1980 In

The Recovery Process in Damaged Ecosystems

J. Cairns (ed) Ann Arbor Science Pub. Inc

Ann Arbor

CHAPTER 6

MULTIVARIATE QUANTIFICATION OF COMMUNITY RECOVERY

Stephen A. Bloom

Department of Zoology
University of Florida
Gainesville, Florida 32611

INTRODUCTION

As the ecological impact of human activities on natural communities has increased, so has the interest in quantifying and understanding changes and patterns of changes within those communities. Given a distinct and temporally limited perturbation, there are three major questions:

1. Does the community recover or change states (*sensu* Sutherland 1974)?
2. If the community recovers, how rapid and how complete is the recovery?
3. What is the path of the recovery after the perturbation?

The form of these questions implies a reasonable knowledge of the community before and after perturbation. The existence of such data will be an underlying assumption of the technique presented here.

Any analytical technique, if it is to be of any real utility, must possess certain features. Chief among these are the maximum use of available information and the ability to be easily grasped. If the user of a model or a technique does not have an intuitive feel for the assumptions, advantages and limitations of the tool that he is using, the potential for misuse is high. While the technique presented here uses relatively sophisticated mathematical procedures, these procedures are readily available as computer packages [Bloom et al. 1977] and are intuitively obvious.

The usual types of quantitative biotic information gathered about natural communities are the number of species, the species composition (the taxa represented), and the number of individuals per species in some defined sample size. There are several techniques addressed to recovery problems which use one or more of these informational types. For example, island biogeography [MacArthur and Wilson 1967; Simberloff 1974] deals with species composition and numbers, while multivariate approaches [Allen et al. 1977] deal with all three informational types.

The technique presented here is an extension of the multivariate approach combined with a geometrical and hypothesis-testing statistical analysis. The basic technique is simple and conforms closely to general ecological concepts of community stability and resilience [Holling 1973]. The similarity (dissimilarity) of all pre- and postperturbation samples are represented as intersample distances in a three-dimensional ordination system. The preperturbation samples are defined as a cluster and the statistical envelope which encloses the cluster is calculated. Sequentially, each postperturbation sample is tested to see whether it lies within the envelope (statistically recovered) and how far it lies from the nearest edge of the envelope (the distance to recovery). The distance to recovery is then plotted through time to reveal patterns of multistable states, successional stages, speed to recovery and the path of recovery in ordination space.

DESCRIPTION OF DATA SET

The development of the multivariate technique requires the use of a data set featuring normal variation, a reasonable time frame, a distinct perturbation and before and after quantitative monitoring of the community. Such a data set appeared to exist in two studies of a sandy intertidal habitat at Courtney Campbell Causeway in Old Tampa Bay, Florida. The first study [Bloom et al. 1972] represented a 9-month survey of the site beginning in September 1968. The second was an intensive 24-month monitoring of the same site after virtual defaunation due to Red Tide in August 1971 [Dauer and Simon 1976]. Because of differences in sampling techniques, sieve sizes and taxonomic expertise, the data from the two studies were not comparable and the studies do not represent a legitimate before-and-after perturbation data set. However, to develop the technique, the second year of data from the Dauer and Simon [1976] study was considered to be "preperturbation". The results presented here are thus not a legitimate test of the technique but are simply an example of how the analysis works. Given the definition of preperturbation, the analysis should show a distinct trend towards "recovery" if the technique is valid.

DESCRIPTION OF ANALYTICAL PROCEDURE

Ordination

The initial data matrix (species by sample with counts per square meter filling the matrix) is reduced and converted to an n-dimensional representation of intersample distances by a principal coordinate analysis (PCOR) [Gower 1966; Sneath and Sokal 1973]. PCOR requires a unit-variance standardization [Whittaker 1973] but does not specify any particular data

transformation. The output of a PCOR includes a table of samples by principal axes. The axes are ranked in descending order by the amount of variance in the original data "explained" by the axes. At the present stage of development, only the first three axes are considered. By plotting the values of the various samples on the three axes, a three-dimensional representation of the intersample distances is generated.

In the data set used here, the 153 species by 24 samples (monthly samples) was so treated initially using a square-root transformation. The results on principal axes I and II are presented in Figure 1. Samples 1 and 12 are defined as preperturbation but are actually the second year of monitoring. Sample 13 represents the community state immediately after perturbation and each successive sample represents the community state in ordination space one month later.

