CLUSTER ANALYSIS OF PHYTOPLANKTON DATA COLLECTED FROM THE NATIONAL STREAM QUALITY ACCOUNTING NETWORK IN THE TENNESSEE RIVER BASIN, 1974-81

by D. W. Stephens and J. B. Wangsgard

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 86-4029

Salt Lake City, Utah
1986

DEPARTMENT OF THE INTERIOR

DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

District Chief
U.S. Geological Survey
Water Resources Division
1016 Administration Building
1745 West 1700 South
Salt Lake City, Utah 84104

Copies of this report can
be purchased from:

U.S. Geological Survey
Books and Open-File Reports
Federal Center, Bldg. 41
Box 25425
Denver, Colorado 80225
[Telephone: (303) 236-7476]

CONTENTS

ILLUSTRATIONS
[Plates are in pocket]

ILLUSTRATIONS--Continued

TABLES

TABLES--Continued

CLUSTER ANALYSIS OF PHYTOPLANKTON DATA COLLECTED FROM THE NATIONAL STREAM QUALITY ACCOUNTING NETWORK IN THE TENNESSEE RIVER BASIN, 1974-81

by D. W. Stephens and J. B. Wangsgard

ABSTRACT

A computer program, Numerical Taxonomy System of Multivariate Statistical Programs (NTSYS), was used with interfacing software to perform cluster analyses of phytoplankton data stored in the biological files of the U.S. Geological Survey. The NTSYS software performs various types of statistical analyses and is capable of handling a large matrix of data. Cluster analyses were done on phytoplankton data collected from 1974 to 1981 at four National Stream Quality Accounting Network stations in the Tennessee River basin. Analysis of the changes in clusters of phytoplankton genera indicated possible changes in the water quality of the French Broad River near Knoxville, Tennessee. At this station, the most common diatom groups indicated a shift in dominant forms with some of the less common diatoms being replaced by green and blue-green algae. There was a reduction in genera variability between 1974-77 and 1979-81 sampling periods. Statistical analysis of chloride and dissolved solids confirmed that concentrations of these substances were smaller in 1974-77 than in 1979-81. At Pickwick Landing Dam, the furthest downstream station used in the study, there was an increase in the number of genera of "rare" organisms with time. The appearance of two groups of green and blue-green algae indicated that an increase in temperature or nutrient concentrations occurred from 1974 to 1981, but this could not be confirmed using available water-quality data.

Associations of genera forming the phytoplankton communities at three stations on the Tennessee River were found to be seasonal. Nodal analysis of combined data from all four stations used in the study did not identify any seasonal or temporal patterns during 1974-81. Cluster analysis using the NTSYS programs was effective in reducing the large phytoplankton data set to a manageable size and provided considerable insight into the structure of phytoplankton communities in the Tennessee River basin. Problems encountered using cluster analysis were the subjectivity introduced in the definition of meaningful clusters, and the lack of taxonomic identification to the species level.

INTRODUCTION

The U.S. Geological Survey established the National Stream Quality Accounting Network (NASQAN) in January 1973 with the following objectives: (1) To account for the quantity and quality of water moving within the United States; (2) to depict areal variability in water quantity and quality; (3) to detect changes in stream water quality; (4) to establish a data base by which future water-quality changes could be evaluated. In addition to data on water chemistry, biological constituents such as chlorophyll, benthic invertebrates, periphyton, and phytoplankton were sampled periodically until about 1981 to provide a more time-integrated record of water quality. A massive volume of data have been collected, and at least two interpretive reports (Steele and others, 1974; Hirsch and others, 1982) dealing with the water chemistry have

been completed under this program. Little analytical work has been done on the biological data as those data are not as easily processed using common numerical techniques. The phytoplankton data base alone consists of 352 genera identified in more than 17,000 samples from 518 stations. The collection of phytoplankton and other biological data was terminated in the summer of 1981.

The identification and quantification of organisms comprising an aquatic community can provide much information on the quality of water. Unfortunately, large data sets are very difficult to interpret because the excessive number of biological identities tends to obscure the community relationships. The initial problem in the analysis of a large volume of biological data is one of reducing the data to a manageable form without losing meaningful information.

This report describes a method of data reduction and analysis useful for biological data. The report is the result of a project having the following objectives:

1. To obtain a comprehensive computer program with multivariate-analytical methods capable of handling large data sets, such as the NASQAN phytoplankton data;

2. Interface this program with the biological data files of the U.S. Geological Survey; and

Utilize the program to analyze a segment of the NASQAN phytoplankton data identify patterns of occurrence of phytoplankton genera and to determine these patterns were changing with time.

METHOD OF ANALYSIS

Multivariate statistical-analyses have been used to evaluate biotic relationships in a variety of communities (Patil and others, 1971). Cluster analysis is one of several multivariate methods which have been used to delimit aquatic community relationships with protozoa (Cairns and Kaesler, 1969), plankton (Brown, 1969), and macroinvertebrates (Crossman and others, 1974). Cluster analysis refers to an assortment of classification schemes used to analyze multivariate arrays of data by numerical methods. The objective of a cluster analysis may be to identify community structure by grouping data elements into clusters that possess a natural affinity among the members. Because members of a cluster share similar attributes, groups can be formed by the presence, absence, or relative abundance of species in the samples. At some point, the clusters then indicate separate communities. Presence and absence of organisms in each community is controlled primarily by the ability of the organism to tolerate conditions existing in the environment. Changes in community structure over a period of time may reflect natural succession or the influence of man. Cluster analysis can yield patterns of community structure in large assemblages of data that otherwise would not be obvious. Excellent reviews of clustering methods have been done by Boesch (1977), and Hellawell (1978), and several computer programs are available which are capable of analyzing relatively small data sets (Bonham-Carter, 1967; Pinkham and others, 1975; Gauch, 1979).

The first step in a cluster analysis is the calculation of a similarity coefficient using one of numerous methods based on the structure of the data and the intended use of the analysis. The Jaccard coefficient is perhaps the most satisfactory of the commonly used coefficients in ecological studies (Clifford and Stephenson, 1975, p. 55) because it does not consider absence of an organism at two collection sites to be indicative of similarity between the sites. Applying cluster analysis using the Jaccard coefficient to a biological-data set such as that collected under the NASQAN program, requires that the data first be arranged in a binary matrix of organism presence-absence for each sample. An example of a binary data matrix is given in table 1.

The meaning of the Jaccard coefficient of similarity between two individuals can be interpreted using the following two-way classification table:

		Individual i	
		+	-
Individual j	+	a	b
	-	c	d

In a series of samples, a = the number of mutual occurrences between i and j (+,+), b = the number of times j occurs but not i (+,-), c = the number of times i occurs but not j (-,+), and d = the number of times neither occur (-,-). The Jaccard coefficient is calculated as:

$$S_{ij} = \frac{a}{a+b+c}$$

Mutual absence (d) is disregarded in calculation of the Jaccard coefficient to prevent a negative match from indicating similarity between two individuals. The values of the coefficient range from zero to one, where a value of one indicates complete similarity and zero indicates complete dissimilarity in the occurrence patterns of the two organisms.

The Jaccard coefficient is computed pairwise for each possible pair of entries in the data matrix. The symmetrical matrix of Jaccard coefficients calculated from data in table 1 is shown in table 2.

The next step in the analysis is the interpretation of the similarity matrix using clustering algorithms. There are several methods of performing a cluster analysis, but the most widely used method consists of a hierarchical, agglomerative, and combinatorial approach (Boesch, 1977, p. 42). Hierarchical methods determine the optimal route from the individual entities to the larger group and results are presented in the form of a dendrogram. Agglomerative clustering progressively joins entities ending with the complete population. Combinatorial methods involve the successive calculation of resemblance

Table 1.--Binary data matrix of the occurrence of phytoplankton genera in monthly collections at a single site

[1 = present 0 = absent]

Organism	Jan. 1	Feb. 1	March 1	April 1	May 1	June 1
Oscillatoria	1	0	1	0	1	1
Melosira	0	0	1	1	0	0
Nitzschia	1	1	0	0	1	1
Chodatella	1	1	1	1	1	0
Gomphonema	0	0	1	0	1	1

Table 2.--Symmetrical matrix of Jaccard coefficients of similarity where values at the intersection of row and column give the similarity values for the organisms as implied by the Jaccard coefficient

	Oscillatoria	Melosira	Nitzschia	Chodatella	Gomphonema
Oscillatoria	1.0	0.20	0.60	0.50	0.75
Melosira	.20	1.0	.0	.40	.25
Nitzschia	.60	.0	1.0	.50	.40
Chodatella	.50	.40	.50	1.0	.33
Gomphonema	.75	.25	.40	.33	1.0

measures from the data matrix, and thus, once the resemblance is computed, it is no longer necessary to retain the raw-data matrix. A hierarchical, agglomerative-cluster analysis groups entities based on their similarity coefficient and produces useful results when the initial data are in binary form. Individuals displaying the largest similarity values are grouped first into their respective clusters. Additional members may join clusters when they most nearly resemble all other combined members of a cluster as the level of similarity for inclusion in a group is lowered. Clusters typically become larger as the degree of similarity within the group becomes less. When an entity is joined in a cluster, it is nonseparable and its attributes become part of the group's attributes. Eventually, groups are considered as entities and they too may join together. The result of a cluster analysis using the data matrix from table 2 is illustrated by a dendrogram displaying the relationships as given in figure 1.

The degrees of similarity depicted by the dendrogram together with corresponding values from the similarity coefficient matrix form a set of ordered pairs of numbers that can be plotted in a two-way scatter diagram. The relationship of points in the scatter diagram may then be evaluated using the cophenetic correlation coefficient (Sokal and Rohlf, 1962; Kaesler, 1970) which is calculated by the method of least squares and ranges from one to zero. A coefficient of one indicates a perfect linear relationship between the dendrogram and matrix values, and that the dendrogram accurately displays the results of a cluster analysis. A coefficient equal to zero means the dendrogram randomly represents the clustering relationships. Intermediate values are subject to interpretation, but values less than 0.70 probably indicate a large number of misclassifications in the clustering process and different clustering techniques should be tried. A scatter diagram produced from the Jaccard coefficient matrix and the resultant dendrogram in figure 1 is shown in figure 2. The cophenetic correlation coefficient for this relationship is 0.84.

The final step in the procedure is the verification and interpretation of the clusters. This is the step most frequently overlooked. The similarity indices and clustering methods simplify complex data, they do not provide ecological interpretations.

NTSYS: A Versatile Analytical Tool

A computer program with the capability of cluster analysis of large volumes of data has been compiled on the computer system of the U.S. Geological Survey in Reston, Virginia. The program, Numerical Taxonomy System of Multivariate Statistical Programs (NTSYS) is a system of algorithms developed for use in numerical taxonomy by F. James Rohlf of the State Univerity of New York (Rohlf, 1985). It has been used considerably in water-resources investigations (Kaesler and Cairns, 1972; Crossman and others, 1974). The program is versatile, containing routines for the computation of a variety of association and similarity coefficients and several methods of cluster analysis. A routine is also available to generate a dendrogram and a scatter diagram presenting the cophenetic correlation coefficient from the results of a cluster analysis. The program will process large matrices of data and is limited only by the amount of computer core storage available. In practice, a matrix as large as 400 by 400 may be easily analyzed. The program documentation is included as a user-accessible file within the NTSYS program.

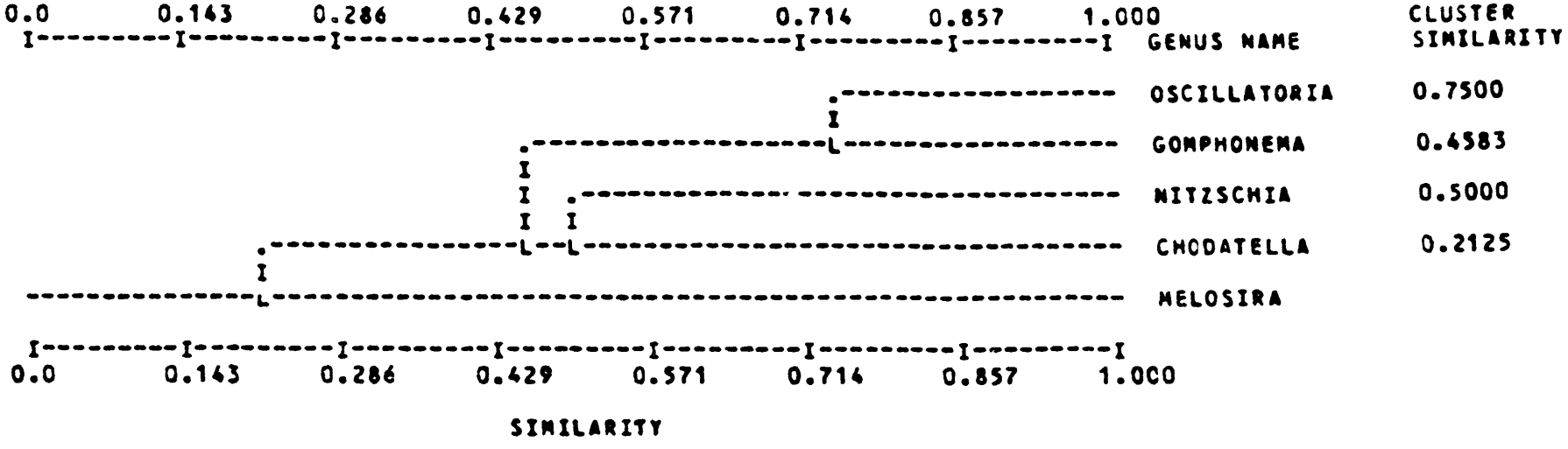


Figure 1.—Dendrogram showing cluster analysis relationships of organisms presented in table 2.

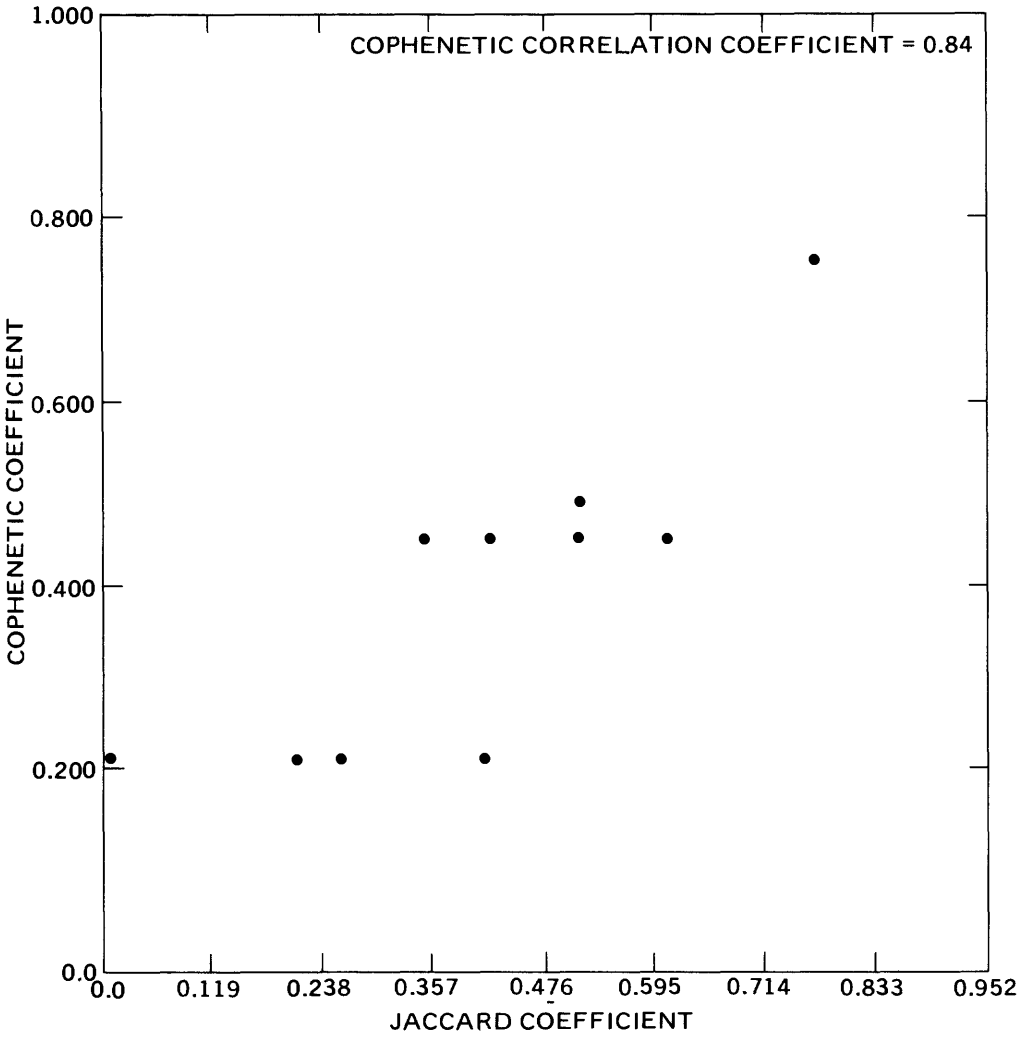


Figure 2.—Scatter diagram for the matrix of similarity values in table 2 and the implied similarity (cophenetic) values in figure 1.

