Measuring the Structural Similarity
of Two Communities Composed of the
Same Species

Robert W. Poole

Illinois Natural History Survey

Urbana, Illinois 61801*

* Present Address

Biometrics Unit, Warren Hall

Cornell University

Ithaca, New York 14850

Introduction

In studying communities of organisms there may sometimes be a need to compare two communities to determine how similar they are in terms of the pattern of interactions between the species of the communities and the species and the physical variables of the ecosystems. If a model of the changes in abundance of the species of a community with time or distance is applied to another community of essentially the same species in another area, it must first be shown that the structural characteristics of the two sites or communities are similar before the model based on one community can be assumed to be adequate for predicting the abundance changes in the second community. The purpose of this paper is to propose tests for structural similarity and given structural differences to derive empirical measures of the structural differences between the two communities.

Structure is an inclusive definition combining several of the characteristics of a community. Some of these characteristics are:

1. The similarity of the species lists between the two communities. Communities with different species or different numbers of species may automatically be defined as structurally different. However, in many cases it may be useful to make exceptions to this rule for species which are taxonomically different but functionally similar. Rare and occasional species should probably be excluded from consideration.
2. The changes in the abundances of the species of a community are due to both causal factors common to other species of the community and to factors specific to a species. Common factors may be interactions among the species or the effects of the physical variables of the ecosystem on the species' abundances. The

pattern of these common factors is reflected in the observed covariances or correlations in abundance among the species of the community. Similar covariance or correlation matrices for two communities indicate that the same pattern of common factors occurs in both communities.

3. Even if the species lists and the covariance matrices are similar, there can still be differences between the two communities in the mean abundances through time of each species. Even though species may fluctuate greatly in abundance, the species of a stable community usually fluctuate around a constant mean density provided the populations of each species are observed over a long enough period of time.

The many various forms of community classification and ordination usually compare a pair of sites entirely on the basis of a comparison of the two species lists. The first, and to some small extent the third, criterion forms the basis of most studies of community diversity, diversity being defined as the number of species and the relative abundances of each species at a single instant in time. By contrast the structural similarity of two sites or communities in this study will be based on the pattern of interactions occurring in the ecosystem and the overall mean abundances of the species. The purpose of the paper will not only be to derive measures of structural difference, but also to try to relate these measures to actual differences in the underlying patterns of interactions in the two communities. Specifically we would like to measure the contribution of each species or physical variable to the observed difference in structure between the two communities.

The first criterion of community structure is straight forward, but the latter two need to be tested statistically. The test of characteristic three sometimes depends to some degree on equal covariance matrices. Pragmatically it may be a rare event indeed if the structural pattern of common factors were identical in both communities, and so the last sections of the paper discuss the derivation of empirical measures of the structural differences of the two communities.

A Comparison of Two Communities of Fruit Flies

Hunter (1966) studied changes with time in the abundances of the species of Drosophila at several sites in Colombia. Two of these sites were a government pine plantation and a city watershed, both near Bogota and nearly adjacent to each other. The six commonest species of Drosophila were the same at both sites, and these two sites will be designated community one and community two respectively. Although adjacent to each other and containing essentially the same species of fruit flies, the two habitats are quite different. The pine woods is a government protected pine plantation. The watershed, here termed Aqueduct, is a near-virgin area of tropical montane vegetation. Hunter sampled the species at bait traps monthly from September, 1961, to December, 1963, at site one and from November, 1961, to December, 1963 at site two so there were a total of 28 observations on community one and 26 observations on community two. In Hunter's tables the commonness of each species is expressed as its percent frequency of the total number of flies. Also listed is the total number of traps used, total number of flies caught, and the average number of flies per trap. Because the number of traps at each site changed from month to month, the percentages were changed to number of flies based on the average number per trap rather than the total number of flies. Obviously the transformation from percentage to numbers based on the average number of flies per trap results in rounding error and the loss of any information obtainable on differences between traps. Because the number of flies in each trap is not given, there is no way to check for differences between traps. However, because of spatial heterogeneity and aggregated spatial patterns differences in abundance between traps probably existed. The species counts of the six species were subjected to a $\log_{10}(x+1)$ transformation in an attempt to normalize the variables. The estimated covariance matrices based on the number of flies caught each month for both sites are shown in Table 1.