Cluster Definition

The preperturbation samples are defined as a cluster. By analyzing the positions of these preperturbation samples, a 95% rejection envelope can be calculated. The envelope defines all points in ordination space which are statistically indistinguishable ($\alpha = 0.05$) from the preperturbation samples. Thus if a postperturbation sample should fall into the envelope, it has statistically recovered, i.e., can no longer be said to be statistically distinct from the preperturbation community state. Obviously, the cohesiveness of the preperturbation cluster will directly affect the probability of any point being found to be statistically distinct from the cluster. If the variance is low between preperturbation samples, the envelope will enclose a relatively small volume of space and a randomly distributed point will most likely fall outside of the cluster. On the other hand, if the preperturbation variance is high, a large volume will be enclosed and virtually any community state will be indistinguishable from the preperturbation community state. The value of replicate sampling in realistically defining the preperturbation state is obvious.

The amount of information and, hence, the importance of any axis is theoretically related to the amount of variance "explained" by the axis. The coordinates of samples on an axis can be weighed by the amount of variance "explained" by that axis by multiplying the coordinate by an axis weighing constant. The axis weighing constant is simply the amount of variance accounted for by that axis divided by the largest amount of variance accounted for by any axis (always axis I), i.e., axis I will always have a constant of 1.0. This procedure is essentially equivalent to weighing each axis by the fraction of total variance accounted for by all axes under consideration. The ratio between the weighing constants will be the same by either method of calculation and the relative adjustment of intersample distances will be the same. Once the coordinates have been appropriately weighed, the rejection envelope can be calculated.

There were two major approaches taken in defining the rejection envelope. The first was to calculate the three-dimensional distances from each

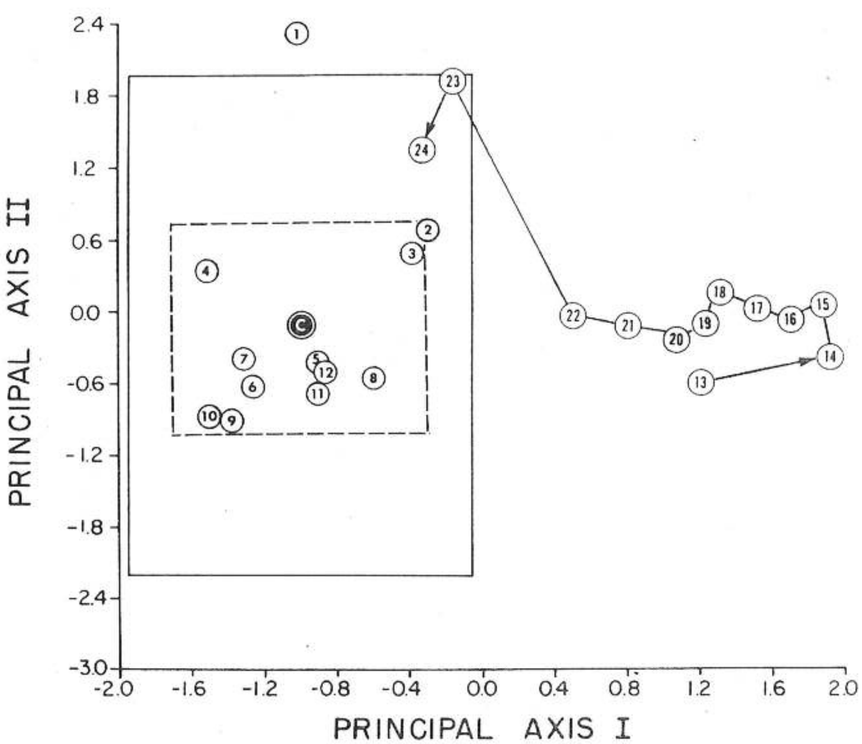

Figure 1. Ordination of a benthic infaunal community after natural defaunation. Samples 1 through 12 represent preperturbation and Samples 13 through 24 represent postperturbation (see text for details).

preperturbation sample to the centroid of the cluster. Using the parametric approach outlined below, a spherical rejection envelope could be calculated. Since this method ignores cluster shape, it was not surprising that the results were confusing. For this reason, the spherical model was rejected.

