature of note for the Ekofisk plot (Fig. 9c) is that the points for Bray–Curtis, its zero-adjusted form and quantitative Kulczynski are coincident, as they are also (in a shifted location) for the P/A form of all three. For relatively subtle assemblage changes, retaining high counts and many species in all samples, the arithmetic mean/harmonic mean differences in Eqs. (1) and (5), and the added denominator constant in Eq. (4) become irrelevant.

The biggest contrast with the Ekofisk second-stage plot is for the Tikus Island corals data (Fig. 9b). Here, the differences in first-stage ordinations seen in Fig. 6 could not be more dramatic, and the second-stage plot collapses into three tight groups. The subsets from two of these (the third is just χ^2 distance) are re-run through the MDS algorithm, to obtain the sub-plots shown. These form similar patterns to the other studies, but because there are now a number of impoverished samples, with very low total cover, harmonic means of these totals differ from arithmetic ones. Kulczynski thus separates from Bray–Curtis, the difference being of similar magnitude (but orthogonal direction) to that between root-transformed data and a P/A analysis using the same coefficient. A comparable difference results from adding a dummy species to either the quantitative or P/A form of Bray–Curtis (this zero-adjustment was seen, objectively, to provide the best analysis for this data, Fig. 6). In contrast, for the boulder-fauna data of Fig. 9a, where the sparseness of the counts is also extreme, the zero-adjustments make the greatest difference of all, and the different transformations make virtually no difference to the standard forms of Bray–Curtis and Kulczynski (and the latter are indistinguishable, as they were in Fig. 9c). Finally, Fig. 9d, as with 9c, demonstrates the irrelevance of the zero-adjustment of Eqs. (4) and (9) in this case, but the Clyde study is one in which the harmonic mean divisor in the Kulczynski coefficient does produce a slightly different outcome from the arithmetic mean divisor in Bray–Curtis, because the total counts in each sample, though generally large, are highly variable.

4.4. In conclusion

What Fig. 9 tells us about the importance of coefficient selection, in relation to other analysis choices, is enlightening. This paper does not primarily address choice of data transformation because that has been extensively examined using second-stage MDS elsewhere (Somerfield and Clarke, 1995; Olsgard et al., 1997, 1998). However, the changed outcomes to analyses resulting from different transformations prior to computation of two of these coefficients, Bray–Curtis and Kulczynski, can be judged from Fig. 9. In Fig. 9a, the two extremes of no transformation and reduction to presence/absence (P/A) were barely distinguishable. In

Fig. 9b and c, the transformation ranged nearly as widely, from square root to P/A, and whilst not of negligible consequence (especially in the Ekofisk study), the effect is demonstrably smaller than that of a change in the coefficient itself. Note that Somerfield and Clarke (1995) and Olsgard et al. (1997, 1998) concluded that different transformation choice generally had a much greater effect on the outcome of a multivariate community analysis than decisions on taxonomic level (whether the taxa involved should be identified to species or only to genus level, family level, etc.). The inescapable consequence is therefore that the primary attention should be directed to identifying a coefficient with the correct properties for the proposed study, secondarily to a relevant transformation, and only then to an appropriate level of taxonomic identification. This is probably in precisely the reverse order to most current biological practice!

The holistic view provided by a second-stage MDS is clearly very powerful in this context and one can envisage it being used to classify, by second-stage cluster analysis perhaps, an even wider set of resemblance measures, thus bringing some order to the anarchy of choosing from the overwhelming number of coefficients defined in the literature. An alternative to clustering would be a more continuous approach, in which a presence/absence-type matrix of guidelines obeyed by each measure could be treated as an associated 'environmental' matrix, by analogy with standard community analysis structures of sample × sample resemblances (based on species variables) and matching environment × sample arrays. Here the 'samples' are coefficients, and the 'environmental variables' are guidelines—a much wider set than were listed in the Introduction to this paper of course. The methodology then exists (e.g. the BIO-ENV procedure of Clarke and Ainsworth, 1993) to determine which guidelines are the important ones in determining different behaviour of coefficients in realistic analyses.

