

Affinities and high-order diversity within landscape mosaics

CONRAD A. ISTOCK and SAMUEL M. SCHEINER*

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA and University of Michigan Biological Station, Pellston, Michigan 49769, USA

Summary

Concepts and methods for use in the study of plant community variation across a single landscape, or for the comparative analysis of two or more landscapes, are presented. The method is called affinity analysis because it produces a sorting and scaling of all the communities in a data set according to degree of differentiation in composition away from an objectively identified subset of modal communities. Modal communities possess maximum affinity with the whole landscape because they contain a large number of the species common throughout the landscape. The modal communities provide a kind of central tendency for the landscape and an anchor for depicting the dispersion of all other communities in it.

To accomplish affinity analysis, the Jaccard similarity coefficient and the Wilcoxon T statistic are used in a two-step transformation of the primary species-by-site data matrix. From these calculations we obtain both a visual image, the S - T graph, and summary statistics for the landscape-wide diversity described by the data. One summary statistic is a high-order diversity measure for the total patchwork of communities. We refer to this measure as mosaic diversity. The analytical results are referred to collectively as the 'metastructure' because they provide a general, abstract, characterization of any set of community data. This abstract rendering of the data allows comparison of pattern and variation between taxonomically, geographically, and temporally different landscapes. Either presence/absence or abundance data may be used.

Examples using artificial and field data are presented. All but one of the field data sets showed a significantly higher mosaic diversity than would be expected from a randomly constructed landscape. We also show how the new methods may be used with ordination to explore intralandscape patterns in more detail than was previously possible.

In addition to statistical matters, ecological and evolutionary interpretations of affinity analysis are discussed. Topics included in this discussion involve reasoning about the influences on diversity arising from micro- and macroevolution, species packing and association, environmental gradients, differential fitness expression among species-populations, continuity in community variation, and the uses of both presence/absence and abundance data in community studies. Among the examples provided, mosaic diversity is independent of β diversity (Whittaker, 1972).

Keywords: Diversity; β diversity; mosaic diversity; ecological communities; landscape mosaics; plant ecology; affinity; community analysis; ordination; community variation; statistical inference; random community arrays; null-hypothesis communities.

*Present address: Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115, USA.

Introduction

Diversity within and among ecological communities continues to provide one of the most puzzling and challenging objects of study. For more than a decade one could sense a widespread ennui concerning this essential area of community ecology. Uncertainty about what it is we should measure and how we should go about collecting, analysing, and interpreting community data lies at the root of the difficulty and heightens the challenge of community studies (Eberhardt, 1969; MacArthur, 1972; Whittaker, 1972; Pielou, 1975). Complexities are inherent in the hierarchy of spatial scales, temporal scales, taxonomic categories, and physiognomic categories (Pielou, 1975). Efforts to develop measures of diversity have concentrated on the lowest levels of the hierarchy, the diversity within single communities or the comparison of diversities among pairs of communities (Whittaker, 1972; Peet, 1974). Some multivariate techniques have been developed which extend the pairwise comparisons to the examination of entire sets of community data (Gauch, 1982). However, even these methods are still based solely on pairwise comparisons of similarity, and pursue only the investigation of internal pattern.

A mosaic of communities, spread over a landscape or archipelago, possesses properties at the level of the entire mosaic which previously have not been well defined and measured. We present here some new ideas and associated measures for the study of overall patterning and diversification within and between mosaics. These ideas and measures provide a logical, conceptual basis for characterizing and comparing general properties of the mosaic which exists in any landscape of plants or animals.

Our goal is to explore a space which represents the set of species present at each site in a landscape. There are as many axes in the space as there are species. Each axis records the presence or absence of a single species. Thus the list of species at a site is described by a single point in the space. We will refer to this space as the similarity space and explore the central tendency and dispersion of sites within it. The concepts and methods of this paper explicitly assess the following properties of landscape data: (1) the central tendency or approximate 'mode' or 'modes' of the ensemble of communities (identical with sites) which the data describe, (2) an ordering of all community samples according to the deviation in composition of each community away from this central tendency, and thus its 'affinity' to the entire ensemble, (3) a single measure of diversity for the entire patchwork of community samples analogous to a variance for the community ensemble – here referred to as 'mosaic diversity' (μ), and (4) the extent of continuity and discontinuity in sample variation. The procedures which accomplish these things are collectively called affinity analysis. Although we present this method in the context of the analysis of landscapes of plant communities, any set of objects which can be arranged as items (columns) and attributes (rows) is amenable to analysis by this method.

Mosaic diversity provides a higher-order extension of the α , β , and γ hierarchy of diversity measures defined by Whittaker (1972). Beta diversity is one measure of high-order diversity. For β diversity Whittaker used two definitions: the turnover of species along a gradient and the mean similarity among sites (\bar{S}). The two definitions are conceptually close, species turnover increases as \bar{S} decreases. Affinity analysis presents a new measure of high-order diversity, μ , which represents the variation and degree of structuring around the mean similarity. The new information captured by μ is not available merely by inspection or ordering of specific details within the data set or of the derived matrix of similarity coefficients. Further, on empirical grounds, as we will show, \bar{S} and μ appear to be independent measures of diversity with different biological meanings. Lower-order diversity measures such as α and γ , i.e. the number of species per site and the number of species in a region, respectively, are less inclusive descriptions of the data matrix.

We will also examine briefly how the overarching assessment obtained through affinity analysis

can be supplemented by and coordinated with depictions of direct, internal, similarities of communities through ordination (Gauch, 1982). While ordination provides one or more images of the internal patterning of a single data array (for many examples see Gauch, 1982), it does not allow general comparisons of patterning and diversification among different landscapes because of the incommensurable scales of ordinations for two or more arrays. The graphical results of affinity analysis and the associated mosaic diversity measurements, all of which we call the 'metastructure', do allow a comparative search for regularities and irregularities of patterning among taxonomically, geographically, or temporally unrelated landscapes. A subsequent paper (Scheiner and Istock, unpublished; see also Appendix) extends the procedures for statistical inference used within affinity analysis. Thus it is possible to test the examples presented here for overall deviations from random variation.

Occasionally, it has been recognized that an evolutionary perspective for diversity studies exists (Whittaker, 1972; Pielou, 1975). However, few ideas about diversity have been extracted from evolutionary thinking. The properties of community mosaics which we explore here have explicit connections with reasoning from evolution. First, there is a connection through our attempts to discern and order larger-scale patterns of species association, such patterns being the present consequence within the community mosaic, or regional vegetation, of many separate genetic and evolutionary directions. The community mosaic is then the coming together of many separately evolved species, and perhaps also small groups of coevolved species, creating shifting patterns of association in space and time among the modern-day representatives of diverse taxa. It is highly unlikely that all modes of ecological adaptation are present all the time everywhere on earth, or perhaps that all such modes have appeared yet. And it is clear from studies in physiological ecology that most species possess quite restricted adaptations to regional environmental variations. If these ideas are correct, it follows that the species and modes of adaptation actually present will engender characteristic patterns of association and exclusion within the community mosaic in response to the prevailing abiotic backdrop. Thus, we would predict that species association and community composition over a landscape will be non-random (Pielou, 1975). The procedures we present here allow a limited testing of this conjecture.

As a second conjecture, related to that just discussed, we might expect that the spread of adaptive modes may often be highly incomplete for historical and biogeographical reasons. The result is likely to be the creation of patches characterized by evolutionary oddities, so-called relict communities in obvious cases. Our procedures allow us to identify less obvious cases of such a phenomenon and measure the extent to which discontinuity arises in the landscape-wide spread of variation in community composition. We find evidence suggesting that limited discontinuity may be a common property of community mosaics, but continuity appears to be the more general property of plant community variation.

A preliminary exploration of these methods for plant ecology appears in Coleman and Istock (1980), and a quite different use of the present version of affinity analysis for study of clinal variation and diversification of chromosomal variants within one fruit-fly species appears in Etges (1984).