Although it is theoretically possible to standardize the axes in such a way as to make the spherical model more reasonable, a more direct approach was taken. The set of coordinates of the preperturbation samples on any axis can be characterized by their mean and standard deviation. By rearranging the "t" test for a single observation [Simpson et al. 1960], the following expression results:

$$x_c = \overline{X} \pm \frac{|t_{(df=N-1:\alpha=0.05)}| \times S}{\sqrt{\frac{N}{N+1}}}$$

By substituting the mean $\overline{X}$, standard deviation S, the number of samples N for any axis and the appropriate t value, the critical values (above and below the mean) which will result in a rejection of the null hypothesis can be calculated. The pair of these critical values for the three axes represent the coordinates of the corners of the 95% rejection polygon. A rejection polygon more closely approximates true cluster shape than any arbitrarily selected geometric solid. The box of solid lines in Figure 1 represents the rejection polygon for the first two axes of the example presented here. If the coordinate of a postperturbation sample falls between the critical values for any axis, the postperturbation sample is within cluster for that axis. If the sample is within cluster on all axes, then the postperturbation point is statistically indistinguishable from the preperturbation state and the community can be said to have recovered.

The distribution of points within the cluster (Figure 1) along axis II appears to be bimodal and is decidedly not normal. Since the test outlined above is based on parametric statistics, the data violates the basic assumption of normality. To counter this legitimate objection, a nonparametric analog to the above procedure was created.

Along any axis, the distance between two coordinates can be represented as a line segment. The line segments between all possible pairs of preperturbation samples for a given axis can be calculated and form the preperturbation data set. In Figure 2, the preperturbation sets for axes I and II are represented by the set of bold line segments parallel and closest to those axes respectively. For any given point on the same axis, line segments between that point and the coordinates of the preperturbation samples can be calculated. If the point is the coordinate of a postperturbation sample, the set of line segments form the postperturbation data set. The postperturbation data sets are represented in Figure 2 by the sets of thinner lines parallel and further from the respective axes. By ranking the line segments of pre- and postperturbation data sets together and performing a Mann-Whitney "U" test [Siegel 1956], the null hypothesis that there is no difference between the data sets can be tested. If a point lies outside the cluster, as does the postperturbation point on axis 1 in Figure 2, the line segments of the postperturbation set will be longer than those of the preperturbation data set and the null hypothesis will be rejected. Conversely, if the point should lie within cluster, as it does along axis II, the sets of line segments will be statistically indistinguishable. The appropriate test is then a one-tailed Mann-Whitney "U" test.

To determine the rejection polygon, an iterative procedure is used. A point is arbitrarily chosen remote to the centroid coordinate (four times the standard deviation away from the centroid coordinate for a given axis). That point is then tested nonparametrically as outlined above. If it is found to be outside of cluster, a second point is chosen at half the distance from the first point to the centroid coordinate. If the first point is found to be within cluster, the second point is placed further from the centroid coordinate by a similar algorithm. By repeating the process until some arbitrary small distance (here set to 0.0001 axis units) no longer exists

