Faith, Minchin & Belbin 1987 Vegetatio 69:57-68

dissimilarity problem in ordination': what dissimilarity measure has a robust, informative, relationship with ecological distance; and what ordination method can take advantage of this relationship in its assumptions?

Methods

Dissimilarity coefficients evaluated

The dissimilarity measures evaluated in this study are listed in Table 1. Examination of two basic measures, Manhattan distance (MAN) and the complement of Kendall's coefficient (KEN), illustrate a problem which prompted consideration of many of the alternative measures in Table 1. MAN is defined by the sum of the absolute differences in abundance over all species. Such an index might be expected to reflect ecological distance, but a problem arises for larger ecological distances. When two sites are sufficiently far apart in ecological space that they share no species, MAN yields a value that depends only on the total site abundances. Values of MAN can therefore suggest that one ecological distance is larger than another when the reverse is true. KEN initially appears to avoid spurious variation due simply to site totals, in that it reaches a constant maximum value when two sites share no species. However, for small ecological distances, when sites will tend to share many species, the actual value of KEN does reflect variation in site totals. Variation in site totals thus obscures predictive information about ecological distance for both simple measures.

A desirable measure would take a value of zero when ecological distance was zero, and some constant maximum value when ecological distance had increased to the point that shared abundance was zero (Beals, 1984). This theoretical argument has led us to focus on a number of measures that all have some form of standardization, such that the above properties are satisfied. These measures (Table 1) are the Kulczynski (QSK), Bray-Curtis (B-C), Chord distance (CHD), Canberra metric (CAN) and Relativized Manhattan (MAN:SAT) measures.

While the measures listed above are well constrained for large

Table 1. Basic measures with reference, abbreviation, standardizations used, and formula for dissimilarity between two objects, j and k, based upon attributes, $i = 1$ to N. Z is the number of attributes that are 0 for j and k. MAX_i is the maximum value of attribute i over all sites; MIN_i is the corresponding minimum. SPM is species adjusted to equal maximum abundance. SAT is sites standardized to equal totals. DBL is SPM followed by SAT. Equivalences of measure-standardization combinations reduced the total number of combinations to 29. For further explanation see text.

Name and reference	Abbreviation	Standardizations	Formula
Kendall (1970)	KEN	SPM	$\sum_i [MAX_i - \text{minimum}\ (X_{ij}, X_{ik})]$
Manhattan (Sokal & Michener, 1957)	MAN	SAT, SPM, DBL	$\sum_i \lvert X_{ij} - X_{ik}\rvert$
Gower metric (Gower, 1971)	GOW	SAT	$\sum_i [\lvert X_{ij} - X_{ik}\rvert / (MAX_i - MIN_i)]$
Euclidean (Sokal & Sneath, 1963)	EUC	SPS, SPM, SAT, DBL	$[\sum_i (X_{ij} - X_{ik})^2]^{1/2}$
Intermediate (Faith, 1984)	INT	SPM	$(1/2)\ [\sum_i \lvert X_{ij} - X_{ik}\rvert + MAX_i - \text{minimum}\ (X_{ij}, X_{ik})]$
Quantitative symmetric (Kulczynski) – See for instance, Hajdu (1981)	QSK	SPM	$1 - (1/2)\{[\sum_i \text{minimum}\ (X_{ij}, X_{ik})/\sum_i (X_{ij})] + [\sum_i \text{minimum}\ (X_{ij}, X_{ik})/\sum_i (X_{ik})]\}$
Bray-Curtis (Bray & Curtis, 1957)	B-C	SPM	$(\sum_i \lvert X_{ij} - X_{ik}\rvert)/[\sum_i (X_{ij} + X_{ik})]$
Chord (sensu Orlóci, 1967)	CHD	SPM	$[2\ (1 - \{(\sum_i X_{ij} X_{ik})/[(\sum_i X_{ij}^2)\ (\sum_i X_{ik}^2)]^{1/2}\})]^{1/2}$
Canberra metric, Adkins form (Lance & Williams, 1967)	CAN	SPM, SAT, DBL	$[1/(N-Z)] \sum_i \lvert X_{ij} - X_{ik}\rvert/(X_{ij} + X_{ik})$
Chi-squared (Chardy *et al.*, 1976)	CSQ	SPM, SAT, DBL	$\{\sum_i (1/\sum_i X_{il})[X_{ij})[\sum_i X_{ij}) - X_{ik}/(\sum_i X_{ik})]^2\}^{1/2}$

