PLANKTON COMMUNITY STRUCTURE: EVALUATION OF
ANALYTICAL METHODS USING MODEL COMMUNITIES

Final Report

Willis B. Hayes

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Department of Zoology
University of Georgia
Athens, Georgia 30602

Report Date - July, 1976

PREPARED FOR THE
U. S. ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
UNDER CONTRACT AT(38-1)852

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

The following pages are an exact
representation of what is in the original
document folder.

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately-owned rights.

I. Manuscripts Submitted
for Publication

(Submitted to Biometrics, May, 1976)

*preprint
removed*

Some Sampling Properties of the Fager
Index for Recurrent Species Groups

Willis B. Hayes
Zoology Dept.
University of Georgia
Athens, Georgia 30602

INTRODUCTION

The general purpose of this project was to provide information that might be of use to biologists who are investigating the effects of effluents on plankton. Although most applicable to non-persistent pollutants, such as thermal effluents, the methodology may be applied in some degree to nondegradable pollutants as well.

An impact analysis of this sort seeks to answer one basic question: is the effluent causing a significant change in the planktonic community? A "significant" change may be taken to be one that is obvious in a well-designed sampling program, and one that is expected to persist for at least several years. This kind of analysis obviously requires a baseline study, followed by periodic monitoring. The detection of a change in the plankton is not always easy, however, since even a good baseline study will usually uncover large fluctuations in biomass with time and gradients in species distributions. These are normal characteristics of natural planktonic communities. Detection of a change, due to an environmental insult, therefore depends on the discovery of one or both of the following: (1) a greater degree of variability in the species lists, biomass, productivity, etc. than could be accounted for by natural phenomena; (2) a persistent gradient in the distribution of certain species, terminating at the point of maximum impact, which was not present in the baseline study but appeared later.

The first approach is much more difficult, since it is rare to have a clear-cut idea of the magnitude of "natural" variability. Thus,

investigators have tended to rely on the second. Two techniques which have been used widely in the past are the search for indicator species and the study of species diversity. The value of diversity studies is presently a matter of some controversy among biologists. Certainly a broad arsenal of tests is available (Pielou, 1975, provides a good review), and, properly conducted, these tests can provide much useful information about the community. It is probably unwise, however, to rely on diversity as the sole measure of the "health" of a community. This is particularly true in the case of plankton, which can demonstrate large diversity changes without any obvious environmental perturbation taking place (Cassie, 1967).

Indicator species are useful, when they occur and can be identified. As impact studies increase in numbers in the future, a greater reliance on indicator species can be expected, if only because manpower is limited, and it is easier to monitor one species than many. Yet, relying on indicator species is equivalent to putting all of one's eggs in a single basket. There is no guarantee that the fluctuations in numbers of the indicator will always be a precise analog of the health of the entire community.

The obvious approach is to do as many kinds of analyses of the data as possible. In this connection, I do not suggest the elimination of either of the above sorts of studies; I would like, however, to point out the potential advantages of association analysis in addition to them.

Association analysis, or affinity testing, is a measure of the degree of similarity between two or more objects, each of which is described by a set of attributes of some sort. The nature of the objects and attributes will depend on the type of study being done. If the objects are sampling

11

sites, each described by a species list, then two stations will have a high degree of affinity if they have a large percentage of their species in common. If the species are the objects, and each species is defined by a set of locations where it occurs, then two species will have a high degree of affinity if they tend to occur in the same localities.

If one is grouping species, caught at a series of sampling sites, it is simple to assign each group an "occurrence index" for each station. Thus, if group A tends to occur only near the outfall, then one has, in effect, defined an indicator group. If this group is monitored, instead of a single indicator species, then much of the stochastic variability associated with any of the individual species in it will be damped out, and trends in the occurrence of the group will become more obvious. If group B occurs only at some distance from the outfall, then the members of that group are immediately pinpointed as being particularly sensitive to the effluent.

It is possible, though more risky, to make assumptions about certain members of the group based on data collected for other members. A group will only occur, and persist, if its members are showing intraspecific interaction, or if they are reacting similarly to some factor in the environment. A copepod and its favorite diatom food source might co-occur, as might two diatoms having similar half-saturation constants for nitrogen, and so on. Bush and Welsh (1972) give an application of this kind of analysis

If one has a series of sampling sites, beginning at the outfall and proceeding to some presumably unpolluted site, then it is possible to use the entire species list as an indicator, by grouping the stations. If the stations near the outfall tend to have different species than the

stations further from it, then those stations strongly influenced by the outfall will form a distinctive group that should persist through time.

If there is no particular outfall effect, then the groups will be transient, and will not relate to the order of stations along the transect.

One problem with affinity tests is that there are a very large number of them in the literature. Greig-Smith (1964), Pielou (1969), Orloci (1975), and Goodall (1973) review some of them. Many appear to have been developed for a specific study, then rarely, if ever, used again. Others have proved useful in a variety of applications. There is however, no strategy for choosing one test over another for a particular application (Pielou, 1969). In most cases, the sampling distributions of the various affinity indices are not even known, and thus their relative biases with different sample sizes cannot be assessed. Goodall (1973) has shown that there are large differences between certain of the indices in this respect.

The purpose of the present study was, therefore, to provide some information about the relative effectiveness of certain representative affinity tests, in circumstances that would be analogous to those encountered in an impact investigation on plankton. In such a study, real data cannot be used, if the purpose is to show the efficiency of a test at detecting groups with known characteristics; artificial data must be employed in which those characteristics are built in. For that reason, I have used a distribution model to generate the data.

METHODS

The Plankton Community

The present study was based on the following general procedure: distribution characteristics were established for a community of phytoplankton species, and a generating program was then written to create the community within the confines of a sampling grid. Samples were taken from the grid (both randomly and in certain specified patterns) and enumerated qualitatively (e.g., only the presence or absence of species at the sampling locations were recorded). The resulting species lists were analyzed by the species-affinity techniques described below.

The original intention was to establish a "complete" community of plankton, that is, microplankton, macrozooplankton and macrophytoplankton. It was later realized, however, that this was both impractical and unjustified for the purposes of the present project. As mentioned above, the object of the exercise was to evaluate methods that might be of use in routine investigations of the effects of power plant effluents on plankton. To the author's knowledge, no such studies attempt to collect all of the members of the planktonic community on a routine basis. The taxonomic analysis of such a sample series would be formidable, requiring the services of numerous specialists for a long period of time. Moreover, it is by no means certain that existing sampling gear can collect all of the plankton in an unbiased fashion, even on a presence-absence basis. For instance, Hamner, et al. (1975) have shown that numerous species of

zooplankton, hand-collected by divers, are quite common in certain marine environments but are rarely, if ever, taken in nets in those regions. This indicates that conventional studies of the effects of effluents and pollutants on some zooplankton communities may be highly biased. Zooplankton were therefore eliminated from the present study.

Microzooplankton and microphytoplankton are also in an ambiguous position. Although they can probably be sampled in a more unbiased fashion than can the macrozooplankton, their taxonomic identification is much more difficult. Small protozoans are often lumped together as "monads", then generally ignored, although their energetic importance in the community may be considerable (Pomeroy, 1974). Certain groups like the tintinnids, radiolaria and foraminifera have been studied, but are not usually included in environmental impact surveys.

The community in the present study is therefore intended to include the groups such as diatoms and dinoflagellates that are normally included in standard phytoplankton samples. A set of 112 species was used, the number being determined by the capacity of the computer in certain of the analytical programs. This number is, however, not unrealistic in terms of the species counts obtained in various studies of nearshore marine plankton (e.g. Strickland, et al., 1970). The species exist only as numbers in the present study, and no attempt was made to relate particular concentrations, distributions, etc., to those of particular species in the literature. The range of such values, however, is chosen so as to be realistic in terms of actual phytoplankton collections.

The "area of study" is diagrammed in Fig. 1. For simplicity, it is assumed that the only environmental perturbation is a thermal outfall located at grid square 100. Square 1, at the maximum distance from 100,

represents an "undisturbed" square, i.e. one where the temperature variability is indistinguishable from that which would occur if the outfall was not present. The phytoplankton species are assumed to be distributed homogeneously within each square at any given instant. Dimensions are not specified for the squares, since this will depend on the excess heat of the outfall water as compared to the ambient water temperature, on the amount of current-induced mixing in the water, and on the water stratification (which will be partially dependent on bottom topography). For instance, a constant longshore current from upper left to lower right would deflect the thermal plume; this could be compensated for by distorting the grid, in the same way that a net could be stretched. This will not affect the analysis, as long as every grid square has the same probability of being sampled, whatever its shape.

Distribution of the various phytoplankton species is accomplished through a combination of four variables:

- 1) The mean concentration of the species (cells/liter);
- 2) The variability in this concentration with time (corresponding to the species dispersion or patchiness);
- 3) The qualitative response of the species to the thermal gradient (unaffected, positively affected, or negatively affected);
- 4) The quantitative response of the species to the gradient (i.e., the degree to which it is affected, if it is affected at all).

Community Generation Model

It should be emphasized strongly that the "model community" created by the community-generation computer program with the above data as input is not a dynamic systems model. Such a model would not be appropriate for this study. Systems models are deterministic, in that they model the behavior of a set of components which have known interconnections and known environmental forcing functions. The present model is not trying to evaluate the degree of change of the various planktonic species over time; rather, it assumes that such change will take place and models it stochastically, through the random number generator of the computer. This results in a generalized distributional picture of the community, which varies with time. The conclusions drawn from this model are valid for the real world, as long as the variability of the model is at least as great as the variability of a natural phytoplankton community; the data for the various species were chosen to produce this effect.

The stochastic character of the community-generation model is an important feature. There would be little purpose in sampling a static model, analyzing the resulting species lists with affinity tests, and then trying to generalize the conclusions to a real sampling program where the community is undergoing random variation. The philosophy behind the present study was that, if the affinity tests could not detect significant built-in affinities in the model community, while that community was undergoing random variation with time, then the tests would probably be of little use in a real sampling program.

Considerable time was wasted during the earlier phases of this study in attempts to generate the artificial community according to the protocol discussed in the original grant proposal. In this approach, species would be loaded into a large three-dimensional matrix in the computer

according to their mutual affinities. A random sample of these "compartments" would produce a series of species lists that could then be analyzed. The disadvantages of this procedure are twofold. First, "affinity" is quantitatively undefinable except in terms of some affinity index. If species affinities are "pre-set", then one is necessarily using a particular index to define them. This biases the later analysis towards that index. Second, the loading procedure requires an enormous amount of computer storage space, and, since samples were often drawn at random, the storage space must be randomly accessed (i.e., tape or disc cannot be used). The size of the matrix, and the number of species that it contains, are therefore limited by the central memory capacity of the computer.

A much more efficient procedure was developed when it was realized that much of the above-mentioned storage matrix would never be used--all of the matrix would never be included in the samples. By assigning each species the four distribution variables mentioned above, it became possible to generate the samples rather than the entire sampling universe. This required comparatively little storage space. The actual procedure is as follows, for the community-generation program SLIST:

It is assumed that the phytoplankton sample is taken by a vertical net tow or plankton pump, and consists of 1 m^3 of water filtered or pumped. The actual sample examined in the laboratory is a small fraction of this; usually a 50 ml aliquot is drawn and the cells are settled in a settling tube and counted. The probability of one or more cells occurring in the aliquot is therefore related to the concentration of cells in the original m^3 of water. If the cells are homogeneously distributed within this m^3 , then the probability of occurrence in the aliquot is:

$$p = 1.0 - (.99995)^c$$

where c represents the concentration in cells/liter. The probability approaches zero when the cell concentration approaches zero and is essentially 1.0 at concentrations greater than 10^5 cells/liter. This probability is referred to as the "abundance probability", or p_{ab} .

The actual concentration of cells of a particular species is calculated in the program from the mean cell concentration and the variability. Randomly distributed cells, which follow a Poisson distribution, will have a variance equal to the mean; evenly distributed cells will have a lower variance, and aggregated cells will have a higher one. Concentration is calculated as:

$$c = (R)(S) + \bar{x}$$

where S is the standard deviation and R is a normally distributed random number between 0 and 1, with a mean of 0.5. S and $\bar{x}$ for each species are given as input data.

This procedure is adequate for those species which are unaffected by the presence of the thermal outfall. Those which are affected by it may be either positively affected (more numerous near the outfall) or negatively affected (less numerous near the outfall). The effect will be dependent on the sensitivity of the species to the outfall, and on the distance from it. That is, a very sensitive species, living near its maximum temperature, might be so affected by the elevated water temperature that it could occur nowhere except at the control station. A less sensitive species might show an exponential drop-off as one moved from the control station to the outfall, but could still occasionally be found at station 100. The reverse might be true of outfall-enhanced species (like certain of the blue-green algae), which might be found

only at the outfall station.

The program therefore modifies p_{ab} by a second probability p_{dist} , representing distance from the outfall, or the control station, depending on whether the species in question is outfall-enhanced or outfall-inhibited. It is calculated as:

$$p_{dist} = e^{-rD}$$

where r is a sensitivity coefficient given as input data for each species and D is the distance of the sampling site from the outfall or the control station. The magnitude of this value is shown in Table 1.

At any given sampling site, therefore, the total probability of finding a given species will be:

$$p_{tot} = p_{ab} \times p_{dist}$$

Some species will have a very small probability of being captured at that station (due to low intrinsic abundance or to the effect of the outfall or to both) and others will have a high probability. The make-up of the species list for that station is determined by comparing p_{tot} for each of the species to a random number between 0 and 1 chosen for that station. If p_{tot} for a species is greater than that random number, then the species is assumed to be "caught." If sampling continues for more than one day (replicate samples at the same locality), new random numbers will be chosen for each station the next day.

Table 2 shows the input data for each of the 112 species. These numbers remained constant throughout this study; the various tests were made by sampling at different sets of stations or for different periods of time.

Affinity Tests

In the present study, the original intention was to investigate a broad spectrum of these methods, but this proved to be impractical in terms of programming time. Many of the older methods had never been programmed, and in fact existed only as isolated formulae. An index formula by itself is of little use, and must be combined with some sort of grouping procedure in order to be ecologically useful. Attention was therefore focused on methods which (a) existed as working programs in the literature, or could be obtained as such from the authors; (b) showed some versatility in their applications, and (c) were representative of the different groups of affinity tests set up by Pielou (1969). This narrowed the choices considerably, and work was begun on three of the most promising procedures.

The first method is the "recurrent group" technique developed by Fager (1957). Using the terminology of Pielou (1969), it is "reticulate, divisive, and polythetic." The procedure for forming the species groups is given in detail in Fager (1957), with a modified index formula given in Fager and McGowan (1963). Basically, it forms the largest possible groups of species in which there is significant affinity between all possible species pairs within the group. The primary program for delineating the groups (REGRO) has been supplemented in the present study by two others: AFFIN uses a Monte Carlo procedure to set an objective level for the affinity index which will indicate significant association. LOCATE uses the groups calculated by REGRO as input, and determines the tendency of each group to occur at each sampling site.

Recurrent group analysis (or the "Fager procedure") uses only presence-absence data, but that is not a handicap here, since that is

all that is generated. The program itself is rather long, but runs quickly. Running time is greatly increased, however, if the affinity index break-point is low, since this increases the number of species pairs that must be examined. Values for the index less than 0.4 could not usually be handled in the present study, since the number of calculations would often exceed 10^7 in such cases. The Fager procedure has been used in a variety of applications, and its sampling distribution has been investigated (see Hayes, 1976, ms attached).

The second procedure used was the sum-of-squares agglomeration method developed by Orloci (1967). This is a "hierarchical, agglomerative, polythetic" procedure. Each object is represented as a point in s -dimensional space, where s is the number of attributes available to describe any of the objects. The points are projected onto the surface of a unit hypersphere centered at the origin, and the affinity between any two of the objects is then a function of the Euclidian distance between their two respective points, on the two-dimensional surface defined by their position vectors. Orloci's original method used straight-line Euclidian distance, but for the present study I wrote the program DISTS to calculate arc distances, along the hypersphere circumference, between the points. This eliminates some of the objections raised by Beals (1973) to the use of the method. Once the distances between object-points are known, the points can be formed into clusters by minimizing the sum of squared distances between every point and the group's centroid. The result is a dendrogram, or trellis diagram, in which groups of points are succesively enlarged until all points have been included. This has the advantage of showing, graphically, the relation of each point or group of points to every other one. It has the

disadvantage of leaving the group boundaries undefined; one must decide subjectively what constitutes a significant group.

The last program on which work was done was the probabilistic similarity index developed by Goodall (1966) and modified by Legendre (1973). This is a very versatile program which can accept almost any sort of data, qualitative or quantitative. The procedure is discussed in the above reference, and will not be repeated here, since I was not able to obtain useful results with the program sent to me by Dr. Legendre. Although the program would run, the answers produced indicated that errors were being made in its operation, and neither I nor the members of the computer center staff could locate them. An inquiry was made to Dr. Legendre, but no reply was received. It should be noted, however, that the program has certain features which tend to make it undesirable. First, as written it is IBM-specific, and will not run on a CDC machine (A CDC Cyber 70 was used for all of the other programs in this study). Second, it is much longer than any of the other programs, perhaps a price that must be paid for its versatility. Third, the runs that were made cost in excess of \$20 each, which makes it a bad program for use in a comparative study that requires many comparative runs. I personally feel that the program may have merit for plankton analysis (Legendre (1973) used it in that application), but I am unable to produce any evidence supporting this assumption.