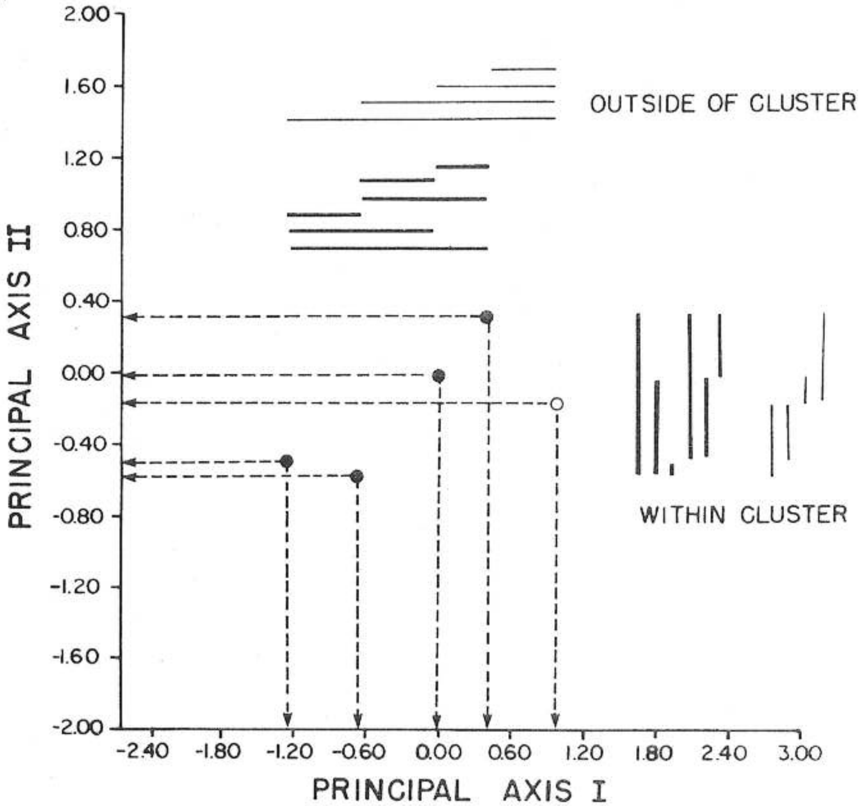

Figure 2. Nonparametric statistical method of determining whether a point is within cluster by creation and comparison of intersample line segments (see text for details).

between the first and second points, the nonparametric equivalent of the critical values of the rejection polygon for the axis is generated. The procedure is repeated for each axis and the three-dimensional rejection polygon is created. The dashed line in Figure 2 represents the nonparametric rejection polygon for axes I and II for the data set examined here.

Distance to the Rejection Envelope

A line can be generated from each postperturbation point in three dimensions to the centroid. Once the rejection polygon is defined, the coordinates of the point at which the line between the postperturbation point and the centroid pierces the nearest surface of the polygon can be determined. By application of an extended Pythagorean theorem, the distance between the postperturbation point and the pierce point can be

calculated. If the postperturbation point is found to be within the cluster on all axes, a line is projected from the centroid, through the point and back to the nearest surface of the polygon. The pierce coordinates are determined and the distance between the postperturbation point and the pierce point is set as a negative to indicate that the distance is back to the surface of the polygon. In the event that the postperturbation point should fall exactly at the centroid, the distance would be the perpendicular to the nearest surface of the polygon.

The reason for the three-dimensional limitation of the technique lies in the calculation of the pierce coordinates. There are six critical zones and six critical lines defined for each axis. Therefore, for three axis there are 1728 possible locations for any point relative to the rejection polygon and the centroid. In its present form, the relative location of each postperturbation point is found by a hierarchical search procedure. Expansion to four or more axes would greatly complicate the programming and dramatically increase the computing cost of the analysis.

Once the distance from each postperturbation point to the nearest surface of the polygon ("distance to recovery") is known, these data can be plotted against time. In Figure 3, the solid line represents the parametric result and the dotted line represents the nonparametric result. As was evident from the relative volumes of the polygons in Figure 2, the nonparametric analysis is less likely to show recovery. However, there is no real choice between the two types of analyses. In that the parametric assumptions are not likely to be met, the nonparametric approach is the valid statistical procedure.

Availability of Programs

Principal Coordinate Analysis is supported by many computer centers and is available as part of a large package [Bloom et al. 1977]. The program for the calculation of the rejection polygon, the pierce coordinates, and the plotting data was written in FORTRAN IV for an Amdahl 470 (compatible with IBM 360/370). A listing of the program is available from the author upon request.

DISCUSSION

As can be seen in Figure 3, the community analyzed here does show a distinct trend towards recovery (as it should given the definition of the preperturbation cluster as being really the second year of recovery). There is an initial divergence from the preperturbation state (most likely due to the presence of several normally rare species which were unusually numerous in the second month after perturbation). There is a gradual fall towards recovery up to month 11 and then a minor divergence at month 12. The analysis thus appears to be a meaningful way to express recovery. The speed of recovery is expressed by the general slope of the falling line. If there