Cluster analysis essentially organizes and simplifies data into useful generalizations for interpretation. This data reduction always results in some loss of information and distortion by summarizing the results (Rohlf, 1970, p. 61). A subroutine in NTSYS provides a method of assessing distortion from the agglomerative approach. Elements of the original similarity matrix are compared to similarity values implied by a clustering scheme; the comparisons are plotted in a bivariate scatter diagram and the cophenetic correlation coefficient is determined by a least-squares regression line.

NTSYS allows for clustering in both "normal" (Q mode) and "inverse" (R mode) phases (fig. 3). In normal clustering, the entities being classified are collections with the taxonomic content as the attributes. In the inverse mode, the individual taxa are the entities and their presence or absence are the attributes. Ecological investigations usually employ normal-mode clustering to determine the relationships among collection sites or dates. Inverse clustering may be used to evaluate the relationships among organism assemblages at different sites or on different dates. If both normal and inverse methods are used with the same data, a two-way coincidence plot arranged by collection and organism groups may be used in a nodal analysis. This identifies misclassifications and enhances the ecological interpretations. Differences in the collection groups (from normal analysis) can be described by the frequency of members in the organism groups (from inverse analysis). Also, the differences in the distribution of the organism groups can be determined by the frequency of the taxa in the collection groups (Boesch, 1977, p. 63).

Interfacing NTSYS with Biological-Data Files

A large quantity of data on periphyton, phytoplankton, and macroinvertebrates are available on the computer files of the U.S. Geological Survey. An interface system was created as a part of this project to access the biological data, code it into a binary presence-absence form, and create a data set compatible with NTSYS. The system (fig. 4) begins with the specification of NASQAN stations within the hydrologic units desired from the Master Water Data Index of the National Water Data Exchange (NAWDEX) file. The NASQAN station identification numbers are then used to access the biological data through the BIOPUNCH program of the U.S. Geological Survey Atlanta Central Laboratory. This provides a printed list and card images of collections of organisms at specified sites. The investigator then selects those sites and dates which are desired as input for cluster analysis. A Statistical Analysis System (SAS) program (Barr and others, 1979) is used to code the desired card images into a binary format compatible with NTSYS. The SAS program prepares properly formatted data sets for entry into NTSYS, which are stored on disk files, and a printout of phytoplankton identification numbers with a matrix of collection dates and organisms. This matrix is very useful in verifying the accuracy of the input. Taxonomic names of the phytoplankton must be manually decoded from the numerical form using an organism code list from the Atlanta Central Laboratory.

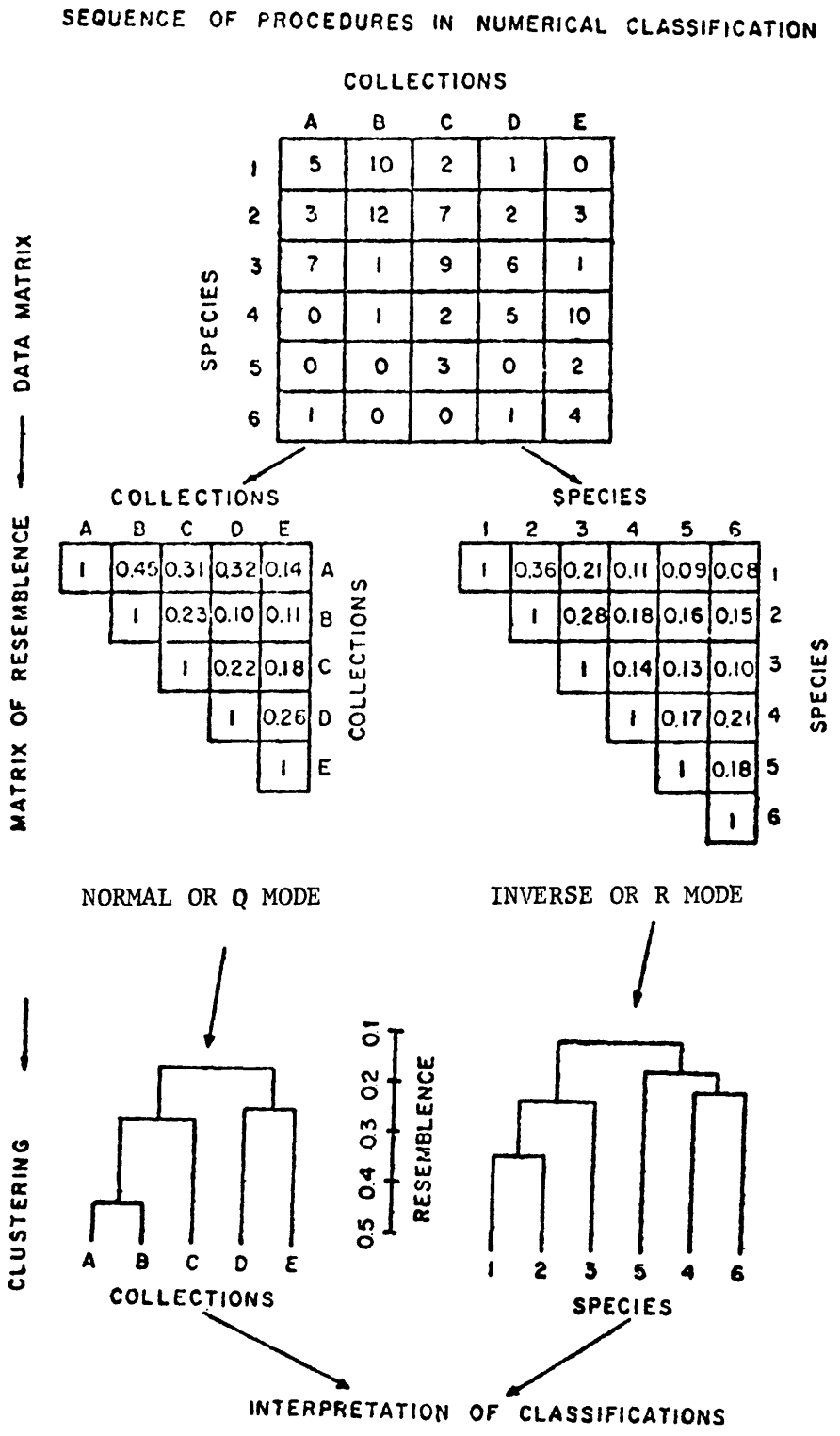
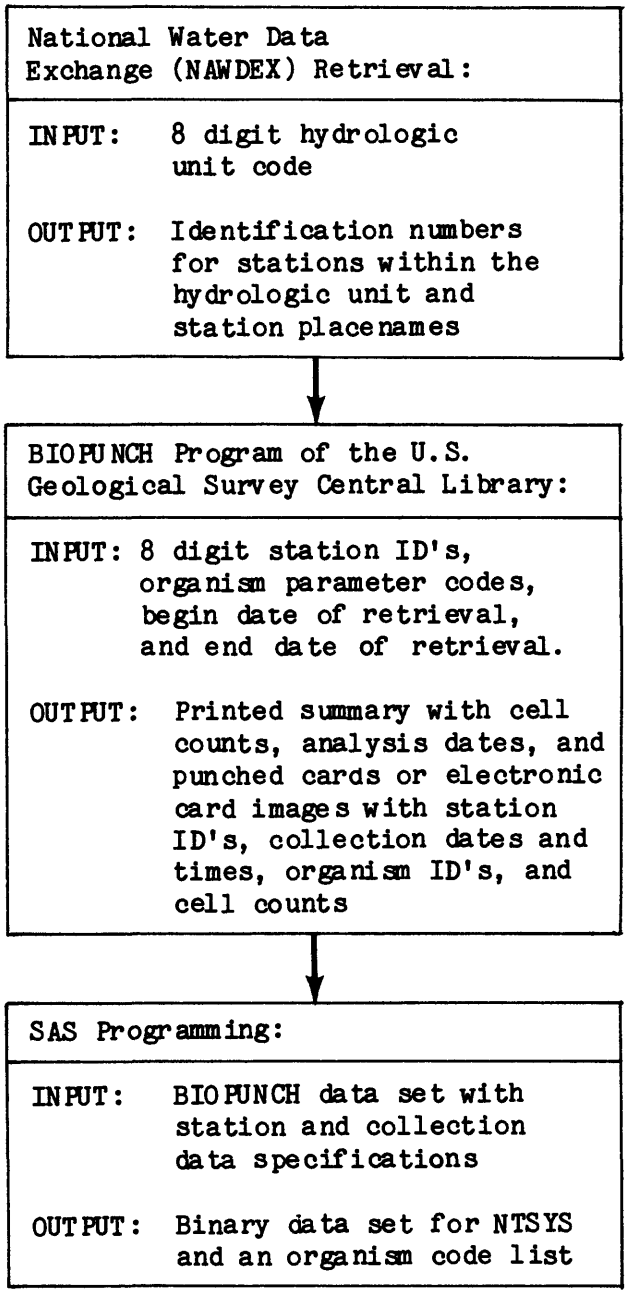


Figure 3.—The sequence of procedures in normal and inverse clustering methods. Adapted from Boesch (1977.)



This is done if there is no previous knowledge of the stations in the area of interest

The cards or card images are the raw data for an analysis, and the printout includes a summary of the biological analysis (collection dates and sites). The dates and stations are used to specify data processing in a Statistical Analysis System (SAS) program to create NTSYS data sets.

The SAS program creates a binary data set that is used in NTSYS. Data are assimilated by NTSYS with a FORTRAN format, and a useful formatted output by the SAS program is: nF2.0, where n is the number of dates for an analysis.

Figure 4.—Interface procedures used to create a biological data set for input to the Numerical Taxonomy System of Multivariate Statistical Programs

Cluster Analysis of Phytoplankton Data Using NTSYS

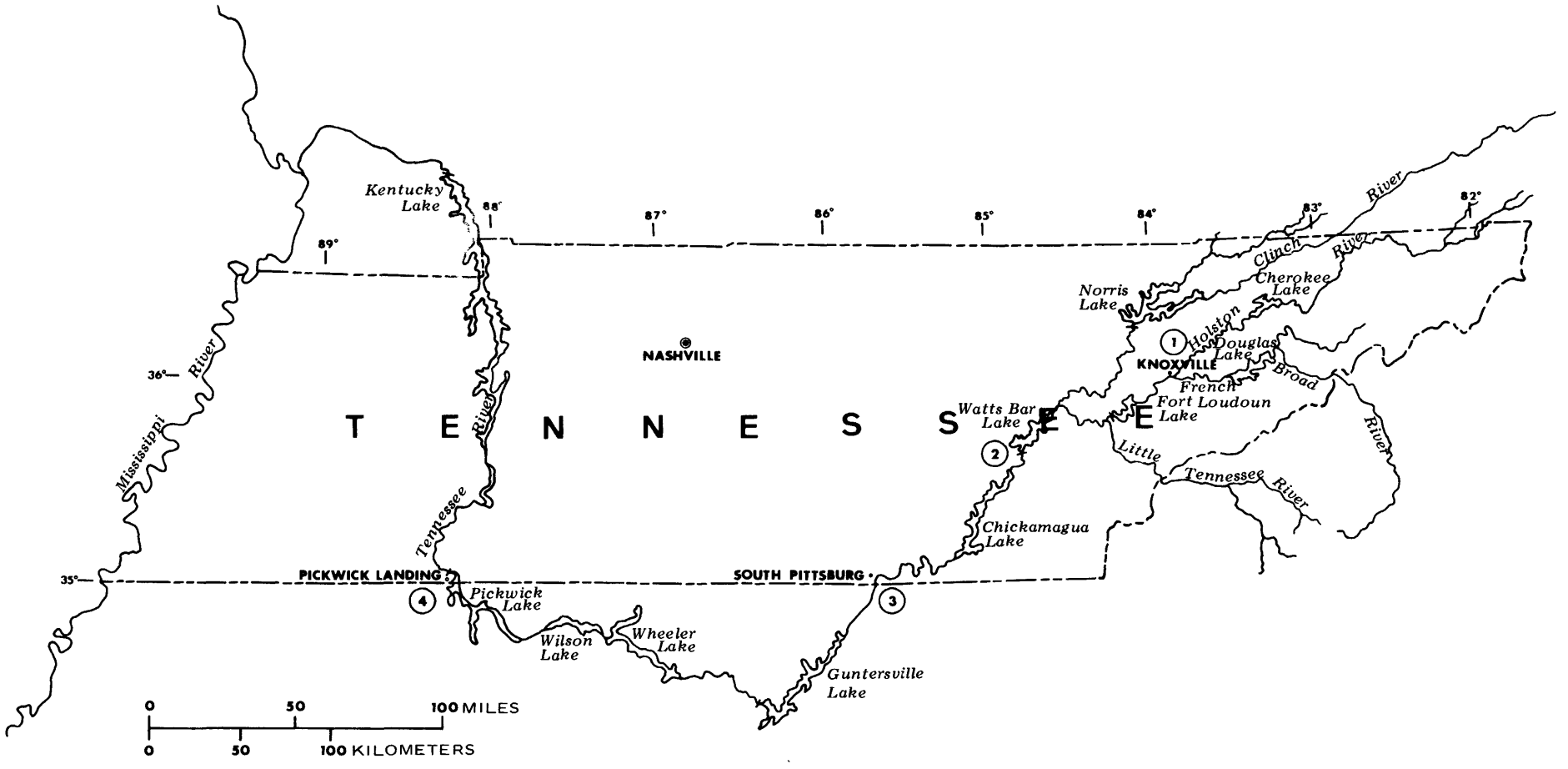
The Tennessee River basin was chosen as a test area because it had an abundance of on-line reservoirs which provide a variety of phytoplankton and a manageable number of sampling stations. Data from NASQAN stations in the basin were selected for analysis using the following criteria: A 7-year period of record; representative seasonal sampling; and a lack of gaps in the sampling record.

A NAWDEX retrieval was requested for all NASQAN stations in Water Resources Council hydrologic unit-region 6. A total of eight stations were identified, consisting of five with data from 1974-81, and three with a data record of five years or less. Four stations were selected, each providing 7 years of phytoplankton data on the Tennessee River and one of its tributaries, the French Broad River. One NASQAN station, the Clinch River at Melton Hill Dam, was not included since the data were neither as extensive nor seasonally representative as data from the other stations. Data from the following list of stations were used in analyses presented in this report. The locations of the stations are shown in figure 5.