Procedures

Affinity analysis can be applied to presence/absence data or to absolute or relative abundance data. The method, in general form, proceeds by a two-step calculation. First, a site-by-site similarity matrix is calculated from primary data in a species-by-site matrix. Second, a site-by-site matrix of Wilcoxon T values with signs retained is calculated from the similarity matrix. It is the matrix of Wilcoxon T values which relates each site to the landscape as a whole in contrast to the

pairwise comparisons embodied in the matrix of similarity values. The statistical assumptions being made by this method are more spartan than those of most multivariate analyses.

Specifically, in the first step, a similarity matrix is generated using the Jaccard (1901) similarity coefficient with presence/absence (incidence) data:

$$S_{ij} = c/(a + b + c) \quad (1)$$

where c is the number of species common between any two samples i and j , and a and b are the number of species found only in sites i and j , respectively. If abundances are involved we have used analogous quantitative similarity coefficients (Mueller-Dombois and Ellenberg, 1974; Goodall, 1973). For each site the mean similarity (\bar{S}_i) is calculated. Our use of presence/absence data accords with previous studies (e.g. Avena *et al.*, 1981) which demonstrated that this type of data yields a sufficient representation of landscape composition. The Jaccard coefficient has been judged one of the best for use with presence/absence data (Mueller-Dombois and Ellenberg, 1974; Janson and Vegelius, 1981; Faith, 1983).

A similarity matrix is not sufficient for our goals. What is the pattern of dispersion of sites (points) in the similarity space? In a randomly occupied space there will be neither significant clusters of sites nor an even distribution of them. To investigate the dispersion of sites in the similarity space we use an adaptation of the Wilcoxon T statistic (Siegel, 1956, sample calculation on p. 79; Snedecor and Cochran, 1967). Calculation of the Wilcoxon T statistics is the second step in affinity analysis.

The following example shows the calculation of a single Wilcoxon T value for two sites i and j using two columns of similarity coefficients. The subscript k denotes the number for any site other than i or j . The order of subtraction for the difference is defined as S_{ik} minus S_{jk} .

k	site i	site j	difference	rank
1	S_{i1}	S_{j1}	d_1	r_1
2	S_{i2}	S_{j2}	d_2	r_2
3	S_{i3}	S_{j3}	d_3	r_3
.
.
.

Each d value is positive or negative. Ranks are assigned according to the absolute values of the differences, but the sign of the difference is given to the rank. Positive and negative ranks are summed separately, and T_{ij} is equal to the sum of ranks with the smaller absolute value. Thus each T_{ij} relates sites i and j to all of the remaining samples collectively. The calculation is repeated for all pairs of sites thus generating a matrix of T_{ij} . Each T_{ij} equals $-T_{ji}$, i.e. across the diagonal, the matrix is symmetrical in absolute values and opposite in signs.

Two sites close together in the similarity space will have differences, d_k , which average close to zero and fluctuate randomly in sign. Hence, their $|T_{ij}|$ will be close to half the sum of the absolute values of all the ranks (Siegel, 1956) and the sign of their T_{ij} will be as often negative as positive. Sites farther apart in the similarity space will have smaller $|T_{ij}|$ values. If a site is close to the mode compared to another, the close one will tend to have the positive T value, the other the negative T .

By taking the algebraic average of all the T_{ij} for any site i , we obtain a general measure of the position of site i within the context of the entire similarity space. This average, \bar{T}_i , becomes then the measure of 'affinity' of site i with the entire landscape. Sites with large positive \bar{T}_i lie close to the central tendency of the similarity space and those with large negative values are distant from it.

The average similarity, \bar{S}_i , and affinity, \bar{T}_i , values for all sites are used to construct the S - T plot shown schematically in Fig. 1, and it is this graph which provides the central image of variation and degree of structuring of the landscape as a whole. Whether the graph rises to the right or to the left is determined by the convention established for positive and negative orders of subtraction in the Wilcoxon T procedure, and is merely an arithmetic detail. On an S - T graph arranged like that of Fig. 1, modal communities fall up and to the left because they are particularly rich in species which are common all over the landscape and they lack the rarer species. Communities which plot progressively downward and to the right are either depauperate or rich in rarer species relative to the mode. Thus the S - T plot is a sorting and scaling of communities from the mode out to those with increasingly divergent and unusual assemblages of species. The graph will always have a weighted average at $\bar{T}_i=0$.

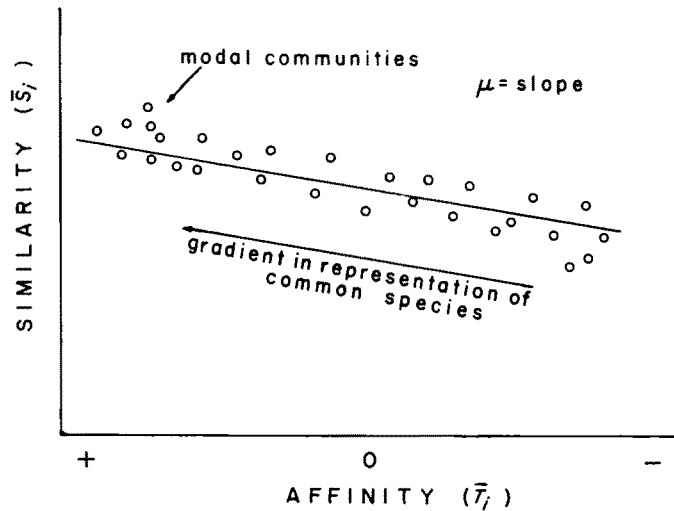


Figure 1. Schematic representation of the S - T graph of affinity analysis. Mean similarity (\bar{S}_i) for each site is on the ordinate and the mean affinity (\bar{T}_i) for each site is on the abscissa. The S - T graph along with its summary statistics are the 'metastructure' of a landscape.

Empirically, we have discovered that the S - T plot is approximately linear, at least over most of its extent. Simulations of randomly formed landscapes yield S - T plots with very shallow slopes. So, the degree of structuring of the regional patchwork may be defined by the slope of the graph; the greater the structuring the steeper the slope. The slope of a linear regression of \bar{S}_i on \bar{T}_i fit to the points on the S - T graph measures the extent of this tilt. Hence we define the mosaic diversity (μ) of the landscape to be this regression slope. This tilt is a function of both the variance in similarities among sites, measured by the variance of \bar{S}_i , and the pattern or structure of those similarities, measured by the variance of \bar{T}_i . The affinities are a function of the similarities. However, different amounts of structure will produce different slopes even if the variance in similarities is the same (see Table 1, examples 2-4). Conversely, since the Wilcoxon procedure uses only the ranked differences in similarity and not the absolute values, different variances in similarity can result in the same variance in affinity.

Now we can obtain a standardized mosaic diversity which allows comparison of high-order structure among unrelated regional mosaics. The data matrix for each regional array has three dimensions: the number of sites, number of species, and proportion of species-by-site entries

Table 1. Summary statistics from affinity analysis and dimensions for examples based on both artificial and field data matrices. All symbols are defined in the text.