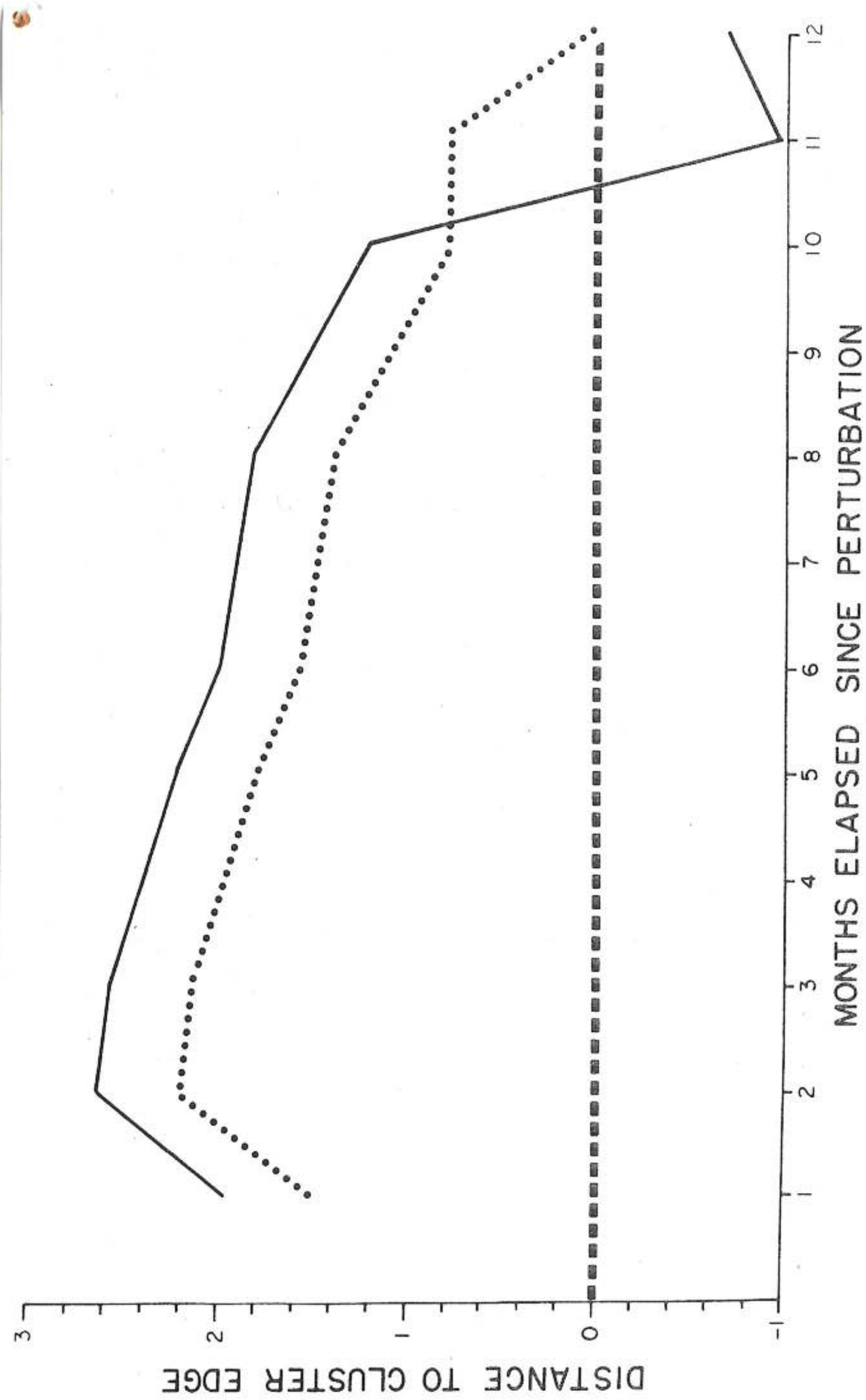

Figure 3. Recovery analysis of a benthic infaunal community after natural defaunation using a square-root transformation with parametric and nonparametric statistical techniques (see text for details).

were multistable points, the line would plateau away from the preperturbation cluster, i.e., a second cluster would form in ordination space. If a series of successional stages were encompassed within the general recovery path, a staircase effect would result. Thus the slope and shape of the line can express meaningful concepts.

The analysis can also be profitably used with experimental data. Control data can form the "preperturbation" cluster while the experimental data can be regarded as the "postperturbation" data. The response of an experimental community to perturbation could thus be monitored and the impact of the experimental perturbation with regard to community alteration can be quantified.