French Broad River near Knoxville, Tennessee (03470500)
Tennessee River at Watts Bar Dam (tailwater), Tennessee (03543005)
Tennessee River at South Pittsburg, Tennessee (03571850)
Tennessee River at Pickwick Landing Dam (lower lock), Tennessee (03593005)

The station identification numbers were then used to request, from the Atlanta Central Laboratory, all phytoplankton data from January 1, 1974 to December 31, 1981. The data consisted of phytoplankton identification numbers, cell counts, dates, and times of collection. The data were examined to determine the sampling frequency, then those data with comparable seasonal coverage among stations were selected for analysis. The data format used in the calculation of a Jaccard coefficient requires that phytoplankton cell counts be converted to the binary form of presence or absence. Phytoplankton cell counts, therefore, were assigned a value of one if they occurred in a given sample regardless of number of cells. If phytoplankton genera were not present in the sample, they were assigned a zero, representing absence.

Binary data sets of phytoplankton occurrence were analyzed with normal and inverse modes of clustering. Similarities were calculated using the Jaccard coefficient, and clustering was done by the unweighted pair-group method with arithmetic averaging (UPGMA). A separate routine was used which summed rows and columns of the presence-absence matrices to provide additional information on the frequency of organism occurrence and diversity in the collection.



EXPLANATION

LOCATION OF NATIONAL STREAM QUALITY ACCOUNTING NETWORK (NASQAN) STATIONS

(1) French Broad River near Knoxville, Tennessee

(2) Tennessee River at Watts Bar Dam (tailwater), Tennessee

(3) Tennessee River at South Pittsburg, Tennessee

(4) Tennessee River at Pickwick Landing Dam (lower lock), Tennessee

Figure 5.—Tennessee River basin showing National Stream Quality Accounting Network (NASQAN) stations used as sources of phytoplankton data for cluster analyses.

INTERPRETATION OF CLUSTER ANALYSES OF PHYTOPLANKTON FROM THE TENNESSEE RIVER BASIN

Inverse Cluster Analysis

The upstream NASQAN station for this analysis of the Tennessee River basin is on the French Broad River (fig. 5) near Knoxville, Tennessee. At this location, the data from three nonoverlapping time periods (1974-76, 1976-77, and 1979-81) were examined to determine if changes in population structure had occurred with time. Inverse analysis was used to identify the types of phytoplankton communities which were present and to identify any changes in their composition. The data for 1974 to 1977 exhibited a variety of several types of phytoplankton with the diatoms Navicula, Nitzschia, and Melosira forming the dominant community (figs. 6, 7). Other diatoms such as Synedra, Gomphonema, and Achnanthes formed secondary groups. In 1979-81 there were fewer meaningful clusters of diatoms and the major association was Navicula-Gomphonema (fig. 8). The blue-greens Aphanizomenon and Lyngbya appeared to form a meaningful group with a similarity level comparable to that of previously mentioned common diatom groups. Navicula was common in all samples from 1974 to 1981. The reduction in diatom-dominated clusters and coincident appearance of groups of blue-green algae indicated a probable deterioration of water quality in the French Broad River near Knoxville after 1979.

Examination of the inverse cluster analysis of the data from the Watts Bar Dam station also indicated a changing composition of the phytoplankton community. The 1976-77 data from the Watts Bar Dam station were characterized by a Cyclotella-Melosira-Nitzschia type community, with other well-defined groups including a cluster of Chroomonas and Stephanodiscus (fig. 9). By comparison, samples collected from 1979-80 had well-defined clusters of Melosira-Ankistrodesmus, Cyclotella-Anacystis, and Chlamydomonas-Scenedesmus (fig. 10). They reflected a shift in the types of commonly occurring organisms. Most notably, the diatom Navicula disappeared in 1979-80, but was fairly common in earlier dates. Both groups of data had considerable richness in phytoplankton diversity with 12 incidences of single-occurrence clusters in each data set. Such forms as the blue-greens Oscillatoria and Anacystis, and the green algae Ankistrodesmus and Chlamydomonas are often used as indicator organisms and all are ranked within the 20 most pollution-tolerant genera by Palmer (1969, p. 79). The fact that these particular organisms were present in increasing numbers of samples and formed well-defined clusters was useful in identifying changes in the phytoplankton community.

Downstream from Watts Bar Dam at South Pittsburg, Tennessee, collections from 1974-76 and 1976-77 had typical communities of Cyclotella-Melosira-Nitzschia (figs. 11, 12). During 1979-81, the phytoplankton from the South Pittsburg station were characterized by the appearance of a Chlamydomonas-Melosira-Nitzschia community (fig. 13). Chlamydomonas occurred much more frequently in samples collected after 1979, and Cyclotella, although still common, did not join clusters with Melosira and Nitzschia at similarity levels which were as large as noted earlier. Gomphonema was common in 1974-76, but disappeared completely after that. Upstream, in the French Broad River, Gomphonema was quite common, particularly in 1979-81. Navicula became rarer with time at South Pittsburg, but it occurred commonly with Gomphonema in the French Broad River during 1979-81.

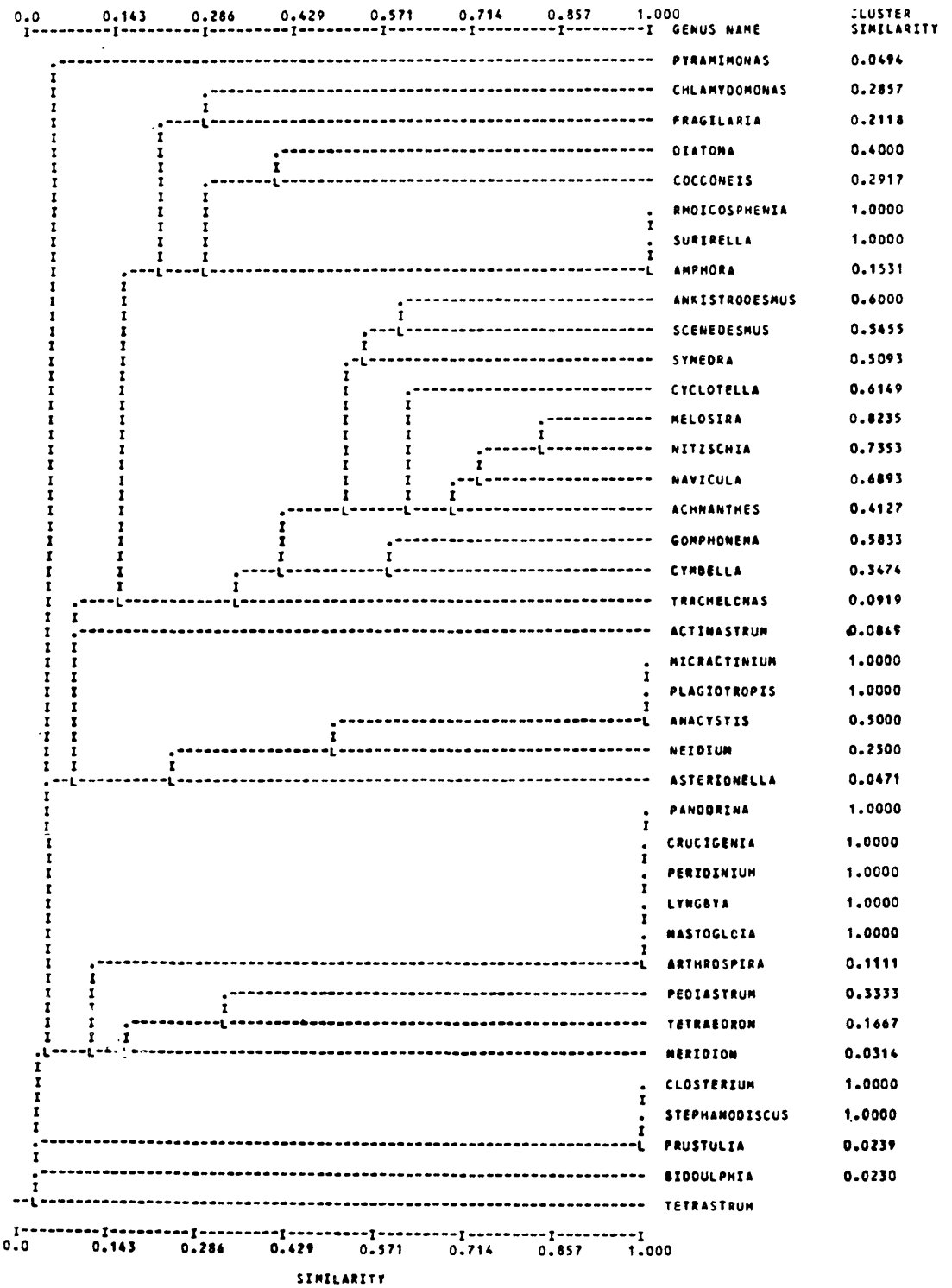


Figure 6.—Relationships of phytoplankton in the French Broad River near Knoxville, Tennessee, 1974-76. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.936. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

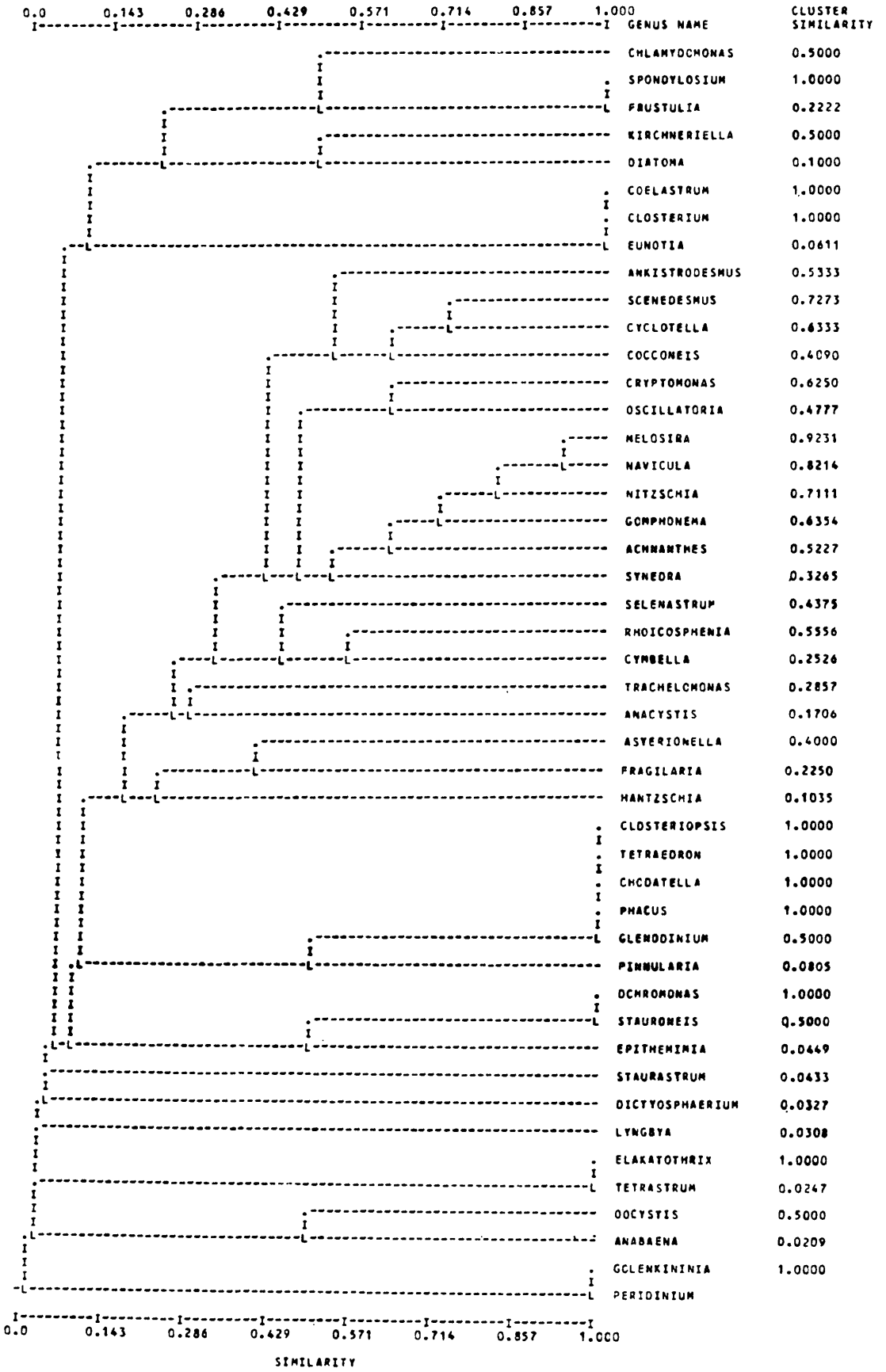


Figure 7.—Relationships of phytoplankton in the French Broad River near Knoxville, Tennessee, 1976-77. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.889. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

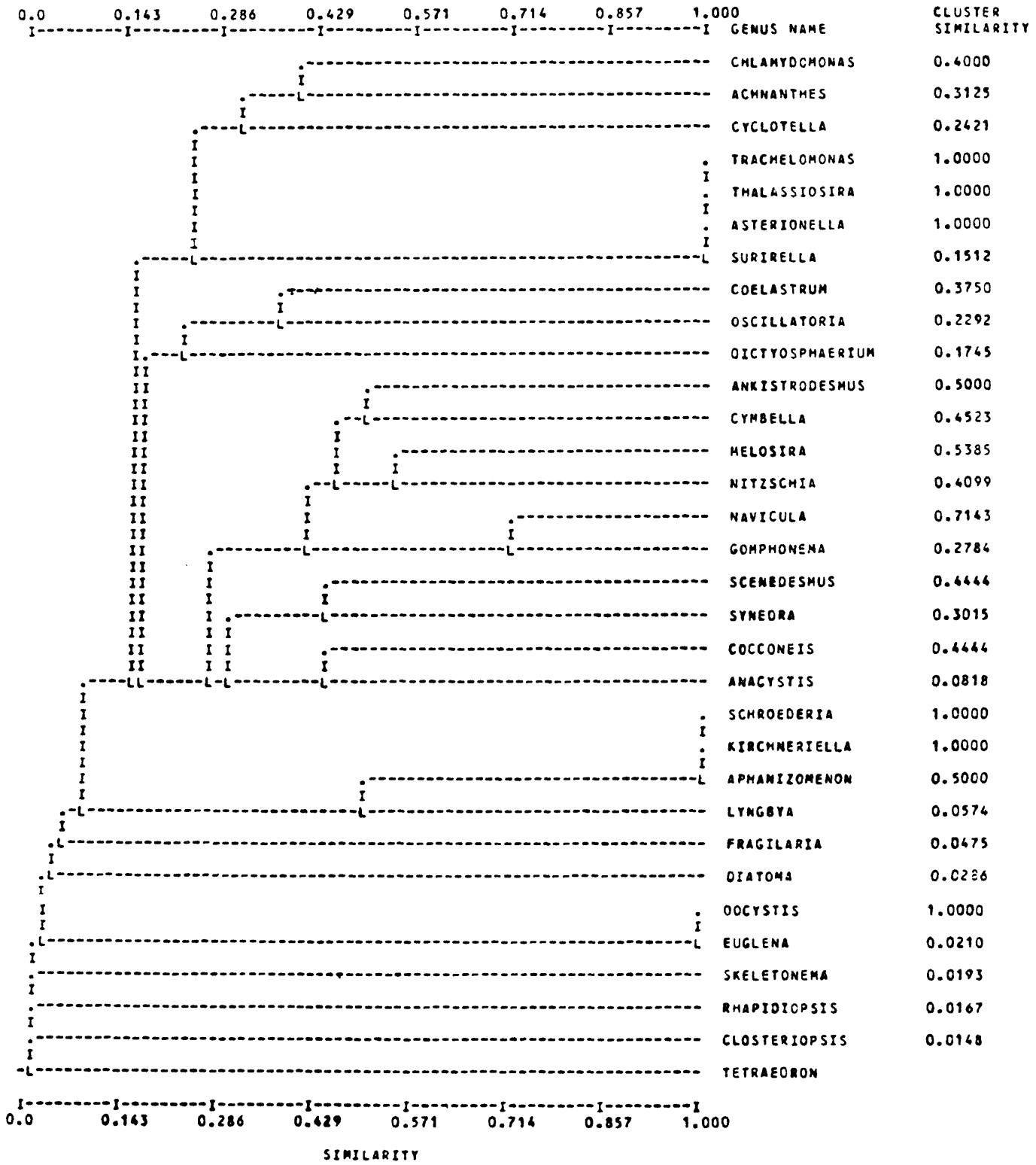


Figure 8.—Relationships of phytoplankton in the French Broad River near Knoxville, Tennessee, 1979-81. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.897. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