Equality of the Covariance Matrices

The first test of structural similarity is a test of the equality of the two covariance matrices. If $\underline{S}_1$ and $\underline{S}_2$ are the estimated covariance matrices of the first and second communities, the null hypothesis is $H_0: \underline{\Sigma}_1 = \underline{\Sigma}_2$ versus $H_1: \underline{\Sigma}_1 \neq \underline{\Sigma}_2$ where $\underline{\Sigma}_1$ and $\underline{\Sigma}_2$ represent the true population covariance matrices estimated by $\underline{S}_1$ and $\underline{S}_2$. A pooled covariance matrix for the two communities may be formed by first calculating the respective matrices of corrected products and cross-products, $\underline{A}_1$ and $\underline{A}_2$. A pooled covariance matrix $\underline{S}$ is

$$\underline{S} = \frac{1}{N_1 + N_2 - 2} (\underline{A}_1 + \underline{A}_2)$$

where N_1 is the number of observations on the first community and N_2 the number of observations on the second. An approximate test of the null hypothesis of equal covariance matrices is the quantity $\underline{MC}^{-1}$ (Box, 1949).

$$M = (N_1 + N_2 - 2) \ln |\underline{S}| - \left[(N_1 - 1) \ln |\underline{S}_1| + (N_2 - 1) \ln |\underline{S}_2| \right] \quad (1)$$

and

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p + 1)} \left[\left(\frac{1}{(N_1 - 1)} - \frac{1}{(N_1 + N_2 - 2)} \right) + \left(\frac{1}{(N_2 - 1)} - \frac{1}{(N_1 + N_2 - 2)} \right) \right] \quad (2)$$

where p is the number of variables in each community. The quantity $\underline{MC}^{-1}$ is approximately distributed as chi-square with $\frac{1}{2}p(p+1)$ degrees of freedom. This is a large sample statistic and is only an approximation for small samples such as

occurred in the two Drosophila communities sampled. Applying the test to the two Drosophila communities' covariance matrices $|S| = 3.5140$, $|S_1| = .1819$, and $|S_2| = 1.0328$. The quantity $MC^{-1} = 100.5992$ with 45 degrees of freedom. The statistic is highly significant, the null hypothesis is rejected, and it is concluded that the two covariance matrices are significantly different.

Equality of the Mean Vectors

The mean densities of each species in the two communities may be summarized in a vector of the means for each of the two communities,

$$\bar{x}_1 = [\bar{x}_{11}, \bar{x}_{12}, \dots, \bar{x}_{1p}]$$

$$\bar{x}_2 = [\bar{x}_{21}, \bar{x}_{22}, \dots, \bar{x}_{2p}]$$

where, for example, $\bar{x}_{23}$ is the mean abundance of the third species in the second community. We can test the null hypothesis $H_0: \mu_1 = \mu_2$ versus the alternate hypothesis $H_1: \mu_1 \neq \mu_2$ where μ_1 and μ_2 are the true mean vectors represented by the sample statistics $\bar{x}_1$ and $\bar{x}_2$. This test involves calculating Hotelling's T^2 statistic

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$$

The quantity

$$F = \frac{N_1 + N_2 - p - 1}{(N_1 + N_2 - 2)p} T^2$$

has the variance ratio F distribution with p and $N_1 + N_2 - p - 1$ degrees of freedom (Morrison, 1967). The decision rule for a test at the $1 - \alpha$ level of probability is to accept the null hypothesis if

$$T^2 \leq \frac{(N_1 + N_2 - 2)p}{N_1 + N_2 - p - 1} F_{\alpha; p, N_1 + N_2 - p - 1}$$

and reject it otherwise.