Late in the course of the study, a publication by Frenkel and Harrison (1974) was discovered in which a series of phytosociological procedures were tested, much as in the present study. Collected data were used, however, rather than computer-generated model data. Some of their programs, however, would provide a useful extension of the present

work at some future time.

Experimental Procedures

The experiments conducted in the present study fell under three categories:

- 1) A series of successive sampling runs were made on different days at the same sampling sites. This provided an estimation of how much the stochastic variability in the community-generation program would affect the species lists.

- 2) The Fager and Orloci procedures were tested comparatively on species lists collected on given days at different numbers of sample sites. This permitted a determination of where the different methods begin to become effective, with minimal data. That is, how many samples are necessary for each method in order to get unambiguous separation of the outfall-influenced stations?

- 3) Species lists were collected from stations along the grid diagonal from station 1 to 100, on a single day and on five successive days. Species groups and station groups were both extracted, using both the Fager and Orloci procedures. The multi-day sampling averaged out some of the random variability, permitting a comparison between the two tests of the group compositions and the degree to which the groups reflected the influence of the outfall as opposed to the day-to-day stochastic variability of the model.

The specific procedures for these three steps are outlined below:

Series 1

An early version of the SLIST program was used, containing 189 species. The data for the individual species were essentially the same

as those shown in Table 2, except that some of the species were replicated. Sampling was done at stations 1, 12, 23, 34, 45, 56, 67, 78, 89, and 100 on a single day. A similar series was made on each of 8 succeeding days. The total number of species collected are shown in Table 3, along with the sampling variability.

Series 2

The input data from Table 2 were used. Five experiments were done, in each of which the sampling was done on a single day, but at different locations.

1) Four sampling stations were chosen from the grid in Fig. 1. The diagonal of stations from 91 to 10 was excluded, and two stations were chosen at random on either side of this diagonal (i.e. the region toward the outfall and the region away from it). The Orloci procedure was then applied to the stations to form them into clusters. Program AFFIN was used to calculate the correct level of the Fager affinity index, and REGRO was done to get recurrent groups of species. Note that the Orloci procedure clusters stations while the Fager procedure forms groups of species. In order to compare the two, it is necessary to get the opposite information from each of them as well, that is, species clusters from ORLOCI and station groups from REGRO. This is most easily accomplished by reversing the data. The program FREV was written to do this; it converts the basic data set from the form in which a station number is followed by a list of species to a form in which a species number is followed by the list of station numbers at which that species is found. The two tests, ORLOCI and REGRO could then be applied again, on the reversed data set.

As a comparison with the randomly selected stations, the four "best" stations were also selected for analysis. These consist of the four stations at which the effects of the outfall, or the absence of such effects, would be the most apparent, namely the outfall station (100) and the control station (1). Since four stations are required, a random choice was made between stations 2 and 11 (which are equally far from 100) and between stations 99 and 90 (which are equally far from 1). The two affinity tests were then applied, as above, to both the normal and reversed data sets from these four stations.

2) This experiment was the same as the previous one, except that 8 stations were sampled. For the "best" set, stations 1, 2, 11, and 12 were compared with stations 99, 100, 89, and 90.

3) In the third experiment, 16 stations were selected. In the random set, 2 stations were selected from the 91-10 diagonal, 7 were chosen from the region below the diagonal and 7 were chosen from the region above it. The "best" set consisted of stations 1, 2, 11, 3, 12, 21, 13, and 22 as opposed to stations 100, 99, 90, 98, 89, 80, 88, and 79. The analysis was conducted as in the previous experiments.

4) 32 stations were selected. The random set contained 2 from the diagonal and 15 from each side of it. The "best" set contained stations 1, 2, 11, 3, 12, 21, 4, 13, 22, 31, 14, 23, 32, 24, 33, and 5 as opposed to stations 100, 99, 90, 98, 89, 80, 97, 88, 79, 70, 87, 78, 69, 77, 68, and 96. The analysis was carried out as above.

5) A set of 64 stations was selected, but not analyzed except to note the number of species collected. The curves for number of species caught vs. number of stations sampled are shown in Fig. 2. It is apparent that the size 64 sample would not produce any new information,

although it would use up more computer time than all of the previous samples combined.

Series 3

This series consisted of a number of experimental runs in which samples were collected from stations 1, 12, 23, 34, 45, 56, 67, 78, 89, and 100, first for a single day, then for five days running. The latter experiment produced a total of 50 samples for analysis. The cumulative species catch for six days is shown in Fig. 2; six days is seen to offer no advantages over five. The data sets for both of the experiments were analyzed as in the previous experiment, and the species and station groups were examined as follows:

- 1) The program LOCATE was used to plot the tendency of each of the Fager groups to occur along the outfall-control axis. This permits one to determine whether the groups produced by this method in fact correspond to groups of species having the same input characteristics.

- 2) The Fager procedure gives, in addition to sets of recurrent groups, sets of "associated species." If a species is associated with a group, then it has significant affinities only with members of that group, but not with all of them (if the latter situation occurred, the associated species would be a member of the group). The analysis in experiment 1 above was therefore repeated, with the associated species for each group (if any) being pooled with the group.

- 3) Groups can be split off from the Orloci dendrogram by drawing horizontal lines across it, at different sum-of-squares values. In this experiment, the LOCATE program (which is not specific for Fager groups) was used to analyze the Orloci groups of species split off by sum-of-squares

break points of 14.

4) This experiment was the same as the previous one, except that a break-point of 6 was used.

5) This experiment was the same as the previous one, except that a break-point of 1 was used.

6) In the final experiment, the groups of stations produced by the Orloci procedure were compared with those produced by the Fager procedure. A statistical test was used to evaluate the type of variability examined by each of the tests.

RESULTS AND DISCUSSION

The data in Table 3, and the cumulative species catches in Fig. 2, indicate that there are some species out of the total number available for sampling, that are never caught. These are primarily those species having a very low mean concentration, and an overdispersed or random distribution pattern. Species with a highly aggregated distribution pattern, but the same low mean concentration, will sometimes be captured, however, since the concentration in the "patches" will be high enough for p_{ab} to be non-negligible on occasion. The actual variability in the catches of species along the principal diagonal of the grid (stations 1-100) is not too high, with the 95% confidence limits being only about $\pm 5\%$ of the mean. This is lower than one would expect for plankton samples in most parts of the open ocean, but probably not unlike that which would be obtained from estuarine or lake plankton.

For the experiments in series 2, the following general questions are being asked:

- 1) Can either of the affinity tests produce species groups in which cosmopolitan, outfall-inhibited, and outfall-enhanced species are distinguished? For the latter of these two, can species groups containing strongly-affected species be distinguished from those containing weakly-affected species?
- 2) Will either of the affinity tests produce station groups in which the stations located near the outfall are grouped separately from stations located a maximum distance from the outfall?

3) Which test detects these differences first (i.e. with the smallest number of stations sampled)?

The data for the experiments in series 2 are shown in Tables 4, 5, 6, and 7. The experiments are summarized in Table 8. The Fager groups are analyzed with the "associated species" included; in general, it appears that such species belong to the same category as the majority of the group species, and should therefore be included. It should be noted that not all of the species captured at a particular station will necessarily be included in one of the Fager groups. The groups include only those species that meet the criteria outlined for a group by Fager (1957), mainly that every species in a group should have significant associations with every other member of the group. A disadvantage of this is that it often leads to "groups" with only two members. Monitoring such a group, as a pollution indicator, is not a great improvement over monitoring a single species, though a group of size 2 can be more readily analyzed for intra-group interactions between the species.

The extraction of groups by the Orloci procedure is somewhat subjective, since there are no clear criteria in the procedure to delineate them, as in the Fager test. The output of the Orloci procedure is simply a dendrogram, showing the within-group sum of squares value for each group fusion. That is, pairs of species are fused together, each such fusion having a particular sum-of squares value which is low if the species are close together on the hypersphere graph and higher if they are further apart. Each pair is then fused with another species, or another pair, with a higher sum of squares for the new group. This continues until all of the species caught in that sampling run are fused into a single group. Obviously, the last all-inclusive group represents a trivial solution to

30

the problem, and the initial two-species groups give only a minimal amount of information. At some intermediate stage (i.e. some intermediate sum of squares) the groups presumably are "real"; in the present study the "real" groups would be those corresponding to outfall-inhibited or outfall-enhanced or cosmopolitan species. In theory, this level should be indicated on the dendrogram by an abrupt change in the fusion sums of squares below and above this level. That is, similar species should be fused into groups with only a moderate rise in the sums-of-squares, since the species are all clustered together in one region of the hypersphere. When all of the similar species have been fused into groups, and two of these dissimilar groups are fused, the sum of squares should rise greatly. In practice, this is not obvious because of the probabilistic nature of the model; species do not always occur where they "should," any more than they do in nature.

Consequently, the groups must be extracted from the dendrogram by eye. In order to make the process somewhat objective, I have chosen not to do this, but rather to define as a "group" in this series any collection of species that fuse with a sum of squares less than 2.0. This level is arbitrary, except that such a division tends to produce groups that are more or less equivalent in size and number to the Fager groups. In some cases, errors are obvious. For instance, in Table 7, Orloci species group 1 for the random stations is clearly composed of three cosmopolitan species (3, 6, 5), and 5 of the code 2, $k = 5$ (species moderately enhanced by the presence of the outfall). Groups 12-15, on the other hand, are split although they belong together. However, if the investigator is dealing with species whose distributional characteristics are unknown, these alterations could not be made, and the only reasonable

way to proceed would be by setting a sum-of-squares limit to define the groups, as I have done.

Each group is either homogeneous (all cosmopolitan species, all outfall-inhibited, or all outfall-enhanced), mixed, or spurious. A spurious group is defined as one in which no component makes up more than 50% of the total species. To facilitate comparison, a homogeneity index was calculated for each group: 1.0 for the homogeneous groups, 0 for the spurious ones, and some intermediate value greater than 0.5 for the mixed groups, the value being obtained by dividing the number of species in the group's largest component by the total number of species in the group. These fractions are given in the columns labeled "character of groups" in Tables 4-7. The means of these indices are calculated for each collection of groups, and are listed in Table 8.

Several interesting points emerge from Table 8. First, it is apparent that both methods can delineate groups of species that are fairly homogeneous in terms of distributional type. If groups were being formed more or less at random, most of them would be spurious, and the homogeneity index would be very low. There is no apparent advantage of one test over the other in terms of the homogeneity index, either with station or species groupings.

In terms of species groups, there is little advantage in sampling the "best" (most extreme) stations rather than picking random stations. The stations group themselves much better, however, when the "best" stations are sampled. This suggests that, in terms of sampling effort, it would be better to take more samples at the outfall and at a control station, and fewer in between, if station grouping is the object, and one wishes only to determine if the outfall is changing the species list.

The Orloci test appears to be slightly better when only 4 sites were sampled. The Fager test is slightly better for the larger samples. It appears that four samples are too few, and 32 shows no advantage over 16, in both the random and best-sample cases. A possible explanation for this is that, as the number of samples is increased, proportionately more samples are taken from the central portion of the grid, where the differences between the distributional types are not as apparent. This "adulteration" of the sample set may tend to obscure some of the differences in types.

The Fager test shows significantly fewer spurious species groups in the random samples, though not the best-station samples. There is no difference between the two for station groups, though neither method grouped stations as well as they did species.

Table 9 shows a comparison of the two tests in terms of their ability to separate weakly-influenced species (+ or -) from strongly-influenced ones. There is a difference in the means, but it is not statistically significant. The best-sample set of values for the Orloci test is significantly worse than the random set, however; the explanation for this is not immediately apparent.

In the last series of experimental runs, the "reality" of the groups formed by each of the tests was examined, both with respect to variation in time, due to repetitive sampling, and to variations in the criteria established for group definition in each test. The results of the first experiment are shown in Table 10.

In this case, sampling was done at each of the stations on the principal diagonal, and only one day's samples were used. The Fager groups were defined as previously, but variations were made in the cut-off levels for the Orloci groups. The first set of species groups (labeled

"primary groups") consists of only those groups that were formed with a sum-of-squares value less than 10^{-4} , the minimum necessary for printing in the program. Species not in these groups were ignored. All of the groups thus formed are completely homogeneous for a single distributional type, with several of them corresponding to the Fager groups. If the sum of squares cut-off is raised to 2.0, the value used in the previous series, all of the species will now be included, but the homogeneity of the groups drops. A further rise in the cut-off value to 6.0 reduces the number of groups to five, and raises the index slightly. It can be seen, however, that if any two adjacent groups in this set were fused into a single group (as the Orloci procedure would do), that group would probably be a spurious one. This cut-off value therefore represents the optimum for this particular run. There would, however, be no way to determine this for a set of real samples in which the distributional characteristics of the species were unknown. This points up the basic weakness of the dendrogram as a method of data presentation.

Both tests separated out identical station groups, with one pair of stations misplaced. The indices are a respectable 0.8, however.

If an investigator was taking a real series of samples, and attempting to separate out species groups for later monitoring as indicators, it would be necessary for him to know the tendency of each group to occur at each sampling site. The program LOCATE has been used with both the Fager and the Orloci species groups in this experiment to determine this. The results are plotted in Fig. 3. The investigator can use plots of this sort to help define the groups, and, in a crude way, to locate the optimum group sizes.

The plots for the Fager groups are hard to interpret. Group 1 is

fairly readily identified as a cosmopolitan group, since it occurs to more or less the same degree at all the sampling sites. With a certain amount of inspiration, the investigator might be able to place group 2 as an outfall-enhanced group and group 3 as an outfall-inhibited one, but the distributions are not clear-cut.

The Orloci groups plot out more neatly, despite the fact that the overall homogeneity index is not greatly different from the Fager test. For the Orloci primary groups, groups 1, 2, and 3 are readily identified as outfall-inhibited, outfall-enhanced, and cosmopolitan, respectively. Moreover, the identification of these three groups means that a total of 35 species have been accounted for (72 were caught in the entire sampling run). The first three groups in the Fager test account for only 23 species.

The plots for the Orloci groups with a sum-of-squares cut-off at 2.0 show that the first three groups are again clearly defined, and that groups 4 and 5 can now be identified with reasonable certainty as outfall-enhanced and outfall-inhibited, respectively. This accounts for 47 species.

The last set of plots shows the Orloci groups with a sum-of-squares cut-off at 6.0. All of the groups are now reasonably well identified as to type. If two of these groups are combined together, however (i.e. the sum-of-squares level is raised again), then the plots will become more variable because a spurious group is being formed. This gives the investigator an idea of where to stop with the group fusions.

The last experiment followed the previous one in format, except that the samples were taken along the principal diagonal stations for five consecutive days, producing a total of 50 samples. It would be

32

expected that this procedure, when analyzed in the same manner as the previous experiment, would provide a clearer picture of the group structure, since day-to-day variability is averaged out. Also, it allows a comparison of the two affinity measures in terms of the type of variability they are reacting to: day-to-day stochastic variability, or station-to-station environmental variability. Obviously, it is more desirable for a test to react to the second of these.

To make visual comparisons easier, the dendrograms for Orloci species clustering and station clustering are shown in Figures 4 and 5 respectively. Since few species came out as "primary groups" in this case, the lowest level for the sum-of-squares cut-off was taken as 1.0. This produced a total of 28 groups, which are listed in Table 11. For each of these groups, the program LOCATE was used to determine its tendency to occur at each of the principal diagonal stations, these "tendencies" being averaged over the five sampling days. The resulting distributions are shown in Fig. 6. Before determining the character of each of the groups, using Table 2, the graphs in Fig. 6 were scored subjectively as cosmopolitan, outfall-inhibited, outfall-enhanced, or indeterminant. These values are shown in Table 11, along with the actual group characters, and the cases in which the two disagree.

A similar procedure was followed for the Orloci species groups with a sum-of-squares break-point at 6.0, and at 14.0, and with the Fager species groups (including associated species). These graphs are shown in Figures 7, 8, and 9 respectively, and they are also summarized in Table 11.

All of this gives a picture much different than that in the previous run. Most of the graphs of group distribution are interpretable, and most of the interpretations came out correctly. Even the smaller

groups do not exhibit the degree of fluctuation shown in the graphs in Fig. 3. The value of repetitive sampling before applying one of the affinity tests is demonstrated clearly.

Changing the level at which the Orloci groups are separated out has little effect on the homogeneity index, as seen in the Table 11 summary. The Fager groups are higher in homogeneity, although they include only a total of 74 species out of the 96 caught. The lowest-level Orloci groups include 88 species, since single, ungrouped species were not included. Comparative runs for the Fager test, with and without the associated species, showed that including them did not change the shape of the distribution graphs much, or affect the homogeneity. Under these circumstances, it is best to include them, since the larger the groups are, the more information they will provide, and the less will be their degree of random fluctuation.

