NOAA Technical Report EDS 19

Separation of Mixed Data Sets into Homogeneous Sets

Washington D. C.
January 1977

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Environmental Data Service

NOAA TECHNICAL REPORTS

Environmental Data Service Series

The Environmental Data Service (EDS) archives and disseminates a broad spectrum of environmental data gathered by the various components of NOAA and by the various cooperating agencies and activities throughout the world. The EDS is a "bank" of worldwide environmental data upon which the researcher may draw to study and analyze environmental phenomena and their impact upon commerce, agriculture, industry, aviation, and other activities of man. The EDS also conducts studies to put environmental phenomena and relations into proper historical and statistical perspective and to provide a basis for assessing changes in the natural environment brought about by man's activities.

The EDS series of NOAA Technical Reports is a continuation of the former series, the Environmental Science Services Administration (ESSA) Technical Report, EDS.

Reports in the series are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: $3.00 paper copy; $1.45 microfiche. When available, order by accession number shown in parentheses.

ESSA Technical Reports

EDS 1 Upper Wind Statistics of the Northern Western Hemisphere. Harold L. Crutcher and Don K. Halligan, April 1967. (PB-174-921)

EDS 2 Direct and Inverse Tables of the Gamma Distribution. H. C. S. Thom, April 1968. (PB-178-320)

EDS 3 Standard Deviation of Monthly Average Temperature. H. C. S. Thom, April 1968. (PB-178-309)

EDS 4 Prediction of Movement and Intensity of Tropical Storms Over the Indian Seas During the October to December Season. P. Jagannathan and H. L. Crutcher, May 1968. (PB-178-497)

EDS 5 An Application of the Gamma Distribution Function to Indian Rainfall. D. A. Mooley and H. L. Crutcher, August 1968. (PB-180-056)

EDS 6 Quantiles of Monthly Precipitation for Selected Stations in the Contiguous United States. H. C. S. Thom and Ida B. Vestal, August 1968. (PB-180-057)

EDS 7 A Comparison of Radiosonde Temperatures at the 100-, 80-, 50-, and 30-mb Levels. Harold L. Crutcher and Frank T. Quinlan, August 1968. (PB-180-058)

EDS 8 Characteristics and Probabilities of Precipitation in China. Augustine Y. M. Yao, September 1969. (PB-188-420)

EDS 9 Markov Chain Models for Probabilities of Hot and Cool Days Sequences and Hot Spells in Nevada. Clarence M. Sakamoto, March 1970. (PB-193-221)

NOAA Technical Reports

EDS 10 BOMEX Temporary Archive Description of Available Data. Terry de la Moriniere, January 1972. (COM-72-50289)

EDS 11 A Note on a Gamma Distribution Computer Program and Graph Paper. Harold L. Crutcher, Gerald L. Barger, and Grady F. McKay, April 1973. (COM-73-11401)

EDS 12 BOMEX Permanent Archive: Description of Data. Center for Experiment Design and Data Analysis, May 1975.

EDS 13 Precipitation Analysis for BOMEX Period III. M. D. Hudlow and W. D. Scherer, September 1975. (PB-246-870)

EDS 14 IFYGL Rawinsonde System: Description of Archived Data. Sandra M. Hoexter, May 1976. (PB-258-057)

EDS 15 IFYGL Physical Data Collection System: Description of Archived Data. Jack Foreman, September 1976.

(Continued on inside back cover)

NOAA Technical Report EDS 19

Separation of Mixed Data Sets into Homogeneous Sets

Harold L. Crutcher and Raymond L. Joiner
National Climatic Center
Asheville N.C.

January 1977

U.S. DEPARTMENT OF COMMERCE
Elliot L. Richardson, Secretary

National Oceanic and Atmospheric Administration
Robert M. White, Administrator

Environmental Data Service
Thomas S. Austin, Director

Stock Number 003-019-00036-5   Price $2.45

ACKNOWLEDGMENTS

Appreciation is expressed to the many who have helped us in the preparation of this paper. Among these are the personnel of the National Climatic Center's Science Advisory Staff, ADP Services Division, Audio-Visual Services Section, and the Library Group. Specific acknowledgment is made to Miss Lisa Green for preparation of the original typescript and to Mrs. Margaret Larabee for the careful preparation of the final typescript.

Acknowledgment is made to Prof. E. S. Pearson and the Trustees of Biometrika to use the data and format displayed in figures 2a and 2b. Acknowledgment is made to University Microfilms, Ann Arbor, Michigan, for permission to reproduce or to modify figures as these appear on figures 4, 5, and 6.

Appreciation is expressed to the U. S. Navy and to Dr. John H. Wolfe of the U. S. Navy Personnel Research and Development Center, San Diego, California, to adapt and use his NORMIX and NORMAP computer programs and for his ever ready response to written, telephonic, and personal visit requests.

Acknowledgment is made to the National Oceanic and Atmospheric Administration for permission to quote material from the *Monthly Weather Review*.

Acknowledgment is made also to Prof. A. Clifford Cohen, Jr. and to Mr. Lee Falls for permission to use their computer program which they furnished and which was adapted to separate two mixed univariate normal distributions.

Appreciation also is expressed to the staff at EDS's Environmental Science Information Center, in particular to Mr. Patrick McHugh, for the care given to the editing of this paper.

CONTENTS

FIGURES

TABLES

Symbols used in this report

d	difference; deviation
d.f.	degrees of freedom
f	function
ft.	feet
gdkm	geodynamic kilometer
i	subscript or superscript
j	subscript or superscript
k	kth point in a sample; number of clusters; kth cluster
km	kilometer
m	mth item; meter
mb	pressure in millibars
mi.	miles
n	nth item; number in a sample
r	sample correlation coefficient
s	sample standard deviation; second; cluster
s^2	sample variance
t	Student's "t"
x	variate
x'	variate transpose
y	variate
C	covariance matrix; Celsius
F	function
G.C.T.	Greenwich Civil Time
$\hat{P}$	probability
$\hat{P}(S\|X_k)$	probability of membership of X_k in the cluster s
R	correlation matrix
$\|R\|$	determinant of the correlation matrix R
S	type; cluster type
X	observed value of variate
$\bar{X}$	mean of variate X
X_k	vector of observations for the kth point in the sample
Y	observed value of variate
$\bar{Y}$	mean of variate Y

α	alpha; proportionality factor ($\hat{\lambda}_s$); probability level of rejection for the null hypothesis
$\hat{\lambda}_s$	lambda hat; mixing proportion for type cluster s
Λ	lambda; the diagonal matrix of eigenvectors
μ	mu; population mean
$\hat{\mu}_s$	mu hat; mean vector for cluster s
π	pi
ρ	rho; population correlation
σ	sigma; population standard deviation
σ^2	population variance
$\hat{\sigma}$	sigma hat; covariance matrix for cluster s
Σ	sigma; summation
ψ	psi
$\hat{}$	hat (caret)
$=$	equal to
$\simeq$	approximately
$-$	overbar; averaging process
$'$	transpose

SEPARATION OF MIXED DATA SETS INTO HOMOGENEOUS SETS

Harold L. Crutcher and Raymond L. Joiner
National Climatic Center
Environmental Data Service, NOAA
Asheville, N.C.

ABSTRACT. In any study, the collection, processing, and storage of data are important. Whether the data are clean, biased or contaminated is also important. Pollution or adulteration of data confuse the investigator.

Data do not necessarily fall into neatly packaged boxes or groups. Usually the data sets are mixtures of several types of phenomena. Some of these are basically deterministic in nature while others are not.

This paper illustrates the use of a clustering technique to separate mixed data sets into subsets which exhibit group characteristics. The investigator then assesses the relative importance of the subsets, the nature of the subsets, and perhaps makes an assumption as to whether a particular subset is biased, contaminated, or adulterated. That is, an assessment of the quality of the data may be made.

The techniques are applicable to any data set which is multivariate normal. Here, they are applied to weather data subsets, (1) land-sea breeze, (2) tropical stratospheric winds, (3) mid-latitude tropospheric winds, (4) mountain pass winds and temperatures, (5) surface marine weather temperatures, dew points and winds, and (6) radiosonde observation of heights, winds, temperatures, and dew points.

1. INTRODUCTION

Prehistoric man differentiated between the good and the harmful, between winter and summer, between drought and floods, and among many variable factors affecting him.

In the real world, the experienced hunter knew the differences between the bear, boar, deer, and turkey. His senses of sight, smell, and hearing aided him when he could not see the animal, but he could sense its presence. When he had some models in mind, he drew symbols or wrote words for the benefit of the inexperienced. He could even provide measurements of a sort. At this stage of the game, he began to deal more with the abstract.

Later, man attempted to record experiences, his thoughts, and his aspirations in pictographs and monuments. (Even today, artists present forms in symbolic representation or as configurative concepts.) Undoubtedly, many numbering systems had also been developed and were then lost in antiquity, for there are some which remain indecipherable today.

The above examples may be generalized to any field. Clear-cut numerical descriptions and configurations remain important whether the field be sociology, psychology, geophysics or medicine, four of an infinite number of fields.

In this paper we illustrate some of the techniques used to differentiate groups within sets of given measurements. Hopefully, the measurements are both accurate and precise. Also, hopefully, the measurements obtained and used are those which will provide good differentiation bases.

As Friedman and Rubin (1967) quote from Bose and Roy (1938): "The problems of discrimination and classification are insistent in sciences."

Once we enter the realm of measurements and numbers in multidimensional space, assumptions and decisions are made as to the better or best characteristics. The metrics that can be used are many. Those chosen ought to provide the most useful differentiation possible.

There are many techniques and procedures used to differentiate groups. These usually involve some measures of central tendencies within the groups, some measures of differences of these central tendencies, variability within and among the groups, group shapes, and scales, etc. As more is known about the normal distribution than other distributions, it is wise, wherever possible, to transform non-normally distributed data sets to data sets which may be approximated by the normal distribution. Transformations have been of interest for many years. In fact, these are represented in the change of base in many counting systems. For example, the logarithm transformation changes a zero bounded positively skewed distribution to an unbounded distribution at both ends where the values more distant from one are scaled downward faster than those nearer one.

If the various features of a data set are each transformed to normal or near normal distributions, then the combined features may be multivariate normal. A multivariate normal distribution has normal marginal distributions. However, the fact that the marginal distributions are normal does not assure multivariate normality. If multivariate normality is assumed, then probabilistic statements can be made. Transformation will be discussed in greater detail later in the paper.

If multivariate normality is assumed and the distribution has one centroid, i.e., it is unimodal, then multivariate regression techniques can be used to produce forecast equations. If the distributions are normally distributed in the multivariate sense but the total set is multimodal, then ordinary linear regression techniques will not serve. Regression equations for clusters must be developed and the better sets of regression equations clustered and examined. The unique or best equation or set of equations can be used. If there is a best equation, it is unique.

Figure 1 illustrates a multimodal trivariate distribution. The various clusters take various ellipsoidal forms such as spheres, ellipsoids, and disks. Each one of these is trivariate normal. Each one represents a centroid or grouping of characteristics.

Some may be equal in all directions, such as in the spheres. Some may be equal in two directions but not the third, such as in the football- and disk-shaped ellipsoids. In others the distribution may be unequal in all directions. In figure 1, no scales are indicated as the illustration is for concept only. The illustration also could be considered as a higher dimensional ensemble projected onto three dimensions. When multimodal features are evident, and if one wishes to study the modes or clusters, then techniques other than regression are required to separate the data into appropriate subsets.

There are many techniques to separate distributions into parts which can be studied individually. Some of these are:

(a) discriminant function analysis
(b) factor analysis
(c) principal cluster analysis
(d) dendritic (tree) analysis
(e) cluster analysis
(f) clumping
(g) numerical taxonomy
(h) unsupervised pattern recognition
(i) typology

Some of these are essentially the same.

Mixtures always present problems when they must be separated. Characteristics may or may not be so noticeable that classification and discrimination can be made. A mixed herd of cattle, sheep, goats, and horses may be easily separated though the sheep and goats may sometimes present a few problems. The herdsman, the separator, or the investigator must have a clear picture in his mind, i.e., a model which includes the necessary characterization(s) of the populations in which he is interested. If he doesn't have his senses of sight, smell, hearing and touch to guide him (for these provide him explicit models) and he has only some measurements of the mixture provided to him, he is in a quandry. He no longer has an explicit model. He has to start somewhere. Good (1965) discusses the philosophical problem of deciding what can be the best beginning in the problem of classification and discrimination. First of all, the investigator decides to accept the characterization of each object by a set of measurements. He believes that there should be some categories or sub-categories which will be helpful in distinguishing group characteristics. Explicitness is lost and there is no external criterion with which to define the categories. An internal criterion (Good, 1965) would be acceptable. That is, the data themselves may suggest "natural categories." The word "suggest" is necessary, for a different beginning in the treatment of the data may lead to slightly different categories. This item will be treated in more detail later.