Of more interest than the acceptance or rejection of the null hypothesis is the construction of confidence intervals about the two mean differences between the two communities. By utilizing the T^2 statistic simultaneous confidence intervals can be placed about the mean differences in species abundances between the two communities. The construction of these confidence intervals is discussed in Morrison (1967) and Anderson (1958).

Applying the T^2 test to the two Drosophila communities $T^2 = 21.7281$ and $F = 2.0428$. Setting α equal to .05 the 95 percent probability value of F with 9 and 44 degrees of freedom is 2.09. Therefore the hypothesis of equal mean vectors is just barely accepted with 95 percent probability. The observed mean differences were $(.2463 \quad -.0742 \quad .1112 \quad .1314 \quad .0841 \quad -.4473) = (\bar{x}_1 - \bar{x}_2)'$. By far the greatest contribution to T^2 is the relatively large difference in log mean abundance of D. viracochi between the two communities.

The T^2 statistic assumes that the two covariance matrices Σ_1 and Σ_2 are equal. If the two covariance matrices are not equal, and they probably will not be in most situations, the test of the equality of mean vectors can be greatly biased and misleading. Ito and Schull (1964) have investigated the large sample properties of the T^2 statistic given that Σ_1 and Σ_2 are not equal. In general if the sample sizes N_1 and N_2 are equal, the inequality of the covariance matrices has no effect upon the size of the type I error provided the sample sizes are

relatively large. If $N_1 > N_2$ the true α probability is increasingly larger or smaller than its given value. In the example in this paper $N_1 = 28$ and $N_2 = 26$. Because N_1 and N_2 are almost equal, the T^2 statistic in this case is probably fairly reliable. Anderson (1958) discusses this multivariate analogue of the Fisher-Behrens problem in more detail.

It might be noted in passing that the quantity $D^2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$ may be used as a measure of the overall difference in mean vectors between the two communities.

Empirical Measures of Structural Similarity

Because of the perversity of nature the two covariance matrices defining the pattern of interactions in their respective communities will, in all probability, rarely be equal. The question then becomes not are they different but how different are they? Even if there are significant differences between the two covariance matrices, it may still be possible to extrapolate from one community to the other if the differences are not great. An intuitively appealing measure of structural inequality between the two communities is the statistic $\underline{MC}^{-1}$ (Eqs. 1 and 2). In the calculation of M the determinants $|S|$, $|S_1|$, and $|S_2|$ are used. These determinants are called generalized variances and are multivariate analogues of the univariate variance estimate s^2 . The larger the determinant the greater the dispersion of the sample points in p -dimensional space. In the example $|S_1| = .1819$ and $|S_2| = 1.0328$ indicating that the dispersion of the sample points is far greater in the second community than in the first. In general the smaller the generalized variance, the larger the covariances between the species and hence the greater the dependence between the changes in abundance of the species variables. The statistic $\underline{M}$ is somewhat analogous to the information statistic used in the information method of plant community classification. Theoretically both communities can have similar generalized variances

but different covariance patterns. However, if the patterns of covariances are different in the two communities, the dispersion of sample points in the pooled data will increase and the generalized variance of the pooled data will be increased by the differences in covariance patterns between the two communities. If the covariance matrices of the communities are identical, the value of $\underline{M}$ is zero, its value increasing as the differences between the two community covariance matrices increases.

In some cases it may be useful to compare two structure difference measures $\underline{MC}^{-1}$ to see if they are significantly different. Because $\underline{MC}^{-1}$ is approximately a chi-square variable with $k = \frac{1}{2}p(p+1)$ degrees of freedom, the ratio of the two chi-square values

$$F = \frac{\chi^2_1/k_1}{\chi^2_2/k_2}$$

has an F distribution with k_1 and k_2 degrees of freedom. Suppose $\underline{MC}^{-1}$ for a second comparison had been 34.8000 with $k = 25$ degrees of freedom. Then $F = (100.5992/45)/(34.8000/25) = 1.6399$ with 45 and 25 degrees of freedom. With a type I error of $\alpha = .05$ this value of F is not significant and it is concluded that the two indices $\underline{MC}^{-1}_1$ and $\underline{MC}^{-1}_2$ are not significantly different.

