Conclusions on comparing resemblance coefficients

Clarke KR, Somerfield PJ, Chapman MG 2006, *J Exp Mar Biol Ecol* 330: 55-80 discuss this analysis (and that for several other data sets) in more detail, but to pick out just four general points:

a) These 2nd stage plots have common features, irrespective of the actual data set, e.g. coefficients which are in what they term as the 'Bray-Curtis family' (including quantitative measures: S17, S18 & Ochiai (quant), matched by pres/abs measures: S8, S13, S14; also Canberra similarity exc 0-0) tend always to cluster on the 2nd stage plot, i.e. produce similar multivariate conclusions, and radically differ from Euclidean distance, even more so when the latter is normalised.

b) Choice of coefficient is much more crucial to a multivariate analysis than transformation (which itself is more important than taxonomic level – see earlier); this is apparent here by noting the relative proximity of the Bray-Curtis and Bray-Curtis P/A (Sorensen) points, and the Kulczynski and Kulczynski P/A points, on the 2nd stage plot (the first of the pair uses a mild square root, and the second is on presence/absence data – the most severe transform possible).

c) The inference of similarity from joint absences for coefficients such as Euclidean distance, S15 Gower etc., has a dramatically adverse effect on their performance in describing gradients of assemblage change where there is a turnover of species (i.e. pres/abs data is informative); this is clear from the above (1st stage) MDS based on Euclidean distance, which places site 6, at the centre of the dumpground, close to the extreme ends of the transect, 1 and 12, when 6 has no species in common with either! Similarity is deemed higher because they share absent species. The radical effect of counting (or not) joint absences is also clear here from: the separation of the Canberra metric from Canberra similarity (the only difference is an adjustment for double zeros, Section 5), and the way the plots splits left, right (counts 0-0, ignores 0-0), with the Faith coefficient intermediate since it counts joint absences, but with less weight than joint presences.

d) Another key feature which separates out the behaviour of coefficients is whether they implicitly or explicitly standardise (or normalise), and whether over samples or species. Chi-squared distance does both, removing all differences in total abundance between samples and also having a divisor of the total abundance of each species across all samples – low density species can be given very heavy weight, leading to problematic behaviour. Normalised Euclidean and Gower also have a species (but not sample) standardisation, giving rare and common species equal weight.

Close the workspace – we shall start a clear workspace next time we meet this data (Section 15).

Revision #4
Created 22 October 2024 19:56:20 by Arden
Updated 25 February 2025 22:17:23 by Abby Miller