# Voices of the Forest (Dead Can Dance)
## Assessing Similarities Between Fossil Assemblages

The distribution of faunas and floras can provide an understanding of the temporal and spatial relationships of the paleoenvironments in which they lived and thrived. For example, climatically sensitive assemblages, such as reefs and peat swamps, can reflect where warm, shallow, sediment-free waters existed, in the former case, and everwet, warm and humid terrestrial conditions existed, in the latter. Additionally, the geographic distribution of fossil organisms also may provide evidence about the relationships between ocean basins and continental land masses through time.

This laboratory is designed to test several relationships of communities and assemblages using several basic ecological parameters, correlation coefficients, and simple multivariate statistical analyses. The data provided are taken from a Carboniferous coal mine in Alabama.

## Diversity Indices

To better compare terrestrial or marine benthic communities in space and time, it is necessary to use and apply modern ecological techniques to fossil assemblages. Whittaker distinguished three types of diversity: (1) alpha diversity – diversity within a particular sample (local community – $km^2$ area), (2) beta diversity – diversity associated with changes in sample composition along an environmental gradient (species turnover among local communities – 100 $km^2$), and (3) gamma diversity – diversity due to differences among samples when they are combined into a single sample (regional diversity – $10^6$ $km^2$). One technique for assessing the components of an assemblage is to measure diversity. Simply stated, diversity is the number and distribution of species in an assemblage. Diversity indices allow for comparisons between two habitats. There are several measures of diversity often used. These include:

1. The number of species in an area can be counted (α diversity). This species diversity (or richness) is an expression of the community structure. The more different species present the more diverse the community.
2. The evenness (equatability) of that assemblage, which is also an important aspect of diversity. For example, a community with 10 species, 80 individuals of one type and 2 of each of the rest, is less even than 10 individuals of each of 10 species.
3. Species Diversity Indices attempt to relate the number of species (S = # species) to the number of individuals (N=#individuals) in any sample.

**Shannon-Wiener Index**

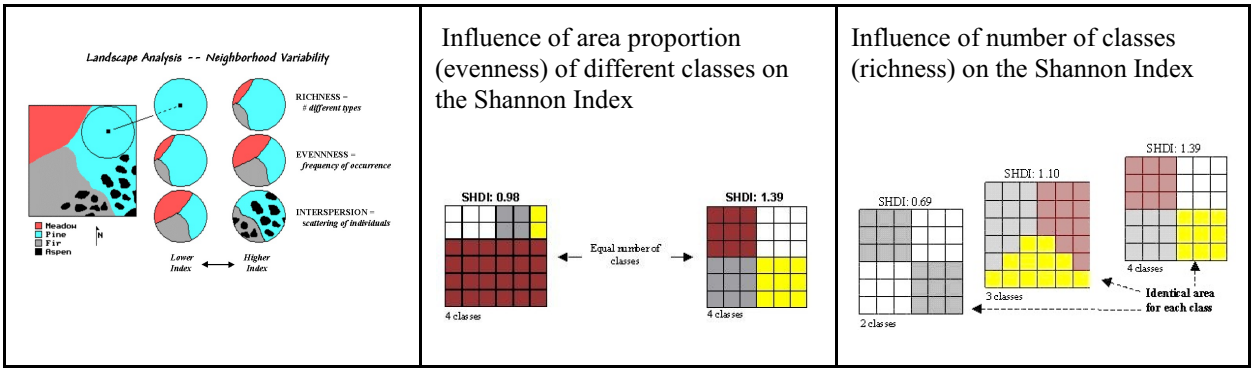
The Shannon-Wiener diversity index is commonly used to describe diversity. This index is a measure of the predictability that a given individual picked at random from a community will belong to a particular species. It is a diversity index that describes in a single number the different types and amounts of animals present in a collection. Shannon-Wiener varies with both the number of species and the relative distribution of individual organisms among the

species. The index ranges from 0 for communities containing a single species to high values for communities containing many species and each with a small number of individuals.
It combines two quantifiable measures:
1. Species richness (the number of species in the community), and
2. Species equatability (how even are the numbers of individuals of each species).
The formula is:

H' = (negative) sum for all species of (p $\log_2$ p)

$$H = -\sum_{i=1}^{s} p_i \left(\ln(p_i)\right)$$

Where p is the proportion of each species and $\log_2$ = the natural log. It contains aspects of both evenness and number of species. The value increases with more species and as they are more evenly distributed.