Although $\underline{MC}^{-1}$ is a convenient measure of the structural difference between two communities, it does not indicate which species or physical variables of the ecosystems are responsible for the observed differences between the two communities. One obvious way of determining which variables are responsible for the observed difference is to compare the variances and covariances of the two covariance matrices $\underline{S}_1$ and $\underline{S}_2$. However, there are $p(p+1)/2$ different elements in a covariance matrix. Each species or physical variable is represented by its variance and covariances with the remaining $p-1$ variables. In addition the covariances of a variable with the remaining $p-1$ variables are not independent, i.e. the observed covariance

between two variables may, and usually is, determined in part by each variable's covariance with other variables of the ecosystem. Therefore each variable is represented by p measures of its contribution to the overall difference between the structure of the two communities and these p indices are all correlated to one degree or another. If p is large, the problem of too many indices and their interdependence can become overbearing. Therefore we should try to replace these $p(p+1)/2$ indices by a smaller number of indices such that if it is not possible to represent the contribution of each species to the structural difference observed by a single figure, at least the measures associated with each species should be uncorrelated and arranged in order of lessening importance. In order to derive such indices, the factor analysis model will be used.

The factor analysis model states that the observed standardized value of a random variable, z_1 , can be expressed as a linear compound of a series of n common factors common to all the species and a specific factor specific to each variable, species or other variable, i.e.

$$z_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + d_1U_1$$

where F_1 are the common factors and U_1 the specific factor. These factors are mathematical constructs only and have no direct biological meaning. They may be derived in an infinite number of ways and the factors extracted depend on the optimizing criterion used. The variance of the specific factors will be called the unique variance and is equal to d_1^2 . The variance of the random variable due to common factors is termed the communality of the variable, h_1^2 . Therefore the unique variance is $d_1^2 = 1 - h_1^2$. The total variance of a variate is

$$\sigma^2_1 = a_{11}^2 + a_{12}^2 + \dots + a_{1m}^2 + d_1^2$$

and the covariance between variables i and j

$$\sigma_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + \dots + a_{im}a_{jm} \quad (17)$$

The parameters a_{ij} are termed the factor loadings and can be represented by the $p \times m$ matrix L. Alternatively the correlation matrix may be used in the estimation of the factor loadings instead of the covariance matrix. The unique variances are a diagonal matrix U. The sample correlation matrix R can be written in matrix notation as $\underline{R} = \underline{L}\underline{L}' + \underline{U}$. The pattern of loadings depends upon the criterion of the factorization of the correlation or covariance matrix. Maximum likelihood factor analysis was used utilizing a program written by Jöreskog (1967). The estimates of L and U are found by an iterative algorithm too lengthy to discuss in this paper. There are an infinite number of solutions to the likelihood equations and an additional constraint is needed to provide a unique solution. In most maximum likelihood factor analysis programs the constraint is the purely mathematical one of making the matrix $\underline{J} = \underline{L}'\underline{U}^{-1}\underline{L}$ diagonal.

Even though the factors arrived at by the maximum likelihood factor analysis should not be construed as having any direct homologues with actual variables of the ecosystem, they are indicative of the underlying pattern of common ecological factors. Therefore if the factor analysis is carried out on two different correlation matrices and the matrices are similar, the pattern of factor loadings should be similar. The greater the differences between the structural characteristics of the two communities, the greater the differences between the loadings.