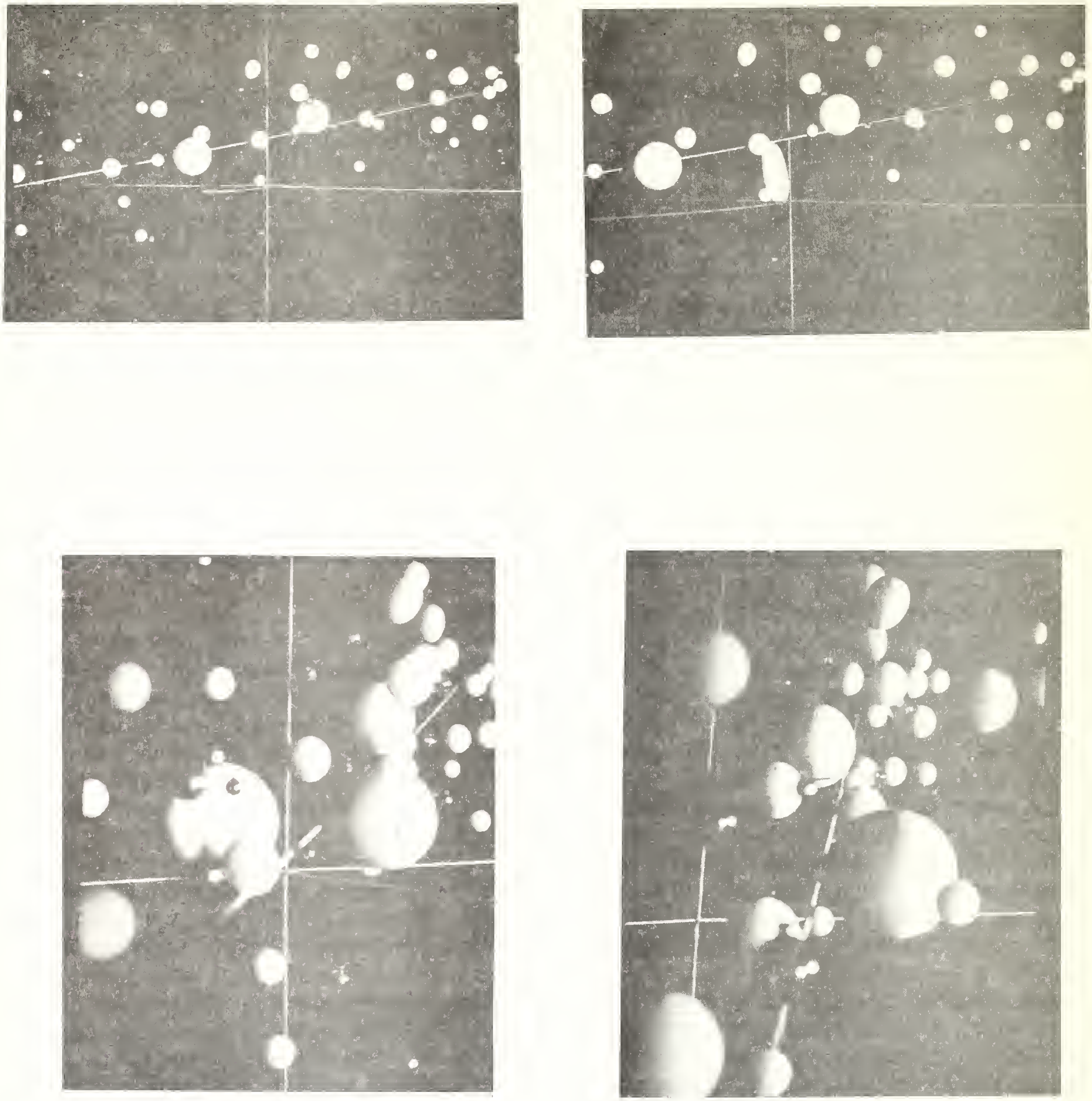


Figure 1 A mixed set of trivariate distributions showing clusters or modes of varying sizes and shapes. These are viewed from four different points in 3-space. In each view, the three axes and a suggested line of best fit are indicated.

The next few sections take the reader through a short discussion of regression techniques and through some of the various techniques used to provide classification and discrimination within heterogeneous mixtures.

2. REGRESSION TECHNIQUES

Inevitably, due to inherent laziness or an inherent desire to get to a goal with the least expenditure of energy, physical or mental, one attempts to read relationships into sets of experiences or of data. This is a canon of science. For example, if one knows that a certain thing will happen provided that something else specific is done, then there is a clear-cut and obvious relationship between the reaction and the prior action. Though the converse may not be true, it is disregarded here.

Galton (1889) and Snedecor and Cochran (1967) noted the height of sons as related to the heights of the fathers. With the heights of the fathers as one set of data and the heights of the sons as a second set of data, Galton plotted the heights of the fathers against the heights of the sons, respectively. He noted that tall fathers did not always produce as tall or taller sons but that the height of a son seemed to be between the height of the father and the mean height of the group. That is, the height of a son regressed towards the group norm. The line of best fit for the data set was then and since then labeled the line of regression for any line relationship between data sets. Pearson and Lee (1903), Galton's associates, collected a set of data (more than a thousand) of stature, cubit, and span in family groups.

Figure 2a shows the regression of sons' statures on the fathers' statures. Please note that the heights of the sons of short fathers also tend (or regress) toward the mean. This figure is taken from Pearson's and Lee's data as illustrated, but is also shown by Snedecor and Cochran (1967). The data are scaled in the metric system here. Also, note that the data scatter uniformly along the line and do not cluster.

Figure 2b illustrates the relationship between a sister's span for a given forearm length of her brother. Both figures are adapted from Pearson and Lee with the kind permission of the Trustees of Biometrika.

Figure 3a illustrates schematically the correlation at one point. The correlation may be said to be perfect on the one hand, but also it can be said to be indeterminant.

Figure 3b illustrates the correlation between two points. The correlation may be said to be perfect on the one hand but for the space in between the points, on the line connecting the points, it may be said to be indeterminant or even zero.

Figure 3c illustrates a correlation of one with three collinear data.

Figure 3d illustrates the case of two clusters each with zero correlation, one superposed on the other.

Figure 3e illustrates the case of two clusters shown in figure 3d where the clusters are slightly separated with cluster 2 moving away on a line of 45 degrees. The correlation coefficient is something greater than zero but much less than one; i.e., $0<r<<1$.

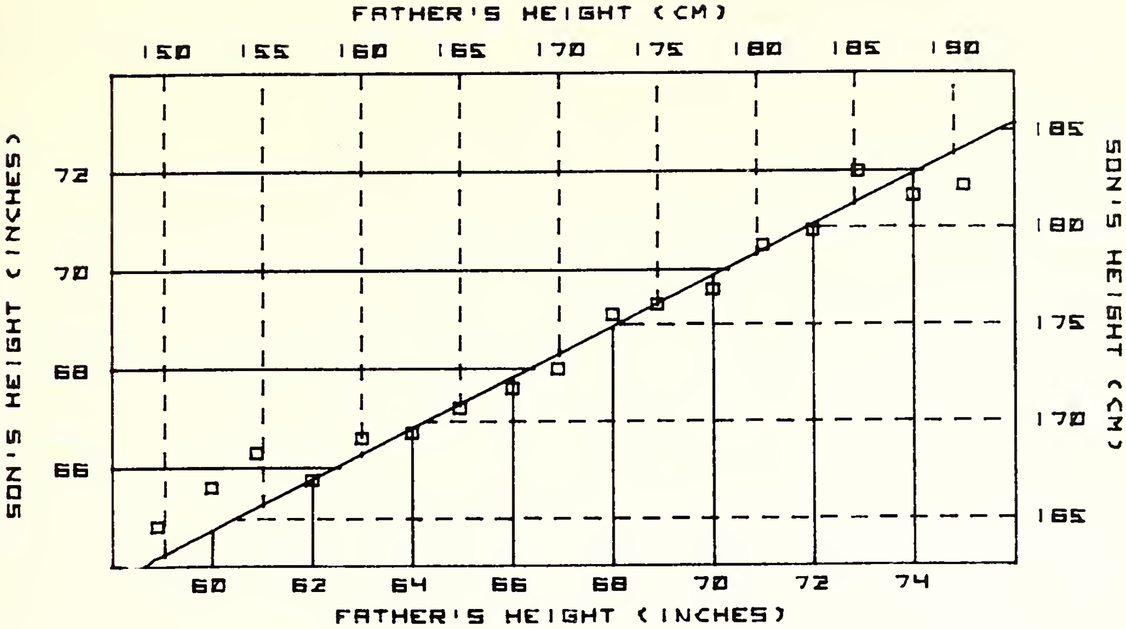


Figure 2a Regression of sons' statures on the fathers' stature, in. and cm., adapted from Galton (1889) and Pearson and Lee (1903).

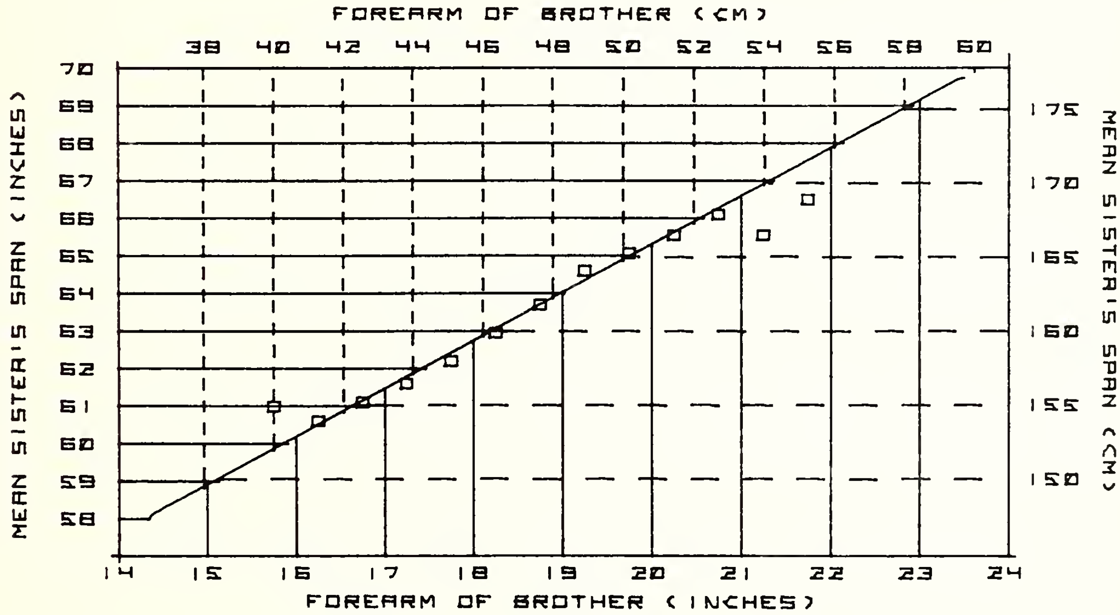


Figure 2b Regression of sister's span, in. and cm., for given forearm of brother, adapted from Galton (1889) and Pearson and Lee (1903).