Data set	Q	R	F	\bar{S}	$E(\bar{S})$	$\mu \times 10^{-3}$	$E(\mu) \times 10^{-3}$	μ^*
<i>Artificial data</i>								
1	24	12	0.54	0.35	0.38	4.42	1.25	2.54
2	22	40	0.38	0.30	0.23	2.52	1.10	1.29
3	22	30	0.50	0.41	0.34	2.84	1.18	1.42
4	22	20	0.75	0.63	0.64	3.86	1.13	2.42
<i>Field data</i>								
5	33	237	0.17	0.24	0.09	5.37	1.50	2.59
6	21	142	0.21	0.23	0.11	1.61	0.58	1.78
7	21	142	0.21	0.30	0.12	2.01	0.40	4.06
8	21	142	0.21	0.40	0.13	4.54	0.72	5.34
9	25	94	0.34	0.41	0.20	1.61	0.58	1.78
10	25	94	0.34	0.57	0.21	1.57	0.45	2.47
11	13	143	0.19	0.23	0.10	4.97	1.12	3.44
12	30	19	0.32	0.30	0.18	0.89	0.76	0.18
13	30	19	0.32	0.36	0.21	0.81	0.61	0.33
14	47	486	0.15	0.22	0.08	0.25	0.03	6.47
15	40	405	0.19	0.27	0.10	0.26	0.07	3.03
16	31	347	0.20	0.23	0.10	0.41	0.14	1.97
17	29	316	0.22	0.26	0.12	0.40	0.19	0.11

1. Matrix with 2 different monocultures, as described in the first example in the text.
- 2-4. A series which varies in F and R .
5. North Carolina piedmont upland and bottomland communities, Oosting (1942), presence/absence data.
6. North Carolina upland old-field and pine communities, Oosting (1942), presence/absence data.
- 7-8. Same as 6 with frequency and density data, respectively.
9. Danube meadows, Mueller-Dombois and Ellenberg (1974), presence/absence data.
10. Same as 9 with density data.
11. Hyuck Preserve moss communities, Coleman and Istock (1980), presence/absence data.
12. Wisconsin tree communities, Peet and Loucks (1977), presence/absence data.
13. Same as 12 with density data.
14. Northern lower Michigan plant communities, unpublished data of Istock and Scheiner (see text), presence/absence data.
15. Same as 14 but with the 'tail' communities deleted.
16. Australian mediterranean-type shrubland communities, Griffin *et al.* (1983), presence/absence data.
17. Same as 16 but with the 'tail' communities deleted.

actually filled. To correct for differing dimensions, a standardized measure of mosaic diversity is calculated:

$$\mu^* = [Obs(\mu) - E(\mu)]/E(\mu) \quad (2)$$

where $E(\mu)$ is the expectation of μ obtained by simulation of a random case (Scheiner and Istock, unpublished).

A distinct and useful deviation from simple linearity sometimes occurs in the form of one or more samples which plot very low on the right and back toward $\bar{S}_i=0$ and $\bar{T}_i=0$, as seen in Fig. 2. We refer to this as an extra 'tail' on the $S-T$ graph. This 'tail' is of ecological interest because it introduces an additional property of landscape variation which we define as *continuity* within the specific context of affinity analysis. As long as variation among communities grades continuously

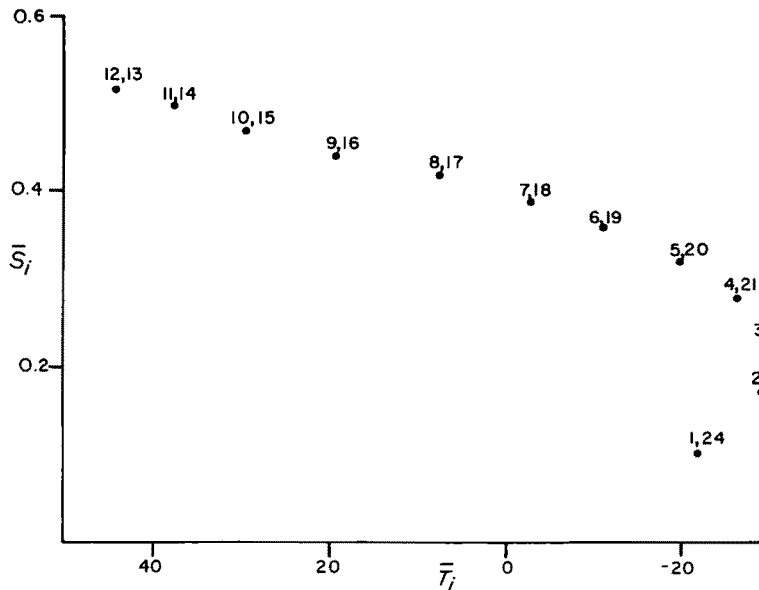


Figure 2. S - T graph from affinity analysis for the artificially structured data set with two different terminal monocultures (sites 1 and 24) described in the text (example 1 of Table 1). This example illustrates the development of a 'tail' of discontinuous sites.

away from the central tendency, only the linear S - T graph arises (Figs 3 and 4), while discontinuity produces the tail. This is to say, no matter how many floristically, environmentally, successional, or stochastically determined directions of divergence there are in the data, the S - T graph remains distinctly linear as long as each direction of divergence is quasi-continuous. Biologically, quasi-continuous means that a relatively smooth substitution of rarer for commoner species is occurring along the linear portion of the graph down and away from the mode. The 'tail' appears because some sites make a rather abrupt shift to a composition which includes both species unique to a single site and a group of species characteristic of widely divergent parts of the landscape. Thus the relationship of \bar{S}_i and \bar{T}_i reverses. The linear nature of the graph remains even if the samples at the ends of the various gradients have no species in common. Any samples sharing no species with all other samples will plot exactly as $\bar{S}_i=0$, $\bar{T}_i=0$. The proper measure of mosaic diversity in data sets with discontinuity revealed by the 'tail' is the slope of the linear regression from a reanalysis with the 'tail' samples deleted. The change in μ provides an approximate measure of the amount of discontinuity.

Through the analysis of artificial data arrays we have found that affinity analysis is robust to non-proportional representation of the communities of the landscape. One must greatly distort a sampling procedure away from a random or stratified method before large changes in the S - T representation of a landscape structure appears. The method is particularly robust with large sample sizes. On the other hand, the method also appears to be rather robust with only a few samples, say twenty or so. In our experience 20 samples seem sufficient to provide a reliable, initial description of the landscape structure.

Alternative similarity coefficients and other non-parametric, rank-order, statistical measures might be used in affinity analysis, though we have not tried them. A non-parametric procedure seems appropriate because the similarity coefficients are not always normally distributed (Wolda,

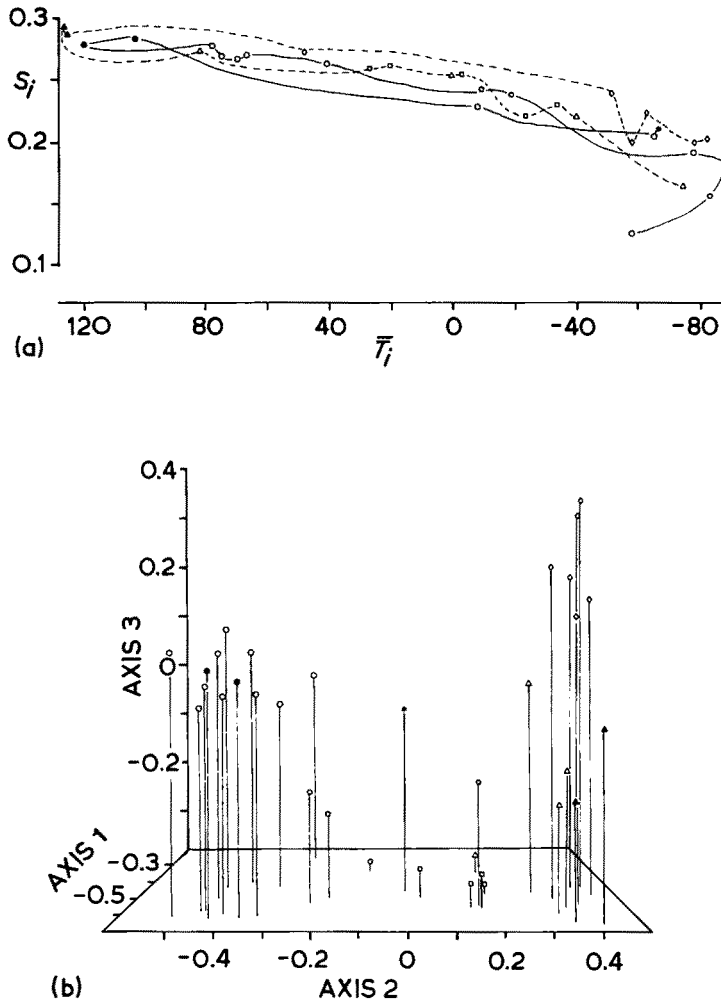


Figure 3. (a) S - T graph and (b) principal-components analysis for Oosting's (1942) data describing upland and bottomland piedmont communities in North Carolina (example 5 of Table 1). The various types of communities, Oosting's classification, are indicated as follows: (Δ) upland pine sites, (\square) bottomland pine sites, (\blacklozenge) bottomland birch sites, (\circ) bottomland mixed hardwood sites, (\diamond) upland climax oak-hickory sites, and ($*$) the bottomland post-climax hardwood site. The model sites are indicated by filled symbols. The various extensions away from the mode disclosed in 3b are traced out on the S - T graph.