The factors of the factor analysis model have two important characteristics in terms of using the model to generate measures of the differences in covariance

structure between the two communities. The factors of the model are orthogonal to each other, i.e. the factors are uncorrelated. The covariance of a variable with a second variable is difficult to interpret because the second variable is in turn correlated to one degree or another with the remaining variables of the ecosystem. However, this difficulty disappears if the correlation between a variable and a factor, i.e. a factor loading, is used as a measure of difference because the factors are uncorrelated. Secondly the factors of the factor analysis model are estimated so that the first factor accounts for the maximum amount of the variance in the sample possible. The second factor is estimated to account for the maximum amount of the residual variance possible with the constraint that it be orthogonal to the first factor. Subsequent factors are derived in the same manner. Therefore, if the factors and factor loadings are used, either directly or indirectly, to measure the structure differences between the two communities, the problem of correlation between variable indices is removed and in addition the $p(p+1)/2$ covariance measures may be greatly reduced in number if a few mathematical factors adequately reproduce the original covariance or correlation matrix.

Given estimates of $\hat{L}$ and $\hat{U}$ the sample correlations between the variables may be estimated as

$$\hat{R} = \hat{L}\hat{L}' + \hat{U}$$

where $\hat{R}$ is the matrix of estimated sample correlations. The number of common factors derived depends upon the number of common factors necessary to adequately regenerate the observed correlation matrix R . In the course of the derivation of the common factors, adding an additional common factor usually explains an increasingly smaller amount of residual variance so that the factors are derived in order of importance. An approximate test of the importance of each factor is the size of the successive latent roots of the correlation matrix.

An approximate test of the hypothesis that m common factor adequately regenerate the observed correlation matrix is the statistic $\frac{U}{m}$

$$U_m = \left[N - 1 - \frac{2p + 5}{6} - \frac{2m}{3} \right] \sum_i \sum_j \frac{(r_{ij} - \hat{r}_{ij})^2}{d_i d_j}$$

where r_{ij} is the observed correlation between variables i and j and $\hat{r}_{ij}$ the estimated correlation regenerated by $\hat{LL}' + \hat{I}$. This statistic is approximately distributed as chi-square with $\frac{1}{2}[(p-m)^2 - p - m]$ degrees of freedom (Bartlett, 1954).

Three possible determining factors in both ecosystems measured by Hunter during the two and a half years of the study were the minimum temperature recorded during the month, monthly rainfall, and monthly rainfall lagged four months. Therefore measurements on these three variables were included in the analysis along with the six species. The factor loadings for both communities with five common factors are listed in Table 2. The estimated communalities are also listed. In both cases three common factors adequately regenerated the correlation matrices as tested by the chi-square test. The five factor matrices were used as measures of structural similarity. However, the higher factors three, four, and five are probably more strongly influenced by random and sampling error than by real structural differences and the interpretation of higher factors becomes increasingly difficult as their importance becomes minor.

Utilizing the factor-loading structures two criteria of similarity between communities can be established. First, are the successive estimated communalities for increasing numbers of common factors similar, and secondly are the estimated factor loadings estimated from both community correlation matrices similar. An examination of variable two in the factor loading matrix for the second community shows that in this community rainfall loads extremely heavily on factor two, i.e. .9957. Because the correlation between factor two and rainfall is so high, rainfall may be regarded

for all practical purposes as common factor two in the second community. This is not true of common factor two in the first community where the loading of rainfall is only $-.25013$. Evidently the importance and response of the species to rainfall is different in the two communities. However, it may be that the difference is merely an artifact of the optimizing constraint used and in fact in this case probably is (see below). A similar, but less graphic, situation occurs for factor two. In community two the loadings of minimum temperature on common factor one is very high, $-.8668$, but less so in the first community. As a purely subjective decision it might be hypothesized that the interactions in the second community are much more dependent on the two physical variables rainfall and minimum temperature than they are in the first community. These are only subjective hypotheses which may be figments of the estimation procedure and in any case would have to be proven experimentally in the field. There is a very large difference in loadings for Drosophila bandeirantorum on the first factor between the two communities.