If we collected one outcrop sample in which there were 256 individuals representing 5 species, the frequency of each of the species would be recorded. Then it is possible to calculate the proportion of each species in the sample (Pi).

| **Fossil Species** | **Frequency** | **$P_i$** | **ln($P_i$)** | **$P_i$*ln($P_i$)** |
|---|---|---|---|---|
| Species #1 | 84 | 0.3281 | -1.1144 | -0.3656 |
| Species #2 | 4 | 0.0156 | -4.1589 | -0.0650 |
| Species #3 | 91 | 0.3555 | -1.0343 | -0.3677 |
| Species #4 | 34 | 0.1328 | -2.0188 | -0.2681 |
| Species #5 | 43 | 0.1680 | -1.7840 | -0.2997 |
| **Sum=** | **256** | **1** | | **-1.3661** |

ln(Pi) is the natural log of that proportion value for each species and the final column is the multiplication of the natural log value and the proportion.

The Shannon-Wiener Diversity Index, H, is then calculated using the equation provided above resulting in a S-W Index H = 1.36.

Where there is a very large sample size with many species, the S-W Index values (H) can range from 0 to ~4.6. A value near 0 indicates that every species in the sample is the same. Conversely, a value near 4.6 would indicates that the number of individuals are evenly distributed between the # of species. Single values in the middle are a toss up - which is an obvious flaw in the index and is the reason that care should be taken when using such a measure.

**NOTE:** To count the number of taxa present in the list that are identified, rather than to count then number of cells in which a number appears, use the COUNTIF function with the modifier ">0" in the formula: COUNTIF(array, ">0").

**Simpson's Index**

The Simpson measure of diversity is sensitive to the abundances of the 1 or 2 most common species of a community and can be regarded as a measure of "dominance concentration". Simpson's index is most appropriately used when the relative degree of dominance of a few species in the community is of primary interest, rather than the overall evenness of the abundance of all species. Simpson's index varies inversely with heterogeneity (i.e., index values decrease [or increase] as diversity increases [or decreases]).

The formula for Simpson's Index is:

$$SI = 1 - \sum_{i=1}^{s} \frac{n_i(n_i - 1)}{N(N-1)}$$

where $n_i$ is the number of individuals/species, and N is the sample size (total number of individuals in the assemblage). It is often calculated using:

$$D = \frac{1}{\sum_{i=1}^{s} p_i \, 2}$$

**NOTE:** In Excel, the carat "^" is used to square a number; for example 5^5 calculates 25.

**Marglef's Index**

This index relates diversity to the number of species (S) to the number of individuals (N).

D = (S-1)/Log N

You can find several Java applets for Ecology and Behavior at:

http://www.hws.edu/aca/depts/bio/Pages/EcoApplets.html

## Similarity Indices

There are many different similarity indices, each of which is designed to produce a metric indicating the proportion of resemblance between two data sets. If the Similarity Index (SI) is relatively high, the two assemblages have a relatively large number of taxa in common. Conversely, if the SI is relatively low, the two assemblages have less taxa in common. This may be due to intrinsic or extrinsic environmental factors. Similarity coefficients come in various permutations. For example, a very simple SI can be defined as the number of genera/species in common between two localities divided by the total number of genera present less the number in common, multiplied by 100 to result in a percentage. Hence,

$$SI = ( CG/(TG-CG)) \times 100$$

Therefore, if 70 genera were collected at one locality and 50 genera collected at another, and if only 20 genera were common to both assemblages, then the SI would be:

$$((20)/(70 + 50-20)) * 100 = 20\%$$

Many data sets consist of binary data (presence/absence) and, hence, binary coefficients must be calculated. These are based on a table of frequency of matches and mismatches of the presence or absence of a single variable. The binary data should be entered into the data matrix as 0 (zero) and 1 (one). Any number that is not zero is also treated as a one, indicating presence.