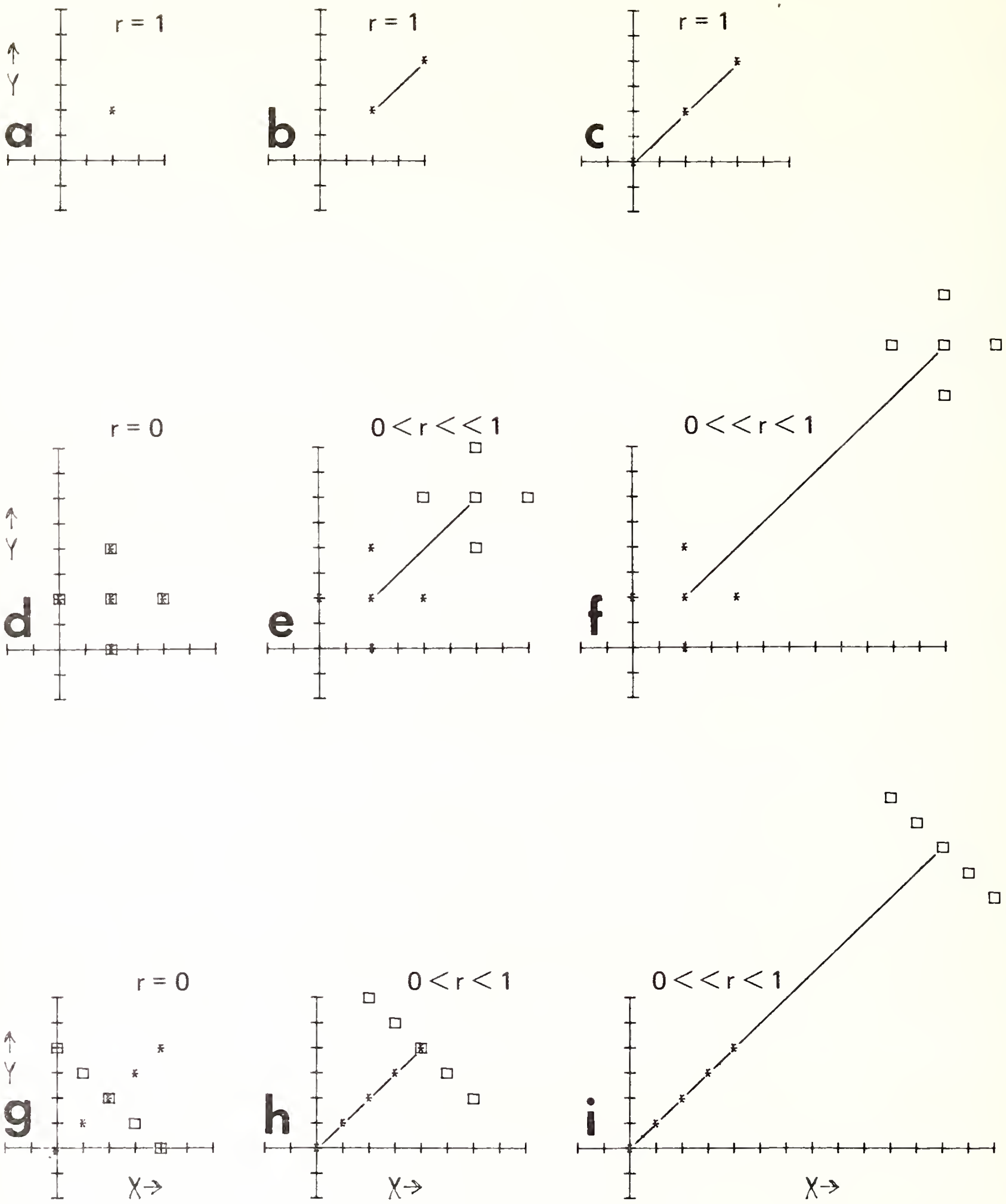


Figure 3 General schematic of points and clusters with implied correlations, r.

Figure 3f illustrates the case of two clusters shown in figure 3d where the clusters are separated still further with cluster 2 moving farther away on the 45 degree line.

Figure 3g illustrates the case of two superposed clusters. One with a correlation of plus one, the other with a correlation of minus one. The total cluster has a correlation of zero.

Figure 3h illustrates the case of two clusters one with a correlation of plus one, the other with a correlation of minus one. The second cluster is moving on a 45-degree line from the first. The distance between cluster centroids is the same as in figure 3c. The group correlation is greater than zero but less than that shown in figure 3e.

Figure 3i illustrates the case of the two clusters shown in figure 3g moving along the 45-degree line farther than shown in figure 3h. The correlation is increasing and the coefficient approaches one but not as rapidly as in figure 3f as the distance between the centroids increases.

The above discussion implies that there may be two or more clusters in any data ensemble. The scatter may be more in one cluster than in another. Also, the internal correlation (or dispersion) within each cluster may be the same or may be different from the other clusters. The total correlation attained thus may be more of the relationship among the clusters than points. As the distance between the centroids increases, the clusters are more like singleton points in the regression analyses. If there is a clustering, i.e., effectively a dearth of observations between groups either real or simply unobserved, then regression analysis whether linear or non-linear will fail. Please refer again to figure 1 which represents clusters in 3-space or clusters in n-space projected into 3-space. The representation here is multinormal. This may not be the case sometimes and some clusters or all clusters may be eccentric in shape such as eggs, starch grains, or bivalve shells.

If the variates are not all normally distributed, it is assumed that the user will transform the variates to normal variates. It is helpful to know by means of the Central Limit Theorem that, though individual data characteristics may not be normally distributed, linear functions of these tend to be normally distributed. Also, it is assumed that the user will extract the deterministic part of the variables wherever possible. Therefore, the problems of nonlinear regression are not discussed here.

The point of the entire discussion above is simply that linear regression analysis ought to be used only with a unimodal univariate or multivariate distribution. Linear predictor equations developed in linear regression models then become more useful and accurate.

Gupta and Sobel (1962) consider this problem. Aversen and McCabe (1975) present some subset selection problems for variances associated with applications to regression analysis.

3. DISCRIMINANT TECHNIQUES

3.1 Discriminant Functions

Many workers realized the problems induced by heterogeneous or mixed distributions. Among the first to attack the problem in a systematic mathematical treatment was Pearson (1894, 1901) in work on univariate distributions. Many others also have worked on this problem.

Barnard (1935) and Fisher (1936) may be considered to have first attacked the problem of classification with discrimination techniques, though the problems of classification had involved many other workers up to that time.

Let us look at a two cluster mixture from the viewpoint of separation or discrimination with subsequent rules for classification. Figure 4 shows an assemblage composed of two clusters. The clusters are shown first within a dashed circle with no axes chosen (a). Suppose that the measurements are made in terms of the x-axis drawn horizontally (b).

Projections of the cluster points are on the x-axis and in (d) on the y-axis. Visually there is separation in the (x, y) space or two space. This separation can also be shown in one space, i.e., linearly. In one space or one dimension chosen first on the x-axis, there is mixture of the projections. In (c) where the x-axis is rotated through an angle to x', there is some separation but two points indicate some mixing. In (d) where the rotation has been carried through 90 degrees so that the x' axis is equivalent to the former y-axis, separation is complete though perhaps not the optimum. There is some angle of rotation which will produce a major separation between the clusters and a minimum variance within the clusters. The computed linear function which describes the above line after rotation is called the linear discriminant function. Brown (1947) applies these techniques to establish the discriminating procedures for azotobacter, the nitrogen-fixing bacteria. Smith (1947) provides some discrimination examples. Crutcher (1960) applies the techniques developed by Rao (1950) to the annual march of temperature and rainfall climates in the United States. Figure 5 (adapted from Crutcher, 1960) shows a two cluster (bimodal) bivariate distribution with the two dimensions shown in three-dimensional form. These three-dimensional forms are projected into two dimensions on the $\bar{x}z$ plane. Linear discrimination is effected along the single axis pointing to the lower right ($\bar{x}\bar{y}$ plane). Figure 6 illustrates the same idea with all projections being made onto the xy plane. Tatsuoka (1971) presents similar ideas to illustrate geometrically how the discriminant function operates.

Figure 7 (Crutcher, 1960) shows three constellations for monthly average temperature galaxies where the basic variability of the constellations is different. The covariance within a galaxy is the same where the circles provide a measure of individual cluster variances and the distances between clusters is a measure of cluster variance. A point indicates one station only. Miller (1962) applies the technique to weather prediction. Applications as indicated by Fix and Hodges (1952) ran into the hundreds.

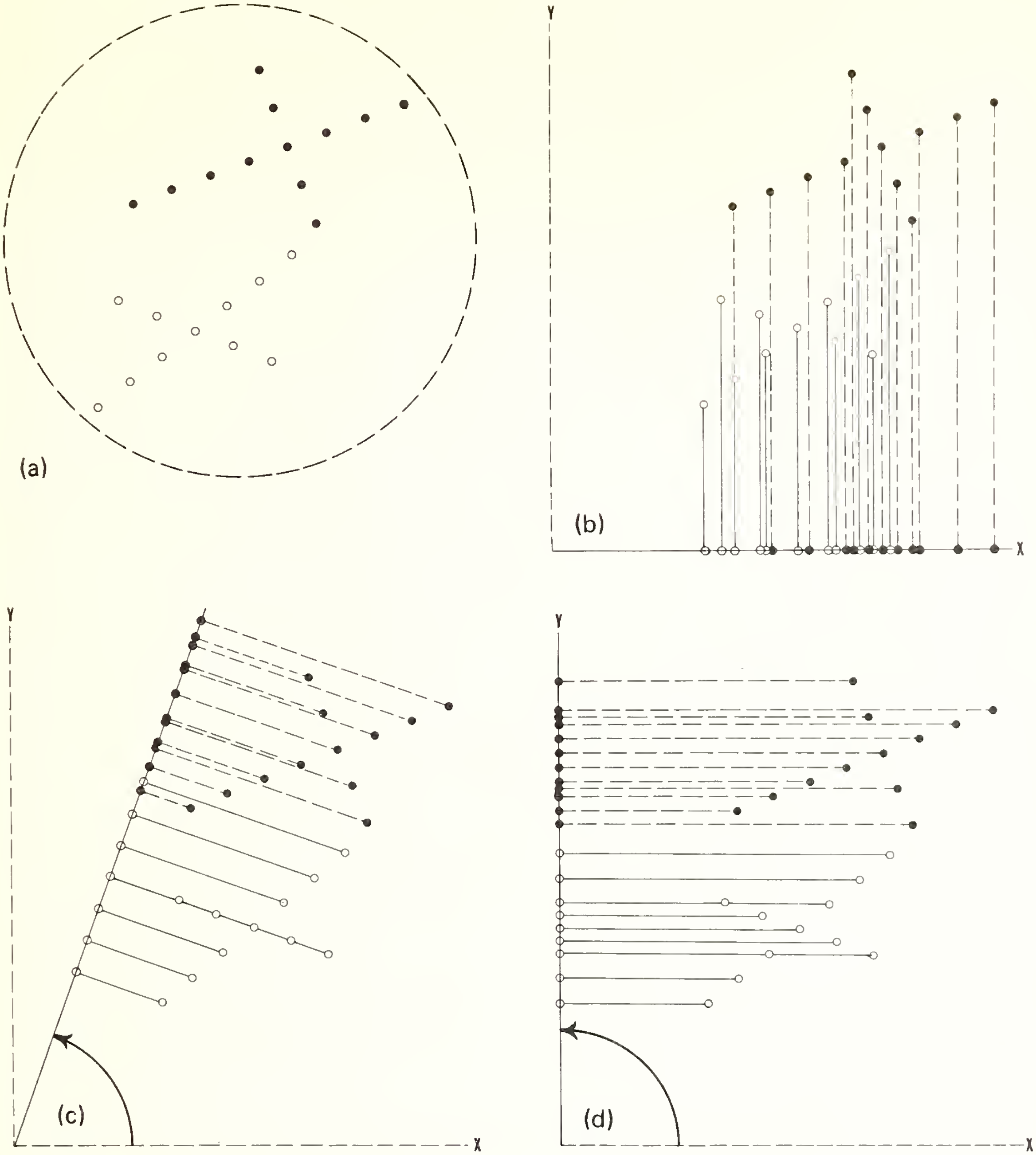


Figure 4 Schematic illustrations of discrimination for two clusters. (a) Two separate and distinct clusters, (b) same two clusters with projections on an arbitrary x-axis, (c) with projections on an angle from the x-axis, and (d) with projections on an axis, y-axis, $\pi/2$ radians from the x-axis.