1981), and the number of sample sites may be small in some cases. These are the motivations in the preliminary development of these methods (Coleman and Istock, 1980).

Listings of computer programs for affinity analysis written in FORTRAN and Microsoft BASIC are available from the authors. The programs provide: a matrix of joint species occurrences, a matrix of pairwise similarity values, a matrix of T_{ij} values, \bar{S}_i and its variance for each site, \bar{T}_i and its variance for each site, the number of species and mean number of species shared by each site, \bar{S} , the among site variances of \bar{S}_i and \bar{T}_i , and μ and the random expectations of \bar{S} , $V(S_{ij})$, $V(\bar{S}_i)$, $V(T_{ij})$, $V(\bar{T}_i)$, and μ .

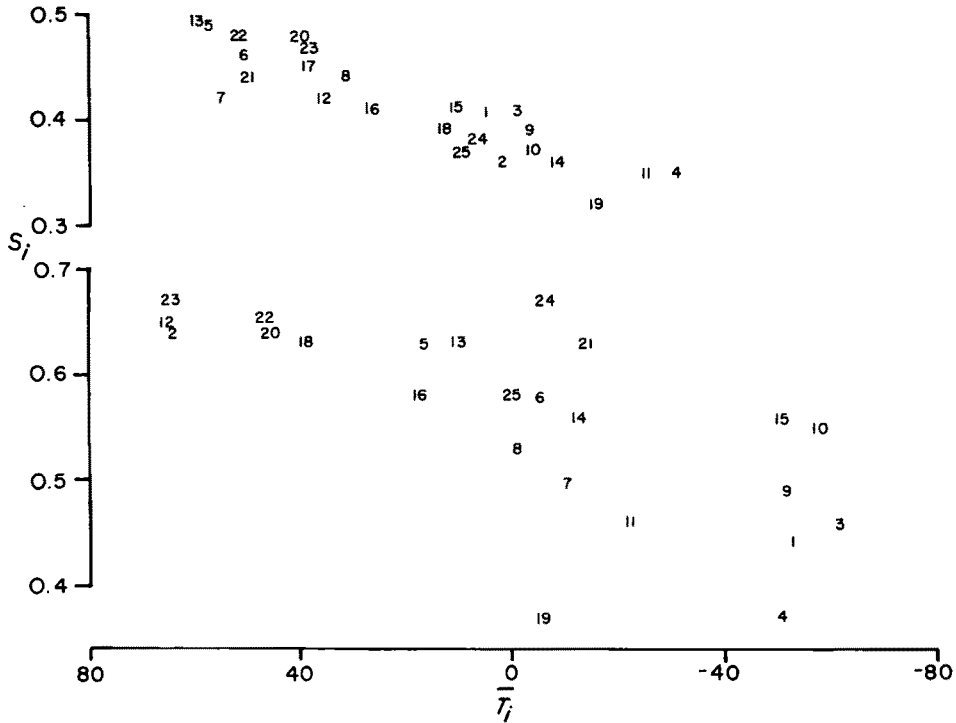


Figure 4. $S-T$ graphs for the Danube meadows data of Mueller-Dombois and Ellenberg (1974), examples 9 and 10 of Table 1. The $S-T$ graphs show presence/absence (upper) and density (lower) versions of the data. Site numbers as in original data.

Examples

Next we present a number of examples of affinity analysis. For each: Q is the number of sites or community samples, R is the number of species, and F is the fraction of all cells in the site by species matrix containing a value for presence (1) or a value for abundance.

The first example uses an artificial presence/absence matrix where $Q=24$, $R=12$, and $F=0.54$ (example 1 in Table 1) having the following shape and with two central sites containing all species (not all species and sites are shown):

		sites														
		1	1	1	1	1	.	1	1	.	0	0	0	0	0	0
		0	1	1	1	1	.	1	1	.	0	0	0	0	0	0
		0	0	1	1	1	.	1	1	.	0	0	0	0	0	0
species

	0	0	0	0	0	.	1	1	.	1	1	1	1	0	0	
	0	0	0	0	0	.	1	1	.	1	1	1	1	1	0	
	0	0	0	0	0	.	1	1	.	1	1	1	1	1	1	

Fig. 2 presents the $S-T$ graph for this matrix. The richest central communities in the matrix share species with all other communities and they plot as the modal communities in the $S-T$ graph. The linear and 'tail' parts of the graph are also clear, although there is not complete

discontinuity and the 'tail' ending in monocultures does not reach completely back to $\bar{S}_i=0$, $\bar{T}_i=0$. The original symmetry of the data is revealed by the fact that the pairs of sites at equal distance from the modal communities plot in exactly the same places. Mosaic diversity as well as other statistics for this and other artificial and natural examples are summarized in Table 1. Mosaic diversity is a complex phenomenon as artificial examples 1–4 of Table 1 illustrate. The increased species richness in examples 2 and 3 actually lowers mosaic diversity.

We have evaluated many other simple and complex artificial data structures representing many kinds of imaginable landscapes (e.g., examples 1–4 in Table 1). All of them lead to the central interpretation depicted in Fig. 1, namely that the S – T graph is controlled by a gradient in representation of common species, that the Wilcoxon T expresses the extent of variation in species composition among communities in a landscape, and that μ diversity measures the degree of structuring among species associations.

North Carolina piedmont

The analysis of Oosting's (1942) data of the North Carolina piedmont appears in the S – T graph of Fig. 3 (example 5 in Table 1), where $Q=33$, $R=237$, and $F=0.17$. The data consist of sites of various ages from both upland and bottomland forests. We used only the presence/absence aspect of the data. The youngest and the oldest sites plot at the extreme right end of the S – T graph while the modal communities are of intermediate age.

Using ordination, in this case a principal-components analysis (Fig. 3b), one can trace arms of relatedness or divergence relative to the mode, as we have done on Fig. 3a. This analysis indicates that the upland climax oak–hickory sites are closely allied with each other, forming the end of one arm. They are related to the younger upland pine sites which, presumably, precede the oak–hickory sites during succession. Two of the older pine sites, 31 and 75 years old, are modal communities. Similarly, the bottomland sites proceed from the youngest mixed hardwood and birch sites at the end of one arm, through mixed hardwoods of intermediate age at the mode and back to a very old, >300 years, post-climax bottomland forest. The bottomland pine sites are in a somewhat anomalous position being more related to the upland pine sites than the other bottomland sites. Thus the arms trace a pair of presumed successional sequences. Any method of ordination or classification describing internal structure can be used in concert with affinity analysis in this manner to augment the interpretation of the S – T graph. Recall, however, that affinity analysis as a whole yields a measure of diversity and other characteristics of the data not available from methods of ordination and classification.

The Oosting (1942) study also contained a detailed analysis of old field to pine forest succession in the Piedmont upland (examples 6–8, Table 1). An analysis of this subset of Oosting's data is included to show the effects of abundance data used with affinity analysis. Addition of abundance data increases both the mean similarity of sites and the mosaic diversity of the landscape. This increase in structure occurs because species with high abundance in one site also tend to have high abundances in other sites that share the same species. Similar sites become even more similar when abundance data is considered. The S – T graph and principal-components analysis (not shown) give a similar tracing of the successional sequence but in the much simpler form of a single pair of arms since only upland habitats were analysed.