The limitations of the proposed technique should be kept firmly in mind. Comparable before-and-after quantitative data are absolutely necessary. Although the first three axes of a PCOR should account for the majority of variation in the data, for many data sets they do not. The extension of the analysis to more than three dimensions is therefore desirable. On theoretical grounds, a principal component analysis (PCA) would seem to be superior in that the axes' rotation could be solely based on the preperturbation data. The postperturbation samples could then be introduced into the axes system after a coordinate transformation [Sneath and Sokal 1973]. In this manner, the rejection polygon would be only influenced by the preperturbation data and would overestimate the actual cluster shape by the smallest amount possible. In a trial run, however, a PCA obliterated any meaningful pattern in the data. Presumably, the noted space-distortion of a PCA [Rohlf 1968] creates more problems than are cured by excluding the postperturbation data from consideration in the formation of the rejection polygon.

In that data transformation is not specified in running a PCOR, a valid concern is how much variation in the final result can be generated by using different transformations. In Figure 4, the results of analyzing the same data set using no transformation, a frequency standardization (the number of individuals of a species at a station divided by the total number of individuals of all species at that station), logarithmic, square-root and cube-root transformations [Boesch 1977] are presented. The various data transformations do little to alter the general pattern of recovery and the last three transformations yield virtually identical results. The analysis thus appears to be quite robust and is relatively insensitive to even major alterations in the initial data matrix.

The technique presented here should be of value in assessing environmental changes, in monitoring of communities under chronic and acute perturbations and in helping analyze experimental perturbation systems. Alterations can be made in the various analytical subsections of the technique such as ordination, the number of axes used, the nonparametric test utilized and the method of calculating distance to recovery. However, the basic framework of representing intersample dissimilarities as projected distances, defining a rejection region based on the preperturbation data, and expressing the distance to recovery through time is not only intuitively simple but places questions of community recovery in a testable statistical frame.