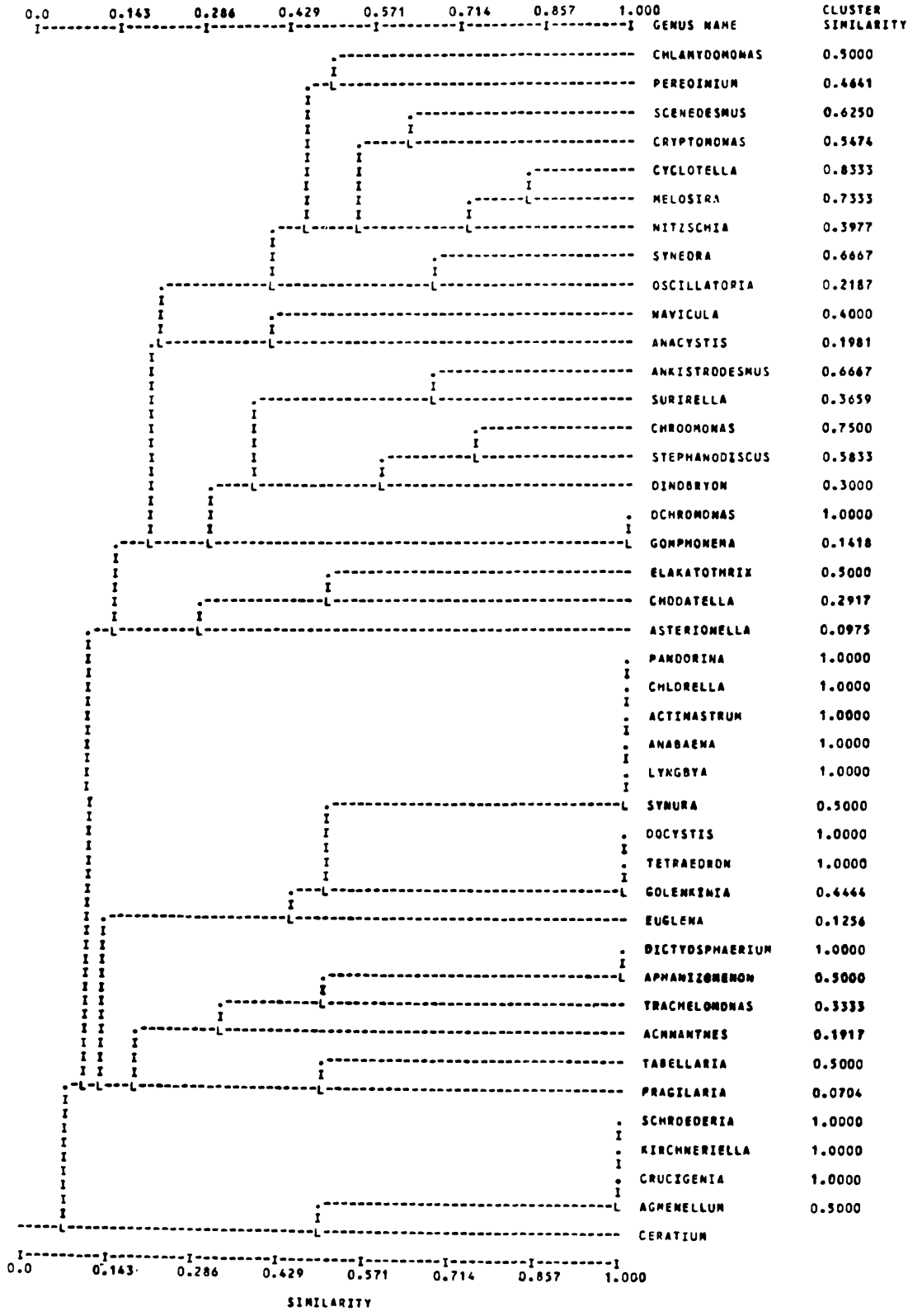


Figure 9.—Relationships of phytoplankton in the Tennessee River at Watts Bar Dam, Tennessee, 1976-77. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.836. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

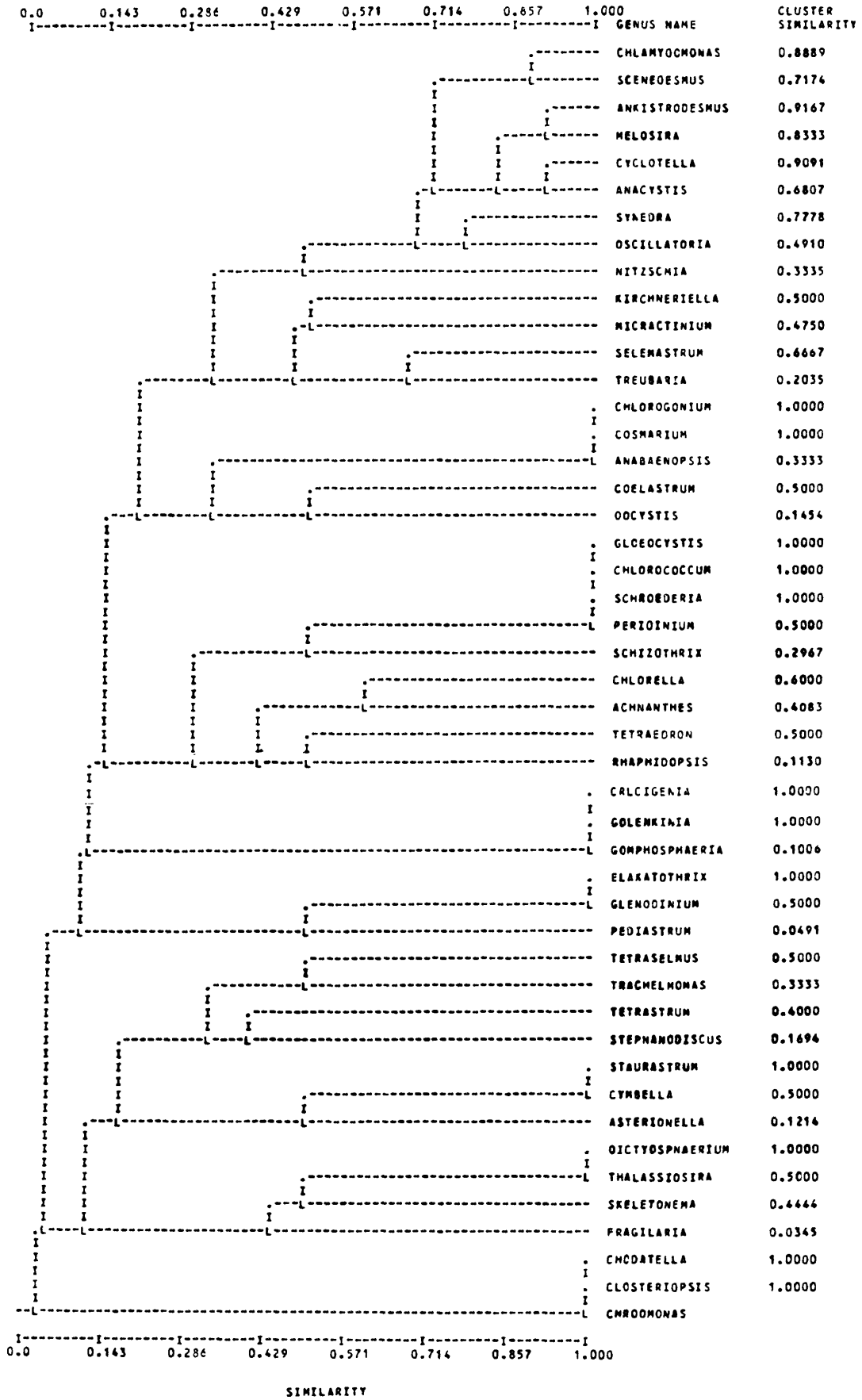


Figure 10.—Relationships of phytoplankton in the Tennessee River at Watts Bar Dam, Tennessee, 1979-80. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.854. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

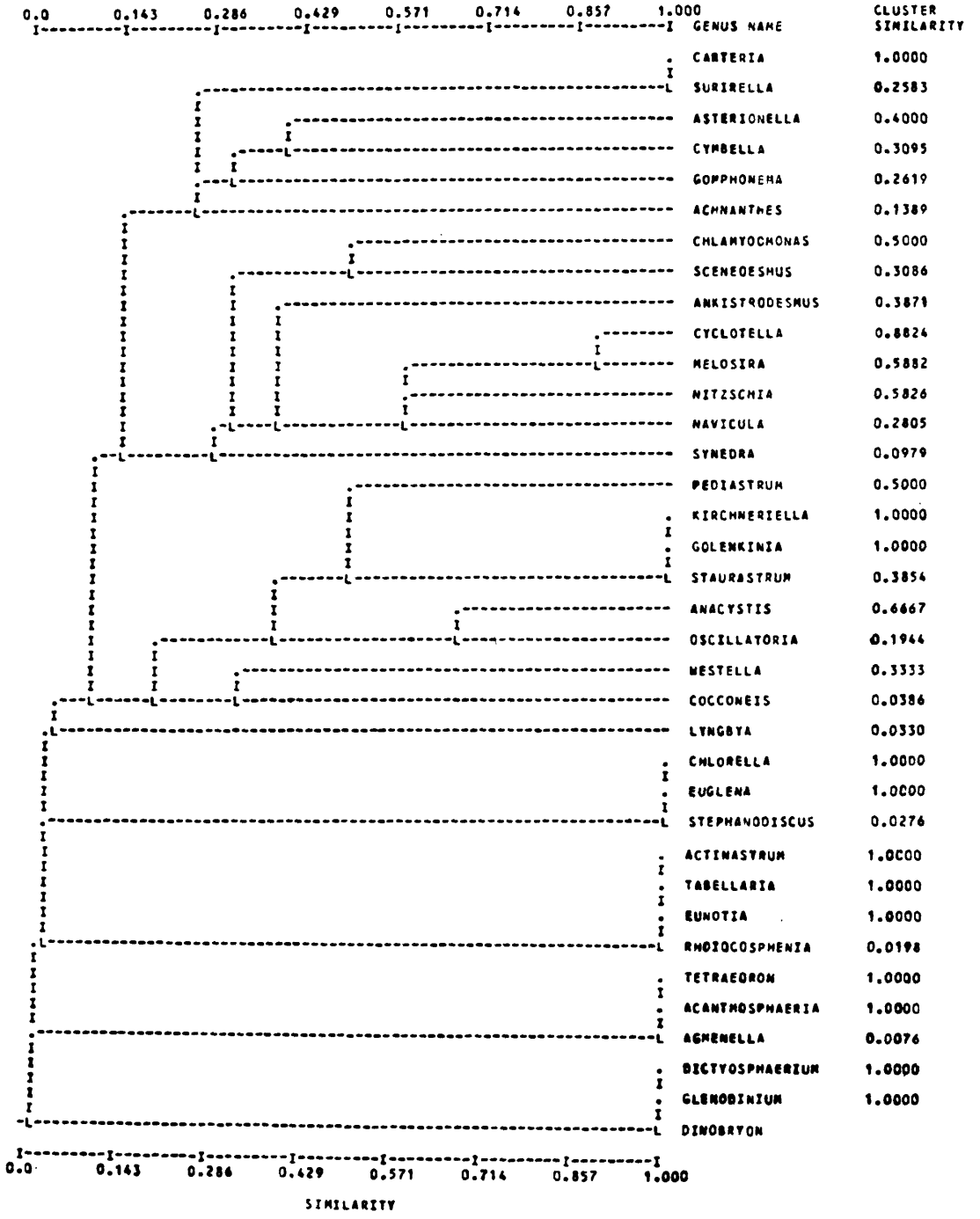


Figure 11.—Relationships of phytoplankton in the Tennessee River at South Pittsburg, Tennessee, 1974-76. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.927. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

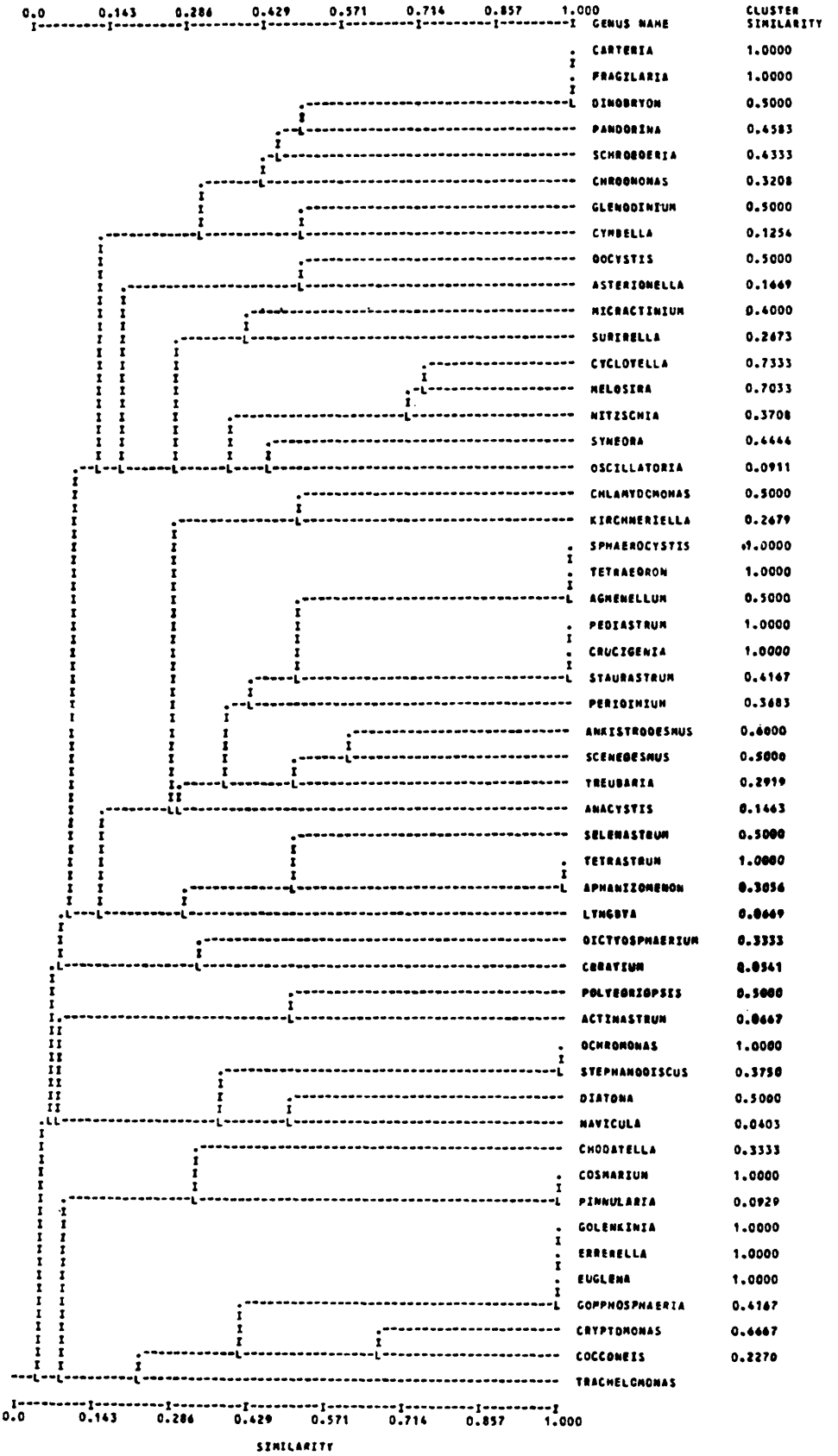


Figure 12.—Relationships of phytoplankton in the Tennessee River at South Pittsburg, Tennessee, 1976-77. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.815. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

