

# Linkage trees – rationale

Another technique for linking sample patterns based on assemblage data to a suite of environmental (or other) explanatory variables was also discussed in Clarke KR *et al* 2008 *J Exp Mar Biol Ecol* 366: 56-69 (see the last topic in Chapter 11, CiMC). The well-established statistical procedure of *Classification And Regression Trees* (CART) was further developed in an ecological context by De'ath G 2002, *Ecology* 83: 1105-1117, termed *Multivariate Regression Trees* (MRT). PRIMER implements a modification of this, in a form which is consistent with the non-metric philosophy underlying the rest of the package. The connection with regression is minimal (and confusing) so the more descriptive term *linkage trees* is used by PRIMER for its variation of the procedure. Its real affinity is with Cluster analysis (Section 6, under heading **Binary divisive clustering**), and it is therefore accessed in PRIMER v7 by running **Analyse>Cluster>LINKTREE**. In fact, it is a form of *constrained* binary divisive clustering in which the successive divisions of the full set of biotic samples, seen in the *unconstrained* divisive clustering of

**Analyse>Cluster>UNCTREE** (Section 6), are limited to those splits of each group (into two new sub-groups) which have an explanation in terms of larger or smaller values of a specific explanatory (typically abiotic) variable – consistently so on either side of that divide. In other words, all constraints are a threshold inequality on a single abiotic variable and this set of inequalities form the possible 'explanation' for the biotic structure.

We have already seen two techniques for linking assemblage patterns to abiotic variables: bubble plots (Section 8) and the above BEST procedure. BEST has the advantage of looking at the abiotic variables in combination, trying to identify a subset which is sufficient to 'explain' all the biotic structure capable of explanation, and the matching procedure takes place in the full high-d space, i.e. on the respective resemblance matrices. But on its own, this falls short of a full interpretation because it does not demonstrate which variables take high or low values for which samples. Bubble plots give the latter but are only satisfactory where the low-d biotic *n*MDS has acceptable stress as an approximation to the full biotic pattern. *Linkage trees* can fill this gap: they can take the subset of abiotic variables identified by BEST, and use them to describe how the assemblage samples are optimally split into groups (in the high-d space), and interpret this, e.g. Group 1 communities have Salinity<23ppt but Group 2 are from >26ppt (with no samples between these salinity thresholds). Group 1 and 2 samples are then each divided into two by a different threshold on the same abiotic variable, or more likely by a different abiotic variable. The result is divisive clustering of the biotic samples, and an environmental interpretation, e.g. for the lagoon diatoms, the cluster of sites 13,14, 15 below has (Salinity<23), (54<PO4<82) and (In-N<965), the only sites to meet those conditions.

---

Revision #3

Created 7 October 2024 21:35:56 by Arden

Updated 12 February 2025 03:09:57 by Abby Miller