A key experiment was done to complete this set of runs. Station clusters were calculated, using both affinity tests, for the 50 stations sampled (10 samples/day for 5 days). These were plotted on a pattern as shown in Table 12. If the station groups formed are mostly along rows in this pattern, then the test is reacting primarily to the random variation in the collections from one day to the next (since each row represents a day). If the tests are reacting primarily to the environmental variation between the sampling sites (column 1 is the control station and column 10 is the outfall), then the groups formed will be primarily along columns.

The simplest way to test for this is to treat the pattern in Table 12 as an analysis-of-variance array, and test for significant differences between columns and between rows, when numbers are inserted into

3

the grid corresponding to the groups to which each of the stations belong. This is done in Table 13, which also shows the set-up for the statistical test. Since a normal distribution of the groups cannot be assumed, a nonparametric test is used, the Kendall Concordance procedure (Tate and Clelland, 1957).

The concordance procedure shows a striking difference between the two affinity tests. The Orloci test shows a highly significant difference between rows in the diagram, and no differences between columns. This indicates that the Orloci groups are largely within rows, and spread across most of the control-outfall gradient, i.e. the groups are reacting primarily to random sampling variability from one day to the next. This is accented by the low homogeneity index for these groups, of 0.57. This is only slightly better than random sorting; in the absence of spuriour groups (averaged as 0.0), a random assortment would give an index value of 0.50.

The Fager groups are much better; in this case, they omit none of the stations sampled. The index is 0.90, and, in the concordance test, it is seen that most of the variability is in the columns of the diagram. That is, the Fager procedure is ignoring the sampling variability and extracting real groups based on the environmental variability.

CONCLUSIONS

The data presented here indicate that affinity tests have potential use in water-quality investigations, where planktonic organisms are collected. The tests do not replace diversity studies, but give valuable supplementary information. They can strengthen studies based on the monitoring of indicator organisms by delineating indicator groups, composed of species having similar reactions to environmental conditions. The use of affinity tests to group sampling sites, based on their species-list similarities is of value in detecting small but persistent changes between control and test points. It is quite sensitive, since it uses the entire species list and not just a few members of it. The tests require no more sampling than that required for diversity studies.

These studies could be extended in a number of ways not touched upon in the present investigation. Ordination of sample sites, with respect to specific environmental variables, represents the next level of sophistication. A few workers have applied these techniques to plankton (e.g. Roux and Reyssac, 1975); although their general ecological application has been questioned (Beals, 1973), they seem well-suited to certain kinds of pollution studies. Nor has the present investigation by any means exhausted the possibilities of existing affinity indices and procedures.

Both of the affinity tests emphasized in this study appear to have some usefulness for the kind of sampling program required in studying a thermal outfall. Since there are various affinity tests related to the ones investigated, this would indicate that they may be useful as

well. In most of the experiments done with the model communities generated here, both the Fager and Orloci procedures performed about equally well. However, in the last sampling series, which employed samples taken along the outfall-control line over a period of several days (the procedure likely to be employed in actual field studies), the Fager procedure proved clearly superior. It was much more able to delineate community differences, masked by sampling variability, than was the Orloci test.

Most authors who review the subject of affinity tests point out the difficulty of choosing suitable ones for a particular study. There is no analytical method for predicting the applicability of a particular method in advance. The procedure employed in the present study appears to be one way around that difficulty. With a minimal amount of information about the distributions of various species in a planktonic community, and about the sampling procedures to be employed, it is possible to construct a distributional model of the community which can be used to test sampling techniques and analytical methods. Such a program can be set up and tested at rather low cost, and will probably save money in the long run, by maximizing the information that can be obtained from collections in the field.

I feel that the present study represents only an initial effort in this direction, and should be extended to other aspects of association analysis, multivariate analysis, ordination, and diversity studies. There is no reason why the general techniques employed here cannot be extended as well to different environments, such as the benthos, although the greater heterogeneity of the environment will make the generation programs somewhat more complex, in those cases.

REFERENCES

- Beals, E.W. 1973. Ordination: mathematical elegance and ecological naivete. *Journal of Ecology* 61:23-35.
- Bush, R.M., and E.B. Welch. 1972. Plankton associations and related factors in a hypereutrophic lake. *Water, Air and Soil Pollution* 1(3):257-274.
- Cassie, R.M. 1967. Mathematical models for the interpretation of in-shore plankton communities. pp. 509-514 in *Estuaries* (G.H. Lauff, ed.), Pub. 83, American Association for the Advancement of Science, Washington, D.C.
- Fager, E.W. 1957. Determination and analysis of recurrent groups. *Ecology* 38(4):586-595.
- Fager, E.W., and J.A. McGowan. 1963. Zooplankton species groups in the North Pacific. *Science* 140(3566):453-460.
- Frenkel, R.E., and C.M. Harrison. 1974. An assessment of the usefulness of phytosociological and numerical classificatory methods for the community biogeographer. *Journal of Biogeography* 1(1):27-56.
- Goodall, D.W. 1966. A new similarity index based on probability. *Biometrics* 22:883-907.
- Goodall, D.W. 1973. Sample similarity and species correlation. pp. 105-156 in R.H. Whittaker, ed., *Handbook of Vegetation Science*. Part V: Ordination and Classification of Vegetation. Dr. W. Junk B.V., The Hague.
- Greig-Smith, P. 1964. *Quantitative Plant Ecology*. Butterworths, London.

- 41
- Hamner, W.M., et al. 1975. Underwater observations of gelatinous zooplankton: Sampling problems, feeding biology, and behavior. *Limnology and Oceanography* 20(6):907-917.
- Legendre, L. 1973. Phytoplankton organization in Baie des Chaleurs (Gulf of St. Lawrence). *Journal of Ecology* 61:135-149.
- Orloci, L. 1967. An agglomerative method for classification of plant communities. *Journal of Ecology* 55:193-205.
- Orloci, L. 1975. *Multivariate Analysis in Vegetation Research*. Dr. W. Junk B.V., The Hague. 276 pp.
- Pielou, E.C. 1969. *An Introduction to Mathematical Ecology*. Wiley, New York. 286 pp.
- Pielou, E.C. 1975. *Ecological Diversity*. Wiley, New York. 165 pp.
- Pomeroy, L.R. 1974. The ocean's food web, a changing paradigm. *BioScience* 24(9):499-504.
- Roux, M., and J. Reyssac. 1975. Essai d'application au phytoplancton marin de methodes statistiques utilisees en phytosociologie terrestre. *Ann. Inst. oceanogr. Paris* 51(1):89-97.
- Strickland, J.D.H., et al. 1970. The ecology of the plankton off La Jolla, California, in the period April through September, 1967. *Bulletin of the Scripps Institution of Oceanography* 17:1-103.
- Tate, M.W., and R.C. Clelland, 1957. *Nonparametric and Shortcut Statistics*. Interstate Printers & Publishers, Danville, Ill. 171 pp.

Table 1

Effect of the intensity coefficient K on the probability of species occurrence in the community-generation program SLIST.

If the probability of distributional occurrence (p_{dist}) = 1.0 at square 1 (for outfall-inhibited species) or at square 100 (for outfall-enhanced species), then the probability is assumed to drop with geometric distance from these points according to the relationship:

$$p_{\text{dist}} = e^{-0.1 K D}$$

(K is input as a whole number to facilitate data manipulation).

Distance (grid units) from point of maximum occurrence	p_{dist} for:		
	K = 1	K = 5	K = 10
0	1.000	1.000	1.000
1	0.368	0.607	0.905
2	0.135	0.368	0.819
4	0.018	0.135	0.670
6	0.002	0.050	0.549
8	< .001	0.018	0.449
10	< .001	0.007	0.368
12	< .001	0.002	0.301
12.726 (max.)	< .001	0.002	0.280

Table 2

Input data for the individual species in the community-generation program (SLIST).

sp. #	code	K	$\bar{x}$	S^2	sp. #	code	K	$\bar{x}$	S^2	sp. #	code	K	$\bar{x}$	S^2
1	0	0	2	1	48	1	5	4	5	95	2	5	7	7
2	0	0	2	2	49	1	5	5	5	96	2	5	7	8
3	0	0	2	3	50	1	5	5	6	97	2	5	8	8
4	0	0	2	4	51	1	5	6	6	98	2	5	8	9
5	0	0	3	2	52	1	5	6	7	99	2	1	2	2
6	0	0	3	3	53	1	5	7	7	100	2	1	2	3
7	0	0	3	4	54	1	5	7	8	101	2	1	3	3
8	0	0	3	5	55	1	5	8	8	102	2	1	3	4
9	0	0	4	3	56	1	5	8	9	103	2	1	4	4
10	0	0	4	4	57	1	1	2	2	104	2	1	4	5
11	0	0	4	5	58	1	1	2	3	105	2	1	5	5
12	0	0	4	6	59	1	1	3	3	106	2	1	5	6
13	0	0	5	4	60	1	1	3	4	107	2	1	6	6
14	0	0	5	5	61	1	1	4	4	108	2	1	6	7
15	0	0	5	6	62	1	1	4	5	109	2	1	7	7
16	0	0	5	7	63	1	1	5	5	110	2	1	7	8
17	0	0	6	5	64	1	1	5	6	111	2	1	8	8
18	0	0	6	6	65	1	1	6	6	112	2	1	8	9
19	0	0	6	7	66	1	1	6	7					
20	0	0	6	8	67	1	1	7	7					
21	0	0	7	6	68	1	1	7	8					
22	0	0	7	7	69	1	1	8	8					
23	0	0	7	8	70	1	1	8	9					
24	0	0	7	9	71	2	10	2	2					
25	0	0	8	7	72	2	10	2	3					
26	0	0	8	8	73	2	10	3	3					
27	0	0	8	9	74	2	10	3	4					
28	0	0	8	10	75	2	10	4	4					
29	1	10	2	2	76	2	10	4	5					
30	1	10	2	3	77	2	10	5	5					
31	1	10	3	3	78	2	10	5	6					
32	1	10	3	4	79	2	10	6	6					
33	1	10	4	4	80	2	10	6	7					
34	1	10	4	5	81	2	10	7	7					
35	1	10	5	5	82	2	10	7	8					
36	1	10	5	6	83	2	10	8	8					
37	1	10	6	6	84	2	10	8	9					
38	1	10	6	7	85	2	5	2	2					
39	1	10	7	7	86	2	5	2	3					
40	1	10	7	8	87	2	5	3	3					
41	1	10	8	8	88	2	5	3	4					
42	1	10	8	9	89	2	5	4	4					
43	1	5	2	2	90	2	5	4	5					
44	1	5	2	3	91	2	5	5	5					
45	1	5	3	3	92	2	5	5	6					
46	1	5	3	4	93	2	5	6	6					
47	1	5	4	4	94	2	5	6	7					

Key

code 0 = cosmopolitan spp.

1 = outfall-inhibited spp.

2 = outfall-enhanced spp.

K = 1: sp. weakly affected (+ or -)

2: sp. moderately " (")

3: sp. strongly " (")

 $\bar{x}$: mean cells per liter (as a power of 10) S^2 : variance of number of cells per liter (as a power of 10)

variance = mean for randomly-distributed spp. (Poisson)

variance < mean for evenly-distributed spp.

variance > mean for aggregated spp.

Table 3

Summary of series 1 - number of species obtained with repetitive sampling, along the 10 principal diagonal sampling sites, on 9 successive days:

<u>Day</u>	<u># species caught</u>
1	132
2	130
3	147
4	125
5	126
6	136
7	131
8	123
9	131

$$\bar{x} = 131.22$$

95% confidence limits =

$$125.72 - 136.72$$

Sampling was from an expanded version of the species list in Table 2, containing 189 species. Thus, an average of 69% (and a maximum of 78%) of the species were captured.

Table 4

Summary of Experiment 1, series 2 (4 sampling sites)

a) Random Stations

Test	Objects Grouped	Groups	Character of Group
Fager	species	1) 13 14 17 18 21 25 2) 16 24 207 209	all cosmopolitan spurious
Orloci	species	1) 4 15 66 68 9 65 67 70 106 2) 20 27 28 3) 8 108 111 16 24 107 109 4) 13 14 17 18 21 25 22 26 5) 10 11 104 105 112 6) 19 61 62 63 64 69	5/9 code 1, k=1 all cosmopolitan 4/7 code 2, k=1 all cosmopolitan spurious 5/6 code 1, k=1
Fager	stations	no groups found	
Orloci	stations	1) 1 99 2) 11 100	spurious spurious

b) Best Stations

Test	Objects Grouped	Groups	Character of Group
Fager	species	1) 4 12 107 109 112 11 13 14 17 21 22 25 26 2) 15 20 28 105 106 111 3) 10 23 104 4) 65 67 69 5) 18 24 6) 9 103	10/13 cosmopolitan spurious 2/3 cosmopolitan all code 1, k=1 all cosmopolitan spurious
Orloci	species	1) 4 12 107 109 112 9 103 11 13 14 17 21 22 25 26 2) 10 23 104 18 24 3) 15 20 28 105 106 111 19 4) 75 77 78 79 81 82 83 91 92 93 94 96 5) 16 27 63 64 66 68 70 65 67 69 6) 34 35 36 37 38 39 40 41 42 47 48 49 51 52 55 56 61 62 108 110	11/15 cosmopolitan 5/6 cosmopolitan 4/7 cosmopolitan all code 2, k=10 or 5 8/10 code 1, k=1 18/20 code 1, k=10 or 5 or 1
Fager	stations	1) 100 99	outfall
Orloci	stations	1) 100 99 2) 1 11	outfall control

Table 4 (cont.)

Key (for Tables 4, 5, 6, and 7)

Numbers of the objects in the groups refer either to species numbers (see Table 2) or station numbers (see Fig. 1)

code 1 = species inhibited by the presence of the outfall (see Table 2)

code 2 = species enhanced by the presence of the outfall (see Table 2)

cosmopolitan = species uninfluenced by the outfall (see Table 2)

k = 1 : species slightly influenced (+ or -) by the outfall (see Tables 1 and 2)

k = 5 : species moderately influenced (+ or -) by the outfall (see Tables 1 and 2)

k = 10 : species strongly influenced (+ or -) by the outfall (see Tables 1 and 2)

spurious group: one in which no single category accounts for more than half of the group members, i.e. a mixed group with no single fraction predominating

underlined species in Fager test: species associated with that recurrent group (see text). These species have been included in the groups.

Table 5

Summary of Experiment 2, series 2 (8 sampling sites)

a) Random Stations

Test	Objects Grouped	Groups	Character of Groups
Fager	species	1) 18 26 107 109 13 17 21 25 28 <u>15</u> 2) 112 105 108 <u>12</u> 3) 9 63 65 4) 110 14 <u>111</u> 5) 11 22	8/10 cosmopolitan 3/4 code 2, k=1 2/3 code 1, k=1 2/3 code 2, k=1 all cosmopolitan
Orloci	species	1) 4 32 33 35 36 37 39 40 41 46 47 49 50 51 53 54 55 60 64 66 68 2) 8 9 26 63 107 109 111 61 110 3) 12 105 108 112 23 27 4) 10 20 16 102 104 5) 11 22 106 13 17 18 21 25 65 6) 14 28 70 15 7) 19 24 67 69 62	20/21 code 1, k=10, 5, 1 spurious spurious 3/5 cosmopolitan 7/9 cosmopolitan 3/4 cosmopolitan 3/5 code 1, k=1
Fager	stations	1) 67 9 76 25 86 <u>48</u>	spurious
Orloci	stations	1) 9 48 25 86 67 33 76 2) 1	spurious control

Table 5 (cont.)