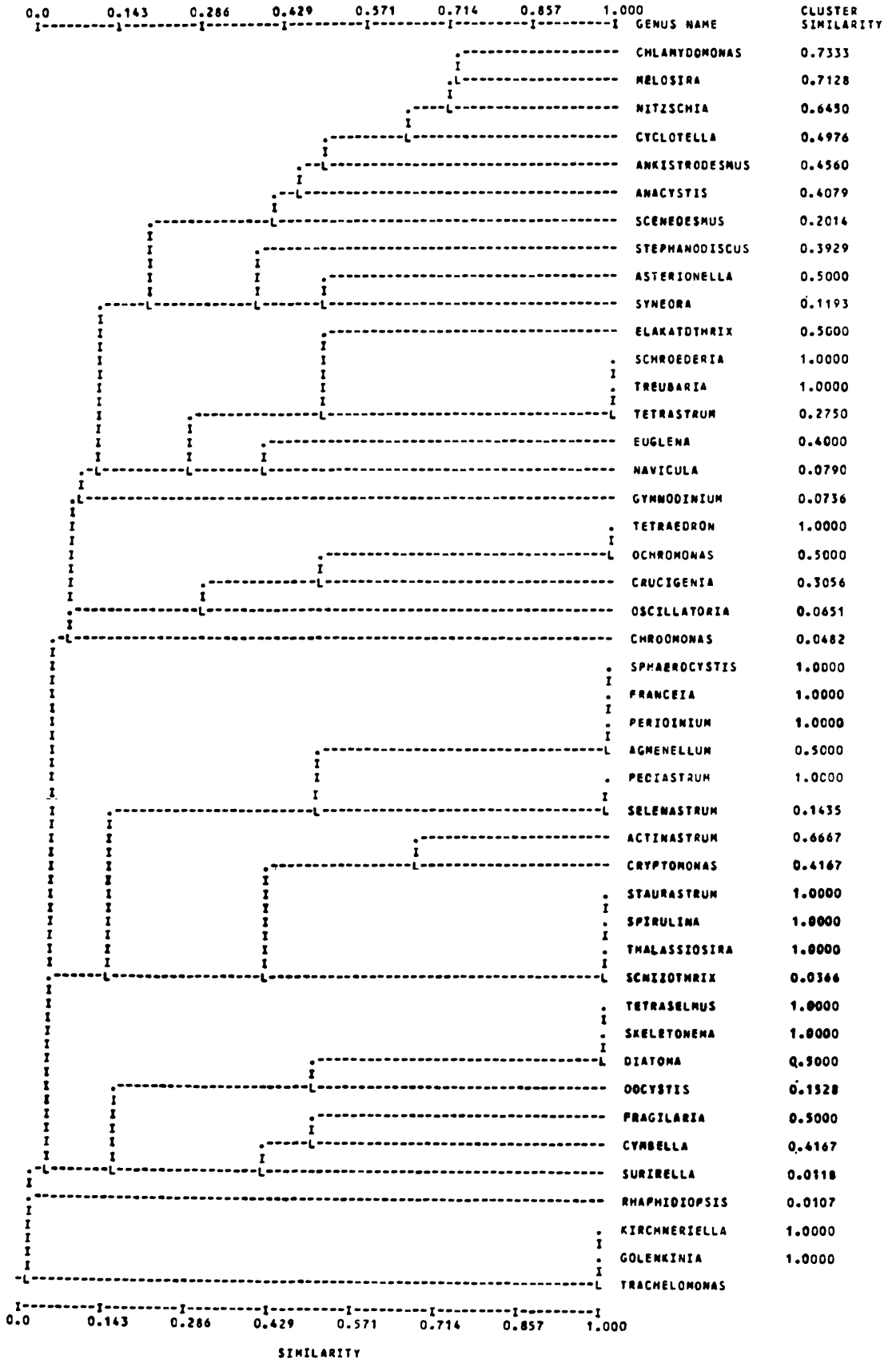


Figure 13.—Relationships of phytoplankton in the Tennessee River at South Pittsburg, Tennessee, 1979-81. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.889. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

Clusters of phytoplankton sampled at the Pickwick Landing Dam station during 1975-77 (fig. 14) tended to be defined by larger similarity coefficients than those collected during 1979-81 (fig. 15). Both periods were dominated by a Cyclotella-Melosira assemblage, but the similarity level for the 1979-81 assemblage was 0.78 compared to 0.96 for 1975-77. There was also a 30 percent increase in the number of singly-occurring groups during 1979-81. The reduction in the overall levels of similarity coupled with the increase in types of organisms suggests an increase in community diversity.

The four NASQAN stations used in this analysis are on, or immediately downstream from reservoirs, and the seven major reservoirs in the Tennessee River system nearly cover the length of the river. Changes in reservoir management may produce large changes in current velocity and result in considerable change in the structure of the phytoplankton community. These changes may or may not alter the water quality. Cluster analysis using the inverse mode indicated there were marked changes in the structure of the phytoplankton community with time at each of the four stations used in this analysis. An overview of the results indicates Nitzschia joined clusters at successively lower levels of similarity with distance downstream; it became more independent of community associations. The typical upstream communities in the French Broad River near Knoxville were dominantly Nitzschia-Navicula-Melosira. Cyclotella-Melosira communities became dominant at downstream stations. There was an increase in the number of rare organisms during 1979-81 at Pickwick Landing Dam and considerable variability in the number of organisms throughout the sampling period in the French Broad River. Navicula was common in the French Broad River near Knoxville, but became rarer with subsequent samples at all stations. Navicula was absent in 1979-81 at all stations but the French Broad River. Owing to the broad range of conditions under which this genus occurs, it is surprising that it was not present at all stations.

In general, shifts in organism groupings indicated possible changes in the water quality, flow regime, or water temperature between 1974 and 1981. The general composition of the clusters in the 1979-80 samples indicated increases in organisms usually associated with organically enriched waters and/or warmer waters.

Cluster analysis using the inverse mode was successful in generalizing community relationships. Certain changes in community structure were noted with time and distance downriver. More detailed interpretation was limited by lack of organism identifications to species, especially for common forms of organisms such as Nitzschia. Nitzschia has several species indicative of both good and poor quality water. Also, when a river is dominated by reservoirs, there is a seasonal appearance of many phytoplankton due to blooms within the impoundments. For example, Chlamydomonas only appeared in late summer and winter (August-February) at Watts Bar Dam. The seasonal nature of phytoplankton distributions can best be determined by examining the results of both normal and inverse modes of cluster analyses.

Analysis of Normal and Inverse Clustering Methods

It is informative to consider the results of normal and inverse analyses together, using a two-way coincidence plot. This is done by plotting a dendrogram for a normal analysis on the X or Y axis and the inverse-analysis

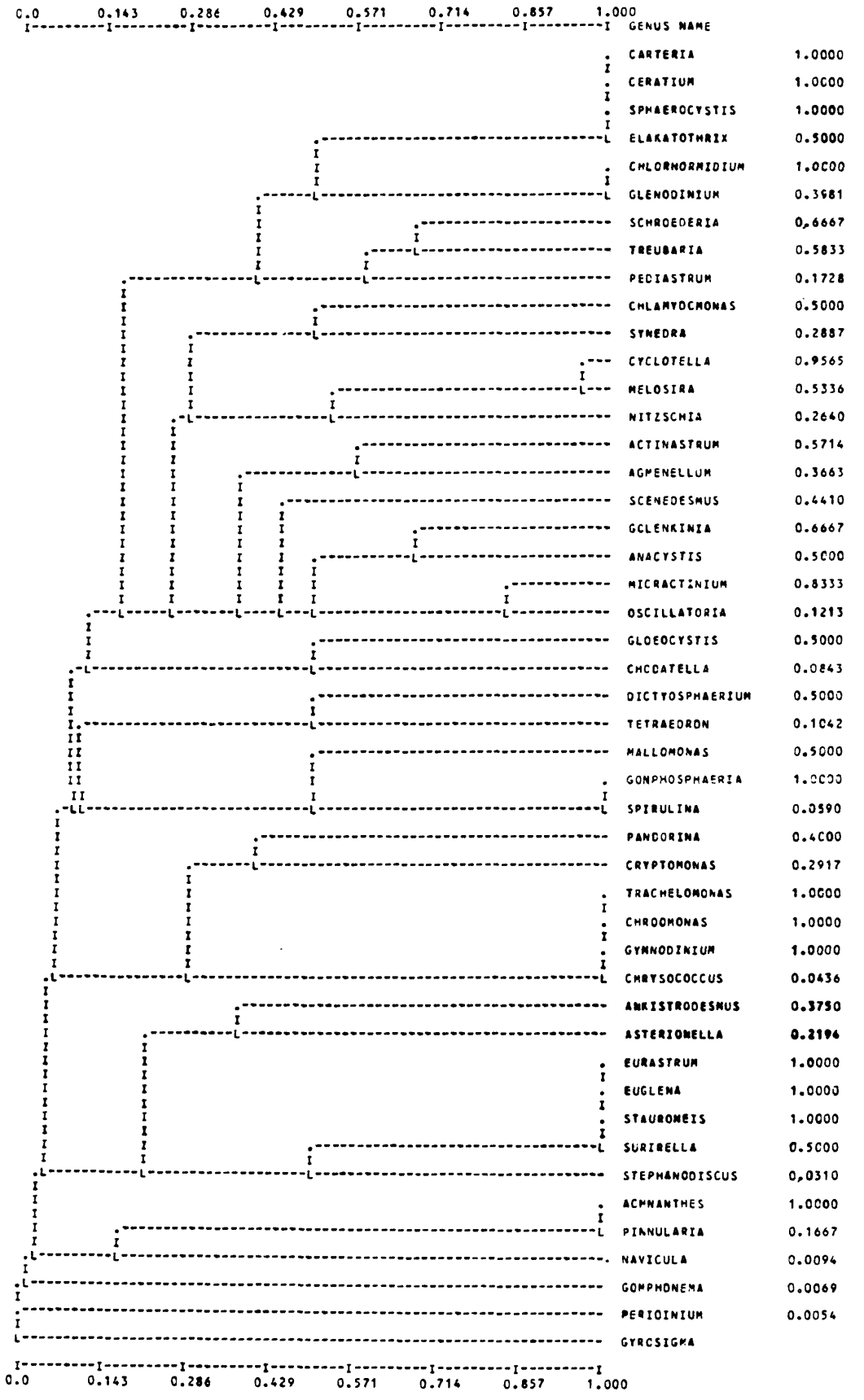


Figure 14.—Relationships of phytoplankton in the Tennessee River at Pickwick Landing Dam, Tennessee, 1975-77. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.885. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

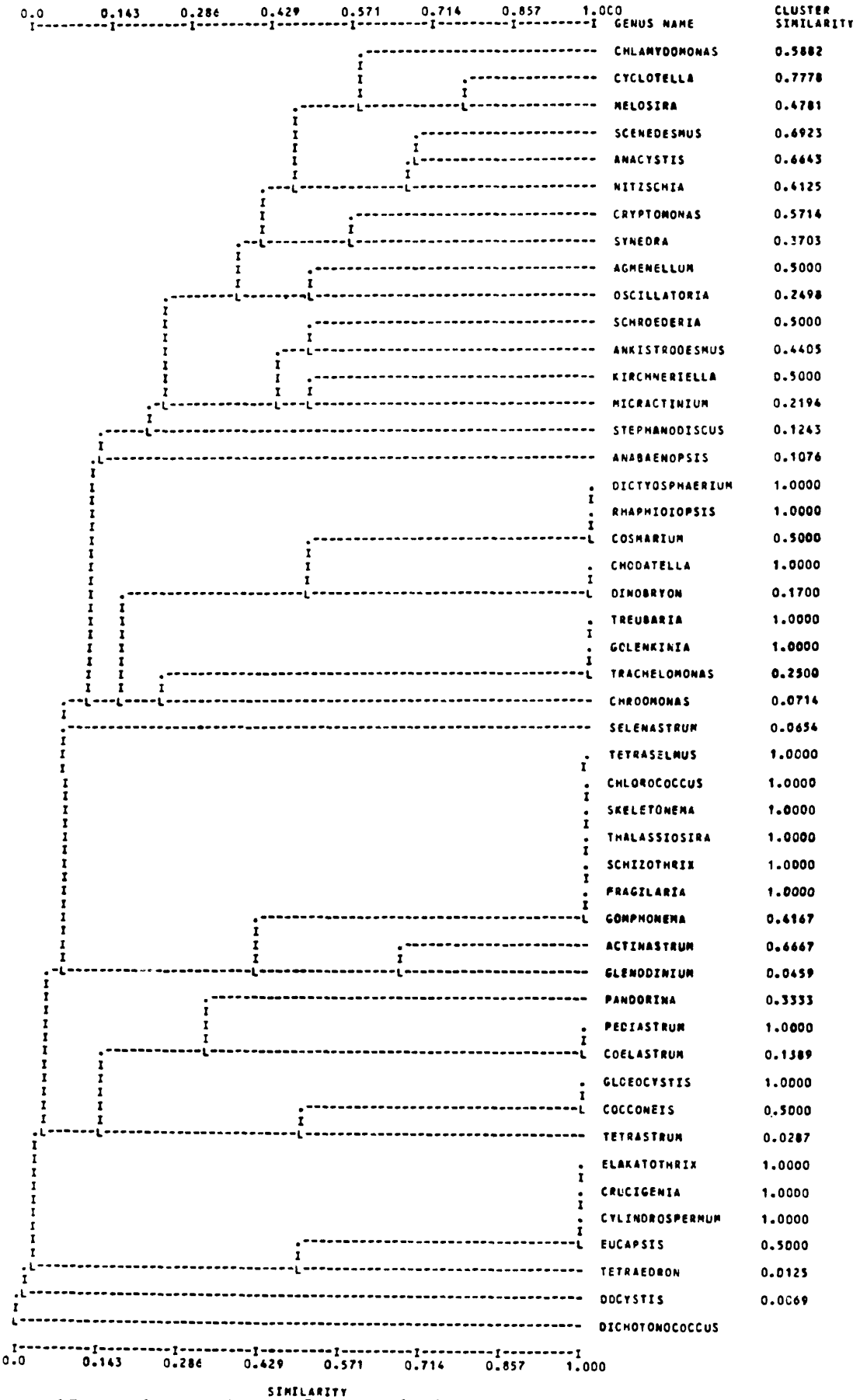


Figure 15.—Relationships of phytoplankton in the Tennessee River at Pickwick Landing Dam, Tennessee, 1979-81. Similarity indicates cluster similarity implied by the dendrogram with a cophenetic correlation coefficient of 0.909. (Scaling of the divisions on the similarity axis is not equal due to rounding.)

dendrogram on the other. An "X" is then placed at intersection points on the graph (nodes) where a specific genus occurs on a collection date. The resulting two-way coincidence plot is then used to provide semi-quantitative information on the composition of each cluster. This is preliminary to a nodal analysis (Boesch, 1977, p. 62) which may be used to identify misclassifications and enhance the ecological interpretation of the data.