A second-stage MDS, however, like all ordination procedures, is only comparative. It has the capacity to tell us which coefficients are essentially behaving in the same way, and which are capable of taking a different view of the data; it does not tell us which of these different behaviours is helpful or unhelpful. The latter is provided by the hypothesis-based structure of specific datasets. Which coefficients best separate the pre- and post-El Niño years in Fig. 6 is determined by the largest values of the ANOSIM R statistic in testing the coral community differences for those years. Which coefficients, in Fig. 7, best delineate the ordered gradient of assemblage change approaching the Ekofisk oilfield centre, is determined by the largest values of the Spearman ρ statistic, rank correlating the community resemblances with a triangular model matrix corresponding to that ordered change. For the Clyde sewage-sludge disposal study of Fig. 8, continuous environmental data is available, characterising the impact on the macrobenthic community from organic enrichment and heavy metals, so the rank correlation ρ between community-based and environment-based resemblances among samples (Clarke and Ainsworth, 1993) can be optimised, by appropriate choice of coefficient. (Coefficients from the Bray–Curtis family again do the best job of matching the assemblage patterns to the strong, monotonic contaminant gradient approaching the disposal centre.) These three examples span the range of hypotheses that might be tested, covering categorical, ordered categorical and continuous change. Together with the second-stage procedure, they provide the methodology for a more robust and broadly based statement on the efficacy of competing resemblance measures in different ecological contexts. It goes without saying, of course, that such techniques should be used to draw out general rules about which contexts are best served by which measures, rather than adopting a 'try it and see' approach, for each new data set, of running through 50+ choices of resemblance measure for the one that best substantiates a desired outcome!

Finally, to return to the starting point of the paper, Bray–Curtis and related coefficients have a favoured position in ecological studies largely through 'received wisdom' that they outperform most other measures with real assemblage data in important practical applications. The data sets of this paper have provided further evidence for this. Though clearly not a universal panacea, the Bray–Curtis coefficient, in its original form or in one of the variants discussed in this paper – modified to cater for denuded samples, or to incorporate higher taxonomic level relationships – has provided, in these cases, simpler, more interpretable and statistically significant analyses, in objective tests, than a wide range of alternatives. To end on a cautionary note, however: we have demonstrated that the 'dummy species' adjustment to Bray–Curtis can improve interpretability of displays and significance of tests, and where there are no impoverished samples the modification is irrelevant. Nonetheless, it would still be unwise to apply this zero-adjustment unthinkingly, as a cure for all sparse data problems. It contains an implicit assumption that two samples which are sparse are similar, because they are denuded for the *same* reason. This will not always be the case. Where samples correspond to replicates at too small a physical scale to be sure of capturing some

organisms in each (perhaps because the species are all heavily clustered in space), then blank replicates could potentially be found in any or all of the treatments/times/sites being compared—this is arguably the situation for the boulder fauna of Fig. 5. It is probably not then biologically meaningful to define such replicates, in unrelated groups, as being 100% similar to each other. Practically speaking also, it would tend to reduce the global ANOSIM *R* statistic when testing these groups. On the other hand, it seems equally undesirable to omit the blank samples altogether, on the grounds that they contain no information: a treatment with several such replicates could reasonably be considered different than one which leads to substantially fewer blank samples. A better solution here is to reduce the replicate variability by taking large enough replicate 'extents' (core diameter, tow length, quadrat size etc.) to ensure that, for unimpacted conditions, a moderate number of organisms are always recorded. Impoverished samples are now an indication of the common impact gradient, not sampling inadequacies, and can reasonably be regarded as highly similar. This raises interesting issues of replication numbers and power in assemblage tests based on resemblance matrices. Given a large, fixed number of independent replicates from each of two groups (a control v impact study, say), and sufficiently large among-replicate variability to generate blank samples in both groups, it seems clearly advantageous to pool the replicates arbitrarily in 2's, 3's, 5's etc. before applying the 1-way ANOSIM test. Heavier pooling will give a larger ANOSIM *R* statistic, though with a smaller number of replicate pools (and thus permutations) to test its difference from zero, the hypothesis of no impact effect. If such a trade-off *can* be beneficial, that would make an interesting contrast with the univariate analogue (the 2-sample *t*-test), in which such pooling is not advantageous and not carried out, in general. To our knowledge, this problem has not been examined, though there is clearly scope here to derive useful guidelines for practitioners, on the desirability and extent of pooling of raw data prior to multivariate display and testing.

Acknowledgements

This work is a contribution to the biodiversity element of the Plymouth Marine Laboratory's core strategic research programme. It was supported by: the UK Natural Environment Research Council (NERC); the UK Department for Environment, Food and Rural Affairs (DEFRA) through projects AE1137, CDEP 84/5/295 and ME3109; the Australian Research Council through its Special Research Centres Programme; and the University of Sydney. We are indebted to Ray Gorley for his work on the PRIMER v6 code. This work was instigated whilst KRC held a visiting professorship at the Centre for Research on Ecological Impacts of Coastal Cities, University of Sydney, and was completed under his current position as an honorary fellow of the Plymouth Marine Laboratory. **[SS]**

References

Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecol. Monogr. 27, 325–349.