| | | Sample j | |
|---|---|---|---|
| | | Presence | Absence |
| Sample i | Presence | a (number present in both assemblages) | b (present in i, absent in j) |
| | Absence | c (absent in i, present in j) | d (absent in i, absent in j) |

Several standard coefficients are used routinely and include:

**Sorensen's coefficient :**

$$Sc_{ij} = \frac{2a}{(2a + b + c)}$$

**Jaccard's coefficient:**

$$Jc_{ij} = \frac{a}{(a + b + c)}$$

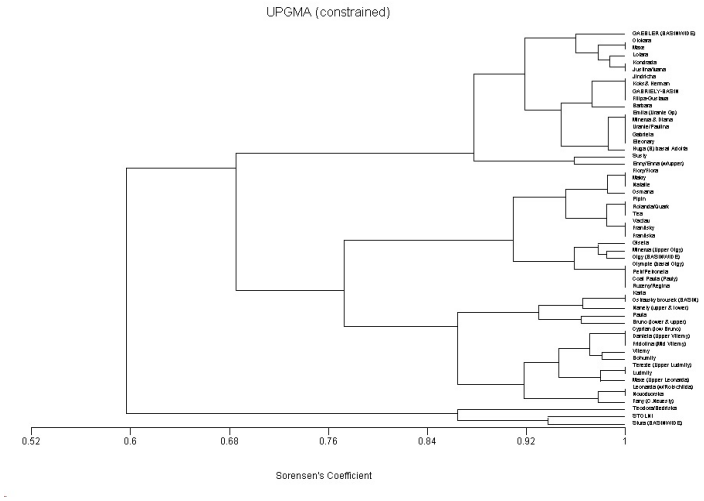
**Simple matching coefficient:**

$$Smc_{ij} = \frac{(a+d)}{(a+b+c+d)}$$

## Cluster Analysis

Cluster analysis is a term used to describe a set of numerical techniques in which the main purpose is to divide the objects of study into discrete groups. These groups are based on the characteristics of the objects and it is hoped the clusters will have some sort of significance related to the research questions being asked.

Cluster analysis is used in many scientific disciplines and a wide variety of techniques have been developed to suit different types of approaches. The most commonly used ones are the agglomerative hierarchical methods. Hierarchical methods arrange the clusters into a hierarchy such that the relationships between the different groups are apparent. The results of this type of analysis generally are presented in a tree-like diagram called a dendrogram. The term agglomerative means that the dendrogram is produced by starting with all the objects to be clustered separate, then successively combining the most similar objects and/or clusters until all are in a single, hierarchical group.




The agglomerative clustering algorithm proceeds as follows:

1. First *the similarity* (see above similarity indices) between each pair of cases must be calculated and placed in a matrix. There are numerous types of similarity and distance measures that can be used.
2. This matrix is then scanned to find the pair of cases with the highest similarity. These will be the most similar cases and should be clustered most closely together.
3. The cluster formed by these two cases can now be considered a single object. The similarity matrix is recalculated such that all the other cases are compared with this new group, rather than the original two cases.

4. The modified matrix is then scanned (as in step 2) to find the pair of cases or clusters that now have the highest similarity. Steps 2 and 3 are repeated until all the objects have been combined into a single group.

The result is a dendrogram that shows the most similar cases linked most closely together. The level of the vertical lines joining two cases or clusters indicates the level of similarity between them. It is important to note that the branching hierarchy and the level of similarity are the only important features of the dendrogram. The exact order of the cases along the vertical axis is not significant. The dendrogram can be envisaged as a mobile that allows the individual clusters to rotate around.

There are two types of cluster analyses that can be conducted – Q-Mode and R-Mode Analyses. In Q-Mode analysis, samples/assemblages are compared. Results from this analysis indicate which assemblages are most similar and, hence, which may be communities or biofacies. Plant ecologists who collect data using a quadrat technique and marine biologists who collect grab-sample data have applied these analyses to their data sets. R-Mode analysis, on the other hand, is where individual taxa are compared in terms of their distribution in the samples. Those taxa that co-occur are grouped together, whereas taxa with distributions that are mutually exclusive are not correlated strongly and are considered to represent different communities (biofacies). Communities identified using these methods are characterized by the species that are common, indicating a high degree of fidelity.