Analysis of the two-way coincidence plot for the combined 1974-81 data from the French Broad River near Knoxville (pl. 1A) revealed a central cluster of genera which was present in most collections throughout the period: Melosira, Navicula, Nitzschia, Gomphonema, Achnanthes, Cyclotella, and Cymbella. A secondary group consisting of Ankistrodesmus, Scenedesmus, and Synedra also formed a relatively dense cluster of common organisms. Most other clusters of genera were formed of organisms occurring only occasionally. The greatest number of genera occurred in collections obtained between 1974 and 1977. Eighty-six percent of all "diverse" collections (those having 16 or more genera), were made prior to 1978. There were only two clusters formed at a similarity level greater than 0.57 in post 1977 collections, which is probably related to the decrease in organism diversity noted earlier.

Analysis of the two-way coincidence plot of data from the Watts Bar Dam station (pl. 1B) revealed a core cluster of very common organisms similar in composition to those from the French Broad River. Cyclotella, Melosira, Ankistrodesmus, Scenedesmus, and Nitzschia were present in almost every collection. Secondary groups which were also very common consisted of Chlamydomonas and Anacystis, and Synedra and Oscillatoria. Most other groups were not well defined due to sporadic occurrences. Seasonal variation was evident in the collections, with a large cluster of winter collections (November-April) and a separate cluster of summer collections (May-September). The number of genera within each collection was fairly uniform throughout 1974-80, and the maximum number of genera occurred during the months of June-September.

Collections obtained from the South Pittsburg station during 1974-81 consisted of the same common cluster of Cyclotella, Melosira, and Nitzschia, with less ubiquitous forms such as Chlamydomonas, Scenedesmus, Ankistrodesmus, Navicula, and Synedra (pl. 1C). The less common genera formed two clusters of organisms which were generally present only in collections obtained from November to April: Achnanthes, Gomphonema, Tetraselmus, Skeletonema, Diatoma, and Fragilaria. A large cluster of 11 genera (the Sphaerocystis-Aphanizomenon group in pl. 1C) were present only in collections obtained from June to August. The greatest number of genera in the collection occurred during July and August.

Fewer genera comprised the clusters of common organisms at the Pickwick Landing Dam station (pl. 1D). Cyclotella and Melosira represented the most common group, with Scenedesmus, Anacystis, Nitzschia, and Chlamydomonas in secondary clusters. There were only 5 collections with 15 or more genera, and one of those (the collection of September 13, 1979) had 7 genera that occurred in no other collection. All of the collections with 15 or more genera were obtained from June to August. There appeared to be no overall reduction in diversity during the period of data collection. In general, all collections which formed with a similarity value greater than 0.47 conformed to seasonal

groups such as winter and summer. The clusters formed in the winter collections typically had small diversity, consisting of only the most common forms.

Nodal Analysis

French Broad River near Knoxville, Tennessee

Nodal analysis represents an extension of the two-way plot in that it is concerned with the appearance of genera (or other taxa) in specific collections (site, season). The analysis is done to produce generalizations of the occurrence of a taxonomic cluster within the confines of a date or site cluster. The presence of indicator groups at specific times or locations then provides information about organism distribution and possibly water quality (Boesch, 1973).

Construction of a nodal two-way diagram allows the calculation of both constancy and fidelity. Constancy is the consistency of occurrence of taxa belonging to a particular taxonomic group in a particular collection group, and is expressed as the relative densities of the nodes. Constancy is arbitrarily graded from very high to very low, based on the proportions of the number of occurrences of taxa in the collection group to the total possible number of occurrences. The constancy value of each node is calculated as:

$$C_{ij} = a_{ij}/(n_i n_j) \tag{2}$$

where a_{ij} is the number of occurrences of members of taxa group i in collection group j, and n_i and n_j are the numbers of entities in the respective groups. The values range from zero, where none of the taxa occurred in the collections, to one where all group members occurred in all collections. Also, the size of each node is proportional to the sizes of the clusters of collections and genera forming them.

Fidelity is the faithfulness of taxonomic groups to collection groups and indicates the extent to which taxa are limited to specific habitats or seasons. The fidelity of taxonomic group i in collection group j is given by Boesch (1977, p. 64) as:

$$F_{ij}=(a_{ij}\sum_{j}^{i} n_j)/(n_j\sum_{j}^{i} a_{ij}) \tag{3}$$

where a_{ij}, and n_j are the same variables as given in the constancy calculation.

The fidelity is one when the constancy of a taxonomic group in a collection group is equivalent to its overall constancy, greater than one when its constancy is greater than the overall, and less than one when its constancy is less than the overall. Values greater than two indicate that a taxonomic group has a strong affinity for a collection group, and values much less than one may indicate an aversion to a collection group. Taxonomic groups may have a large constancy value in some collection groups, but may not be faithful in any (Boesch, 1977, p. 66).

The nodal constancy and fidelity diagrams for phytoplankton data collected from the French Broad River near Knoxville during 1974-81 are shown in plate 2A and B. The normal-mode clustering axis (collection groups) was divided into 21 groups from 52 collection dates. The inverse-mode clustering axis (genera groups) was divided into 19 groups from 67 genera. The group divisions for both axes are given in table 3. The branches on the dendrogram are identified by beginning and ending dates or genera names for the nodal group they represent (table 3).

In general, constancy of the collection groups indicated a division into two periods, 1974-77 and 1979-81, with very weak linkage between them. Discharge of the French Broad River during 1981 was only 54 percent of the annual average for 1974-81. Discharge of the river during 1979-80 was about the same as the 1974-81 annual average. Because of the smaller discharge during 1981, the water quality during 1979-81 was slightly different than during 1974-77. There was a significantly smaller (t test on means at the 95 percent confidence limit) concentration of dissolved substances as indicated by chloride and dissolved solids during 1974-77 than during 1979-81. This may have been a factor in the separation of collections 1 to 10 and 11 to 21 as distinct groups (pl. 2A, B). In general, collections 11 to 21 were obtained during relatively large discharge conditions when the water contained fewer dissolved solids. Phytoplankton populations in those collections indicated high constancy for the common diatom group C (pl. 2A). Fidelity of that group (pl. 2B) was also greater for those collections. A smaller, but well-defined cluster, consisting of collections 17 to 21, reflected high fidelity for many genera groups and high constancy for the common genera groups B and C (pl. 2A, B). Within this cluster, only 3 of the 17 collection dates were during 1980-81, and only the May 13, 1981 collection (in collection group 18) was representative of relatively small discharge conditions.

Genera groups B, C, and D dominated the typical communities in the French Broad River (pl. 2A, B). Genera group B consisted of Chlamydomonas, Ankistrodesmus, Scenedesmus, and Synedra, which are genera listed within the top 10 pollution-tolerant genera given in Palmer (1969, p. 79). In samples collected during August and September of the relatively small discharge years, group B had the largest constancy and fidelity. August and September are typical months of "nuisance" algal blooms. During 1974-77, constancy and fidelity of the same group was high in spring and winter when fidelity of several other groups was high; this may indicate a more diverse phytoplankton community. Group C (all diatoms) was very common at the site and generally displayed high constancy in the 1974-77 observations. There was a slight decrease in constancy of the group C genera during 1979-81. A third major genera group, D, was less constant than groups B or C, but it was a regularly occurring group. Genera group D displayed moderate to low constancy only, with a higher constancy during 1974-77 than during 1979-81.

The remainder of the genera groups consisted of phytoplankton which occurred rarely. Collectively, these "rare" types occurred more frequently during 1974-77. The nodal constancy and fidelity figures show that 1974-77 was characterized by Closteriopsis (group E), Staurastrum (group P), Skeletonema (group S), and genera groups F and L. Rhaphidiopsis (group R) was the only "rare" genus restricted to the 1979-81 period. Genera groups G, H, I, J, K,

Table 3.--Nodal groups formed of data from the French Broad River near Knoxville, Tennessee

Normal cluster analysis groups

1	7	13	18
September 2, 1981	August 20, 1977	November 14, 1977	May 13, 1981
September 8, 1980			May 19, 1980
August 28, 1979	8	14	May 19, 1975
2	November 17, 1980	May 4, 1976	July 15, 1975
September 27, 1977		April 1, 1975	19
August 11, 1976	9	June 3, 1975	September 9, 1975
3	September 21, 1981	15	November 4, 1975
March 15, 1979	June 30, 1980	May 17, 1979	December 9, 1975
4	July 30, 1979	January 10, 1977	20
August 11, 1981	10	February 22, 1977	March 4, 1980
5	September 3, 1980	November 22, 1976	January 21, 1976
August 6, 1980	March 16, 1976	16	December 7, 1976
6	December 18, 1974	February 23, 1976	October 6, 1975
September 14, 1981	11	July 13, 1976	21
July 30, 1981	May 24, 1977	January 28, 1975	June 28, 1977
	July 26, 1977	April 21, 1975	April 13, 1976
	12	August 14, 1975	June 14, 1976
	November 28, 1979	17	November 2, 1976
	February 25, 1975	September 13, 1976	November 25, 1974

Table 3.---Nodal groups formed of data from the French Broad River near Knoxville, Tennessee--Continued

Inverse cluster analysis groups

A	F	K	Q
Pyramimonas	Chodatella	Thalassiosira	Oocystis
	Phacus	Surirella	Euglena
B	Glenodinium	Amphora	Anabaena
Chlamydomonas	Selenastrum		
Ankistrodesmus	Rhoicosphenia	L	R
Scenedesmus	Ochromonas	Pandorina	Rhaphidiopsis
Synedra	Stauroneis	Crucigenia	
	Pinnularia	Mastogloia	S
C	Epithemia	Arthrospira	Skeletonema
Cyclotella		Peridinium	
Melosira	G	Lyngbya	
Navicula	Hantzschia	Golenkinia	
Nitzschia			
Gomphonema	H	M	
Achnanthes	Schroederia	Pediastrum	
Cymbella	Aphanizomenon	Tetraedron	
	Kirchneriella	Meridion	
D	Coelastrum		
Cocconeis	Eunotia	N	
Trachelomonas		Micractinium	
Anacystis	I	Plagiotropis	
Cryptomonas	Closterium	Neidium	
Oscillatoria	Stephanodiscus		
Asterionella	Spondylosium	O	
Fragilaria	Frustulia	Elakatothrix	
Dictyosphaerium		Tetrastrum	
	J		
E	Actinastrum	P	
Closteriopsis	Biddulphia	Staurastrum	
	Diatoma		

Table 4.--Rare genera characterizing two periods for the French Broad River near Knoxville, Tennessee

1974-77			
Group E:	Closteriopsis	Group L:	Pandorina
Group F:	Chodatella		Crucigenia
	Phacus		Mastogloia
	Glenodinium		Arthrospira
	Selenastrum		Peridinium
	Rhoicosphenia		Lyngbya
	Ochromonas		Golenkinia
	Stauroneis	Group P:	Staurastrum
	Pinnularia	Group S:	Skeletonema
1979-81			
Group R:	Rhaphidiopsis		

M, N, O, and Q were distributed throughout 1974-77 and 1979-81 (pl. 2A, B). The rare organisms present during the two periods are given in table 4.

The nodal constancy diagram indicated a change in the phytoplankton community structure in the French Broad River near Knoxville. The nodal analysis generally indicated decreases in generic diversity, nodal constancy, and nodal fidelity with time, and there was a change in the types of rare organisms observed in the French Broad River. This trend was observed in the inverse analyses of three non-overlapping periods mentioned earlier. Some of the differences could be due to sampling or to seasonal factors, but there appeared to be a general decrease in numbers of genera with time. The shift in types of rare organisms showed a decrease in the number of diatoms observed and an increase in green and blue-green algae. These trends may have been associated with changing water quality in the French Broad River.

Tennessee River at Pickwick Landing Dam, Tennessee

Nodal constancy and fidelity for phytoplankton data collected from the Tennessee River at Pickwick Landing Dam during 1975-81 is presented in plate 2C and D. There are a total of 42 collection dates represented in the 12 collection groups of the normal analysis and 68 genera in the 19 groups of the inverse analysis (table 5). Node-forming cluster groups in the normal and inverse analyses consist of one to eight members. Genera listed on the dendrogram branches for each group represent the initial and final members of the group as taken from the inverse analysis.

The collection groups represented seasonal trends, some well defined, and some poorly defined (mixed). Group 3 was a spring group, and groups 5-7 were summer groups. Group 8 was the only definite winter group among the nodal divisions. Although seasonally mixed, groups 11 and 12 may be indicative of temperate conditions common to both fall and spring. Groups 1, 2, 4, and 9 were individual collections that did not appear to be closely related to other clusters, although group 9 could possibly have been combined with group 10. Groups 1, 2, and 4 consisted of genera that were observed rarely in 1975-81 data, and were unique in composition. No attempt was made to reallocate collections or genera into other clusters to improve interpretation of the relationships.

Genera group A had low to very low constancy throughout the collection period, and the nodal fidelity diagram indicated that this group had a preference for spring and summer (groups 1, 3 and 6). That genera group, however, had low fidelity in collection group 7, which was composed of spring and summer collections. Those collections were very similar to group 6. This apparent dichotomy can be resolved by inspection of the two-way coincidence table (pl. 1D), which shows that one collection in group 6 (June 21, 1977) had an unusually high proportion of organisms that otherwise were rarely observed during the entire collection period; this resulted in a large fidelity value for that particular node. The genus Glenodinium accounted for five of the twelve occurrences of organisms comprising group A, and it was present in collection group 1. Otherwise, the distribution of genera group A was very sparse and was limited to spring and summer collections. Genera group A was formed almost solely because of the mutual occurrences of genera in the June 21, 1977 collection.