Test	Objects Grouped	Groups	Character of Group
Fager	species	1) 14 105 109 13 17 18 21 22 25 107 <u>10</u> <u>112</u> 2) 33 35 37 40 41 47 49 51 53 54 3) 106 81 83 91 95 96 97 98 4) 65 103 9 61 63 5) 67 70 69 <u>24</u> 6) 75 79 89 7) 76 77 93 8) 110 111 9) 19 16 10) 68 11 <u>15</u> <u>64</u> 11) 28 20 12) 23 26 <u>66</u>	8/12 cosmopolitan all code 1, k=10 or 5 all code 2, k=10,5,1 3/5 code 1, k=1 3/4 code 1, k=1 all code 2, k=10,5 all code 2, k=10,5 all code 2, k=1 all cosmopolitan spurious all cosmopolitan 2/3 cosmopolitan
oci	species	1) 4 8 60 84 108 112 2) 5 110 111 10 107 14 105 109 9 61 63 64 65 103 106 3) 13 17 18 21 22 25 16 19 23 26 20 28 4) 11 68 15 24 67 70 66 69 62 5) 12 104 75 79 89 59 74 78 88 6) 27 82 90 76 77 93 80 81 83 91 95 96 97 98 7) 6 7 34 36 39 42 48 50 55 56 58 8) 31 45 46 52 101 33 35 37 40 41 47 49 51 53 54	spurious spurious all cosmopolitan 6/9 code 1, k=1 6/9 code 2, k=10,5 13/14 code 2, k=10,5 9/11 code 1, k=10,5,1 14/15 code 1, k=10,5
Fager	stations	1) 90 99 100 2) 1 2 3) 12 89	outfall control spurious
oci	stations	1) 1 2 2) 11 90 99 3) 12 89 100	control 2/3 outfall 2/3 outfall

Table 6

Summary of Experiment 3, series 2 (16 sampling sites)

a) Random Stations

Test	Objects Grouped	Groups	Character of Groups
Fager	species	1) 13 17 21 25 10 9 69 67 63 66 <u>62</u> <u>64</u>	spurious
		2) 26 22 18 28 <u>8</u> <u>12</u> <u>20</u> <u>24</u> <u>14</u> <u>19</u>	all cosmopolitan
		3) 49 51 53 55	all code 1, k=5
		4) 61 65 70	all code 1, k=1
		5) 109 107 106 <u>105</u> <u>111</u>	all code 2, k=1
		6) 46 48 56	all code 1, k=5
		7) 27 16	all cosmopolitan
		8) 108 112	all code 2, k=1
Orloci	species	1) 4 16 27	all cosmopolitan
		2) 11 14 109 111	spurious
		3) 104 105 107	all code 2, k=1
		4) 103 108 112	all code 2, k=1
		5) 8 20 12 15	all cosmopolitan
		6) 3 17 21 25 26 18 22 28 24	all cosmopolitan
		7) 19 23	all cosmopolitan
		8) 6 102 60 110	spurious
		9) 7 68	spurious
		10) 9 10 106 61 65 70	spurious
		11) 62 63 69 66 67 64	all code 1, k=1
		12) 33 35 39 40 41 42	all code 1, k=10
		13) 46 48 56 47 50	all code 1, k=5
		14) 49 52 51 53 55 54	all code 1, k=5
		15) 75 78 79 80 81 82 83 84 89 90 92 93	
Fager	stations	95 96 97 98 101	all code 2, k=10, 5, 1
		1) 39 7 14 9 34 39 22 <u>98</u>	5/7 control
		2) 80 64 50 56	3/4 outfall
		3) 55 77	spurious
Orloci	stations	4) 5 1	all control
		1) 1 14 22 5	all control
		2) 7 34 9 39 98	3/5 control
		3) 64 50 56 80 77 55	4/6 outfall

Table 6 (cont.)

b) Best Stations

Test	Objects Grouped	Groups	Character of Groups
Fager	species	1) 77 79 81 82 83 90 93 94 95	all code 2, k=10, 5
		2) 13 17 18 21 25 22 26 15 <u>14</u> <u>24</u>	all cosmopolitan
		3) 49 53 34 35 37 39 41 47 <u>50</u>	all code 1, k=10, 5
		4) 107 111 105 <u>106</u>	all code 2, k=1
		5) 56 61 55 51	all code 1, k=5, 1
		6) 67 69 64 <u>9</u> <u>65</u>	4/5 code 1, k=1
		7) 91 103	all code 2, k=5, 1
		8) 16 109 <u>110</u>	2/3 code 2, k=1
		9) 20 7 <u>27</u>	all cosmopolitan
		10) 92 98	all code 2, k=5
		11) 11 28	all cosmopolitan
		12) 10 23	all cosmopolitan
Orloci	species	1) 4 112 91 103 96 92 98	6/7 code 2, k=5, 1
		2) 7 20 9 67 65 66	spurious
		3) 12 27 68	2/3 cosmopolitan
		4) 76 78 80	all code 2, k=10
		5) 10 23 16 109 104 108	spurious
		6) 105 107 111 106 110	all code 2, k=1
		7) 11 22 28	all cosmopolitan
		8) 13 17 18 21 25 26 15 24	all cosmopolitan
		9) 75 84 97 77 79 81 82 83 90 93 94 95	all code 2, k=10, 5
		10) 6 36 38 102	spurious
		11) 33 42	all code 1, k=10
		12) 34 35 37 39 41 47 60	all code 1, k=10, 5, 1
		13) 46 50 52 49 53 55	all code 1, k=5
		14) 8 40	spurious
		15) 14 64 69 62	3/4 code 1, k=1
		16) 19 63 70	2/3 code 1, k=1
		17) 48 54 51 56 61	all code 1, k=5, 1
Fager	stations	1) 80 3 21 13	3/4 control
		2) 79 99 98 100 <u>88</u>	all outfall
		3) 12 1 11	all control
		4) 89 90	all outfall
		5) 2 22	all control
Orloci	stations	1) 1 12 2 11	all control
		2) 3 13 89 21 90	3/5 control
		3) 22 80 79 88	3/4 outfall
		4) 100 99 98	all outfall

Table 7

Summary of Experiment 4, series 2 (32 sampling sites)

a) Random Stations

Test	Objects Grouped	Groups	Character of Groups
Fager	species	1) 13 17 21 25 26 14 105 107 109 <u>111 11</u> 2) 48 49 51 35 37 39 54 55 3) 65 9 61 63 67 69 <u>10 64 102 62</u> <u>66 68 110</u> 4) <u>18 22 23 24</u> <u>12 15 20 19</u> 5) 47 53 103 6) 3 6 7) 58 101 8) 38 41 9) 8 16 10) 27 28 <u>106</u>	7/11 cosmopolitan all code 1, k=10, 5 9/13 code 1, k=1 all cosmopolitan 2/3 code 1, k=5 all cosmopolitan spurious all code 1, k=10 all cosmopolitan 2/3 cosmopolitan
Orloci	species	1) 3 6 5 89 90 91 93 94 2) 4 62 59 60 3) 47 53 103 48 49 51 54 4) 35 37 39 58 101 55 56 5) 38 41 42 6) 31 33 40 44 50 52 57 100 7) 7 64 9 69 63 65 67 8) 61 102 9) 20 66 68 10) 10 105 109 111 107 106 110 <u>112</u> 11) 11 104 108 70 12) 8 16 13) 12 15 14) 13 17 21 25 26 18 14 22 23 24 19 15) 27 28	5/8 code 2, k=5 3/4 code 1, k=1 6/7 code 1, k=5 6/7 code 1, k=10, 5, 1 all code 1, k=10 7/8 code 1, k=10, 5, 1 5/7 code 1, k=1 spurious 2/3 code 1, k=1 7/8 code 2, k=1 spurious all cosmopolitan all cosmopolitan all cosmopolitan all cosmopolitan
Fager	species	1) 82 68 59 69 95 8 92 24 45 26 28 65 29 94 39 70 2) 56 62 44 61 30 8 9 99 3) 4 41 23 11 2 <u>76 34 90</u>	11/16 outfall 5/8 control 6/8 control
Orloci	stations	1) 8 69 68 2) 8 92 94 95 3) 62 24 90 65 39 4) 30 56 5) 29 76 6) 26 70 7) 45 4 41 44 23 34 8) 2 61 9) 9 11 59 28 82 99	2/3 outfall 3/4 outfall 3/5 outfall all outfall all outfall spurious all control all control spurious

Table 7 (cont.)

b) Best Stations

Test	Objects Grouped	Groups	Character of Groups
Fager	species	1) 26 13 27 21 25 65 69 67 20 <u>12</u> <u>19</u> <u>62</u> <u>70</u> 2) 39 41 49 53 37 42 <u>51</u> 3) 77 82 91 96 97 4) 107 111 9 103 10 61 5) 18 22 14 23 <u>11</u> <u>15</u> <u>27</u> <u>28</u> 6) 5 6 101 102 7) 76 78 81 93 8) 63 68 9) 40 52 10) 47 56 11) 108 106 <u>105</u> 12) 36 54 13) 46 50 14) 89 92 15) 110 109 <u>112</u> 16) 24 64	8/13 cosmopolitan all code 1, k=10, 5 all code 2, k=10, 5 spurious all cosmopolitan spurious all code 2, k=10, 5 all code 1, k=1 all code 1, k=10, 5 all code 1, k=5 all code 2, k=1 all code 1, k=10, 5 all code 1, k=5 all code 2, k=5 all code 2, k=1 spurious
Orloci	species	1) 4 104 2) 9 10 103 3) 105 106 108 107 111 109 110 112 4) 8 24 61 64 66 5) 20 70 65 69 67 6) 27 63 68 7) 11 16 8) 12 62 15 19 28 9) 13 17 21 25 26 18 22 14 23 10) 7 59 38 48 11) 39 41 49 53 40 52 46 50 47 56 12) 31 32 35 37 13) 33 36 54 37 42 55 14) 5 6 101 102 60 73 74 75 79 87 15) 76 78 81 93 83 94 98 16) 77 82 91 96 97 89 92 80 88	spurious 2/3 cosmopolitan all code 2, k=1 3/5 code 1, k=1 4/5 code 1, k=1 2/3 code 1, k=1 all cosmopolitan 4/5 cosmopolitan all cosmopolitan 3/4 code 1, k=10, 5, 1 all code 1, k=10, 5 all code 1, k=10, 5 all code 1, k=10, 5 spurious all code 2, k=10, 5 all code 2, k=10, 5
Fager	stations	1) 13 12 98 4 5 14 24 79 80 87 33 70 2) 77 96 78 89 3 32 11 23 3) 97 69 68 90 4) 21 1 22 2 <u>31</u> 5) 88 100 <u>99</u>	8/12 control spurious all outfall all control all outfall
Orloci	stations	1) 1 2 22 21 2) 11 4 32 97 96 77 3 69 3) 13 23 78 4) 31 99 5) 100 89 88 6) 12 87 24 98 33 79 5 7) 14 80 68 90 70	all control spurious 2/3 control spurious all outfall 4/7 control 4/5 outfall

Table 8

Summary of Series 2

Random Stations

Species groupings		Number of sampling sites				$\bar{x}$	Standard Error
Test		4	8	16	32		
Fager	no. of groups	2	5	8	10		
	% spurious groups	5	0	13	10	7.00	2.86
	index	.5	.78	.88	.77	0.73	0.08
Orloci	no. of groups	6	7	15	15		
	% spurious groups	17	29	27	13	21.5	3.86
	index	.66	.61	.73	.75	0.69	0.03
Station groupings		Number of sampling sites				$\bar{x}$	Standard Error
Test		4	8	16	32		
Fager	no. of groups	0	1	4	3		
	% spurious groups	0	100	25	0	31.25	23.66
	index	0	0	.62	.69	0.33	0.19
Orloci	no. of groups	2	2	3	9		
	% spurious groups	100	50	0	22	43.0	21.58
	index	0	.5	.76	.67	0.48	0.17

Best Stations

Species groupings		Number of sampling sites				$\bar{x}$	Standard Error
Test		4	8	16	32		
Fager	no. of groups	6	12	12	16		
	% spurious groups	33	8	0	19	15.0	7.15
	index	.57	.81	.96	.79	.78	0.08
Orloci	no. of groups	6	8	17	16		
	% spurious groups	0	25	24	13	15.5	5.84
	index	.81	.63	.70	.77	0.73	0.04

Table 8 (cont.)

Station groupings

		Number of sampling sites				$\bar{x}$	Standard Error
Test		4	8	16	32		
Fager	no. of groups	1	3	5	5		
	% spurious groups	0	33	0	20	13.25	8.10
	index	1.0	.67	.95	.73	0.84	0.08
Orloci	no. of groups	2	3	4	7		
	% spurious groups	0	0	0	29	7.25	7.25
	index	1.0	.78	.84	.58	0.80	0.09

Table 9

Fraction of species groups in series 2, predominantly code 1 or code 2, which are homogeneous for a single value of K (10, 5, or 1)

Test	Set	Number of sampling sites				$\bar{x}$	Standard Error
		4	8	16	32		
Fager	Random	0	1.0	1.0	.75	0.69	0.24
	Best	1.0	.43	.50	.55	0.62	0.13
Orloci	Random	1.0	1.0	1.0	.78	0.95	0.06
	Best	.67	.20	.60	.40	0.47	0.11

Table 10

Summary of Experiment 1, series 2. (10 sampling sites, 1 day)

Test	Objects Grouped	Groups	Character of Groups	Homogeneity Index
Fager	species	1) 13 17 21 25 26 14 22 18 16 <u>11</u> 2) 105 104 107 109 111 <u>112</u> 3) 64 69 67 10 <u>63 65 70</u> 4) 24 28 <u>108</u> 5) 7 9 6) 102 103 7) 19 15 <u>20</u>	all cosmopolitan all code 2, k=1 6/7 code 1, k=1 2/3 cosmopolitan all cosmopolitan all code 2, k=1 all cosmopolitan	0.93
Orloci	species (primary groups)	1) 30 33 35 37 38 39 41 42 46 49 51 54 55 60 61 62 2) 75 78 79 81 82 83 84 89 90 91 93 94 96 98 3) 13 17 21 25 26 4) 7 9 5) 104 105 6) 64 69 7) 102 103	all code 1, k=10,5,1 all code 2, k=10,5 all cosmopolitan all cosmopolitan all code 2, k=1 all code 1, k=1 all code 2, k=1	1.00
Orloci	species (break at s.s.=2)	1) 30 33 35 37 38 39 41 42 46 49 51 54 55 60 61 62 110 2) 75 78 79 81 82 83 84 89 90 91 93 94 96 98 102 103 3) 13 17 21 25 26 22 16 18 19 20 24 28 4) 104 105 108 112 107 109 111 5) 64 69 70 65 68 6) 11 14 15 23 7) 4 7 9 8) 10 67 63 9) 8 12 106 10) 27 66	16/17 code 1, k=10,5,1 all code 2, k=10,5,1 all cosmopolitan all code 2, k=1 all code 1, k=1 all cosmopolitan all cosmopolitan 2/3 code 1, k=1 2/3 cosmopolitan snurious	0.83
Orloci	species (break at s.s.=6)	1) 8 12 106 75 78 79 81 82 83 84 89 90 91 93 94 96 98 102 103 2) 30 33 35 37 38 39 41 42 46 49 51 54 55 60 61 62 110 3) 11 14 15 23 13 17 21 25 26 22 16 18 19 20 24 28 4) 4 7 9 104 105 108 112 107 109 111 5) 10 67 63 64 69 70 65 68 27 66	17/19 code 2, k=10,5,1 16/17 code 1, k=10,5,1 all cosmopolitan 7/10 code 2, k=1 8/10 code 1, k=1	0.87
Fager	stations	1) 56 45 89 67 <u>100</u> 2) 78 23 12 34 <u>1</u>	4/5 outfall 4/5 control	0.80
Orloci	stations (break at s.s.=2)	1) 1 12 34 23 78 2) 45 56 89 100 67	4/5 control 4/5 outfall	0.80

Table 11

Summary of Experiment 2, series 2.