Danube valley meadows

Data from Mueller-Dombois and Ellenberg (1974) describing an array of meadow samples in the Danube River valley (examples 9 and 10 in Table 1) provide a third example. The dimensions are $Q=25$, $R=94$, and $F=0.34$. S – T graphs with both presence/absence and density

data appear in Fig. 4. Again linearity is apparent on both $S-T$ graphs and the only sample with a tendency to discontinuity is site 19 (as discussed by Mueller-Dombois and Ellenberg, 1974). Overall, the addition of density information to the analysis does not produce pervasive change in the way the sample sites assort by affinity with the mode, but there are some separate, dramatic changes of position, as for example sites 5, 6, 7, and 13.

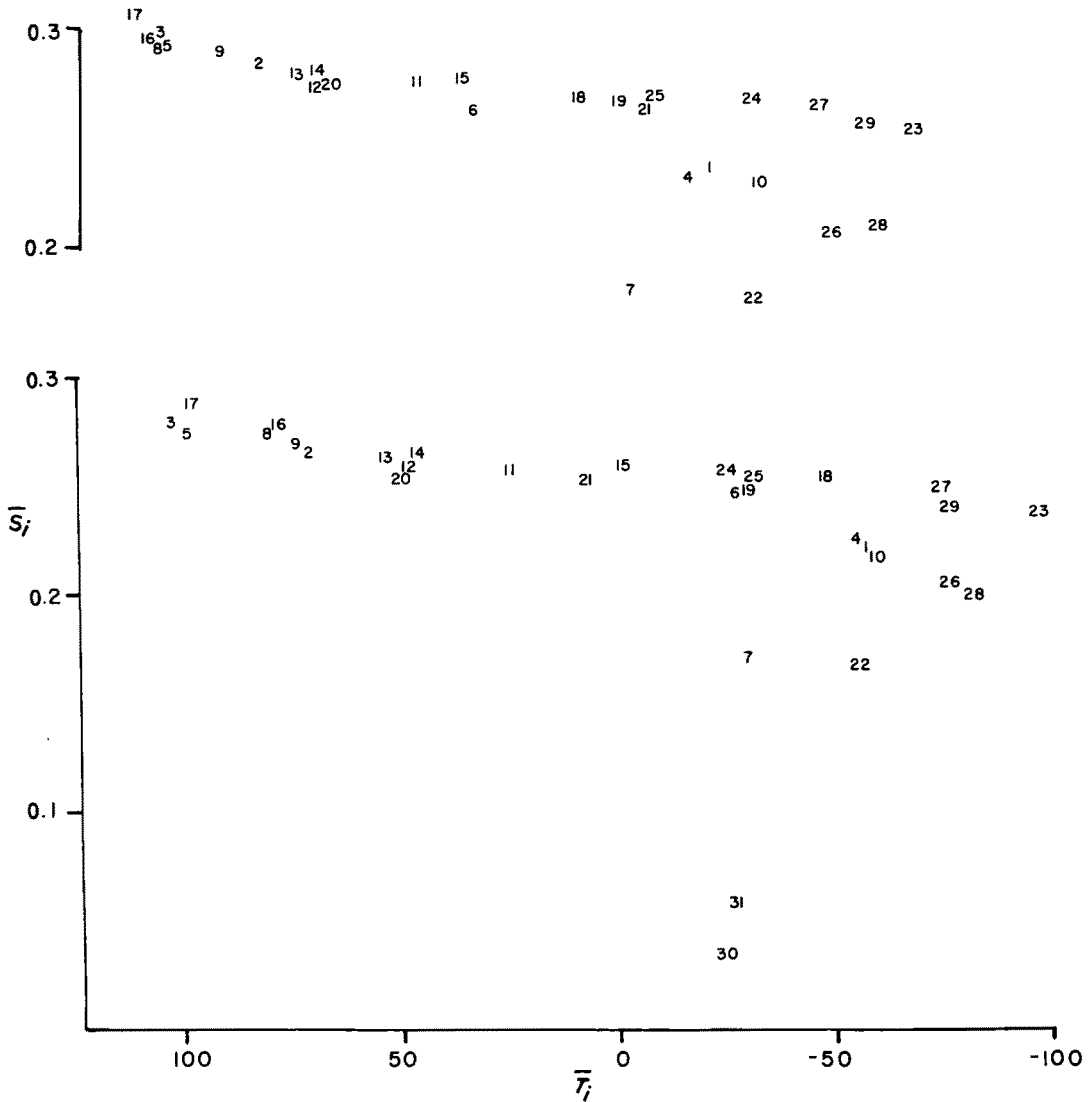


Figure 5. $S-T$ graphs of Australian mediterranean-type shrublands data of Griffin *et al.* (1983), presence/absence version (examples 16 and 17 of Table 1). The lower graph is based on the complete data and the upper graph on the data when two extreme sites, 30 and 31, are deleted.

Australian shrubland

Another example (numbers 16 and 17 in Table 1) provides a summary from the affinity analysis of data describing a mediterranean-type shrubland landscape in south-western Australia (Griffin *et al.*, 1983). Only the presence/absence aspect of the data has been used. The μ^* of 1.97 is close to the values from the North Carolina and Danube data. If the two communities acknowledged by the authors to be extreme are removed, the value of μ^* falls to 1.11, indicating a quite homogeneous array. Perhaps most interesting from the standpoint of confirming our general interpretation of affinity analysis, the geographically central sites, which are also floristically central according to the original analysis of Griffin *et al.* (1983), are the ones plotting in the mode of the S - T graph (Fig. 5 here, and see Fig. 3 of Griffin *et al.*, 1983).

Michigan forests

A final example (numbers 14 and 15 in Table 1, and Fig. 6) comes from our recent study of compositional variation within the hemlock-white pine-northern hardwoods forest (Nichols, 1935) of northern lower Michigan. In the S - T graph for these data a 'tail' extending almost to $\bar{S}_i=0$, $\bar{T}_i=0$ is present, allowing us to assess quantitatively the effect of discontinuity. At present, our data describing this hemlock-white pine-northern hardwoods landscape include 47 sample sites and 486 species of vascular plants. Among the sample sites are wet swamp forests, upland hardwoods, red and white pine forests, mixed hardwoods and pine, aspen and pine, natural and artificial burns, jack pine woodlands on wet and dry sites, spruce-fir forests, birch forests, bogs, and sand dunes. When the seven sites (dunes, dry jack pine, bog, and larch swamp) which form the 'tail' are eliminated the mosaic diversity reflected in μ^* decreases markedly. The reality of the

