Species sets ‘explaining’ the overall pattern

The main application area for the BVStep routine introduced by Clarke KR & Warwick RM 1998, *Oecologia* 113: 278-289, is what might be termed *Bio-Bio*, namely searching for subsets of species whose resemblance matrix best matches that of another (fixed) set of species. One can envisage this used on different faunal (taxonomic- or trophic-based) groups to elucidate potential interactions but the most obvious context is when the two biological matrices are from the same data. That is, the input similarity matrix is computed from the full set of species, and the secondary data sheet from which species are selected is the same full species data. Now, the idea is not to maximise ρ, since it can always be made equal to 1 by choosing a subset which is the full set of species, but to find the smallest possible subset of species which, in combination, describe most of the pattern in the full data set. ‘Most’ in this context is taken to be a conventional, and somewhat arbitrary, ρ>0.95. Once ρ gets to about this level, two multivariate patterns (e.g. as seen in 2-d ordinations) are effectively indistinguishable, and would not lead to different interpretations.

The procedure can be thought of as a generalisation of the SIMPER approach (Section 10) to the case of continuous multivariate patterns, rather than a clearly-defined clustering of samples. For example, in the Morlaix MDS of the time series of 21 samples, seen earlier in this section, SIMPER could perhaps be run on three groups of times – before and immediately after the oil-spill, and the partial recovery phase, to identify all species contributing to the dissimilarity between each pair of those groups. The BVStep procedure, however, asks a subtly different question, namely, is there a subset of species which between them account for the whole continuous pattern: the structure of initial seasonal cycle, a period of marked change following the oil-spill, then a gradual recovery with the re-establishment of the seasonal cycle? Not only does this provide a more holistic answer than SIMPER (and, importantly, one that can be applied whatever the chosen resemblance matrix), it is also more parsimonious in identifying indicator species: if several species are contributing to the pattern in exactly the same way, BVStep will only need to select one of them, but SIMPER will identify all as contributing something to the average between-group dissimilarity. A next question is then to ask whether the identified set of species is the only subset which is capable of accounting for this multivariate impact, recovery and seasonal pattern (i.e. would constitute a good set of indicators for this time series). In other words, is the same pattern reinforced in the matrix over several sets of species? – what might be termed *structural redundancy*.

Revision #3
Created 26 October 2024 03:55:16 by Arden
Updated 25 February 2025 23:04:29 by Abby Miller