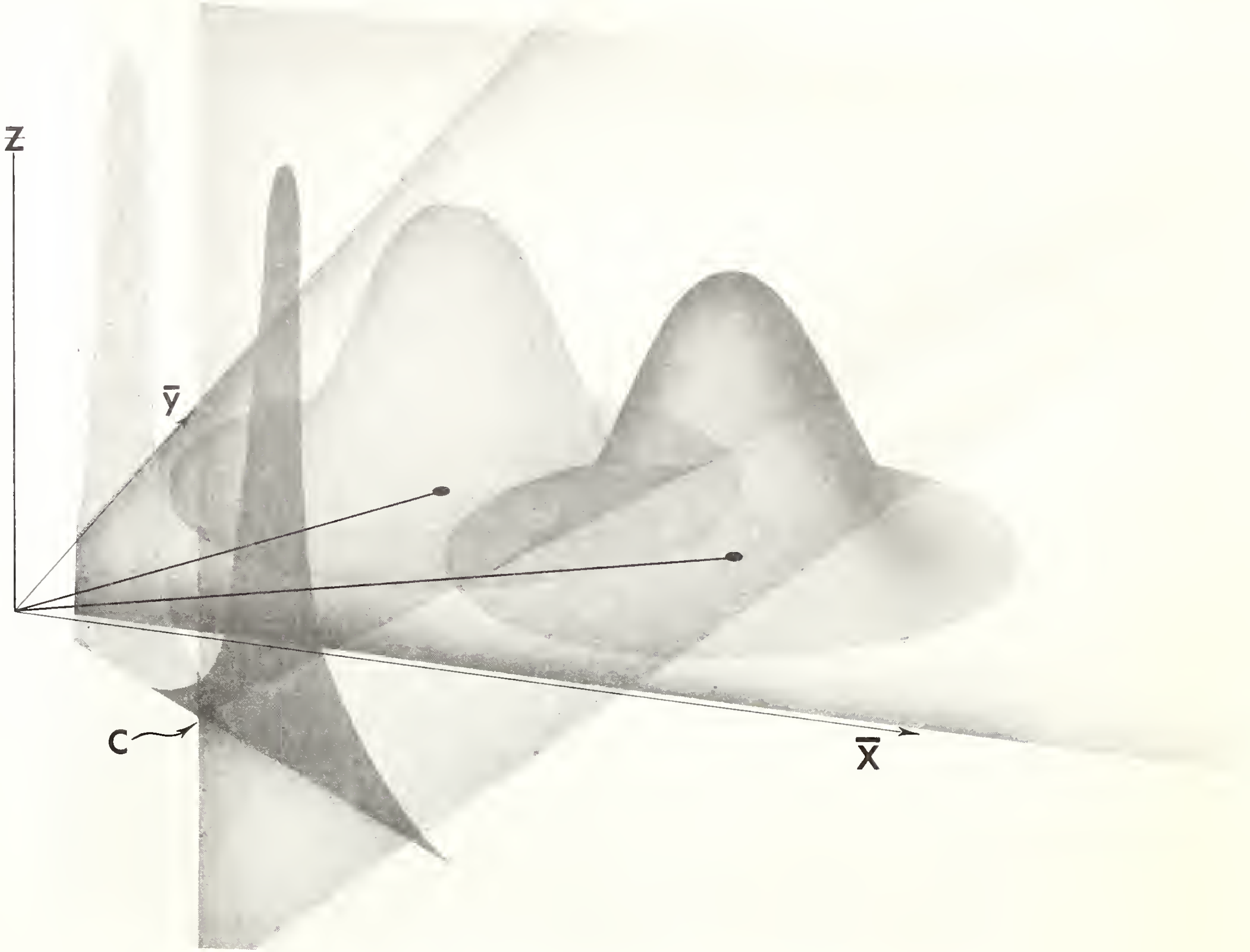


Figure 5 Schematic illustration of discrimination in two-dimensional form with projection onto plane and onto one axis for linear discrimination. Adapted from Crutcher (1960).

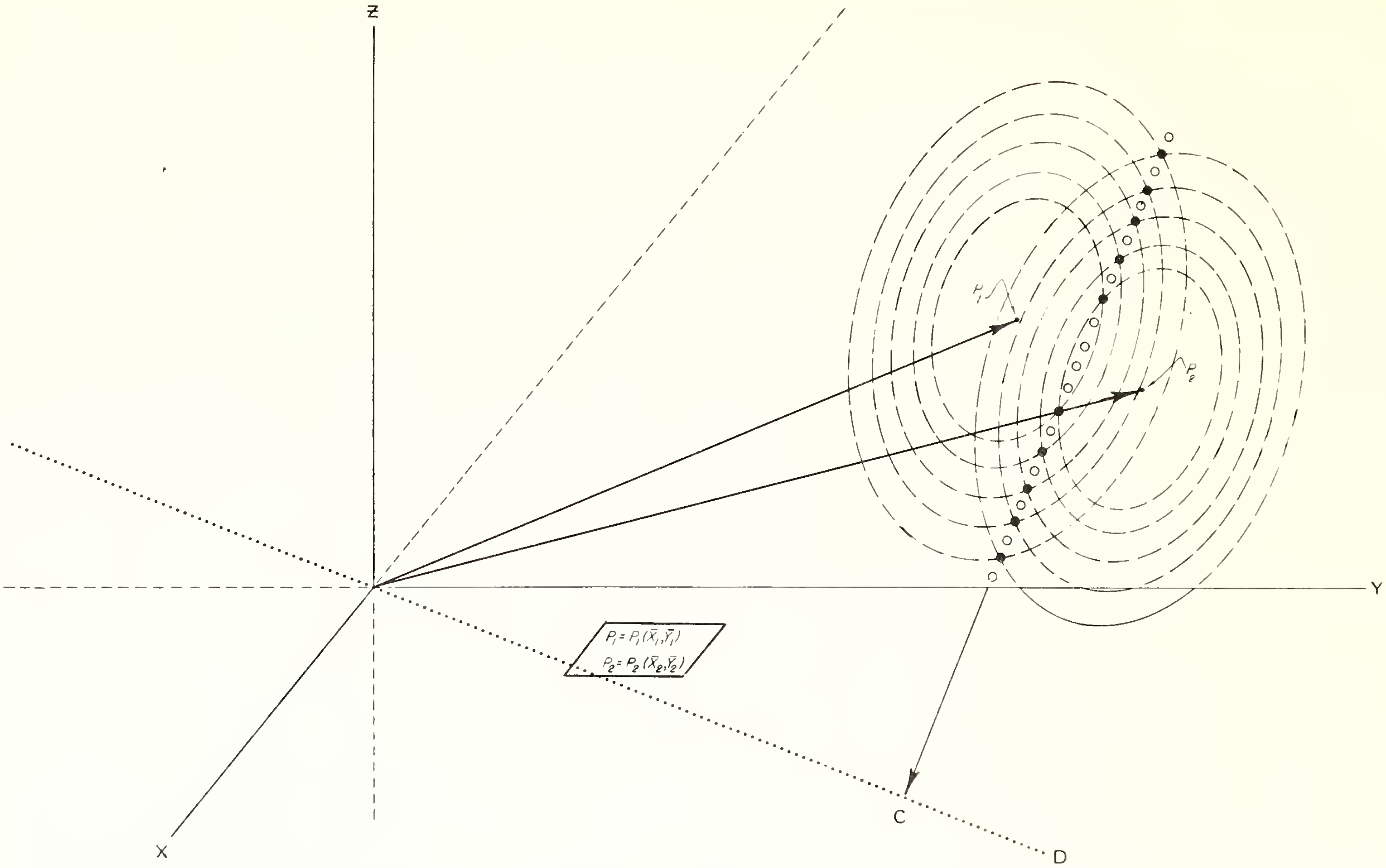


Figure 6 Schematic representation of the discriminant function (dotted) line when the two bivariate populations, P_1 and P_2, have unequal means but equal variances and covariances. If the two-dimensional function is linearized so that the data are projected onto the D axis (dotted), the discriminant criterion, C, is shown at the head of the arrow. All projections to the left of C are assigned to P_1. Those to the right of C are assigned to P_2. Adapted from Crutcher (1960).

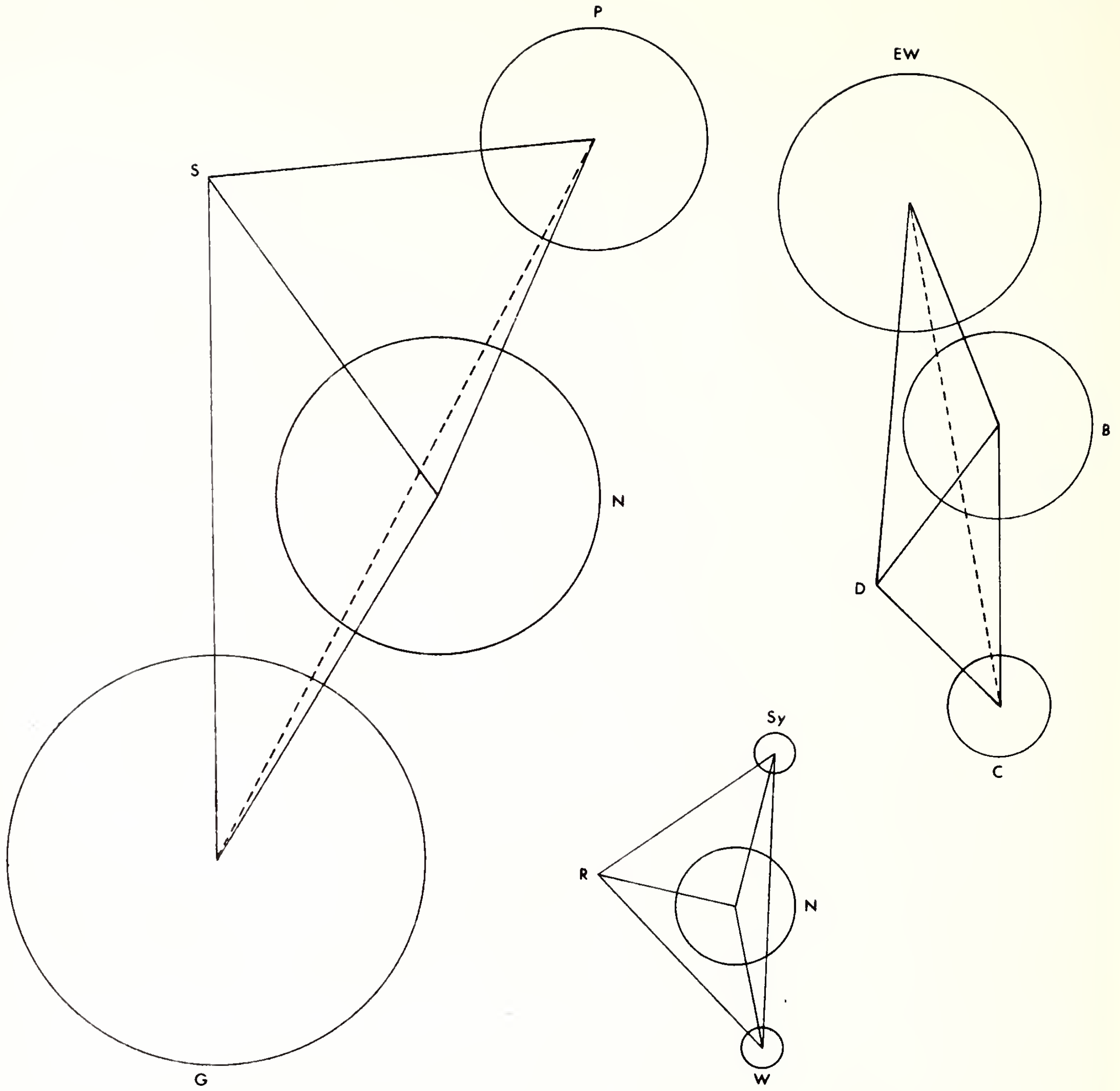


Figure 7 Schematic "D" illustration of constellations of annual temperature climates for North American stations. The constellations and clusters are based on the a_0, a_1, and a_2 harmonic coefficients. Dispersion matrices within constellations are not significantly different but are different between constellations. Crutcher (1960).

The works of Hotelling (1931) and Mahalanobis (1936) are related closely to Fisher's (1936) discriminant criterion. Study of these papers provides excellent background. Hotelling describes these relationships in 1954. Mahalanobis' work in a sense deals with a set of multi-dimensional data for which the ensemble variance is reduced to one, i.e., the total ensemble is standardized. Euclidean measurements then are used.

There are other techniques to categorize and classify data. All of these require some transformation of data, the selection of those measurements which provide the greatest amount of information, and the elimination of those that provide the least if these may be termed unimportant. Usually, their importance is assessed in how much they contribute to the overall variability.

3.2 Factor Analysis

The basic mean structure and variance-covariance structure, i.e., the matrices of means and covariances, are the bases for factor analysis and principal component analysis. It is in the details and interpretation, some complete and some incomplete, of the internal structures that the techniques differ.

The term "factor analysis" comes from factoring procedures and techniques. There are many variants each of which is a special case of the general method of independent dimensional analysis (Tryon, 1959, Tryon and Bailey, 1970). Briefly, as with the discriminant function analysis described previously, the correlation (or covariance) matrix is the initial starting point. The methods used to extract information from this matrix are many and varied. Some are more complex than others. Some use weighting schemes either based on "a priori" knowledge and experience or physical constraints and bases. Some simply attempt to let the matrix itself determine the factors. All this leads to some confusion and some competition among the proponents of the various systems.

Multiple factor analysis with rotation to simple structure is the usual procedure in factor analysis.

The purpose of factor analysis is to explain the matrix of covariances of a multidimensional set by the least number of hypothetical factors. The correlation matrix is used. First of all, the matrix is examined to determine whether it is significantly different from zero (the identity matrix). If so, the technique then identifies, extracts, and weights proportional amounts of the correlation until the residual matrix is not significantly different from zero. Factor analysis stems mainly from the initial work of Spearman (1904, 1926). Essentially, the technique serves to study the similarities in a set of data.

Other pertinent references on factor analysis techniques are Thurstone (1947), Kendall and Smith (1950), Cattell (1952), Bartlett (1953), Fruchter (1954), Harman (1967), Sokal and Sneath (1963), Lawley and Maxwell (1963), Mulaik (1972), and Anderson and Rubin (1956).

3.3 Principal Component Analysis

Karl Pearson (1901) first proposed an empirical method for the reduction of a large body of data so that a maximum of variance could be extracted. Hotelling (1933) developed this method fully as the principle component method. This under some conditions is identical to the discriminant function [Kullback (1968), Anderson (1958), and Girshick (1936)].

