Cluster Analysis

Cluster analysis defines groups of stations with similar community composition. The results are displayed in a hierarchical tree-like structure called a dendrogram. On the dendrogram, two groups are first defined, and within these groups subgroups are defined. Subsequently, subgroups within the subgroups are defined. This process is continued until all stations are a separate subgroup. The hierarchical nature of the dendrogram allows the analyst to choose groups of stations that represent a scale of community differences relevant to the present project.

Cluster analysis is also be used to define groups of species that tend to have similar distributional patterns among the stations.

Dissimilarity Index

Both the ordination and cluster analyses require the imput of a dissimilarity matrix, which quantifies the (biological community) dissimilarity between all pairs of stations. We used the Bray-Curtis dissimilarity index (Bray and Curtis 1957) with the stepacross procedure (Williamson 1978, Bradfield and Kenkel 1987). Before computation of the dissimilarity index, the species abundance data were transformed by a square root and standardized by a species mean of abundance values greater than zero. The square root transformation tends to dampen some of the noise often found positively skewed species abundance data. The Bray-Curtis index has been shown to perform well when used with a species standardization (Faith et al. 1987, Smith 1976). Smith (1976) demonstrates how the species mean standardization in particular should best emphasize species abundance counts that change commensurate to changes along community gradients.

All dissimilarity indices are incapable of properly measuring community change for highly dissimilar stations (Swan 1970, Beals 1973). This is because once two stations have no species in common, the dissimilarity index values cannot continue to increase in value as stations become more dissimilar in community composition. The non-monotonic pattern of species abundance values along community gradients also contributes to this lack of index sensitivity for relatively large amounts of community change. The stepacross procedure applied to the computed dissimilarity matrix corrects for this deficiency of the dissimilarity index. Here the larger dissimilarity values (>.8 on a scale of 0 to 1) are reestimated from the shorter dissimilarity values, resulting in larger dissimilarity values that are more commensurate with the degree of actual community changes.

Two-way Coincidence Table

A two-way coincidence table is the station-species abundance data matrix displayed as a table of symbols indicating the relative abundances of the species at the stations. The rows and columns of the table are arranged to correspond to the order of stations and species along the respective station and species dendrograms. Since similar entities (stations or species) will tend to be closer together along a dendrogram, the row and column orders will efficiently show the pattern of species over the stations and station groups.

Since the rows and columns of the two-way coincidence table are ordered according to the dendrograms, the two-way coincidence table is also used to help delimit the station and species groups defined by the cluster analyses. At each potential separation of subgroups defined by the dendrogram, the two way coincidence table is examined to see the corresponding group differences in terms of species presences and abundances. This allows the analyst to choose groups with a level of community differences consistent with the goals of the project.

Calculation of Symbols in the Two-Way Coincidence Table

The symbols are based on square-root transformed, species maximum standardized data values. For example, a raw data matrix:

	species		
station	A	B	C
1	0	4	9
2	1	5	7
3	7	2	12
4	4	0	0

Data transformed by square root:

	species		
station	A	B	C
1	0	2	3
2	1	2.2	2.6
3	2.6	1.4	3.5
4	2	0	0
species maximum	2.6	2.2	3.5

Data standardized by species maximum:

station	species A	B	C
1	0	.91	.86
2	.38	1	.74
3	1	.64	1
4	.77	0	0

Transformed and standardized data as symbols:

station	species A	B	C
1		*	*
2	-	*	+
3	*	+	*
4	*		

where 0 = blank
>0 - .25 = .
>.25 - .50 = -
>.50 - .75 = +
>.75 = *

Thus, the symbols represent the (square-root transformed) abundance of the species relative to the (square-root transformed) maximum species abundance. The standardized values will range between 0 and 1. We need to put all species on a common scale so we can use a single set of symbols for all species.