A comparison of the vectors of communalities leads to essentially the same conclusions. In the first community the communalities are highest for the species, but in the second community the high communalities are associated with the physical variables rainfall and minimum temperature. Again as a subjective hypothesis it may be that in the pine woods the interactions among the species are very strong, but at the aqueduct locality the influence of the physical variables on the species abundances is considerably more important, submerging to some extent the effect of interspecies interaction. However, again the differences in communalities may be due to the estimation procedure and not to true differences between the structural patterns of the two communities.

Procrustes

The differences in factor loadings and communalities between the two communities are difficult to interpret in a meaningful way. These measures point out only the differences or similarities between factor structures and thus, hopefully, between the correlation and covariance matrices. However, as already suggested, the differences observed may not be real but only artifacts of the estimation procedure. The maximum likelihood equations have an infinite number of solutions and in order to derive a unique solution the constraint that $\underline{J} = \underline{L}'\underline{U}^{-1}\underline{L}$ be diagonal was imposed. However there is no biological reason for preferring this solution over any other solution that can be achieved by a rotation of each of the factor axes through some given angle. All solutions are equally correct. The question then becomes; given the factor loading matrices, here termed $\underline{A}$ and $\underline{B}$, how close can the two matrices $\underline{A}$ and $\underline{B}$ be made by a rotation of the factor axes of one of the factor loading matrices? The problem can be expressed as

$$\underline{AT}^* = \underline{B} + \underline{E}$$

where $\underline{T}^*$ represents a transformation matrix of the angles of rotation of the factor axes of $\underline{A}$. The matrix $\underline{E}$ represents a matrix of residuals or differences between the matrix $\underline{AT}^*$ and $\underline{B}$ and provides measures of how closely the two factor structures can be matched. In this way the arbitrariness of $\underline{J}$ being diagonal is removed and the residuals should indicate the degree to which each variable is responsible for the observed differences in community structure. In factor analysis this type of analysis is termed Procrustes (see Harman, 1967). The transformation matrix $\underline{T}^*$ is solved for by least squares, i.e. by minimizing $\text{tr}(\underline{E}'\underline{E})$ so that

$$\underline{T}^* = (\underline{A}'\underline{A})^{-1}\underline{A}'\underline{B}$$

The matrix $\underline{T}^*$ is then normalized by columns to yield $\underline{T}^* \underline{D} = \underline{T}$ so that $\text{diag}(\underline{T}' \underline{T}) = \underline{I}$. The matrix $\underline{A} \underline{T}$ is then computed which in a loose sense may be regarded as a least squares fit of $\underline{A}$ to $\underline{B}$ under the restriction that $\text{diag}(\underline{T}' \underline{T}) = \underline{I}$.

The matrix of interest is $\underline{E}$, the discrepancies of the least squares fit of $\underline{A}$ to $\underline{B}$ (Table 3). For the first factor the only really large discrepancy is caused by variable 5, Drosophila bandeirantorum. For factor two, rainfall and D. viracochi have reasonably large residuals. The differences in the remaining three factors are not as clear-cut, but for factor four D. pseudoobscura has quite a large residual. Therefore as a tentative set of hypotheses to be proven by subsequent field work it is possible that the major difference between the two communities is the response of Drosophila bandeirantorum to the ecological factors making up common factor one, and of lesser importance a different importance of rainfall and the response of D. viracochi to the ecological factors represented by common factor two. The supposed major differences in the importance of rainfall and minimum temperature between the two communities are apparently artifacts of the estimation procedure as the differences disappear when the Procrustes transformation is used.

Because three factors adequately regenerated the estimated covariance matrices in both communities, the use of Procrustes residuals has reduced substantially the number of comparisons made from $(9)(10)/2 = 45$ to $(3)(9) = 27$. In addition the factors, because they are orthogonal and in order of decreasing importance, greatly simplify the interpretation of the differences between the estimated covariance we have observed.