The example is of a data set from the Silesian coal basin, Czech Republic, analyzed using R-Mode cluster analyis and the Sorenson's similarity coefficient.

# EXERCISES

**Diversity Measures**

If you were to take a walk through the Colby Arboretum or any vegetated area in Maine, you would see that the forest is not a monoculture. True, monocultures do exist in the state, mainly planted by the pulp-and-paper industry for economic reasons. But, monocultures are rare in nature, generally restricted to sites where the plants are subjected to some environmental stress. Hence, the biodiversity of any single locality is a function of the biota that lives there at any point in time. The function of this part of the exercise is to evaluate biodiversity measures on a Carboniferous forest.

The attached data set is taken from an autochthonous forest litter preserved above the Blue Creek coal in the Black Warrior Basin, Alabama (Bolsovian [Early Pennsylvanisn] in age). This forest was sampled using a quadrat technique following subsequent drag-line cuts along the mine highwall. Hence, each locality (designated by latitude and longitude) has a spatial relationship to each other locality; there is no difference in temporal relationships. This is a $T_0$ assemblage.

- Using the data, calculate the Shannon-Weiner, Simpson's, and Marglef's Diversity Indices.
- Provide an explain as to why there is variation in the different diversity measures, and the various patterns within the data set.

**Assemblage Similarity**

Using only the Simple Matching coefficient ($SMc_{ij} = (a+d)/(a+b+c+d)$), calculate the similarity coefficients of each of the forest localities as pairwise comparisons.

- That is, calculate the similarity coefficients for: locality 1 vs. locality 6; locality 1 vs. locality 8, locality 1 vs. locality 13, locality 1 vs. locality 14, locality 6 vs. locality 8, etc. See the spreadsheet for the matrix.
- Why is the Smc a logical choice for this data set?

**Cluster Analsyis**

Using cluster analysis techqniues, determine the assemblage relationships (Q-Mode) between the five collection sites in the Blue Creek coal. Use only the Simple Matching coefficient ($SMc_{ij} = (a+d)/(a+b+c+d)$) for the analysis. Above you have already calculated the similarity coefficients of each of the forest localities as pairwise comparisons.

- Find the pairwise comparison that has the highest similarity coefficient. These two collections are then linked at the highest level in the analysis.
- The data in these two collections, then, are combined as a new datum (presence of taxon is now considered to be common to the group) and this datum is then used to calculate the similarity coefficients along with the remaining localities. For example, if localities 6 & 8 are grouped together at the highest level, then the common data pool would be used to calculate 6&8 vs. locality 1, 6&8 vs. locality 13, 6&8 vs. locality 14, locality 1 vs. locality 13, locality 1 vs. locality 14, locality 13 vs. locality 14.
- Again, find the pair-wise comparison that has the highest similarity coefficient. These two collections are then linked at the next highest level in the analysis.
- The data in these two data sets are then combined as a new datum, and the procedure repeated.
- When you've completed all of the analyses, draw a dendrogram using the percent similarity coefficient numbers along the X-axis (proportion similar), arrange the localities along the Y axis, and draw lines linking the groups of collections.
- What community assemblages can you interpret from the cluster analysis? That is, what assemblages cluster close together and which cluster farther apart?
- What plants in these assemblages are responsible for the clustering relationships?

Using the attached plan of the sampling sites, the basic plant architectures for each of the major growth habits (canopy, subcanopy, groundcover), the proportion of litter in each site and the relationships between sites as determined from cluster analysis, draw a reconstruction of the Carboniferous forest across the five-site transect.