Table 5.--Nodal groups formed of data from the Tennessee River at Pickwick Landing Dam, Tennessee

Normal cluster analysis groups

1	7	11
September 13, 1979	August 26, 1975	September 30, 1975
	August 25, 1977	April 10, 1979
2	August 20, 1980	May 15, 1979
September 16, 1981	July 27, 1977	
	July 16, 1981	12
3	August 13, 1981	May 21, 1975
March 6, 1980		March 8, 1976
May 21, 1981	8	November 24, 1975
March 19, 1981	January 6, 1976	June 16, 1976
	February 10, 1976	May 15, 1980
4	February 15, 1977	September 14, 1976
January 25, 1977	November 1, 1976	November 16, 1977
	November 19, 1980	November 16, 1979
5	9	
August 16, 1976	July 20, 1976	
June 12, 1979		
July 11, 1979	10	
	July 14, 1975	
6	April 5, 1976	
May 24, 1977	November 4, 1975	
June 21, 1977	May 4, 1976	
August 15, 1979		
June 18, 1981		
July 17, 1980		
June 19, 1980		

Table 5.--Nodal groups formed of data from the Tennessee River at Pickwick Landing Dam, Tennessee--Continued

Inverse cluster analysis groups:

A	I	O
Carteria	Tetraedron	Gloeocystis
Ceratium	Crucigenia	Cocconeis
Sphaerocystis	Cylindrospermum	Coelastrum
Elakatothrix	Eucapsis	Tetrastrum
Chlorhormidium	Selenastrum	
Glenodinium		P
	J	Eurastrum
B	Mallomonas	Euglena
Chlamydomonas	Gomphosphaeria	Stauroneis
Scenedesmus	Spirulina	Surirella
Anacystis	Arthrospira	Asterionella
		Stephanodiscus
C	K	Anabaenopsis
Cyclotella	Achnanthes	
Melosira	Pinnularia	Q
Nitzschia	Navicula	Dichotomococcus
D	L	R
Cryptomonas	Tetraselmis	Peridinium
Synedra	Chlorococcum	
	Skeletonema	S
E	Thalassiosira	Gyrosigma
Ankistrodesmus	Schizothrix	
Actinastrum	Fragilaria	
Agmenellum	Gomphonema	
	Pandorina	
F		
Pediastrum	M	
Treubaria	Oocystis	
Golenkinia		
Micractinium	N	
Oscillatoria	Trachelomonas	
	Gymnodinium	
G	Chrysococcus	
Schroederia	Chroomonas	
Kirchneriella		
H		
Dictyosphaerium		
Chodatella		
Dinobryon		
Cosmarium		
Rhaphidiopsis		

Genera groups B, C, and D consisted of frequently occurring algae (pl. 2C, D). Genera group B had high to very high constancy in collection groups 1, 3, 5, 6, and 7; these were spring and summer groups, and with only two exceptions, all were collections made in 1977-81. Collection groups 8-10 and 12 were mainly 1975-76 collections where genera group B had only low or very low constancy. Group B had moderate constancy and low fidelity (between 0.5 and 1.0) in collection group 11 (seasonally temperate). Fidelity for genera group B indicated that it was most prevalent in spring and summer collections 3, 5, 6, and 7. Genera group C generally had high to very high constancy throughout the collection period, and it formed the nucleus of most collections. Its ubiquity, however, limits its use as an indicator. The group also had an average fidelity (between 0.5 and 1.5) for the major seasonal groups indicating that this group was likely to be found in most samples. Genera group D has high to very high constancy in collection groups 4, 5, and 6. It also had fidelity greater than 1.0 in collection group 7, and greater than 2.0 in groups 4, 5 and 6 (pl. 2D). Groups 5-7 were summer collections and group 4 was a winter collection, but occurrences of the genera in genera group D were most likely in summer. Genera group D had 2 genera and collection group 4 consisted of a single collection. Therefore, the fidelity expressed for collection group 4 was not very meaningful because the occurrence of only one genera in the sample represented 50 percent of the genera group. Also, genera group D had low constancy, and fidelity of less than 0.5 in collection group 8 (winter), indicating that there were fewer numbers of these genera in colder water.

Genera group E had scattered occurrences in all seasonally defined collections, as shown in the two-way coincidence plot, and generally moderate or very low constancy throughout the collection period. The highest fidelity for this group occurred in spring and summer collections (groups 1, 5, 6, and 9). The coincidence plot (pl. 1D) shows that the genus Ankistrodesmus occurred in all collections comprising groups 8 and 4 (winter collections), resulting in a fidelity greater than 1.5. Other members of genera group E, however, were absent in those nodes. Genera group E also had a fidelity greater than 1.5 in collection groups 1, 2, and 9 (July and September collections) due to the occurrence of other genera in the nodal group (pl. 2D). The pattern of occurrences in the coincidence plot (pl. 1D) indicates that Ankistrodesmus was tolerant of a wider range of seasonal conditions than the other members of genera group E (Actinastrum and Agmenellum), which appeared to be acclimated to more temperate conditions.

Genera group F showed moderate to high constancy in the summer collection groups 6-7, and genera group G had high constancy in collection groups 1 and 6. Analysis of the two-way coincidence plot shows that both genera groups F and G were uncommon in collection groups 2, 3, 5, and 8-12, where they have very low constancy (pl. 2C). Two out of five organisms from genera group F were observed in collection group 4 (January 25, 1977), and one out of two organisms in genera group G was found in collection group 1 (September 13, 1979). This indicates, respectively, a moderate and high constancy for these nodes. The fidelity of these genera groups indicates that they tended to occur more often in the summer collections. Observation of the nodes in the two-way coincidence plot confirms that organisms in genera groups H-S occurred sporadically, and as a result, they exhibited low to very low constancy in the nodal diagram except in single-date collections (groups 1, 2, and 4), where they contributed to the uniqueness of the individual-collection groups (pl.

2C). Values of fidelity for these genera indicate that genera groups K, M, and P had an affinity for winter collections, and all other genera in groups H-S were more likely to be associated with spring and summer collections.

The nodal fidelity diagram (pl. 2D) portrays the seasonal nature of the genera groups, and it shows that a large proportion of the genera groups had higher fidelity in collection groups 3-7. The normal analysis dendrogram and both the nodal fidelity and constancy diagrams indicate a division between collection groups 3-7 and 8-12 which respectively approximate the 1977-81 and 1975-76 collection periods. With respect to this division, the two-way coincidence plot and nodal constancy diagrams show a greater diversity and abundance of genera in the 1977-81 collections than in the 1975-76 collections. Groups B and E consisted of genera of green and blue-green algae which tended to form blooms indicative of large nutrient concentrations. Their relative scarcity in 1975-76 and increased appearance in 1977-81 suggests an increase in water temperature or nutrient concentrations with time at Pickwick Landing Dam, however, compilation of data on temperature and concentrations of total forms of ammonia, nitrite plus nitrate, and phosphorus indicated there was no significant increase in mean concentrations of these nutrients between 1975-76 and 1977-81. Other genera groups at this site occurred either too commonly, or were composed of such a large variety of genera (with widespread tolerances of individual species), that interpretation is difficult.

Tennessee River Basin

The nodal constancy and fidelity diagrams for all sites combined are illustrated in plate 2E and F. There is a total of 185 collections for the four sampling stations in which 104 genera occurred at various times. Forty-nine of the genera are listed in the 60 most pollution-tolerant genera of algae (Palmer, 1969, p. 79). There are 31 collection groups and 27 genera groups that best depict the normal and inverse clustering relationships. The components of the nodal groups are listed in table 6.

Jaccard coefficients for both the normal and inverse analyses were often smaller than 0.60, which implies weak relationships among members of clusters, however, the cophenetic correlation coefficient for the inverse analysis was 0.83. This indicates that the inverse analysis dendrogram was a good representation of genera similarities based on the values represented by the Jaccard coefficient of similarity. The cophenetic correlation coefficient was 0.69 for the normal analysis, indicating that a number of collections had larger Jaccard similarity coefficients than implied by the dendrogram. Because the majority of the Jaccard values were below 0.50, however, no attempt was made to reallocate the collection members.

There were only weak tendencies among the collection groups to form clusters based on seasonal, temporal, or collection site affinities. Collection groups 27-31 formed a weakly defined group dominated by data from the French Broad River. In this assemblage, 79 percent of the collections were from the French Broad River and 21 percent from the Tennessee River (16 percent from South Pittsburg, 5 percent from Watts Bar Dam, and none from Pickwick Landing Dam). These collections seemed to represent an upstream assemblage and were not seasonally differentiated.

Table 6.--Nodal groups formed of data from the Tennessee River basin, 1974-81

(FB-French Broad, WB-Watts Bar Dam, SP-South Pittsburg, PL-Pickwick Landing)

Normal cluster analysis groups

1	2	3	4	5	6	7
FB 9-2-81	SP 9-28-77	FB 8-11-76	PL 9-16-81	SP 11-24-80	SP 9-15-76	FB 9-3-80
			FB 8-11-81	WB 7-27-77	WB 9-29-77	SP 7-13-79
			SP 6-30-77	SP 11-3-76	SP 7-28-77	FB 11-17-80
			FB 8-6-80		FB 7-26-77	WB 5-20-75
			FB 8-28-79			FB 9-27-77

8	9	10	11	12
WB 3-6-80	SP 6-16-76	SP 8-26-81	WB 12-3-79	WB 1-11-77
PL 9-13-79	FB 5-24-77	PL 7-20-76	WB 5-30-79	WB 11-24-76
SP 5-26-77		PL 5-21-81	SP 4-8-81	SP 11-23-76
		WB 11-4-76	WB 12-9-76	
		PL 3-6-80	SP 5-31-79	
		SP 5-28-81	WB 9-5-80	
		PL 5-15-79		
		PL 4-10-79		
		FB 9-8-80		
		PL 3-19-81		

13	14	15	16	17	18
WB 8-2-77	SP 8-5-76	PL 6-18-81	SP 8-29-79	PL 8-20-80	PL 6-21-77
WB 9-10-75		PL 6-19-80		PL 8-25-77	SP 8-27-80
FB 8-20-77		WB 11-15-77		PL 8-26-75	PL 7-11-79
		SP 6-26-80		SP 8-30-77	PL 6-12-79
		WB 3-29-79			SP 5-29-80
		SP 7-21-75			WB 6-25-80
					WB 8-30-79
					PL 8-13-81
					PL 7-27-77
					PL 7-17-80
					SP 9-9-80
					WB 2-26-75
					WB 8-5-80
					WB 5-28-80
					WB 9-26-79
					WB 11-5-75
					PL 7-16-81
					SP 9-11-75
					SP 7-15-76
					SP 8-6-75
					WB 7-12-79
					PL 8-15-79
					WB 9-14-76
					FB 9-14-81
					PL 5-24-77

Table 6.--Nodal groups formed of data from the Tennessee River basin, 1974-81--Continued

19
WB 8-12-76
WB 7-14-76
WB 6-29-77
FB 9-9-75

20
SP 12-8-76

21
PL 9-14-76
WB 10-7-75
SP 4-15-76

22
PL 11-16-77
PL 11-16-79
PL 3-8-76
PL 5-21-75
PL 5-15-80
PL 6-16-76
PL 11-24-75
WB 3-17-76
FB 11-28-79

23
SP 3-18-76
WB 11-20-80
SP 9-16-81
SP 6-25-81
PL 11-19-80
PL 11-1-76
SP 5-21-75
FB 3-15-79

24
WB 5-5-76
SP 1-30-75
PL 1-25-77
SP 2-24-77
WB 12-19-74
WB 2-23-77
SP 2-27-80
PL 2-15-77
SP 4-23-75
FB 3-16-76

25
PL 5-4-76
PL 11-4-75
WB 4-22-75
SP 11-17-77
SP 9-30-81
FB 9-21-81
FB 7-30-79

26
FB 5-4-76
PL 7-14-75
PL 9-30-75
SP 11-27-74
SP 11-29-79
SP 3-28-79
WB 3-26-75
PL 2-10-76
SP 12-11-75
WB 11-26-74
SP 10-8-75
WB 4-14-76
SP 5-6-76
SP 6-5-75
PL 1-6-76
SP 11-6-75
FB 11-4-75
PL 8-16-76
WB 2-19-76
FB 10-6-75
FB 12-18-74

27
FB 6-30-80
SP 12-20-74
FB 2-25-75

28
FB 11-14-77

29
SP 3-27-75
FB 5-4-76
SP 2-25-76
FB 6-3-75
FB 4-1-75

30
FB 2-22-77
FB 1-10-77
FB 5-17-79
FB 2-23-76
FB 7-13-76
FB 8-14-75
WB 1-29-75
SP 2-27-75
FB 4-21-75
FB 1-28-75
FB 11-22-76

31
SP 1-12-77
FB 9-13-76
FB 7-30-81
SP 1-20-76
FB 5-13-81
FB 5-19-80
FB 7-15-75
FB 5-19-75
WB 6-15-76
FB 12-7-76
FB 3-4-80
FB 1-21-76
FB 6-28-77
FB 12-9-75
FB 11-2-76
FB 6-14-76
FB 4-13-76
FB 11-25-74

Table 6.—Nodal groups formed of data from the Tennessee River basin, 1974-81--Continued

Inverse cluster analysis groups

A	H	Q	V
Pyramimonas	Synedra	Tetraselmus	Chlorogonium
	Oscillatoria	Skeletonema	Anabaenopsis
B	Cryptomonas	Chlorococcum	Cosmarium
Carteria		Schizothrix	Phormidium
Chlorhormidium	I	Thalassiosira	
Glenodinium	Achnanthes	Tetrastrum	W
Pandorina	Navicula		Closterium
Actinastrum	Gomphonema	R	Eunotia
	Cymbella	Chlorella	Tabellaria
C	Cocconeis	Euglena	Spondylosium
Schroederia		Ceratium	Frustulia
Treubaria	J	Synura	
Micractinium	Trachelomonas	Rhopalodia	X
Kirchneriella		Anabaena	Biddulphia
Chodatella	K		Diatoma
Golenkinia	Fragilaria	S	Amphora
		Pteromonas	
D	L	Ulothrix	Y
Sphaerocystis	Stephanodiscus	Chrysococcum	Eurastrum
Franceia		Gloeocystis	Stauroneis
Pediastrum	M	Staurastrum	Epithemia
Agmenellum	Asterionella	Chroomonas	Hantzschia
Tetraedron	Surirella	Dinobryon	
Crucigenia		Gymnodinium	Z
Peridinium	N		Neidium
	Coelastrum	T	Plagiotropis
E	Aphanizomenon	Elakatothrix	
Oocystis	Lyngbya	Eucapsis	∝
	Dictyosphaerium	Cylindrospermum	
F			Gyrosigma
Chlamydomonas	O	U	Westella
Anacystis	Ochromonas	Errerella	Meridion
Ankistrodesmus	Rhaphidiopsis	Mallomonas	Polyedriopsis
Scenedesmus		Gomphosphaeria	Acanthosphaera
	P	Spirulina	Dichotomococcus
G	Closteriopsis	Mastogloia	
Cyclotella	Phacus	Arthrospira	
Melosira	Pinnularia		
Nitzschia	Selenastrum		
	Rhoicosphenia		

A well-defined cluster of genera was dominant in this assemblage, which had a very high constancy in collection groups 28-31, and was formed of genera primarily from the French Broad River. This upstream genera group (I) consisted of the diatoms Achnanthes, Navicula, Gomphonema, Cymbella, and Cocconeis. This group was also associated with 10 other collection groups, where it had low to moderate nodal constancy and it occurred in collection groups that included French Broad River data in all but two of the collection groups. The fidelity diagram (pl. 2F) indicated that group I was most faithful to collection groups 27-31.

Collection groups 8-24 were almost exclusively from sites downstream of the confluence of the French Broad and Tennessee Rivers. Only 8 of the 99 collections in those groups were from the French Broad River near Knoxville, Tennessee. Within those downstream groups, collections appeared to be seasonally differentiated rather than site differentiated. There were well defined seasonal groups consisting of summer (groups 9, 13-19), winter (group 12), and two groups (groups 8 and 10) that were seasonally mixed but represented temperate conditions (pl. 2E, F). These collections consisted of a variety of genera but were dominated by genera groups B-H. Genera group I, which was indicative of genera from the French Broad River, lacked fidelity and had low to very-low constancy in these collections.

Of the 27 genera groups in the nodal constancy diagram (pl. 2E), there was one that maintained moderate to very high constancy throughout the collection groups, and was equally faithful to upstream and downstream stations. Genera group G consisted of the diatoms Cyclotella, Melosira, and Nitzschia, which were found to be the most commonly occurring genera in all of the cluster analyses performed. Genera group F was also quite common among collections, although it was not as common as group G. Both groups F and G were listed in the top 20 most pollution-tolerant genera (Palmer, 1969, p. 79), and the presence of these organisms may be indicative of organic pollution (Palmer, 1969, p. 81), however, the seasonal and temporal persistance of both groups throughout the study period did not indicate any pronounced changes in water-quality with time. All other genera groups tended to have lower constancies and did not distinguish collection sites or trends in the nodal diagram.