For example, principal components according to Anderson (1958) are linear combinations of random or statistical variables which have specified properties in terms of variances. The first principal component is the normalized linear combination with maximum variance. In a swarm of points, each point representing an n-component vector, the distribution may be ellipsoidal. Unless the distribution is spherical, there will be an axis which is as long or longer than any other axis. This axis is called the major or principal axis. A plot of these points will reveal the ellipsoid. It is difficult to represent such an ellipsoid in more than three dimensions. Measures of the components may not directly reveal the ellipsoidal nature of the swarm. However, the variance-covariance or correlation matrix may be rotated in space such that new axes are obtained where the components, projections, or linear functions along these new axes are not correlated. The directions of these axes are known variously as characteristic vectors, latent vectors, direction cosines, or eigenvectors. If there are observations in n-dimensions, each observation as an n-dimensional vector may be projected onto each of the n-orthogonal (mutually uncorrelated) axes obtained above. The variability of these components along each axis may be determined. These variances are known respectively as characteristic roots, latent roots or eigenvalues. The sum of these is the total variance or trace of the matrix. Their product is the determinant. Each one divided by the total provides the proportional amount of variance contributed by the components along each axis. The largest axis is the major axis. The principal axes are those axes whose sum, from the largest in sequence to the next largest, accounts for all or some preselected or specified proportion, say 95 percent. It is the components along each axis that are used. Thus, it is possible to reduce a large dimensional problem to a small dimensional problem. It is also possible to select from the same data a dimensional subset (n-1 or less) to eliminate that (those) dimension(s) which contribute(s) little to the total overall variability.

There are tests to determine which principal components may be considered to be significantly different from the others if indeed they are. Hotelling (1933) and Bartlett (1950, 1951) provide subtests. Mulaik (1972) presents a good discussion of this problem. However, with all this, one must heed the advice of Hotelling (1957) that there may be difficulties involved in neglecting one principal axis even though it is the least in variance. Such an omission may change radically the multiple correlation used in regression. Only adequate testing will reveal whether an axis may be neglected.

In addition, the components along any one axis may be checked for evidence of heterogeneity or mixtures. If multimodality is present on any principal axis, then separation or clustering can be done for these data.

The principal component analysis is distinguished from the factor analysis in that the principal component analysis studies the data structure from the viewpoint of differences rather than similarities.

Any degeneracy which exists in a multivariate distribution is revealed in a principal component analysis [Tatsuoka (1971)]. Therefore, principal component analysis perhaps ought to be considered to be a first stage in factor analysis, though this does not meet the views of the proponents and the opponents of the two techniques.

3.4 Multivariate Statistical Methods

As mentioned previously, the ways in which factor analyses are developed and used lead to newer systems with their respective proponents and opponents. This is so because the data sets must be discussed on the basis of each set. From set to set the bases may be different. Hotelling (1936a & b, 1957) discusses the relations between two sets of variates, simplified calculation of principal components, and then the newer multivariate statistical methods to factor analysis.

Up to the time of Hotelling's (1957) paper, factor analyses of the usual kinds were often inferior to other procedures. Unless the research worker determines and uses (an) invariant statistic(s), the results always will be difficult to assess. However, as in all investigative work of this nature, these analyses may have only heuristic or suggestive value. Hypotheses may be exposed which may be better tested by other methods.

In examination of the use of statistics in the various scientific fields, it is readily apparent that each field develops its own names and "jargon" for statistical terms within their specialized fields. Also, workers trying to reach a common goal will independently develop similar techniques. Most such techniques are beset with "nuisance" parameters induced by an arbitrarily chosen statistic. As Hotelling (1957) points out, it is the invariance of the multiple correlation coefficient as deduced by Fisher (1928) that illustrates the possibility of eliminating the "nuisance" parameters which can only becloud the issue. This puts a premium on the use of invariant statistics.

Student's (1925) "t" distribution is well known for its use to test the significance of means and difference of means and to establish confidence intervals with known probabilities. Its very usefulness in the one dimensional case led to much work towards generalizing this procedure to the multivariate case.

Hotelling (1933) in his work on principal components led to such a generalization in the "T^2" test. The value of "T^2" is invariant under all non-singular linear transformations among the variates. The trace or the sum of the diagonal variances of a covariance matrix is invariant. The "T^2" distribution is a beta distribution or a variance ratio distribution.

The "D^2" stability of Mahalanobis (1936) is closely allied to the "T^2" stability of Hotelling (1931). This also was an attempt to arrive at the use of invariant statistics. Wilks (1932) considered generalization in multivariate analysis. Roy (1939, 1942a & b), Hsu (1938), Bartlett (1950), and Bose and Roy (1938) did considerable work in the field of multivariate analysis.

In the application of principal components in factor analysis, all components must be used. The exclusion of even the smallest component may lead to a complete change in the obtained functions to such an extent that interpretations or decisions may be quite erroneous.

Ridge regression is mentioned as an evolving technique which should receive the attention of some readers [Hoerl (1962), Hoerl and Kennard (1970a,b), and Bannerjee and Carr (1971)]. From the viewpoint of response surfaces, Davies (1956), Draper (1963), and Myers (1971) may be consulted. Though these have their impact on the problems of clustering and classification, they are not discussed further.

3.5 Multivariate Logit

Distribution of quantal responses to drugs or poisons may be better described by distributions other than the normal. The logistic curve is one of these. As Kendall and Stuart (1968) state, "The Probit and Logit transformations of percentages, respectively to normal and logistic distribution deviates, arise mainly in biological contexts and are discussed by Finney (1952)." It has not found wide application in the geophysical field. Anderson (1958) and Dempster (1973) discuss the Logit model. This would be most appropriate if the mixture is composed of a set of variables one or more of which would be considered fixed while the others are normal. This is beyond the scope of the present paper as we here utilize only those distributions which from experience are known to be approximated by the normal distribution. The use of the multivariate Logit distribution mixture separation must be deferred to later research. Hopefully, this will be within the next five years.

4. CLUSTERING

Grouping of individuals with similarities and separation of individuals with dissimilarities is a continual process. This process is evident from the smallest to the largest features in any ensemble whether it be in the universe in the development and dissolution of galaxies, in the earth in its geologic processes, or in life on earth in whatever form it may be. It is evident in the abstract world, too. Readers of Aristotle and the followers of Hippocrates and Linnaeus will recognize attempts to order chaos in the sense of clustering and differentiation, whether it be the assignment of some attribute or the recognition of some measurable quality or quantity. Much has been done and much has been written. The techniques of clustering and differentiation are as varied as those who attempt to perform those functions. The logic and the metrics may vary but the insistent theme is to group those that are alike and separate those that are unlike. Each individual is, of course, an entity in itself. A set of measurements on that individual undoubtedly constitutes a group in itself. This much is acknowledged by any worker. The job is to group those individuals together whose sets of measurements are not too far different and to put aside those whose measurements are different at some level of perception and thinking to another group or groups. Much sometimes depends on the individual whose measurements are used as the starting point. In a sense, those measurements of the first individual are in an "a priori" sense weighed most heavily in the clustering process. Coalescence begins on this individual.

Clustering is simply another term among the many in the general field of taxonomy. Taxonomy is the scientific ordering and classification of information. Taxonomic systems must be simplified representations of group characteristics and all their interrelationships. Because these are most general and therefore quite usable, perfection cannot be attained except in the case of an individual. The main function of any such system is to reduce the complex problem to a simpler problem, thus reducing the requirements of memory. References, some of which have been given previously, are Sokal and Sneath (1963), Tryon and Bailey (1970), Duda and Hart (1973), Fisher (1936), and Anderson (1958). Hartigan (1975) offers a considerable number of clustering algorithms.

Many terms are used in the grouping concept. Some are more suggestive of the process than others. These terms depend on the field of endeavor and on the basic background of the investigators. Some of the techniques are the same though the words are different. Let us look at a few of them: clusters, clumps, coalescence, condensations, aggregations, agglomerations, divisions, similarities, cohesions, linkages, hierarchies, communalities, types, and affinities. The classes obtained may be given names in the specialized fields. In the biological sciences these are the families, genera, species, subspecies, etc. In the field of the natural sciences, though not restricted to these fields, these may be clusters, universes, constellations, galaxies, etc. In the latter example, the order may be interchanged depending on the viewpoint of the problem.

In any taxonomic problem dealing with measurements, there must be some way to collect the like individuals together and to isolate, reject, and form new groups with those which are most alike among the unlike. Some measures of likeness or similarity must be developed and accepted. As individuals are collected into small groups and as the smaller groups are collected into larger and larger ensembles, a tracing of the procedures may be kept. As diagrams or the concepts of the trace resemble some feature of the known world, these may be called tree-diagrams (branch diagrams), root diagrams (dendrodiagrams or dendritic diagrams), link diagrams, or hierarchical diagrams. These may be shown in two or three dimensions or left in abstract form.

Baker and Hubert (1975) consider procedures to measure the power of hierarchical cluster analysis. This important feature in studying the various techniques and their alternatives is not examined here in further detail.

The clustering procedures are allied closely to the previously discussed techniques of factor analysis, principal component analysis, and discriminant function analysis. Canonical analysis also may be used. The last is simply the techniques which maximize the correlations among linear functions of the data.

The specific clustering techniques used here are discussed in more detail in Wolfe's (1971b) NORMIX program, section 7.

5. TRANSFORMATIONS

The multivariate normal distribution is only one of an infinite number of useful multivariate distributions. The univariate normal distribution has been exploited more than any other because the necessary statistical tools have been developed. It is nearly so with the multivariate normal. If a distribution is multivariate normal, then any of the subspace marginal distributions are normal. Though the normality of all marginal distributions does not guarantee multivariate normality, such normality is a necessary condition for multivariate normality.

If prior experience indicates that a certain measure is usually normally distributed and tests imply non-normality, then one inference that can be made is that the data set is mixed. It is precisely the thrust of this report to separate such mixed distributions into their homogeneous parts. If any marginal distribution is mixed, then certainly any higher order is mixed or any lower order variate distribution is or may be mixed. If the set of projections onto any principal axis is mixed (multimodal) and the set of projections is a linear function of the multivariate set, then this set can be used to establish estimates of the parameters of the various homogeneous collectives.

With normality of distribution, the usual tests of significance or tests of hypothesis may be made. Otherwise, decisions based on such tests may be invalid or questionable. The robustness encountered, though, usually is such that a lack of normality does not impair the decisions very much. Far more important perhaps is the practical significance of the decision.

If it is known that an unmixed marginal distribution is not normal, then some transformation towards normality of that distribution should be sought. A marginal distribution is a subset of the higher order distribution. If the marginal distribution itself is not normal in the multivariate sense, then the still lower order marginals or smaller subsets should be checked. It may be that only one of the lowest order marginals (a one dimensional distribution) is the non-normal. If this is the case, then transformation of this distribution should be sought.

The above does not imply that it is impossible to use the non-normal distributions. It does imply that the distributions must be normally distributed or mixed normal in their distribution if the techniques discussed in this paper can be used to provide valid results.

There may be some cases where appropriate normality cannot be achieved through any transformation (Graybill, 1961, pg. 318).

For those distributions that can be normalized, the transformation process is often carried forward in graphical procedures. Boehm (1974) uses this procedure and has an electronic computer program to provide such transnormalization.

The methods of factor analysis, principal components, and discriminant functions are techniques to linearize a multidimensional situation to a single univariate situation. Hopefully the final distribution or distributions in the components, factors, or clusters will allow linearization within these to the univariate problem.

There are many tests for univariate normality. Some of these (Graybill, 1961) are (1) likelihood ratio tests associated with transformation toward normality, (2) skewness and kurtosis tests, (3) omnibus tests, and (4) normal probability plots.

A useful transformation to improve normality is that proposed by Box and Tidwell (1962) for estimating a shifted power transformation $(\chi + \xi)^{\gamma}$ of a single variable. This is a generalization of many tests developed through the years [Tukey (1957) and Moore (1957)]. Simultaneous transformation of a multiple set of variables has been the subject of much research. Andrews et al. (1971) discuss the problem in more detail for the bivariate distribution. Extensions to the multivariate case are indicated.

Transformations are not discussed further here as the climatological examples used in this report do not require transformation to normality. The clustering technique used here will be applied to other climatological and geophysical problems. Many of the data distributions of the future study will require transformation to normality. The two techniques described above then will be treated in detail.

6. SEPARATION OF MIXTURES

6.1 Univariate Mixtures

Karl Pearson (1894), in his paper on "Contributions to the Mathematical Theory of Evolution," gives the procedure to mathematically dissect a mixture of two univariate normal distributions. In the general case the five parameters to be estimated are two means, two variances, and a proportionality factor. It is necessary to find a particular solution of a ninth degree polynomial equation, a nonic.