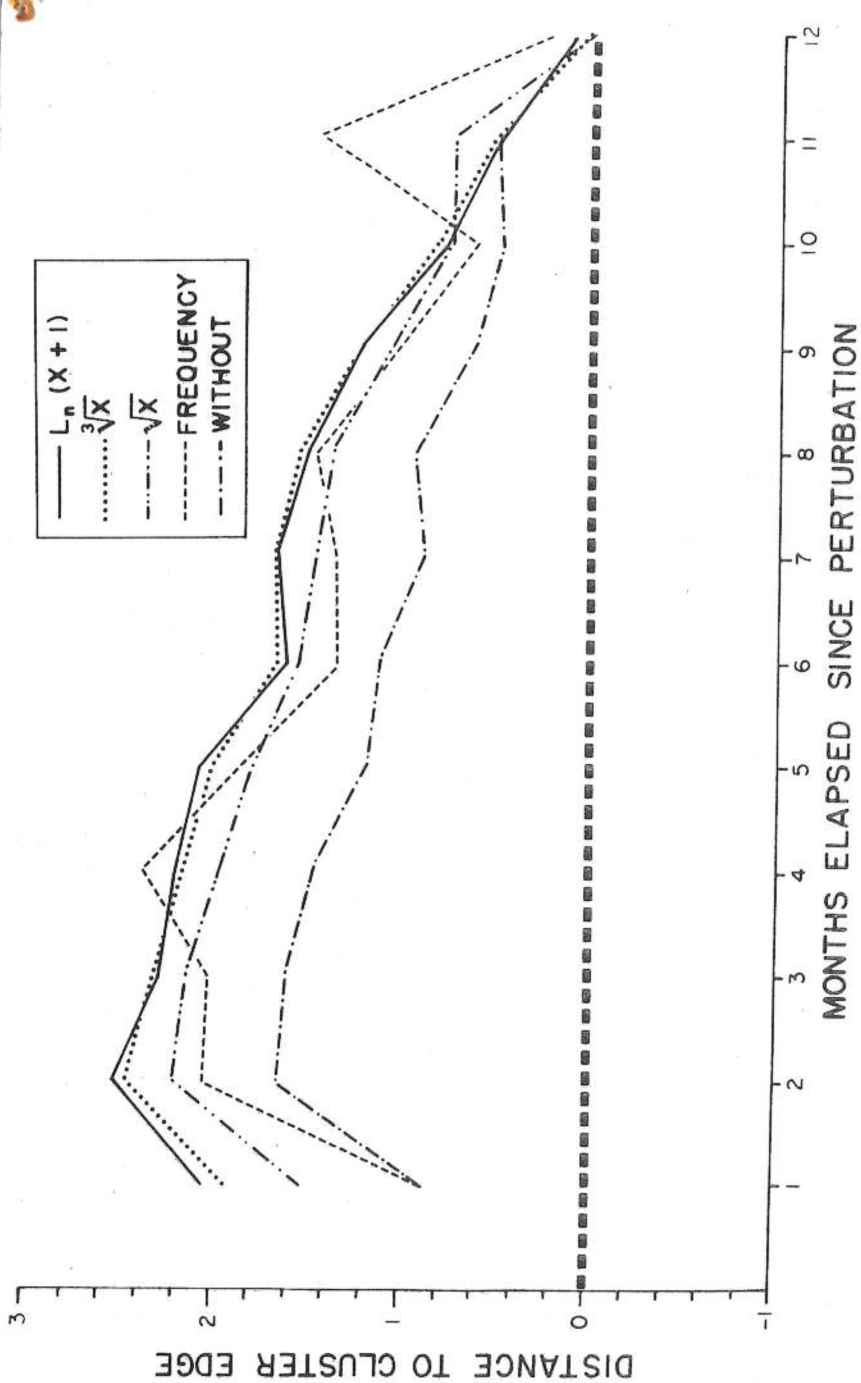

Figure 4. A comparison of recovery analyses of a benthic infaunal community after natural defaunation using a nonparametric statistical technique and five different data transformations.

CONCLUSIONS

A multivariate, nonparametric statistical and geometrical technique that converts before and after perturbation data into a meaningful representation of the distance to recovery through time is described. An example is analyzed and the limitations and areas for improvement are specified. The technique can be used in both descriptive and experimental systems and can be used to quantify recovery, successional transitions and the existence of multistable states.

ACKNOWLEDGMENTS

I thank Dr. Joseph L. Simon for making the data set available, Dr. Stuart Santos for providing a sounding-board, Mr. Robert Smith for his cogent criticisms and suggestions and the University of Florida Department of Zoology for making computer and support funds available.

REFERENCES

Allen, T. F. H., S. M. Bartell and J. F. Koonce. "Multiple stable configurations in ordinations of phytoplankton community change rates," *Ecology* 58:1076–1084 (1977).

Boesch, D. F. "Application of numerical classification in ecological investigations of water pollution," Spec. Sci. Report 77, VIMS; Ecol. Res. Series EPA-600/3-77-033 (1977), 115 pp.

Bloom, S. A., S. Santos and J. G. Field. "A package of computer programs for benthic community analyses," *Bull. Mar. Sci.* 27:577–580 (1977).

Bloom, S. A., J. L. Simon and V. D. Hunter. "Animal-sediment relations and community analysis of a Florida estuary," *Mar. Biol.* 13:43–56 (1972).

Dauer, D. M., and J. L. Simon. "Repopulation of the polychaete fauna of an intertidal habitat following natural defaunation: species equilibrium," *Oecologia* 22:99–117 (1976).

Holling, C. S. "Resilience and stability of ecological systems," *Ann. Rev. Ecol. Syst.* 4:1–25 (1973).

MacArthur, R. H., and E. O. Wilson. *The Theory of Island Biogeography* (Princeton, NJ: Princeton University Press, 1967), 203 pp.

Rohlf, F. J. "Stereograms in numerical taxonomy," *Syst. Zool.* 17:246–255 (1968).

Siegel, S. *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill Book Company, 1956), 312 pp.

Simberloff, D. S. "Equilibrium theory of island biogeography and ecology," *Ann. Rev. Ecol. Syst.* 5:161–182 (1974).

Simpson, G. G., A. Roe and R. C. Lewontin. *Quantitative Zoology* (New York: Harcourt, Brace & World, 1960), 440 pp.

Sneath, P. H. A., and R. R. Sokal. *Numerical Taxonomy* (San Francisco: W. H. Freeman & Company, 1973), 573 pp.

Sutherland, J. P. "Multiple stable points in natural communities," *Am. Nat.* 108:859–873 (1974).

Whittaker, R. H. *Ordination and Classification of Communities* (The Hague: W. Junk B. V. Publishers, 1973), 737 pp.