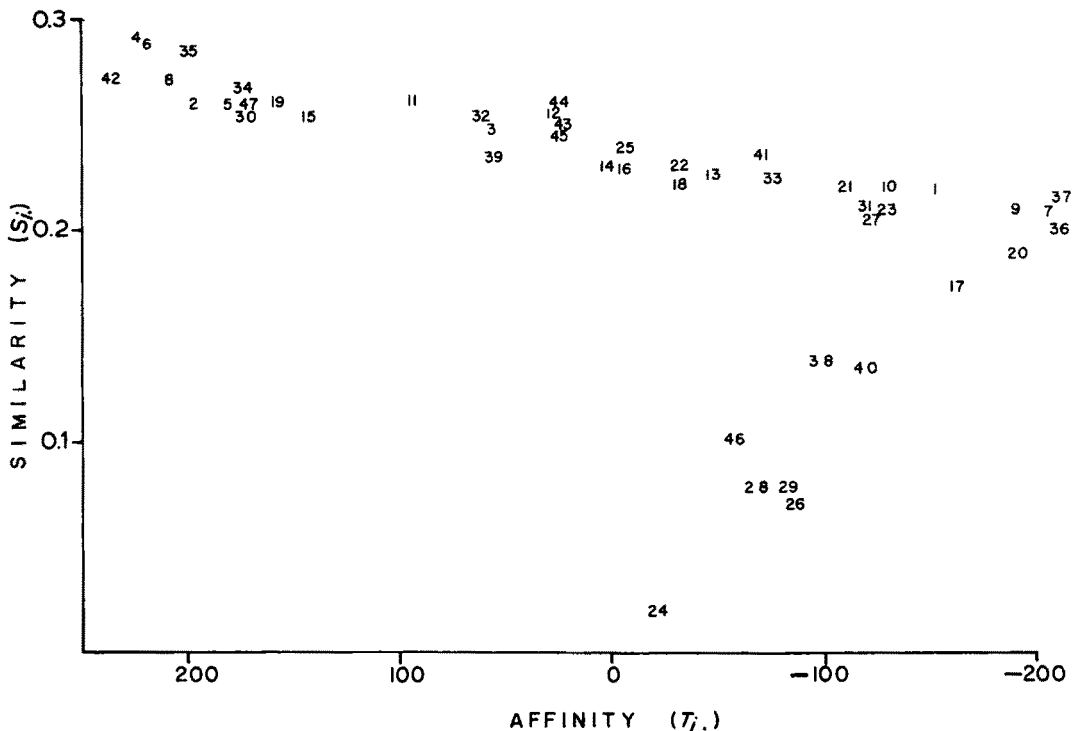


Figure 6. S - T graph of hemlock-white pine-northern hardwood forest of northern lower Michigan (example 14 of Table 1).

'tail' was tested empirically by intentionally sampling additional sites intermediate (bog forest (46), larch swamp (38), and dry jack pine (40)) between the linear part of the graph and the most extreme sites in the 'tail', i.e. the bog (24) and the dunes (26, 28, 29). These additional sites are clearly intermediate along the 'tail' on Fig. 6. Full presentation of the Michigan study will appear in a separate paper.

Comparing landscapes

The examples from field studies show how the method can be used to compare the structuring of unrelated landscapes. For each example, we present two measures of high-order diversity, the mean similarity of all sites (\bar{S}), a measure of β diversity (Whittaker, 1972), and μ diversity. We have also obtained random expectation for \bar{S} and μ using a null model based on a bootstrapping routine (Efron, 1981) and computer simulations (Scheiner and Istock, unpublished). Currently, the best statistic for comparing the mosaic diversities of unrelated landscapes is μ^* ; it varies between 1 and 3.5 in most examples for which Q , R , and F are in the acceptable ranges for reliable behavior of affinity analysis (acceptable ranges derived empirically in Scheiner and Istock, unpublished). Values for Q , R , and F are not within acceptable ranges for the Wisconsin tree community data.

The mosaic diversity of the moss communities on the Hyuck Preserve in New York (Coleman and Istock, 1980) is much greater than for the studies with taxonomically more comprehensive species lists (number 11 in Table 1). The moss S - T graph is linear, lacking any 'tail'. It is possible that taxonomically narrow data tend to produce higher μ values, but there are presently too few comparisons to support this speculation.

The Wisconsin tree data (numbers 12 and 13 in Table 1) provide an example strikingly different from all the others. This case is the only one in which mosaic diversity is not significantly different from randomness (see Appendix). We urge caution in the interpretation of these data, however, because the relation of sample number to species number in the Wisconsin tree data falls in the range of dimensions where affinity analysis is not robust (Scheiner and Istock, unpublished).

Discussion

A web of uncertainties has spread over the part of ecology devoted to the study of diversity (cf. Peet, 1974). Point or α diversity is well measured by various indices, but we do not know what such measurements can tell us about the origin, development, dynamics and structure of natural communities, or about the existence of strong regularities in community processes and structure. It sounds good to say that the central questions are: how are communities created, patterned, and governed? But, it is unclear what, if anything, the measuring of α diversities has to do with such sweeping questions. We certainly have not yet achieved a truly comparative study of ecological diversity. At the level of α -diversity studies, the mere delineation of the object we wish to call 'a community' is operationally unclear. Hence, in part, the ennui we mentioned at the beginning of this paper. Such weariness and uncertainty was expressed by R. H. Whittaker (1972) more than a decade ago.

The development of diversity theory in the past 15 years may have disappointed those who expected the construction in ecology of an area of neat, tight, rigorously formulated relationships as bases of prediction. . . . There is some hazard in applying complex measurements and mathematical treatments to community data, of obscuring rather than clarifying relationships. . . . If diversity is recognized as an evolutionary product, it may cause no surprise that no single measurement serves all purposes.

Little has changed since Whittaker wrote these words.

At the scale of whole landscapes there is usually a visible mosaic, and we used this commonplace observation as a point of conceptual departure for the development of affinity analysis.

Non-randomly structured landscapes

In the Appendix, and in a companion paper (Scheiner and Istock, unpublished), we determine empirical formulae for the null hypotheses of affinity analysis. These formulae allow several types of significance testing. The tests which are now possible include: (1) deviation of a data set from its null (random) expectation, (2) deviation of any single site (sample) from the mode, (3) significant similarity between any pair of sites, and (4) significant affinity between any pair of sites.

All of the field data sets in Table 1, except for the Wisconsin tree data, deviate significantly from their null expectations by more than 3.5 standard deviations (with corresponding null, or random, data structures simulated without constraints on marginal totals). We determined the fiducial limits using the formula for z in the Appendix. All the significant μ^* values were steeper than random. This result indicates that, with the exception of the Wisconsin data, all the landscapes we have analysed have a significant higher-order organization of the sort we refer to as the 'metastructure'. Undoubtedly, experienced plant ecologists would suspect that such a structure exists. Nevertheless, it seems worthwhile to have the methods of affinity analysis to confirm it objectively, and to provide ways of representing, measuring and comparing such a higher-level organization.

A significant μ^* could mean that one or more clusters of similar communities exist within the landscape. We take up the question of clustering of communities versus continuity in community variation below.

Unique aspects of affinity analysis

As in the study of α diversity, the definition of 'a sample' = 'a community' is a difficult and essential one in the study of diversity in regional mosaics. However, when the data are to be used with affinity analysis the problem becomes less severe for two reasons.

First, we can operationally define the sample as *any* area which, though very small relative to the entire landscape, is extensive enough to have the largest abundant life forms of that locale repeated many times within the sample site. Ample opportunity will then exist for the dominant forms to be 'diversified' with respect to rarer and rarer forms of similar size. Smaller forms then too will have ample spatial scope to be repeated and diversified. Thus variation owing simply to statistical fluctuations will be reduced. If such a procedure for delineating sample sites is followed, it is not necessary for the sites to have equal areas. Visibly homogeneous areas of one to a few hectares, and lacking gradients, are suitable for vegetation studies in most parts of the world.

Second, as increasing numbers of samples are brought into the analysis, and the $S-T$ plot emerges, an increasingly well defined and stable 'metastructure' appears and actually accommodates to vagaries of sampling. For example: redundancy, or over-representation, of one sample type does not warp the structure, since all such samples merely plot in the same place. As samples intermediate in species composition between previous samples come into the data, they plot as intermediate between the sites they resemble. Sites which break strongly in their compositional affinity become organized into the 'tail' portion of the $S-T$ graph. The position of a site on the $S-T$ graph is controlled by the proportion of its species common to the whole landscape. The greater this proportion, the nearer the site to the mode and thus the greater the 'affinity' to the landscape. Once enough data exist to define the 'metastructure' clearly and stably, the central statistic for the structuring of the mosaic, μ , and its standardized form, μ^* , are obtained and a comparison of the overall structuring of different landscapes becomes possible.