Brown, A.R., Shillabeer, N., 1997. Development of a biologically based environmental quality standard from a long-term benthic monitoring programme in the North Sea. Oceanol. Acta 20, 275–282.

Cao, Y., Bark, A.W., Williams, W.P., 1997. A comparison of clustering methods for river benthic communities. Hydrobiologia 347, 25–40.

Chapman, M.G., 2002. Patterns of spatial and temporal variation of macrofauna under boulders in a sheltered boulder field. Aust. Ecol. 27, 211–228.

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

Clarke, K.R., Ainsworth, M., 1993. A method of linking multivariate community structure to environmental variables. Mar. Ecol. Prog. Ser. 92, 205–219.

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

Clarke, K.R., Green, R.H., 1988. Statistical design and analysis for a 'biological effects' study. Mar. Ecol. Prog. Ser. 46, 213–226.

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

Clarke, K.R., Warwick, R.M., 1998b. A taxonomic distinctness index and its statistical properties. J. Appl. Ecol. 35, 523–531.

Clarke, K.R., Warwick, R.M., 2001. Change in Marine Communities: An Approach to Statistical Analysis and Interpretation, 2nd edn. PRIMER-E, Plymouth.

Connell, J.H., 1978. Diversity in tropical rain forests and coral reefs. Science 199, 1302–1310.

Craig, N.C.D., Shillabeer, N., Parr, J.C.W., 1993. Long-term benthic monitoring of a dense effluent discharge to sea. J. Inst. Wat. Environ. Manag. 7, 262–269.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57–68.

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.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338.

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27, 857–871.

Gower, J.C., Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classif. 3, 5–48.

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.

Hajdu, L.J., 1981. Geographical comparison of resemblance measures in phytosociology. Vegetatio 48, 47–59.

Izsak, C., Price, A.R.G., 2001. Measuring β-diversity using a taxonomic similarity index, and its relation to spatial scale. Mar. Ecol. Prog. Ser. 215, 69–77.

Kendall, M.A., Widdicombe, S., 1999. Small scale patterns in the structure of macrofaunal assemblages of shallow soft sediments. J. Exp. Mar. Biol. Ecol. 237, 127–140.

Krzanowski, W.J., Marriott, F.H.C., 1994. Multivariate Analysis. Part I: Distributions, Ordination and Inference. Edward Arnold, London.

Kulczynski, S., 1928. Die Pflanzenassoziationen der Pieninen. Bull. Int. Acad. Pol. Sci. Lett. Cl. Sci. Math. Nat. Ser. B Suppl. II, 57–203.

Lance, G.N., Williams, W.T., 1967. A general theory of classificatory sorting strategies: 1 Hierarchical systems. Comp. J. 9, 373–380.

Legendre, P., Legendre, L., 1998. Numerical Ecology, 2nd Engl edn. Elsevier, Amsterdam.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82, 290–297.

Ochiai, A., 1957. Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. Bull. Jpn. Soc. Sci. Fish 22, 526–530.

Odum, E.P., 1950. Bird populations of the Highlands (North Carolina) plateau in relation to plant succession and avian invasion. Ecology 31, 587–605.

Olsgard, F., Somerfield, P.J., 2000. Surrogates in marine benthic investigations—which taxonomic unit to target? J. Aquat. Ecosyst. Stress Recovery 7, 25–42.

Olsgard, F., Somerfield, P.J., Carr, M.R., 1997. Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. Mar. Ecol. Prog. Ser. 149, 173–181.

Olsgard, F., Somerfield, P.J., Carr, M.R., 1998. Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. Mar. Ecol. Prog. Ser. 172, 25–36.

Orloci, L., 1978. Multivariate Analysis in Vegetation Research, 2nd ed. Junk BV, The Hague.

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

Somerfield, P.J., Clarke, K.R., 1995. Taxonomic levels, in marine community studies, revisited. Mar. Ecol. Prog. Ser. 127, 113–119.

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

Sørensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. Biol. Skr. 5, 1–34.

Stephenson, W., Williams, W.T., Cook, S.D., 1972. Computer analyses of Petersen's original data on bottom communities. Ecol. Monogr. 42, 387–415.

Ter Braak, C.F.J., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167–1179.

Warton, D.I., Hudson, H.M., 2004. A MANOVA statistic is just as powerful as distance-based statistics, for multivariate abundances. Ecology 85, 858–874.

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., Clarke, K.R., 1995. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Mar. Ecol. Prog. Ser. 129, 301–305.

Warwick, R.M., Clarke, K.R., Suharsono, 1990. A statistical analysis of coral community responses to the 1982–3 El Niño in the Thousand Islands, Indonesia. Coral Reefs 8, 171–179.