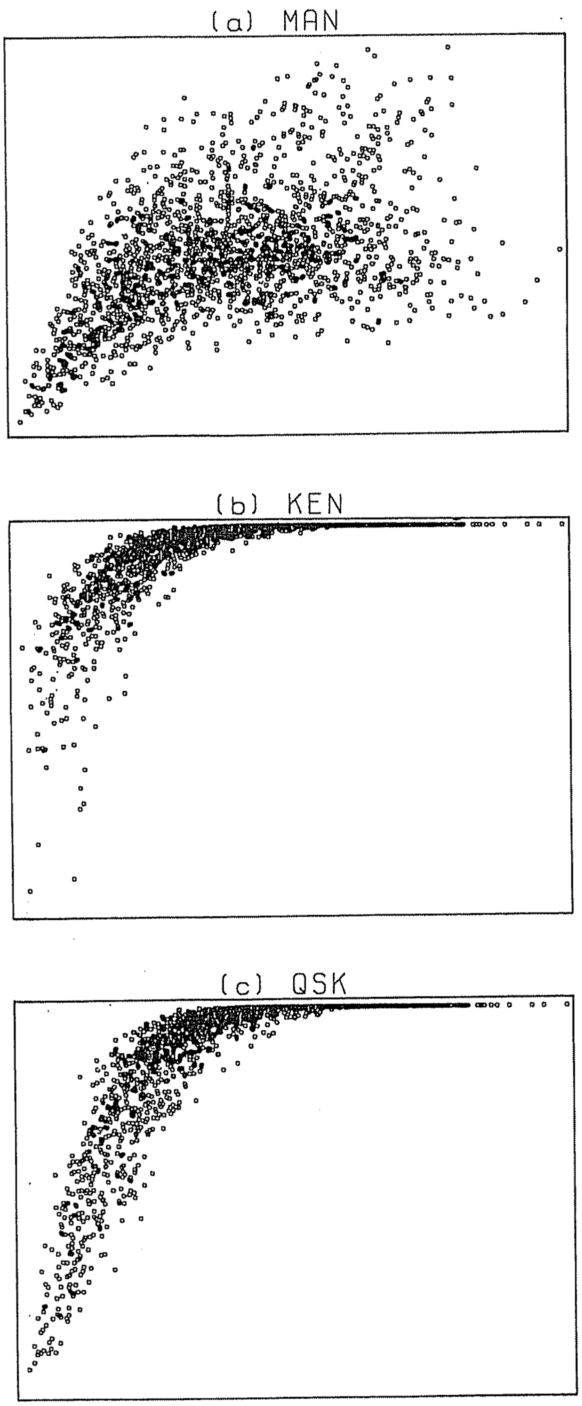

Fig. 1. Relationship between compositional dissimilarity value (vertical axis) and 'target' ecological distance (horizontal axis) for three measures: (a) MAN; (b) KEN; and (c) QSK. Scales of axes are arbitrary. Each circle represents the dissimilarity value – distance value combination for a single pair of sites, for one simulated model. For further explanation see text.

Table 4. Dissimilarity measures ranked in order of (a) mean rank correlation with ecological distance over all 306 two-dimensional models and (b) mean linear correlation with ecological distance over the 102 two-dimensional models in which the beta diversity of the longest gradient did not exceed 0.5R. The abbreviations for the dissimilarity measures are explained in Table 1.

(a) Mean rank correlation			(b) Mean linear correlation		
1	QSK:SPM	0.8925	1	QSK:SPM	0.8722
2	B-C:SPM	0.8924	2	MAN:DBL	0.8684
3	MAN:DBL	0.8923	3	B-C:SPM	0.8682
4	CHD:SPM	0.8861	4	GOW:DBL	0.8491
5	CAN:SAT	0.8812	5	QSK	0.8325
6	CAN:DBL	0.8794	6	MAN:SAT	0.8321
7	CAN	0.8791	7	INT:SPM	0.8308
8	KEN:SPM	0.8637	8	CAN:SAT	0.8304
9	MAN:SAT	0.8551	9	CHD:SPM	0.8272
10	QSK	0.8550	10	CAN	0.8265
11	B-C	0.8548	11	CAN:DBL	0.8262
12	CHD	0.8158	12	B-C	0.8243
13	KEN	0.8157	13	GOW:SAT	0.8025
14	GOW:DBL	0.7487	14	KEN:SPM	0.7890
15	INT:SPM	0.7313	15	INT	0.7709
16	GOW:SAT	0.6917	16	EUC:DBL	0.7695
17	GOW	0.6523	17	CHD	0.7611
18	MAN:SPM	0.6522	18	GOW	0.7565
19	INT	0.6423	19	MAN:SPM	0.7564
20	CSQ:SAT	0.6325	20	CSQ:SAT	0.7539
21	EUC:SPM	0.6136	21	EUC:SAT	0.7442
22	CSQ	0.6087	22	KEN	0.7207
23	EUC:DBL	0.6046	23	EUC:SPM	0.7194
24	EUC:SAT	0.6022	24	CSQ	0.7099
25	CSQ:DBL	0.5926	25	MAN	0.6874
26	MAN	0.5617	26	CSQ:SPM	0.6636
27	CSQ:SPM	0.5605	27	CSQ:DBL	0.6625
28	EUC:SPS	0.5471	28	EUC:SPS	0.6424
29	EUC	0.4657	29	EUC	0.6079

lations (Table 4b) included all forms of EUC and CSQ, together with MAN, MAN:SPM, KEN, GOW, CHD and INT. Once again, the best measures included some type of standardization by species. Of the measures without species standardization, QSK and MAN:SAT had the highest mean linear correlations. Next best was B-C.

The relative performance of those measures which had the highest mean rank and linear correlations was examined in more detail using ANOVA. As an example, a summary of the analysis of the *difference* in rank correlation between QSK and CHD for models with symmetric and skewed response shapes is given in Table 5. In this case, the analysis indicates an interaction between beta