The following list is neither exhaustive nor the most important but it does provide sufficient references for the reader to pursue the subject further: Charlier (1906), Charlier and Wicksell (1924), Burrau (1934), Strömgren (1934), Essenwanger (1954), Schneider-Carius and Essenwanger (1955), Cohen (1965), and Cohen and Falls (1967). Essenwanger's work is applied to meteorology while Cohen's and Fall's work provides electronic computer routines for dissection of heterogeneous mixtures.

Hald (1952) discusses the subject of heterogeneity in the univariate case. Graphical as well as analytical techniques are discussed. The general case may be written as

$$f(x) = \sum_{i=1}^{m} \alpha_i \, f_i(x); \quad \Sigma\alpha_i = 1 \, ,$$

where

$$f_i(x) = (2\pi\sigma_{xi}^2)^{-\frac{1}{2}} \exp\{-[(X-\mu_{xi})/\sigma_{xi}]^2/2\} \, .$$

Even for a mixture of two distributions the solution of the nonic is a big task. Cohen and Falls (1967), using procedures developed by Charlier and Wicksell (1924) and a modern electronic computer, provide procedures to effect the requisite dissection and to obtain estimates of the parameters. Discriminant criteria can then be developed to enable classification of a new measure to one of the two groups or to classify each datum from the original data set.

The computer program, kindly lent to the authors by Cohen and Falls, has been used to dissect a mixed distribution in Section 9.

6.2 Bivariate Mixtures

Hartley (1959) provides techniques to separate two mixed bivariate normal wind distributions. Crutcher and Clutter (1962) apply these to wind distributions but encounter difficulty when the covariance matrices are singular or near singular or when the variances or covariance matrices are assumed to be different.

In the bivariate case of a unimodal distribution there are five parameters to be estimated. These are the two means, the two variances, and the correlation. When there is a mixture of two bivariate normal distributions, there are eleven parameters to be estimated. These are the two means in each group, the two variances in each group, the correlation in each group, and the mixture parameter. If two assumptions are made, (1) the variances are all equal one to each of the other three and (2) the correlation is zero, then the simplest mixture of two circular normal distributions is obtained. If the component variances within each are equal, but unequal from one group to the other, the more complicated mixture of two circular normal distributions with different variances is obtained.

If there are "a priori" reasons to suspect that there are two groups and if the assumption of circularity provides meaningful and useful results, then it is suggested that Hartley's procedures be used. Computing time will be much less than in the clustering program discussed here.

The authors know of no other computer program available specifically to take a set of mixed bivariate normal data and dissect it even for the simplest case of two subsets with unequal means but equal variances, component and vector, and zero correlation. The cluster program used here will do the above as well as the general case. But this clustering program is far from an optimum one in terms of least time and cost. For the one or two time case, the clustering program could be used. However, for repetitive processing of many many sets of data it would be advisable to develop an optimized specific program based on Hartley's procedures. This, the authors plan to do.

The mathematics may be written for the simplest bimodal bivariate circular case as

$$f(x,y) = \sum_{i=1}^{2} \alpha_i f_i(x,y); \quad \sum_{i=1}^{2} \alpha_i = 1 ,$$

where

$$f_i(x,y) = (2\pi\sigma_i^2)^{-1} \exp\{-[(X-\mu_{xi})^2/\sigma_{xi}^2 + (Y-\mu_{yi})^2/\sigma_{yi}^2]2^{-1}\}$$

and the α_i are the proportionality factors. In the slightly more general case,

$$f_i(x,y) = [2\pi\sigma_x\sigma_y(1-\rho_{xy}^2)]^{-\frac{1}{2}} \exp\{-[((X-\mu_{xi})^2/\sigma_{xi}^2)$$

$$-2\rho_{xy}((X-\mu_{xi})/\sigma_{xi})((Y-\mu_{yi})/\sigma_{yi})$$

$$+((Y-\mu_{yi})^2/\sigma_{yi}^2)] [(1-\rho_{xy}^2)^{-1}] [2^{-1}]\} .$$

6.3 Multivariate Mixtures

The general multivariate-multimodal case may be written as

$$f(x,y,\ldots) = \sum_{i=1}^{n} \alpha_i f_i(x,y,\ldots); \quad \Sigma\alpha_i = 1 ,$$

where

$$f_i(x,y,\ldots) = [(2\pi\sigma_{xi},\sigma_{yi},\ldots)|R|]^{-\frac{1}{2}} \exp-(\psi^2/2) ,$$

α is a proportionality factor, R is the correlation matrix, $|R|$ is the determinant of the correlation matrix, $R_{i,i}$ is the respective cofactor, and

$$\psi^2 = |R|^{-1} [R_{xi,\, xi}((X-\mu_{xi})/\sigma_{xi})^2 + R_{yi,\, yi}((Y-\mu_{yi})/\sigma_{yi})^2 + \ldots.$$

$$-2R_{xi,yi}((X-\mu_{xi})/\sigma_{xi})((Y-\mu_{yi})/\sigma_{yi}) - \ldots] .$$

Sample estimates such as $\overline{X}$, $\overline{Y}$,..., s_{xi}, s_{yi},..., r_{xy},..., replace the population parameters μ_{xi}, μ_{yi},..., σ_{xi}, σ_{yi},..., ρ_{xy},..., where the bar represents an averaging process.

7. WOLFE - NORMIX 360 COMPUTER PROGRAM

Culminating over a decade of work, Wolfe (1971b) published his NORMIX computer program. Earlier versions and other pertinent papers by Wolfe appeared in 1965, 1967 a and b, 1968, 1970, and 1971a. He provides background discussion, the necessary programs, and an example. The example chosen is the old standby example of Fisher (1936) which so many investigators use.

7.1 Maximum Likelihood Estimation

Unless "a priori" conditions indicate otherwise, the first hypothesis made is that there is a specified number of types and the probability of a datum being assigned to any one of the types is the same as to any other type. With two groups, the probability of being assigned to one of the two types is 0.50. With four groups, the probability of being assigned to one of the four types is 0.250. With no "a priori" indication, a default option of the program then provides equal mixing proportions as a first guess.

As shown by Wolfe (1970), the maximum likelihood estimates of the mixing proportion ($\hat{\lambda}_s$), mean ($\hat{\mu}_s$), and covariance ($\hat{\sigma}_s$) of the cluster (type s) in a mixture satisfy the following conditions:

$$\hat{\lambda}_s = (1/n) \sum_{k=1}^{n} \hat{P}(S|X_k)$$

$$\hat{\mu}_s = (1/(n\hat{\lambda}_s)) \sum_{k=1}^{n} X_k \hat{P}(S|X_k)$$

$$\hat{\sigma}_s = (1/(n\hat{\lambda}_s)) \sum_{k=1}^{n} (X_k-\hat{\mu}_s)(X_k-\hat{\mu}_s)' \hat{P}(S|X_k)$$

where the prime indicates a transpose, X_k is the vector of observations for the kth point in the sample, and $\hat{P}(S|X_k)$ is the probability of membership of X_k in clusters and is equal to $\hat{\lambda}_s$ times the ratio of the normal density of type s at X_k to the density of the mixture. In the special case where the clusters have a common covariance matrix, $\hat{\sigma}_s$ is $\hat{\sigma}$ and may be written as

$$\hat{\sigma} = (1/(n\hat{\lambda}_s)) \sum_{k=1}^{n} X_k X_k' - \hat{\mu}_s \hat{\mu}_s'.$$

As the unknown parameters appear on both sides of the equation, it is necessary to use an iterative process to solve the equations. For this reason, if "a priori" estimates are used the iterative processing is diminished. Computing costs therefore are much lower. Several sets of values may satisfy the equations, and the results may depend on the starting values for the iteration process. As indicated previously, a change in the order of the input data will usually change the initial estimates obtained.

Local or relative maxima or saddle points may be obtained. The procedure does not guarantee to find an absolute maximum. The initial estimates are simply modified and improved upon. Bad initial estimates or diverging iterations can cause strange estimates in the parameters such as negative variances or singular correlation matrices. When this happens, re-initialization or change of initial estimates may resolve the problem. The program also attempts to resolve the problem of singular matrices by adding a small normal random number to the diagonal terms.

7.2 Initial Estimation

It is assumed that the problem involves clustering within a multimodal multivariate data set. All measurements are standardized. That is, estimates of the individual component means and variances are computed. The squares of the deviations from the means divided by the variances produce the square of a standardized deviate. Thus, the set is reduced to a multimodal multivariate set with a zero mean and a variance of one. This is the basic Mahalanobis distance technique.

Within the above framework, if initial estimates of the cluster means and covariances are not provided, the program generates initial estimates in a KMEAN subroutine which was adapted from Ward, Hall, and Buchhorn's (1967) hierarchical grouping for minimum variance. The variance which is minimized is the sum of Mahalanobis distances between points within a cluster given by

$$d_{ij}^{2} = (X_i - X_j)' \, C_1^{-1} \, (X_i - X_j) \ ,$$

where X_i and X_j are points in the same cluster and C_1 is the covariance matrix. The prime indicates a transpose and the superscript indicates an inverse.

The subroutine first transforms the data to principal axis factor scores,

$$Y = \Lambda^{-\frac{1}{2}} F'X \ ,$$

where F is the eigenvector matrix of C_1 and Λ is the diagonal matrix of eigenvectors. (See discussion on principal component analysis.) Then

$$d_{ij}^{2} = (Y_i - Y_j)' \, (Y_i - Y_j)$$

which is easier to compute.

After hierarchical grouping, the within-group covariance matrix C_2 for ten clusters is determined. The number ten is arbitrary and could be changed. With this modified matrix, hierarchical grouping is performed again. A third covariance matrix C_3 within the ten clusters is obtained. New distances are obtained which are used for the third and final hierarchical grouping.

The program permits the assignment of an arbitrary number of clusters or means known as KMEANS. If the number setup is 150 or less, the procedure uses the number established. If the number of input data is greater than 150, a leading subroutine for the KMEANS (MacQueen, 1967) precedes the hierarchical grouping. The first 150 input data points form the centroids of 150 clusters. The two closest points (clusters) are merged into a two-point cluster with a cluster mean in the location of the first of the two points read. The 151st datum is moved into the place vacated by the second of the two points. The next two closest cluster centroids are merged and the process continues until all data points have been collected into 150 clusters which then are grouped hierarchically.

7.3 Significance Tests for Number of Clusters

The user has some idea that clusters of data exist in the data set or he would not be concerned with this report or its uses in his field of endeavor. Suppositions or hypotheses concerning the number of types in the sample are listed and control is placed on each of these in the program. For example, he suspects two groups at least or perhaps four at most. Just to make sure, as far as this technique is concerned, he might place an upper limit of six. The program reaches a solution in each case which provides a relative (local) maximum to the likelihood function. The ratio of the likelihoods for any two of the hypotheses above 2, 4, or 6 provides a basis for a significance test for rejecting the null hypothesis that there is no difference between the two hypotheses, i.e., the smaller number of types against the alternative of a larger number of types.

7.4 Strategy of Use

As computing time increases with number of variates, number of input data, and null hypotheses to be checked, care should be exercised to keep time, which is transformable to cost, to a minimum. There are certain strategies for use which Wolfe (1971b) discusses.

7.5 Usage

It is appropriate that, since the program can be obtained from Wolfe (1971b), a few usage comments from his program be placed here (with his permission).

7.5.1 Storage Requirements

Storage requirements are variable, depending on the data. The arrays are dimensional at execution time. A minimal problem would require 120,000 bytes of storage (IBM 360). The first two lines of the printout give the storage required for a particular problem.

7.5.2 Restrictions

The number of types or clusters must not exceed 20. There are no fixed limits on the number of variables or sample size. However, all data and arrays must fit into core.

7.5.3 Error Messages

Bad initial estimates or diverging iterations can cause strange estimates of the parameters, such as negative variances or singular correlation matrices which result in system diagnostics. In such a case, the user may have to specify his own initial estimates on the input form or rescale the data to eliminate dichotomous variables.

7.5.4 Input Deck

Input Deck set up cards are:

Input Form, Cards 01-11
Data Cards (omit if data are on tape)
Input Form, Card 12
Initial estimates, if any, for one hypothesis
Input Form, Card 12
Initial estimates, if any, for a greater number of types
...
...
Blank Card

For example:

Card 1 provides the user and the data used.
Card 2 provides the title.
Card 3 provides room for comments.
Card 4 provides room for comments.
Card 5 provides for the number of variables, the sample size, and input tape.
Card 6 provides hypotheses for number of clusters.
Card 7 provides space for assumption of same or different covariance matrices and the minimum number to be accepted into a cluster.
Card 8 provides for the significance level for rejection of the null hypothesis.
Card 9 provides for the entry of the number of KMEANS to be first examined and whether the iterations are to be printed.
Card 10 provides for the establishment of a time limit and a limit on the number of iterations.
Card 11 provides for the data format.
Card 12 provides for the initial estimates, if any, for the number of types, the means, the standard deviations, and the correlations.