Discussion

The structure of a community as defined in this paper consists of three fundamental characteristics of an ecosystem; 1) the species of the community, 2)

have not yet been asked. Perhaps it is a delusion to believe that given the limited amount of data available in studies such as the Drosophila experiment anything more than intuitive hypotheses and empirical results can be derived. However, it should be remembered that even the "limited" amount of data in this study represents hundreds of hours of field work, and conceivably in such complex situations this type of data will be the only information that realistically be gathered.

The factor analysis model has been used to empirically measure the effect of each variable on the overall observed difference between the structural makeup of two communities. In this sense these empirical measures of structural difference are akin to measures of distance in classification problems. Because of the indeterminacy of the factor analysis model, however, inferences about causality suggested by the factor loadings are purely subjective and should not be assumed to be more than that. Although the factor analysis approach is helpful in unraveling the complex skein of interactions, it does not set the interactions out in a complete and totally comprehensible form. In fact, it probably would be impossible to comprehend such a pattern even if it were discovered because of its very complexity. In the last analysis, however, the true test of the measures will be whether they prove to be useful or not.

2) the mean abundances of the species, and 3) the interspecific interactions and the effects of the physical variables of the environment on the abundances of each species. If different species occur in the two communities, the two communities are, by definition different except in those cases where a species in one community has the same functional position as a second species in the second community. Perhaps the most interesting comparison between the two communities consisting of the same species is between the skein of interactions between the abundances of the species and the physical variables of the ecosystem. In earlier sections of the paper a measure of the overall difference between the structure of the two communities has been proposed, and in addition a second group of empirical measures have been given which try to measure the relative contribution of each species or physical variable to the overall difference observed.

The $\underline{MC}^{-1}$ statistic and the Procrustes residuals suffer from two defects. Firstly the measures are not directly biologically meaningful. The use of these measures generates hypotheses which, hopefully, can be tested in the field. Secondly all of the empirical measures of structural difference are based on the multivariate normal distribution to one degree or another. In practice ecological variables such as abundance are rarely normally distributed, although the use of appropriate transformations may help to reduce the non-normality of the data somewhat. In any case these measures should be used with some care, as should any statistic based on normal theory.

The Procrustes residuals matrix and $\underline{MC}^{-1}$ pose the same dilemma as the more familiar measures of species diversity such as $\underline{H}$. Although the diversity measurements can be computed, and in some cases two diversity measurements can be statistically compared, the investigator is still left with the problem of making biological sense out of something which is essentially a mathematical construct. Diversity measurements have rightfully been termed answers to which the questions

Summary

The primary purpose of this paper is to propose empirical measures of the structural differences between two communities of plants or animals composed of the same species. Structure is defined to consist of; 1) the species in the community, 2) the pattern of interactions as represented by the covariance or correlation matrix of successive observations on each species, and 3) the mean abundances of each species in each of the two communities. Statistical tests are proposed for testing whether the covariance matrices and the vectors of mean densities for each community are equal and empirical measures of the differences between the covariance matrices and mean vectors are proposed. Given unequal covariance or correlation matrices the factor analysis model is used to derive empirical measures of the degree to which each variable of the ecosystem is responsible for the observed differences in the pattern of interactions in each community. These tests and measures were applied to data gathered by Hunter (1966) on the abundances of six species of Drosophila censused monthly over a period of approximately two and a half years in two adjacent, but different habitats near Bogota, Colombia. The two covariance matrices were significantly different indicating different patterns of interactions in the two Drosophila communities.