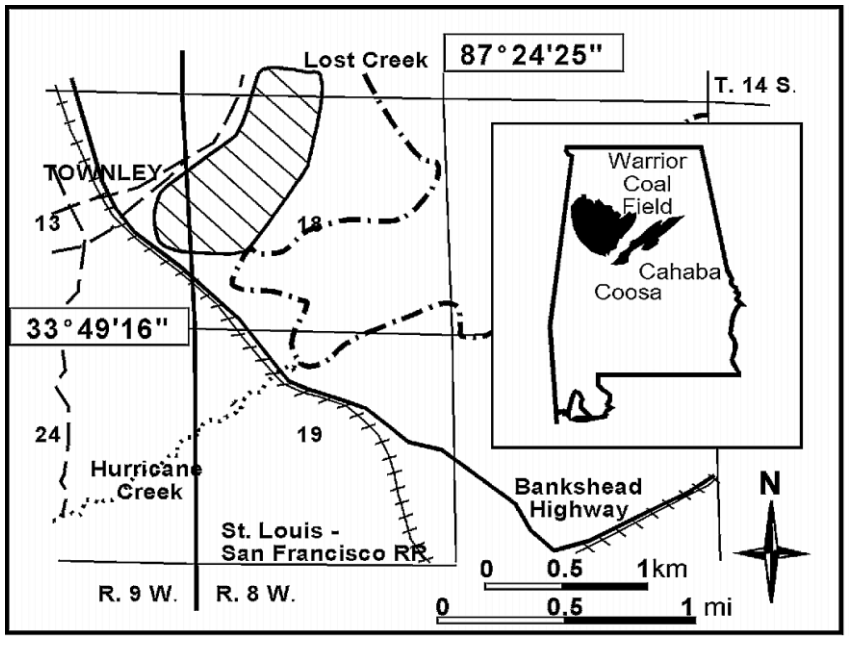
| | | | | | | |
|---|---|---|---|---|---|---|
| Latitude | 8725.113 | 8725.334 | 8725.486 | 8725.512 | 8725.287 | |
| Longitude | 3349.291 | 3349.57 | 3349.671 | 3349.654 | 3349.224 | |
| Blue Creek Forest | Locality1 | Locality6 | Locality8 | Locality13 | Locality14 | |
| Lepidodendron aculeatum | 12 | 0 | 44 | 46 | 8 | CANOPY |
| Lepidodendron obovatum | 0 | 0 | 38 | 18 | 0 | |
| Lepidophloios laricinus | 25 | 45 | 0 | 37 | 99 | |
| Sigillaria elegans | 0 | 8 | 6 | 7 | 0 | |
| Sigillaria ichthyolepis | 0 | 8 | 6 | 9 | 0 | |
| Sigillaria scutellata | 0 | 16 | 0 | 9 | 0 | |
| Calamites cisti | 10 | 7 | 20 | 28 | 45 | |
| Calamites suckowi | 0 | 20 | 8 | 28 | 10 | |
| Artisia | 1 | 0 | 0 | 1 | 0 | |
| Pecopteris arborescens | 7 | 0 | 4 | 2 | 3 | UNDER |
| Cardiopteridium | 8 | 1 | 0 | 0 | 0 | STORY |
| Eremopteris Rhodea type | 0 | 0 | 0 | 0 | 2 | |
| Eremopteris sp. | 0 | 1 | 0 | 0 | 1 | |
| Eusphenopteris lobata | 0 | 2 | 2 | 3 | 3 | |
| Sphenopteris brongniarti | 8 | 5 | 7 | 11 | 5 | |
| Alethopteris cf. valida | 0 | 7 | 0 | 0 | 0 | |
| Alethopteris lonchitica | 0 | 7 | 7 | 13 | 0 | |
| Neuralethopteris elrodi | 3 | 0 | 81 | 13 | 0 | |
| Neuralethopteris pocahontas | 23 | 0 | 18 | 51 | 8 | |
| Neuralethopteris schlehani | 12 | 0 | 7 | 0 | 0 | |
| Neuralethopteris smithsii | 5 | 0 | 0 | 2 | 0 | |
| Neuropteridium | 8 | 0 | 0 | 0 | 0 | |
| Alloiopteris | 0 | 5 | 0 | 0 | 0 | GROUND |
| Diplothmema | 2 | 0 | 0 | 0 | 0 | COVER |
| Lyginopteris hoeninghausii | 1 | 27 | 24 | 12 | 19 | |
| Palmatopteris furcata | 0 | 9 | 1 | 13 | 0 | |
| Sphenophyllum emarginatum | 0 | 6 | 0 | 5 | 0 | |
| Sphenophyllum cuneifolium | 1 | 1 | 0 | 0 | 0 | |
| Sphenopteris cf. schatzlarensis | 3 | 16 | 0 | 0 | 7 | |
| Sphenopteris herbacea | 1 | 0 | 3 | 0 | 2 | |
| Sphenopteris pseudocristata | 1 | 13 | 0 | 0 | 2 | |