A statistical summary of selected chemical and physical characteristics of the water for dates in the major collection groups listed in table 6 is presented in table 7. Data were not included for collection groups represented by less than three collection dates. The chemical and physical data indicate that collection groups 27-31 were representative of water with generally lower mean temperatures than collection groups 8-24 (collection groups 12 and 24 were collected at the lowest mean water temperatures), but other water-quality properties are not significantly different.

The nodal diagrams for data combined from all sites did not indicate any ecological changes in the Tennessee River system. Since a large number of genera observed in the collections are tolerant of a wide range of conditions, it is not likely that subtle changes in water quality could be detected with those organisms. The diagrams mildly differentiated the collection site on the French Broad River from other stations. This may be due to the geographic position of the collection sites, with the French Broad River near Knoxville

Table 7.--Statistical summary of selected water-quality characteristics for major collection groups from the Tennessee River basin

	Temperature °C	Dissolved Chloride (mg/L)	Total dissolved solids (mg/L)	Total Nitrite + nitrate (mg/L as N)	Total Phosphorus (mg/L as P)
Group 4					
Mean	16.6	7.7	93.3	0.36	0.060
Standard deviation	6.9	2.3	17.9	.12	.030
Number of samples	8	8	7	8	8
Group 5					
Mean	17.3	7.7	107.3	.33	.053
Standard deviation	8.0	1.0	18.8	.05	.035
Number of samples	3	3	3	3	3
Group 6					
Mean	25.7	8.5	93.0	.36	.030
Standard deviation	1.5	.8	7.9	.13	.010
Number of samples	4	4	4	4	4
Group 7					
Mean	20.6	14.0	113	.40	.040
Standard deviation	5.4	6.3	19.1	.19	.030
Number of samples	5	4	4	4	4
Group 8					
Mean	14.3	5.7	89.7	.37	.030
Standard deviation	11.7	.5	9.5	.15	.010
Number of samples	2	3	3	3	3
Group 10					
Mean	18.5	7.7	96.3	.46	.040
Standard deviation	7.9	3.1	19.4	.15	.010
Number of samples	9	8	7	6	4
Group 11					
Mean	14.8	6.6	88.6	.35	.080
Standard deviation	5.3	2.4	12.3	.60	.110
Number of samples	5	5	5	5	5

Table 7.--Statistical summary of selected water-quality characteristics for major collection groups from the Tennessee River basin--Continued

	Temperature °C	Dissolved Chloride (mg/L)	Total dissolved solids (mg/L)	Total Nitrite + nitrate (mg/L as N)	Total Phosphorus (mg/L as P)
Group 12					
Mean	8.0	8.3	92.3	0.37	0.030
Standard deviation	3.5	1.2	2.5	.05	.010
Number of samples	3	3	3	3	3
Group 13					
Mean	27.0	9.5	102	.24	.030
Standard deviation	.9	3.9	14.8	.05	.010
Number of samples	3	3	3	3	3
Group 15					
Mean	21.6	6.4	90.2	.36	.040
Standard deviation	7.1	1.8	8.3	.10	.010
Number of samples	6	6	6	6	6
Group 17					
Mean	29.4	7.3	91.7	.18	.050
Standard deviation	.7	.78	12.0	.11	.010
Number of samples	4	3	3	3	3
Group 18					
Mean	24.7	6.6	93.3	.28	.030
Standard deviation	4.6	2.6	13.5	.10	.010
Number of samples	26	24	24	23	3
Group 19					
Mean	24.9	--	--	--	--
Standard deviation	1.3				
Number of samples	4				
Group 21					
Mean	19.7	6.2	90.5	.34	.030
Standard deviation	5.3	1.1	14.8	.10	0
Number of samples	3	2	2	2	2

Table 7.--Statistical summary of selected water-quality characteristics for major collection groups from the Tennessee River basin--Continued

	Temperature °C	Dissolved Chloride (mg/L)	Total dissolved solids (mg/L)	Total Nitrite + nitrate (mg/L as N)	Total Phosphorus (mg/L as P)
Group 22					
Mean	16.4	--	--	--	--
Standard deviation	5.0				
Number of samples	9				
Group 23					
Mean	16.6	7.7	93.3	0.36	0.060
Standard deviation	6.9	2.3	17.9	.12	.030
Number of samples	8	8	7	8	8
Group 24					
Mean	8.3	6.6	91.6	.40	.040
Standard deviation	4.3	1.7	16.2	.11	.020
Number of samples	10	10	10	9	9
Group 25					
Mean	18.4	8.8	94.5	.36	.030
Standard deviation	3.6	3.8	24.3	.13	.010
Number of samples	7	5	4	5	5
Group 26					
Mean	14.4	7.9	90.8	.40	.030
Standard deviation	6.7	3.4	13.4	.09	.010
Number of samples	21	14	14	14	14
Group 27					
Mean	10.5	7.0	87.0	.47	.033
Standard deviation	6.5	1.0	1.7	.10	.021
Number of samples	3	3	3	3	3
Group 29					
Mean	12.1	5.5	75.2	.43	.030
Standard deviation	2.9	2.5	15.1	.06	.030
Number of samples	5	5	5	5	5

Table 7.--Statistical summary of selected water-quality characteristics for major collection groups from the Tennessee River basin--Continued

	Temperature °C	Dissolved Chloride (mg/L)	Total dissolved solids (mg/L)	Total Nitrite + nitrate (mg/L as N)	Total Phosphorus (mg/L as P)
Group 30					
Mean	10.6	7.8	91.2	0.50	0.040
Standard deviation	7.0	3.5	17.2	.15	.020
Number of samples	11	11	11	11	11
Group 31					
Mean	13.6	8.4	88.1	.45	.030
Standard deviation	7.5	3.0	11.3	.13	.010
Number of samples	18	17	16	16	17

being further upstream and the downstream stations being linked by a nearly continuous stretch of reservoirs which receive additional inflow from the Clinch and Little Tennessee River systems.

SUMMARY AND CONCLUSIONS

Data obtained at four NASQAN stations in the Tennessee River basin from 1974 to 1981 were examined to identify patterns of occurrence of phytoplankton genera and to determine if these patterns were changing with time. Stations were selected that represented nearly continuous monthly sampling with an equivalent period of record for each station. Inverse cluster analyses were performed on non-overlapping periods of collection at each station to determine if groups of genera formed recognizable community structures. The data for each station were analyzed using normal and inverse clustering methods. Two-way coincidence plots were constructed and used to investigate trends in the occurrence of phytoplankton communities. Nodal constancy and fidelity diagrams were constructed from two-way plots to present the findings in a semi-quantitative form. Finally, the data from the four stations were combined into one data set to explore trends for the Tennessee River basin through nodal-analysis of phytoplankton occurrence.

Inverse cluster analyses of data from the stations indicated that some well-defined groups of phytoplankton genera could be distinguished in the collections. The initial analysis showed changes in patterns of occurrence of both dominant and "rare" genera. Comparison of the data at each station showed that the composition of phytoplankton communities changed with time and distance downstream. Upstream communities from the French Broad River near Knoxville to the Tennessee River at South Pittsburgh were dominated by the genera Nitzschia, Navicula, Melosira, and downstream at Pickwick Landing Dam, Cyclotella-Melosira dominated communities became more prevalent. Data from upstream stations indicated a shift in the rare organisms with decreasing numbers of diatoms and increasing numbers of green and blue-green algae. However, further downstream at Pickwick Landing Dam, there was an increase in the number of genera of rare organisms with time rather than a shift in their occurrence.

Normal and inverse cluster analyses of phytoplankton data at each station indicated seasonal patterns of diversity in the Tennessee River at Watts Bar Dam, South Pittsburg, and Pickwick Landing Dam and reinforced the hypothesis of a changing community structure in the French Broad River near Knoxville.

Associations of phytoplankton genera at the three stations in the Tennessee River were found to be strongly seasonal. Winter and summer phytoplankton communities were distinguished from each other in the samples. There was greater diversity of phytoplankton in the samples taken in the summer (June-September). The two-way coincidence plot of data from the French Broad River near Knoxville did not indicate seasonal patterns of phytoplankton communities. At this station, a central cluster of genera was found to be common in most collections throughout 1974-81. Two periods, 1974-77 and 1979-81, could be distinguished from each other by the diversity of the phytoplankton communities. Clusters of 16 or more genera were found to be more commmon in the 1974-77 collections than in the 1979-81 collections.

Construction of nodal analysis diagrams using data from the French Broad River near Knoxville helped to refine observations indicated in the two-way plot. Diatoms were common in most samples from that station, and as indicated by the nodal constancy diagram, those diatoms were found throughout 1974-81. However, their frequency of occurrence was reduced in the late summer months, and there was a slight decrease in constancy during 1979-81. Also, the diagrams showed that there was a greater number of genera in the 1974-77 collections than in the 1979-81 collections. Concentrations of chloride and dissolved solids were significantly smaller during 1974-77 than during 1979-81.

Analysis of the nodal diagrams of phytoplankton data from Pickwick Landing Dam confirmed that clusters were formed seasonally, as indicated by analysis of the two-way coincidence plot. Also, the diagrams showed presence of two groups of green and blue-green algae. Those two groups were relatively scarce in the 1975-76 collections from this station. They became more common in the 1977-81 collections, but these changes did not correspond to changes in water temperature or concentrations of the nutrients nitrogen and phosphorus.

Nodal analysis diagrams were not constructed using data from the Watts Bar Dam and South Pittsburg stations because the authors felt that the two-way coincidence plots presented the data in a suitable form for interpretation, and little or no additional information would be gained from the nodal analyses.

The data from all four stations were combined into a single data set. Inverse and normal clustering methods were applied to the data set and nodal diagrams were prepared. This analysis did not indicate that phytoplankton communities in the Tennessee River basin were changing with time. The nodal analysis did differentiate data from the French Broad River from all others, but seasonal and generic assemblages specific to a single station were not primary factors in forming clusters in the majority of instances.

The cluster analysis described in the report provided an effective means of reducing and summarizing the large amount of phytoplankton data collected during a seven year period at four NASQAN stations in the Tennessee River basin. The analysis of various data sets indicated some changes in the community structure at two of the stations examined, which may indicate a deterioration in water quality. At the upstream station (French Broad River near Knoxville, Tennessee), there was evidence that the diatom dominated community was being replaced by increasing numbers of genera of green and blue-green algae. The station furthest downstream (Pickwick Landing Dam, Tennessee) also displayed an increasing frequency of genera of green and blue-green algae during the collection period.

Inverse cluster analyses of data from the two intermediate collection stations for successive two-year periods indicated changing community structure, but a more complete analysis of the data for those stations using both normal and inverse clustering methods and a two-way coincidence plot indicated that observed community changes were closely related to seasonal changes.

Dendrograms, two-way coincidence plots, and nodal diagrams were used to summarize patterns of occurrence of phytoplankton. Each presents the data in a different manner and different conclusions can be gained from each presentation. Small data sets can be adequately interpreted and summarized through the use of dendrograms, but in larger and more complex data sets, two-way coincidence plots and nodal diagrams may be more suitable for investigating the nature of the data.

Nodal analysis, while providing a data reduction and interpretive method, does have some drawbacks. In attempting to summarize the data from the four stations using a single combined data set, station-specific data were lost. Differences in distribution of the genera due to seasonality and site characteristics became less evident as the size of the data set increased. For the combined data set, a seven-year period of data collection probably was not sufficient to detect significant changes in community structure, as was indicated in the analysis of individual data sets. Also, the Tennessee River system may have been a comparatively stable aquatic environment during 1974-81. Moreover, subjectivity is inescapable in interpreting cluster analyses, and conclusions of one investigator may differ from those of other researchers.

Based on the results obtained in this study, it is conceivable that the effectiveness of water-management practices can be qualitatively analyzed using biological data collected during a given period of time, and longer periods of data collection can provide a better baseline from which to evaluate changes in water quality. Taxonomic determinations need to be done to the species level to permit more definitive interpretation of the data. The wide range of tolerances of individual species within a genus precludes accurate use of genera as indicators of water quality. Cluster analysis provides a usable tool for the reduction and visualization of those data, but ultimately, it is the individual who must interpret the results without introducing an objectionable degree of subjectivity.

REFERENCES CITED

Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T., 1979, SAS User's Guide: SAS Institute, Inc., Raleigh, N. C., 494 p.

Boesch, D. F., 1973, Classification and community structure of macrobenthos in the Hampton Roads Area, Virginia, Marine Biology, vol. 21, p. 221-244.

_____1977, Application of numerical classification in ecological investigations of water pollution: U.S. Environmental Protection Agency, Report EPA-600/3-77-033, 114 p.

Bonham-Carter, G. F., 1967, Fortran IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers: Kansas State Geological Survey Computer Contribution no. 17, 28 p.

Brown, S. D., 1969, Grouping plankton samples by numerical analysis: Hydrobiologia, vol. 33, p. 289-301.

Cairns, John Jr., and Kaesler, R. L., 1969, Cluster analysis of Potomac River survey stations based on protozoan presence-absense data: Hydrobiologia, vol. 34, p. 414-432.

Clifford, H. T., and Stephenson, William, 1975, An introduction to numerical classification: New York, Academic Press, 229 p.

Crossman, J. S., Kaesler, R. L., and Cairns, John Jr., 1974, The use of cluster analysis in the assessment of spills of hazardous materials: American Midland Naturalist, vol. 92, no. 1, p. 94-114.

Gauch, H. J., 1979, COMPCLUS--A Fortran program for rapid initial clustering of large data sets: Dept. of Ecology and Systematics, Cornell University, Ithaca, New York, 59 p.

Hellawell, J. M., 1978, Biological surveillance of rivers: Herts, England, Water Research Centre, 331 p.

Hirsch, R. M., Slack J. R., and Smith, R. A., 1982, Techniques of trend analysis for monthly water quality data: Water Resources Research, vol. 18, no. 1, p. 107-121.

Kaesler, R. L., 1970, The cophenetic correlation coefficient in paleoecology: Geological Society of America Bulletin, vol. 81, p. 1261-1266.

Kaesler, R. L., and Cairns, John Jr., 1972, Cluster analysis of data from limnological surveys of the upper Potomac River: American Midland Naturalist, vol. 88, no. 1, p. 56-57.

Palmer, C. M., 1969, A composite rating of algae tolerating organic pollution: Journal of Phycology, vol. 5, p. 78-82.

Patil, G. P., Pielou, E. C., and Waters, W. E., eds., 1971, Statistical ecology, volume 3--Many species populations, ecosystems, and systems analysis: University Park, Pennsylvania State University Press, 462 p.

Pinkham, C. F. A., Pearson, J. G., Clontz, W. L., and Asaki, A. E., 1975, A computer program for the calculations of measures of biotic similarity between samples and the plotting of the relationship between these measures: U.S. Army, Edgewood Arsenal Technical Report EB-TR-75013, 38 p.

Rohlf, F. J., 1970, Adaptive heirarchial clustering schemes: Systematic Zoology, vol. 18, p. 58-82.

_____1985, NTSYS: Numerical taxonomy system of multivariate statistical programs. Technical report, Department of Ecology and Evolution, State University of New York, Stony Brook, 109 unnumbered pages.

Sokal, R. R., and Rohlf, F. J., 1962, The comparison of dendrograms by objective methods: Taxon, vol. 9, no. 2, p. 31-40.

Steele, T. D., Gilroy, E. J., and Hawkinson, R. O., 1974, An assessment of areal and temporal variations in streamflow quality using selected data from the National Stream Quality Accounting Network: U.S. Geological Survey, Open-File Report 74-217, 210 p.