Classifying

Numerical Taxonomy. Proceedings of a colloquium, St. Andrews, Scotland, Sept. 1968. A. J. COLE, Ed. Academic Press, New York, 1969. xvi + 328 pp., illus. $9.

This volume presents the 18 papers given at a colloquium in which discussions centered on how one should classify numerically and not whether one should.

There are papers on the application of the techniques of numerical taxonomy to particular fields such as ecology, business, botany, paleontology, social geography, and molecular biology. Two papers are concerned with implementation of various techniques on digital computers. In addition, author and subject indexes and an appendix giving short descriptions of the programs used by the speakers are furnished.

A number of papers are of special significance because they give good reviews and evaluations of the techniques of numerical taxonomy or because they present new ideas.

A. J. Boyce presents a useful although not critical review of multidimensional techniques for displaying the patterns of diversity found in a numerical taxonomic study. He discusses similarity and distance coefficients, cluster analyses, and principal components analyses.

A. V. Hall describes his heterogeneity functions which have been developed for two-state and quantitative data and contrasts their use with that of Euclidean distances.

D. M. Jackson deals with techniques for comparing two classifications by measuring the extent to which each agrees with the initial data. The technique discussed is based upon a model of binary (presence or absence) characters. Class membership is, like the characters, a two-valued relation, in that an object either belongs or does not belong to a given class. Thus both the original information and the classification may be represented completely by binary arrays. Basically the method measures the monotonicity of similarity functions computed from the raw data and from the classification. Rather than computing a total index of discrepancy one uses a threshold so that one has a measure of discrepancy in classifications for varying values of similarity.

I. C. Lerman has investigated techniques for finding the partition of the total set of OTU's (operational taxonomic units) which best approximates the similarities among the objects (that is, which best separates the OTU's into two sets with maximum within-group homogeneity). He compares two criteria for the partitioning of a set.

L. Orloci presents a general introduction to information theory and its applications to taxonomy, both in nonhierarchic sorting and in hierarchial cluster analyses.

P. H. A. Sneath discusses the various concepts that could be used to develop techniques for evaluating clustering methods. He distinguishes between external criteria (where one compares a classification to some absolute standard) and internal criteria (where one performs some sort of comparison between the classification and the original data upon which it is based). He distinguishes the three general concepts (mass, density, and networks) involved in the definition of a cluster and stresses the desirability of statistical tests for the presence of clusters as well as considerations of the feasibility of actually carrying out the clustering algorithms on present-day computing equipment.

A paper by D. Wishart is an important contribution to numerical taxonomy. Wishart gives a concise survey of various clustering techniques that possess minimum-variance constraints on the clusters. The underlying axiom of these techniques is the intuitive idea that a resultant group of individuals should be homogeneous in relation to the total set of variables. Described in geometric terms, the cloud of points which constitutes a minimum-variance cluster should be hyperspheroidal in shape. The effects upon the minimum-variance solution of adding nondiagnostic and irrelevant variables to the character set are discussed. The major objection to partitions obtained by a minimum-variance method is that they may easily cut across a dense swarm of points with the result that on either side of the partition there will be a large number of individuals that are practically identical (which of course defeats the objective of the analysis). There are relatively few techniques that do not have this dependency. The most widely known exception is the technique of single-linkage cluster analysis, but this technique suffers from "chaining effect" when the clusters are very fuzzy as a result of a large amount of noise. Wishart proposes a technique called mode analysis which differs from previous techniques in that the data points which surround each OTU at a specified distance threshold are first counted and all the non-dense points—those for which the frequency is less than some critical value—are temporarily removed from the analysis. The remaining, "dense" points are clustered by means of single linkage cluster analysis, forming the mode nuclei. Then each non-dense point is allocated to a cluster according to some criterion (for example, each non-dense point may be assigned to the cluster associated with the nearest dense point). Unfortunately no example of the use of this program is given in the paper. But the technique has sufficient theoretical interest to warrant its being tested on numerous data sets.

F. JAMES ROHLF
Department of Biology, State University of New York, Stony Brook

Transport Phenomena

Kinetic Processes in Gases and Plasmas. A. R. HOCHSTIM, Ed. Academic Press, New York, 1969. xiv + 458 pp., illus. $22.50. Reentry Physics Series.

This book consists of ten reviews by different authors of various aspects of the kinetic theory of gases and plasmas. It is written primarily to acquaint persons generally familiar with the field with recent developments and current research.

The subject matter of the book may be divided into three parts. The first consists of "textbook-like" derivations of the conventional sets of transport equations, among them the Chapman-Enskogg method, the Fokker-Planck equation, and the Lenard-Balescu equation. These derivations are lucid and carefully developed.

The second part consists of chapters by E. A. Mason and by A. R. Hochstim and G. A. Massel. Mason discusses applications of transport theory to neutral gases, whereas Hochstim and Massel describe applications to ionized gases. The chapter by Mason gives an excellent review of the role of inelastic collisions, including rotational excitation, in transport phenomena. This includes a survey of various calculational recipes and accuracy. Hochstim and Massel give a very detailed review of transport phenomena in ionized gases, including extensive numeri-