Despite the present paucity of analysed data sets, we have some strong evidence that μ diversity and β diversity are independent measures of landscape level phenomena. This evidence is shown in Fig. 7, where the measure of β diversity is \bar{S} . For the analyses thus far, μ varies independently of \bar{S} (i.e. the two are uncorrelated), and thus μ measures a different property of landscape organization. Specifically, μ differs from \bar{S} because it is sensitive to all of the details of variation in community composition simultaneously, something which a mean similarity coefficient like \bar{S} cannot be. For instance, the high mean similarity of the Danube meadow samples (9 in Fig. 7) is almost certainly due to the restriction of sampling to a single habitat type. Fig. 7 also suggests that for extensive, presence/absence, community data sets \bar{S} is conservative over a range of approximately 0.22 to 0.27 while μ varies more widely.

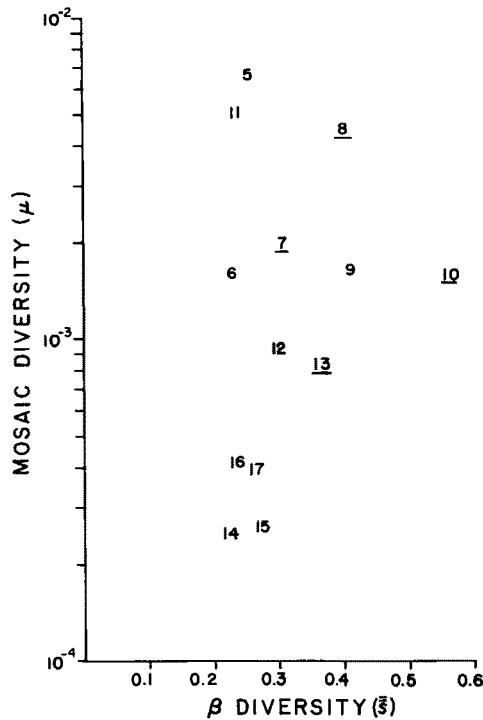


Figure 7. Graph of mosaic diversity against β diversity for all the natural data sets of Table 1. Underlining indicates abundance data. Mosaic diversity (μ) varies independently of β diversity (\bar{S}).

In the future it will be interesting to ask whether there are geographical, climatological, successional, taxonomic, or evolutionary correlates of variation in mosaic diversity. Mosaic diversity should increase over time in regions isolated by lack of dispersal if evolution creates a strengthening of species associations through specific adaptations (including speciation), extinction, and coevolution of species interactions. At present, affinity analysis provides the only way available to search for such regularities.

A simple way to envision the ecological meaning of the S - T graph, and one which suggests that β and μ diversity should be independent, is as follows. Let the number of species in a region be constant. Move the cloud of points in Fig. 1 up and down the ordinate, with its slope unchanged, β diversity changes while μ diversity remains the same. Biologically such change in β diversity

signifies a change in species packing (Whittaker, 1972) which is equivalent to adding or subtracting a constant to all pairwise similarity values. Conversely, when the tilt, μ , is changed without changing β , i.e. mean species packing is left constant, the biological process involved is a change in the degree of structuring among the species associations. Thus, the extraction from vegetation data of the S - T graph and μ — i.e. abstraction into the ‘metastructure’ — rests on biologically motivated concepts and methods.

Ecological and evolutionary ramifications

Community change over any part of the earth is a kind of variation. Affinity analysis emulates the reduction of variation familiar in univariate statistics by establishing representations of central tendency and dispersion for the more complex community data. The ‘tail’ in the S - T plot, of course, introduces a striking non-linear property. Biologically, the method in its presence/absence form treats each species as an evolved genotypic array of monitors, or living sensors, of the surrounding abiotic and biotic environmental variation. With this evolutionary motivation in mind, presence/absence data which give equal weight to a large set of these ‘physiological probes’, while abundance data which represent the species quite unequally and usually less completely, might be expected to provide different kinds of information. Presence/absence data give macroevolutionary diversification at and above the species level the major role in shaping the S - T plot and determining μ . Abundance data necessarily carry much more of a microevolutionary influence derived from both landscape-wide and local expressions of high fitness for particular species. It is interesting that our comparisons, thus far, with the two types of data have shown only modest differences in the positioning of sites within the S - T plot (e.g., Fig. 4). However, all the relevant examples (data sets 6–7–8, 9–10, 12–13 in Table 1) show an increase in μ^* when abundance is included. Taken together, these last two observations suggest: (1) that the internal ordering of sites within the S - T plot is mainly an expression of evolutionary diversification among species and higher taxonomic categories, and (2) that additionally separate μ^* measurements can be used to measure comparatively the overall influence of variations among species in expressed fitness versus macroevolutionary differentiation on the structuring of landscapes when both presence/absence and abundance data are available for the same region.

While we have found discontinuity in several of the natural examples, the rule is clearly that continuity pervades most of the variation in community composition when it is cast in the form of the S - T plot. This observation plus our studies of ordination plots, not presented here, suggest to us that continuity is probably also the rule in the similarity space itself. Hence, the classification and naming of communities to finer and finer levels may never be very useful. A useful classification would arise naturally if there were much clustering and discontinuity in the distribution of sites in the similarity space.

This issue of continuity points up a fundamental difference between our approach and that of Pielou (1975, ch. 5). Her treatment assumes *a priori* that a finite set of vegetational phases exists and that the phases have all been identified prior to a sampling analysis of mosaic patterning. In our case we sample first and then solve empirically for the overall mosaic diversity without knowing beforehand whether the variation in community compositions (phases) is more nearly continuous or discontinuous. It is possible that, in the future, we may be able to use affinity analysis in conjunction with ordination to define the ‘phase’ of any particular site even in complex situations where the mosaic is not completely subdividable on *a priori* grounds, and thus gain access to Pielou’s log likelihood method for testing randomness in vegetation mosaics. Such an approach might provide a check on our use of simulation to achieve statistical inference. Either \bar{T}_i

or \bar{S}_i and \bar{T}_i together can be used as an index for any community relative to the mode of the landscape. Hence, these indices can provide a calibration, an ecological metric, for both successional and ordination studies.

Without a logical basis such as that of affinity analysis and its realization as a method of analysis, the study of large-scale variation and change cannot advance. The human mind falters upon direct inspection of an array of community data. The Michigan data, for example, is a seemingly impenetrable matrix of zeros and ones, nearly 500 by 50 strong, and it is not as large a data set as those we are likely to contend with in the future. More palpable images of internal regional variation come when the primary data is reduced to a diagonally symmetrical matrix of similarity coefficients, say 50 by 50 strong, and various techniques of ordination are available to provide images of internal patterning. At the next level of data reduction with affinity analysis, more easily grasped images and insights emerge in the S - T plot and a small number of descriptive statistical parameters. Quantification of regional diversity, comparisons between data sets, and the search for regularities become possible.

The sorting with respect to the mode in the S - T graph could be of use in matters of conservation by allowing us to express the degree (S - T position coordinates) to which a given site is typical or unusual within its contemporary landscape. Repeated use of affinity analysis with data from sequential sampling through time of the same sites in the same region should also provide information about total change in a landscape mosaic, as well as directions of successional change for specific points in a landscape.

It is interesting how the succession-like chains of communities shown in Fig. 3 for Oosting's data are arranged within the S - T plot. One wonders if major successional trajectories tend to course neatly through the mode of the metastructure, and whether succession does or does not often involve crossovers between such trajectories. The simple internal structures of Fig. 3 may or may not represent dominant successional trajectories, but if they do, the T -axis can provide a quantification of position and perhaps assist in describing motion along paths of succession. Affinity analysis also provides a formal framework for simulations of succession, as well as for simulations of many related ecological processes. Various assumptions about rules or processes operating during the formation, maintenance, and change of structure and diversity of landscape mosaics can now be explored. Explicit descriptions of species dispersal, competition, herbivory, population dynamics, and coevolution, as well as successional transitions, could be studied through computer simulation, with affinity analysis providing the diagnosis of major effects on, and patterns of, landscape structure. Such simulations would proceed by creating dynamical algorithms reshaping the species, or species-abundance, by site matrix, with affinity analysis and ordination then serving to provide a coherent summary of many of the ensuing consequences for the overall organization of a landscape.