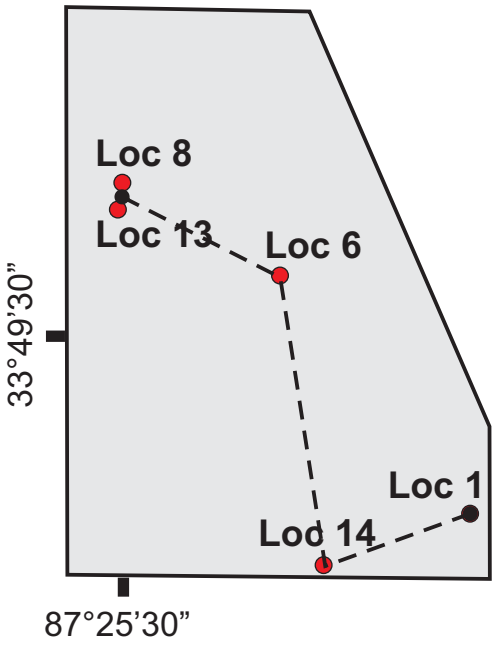
| | | | | | |
|---|---|---|---|---|---|
| Shannon Weiner Index | | | | | |
| Simpson Index | | | | | |
| Marglef Index | | | | | |

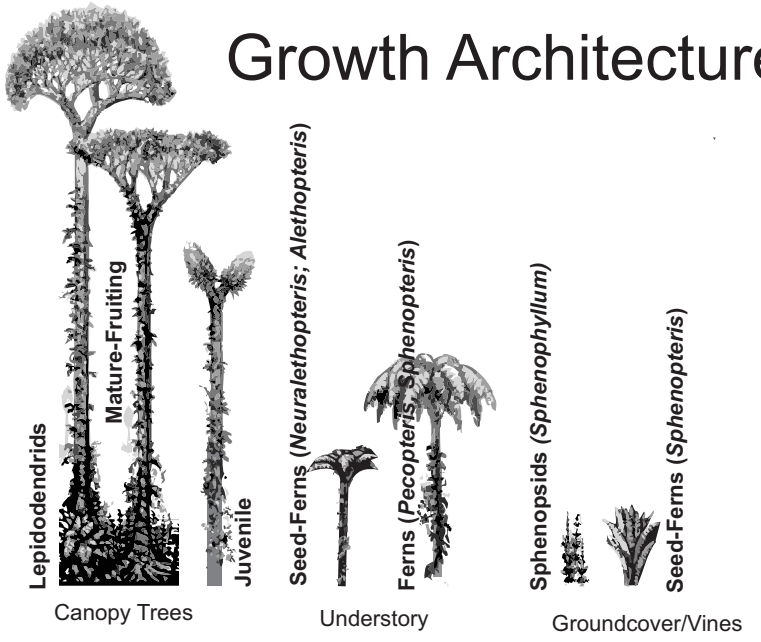
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | locality 1 | | | locality 1 | | | locality 1 | | |
| locality 6 | P1 P6 | A1 P6 | locality 6 | | | locality 8 | | | |
| | A6 P1 | A1 A6 | | | | | | | |
| | locality 1 | | | locality 1 | | | locality 6 | | |
| locality 13 | | | locality 14 | | | locality 8 | | | |
| | | | | | | | | | |
| | locality 6 | | | locality 6 | | | locality 8 | | |
| locality 13 | | | locality 14 | | | locality 13 | | | |
| | | | | | | | | | |
| | locality 8 | | | locality 13 | | | | | |
| locality 14 | | | locality 14 | | | | | | |
| | | | | | | | | | |

Simple Matching Coefficient

| | locality 1 | locality 6 | locality 8 | locality 13 | locality 14 |
|---|---|---|---|---|---|
| locality 1 | | | | | |
| locality 6 | | | | | |
| locality 8 | | | | | |
| locality 13 | | | | | |
| locality 14 | | | | | |




Loc 8/13 Loc 6 Loc 14 Loc 1