7.5.5 Validation Examples

Wolfe (1967, 1971) uses two examples to test his program. The first is the Fisher (1936) treatment of Anderson's (1935) measurements on Irises. Nearly all discrimination papers or clustering papers refer to this classic paper. The second example is an artificial set. These and another set of data on azotobacter, the nitrogen fixing bacteria [Cox and Martin (1937)], were used to validate the Wolfe program adapted for use at the National Climatic Center.

The NORMIX program has been adapted for use at the National Climatic Center. Considerable difficulty was experienced in adapting the program for use on the Univac Spectra 70/45 due to language configurations. One adaptation is an output routine to furnish a sequential leading copy of the input data. This permits ready referencing as to a datum and its assignment to a cluster. Some modification of the output sequences were also made.

Though this point is made by Wolfe, it is given here again for emphasis. The techniques apply to only normally or near normally distributed variates. Any significant departure from normality will invalidate the probability decision levels. Therefore, if data distributions cannot be approximated by the normal distribution, transformations to normality or near normality of distribution must be made before input to the program.

A brief summary by Wolfe (1971b) is quoted below:

> "The problem was to develop a computer program for cluster analysis and unsupervised pattern recognition of types with different cluster shape....
>
> "The program will cluster a sample of thousands of objects measured on many continuous variables....
>
> "The approach seeks maximum likelihood estimates of the parameters of multivariate distributions. The likelihood equations are solved iteratively by continually re-estimating the probability membership of each sample point in each cluster until the likelihood reaches a relative maximum. The initial estimates are derived from a minimum variance hierarchical grouping subroutine, which itself is iterative in seeking an appropriate distance function. The program prints out the means, standard deviations, and intercorrelations of the variables within clusters and the proportions of the population for each cluster. The probabilities of membership for each cluster are also printed."

Briefly some points mentioned in the foregoing section are iterated. In clustering procedures, unless there is "a priori" knowledge as to the characteristics of the clusters in the data set, the data themselves determine the clustering. This is called the unlearned procedure or learning process. The use of "a priori" estimates is termed the learned procedure. There is, then, the necessary decision as to just where to start. This is true of any discriminating procedure whether it be any of the allied techniques of analysis such as discriminant function, factor, principal component, or any other analysis.

Any change in the order of the input data will produce a difference in the output cluster characteristics (MacQueen, 1967, p. 290). Therefore, one run only provides an estimate of the clusters. A number of runs with a different ordering of the data will produce different results. Hopefully, these will not be too different and usually are not too different even though their covariance matrices may be unequal. The more distinctly different the clusters are, the more the various runs will provide the same estimates. This is not unexpected. There will be some data which will not be assigned to the same cluster each time.

8. EXAMPLES

8.1 Introduction

The techniques used here are applicable in any field as they reduce all problems to non-dimensional ones in the sense of units. Here, climatological (weather) observations are used. These were obtained from the archives of the National Climatic Center (NCC) in Asheville, N.C.

We present six data sets from among the many that could have been used. The first four present situations which are recognized as producing mixed distributions. Obviously, these are selected to demonstrate the usefulness of this procedure.

Output of this program for the six data sets used for this paper is not presented in total. The data sets are:

1. Land-sea breeze data
2. Tropical stratospheric wind data
3. Mid-latitude tropospheric wind data
4. Mountain pass wind data
5. Marine surface data
6. Radiosonde and rawinsonde data

In the first and second data sets, only selected tabular output data and computer-drawn, 0.50 probability ellipses are shown. In the third data set, a total input-output of the Wolfe (1971b) program adapted for use at the NCC is shown. Except for the input data and the modified eigenvalue-eigenvector output, the format is essentially that of the Wolfe program. Also, a considerable number of 0.50 probability ellipses are shown. The computer plot routines for these exist in programs written for the CALCOMP drum plotter and Computer Output Microfilm and the Hewlett-Packard desk-top plotter at the National Climatic Center. The computer print plot of the discriminant function for the common covariance assumption is part of the Wolfe program. The last three data sets simply illustrate selected and re-arranged output as well as the 0.25 probability ellipses.

Winds are used in all six data sets as upper winds usually are distributed in the bivariate normal sense. In the second, fifth, and sixth data sets, additional weather elements are used. Where mixtures are evident, the usual unimodal distribution cannot be used. These mixtures occur in the planetary boundary layer, in the trade-wind inversion region, in the tropopause, or in the monsoon change region. It is precisely the advisability of separating these and other mixtures that prompted the investigation of the problem and the publication of these results.

8.2 Brief Descriptions of Data and Locations

8.2.1 Land-Sea Breeze Data

San Juan, Puerto Rico, U.S.A., is a seaport located on the northern shores of the Island in its eastern portion. It is a sub-tropical station at 18°26' north latitude and 66°00' west longitude. The data are taken from the observations made at the airport at Isle Verde. Its elevation is 20 m. This location was selected because it is a logical one for a land-sea breeze effect. The hours selected, 0600-0800 and 1200-1400, are in local standard time. The month of October was selected though any other period could have been used. One hundred observations for each hour group are used for October 1955 and the first portion of the first 3 days of October 1956. The early morning hours should show the land breeze or a balanced condition. The 1200-1400 observation hour should begin to show the effects of the sea breeze though perhaps not as strongly as later hours. The hours of 0900-1100 were not used so as to eliminate some of the overlapping of the land-sea breeze effect if any existed. This would provide a clearer separation for the technique demonstration.

8.2.2 Tropical Stratospheric Wind Data

Canton Island is in the Southern Hemisphere at latitude 2°46' and longitude 171°43' west, elevation 4 m. This location was selected because of its known, distinct quasi-biennial wind oscillation in the stratosphere. During the preparation of Technical Paper 34 (Crutcher, 1958) and U.S. NAVAER 50-1C-535 (Crutcher, 1959), the apparent biennial oscillation of the tropical tropospheric winds had been noted. U.S. NAVAER 50-1C-535 is apparently the first atlas in chart form to use the elliptical bivariate normal distribution to describe the distribution of upper winds, particularly those of the stratosphere. This followed the very important atlases of the British groups [Brooks and Carruthers (1950)] using the circular bivariate distribution and preceded their update, [Heastie and Stephenson (1960) and Tucker (1960)]. The British groups assume the circular distribution to adequately describe an upper wind distribution. This appears to be a good assumption as a first approximation if the winds are not mixed. The extension (Crutcher, 1957) of this assumption to that of the elliptical normal as a second approximation increased the representativeness of that bivariate normal from about 70 percent of the cases to about 90 percent in the troposphere. It was noted during the preparation of NAVAER 50-1C-535 that, in the tropics at the higher altitudes (lower pressure levels), the ratio of the major axis to the minor axis was seemingly extraordinarily large, near four ranging up to ten at times. An examination of the distributions generally revealed two clusters with the easterly winds being stronger and with a higher constancy than those from the west. In the meantime other investigators also were studying the problems. McCreary (1959) reported that in the stratosphere during October 1956 - July 1957 at Christmas Island, westerly winds overlay the stratospheric easterlies. This was a reversal of what was considered to be the usual pictures of easterlies over westerlies. Graystone (1959) then reported on a year to year reversal of the equatorial stratospheric winds. In the following year Reed (1960), Ebdon (1960), and Veryard (1960) provided the impetus for a great amount of research into the phenomenon now known as the quasi-biennial oscillation (QBO) of the winds in the tropical stratosphere. These same authors soon produced further work in the field [Reed et al., (1961) and Reed and Rogers (1962) in the U.S. and Veryard and Ebdon (1961a, b, and 1963) in the U.K.]. Later, an oscillation in the temperatures associated with the winds was soon reported as were oscillations in the tropopause heights, ozone, and other phenomena. Newell et al., (1974) discuss the development of research in this field in the second volume of an important work on the general circulation of the tropical atmosphere and interactions with extratropical latitudes.

In these very important studies, the oscillation was shown to start at the higher altitude of the 10-mb level and propagate downward through the stratosphere to the 100-mb level.

Clayton (1885), Berlage (1956), and Landsberg et al., (1963) had noted this feature in surface phenomena. This feature now occupies a rather solid part in the literature of the atmospheric circulation.

Upper winds during the months of July and January for the years 1954-1964 at Canton Island at the 50-, 30-, 20-, and 10-mb pressure levels were used. For the period 1957-1967, heights and temperatures were added. Data from this station have been used many times in studies. The quasi-biennial oscillation (QBO) is barely evident at 100 mb and becomes strongest at about 30 to 25 mb. Apparently it weakens at lower pressure levels (higher altitudes), such as 10 mb, though this effect may be due to loss of observations. There is a time dependence as noted by Reed et al., (1961), Reed and Rogers (1962), Veryard and Ebdon (1961a, b), Angell and Korshover (1962), and later works.

The purpose of this paper is to provide the techniques to dissect such distributions. Dynamic or synoptic considerations are not made here except in the sense that these do effect the QBO. Newell et al., (1974) discuss the dynamic properties of the phenomenon. Dynamic considerations do create the separate distributions insofar as the calendar year dating system is involved. The results may be used to assess the extent of these effects. Only the horizontal winds and temperatures are used at the individual levels. An analytic solution or more detailed examination of the atmospheric circulation and weather is not attempted here. For example, the problem could have been made much more multidimensional by including all stratospheric levels and concomitant surface weather. The results of this section clearly demonstrate the usefulness of this separation technique.

8.2.3 Mid-Latitude Tropospheric Wind Data

A continental U.S. station was needed. Rantoul, Illinois, was selected since, in an earlier unpublished work, Crutcher and Clutter (1962) experienced some difficulty in applying the techniques of Hartley (1959). The difficulty involved singularity problems in the data when assumptions of unequal variance in the groups were made. Its latitude is 40°18' north, longitude is 88°09' west, and elevation is 227 m. There are some data sets which produce such near singular matrices that the techniques cannot be used in their present form. This is occasionally true of the present technique here. Ordinarily, the singularity problem is resolved in the technique by the addition of a random small amount to the matrix diagonal elements. Even then, negative variances may be encountered. Re-initialization or change in order of data entry may resolve the problem.

The period of record of data used is the month of October for the years 1950-1955. The upper wind data used are those at the 700-, 500-, and 300-mb levels. Both input and output data are shown to better illustrate the procedures. The first two data sets described were two-dimensional. This data set example illustrates a distribution in six dimensions, two for each of three levels. For this purpose there had to be a simultaneous wind observation at each of the three levels. Therefore, each input data vector given was composed of the vector formed by the zonal and meridional component of the wind at each level, i.e., a vector composed of six components. The two-dimensional illustration for a given level then is comparable to that of another level for each datum at the given level having a matching datum at the other level.

As more than three space is difficult to illustrate except for linear reduction to three space or less, output for the six-dimensional forms is given in eigenvector and covariance (correlation matrix) form.

No attempt is made here to correlate the wind data with current or later weather. This can be done by simply extending the six-dimensional vectors to a greater number of dimensions by including the desired surface weather data.

8.2.4 Mountain Pass Wind Data

Stampede Pass, Easton, Washington, U.S.A., was selected to demonstrate the existence of two or more distinct clusters of wind, and of wind with temperature regimes, in a constricted geographic location. Its latitude is 47°17' north, its longitude is 121°20' west, and its elevation is 1206 m. It is almost in the lowest part of a saddle-back with a ridge running north-south upwards from the station location to a height of about 200 m above the station at a distance of about 3 km. North of the saddle, the ridges rise in an east-west fashion to heights near 3,000 m at distances ranging from 20 to 80 km.

8.2.5 Marine Surface Data

Weather observations over oceanic areas constitute an important part of the archives at the National Climatic Center. These have been the basis for the Marine Climatic Atlases program of the U.S. Navy since 1950. In some regions the weather is composed of quite distinct weather regimes over the year. Within shorter time periods there may or may not be such distinctness. There may or may not be distinct separation of weather factors between the traveling cyclones and anti-cyclones.

Observations from transient ships offer no real opportunity for time studies. For a particular spot, over time, this may or may not be important. For reasons of time continuity and a measure of pressure gradients, however, an ocean station vessel location was selected. The ocean station vessel is OSV "C" (Charlie) at 52°45' north latitude and 35°30' west longitude. The month of February was selected for the years of 1965-1970. The 1200Z observations are used. The elements selected are the surface winds in $m \cdot s^{-1}$, the temperatures (°C), the sea level atmospheric pressures (mb), and the dew-point temperatures (°C). Other elements such as wave height, wave frequency, visibility, cloud, cloud height, precipitation, and other obstructions to vision can be used. However, some of these have a lower bound of zero or are dichotomous, i.e., a yes or no condition. Transformation to near normality of the distributions of these elements should be made before they are used with this technique.