Literature Cited

- Anderson, T.W. 1958. An Introduction to Multivariate Statistical Analysis.
John Wiley and Sons, New York.
- Bartlett, M.S. 1954. A note on the multiplying factors for various chi-square approximations. *Journal Royal Statistical Society, Series B*, 16:296-298.
- Box, G.E.P. 1949. A general distribution theory for a class of likelihood criteria. *Biometrika* 36:317-346.
- Harman, H.H. 1967. Modern Factor Analysis. Second Edition. University of Chicago Press. Chicago.
- Hunter, A.S. 1966. High-altitude Drosophila of Colombia (Diptera: Drosophilidae). *Annals Entomological Society America* 59:413-423.
- Ito, K., and W.J. Schull. 1964. On the robustness of the T_0^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. *Biometrika* 51:71-82.
- Jöreskog, K.G. 1967. UMFLA -- a computer program for unrestricted maximum likelihood factor analysis. Research Memorandum 66-20. Princeton, New Jersey: Educational Testing Service.
- Morrison, D.F. 1967. Multivariate Statistical Methods. McGraw-Hill Book Co. New York.

Table 1. The estimated covariance matrices among the six species of Drosophila at Pine Woods (Community 1) and Aqueduct (Community 2). The estimates are based on 28 and 26 monthly observations respectively. The variables are: 1) Drosophila pseudoobscura, 2) D. bandeirantorum, 3) D. immigrans, 4) D. viracochi, 5) D. mesophragmatica, and 6) D. gasici. The species counts were subjected to a $\log_{10}(x+1)$ transformation.

Community 1						
variable	1	2	3	4	5	6
1	.17185					
2	.10553	.09226				
3	.05585	.05014	.07685			
4	.04803	.06398	.02153	.13216		
5	.05020	.03509	.01764	.05488	.17190	
6	-.01957	-.00460	.02942	-.00092	.00861	.10353
Community 2						
variable	1	2	3	4	5	6
1	.16127					
2	.00528	.09569				
3	.05577	.03144	.07158			
4	.03054	.02988	.01842	.08944		
5	.02001	.04432	.02027	.03301	.17923	
6	-.06820	.04212	-.00427	.03431	.03657	.20774

Table 2. The factor-loading matrices for communities one and two with five common factors. The communalities in each case are also listed.

Community one - Pine Woods						
Variable	Common factors					Communality
	1	2	3	4	5	
1	.12307	.03949	.03948	.29866	.69047	.58
2	.16897	-.25013	-.38469	-.27427	.04563	.32
3	.62096	.14823	.26456	.18696	.19728	.55
4	-.62824	.22486	-.55943	-.09289	.30681	.86
5	-.81812	.39004	-.42255	0.0	0.0	1.00
6	-.42115	.20998	-.39993	.41945	-.08654	.56
7	-.85765	.19987	.47381	0.0	0.0	1.00
8	-.55146	-.75295	-.09312	0.0	0.0	1.00
9	.00175	-.09029	.02466	.86052	-.07722	.76
Community two - Aqueduct						
Variable	Common factors					
1	-.18168	-.11056	.23122	-.06815	-.28073	.18
2	-.00551	.99574	-.09202	0.0	0.0	1.00
3	-.86678	.09024	.49046	0.0	0.0	1.00
4	.86626	.09663	.49017	0.0	0.0	1.00
5	.18924	-.35496	-.17774	-.77719	.06681	.80
6	.50469	-.11688	.19013	-.35101	.09391	.44
7	.44952	-.36075	-.20459	-.12816	-.40545	.56
8	.40927	.15390	-.51354	-.28123	.01604	.54
9	-.16426	-.13692	-.44288	-.29534	-.35212	.45

Table 3. The matrix $\underline{E}$ of differences between the matrix $\underline{B}$ and the least squares fit of $\underline{A}$ to $\underline{B}$, $\underline{A}\underline{A}$.

Factor	1	2	3	4	5
Variable					
1	.04352	-.07492	-.16782	-.04264	.22638
2	-.12609	.44634	.16229	-.16431	-.37874
3	-.14903	.16176	.19261	-.34208	.17631
4	.09429	.04621	.19664	.50394	-.06315
5	-.65624	-.08779	-.28164	-.09349	.01249
6	.07234	.07571	.24644	.37844	.10427
7	-.03719	.35830	.03129	.18862	.12205
8	-.27638	-.11259	.15387	.03613	.18898
9	.01972	.14464	-.11767	.28747	.03645