Orloci species groups

a) Sum of squares break-point =1.0

Group	Character of Group	Character of Group's Location Graph (Fig.6)	Error
1) 13 17 21 25 22 26 18 14	all cosmopolitan	cosmopolitan	
2) 9 10 63 65 67 69	4/6 code 1, k=1	inhibited	
3) 105 107 109 111 108	all code 2, k=1	enhanced	
4) 3 44 5 59	spurious	indeterminant	
5) 6 76 88 101	spurious	indeterminant	
6) 79 95 94 80	all code 2, k=10,5	enhanced	
7) 37 39 41 48	all code 1, k=10,5	inhibited	
8) 49 53 55 54	all code 1, k=10,5	inhibited	
9) 78 86 98	all code 2, k=10,5	enhanced	
10) 74 75 90	all code 2, k=10,5	enhanced	
11) 84 91 93	all code 2, k=10,5	enhanced	
12) 81 89 92	all code 2, k=10,5	enhanced	
13) 7 64 103	spurious	cosmopolitan	X
14) 12 16 19	all cosmopolitan	enhanced	X
15) 31 32 35	all code 1, k=10	inhibited	
16) 34 36 40	all code 1, k=10	inhibited	
17) 50 51 52	all code 1, k=5	inhibited	
18) 77 83	all code 2, k=10	enhanced	
19) 96 97	all code 2, k=5	enhanced	
20) 4 47	spurious	indeterminant	
21) 66 70	all code 1, k=1	inhibited	
22) 20 61	spurious	inhibited	X
23) 11 15	all cosmopolitan	indeterminant	X
24) 62 68	all code 1, k=1	inhibited	
25) 104 112	all code 2, k=1	cosmopolitan	X
26) 8 24	all cosmopolitan	cosmopolitan	
27) 27 28	all cosmopolitan	cosmopolitan	
28) 33 42	all code 1, k=10	inhibited	

b) Sum-of-squares break-point = 6.0

1) 8 24 13 17 21 25 22 26 18	all cosmopolitan	cosmopolitan	
14 23 12 16 19 27 28			
2) 77 83 84 91 93 96 97 82 79	all code 2, k=10,5	enhanced	
95 94 80 81 89 92			
3) 9 10 63 65 67 69 66 70 20	9/14 code 1, k=1	inhibited	
61 11 15 62 68			
4) 104 112 105 107 109 111 108	all code 2, k=1	inhibited	
110			
5) 104 112 105 107 109 111 108	all code 2, k=1	indeterminant	X
110 106			
6) 38 49 53 55 54 56 50 51 52	all code 1, k=10,5	inhibited	
7) 6 76 88 101 78 96 98	6/8 code 2, k=10,5,1	indeterminant	X
8) 4 47 102 7 64 103	spurious	indeterminant	
9) 3 44 5 59	spurious	indeterminant	
10) 74 75 90	all code 2, k=10,5	enhanced	
11) 31 32 35	all code 1, k=10	inhibited	

Table 11 (cont.)

c) Sum-of-squares break-point = 14.0

Group	Character of Group	Character of Group's Location Graph (Fig. 6)	Error
1) 31 32 35 33 42 60 34 36 40 37 39 41 48 38 49 53 55 54 56 50 51 52	all code 1, k=10,5,1	inhibited	
2) 4 47 102 7 65 103 9 10 63 65 67 69 66 70 20 61 11 15 62 68	11/20 code 1, k=5,1	inhibited	
3) 74 75 90 77 83 84 91 93 96 97 82 79 95 94 80 81 89 92	all code 2, k=10,5	enhanced	
4) 8 24 13 17 21 25 22 26 18 14 23 12 16 19 27 28	all cosmopolitan	cosmopolitan	
5) 3 44 5 59 6 76 88 101 78 86 98	6/11 code 2, k=10,5	indeterminant	X
6) 104 112 105 107 109 111 108 110 106	all code 2, k=1	indeterminant	X

Fager species groups (associated species included)

1) 13 17 21 15 16 22 67 107 14 105 109 9 10 63 65 69 4 23 66 70 8 15 61 106 103 20	14/26 cosmopolitan	cosmopolitan	
2) 49 53 51 55 54 56 50 52	all code 1, k=5	inhibited	
3) 91 93 95 96 97 84 77	all code 2, k=10,5	enhanced	
4) 18 19 16 11 64 12 24	6/8 cosmopolitan	cosmopolitan	
5) 37 39 34 36 41 48	all code 1, k=10,5	inhibited	
6) 81 94 89 92	all code 2, k=10,5	enhanced	
7) 76 88 101	all code 2, k=10,5,1	indeterminant	X
8) 108 110 111	all code 2, k=1	enhanced	
9) 32 35	all code 1, k=10	inhibited	
10) 79 83	all code 2, k=10	enhanced	
11) 27 28	all cosmopolitan	cosmopolitan	
12) 62 68	all code 1, k=1	inhibited	
13) 104 112	all code 2, k=1	indeterminant	X

Summary of Species Groupings

Test	Sum-of-squares break-point	Homogeneity Index	Fraction of Groups Classified Correctly by the Shape of the Location Graphs
Orloci	1.0	0.81	0.82
	6.0	0.77	0.82
	14.0	0.85	0.67
Fager	-	0.95	0.85

Table 12

Pattern of station numbers used in experiment 2, series 2.

Stations along principal diagonal of Fig. 1

	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
Days 2	11	12	13	14	15	16	17	18	19	20
3	21	22	23	24	25	26	27	28	29	30
4	31	32	33	34	35	36	37	38	39	40
5	41	42	43	44	45	46	47	48	49	50

Stations 1, 11, 21, 31, and 41 are located at the control point. Stations 10, 20, 30, 40, and 50 are located at the outfall.

Table 13

Concordance test on station groups for experiment 2, series 2.

a) Orloci groups:	1) 1 39 2 48 3 6 10 29 4 5 7
(sum-of-squares	2) 9 15 20 16 22 19 23 25 28 30 31
break-point =	3) 8 12 35 32 33 11 26 36 17 34 37 38 13 40 41
10.0)	4) 14 50 18 44 21 43 27 45 49 24 42 46 47

$\bar{x}$ index = 0.57

Grid from Table 12 with group numbers 1-4 inserted:

1	1	1	1	1	1	1	3	2	1
3	3	3	4	2	2	3	4	2	2
4	2	2	4	2	3	4	2	1	2
2	3	3	3	3	3	3	3	1	3
3	4	4	4	4	4	4	1	4	4

Each row is ranked:

4.5	4.5	4.5	4.5	4.5	4.5	4.5	10	9	4.5
6.5	6.5	6.5	9.5	2.5	2.5	6.5	9.5	2.5	2.5
9	4	4	9	4	7	9	4	1	4
2	6.5	6.5	6.5	6.5	6.5	6.5	6.5	1	6.5
2	4	4	4	4	4	4	1	4	4

column sums: 24 25.5 25.5 33.5 21.5 24.5 30.5 31 17.5 21.5 | 255

then the between-columns sum of squares = $(24)^2 + (25.5)^2 + \dots + (21.5)^2 - \frac{(255)^2}{10}$
 ≈ 218.5

Then $\chi^2_9 = \frac{(218.5)(12)}{(5)(10)(11)} = 4.77$ $0.75 < p < 0.90$; no significant difference between columns

Now, each column is ranked

1	1	1	1	1	1	1	3.5	3.5	1	row sums
3.5	3.5	3.5	4	2.5	2	2.5	5	3.5	2.5	15
5	2	2	4	2.5	3.5	4.5	2	1.5	2.5	32.5
2	3.5	3.5	2	4	3.5	2.5	3.5	1.5	4	29.5
3.5	5	5	4	5	5	4.5	1	5	5	30
										43
										150

then the between-rows sum of squares = $(15)^2 + (32.5)^2 + \dots + (43)^2 - ((150)^2/5)$
 $= 400.5$

Then $\chi^2_4 = \frac{(400.5)(12)}{(10)(5)(6)} = 16.02$ $p < .055$; highly significant difference between rows

Table 13 (cont.)

b) Fager groups, including associated species:

- 1) 45 26 5 36 3 6 47 16 46 23 24 13 17 15
- 2) 28 14 33 35 37 4 7
- 3) 43 11 22 31 34
- 4) 40 20 10 30
- 5) 8 38 49 50
- 6) 32 42 12 21
- 7) 25 39 9 19
- 8) 1 41
- 9) 27 29
- 10) 48 18
- 11) 2 44

$\bar{x}$ index = 0.90

Grid from Table 12 with group numbers 1-11 inserted:

8	11	1	2	1	1	2	5	4
3	6	1	2	1	1	1	10	4
6	3	1	1	7	1	9	2	4
3	6	2	3	2	1	2	5	4
8	6	3	11	1	1	1	10	5

Column differences are tested as before, by ranking rows and summing columns.

$$\chi^2_9 = \frac{1150.6(12)}{(5)(10)(11)} = 25.10 \quad p < .005; \text{ highly significant difference between columns}$$

Row differences are tested as before, by ranking columns and summing rows:

$$\chi^2_4 = \frac{(60.5)(12)}{(10)(5)(6)} = 2.42 \quad 0.50 < p < 0.75; \text{ no significant difference between rows}$$

FIGURE LEGENDS

- Fig. 1 - Station numbers and grid pattern used in the present study.
- Fig. 2 - Catch curves for samples generated by program SLIST.
- Fig. 3 - Group location graphs for samples collected along the principal grid diagonal on one day (Series 3, experiment 1).
- Fig. 4 - Species dendrogram for the Orloci test with data from species collected along the principal grid diagonal on five successive days.
- Fig. 5 - Station dendrogram for the Orloci test with data from species collected along the principal grid diagonal on five successive days (Series 2, experiment 2).
- Fig. 6 - Group location graphs for the lowest-level Orloci groups (sum-of-squares cut-off at 1.0), taken from the dendrogram in Fig. 4.
- Fig. 7 - Group location graphs for the Orloci groups with a sum-of-squares cut-off at 6.0, taken from the dendrogram in Fig. 4.
- Fig. 8 - Group location graphs for the Orloci groups with a sum-of-squares cut-off at 14.0, taken from the dendrogram in Fig. 4.
- Fig. 9 - Group location graphs for the Fager groups extracted from the data of Series 2, experiment 2.

Fig. 1

Outfall

91	92	93	94	95	96	97	98	99	100
81	82	83	84	85	86	87	88	89	90
71	72	73	74	75	76	77	78	79	80
61	62	63	64	65	66	67	68	69	70
51	52	53	54	55	56	57	58	59	60
41	42	43	44	45	46	47	48	49	50
31	32	33	34	35	36	37	38	39	40
21	22	23	24	25	26	27	28	29	30
11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	6	7	8	9	10

**Control
Point**

Fig. 2

Fig. 3

Fig. 5

Fraction of Group Present

Fig. 6

C = control station

O = outfall station

Fig. 7

Fig. 8

Fraction
of
Group
Present

APPENDIX A

COMPUTER PROGRAMS USED IN THE STUDY

Note: These programs were written in FORTRAN IV (extended) and run interactively through the NOS system, using the University of Georgia's CDC Cyber 70 computer. Some of the statements and library routines in the following programs are specific for this system.

The author will be glad to assist anyone wishing to use these programs on other systems, or to adapt them for other purposes.

PROGRAM SLIST

Purpose: Generates a series of species lists representing collections of species taken from specified sample sites on a grid (see Fig. 1).

Input:

line 130 - ISTATS = number of stations to be sampled on any day
(64 max.)

IDAY = number of days this sampling program is to be repeated

lines 140-159 - I = 1, ISTATS. Data consist of (x, y) coordinates of the sample sites on the grid.

lines 160-200 - species distribution data (here, the numbers in Table 2, read in as columns).

Limits: 112 species, 64 stations. Can be increased by changing the DIMENSION statements, but the program uses a different algorithm (line 530), designed for zooplankton samples, for species beyond 112.

Output: Data file FDATA, which lists the collection data in the following format (11I3):

1	14	1	4	7	23	24	38	51
1	14	66	90	101	103	112	0	0
↑	↑							
day	sample	species collected						
	site							

Each record begins with a day number and site number, although several such records may be required to list all of the species numbers collected.

at that site. If the number of the last species falls in the last position of a record, the program will insert a blank record, with only the day and site numbers, before beginning to list the next station. The last record of a sampling run is completely blank. These formats are required by programs that follow.

The program also prints out a list of all the different species numbers encountered in the sampling run.


```

0100 PROGRAM SLIST(FDATA,OUTPUT,TAPE6=FDATA)
0110 DIMENSION ISP(112,112),RANS(100),IGRID(64,2),LIST(112),
0120+ITOT(6400),IDIF(112),KDIF(112),IDUM(9)
0130 DATA ISTATS,IDAY,MARK1/64,1,1/
0140 DATA((IGRID(I,J),J=1,2),I=1,64)/1,1,2,1,1,2,3,1,2,2,1,3,4,1,3,2,
0150+2,3,1,4,5,1,4,2,3,3,2,4,1,5,6,1,5,2,4,3,3,4,2,5,1,6,5,3,4,4,3,5,
0155+6,2,6,3,5,4,5,5,4,5,3,6,2,6,4,6,10,10,9,10,10,9,8,10,9,9,10,8,
0156+7,10,8,9,9,8,10,7,6,10,7,9,8,8,9,7,10,6,5,10,6,9,7,8,8,7,9,6,
0157+10,5,6,8,7,7,8,6,5,9,5,8,6,7,6,6,7,6,8,5,9,5,5,7/
0160 DATA((ISP(I,J),I=1,112),J=1,4)/
0170+28*0,42*1,42*2,28*0,14*10,14*5,14*1,14*10,14*5,14*1,
0180+4*2,4*3,4*4,4*5,4*6,4*7,4*8,6*(2,2,3,3,4,4,5,5,6,6,
0190+7,7,8,8),1,2,3,4,2,3,4,5,3,4,5,6,4,5,6,7,5,6,7,8,
0200+6,7,8,9,7,8,9,10,6*(2,3,3,4,4,5,5,6,6,7,7,8,8,9)/
0230 CALL RANSET(TIME(A))
0240 DO 100 I=1,IDAY
0250 DO 10 K=1,100
0260 10 RANS(K)=RANF(SEED)
0270 DO 90 K=1,ISTATS
0280 IX=IGRID(K,1)
0290 IY=IGRID(K,2)
0300 CX=IX
0310 CY=IY
0320 NSTA=(10*(IY-1)+IX)
0330 KCUNT=0
0340 IF(K.GT.ISTATS)100,15
0350 15 DO 80 NSP=1,112
0360 IF(ISP(NSP,1).LT.1)20,55
0370 20 PDIST=1.
0380 21 RSUM=0.
0390 DO 25 M=1,12
0400 25 RSUM=RSUM+RANF(SEED)
0410 RSUM=RSUM-6.
0420 XBAR=FLOAT(10**ISP(NSP,3))

```

```

0430 SD=FLOAT(10**ISP(NSP,4))
0440 CON=RSUM*SD+XBAR
0450 IF(CON.LT.0.)CON=0.
0460 IF(CON.GT.1000000.)251,252
0470 251 BUGS=1000000.
0480 GO TO 27
0490 252 BUGS=CON
0500 27 IF(NSP.LT.113)30,35
0510 30 PAB=1.-0.99995**BUGS
0520 GO TO 40
0530 35 PAB=1.-0.984375** (100.*BUGS)
0540 40 PTJT=PAB*PDIST
0550 IF(PTJT.LT.RANS(NSTA))80,45
0560 45 KOUNT=KOUNT+1
0570 LIST(KOUNT)=NSP
0580 GO TO 80
0590 55 IF(ISP(NSP,1).GT.1)60,65
0600 60 DIST=SQRT((10.-CX)**2+(10.-CY)**2)
0610 GO TO 70
0620 65 DIST=SQRT((1.-CX)**2+(1.-CY)**2)
0630 70 COEFF=-0.1*FLOAT(ISP(NSP,2))
0640 PDIST=EXP(COEFF*DIST)
0650 GO TO 21
0660 80 CONTINUE
0670 MOR=KOUNT+1
0680 DO 81 MZ=MOR,112
0690 81 LIST(MZ)=0
0700 MARK2=MARK1+KOUNT-1
0710 KK=1
0720 DO 82 LL=MARK1,MARK2
0730 ITOT(LL)=LIST(KK)
0740 KK=KK+1

```

```
0750 82 CONTINUE
0760 MARK1=MARK2+1
0770 KEEP=1
0780 DO 89 LL=1,21
0790 L=LL-1
0800 LO=L*9+1
0810 IHI=LO+8
0820 IF(LIST(LO))88,83
0830 83 IF(KEEP)84,90
0840 84 WRITE(6,85)I,NSTA,(IDUM(KWIK),KWIK=1,9)
0850 85 FORMAT(11I3)
0860 GO TO 90
0870 88 WRITE(6,85)I,NSTA,(LIST(M),M=LO,IHI)
0880 KEEP=LIST(IHI)
0890 89 CONTINUE
0900 90 CONTINUE
0910 100 CONTINUE
0920 JUMP=1
0930 IDIF(1)=ITOT(1)
0940 DO 110 I=2,6400
0950 K=I-1
0960 IF(ITOT(I).EQ.0)115,102
0970 102 DO 105 J=1,K
0980 IF(ITOT(I).EQ.ITOT(J))110,105
0990 105 CONTINUE
1000 JUMP=JUMP+1
1010 IDIF(JUMP)=ITOT(I)
1020 110 CONTINUE
1030 115 L=1
1040 DO 150 K=1,112
1050 120 DO 140 M=1,JUMP
1060 IF(IDIF(M).EQ.K)130,140
1070 130 KDIF(L)=K
```

```
1080 L=L+1
1090 140 CONTINUE
1100 150 CONTINUE
1110 160 PRINT 165,JUMP,(KDIF(I),I=1,JUMP)
1120 165 FORMAT(/,1X,13HTHE FOLLOWING,I4,1X,
1130+23HDIFFERENT SPECIES FOUND,/,10(2X,I3))
1140 170 I=0
1150 NSTA=0
1160 WRITE(6,85)I,NSTA,(IDUM(K),K=1,9)
1170 STOP
1180 END
```

PROGRAM DISTS

Purpose: Calculates the arc distances, along a unit hypersphere surface, between the position vectors for any two objects (e.g. stations).

The data are in the form of the upper half of a distance matrix, including the principal diagonal. That is, if there are 3 objects, the distance matrix will be:

		Object		
		a	b	c
Object	a	o	a → b	a → c
	b	b → a	o	b → c
	c	c → a	c → b	o

Program output will be in the format (F10.8):

```
o      a → b      a → c      o      b → c      o
```

Input: Data file FDATA, read as TAPE5.

line 130 - NSAMP = # of objects (e.g. stations)

line 140 - NSP = # of attributes (e.g. species, in the entire sample set).

line 145 - If KPR is set > 0, the output will be printed as well as written in a data file. KPR = 0 suppresses printing.

Limits: 112 objects, 112 attributes.

