Expanding an (abiotic) data matrix

A RELATE test could equally well have been carried out between the Ekofisk community pattern and a matching (abiotic) resemblance matrix computed not from the surrogate for increased impact – the nearness of the sites to the oil-field centre – but from a set of contaminant levels themselves, as measured at each site. (For the tests of this section, we assume that this set is fixed – we are not allowing selection of a subset of contaminant variables which appears to best match the observed community pattern, i.e. the BEST(Bio-Env) procedure of the previous section. RELATE tests do not allow for this selection bias). All that is necessary for a simple RELATE test of community to a fixed environmental variable suite is that we have one-to-one matching of the abiotic data to each community sample. The active sheet for **Analyse>RELATE** would logically be the biological resemblance coefficient and the (Secondary data•Resemblance/model matrix) would typically be Euclidean distance on a selectively transformed then normalised abiotic data matrix (though the test would be the same if the matrices were the same size and entered in the opposite order). But where the community data consists, for example, of replicate samples at a number of sites, and the abiotic matrix consists of a single value for each of the suite of variables (which may itself be an average over replicate abiotic measurements, but not matched to the community replicates) then the abiotic matrix needs to be expanded to the same dimensions as the biological matrix, and its entries repeated appropriately. This is achieved by the **Tools>Expand Samples** routine operating on the active matrix of the abiotic data. It is not cheating – at least, not necessarily! It depends on what is then done with the expanded matrix. If we pretend that the repeated readings are independently measured – by running an ANOSIM test on them for example – then of course we are heading for trouble. But in this context the requirement is an expansion of the *Seriation with replication* test of the previous page – we want to test the null hypothesis that there are no differences among sites against the specific alternative that there are such differences and that they are determined by the environmental structure among sites (in statistical parlance we *condition* on this, so the situation becomes no different than if we were testing against a design structure, e.g. seriation or treatment levels). So, the test is no longer of seriation with replication but of a more complex environmental relationship among the sites, but it has the same characterising feature that the resulting RELATE ρ value will capture both whether the sites differ at all and whether they do so in a way that matches the (multivariate) abiotic relationships among sites. A high ρ can only be obtained if both are true. An alternative would be to average up the community replicates to the site level and carry out a simple RELATE test to the abiotic data at that level. However, this might have very little power if there are few sites and it misses the important comparison of whether ρ for this specific alternative is greater than ρ for the unordered test (the 1-way ANOSIM-type model matrix of 0’s and 1’s).

Revision #1
Created 15 October 2024 21:24:46 by Arden
Updated 15 October 2024 21:33:19 by Arden