Perhaps the most important result we have presented is the identification of mosaic diversity, μ , as a measure of high-level diversity independent of β diversity measured by \bar{S} . Since β diversity should reflect the consequences for community and landscape structuring arising from species turnover along environmental gradients, the independence of μ and β indicates that the similarity and dissimilarity of communities due to gradients does not have a common influence on these two forms of diversity. The relation of μ or μ^* to α diversity remains obscure. To this stage in the development of affinity analysis it has been impossible for us to establish either a theoretical or empirical relationship across this wide conceptual span within the general study of ecological diversification.

We believe that as we go beyond calculation of β diversity in affinity analysis, and use the T -statistic to accomplish a study of the 'similarities of similarities', we leave behind much of the

specific imprint of environment gradients. Or perhaps it would be more accurate to think that we thereby obtain a procedure which solves for a more general form of gradation, a form of gradation summarizing many physiological performances by many species in many ecologically different settings, and encompassing, in some substantial degree, the large set of events and processes which create the floristic structure of a landscape.

Appendix

Affinity analysis can be used for statistical inference at several different levels of comparison: the whole landscape, single sites to the whole landscape, and among sites. Expectations and sampling variances of the indices of affinity analysis were generated by random simulation. The null model assumed, given the number of sites (Q), number of species (R), and matrix filling (F), that all species had an equiprobable chance of being in all sites. There were no constraints on either row or column totals. A regression procedure was then performed to obtain random expectations for \bar{S} , μ , and the other affinity analysis indices corrected for data matrix attributes. A complete description of this procedure will be given in a companion paper (Scheiner and Istock, unpublished).

The results of the random simulation procedure can be used for statistical inference in the following ways. For example, the formula for the random expectation of μ was found to be:

$$E(\mu) = 10 \exp[1.20 + \sqrt{Q}\{-0.64 + (0.073) \sqrt[4]{R/Q}\} + (0.63)(F) - (0.23)(F) \sqrt[4]{R/Q}\} + \sqrt[4]{R/Q}\{-1.31 + (2.06)(F)\} - (4.18)(F)] \quad (\text{A1})$$

and the coefficient of variation was found to be less than 0.3. Thus, a conservative test of whether the mosaic diversity of a landscape is significantly different from random expectation is given by:

$$z > [Obs(\mu) - E(\mu)] / 0.3 E(\mu). \quad (\text{A2})$$

Probability values for z can be obtained from any standard statistics text (e.g., Snedecor and Cochran, 1967). As the distribution of μ is positively skewed, we recommend using 3 standard deviations as a minimum test of non-randomness. Most of the natural data sets in Table 1 have mosaic diversities that are at least $3\frac{1}{2}$ standard deviations greater than random expectation.

Significant deviations of individual sites from the modal communities can be identified by use of the measured values of \bar{T}_i and its variance [$V(\bar{T}_i)$]. For example, in the Australian data set (Fig. 5), $V(\bar{T}_i) = 4148$ and $\bar{T}_{23} = 92$, which is 3 standard deviations away from the modal sites.

One type of comparison among sites is based on the pairwise similarity values. Again, the random simulations provided random expectations for the mean similarity and variance of similarities.

$$E(\bar{S}) = F - (0.16) \sin(\pi F)$$

and

$$E[V(S_{ij})] = 10 \exp[-1.00 - (0.16) \sqrt{Q} - (0.32) \sqrt{R/Q}]$$

giving

$$z = [S_{xy} - E(\bar{S})] / \sqrt{E[V(S_{ij})]}, \quad (\text{A3})$$

where S_{xy} is the measured similarity of any sites x and y . For example, in the Australian data samples $E(\bar{S}) = 0.23$ (Table 1) and $E[V(S_{ij})] = 0.0033$. Thus, sites 1 and 10, although they plot next to each other on the S - T graph (Fig. 5) are not more similar than would be expected at random ($S_{1,10} = 0.20$, $z = 0.52$, $P < 0.61$). A similar procedure can be used to test for significant differences in pairwise affinities.

Acknowledgements

We are grateful to James Teeri, John Cairns, William Etges, David Draper, Thomas Caraco, Ruben Gabriel, Joseph Warnick, colleagues at University of Michigan Biological Station, the University of Rochester, the University of Chicago, the University of Arizona and several anonymous reviewers, for many discussions and suggestions about this work. Dr David Gates generously made the resources of University of Michigan Biological Station available. Numerous improvements in the paper resulted from editing and from suggestions for revision by both Stuart Pimm and Michael Rozenzweig. Financial support for parts of this work came from NSF grants DEB 80-20821 and DEB 80-22182, NIH grant GM 07197, the Hinds Fund of the University of Chicago, the Arco Fund, the Naturalist Ecologist Training Program of the University of Michigan Biological Station, and Ms Judy Scheiner.

References

- Avena, G., Blasi, C., Feoli, E. and Scoppola, A. (1981) Measurement of the predictive value of species lists for species cover in phytosociological samples. *Vegetatio* **45**, 77–84.
- Coleman, B. B. and Istock, C. A. (1980) Ecological pattern in the moss species ensembles of a hemlock–hardwood forest mosaic. *Evolutionary Theory* **5**, 1–27.
- Eberhardt, L. L. (1969) Some aspects of species diversity models. *Ecology* **50**, 503–5.
- Efron, B. (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* **68**, 589–99.
- Etges, W. J. (1984) Genetic structure and change in natural populations of *Drosophila robusta*: systematic inversion and inversion association frequency shifts in the Great Smoky Mountains. *Evolution* **38**, 675–88.
- Faith, D. P. (1983) Assymmetric binary similarity measures. *Oecologia* **57**, 287–90.
- Gauch, H. G. (1982) *Multivariate Analysis in Ecology*. Cambridge University Press, New York.
- Goodall, D. W. (1973) Sample similarity and species correlation. In *Ordination and Classification of Communities* (H. Whittaker, ed.) pp. 105–56. W. Junk, The Hague.
- Griffin, E. A., Hopkins, A. J. M. and Hnatiuk, R. J. (1983) Regional variation in mediterranean-type scrublands near Eneabba, south-western Australia. *Vegetatio* **52**, 103–27.
- Jaccard, P. (1901) Distribution of the flore alpine dans le Bassin des Dranes et dans quelques régions voisines. *Bulletin de la Société vaudoise des Sciences Naturelles* **37**, 241–72.
- Janson, S. and Vegelius, J. (1981) Measures of ecological association. *Oecologia* **49**, 371–8.
- MacArthur, R. H. (1972) *Geographical Ecology*. Harper & Row, New York.
- Mueller-Dombois, D. and Ellenberg, H. (1974) *Aims and Methods of Vegetation Ecology*. John Wiley, New York.
- Nichols, G. E. (1935) The hemlock–white pine–northern hardwoods region of eastern North America. *Ecology* **16**, 403–22.
- Oosting, H. J. (1942) An ecological analysis of plant communities of Piedmont, North Carolina. *Amer. Midl. Nat.* **28**, 1–26.
- Peet, R. K. (1974) The measurement of species diversity. *Annual Review of Ecology and Systematics* **5**, 285–307.
- Peet, R. K. and Loucks, O. L. (1977) A gradient analysis of southern Wisconsin forests. *Ecology* **58**, 485–99.
- Pielou, E. C. (1975) *Ecological Diversity*. Wiley-Interscience, New York.
- Siegel, S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- Snedecor, G. W. and Cochran, W. G. (1967) *Statistical Methods*. Iowa State University Press, Ames, Iowa.
- Whittaker, R. H. (1972) Evolution and measurement of species diversity. *Taxon* **21**, 213–51.
- Wolda, H. (1981) Similarity indices, sample size and diversity. *Oecologia* **50**, 296–302.