Output: Data file ODATA, in the format as given above.

```

0100 PROGRAM DISTS(FDATA,ODATA,OUTPUT,TAPE5=FDATA,TAPE6=ODATA)
0110 DIMENSION NET(112,112),KSP(9),KKSP(9),VECT(112),SDIST(112,112)
0120 REWIND 5
0125 KURV=1
0130 NSAMP=82
0140 NSP=32
0145 KPR=0
0150 DO 10 I=1,112
0160 DO 10 J=1,112
0170 10 NET(I,J)=0
0180 KSAM=1
0190 READ(5,20)IDAY1,LOC1,(KSP(I),I=1,9)
0200 15 READ(5,20)IDAY2,LOC2,(KKSP(I),I=1,9)
0210 20 FORMAT(11I3)
0220 30 DO 40 I=1,9
0230 IF(KSP(I))35,40
0240 35 NET(KSAM,KSP(I))=1
0250 40 CONTINUE
0260 IF(IDAY2)45,80
0270 45 IF(LOC2.EQ.LOC1)50,55
0280 50 IF(KKSP(1))60,70
0290 55 KSAM=KSAM+1
0300 60 IDAY1=IDAY2
0310 LOC1=LOC2
0320 DO 65 I=1,9
0330 65 KSP(I)=KKSP(I)
0340 GO TO 15
0350 70 READ(5,20)IDAY2,LOC2,(KKSP(I),I=1,9)
0360 IF(IDAY2)55,80
0370C DISTANCE CALCULATIONS
0380 80 DO 90 I=1,NSAMP
0390 KSUM=0
0400 DO 85 J=1,NSP
0410 KSUM=KSUM+NET(I,J)

```

0420C ALL VALUES EITHER 0 OR 1 SO SQUARING UNNECESSARY

0430 85 CONTINUE

0435 IF(KSUM)87,88

0440 87 VECT(I)=SQRT(FLOAT(KSUM))

0445 GO TO 90

0447 88 VECT(I)=0.

0450 90 CONTINUE

0460 NMIN=NSAMP-1

0470 DO 150 I=1,NMIN

0480 NPLUS=I+1

0490 DO 140 L=NPLUS,NSAMP

0500 SUM=0

0510 DO 130 J=1,NSP

0512 IF(VECT(I))100,95

0514 95 TERM1=0.

0516 GO TO 105

0518 100 TERM1=FLOAT(NET(I,J))/VECT(I)

0520 105 IF(VECT(L))115,110

0522 110 TERM2=0.

0524 GO TO 120

0526 115 TERM2=FLOAT(NET(L,J))/VECT(L)

0528 120 BITE=(TERM1-TERM2)**2

0530 SUM=SUM+BITE

0540 130 CONTINUE

0550 SDIST(I,L)=SQRT(SUM)

0560 SDIST(L,I)=SDIST(I,L)

0570 140 CONTINUE

0580 SDIST(I,I)=0.

0590 150 CONTINUE

0600 IF(KURV.GT.0)180,150

0610 160 DO 170 M=1,NSAMP

0620 DO 170 N=M,NSAMP

0630 WRITE(6,165)SDIST(M,N)

0640 165 FORMAT(F10.8)

0645 IF(KPR)168,170

0650 168 PRINT 165,SDIST(M,N)

0660 170 CONTINUE

0670 GO TO 200

0680 180 DO 190 M=1,NSAMP

0690 DO 190 N=M,NSAMP

0700 SDIST(M,N)=ACOS(1.-(SDIST(M,N)**2)/2.)

0710 190 CONTINUE

0720 GO TO 160

0730 200 STOP

0740 END

PROGRAM ORLOCI

Purpose: Carries out the Orloci sum-of-squares agglomeration on the objects whose relative distances were determined by DIST. See Orloci (1967).

Input: Data file ODATA, read as TAPE5.

Line 130 - NSAMP = # of objects being grouped

Limits: 112 objects. This program requires large random-access arrays, and is usually limited by the size of the computer system.

Output: A set of data is printed for each fusion of species or groups; there will be $N-1$ such fusions, where N = number of objects being grouped. Each set of data contains the number of the fusion, the groups being fused (each identified by the first object in it), the within-group sum of squares for the newly-fused group, and a list of all the species in the new group.

When drawing the dendrogram for the groups, the objects should be listed across the x-axis in the order in which they are given in the last fusion. Note that this program lists objects from 1 to N , e.g. 1, 2, 3, ... N . If the objects are stations, and the stations have grid identifications (from SLIST), these numbers must be inserted afterwards, since they are not part of the input for ORLOCI. e.g., "object No. 50" in ORLOCI may represent "sample site 99" or whatever, using the notation in Fig. 1.

```

0100 PROGRAM ORLOCI(ODATA=65,OUTPUT=65,TAPE5=ODATA)
0110 DIMENSION D(112,112),IR(112,112),N(112),Q(112),A(112,112),IX(112)
0120 REWIND 5
0130 NSAMP=115
0140 PRINT 10,NSAMP
0150 10 FORMAT(1X,19HNUMBER OF OBJECTS =,I3)
0160 DO 40 J=1,NSAMP
0170 DO 30 K=J,NSAMP
0180 READ(5,20)EUCLID
0190 20 FORMAT(F10.8)
0200 EL=EUCLID**2
0210 D(J,K)=EL/2.
0220 D(K,J)=EL/2.
0230 A(J,K)=EL
0240 A(K,J)=EL
0250 30 CONTINUE
0260 40 CONTINUE
0270C SUM OF SQUARES IS NOW IN ARRAY D
0280 ID=0
0290 DO 60 I=1,NSAMP
0300 N(I)=1
0310 Q(I)=0.
0320 DO 50 J=1,NSAMP
0330 IR(I,J)=0
0340 50 CONTINUE
0350 60 CONTINUE
0360 DO 70 I=1,NSAMP
0370 70 IR(I,1)=I
0380C SEARCH FOR VALID FUSIONS
0390 72 ID=ID+1
0400 QLIM=10.**10
0410 LESS=NSAMP-1
0420 DO 100 J=1,LESS
0430 IF(IR(J,1).EQ.0)100,75

```

```

0440 75 JPLUS=J+1
0450 DO 90 K=JPLUS,NSAMP
0460 IF(IR(K,1).EQ.0)90,80
0470 80 W=D(J,K)-Q(J)-Q(K)
0480 IF(W.GE.QLIM)90,85
0490 85 QLIM=W
0500 L=J
0510 M=K
0520 90 CONTINUE
0530 100 CONTINUE
0540 KOUNT=0
0550 LO=N(L)+1
0560 IHI=N(L)+N(M)
0570 DO 110 I=LO,IHI
0580 KOUNT=KOUNT+1
0590 IR(L,I)=IR(M,KOUNT)
0600 110 IR(M,KOUNT)=0
0610 PRINT 120,ID
0620 120 FORMAT(/,1X,15HCLUSTERING PASS,1X,I2)
0630 N(L)=N(L)+N(M)
0640 N(M)=0
0650 Q(L)=D(L,M)
0660 Q(M)=0.
0670 DO 130 I=1,NSAMP
0680 D(M,I)=0.
0690 130 D(I,M)=0.
0700 PRINT 140,L,M
0710 140 FORMAT(1X,18HGROUPS IN FUSION (,I2,3H + ,I2,1H))
0720 PRINT 150,N(L)
0730 150 FORMAT(1X,28HNUMBER OF OBJECTS IN GROUP =,1X,I2)
0740 PRINT 160,Q(L)
0750 160 FORMAT(1X,16HSUM OF SQUARES =,1X,F10.4)

```

```
0760 NEND=N(L)
0770 PRINT 170,(IR(L,J),J=1,NEND)
0780 170 FORMAT(1X,7HOBJECTS,(/,10(3X,I2)))
0790C GENERATE NEW D MATRIX
0800 DO 250 I=1,NSAMP
0810 SUM=0.
0820 IF(IR(I,1).EQ.0)250,180
0830 180 IF(I.EQ.L)250,190
0840 190 NEND=N(I)
0850 DO 200 J=1,NEND
0860 KSTOR=J
0870 200 IX(J)=IR(I,J)
0880 NEND=N(L)
0890 DO 210 IH=1,NEND
0900 KSTOR=KSTOR+1
0910 210 IX(KSTOR)=IR(L,IH)
0920 NEND=N(I)+N(L)-1
0930 DO 230 J=1,NEND
0940 LA=IX(J)
0950 JPLUS=J+1
0960 JEND=N(I)+N(L)
0970 DO 220 IH=JPLUS,JEND
0980 LB=IX(IH)
0990 SUM=SUM+A(LA,LB)
1000 220 CONTINUE
1010 230 CONTINUE
1020 D(I,L)=SUM/FLOAT(N(I)+N(L))
1030 D(L,I)=D(I,L)
1040 250 CONTINUE
1050 IF(ID.EQ.(NSAMP-1))260,72
1060 260 STOP
1070 END
```

PROGRAM FREV

Purpose: changes data file FDATA from one in which a station number is followed by a list of species, to one in which a species number is followed by the list of stations at which it occurs.

Input: Data file FDATA, read as TAPE5.

Line 125 -if KPRNT is set > 0, the output will be printed as well as read onto a file. KPRNT = 0 suppresses printing.

Line 130 - NSAMP = no. of samples in FDATA

Line 140 - NSP = no. of species in FDATA

Limits: 112 stations, 112 species, 63 species in any one station.

Output: Data file FDATR, in the same format as FDATA. Note that the new file must be renamed FDATA before it can be used as input for programs such as DISTS.

```

0100 PROGRAM FREV(FDATA,FDATR,OUTPUT,TAPE5=FDATA,TAPE6=FDATR)
0110 DIMENSION NET(112,112),LOT(63),NONE(11),KSP(9),KKSP(9)
0115 DATA NONE/11*0/
0120 REWIND 5
0125 KPRNT=0
0130 NSAMP=32
0140 NSP=112
0145 LEV=1
0150 DO 10 I=1,NSAMP
0160 DO 10 J=1,NSP
0170 10 NET(I,J)=0
0180 KSAM=1
0190 READ(5,20)IDAY1,LOC1,(KSP(I),I=1,9)
0200 15 READ(5,20)IDAY2,LOC2,(KKSP(I),I=1,9)
0210 20 FORMAT(11I3)
0220 30 DO 40 I=1,9
0230 IF(KSP(I))35,40
0240 35 NET(KSAM,KSP(I))=1
0250 40 CONTINUE
0260 IF(IDAY2)45,80
0270 45 IF(KKSP(1))50,70
0280 50 IF(LOC2.EQ.LOC1)60,55
0290 55 KSAM=KSAM+1
0300 60 IDAY1=IDAY2
0310 LOC1=LOC2
0320 DO 65 I=1,9
0330 65 KSP(I)=KKSP(I)
0340 GO TO 15
0350 70 READ(5,20)IDAY2,LOC2,(KKSP(I),I=1,9)
0360 IF(IDAY2)55,80
0370 80 DO 300 I=1,NSP
0380 KOUNT=0
0390 DO 85 K=1,63
0400 85 LOT(K)=0

```

```
0410 DO 100 J=1,NSAMP
0420 IF(NET(J,I))90,100
0430 90 KOUNT=KOUNT+1
0440 LOT(KOUNT)=J
0450 100 CONTINUE
0455 IF(LOT(1))110,300
0460 110 WRITE(6,115) LEV,I,(LOT(J),J=1,9)
0465 IF(KPRNT)111,112
0470 111 PRINT 115,LEV,I,(LOT(J),J=1,9)
0480 115 FORMAT(11I3)
0490 112 IF(KOUNT.LT.9)300,120
0500 120 IF(KOUNT.EQ.9)290,130
0510 130 WRITE(6,115) LEV,I,(LOT(J),J=10,18)
0515 IF(KPRNT)131,132
0520 131 PRINT 115,LEV,I,(LOT(J),J=10,18)
0530 132 IF(KOUNT.LT.18) 300,140
0540 140 IF(KOUNT.EQ.18)290,150
0550 150 WRITE(6,115)LEV, I,(LOT(J),J=19,27)
0555 IF(KPRNT)151,152
0560 151 PRINT 115,LEV,I,(LOT(J),J=19,27)
0570 152 IF(KOUNT.LT.27)300,160
0580 160 IF(KOUNT.EQ.27)290,170
0590 170 WRITE(6,115)LEV, I,(LOT(J),J=28,35)
0600 IF(KPRNT)171,172
0610 171 PRINT 115,LEV,I,(LOT(J),J=28,36)
0620 172 IF(KOUNT.LT.36)300,180
0630 180 IF(KOUNT.EQ.36)290,190
0640 190 WRITE(6,115)LEV, I,(LOT(J),J=37,45)
0650 IF(KPRNT)191,192
0660 191 PRINT 115,LEV,I,(LOT(J),J=37,45)
0670 192 IF(KOUNT.LT.45)300,200
0680 200 IF(KOUNT.EQ.45)290,210
```

0690 210 WRITE(6,115)LEV,I,(LOT(J),J=46,54)

0700 IF(KPRNT)211,212

0710 211 PRINT 115,LEV,I,(LOT(J),J=46,54)

0720 212 IF(KOUNT.LT.54)300,220

0730 220 IF(KOUNT.EQ.54)290,230

0740 230 WRITE(6,115)LEV,I,(LOT(J),J=55,63)

0750 IF(KPRNT)231,290

0760 231 PRINT 115,LEV,I,(LOT(J),J=55,63)

0770 290 WRITE(6,115)LEV,I,(NONE(J),J=1,9)

0780 IF(KPRNT)291,300

0790 291 PRINT 115,LEV,I,(NONE(J),J=1,9)

0800 300 CONTINUE

0810 WRITE(6,115)(NONE(J),J=1,11)

0820 IF(KPRNT)301,302

0830 301 PRINT 115,(NONE(J),J=1,11)

0840 302 STOP

0850 END

PROGRAM AFFIN

Purposes: Uses a Monte Carlo procedure to calculate a cumulative frequency distribution of affinity indices. The program calculates the number of occurrences of each species in the sample set. Using a random number generator, it then distributes each species independently among an equivalent number of storage locations, using the same frequency distribution. Affinity indices are then calculated for all possible species pairs, and the percentage of them in each of twenty size categories is calculated. An affinity index can then be chosen for REGRO which has a known probability of occurring by chance alone.

Input: Data file FDATA read as TAPE5

Line 120: the dimensions of NET must be set to (# samples, # species)

Line 130: N = # species

NSAM = # samples

NRUN = # of random runs (10 is usually sufficient)

Limits: 112 species, 112 samples

Output: Prints the number of occurrences of each species in the sample set. Gives a table showing the percentage of the calculated affinity indices that are less than .05, .10, .15, ..., .95, 1.00. For instance, the level of the affinity index opposite .95 will represent the level at which there is only a 5% probability of that index occurring by chance alone.

```

0100 PROGRAM AFFIN(FDATA,OUTPUT,TAPE5=FDATA)
0110 DIMENSION XSP(112),CUM(20),SAM(112,112),NS(112),SP(112,112),
0120+FR(20),JSP(112),NET(69,4),KSP(9),KKSP(9)
0130 DATA N,NSAM,NRUN/4,69,10/
0132 DATA(CUM(I),I=1,20)/.05,.1,.15,.2,.25,.3,.35,.4,.45,.5,.55,
0134+.6,.65,.7,.75,.8,.85,.9,.95,1.0/
0140 REWIND 5
0145 CALL RANSET(TIME(A))
0150 DO 200 I=1,NSAM
0160 DO 200 J=1,N
0170 200 NET(I,J)=0
0180 KSAM=1
0185 L=1
0190 READ(5,220)IDAY1,LOC1,(KSP(I),I=1,9)
0200 215 READ(5,220)IDAY2,LOC2,(KKSP(I),I=1,9)
0210 220 FORMAT(11I3)
0220 230 DO 240 I=1,9
0230 IF(KSP(I))235,240
0240 235 NET(KSAM,L)=KSP(I)
0242 L=L+1
0250 240 CONTINUE
0260 IF(IDAY2)245,280
0270 245 IF(KKSP(1))250,270
0280 250 IF(LOC2.EQ.LOC1)260,255
0290 255 KSAM=KSAM+1
0295 L=1
0300 260 IDAY1=IDAY2
0310 LOC1=LOC2
0320 DO 265 I=1,9
0330 265 KSP(I)=KKSP(I)
0340 GO TO 215
0350 270 READ(5,220)IDAY2,LOC2,(KKSP(I),I=1,9)
0360 IF(IDAY2)255,280
0370 280 DO 130 I=1,NSAM

```

```

0380 DO 120 J=1,N
0390 IF(NET(I,J))110,130
0400 110 JSP(NET(I,J))=JSP(NET(I,J))+1
0410 120 CONTINUE
0420 130 CONTINUE
0430 I=1
0440 L=-1
0450 DO 170 L=1,NSAM
0455 K=NSAM+1-L
0460 DO 160 M=1,112
0470 IF(JSP(M).EQ.K)140,160
0480 140 XSP(I)=FLOAT(K)
0490 I=I+1
0500 160 CONTINUE
0510 170 CONTINUE
0520 NUM=I-1
0550 PRINT 35,N,NSAM,NRUN
0560 35 FORMAT(5X,I3,1X,7HSPECIES,10X,I3,1X,7HSAMPLES,10X,I3,
0570+1X,4HRUNS,/)
0580 IF(NUM.EQ.N)360,180
0590 180 PRINT 185,NUM,(XSP(K),K=1,NUM)
0600 185 FORMAT (1X,14HCOUNTING ERROR,5X,13HTHE FOLLOWING,1X,I3,
0610+1X,17HFREQUENCIES FOUND,/,10(2X,F4.0))
0620 360 PRINT 36,(XSP(I),I=1,N)
0630 36 FORMAT(5X,19HSPECIES FREQUENCIES,/,10(2X,F4.0))
0640 XS=NSAM
0650 SX=XS/2.
0660 SX=SX+0.5
0670 MSP=SX
0680 TOTAL=0.
0690 NR=1
0700 34 DO 4 I=1,NSAM

```

```
0710 NS(I)=0
0720 DO 4 J=1,N
0730 4 SAM(I,J)=I.
0740 DO 5 I=1,N
0750 NSP=XSP(I)
0760 IF(NSP.GT.MSP)7,6
0770 7 K=NSAM-NSP
0780 IF(K.EQ.0)5,9
0790 9 KT=1
0800 37 X=RANF(SEED)
0810 X=X*XS
0820 X=X+0.5
0830 NX=X
0840 IF(NX.LT.1)43,44
0850 43 NX=1
0860 44 DO 8 J=1,KT
0870 IF(NS(J).EQ.NX)37,8
0880 8 CONTINUE
0890 NS(KT)=NX
0900 KT=KT+1
0910 IF(KT.GT.K)10,37
0920 10 KT=KT-1
0930 DO 11 II=1,KT
0940 KK=NS(II)
0950 11 SAM(KK,I)=0.0
0960 GO TO 5
0970 6 K=NSP
0980 KT=1
0990 13 X=RANF(SEED)
1000 X=X*XS
1010 X=X+0.5
1020 NX=X
1030 IF(NX.LT.1)45,46
```

```
1040 45 NX=1
1050 46 DO 12 JJ=1,KT
1060 IF(NS(JJ).EQ.NX)13,12
1070 12 CONTINUE
1080 NS(KT)=NX
1090 KT=KT+1
1100 IF(KT.GT.K)14,13
1110 14 KT=KT-1
1120 DO 15 JI=1,NSAM
1130 SAM(JI,I)=0.
1140 DO 16 KI=1,KT
1150 IF(NS(KI).EQ.JI)17,16
1160 17 SAM(JI,I)=1.
1170 GO TO 15
1180 16 CONTINUE
1190 15 CONTINUE
1200 5 CONTINUE
1210 K=N-1
1220 DO 18 I=1,NSAM
1230 NA=1
1240 22 IF(SAM(I,NA).GT.0.)21,23
1250 21 NB=NA+1
1260 DO 19 J=NB,N
1270 IF(SAM(I,J).GT.0.)20,19
1280 20 SP(NA,J)=SP(NA,J)+1.
1290 19 CONTINUE
1300 23 NA=NA+1
1310 IF(NA.GT.K)18,22
1320 18 CONTINUE
1330 DO 24 I=1,K
1340 J=I+1
1350 DO 25 II=J,N
```

```

1360 AFF=XSP(I)*XSP(II)
1370 AFF=SQRT(AFF)
1380 AFF=SP(I,II)/AFF
1390 SP(I,II)=0.
1400 IF(XSP(I).GE.XSP(II))26,27
1410 26 COR=SQRT(XSP(I))
1420 GO TO 28
1430 27 COR=SQRT(XSP(II))
1440 28 COR=1.7/(2.*COR)
1450 AFF=AFF-COR
1460 DO 29 JJ=1,20
1470 IF(AFF.LT.CUM(JJ))30,29
1480 30 FR(JJ)=FR(JJ)+1.
1490 TOTAL=TOTAL+1.
1500 GO TO 25
1510 29 CONTINUE
1520 25 CONTINUE
1530 24 CONTINUE
1540 NR=NR+1
1550 IF(NR.GT.NRUN)33,34
1560 33 PRINT 100
1570 100 FORMAT(///)
1580 PRINT 41,TOTAL
1590 41 FORMAT(5X,F10.0,2X,16HAFFINITY INDICES)
1600 DO 42 I=1,19
1610 J=I+1
1620 42 FR(J)=FR(J)+FR(I)
1630 DO 31 I=1,20
1640 FR(I)=FR(I)/TOTAL
1650 PRINT 32,CUM(I),FR(I)
1660 32 FORMAT(/,5X,16HPROPORTION BELOW,F7.2,5X,F7.3)
1670 31 FR(I)=0.
1680 39 CONTINUE

1690 STOP
1700 END

```

PROGRAM REGRO (REGROUP)

Purposes: Determines the set of groups of objects as defined by Fager (1957) and Fager and McGowan (1963).

Input: Data file FDATA, read as TAPE5.

Line 7: NX = number of species to be grouped

Z = cut-off value of the affinity index (from AFFIN)

Line 237: (in subroutine ORDERB)

MSKIP set greater than 0 will cause printing of a list showing the number of occurrences and number of affinities of each species.

First statement of subroutine CBIN:

CMAX is a power of 10 defining the maximum number of calculations that will be attempted without an error message.

Limits: 112 species as written. Can be extended to 200 by changing the dimension of A so it is no less than $N(N-1)/2$, where N is the number of species.

Output: Lists each recurrent group as it is calculated, and associated species, if any. Error messages will be printed if the number of species encountered in FDATA is different than that given in line 7. An error message will be printed if the number of calculations exceeds that defined by CMAX.

```

PROGRAM REGROUP(FDATA,OUTPUT,TAPE5=FDATA,TAPE6=OUTPUT)
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1  NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2  NSP(200), LX, JX
3  /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4  /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
DATA NX,Z,LRUN/32,.45,0/
REWIND 5
1000 PRINT 10,NX,Z
10 FORMAT (1H0, 10X, 17HNUMBER OF SPECIES, 14, 1X, 18HSIGNIFICANCE PO
1 INT, F5.4, /)
NY = (NX * (NX-1)) / 2
DO 3 NN = 1, NY
3 A(NN) = 0.0
DO 4 KK = 1, NX
4 KOUNT(KK) = 0
CALL INPUTB
CALL TESTA
LEVEL = 0
CALL ROWSUM
8 CALL ZEROWS
CALL ORDERB
CALL SIZEG
CALL LISTG
CALL RJOINT
CALL ZERONN
DO 6 NN = 1, NY
Q = ABS(A(NN) - Z)
IF (A(NN) .GE. Z .OR. Q .LT. 0.000001) 5, 6
5 CALL ROWSUM
CALL RELATED
GO TO 8
6 CONTINUE
DO 11 I = 1, NO
11 NAFF(I) = 0
CALL RELATED
IF(LRUN) 2000, 2000, 1000
2000 PRINT 7
7 FORMAT (1H0, 20X, 18HGROUPING COMPLETED)
END
SUBROUTINE INPUTB
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(400),
1  MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2  NSP(200), LX, JX
3  /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4  /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
NO = 0
NCOUNT = 0
READ (5,300) KR, IHAUL, (NSP(I), I = 1, 9)

300 FORMAT(11I3)
301 NKR = KR
NHAUL = IHAUL
LTEMP = 0
IF(NKR) 302, 316
302 DO 304 KK = 1, NX
304 LATCH(KK) = 0
303 DO 305 I = 1, 9
IF(NSP(I)) 306, 305
306 IF (NJ) 308, 307
308 DO 309 KK = 1, NO
IF (NSP(I) .EQ. NSPD(KK)) 310, 309
310 IF (LATCH(KK)) 305, 311
309 CONTINUE
307 KK = NJ + 1
NSPD(KK) = NSP(I)
NO = KK

```



```

311 IF (NO .GT. NX) 318, 311
    KOUNT(KK) = KOUNT(KK) + 1
    LATCH(KK) = 1
    LTEMP = LTEMP + 1
    ITEMP(LTEMP) = KK
305 CONTINUE
317 READ (5,300) KR, IHAUL, (NSP(I), I = 1, 9)
    IF (KR .EQ. NKR) 313, 312
313 IF (IHAUL .EQ. NHAUL) 303, 312
312 NJ = LTEMP - 1
    NCOUNT = NCOUNT + 1
    IF (NJ) 315, 301
315 DO 320 J = 1, NJ
    K8 = J + 1
    DO 320 I = K8, LTEMP
    NN = NUM(ITEMP(I), ITEMP(J))
320 A(NN) = A(NN) + 1.0
    GO TO 301
318 CALL EXTRASP
    STOP
316 PRINT 321, NCOUNT
321 FORMAT (1H0, 27X, 23HTOTAL NUMBER OF SAMPLES, I5)
    RETURN
END
SUBROUTINE EXTRASP
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(400),
1 MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4 /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
    IF (NO .GT. 200) 304, 301
304 PRINT 302
302 FORMAT (1H0, 5X, 35HMORE THAN TWO HUNDRED SPECIES FOUND)
    STOP
301 DO 305 I = 1, 9
    IF (NSP(I)) 308, 305
308 DO 309 KK = 1, NO
    IF (NSP(I) .EQ. NSPD(KK)) 305, 309
309 CONTINUE
    KK = NO + 1
    NSPD(KK) = NSP(I)
    NO = KK
    IF (NO .GT. 200) 304, 305
305 CONTINUE
    READ (5,300) KR, IHAUL, (NSP(I), I = 1, 9)
300 FORMAT(11I3)
    IF (KR) 301, 310
310 PRINT 311, NO
311 FORMAT (1H0, 5X, 13HTHE FOLLOWING, I4, 1X, 13HSPECIES FOUND)
    PRINT 312, (NSPD(I), I = 1, NO)
312 FORMAT(1H0,10I5)
    RETURN
END
FUNCTION NUM(IDUM, JDUM)
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1 MAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4 /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
    IF (IDUM .GE. JDUM) 482, 481
481 MAX = JDUM
    MIN = IDUM
    GO TO 483
482 MAX = IDUM
    MIN = JDUM

```

```

483 NUM = 0
   IF (MIN .GT. 1) 486, 485
486 IP = NX - MIN
   NP = (IP * (IP+1)) / 2
   NUM = NY - NP
485 NUM = NUM + MAX - MIN
   RETURN
   END
   SUBROUTINE TESTA
   COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1   NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2   NSP(200), LX, JX
3   /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4   /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
   IF (NO .EQ. NX) 412, 414
414 PRINT 415, NO
415 FORMAT (1H0, 5X, 18HONLY THE FOLLOWING, I4, 1X, 13HSPECIES FOUND)
   PRINT 416, (NSPD(I), I = 1, NO)
416 FORMAT(1H0,10I5)
   STOP
412 IROW = 1
   JCOL = 2
   ZA = 1.0 / (Z**2)
   DO 400 NN = 1, NY
   IF (A(NN)) 403, 409
403 IF (KOUNT(JCOL) .LT. KOUNT(IROW)) 405, 406
405 D = KOUNT(IROW)
   C = KOUNT(JCOL)
   GO TO 407
406 D = KOUNT(JCOL)
   C = KOUNT(IROW)
407 IF ((D/C) .LT. ZA) 408, 409
408 IF ((A(NN)/C) .LT. Z) 409, 410
409 TEST = 0.0
   GO TO 411
410 TEST = (A(NN) / (SQRT(C*D))) - (1.0 / (2.0 * SQRT(D)))
411 A(NN) = TEST
   JCOL = JCOL + 1
   IF (JCOL .GT. NO) 413, 400
413 IROW = IROW + 1
   JCOL = IROW + 1
400 CONTINUE
   RETURN
   END
   SUBROUTINE ROWSUM
   COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1   NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2   NSP(200), LX, JX
3   /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4   /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
   DO 55 K = 1, NO
55 NAFF(K) = 0
   IB = 1
   JB = 2
   DO 56 NN = 1, NY
   Q = ABS(A(NN) - Z)
   IF (A(NN) .GE. Z .OR. Q .LT. 0.000001) 57, 58
57 NAFF(IB) = NAFF(IB) + 1
   NAFF(JB) = NAFF(JB) + 1
58 JB = JB + 1
   IF (JB .GT. NO) 59, 56
59 IB = IB + 1
   JB = IB + 1
56 CONTINUE
   RETURN

```

```

END
SUBROUTINE ZEROWS
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1  NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2  NSP(200), LX, JX
3  /2/ LATCH(200) /3/ LAFF(200) /4/ LARK(200)
4  /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT.
NNOT = 0
JCT = 0
LCT = 0
DO 25 I = 1, NO
IF (NAFF(I)) 26, 23
23 IF (LEVEL) 25, 24
24 LCT = LCT + 1
LX = LCT
LARK(LCT) = I
LGROUP(LCT) = NSPD(I)
GO TO 25
26 JCT = JCT + 1
JX = JCT
MAFF(JCT) = NAFF(I)
MARK(JCT) = I
NNOT = NNOT + 1
25 CONTINUE
IF (NNOT) 30, 28
28 PRINT 29
29 FORMAT (26HNO SPECIES HAVE AFFINITIES)
STOP
30 RETURN
END
SUBROUTINE ORDERB
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1  NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2  NSP(200), LX, JX
3  /2/ ASAV(200) /3/ LAFF(200) /4/ LARK(200)
4  /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
5  /AB/ MGROUP(200), ISP(3,200), NNSP(200)
DIMENSION LCOUNT (200), MCOUNT(200)
DATA MSKIP/0/
LEVEL = LEVEL + 1
201 IK = 0
NS = JX - 1
106 DO 104 I = 1, NS
K = I + 1
IF (MAFF(I) .LT. MAFF(K)) 105, 104
105 NSAV = MAFF(I)
MAFF(I) = MAFF(K)
MAFF(K) = NSAV
MSAV = MARK(I)
MARK(I) = MARK(K)
MARK(K) = MSAV
IK = IK + 1
104 CONTINUE
IF (IK) 101, 150
101 NS = NS - 1
IK = 0
GO TO 106
150 DO 112 I = 1, JX
MCOUNT(I) = KOUNT(MARK(I))
112 MGROUP(I) = NSPD(MARK(I))
IF (LEVEL .EQ. 1) 151, 102
151 DO 109 I = 1, LX
LAFF(I) = 0
109 LCOUNT(I) = KOUNT(LARK(I))
IF (MSKIP) 1090, 1240

```

2

```

1090 PRINT 110
110 FORMAT (1H0, 10X, 4HCODE, 5X, 11HOCCURRENCES, 5X,
1 10HAFFINITIES, /)
DO 121 I = 1, JX
DO 122 K = 1, NO
IF (NNSP(K) .EQ. MGROUP(I)) 123, 122
122 CONTINUE
123 PRINT 111, MGROUP(I), MCOUNT(I), MAFF(I)
111 FORMAT (1H, 10X, 14, 8X, 15, 11X, 14)
121 CONTINUE
DO 124 I = 1, LX
DO 125 K = 1, NO
IF (NNSP(K) .EQ. LGROUP(I)) 126, 125
125 CONTINUE
126 PRINT 111, LGROUP(I), LCOUNT(I), LAFF(I)
124 CONTINUE
1240 NM = JX + 1
J = 1
DO 131 I = NM, NO
MARK(I) = LARK(J)
MGROUP(I) = LGROUP(J)
131 J = J + 1
DO 113 I = 1, JX
DO 114 J = 1, NO
IF (MARK(J) .EQ. MARK(I)) 115, 116
116 NN = NUM(MARK(I), MARK(J))
ASAV(J) = A(NN)
GO TO 114
115 ASAV(J) = 1.0
114 CONTINUE
113 CONTINUE
GO TO 102
102 RETURN
END
SUBROUTINE SIZEG
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1 NAFF(200), MAFF(200), ITEM(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ ASAV(200) /3/ LAFF(200) /4/ LARK(200)
4 /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
KSIZE = 0
JCT = JX
DO 96 I = 1, JCT
KSIZE = KSIZE + 1
IF (KSIZE .GT. MAFF(I)) 97, 96
97 IF (KSIZE .EQ. (MAFF(I) + 1)) 98, 99
96 CONTINUE
98 MG = KSIZE - 1
GO TO 95
99 MG = KSIZE - 2
95 RETURN
END
SUBROUTINE LISTG
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1 NAFF(200), MAFF(200), ITEM(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ ASAV(200) /3/ LAFF(200) /4/ LARK(200)
4 /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
IF (MG) 94, 92
92 PRINT 89
89 FORMAT (1H0, 5X, 18HGROUP SIZE IS ZERO)
STOP
94 LTEMP = 0
INDEXA = 0
JCT = JX

```

```

DO 90 I = 1, JCT
IF (MAFF(I) .LT. MG) 93, 91
91 LTEMP = LTEMP + 1
   LX = LTEMP
   ITEMP(LTEMP) = MARK(I)
90 CONTINUE
93 MGM = MG + 1
   RETURN
END
SUBROUTINE RJOINT
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1   NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2   NSP(200), LX, JX
3   /2/ ASAV(200) /3/ LAFF(200) /4/ LLIST(200)
4   /5/ MG, INDEXA, INP(100), IGUT(100), JJ, IA, NCOUNT
648 LTEMP = LX
DO 630 I = 1, LTEMP
   NAFF(I) = 0
   LAFF(I) = 0
630 LGROUP(I) = 0
   IL = 0
   NJ = LTEMP - 1
DO 631 I = 1, NJ
   KA = I + 1
DO 631 J = KA, LTEMP
   NN = NUM(ITEMP(I), ITEMP(J))
   Y = ABS(A(NN) - Z)
   IF (A(NN) .GE. Z .OR. Y .LT. 0.000001) 632, 631
632 NAFF(I) = NAFF(I) + 1
   NAFF(J) = NAFF(J) + 1
631 CONTINUE
DO 633 I = 1, LTEMP
   IF (NAFF(I) .GE. MG) 634, 633
634 IL = IL + 1
   LX = IL
   LAFF(IL) = NAFF(I)
   LGROUP(IL) = ITEMP(I)
633 CONTINUE
   IF (IL .GE. (MG+1)) 635, 636
636 MG = MG - 1
   CALL LISTG
   GO TO 648
635 IF (IL .EQ. LTEMP) 638, 639
639 DO 640 I = 1, IL
640 ITEMP(I) = LGROUP(I)
   GO TO 648
638 IF (IL .EQ. (MG+1)) 641, 642
641 DO 643 I = 1, IL
643 LLIST(I) = NSPD(LGROUP(I))
   PRINT 644, LEVEL
644 FORMAT (1H0, 5X, 43H THE FOLLOWING GROUP HAS BEEN FOUND AT LEVEL,
1   I3)
   PRINT 645, (LLIST(I), I = 1, IL)
645 FORMAT (1H0, 5X, 10I4)
DO 647 I = 1, IL
647 ITEMP(I) = LGROUP(I)
   INDEXA = 1
   GO TO 646
642 CALL PERFECT
   IF (INDEXA) 646, 636
646 RETURN
END
SUBROUTINE PERFECT
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1   NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),

```

```

2   NSP(200), LX, JX
3   /2/ ISAV(200) /3/ LAFF(200) /4/ LLIST(200)
4   /5/ MG, INDEXA, INP(100), IOUT(100), JJ, IA, NCOUNT
   LCOUNT = 0
   MCOUNT = 0
   IL = LX
   DO 800 I = 1, IL
   IF (LAFF(I) .EQ. (IL-1)) 801, 802
801  LCOUNT = LCOUNT + 1
   LLIST(LCOUNT) = NSPD(LGROUP(I))
   NSP(LCOUNT) = LGROUP(I)
   GO TO 799
802  MCOUNT = MCOUNT + 1
   ITEMP(MCOUNT) = LGROUP(I)
799  LGROUP(I) = 0
800  LAFF(I) = 0
   NL = LCOUNT + 1
   NM = LCOUNT + MCOUNT
   IF (LCOUNT .EQ. MG) 828, 829
828  PRINT 832, LEVEL
832  FORMAT (1H0, 5X, 41HTHESE RECURRENT GROUPS ARE FOUND AT LEVEL, I4)
   EDCOUNT = 0.0
   DO 830 I = 1, MCOUNT
   LLIST(NL) = NSPD(ITEMP(I))
   PRINT 817, (LLIST(K), K = 1, NL)
   DDCOUNT = 0.0
   DO 840 J = 1, LCOUNT
   NN = NUM(ITEMP(I), NSP(J))
840  DDCOUNT = DDCOUNT + A(NN)
   IF (DDCOUNT .GT. EDCOUNT) 831, 830
831  EDCOUNT = DDCOUNT
   MLIST = NSPD(ITEMP(I))
   NSP(NL) = ITEMP(I)
830  CONTINUE
   LLIST(NL) = MLIST
   INDEXA = 1
   GO TO 821
829  MMG = MG - LCOUNT
812  NK = MCOUNT - 1
   ML = 0
   DO 803 I = 1, NK
   KB = I + 1
   DO 803 J = KB, MCOUNT
   NN = NUM(ITEMP(I), ITEMP(J))
   Y = ABS(A(NN) - Z)
   IF (A(NN) .GE. Z .OR. Y .LT. 0.000001) 804, 803
804  LAFF(I) = LAFF(I) + 1
   LAFF(J) = LAFF(J) + 1
803  CONTINUE
   DO 805 I = 1, MCOUNT
   IF (LAFF(I) .GE. MMG) 806, 805
806  ML = ML + 1
   LGROUP(ML) = ITEMP(I)
805  CONTINUE
   IF (ML .GE. (MMG + 1)) 807, 808
808  INDEXA = 0
807  IF (ML .EQ. MCOUNT) 809, 810
810  DO 811 I = 1, ML
   ITEMP(I) = LGROUP(I)
   LGROUP(I) = 0
811  LAFF(I) = 0
   MCOUNT = ML
   GO TO 812
809  IF (ML .EQ. (MMG+1)) 813, 814
813  J = 1

```

```

      DO 815 I = NL, NM
      LLIST(I) = NSPD(LGROUP(J))
      NSP(I) = LGROUP(J)
815  J = J + 1
      INDEXA = 1
      PRINT 816, LEVEL
816  FORMAT (1H0, 5X, 43H THE FOLLOWING GROUP HAS BEEN FOUND AT LEVEL,
1     I3)
      PRINT 817, (LLIST(I), I = 1, NM)
817  FORMAT (1H0, 5X, 10I4)
      GO TO 825
814  JJ = MCOUNT
      IA = MMG + 1
      LX = LCOUNT
      INDEXA = 0
      DO 820 I = 1, JJ
820  INP(I) = LGROUP(I)
      CALL CBIN
      IF (INDEXA) 821, 827
821  LX = MG + 1
      PRINT 823
823  FORMAT(1H0, 5X, 40H THE FOLLOWING GROUP WAS SELECTED BECAUSE,
      /*, 10X, 48HITS SPECIES HAVE THE LARGEST TOTAL OF AFFINITIES)
      PRINT 824, (LLIST(I), I = 1, LX)
824  FORMAT (1H0, 5X, 10I4)
      GO TO 826
825  LX = MG + 1
826  DO 818 I = 1, LX
818  ITEMP(I) = NSP(I)
827  RETURN
      END
      SUBROUTINE CBIN
      COMMON /1/ NY, NX, ND, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1     NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2     NSP(200), LX, JX
3     /2/ ISAV(200) /3/ LAFF(200) /4/ LLIST(200)
4     /5/ MG, INDEXA, INP(100), IOUT(100), JJ, IA, NCOUNT
      DIMENSION IN(100)
      DATA IKEEP, CMAX /0, 6.0/
      N = JJ
      K = IA
      L = N - K
      DO 9 I = 1, JJ
9     ISAV(I) = NSPD(INP(I))
      CN = FLOAT (N)
      CK = FLOAT (K)
      CL = FLOAT (L)
      CCN = FLOAT (N) + 0.500
      CCK = FLOAT (K) + 0.500
      CCL = FLOAT (L) + 0.500
      CD = (CCN * ALOG(CN)) - (CCK * ALOG(CK)) - (CCL * ALOG(CL)) -
1     CN + CK + CL
      CA = (CD / 2.3026) - 0.400
19  IKEEP = IKEEP + 1
      IF (CA .GT. CMAX) 14, 15
14  PRINT 17, CA, CMAX
17  FORMAT(1H0, 5X, 14H THE TEN TO THE, F8.3, 21H TH POWER COMBINATIONS,
1     1X, 41H EXCEED THE ASSIGNED MAXIMUM OF TEN TO THE, F8.3, 8H TH PO
2     WER)
      STOP
15  DO 10 J = 1, 100
10  IN(J) = 0
      I = 1
      IN(I) = 1
1     IOUT(I) = INP(IN(I))

```

```

3 IF (I.EQ.K) 2, 3
3 IN(I+1) = IN(I) + 1
IF (IN(I+1).GT.(L+I+1)) 4, 5
5 I = I + 1
GO TO 1
4 I = I - 1
IF (I) 6, 7
7 RETURN
6 IN(I) = IN(I) + 1
IF (IN(I).GT.(L+I)) 4, 1
2 CALL FINDG
GO TO 6
END
SUBROUTINE FINDG
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1 NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ ISAV(200) /3/ KLIST(200) /4/ LLIST(200)
4 /5/ MG, INDEXA, INP(100), IOUT(100), JJ, IA, NCOUNT
DO 701 I = 1, IA
701 ISAV(I) = IOUT(I)
ACOUNT = 0.0
L = IA
NL = L - 1
703 DO 702 LL = 1, NL
NN = NUM(ISAV(LL), ISAV(L))
Y = ABS(A(NN) - Z)
IF (A(NN).GE.Z.OR.Y.LT.0.000001) 702, 704
702 ACOUNT = ACOUNT + A(NN)
L = L - 1
NL = NL - 1
IF (NL) 703, 705
705 IF (INDEXA) 711, 710
710 CONTINUE
BCOUNT = 0.0
INDEXA = 1
IF (LX) 716, 711
716 DO 706 M = 1, LX
706 KLIST(M) = LLIST(M)
711 NL = LX + 1
NM = LX + IA
J = 1
DO 709 I = NL, NM
KLIST(I) = NSPD(ISAV(J))
709 J = J + 1
IF (ACOUNT.GT.BCOUNT) 714, 704
714 BCOUNT = ACOUNT
J = 1
DO 715 I = NL, NM
LLIST(I) = KLIST(I)
NSP(I) = ISAV(J)
715 J = J + 1
704 RETURN
END
SUBROUTINE ZERONN
COMMON /1/ NY, NX, NO, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1 NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2 NSP(200), LX, JX
3 /2/ ISAV(200) /3/ KLIST(200) /4/ LLIST(200)
4 /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCOUNT
IL = LX
DO 550 I = 1, IL
IDUM = ITEMP(I)
DO 550 J = 1, NO
JDUM = J

```



```

551 IF (IDUM .EQ. JDUM) 550, 551
   NN = NUM(IDUM, JDUM)
   A(NN) = 0.0
550 CONTINUE
   RETURN
   END
   SUBROUTINE RELATED
   COMMON /1/ NY, NX, NJ, Z, LEVEL, A(6500), KOUNT(200), NSPD(200),
1     NAFF(200), MAFF(200), ITEMP(200), MARK(200), LGROUP(200),
2     NSP(200), LX, JX
3     /2/ ISAV(200) /3/ JLIST(200) /4/ LLIST(200)
4     /5/ MG, INDEXA, X(100), Y(100), JJ, IA, NCDUNT
   JK = 0
   IL = LX
   DO 224 I = 1, NO
   IF (NAFF(I)) 224, 227
227 DO 222 K = 1, IL
   IF (ITEMP(K) .EQ. 1) 224, 222
222 CONTINUE
   JCT = JX
   DO 228 J = 1, JCT
   IF (MARK(J) .EQ. 1) 226, 228
228 CONTINUE
   GO TO 224
226 JK = JK + 1
   JLIST(JK) = NSPD(MARK(J))
224 CONTINUE
   PRINT 220, JK
220 FORMAT(1H0, 5X, 13H THE FOLLOWING, I3, 1X, 7HSPECIES,
* /, 10X, 42H HAVE AFFINITIES ONLY WITH PRECEDING GROUPS)
   IF (JK) 223, 225
223 PRINT 221, (JLIST(I), I = 1, JK)
221 FORMAT(1H0, 5X, 10I4)
225 RETURN
   END

```

PROGRAM LOCATE

Purposes: Given a set of groups of species (calculated by any procedure) and a set of species lists from different stations, the program calculates the fraction of each group that is found at each station. If all members of the group occur at a station, the value will be 1.0; if none occur, no print-out for that group will be given.

Input: data file FDATA, read as TAPE5

Line 130: NUMGP = number of groups

MAXSP = number of species in largest group

NSAMP = number of samples

ICHEK set greater than 0 causes printing of each group and a listing of the possible occurrence values for each group.

Line 140: I = 1, NUMGP. List the number of species in each group, in descending order of size.

Lines 150-159: I = 1, (sum of values in previous line). List the code numbers for each species in each group, as used in FDATA. The groups must be in the same order as in line 140; the species may be in any order within a group.

Limit: 20 groups, 50 species in the largest group, 112 species, with the largest code number = 112.

Output: For each station, identified by a day and cruise number (the first two numbers in each record of FDATA), groups are listed

in which one or more species occur. The fraction of the group occurring at that point is also given.

```

0100 PROGRAM LOCATE(FDATA,OUTPUT,TAPE5=FDATA)
0110 DIMENSION NSP(20),ICODE(20,50),LEVEL(20),GOT(20,51),LIST(112),
0120+KOUNT(20),VALUE(20),KODES(112)
0130 DATA NUMGP,MAXSP,NSAMP,ICHECK/10,17,10,1/
0140 DATA(NSP(I),I=1,10)/17,16,12,7,5,4,3,3,3,2/
0150 DATA(KODES(I),I=1,72)/30,33,35,37,38,39,41,42,46,49,51,54,55,60,
0152+61,62,110,75,78,79,81,82,83,84,89,90,91,93,94,96,98,102,103,13,
0154+17,21,25,26,22,16,18,19,20,24,28,104,105,108,112,107,109,111,
0156+64,69,70,65,68,11,14,15,23,4,7,9,10,67,63,8,12,106,27,66/
0160 REWIND 5
0162 K=1
0164 DO 50 I=1,NUMGP
0166 KWIT=NSP(I)
0168 DO 40 J=1,KWIT
0170 ICODE(I,J)=KODES(K)
0172 K=K+1
0174 40 CONTINUE
0190 50 CONTINUE
0200 MPLUS=MAXSP+1
0210 DO 70 K=1,MPLUS
0220 DO 70 I=1,NUMGP
0230 LEVEL(I)=NSP(I)-(K-1)
0240 IF(LEVEL(I))55,60,60
0250 55 LEVEL(I)=0
0260 60 GOT(I,K)=FLOAT(LEVEL(I))/FLOAT(NSP(I))
0270 70 CONTINUE
0290 80 FORMAT(///)
0300 IF(ICHECK)85,112
0310 85 DO 110 I=1,NUMGP
0320 KWIT=NSP(I)
0330 KSTOP=KWIT+1
0340 PRINT 90,I,(ICODE(I,J),J=1,KWIT)
0350 90 FORMAT(///,1X,15HRECURRENT GROUP,1X,I3,1X,
0360+35HCONSISTING OF THE FOLLOWING SPECIES,/,10(2X,I3),/)

```

```

0370 PRINT 100, (GOT(I,K), K=1, KSTOP)
0380 100 FORMAT(/, 5X, 40H CAN TAKE ON OCCURRENCE VALUES AS FOLLOWS,
0390+ //, (5X, 5F10.4))
0400 110 CONTINUE
0410 PRINT 80
0420 112 DO 450 I=1, NSAMP
0430 IL=1
0440 115 IR=IL+8
0450 READ(5, 120) KRUZE, ISTAT, (LIST(J), J=IL, IR)
0460 120 FORMAT(11I3)
0470 IF(KRUZE) 470, 470, 122
0480 122 IF(LIST(IR)) 130, 130, 125
0490 125 IL=IR+1
0500 GO TO 115
0510 130 DO 135 K=1, NUMGP
0520 135 KOUNT(K)=0
0530 MPLUS=MAXSP+1
0540 DO 400 J=1, MPLUS
0550 DO 150 K=1, NUMGP
0560 LEVEL(K)=NSP(K)-(J-1)
0570 IF(LEVEL(K)) 140, 150, 150
0580 140 LEVEL(K)=0
0590 150 CONTINUE
0600 DO 160 L=1, NUMGP
0610 IF(KOUNT(L)-999) 155, 160, 155
0620 155 KOUNT(L)=0
0630 160 CONTINUE
0640 DO 300 MSPEC=1, IR
0650 DO 290 MGRUP=1, NUMGP
0660 LIMIT=NSP(MGRUP)
0670 DO 170 N=1, LIMIT
0680 IF(LIST(MSPEC)-ICODE(MGRUP, N)) 170, 180, 170

```

```
0690 170 CONTINUE
0700 GO TO 290
0710 180 IF(KOUNT(MGRUP)-999)190,300,190
0720 190 KOUNT(MGRUP)=KOUNT(MGRUP)+1
0730 IF(KOUNT(MGRUP)-LEVEL(MGRUP))300,200,200
0740 200 VALUE(MGRUP)=GOT(MGRUP,J)
0750 KOUNT(MGRUP)=999
0760 GO TO 300
0770 290 CONTINUE
0780 300 CONTINUE
0790 400 CONTINUE
0810 PRINT 410,KRUZE,ISTAT
0820 410 FORMAT(1X,6HCRUISE,1X,I3,10X,7HSTATION,1X,I3,/,
0830+5X,51HTHE FOLLOWING GROUPS HAD NON-ZERO OCCURRENCE VALUES,P
0840+//,10X,5HGROUP,10X,16HOCCURRENCE VALUE,/)
0850 DO 420 KK=1,NUMGP
0860 IF(VALUE(KK))420,420,414
0870 414 PRINT 415,KK,VALUE(KK)
0880 415 FORMAT(11X,I3,16X,F5.3)
0890 420 CONTINUE
0900 DO 430 KK=1,NUMGP
0910 430 VALUE(KK)=0.
0920 450 CONTINUE
0930 470 STOP
0940 END
```