The input data are not shown. Only the pertinent output data are shown, i.e., the means, variances or standard deviations, the correlation matrices, and the eigenvalue-eigenvector and mixture proportions.

8.2.6 Radiosonde and Rawinsonde Data

The foregoing examples clearly illustrate the usefulness of Wolfe's NORMIX and NORMAP techniques to separate mixed multivariate normal distributions. Two- or higher-dimensional vectors have been used. A radiosonde observation is composed of observations at levels in the atmosphere where significant changes are made from level to level. Also included are observations at mandatory levels so that all stations transmit data for the same levels. These can be used to produce synoptic charts of the data. Ordinarily, these charts can be studied by element and level by level or by changes within and between levels. Also, a complete observation can be studied as one vector in multidimensional space. Only the time and cost limitations of the computers restrict the problem. For example, if the winds, temperatures, dew points, and heights of selected pressure levels are used for 30 levels, an observation may be characterized by a vector of 150 components, i.e., a 150-dimension problem. Extension of this to other levels and features as well as differences level to level such as lapse rates, shears, etc., would extend the vector to still higher dimensions.

As an illustration, radiosonde and rawinsonde observations at Balboa, C.Z., Panama, are used. The 1500Z observations in the month of July for the years 1961-70 were selected. The levels selected were the surface, 950-, 850-, and 700-mb. The data used are winds in $m \cdot s^{-1}$, dry bulb and dew-point temperatures in °C, and heights in gdkm. Each vector then is a 4 x 5 or 20-dimensional vector. Assumption of both equal and unequal covariance matrices is made. No input data are shown. Only the output data in terms of means, variance (standard deviations), covariance (correlation) matrices, eigenvalue and eigenvector, and mixtures are shown.

8.3 Selected Data

8.3.1 Land-Sea Breeze Data Set

8.3.1.1 Input Information.

a. San Juan, Puerto Rico
b. The period of record is October 1-31, 1955, and October 1-3, 1956.
c. The data are surface winds.
d. The number of variables is two; these are zonal and meridional wind components.
e. The number in the sample is 200; the first 100 are from the 0600-1800 local standard time (l.s.t.) while the second 100 are from the 1200-1400 l.s.t.
f. The minimum number to be accepted into a cluster is three.

g. The null hypotheses are made that (k + 1) clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of (k + 1) clusters.
h. The first 40 two-dimensional vector entries are set up as the 40 means of 40 separate and individual clusters. These are 40 points in two dimensions.
i. Two assumptions are made. The first is the equality of covariances while the second is the non-equality of covariances.

8.3.1.2 Tables - Output Information. The program computes the necessary statistics for the three-cluster versus two-cluster including the discriminant functions and classification of entries into the clusters.

The output statistics are now shown in tables 1 and 2 for the two-cluster and the three-cluster even though the hypothesis that three types and two types were not significantly different was not rejected.

8.3.1.3 Figures and Discussion. San Juan, Puerto Rico, was selected because this is a known land-sea breeze effect location. The hour groups of 0600-0800 and 1200-1400 were selected when the land-sea breeze effect would most likely be operating. The periods of October 1-31, 1955, and October 1-3, 1956, were selected so as to provide 100 observations of the wind at each of the hour groups. The wind direction and speed are used to provide the zonal and meridional components of the wind. The first variable is the x-component or zonal (west-east) component of the wind while the second variable is the y-component or meridional (south-north) of the wind. Minus signs indicate components from the east or north.

The procedural techniques of the Wolfe NORMIX-NORMAP electronic computer routines (1971) worked well in this case.

Figure 8a shows a two-dimensional representation for the entire group of 200 observations. This may or may not be an adequate representation. No plot is made of data to visually assess the fit of the mathematical elliptical form. This form is obtained under the assumption that there is one single group. Here it is realized that such an assumption is not valid, yet it is shown for illustrative purposes. Under the assumption of two groups, the land and the sea breezes, and the equality of covariances, the breakout into two groups is shown. Elliptical error probable (e.e.p.) ellipses are shown. These are the 0.50 probability ellipses. When the axes of the ellipses are equal, the ellipses are circles and the 0.50 probability circle is called the circular error probable (c.e.p.). Figure 8b then shows a breakout into three clusters. The assumption of equality of covariances is exemplified by the fact that the ellipses all have the same shape, size, and orientation. This is not the case as shown in figures 8c and 8d where the covariances are assumed to be different. In these figures, the shape, size, and orientation may be different from ellipse to ellipse. The authors believe that assumption of unequal covariances is valid. The component means of the two-cluster configuration under the assumption of equal versus unequal covariance matrices are in the first type -0.4620 versus -0.5764 and 1.4574 versus 1.3866 and in the second type -4.7149 versus -4.7913 and -2.3420 versus -2.4596 $m \cdot s^{-1}$. The proportions assigned to the

Table 1 Surface wind statistics for San Juan, Puerto Rico, October 1-31, 1955, and October 1-3, 1956, 0600-0800 and 1200-1400 l.s.t. The sample size is 200; 100 from each hour group. The assumption is that the covariance matrices are the same in any breakdown. The variables are (1) the zonal component of the wind, positive from the west, and (2) the meridional component of the wind, positive from the south.

Logarithm of the likelihood;	chi-square with 4 d.f.;	probability of null hypothesis.
(a) 2 to 1 types 32.478160	(a) 63.82	(a) 0.00000000
(b) 3 to 2 types 2.6044397	(b) 5.10	(b) 0.27672217*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample (units are $m \cdot s^{-1}$)

	1	2			1	2
Means	-2.1114	-0.1054				
Std. Dev.	2.9022	2.3077				
Corr.	1.0000	0.4922				
	0.4922	1.0000				
<u>Within Group</u>	Two Clusters				Three Clusters	
Std. Dev.	2.0108	1.3527			1.8937	1.3288
Corr.	1.0000	-0.2267			1.0000	-0.1766
	-0.2267	1.0000			-0.1766	1.0000
<u>Clusters</u>	Two Clusters				Three Clusters	
Lambda (Proportion)	Means				Means	
(1) 0.5887	-0.4620	1.4574	(1)	0.5884	-0.4572	-1.4569
(2) 0.4113	-4.7149	-2.3420	(2)	0.3912	-4.9544	-2.2436
			(3)	0.0204	-0.1962	-4.1576
Discriminant Functions						
Variable (1)	0.3621				0.3785	0.3802
Variable (2)	0.6432				0.6286	-0.4352

Table 2 Surface wind statistics for San Juan, Puerto Rico, October 1-31, 1955, and October 1-3, 1956, 0600-0800 and 1200-1400 l.s.t. The sample size is 200; 100 from each hour group. The assumption is the covariance matrices are not equal. The variables are (1) the zonal component of the wind, positive from the west, and (2) the meridional component of the wind, positive from the south.

Logarithm of the likelihood;	chi-square with 4 d.f.;	probability of null hypothesis.
(a) 2 to 1 types 37.528096	(a) 73.74	(a) 0.00000000
(b) 3 to 2 types 12.212320	(b) 23.94	(b) 0.00777169
(c) 4 to 3 types 0.80050798	(c) 1.56	(c) 0.99871828*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample (units are m·s^{-1})

	1	2		1	2
Means	-2.2114	-0.1054			
Std. Dev.	2.9022	2.3077			
Corr.	1.0000	0.4922			
	0.4922	1.0000			
Characteristics of Type 1	Proportion 0.612			Proportion 0.635	
Means	-0.5764	1.3866		-0.6423	1.2772
Std. Dev.	2.1589	1.3792		2.1579	1.4784
Corr.	1.0000	-0.0514		1.0000	0.0092
	-0.0514	1.0000		0.0092	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
4.6695	0.9985	0.0552	4.6571	0.9999	-0.0118
1.8937	-0.0552	0.9985	2.1552	0.0118	0.9999

(Table 2 continued)

Characteristics of Type 2	Proportion 0.388		Proportion 0.255	
Means	-4.7913	-2.4596	-5.3239	-2.0881
Std. Dev.	1.8668	1.2932	1.0169	1.2851
Corr.	1.0000	-0.4905	1.0000	-0.5600
	-0.4905	1.0000	-0.5600	1.0000

Eigen-Values	Vectors		Eigen-Values	Vectors	
4.0697	0.8966	0.4428	2.1371	-0.5529	0.8333
1.0875	-0.4428	0.8966	0.5485	0.8333	0.5529

Characteristic of Type 3	Proportion 0.110	
Means	-4.0398	-3.4767
Std. Dev.	2.6959	0.6627
Corr.	1.0000	-0.8456
	-0.8456	1.0000

Eigen Values	Vectors	
7.5874	0.9784	0.2068
0.1198	-0.2068	0.9784

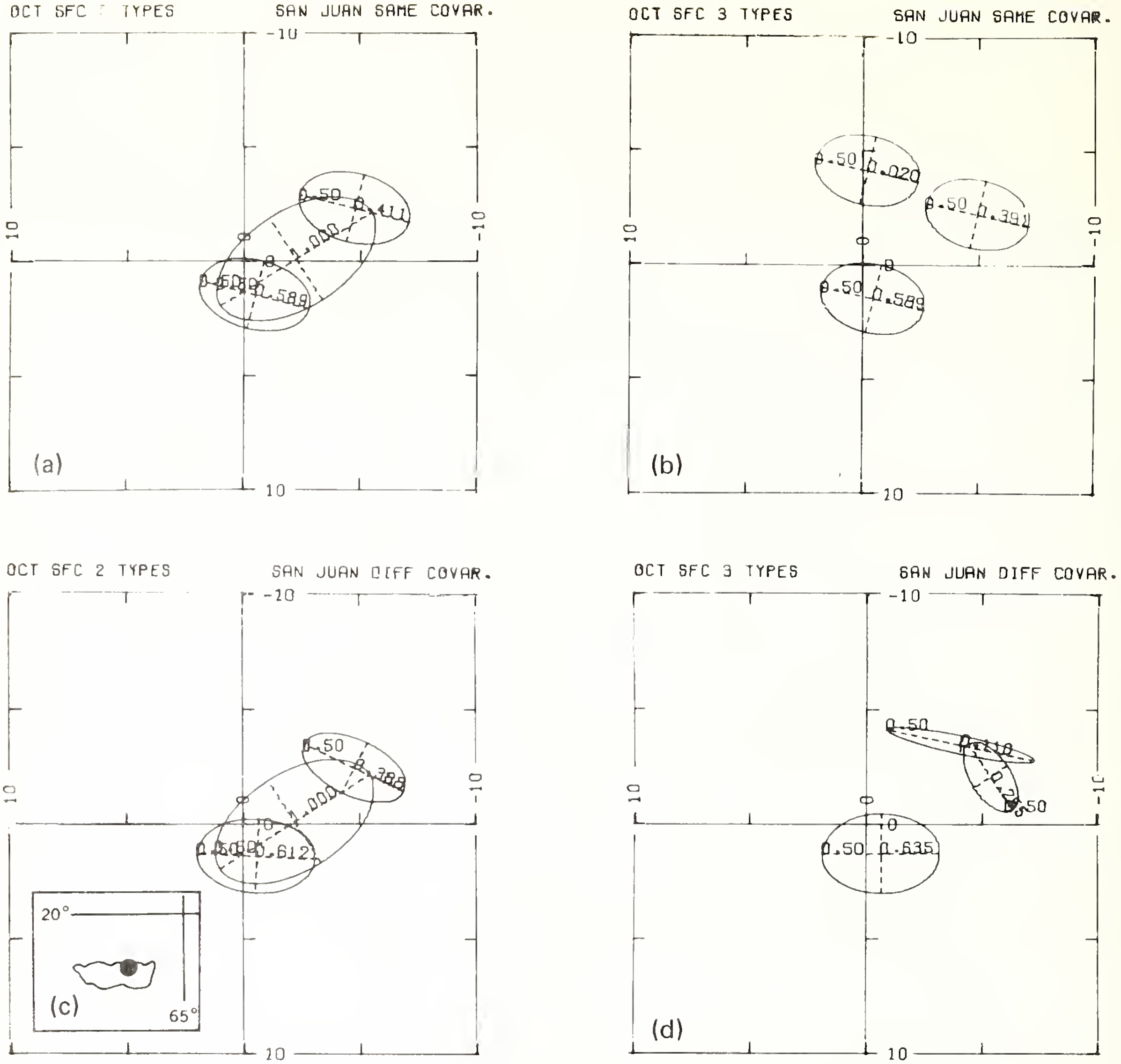


Figure 8 San Juan, Puerto Rico, surface wind distributions; period of record October 1-31, 1955, and October 1-3, 1956, hours 0600 and 0800 and 1200-1400 local standard time; n = 200, 100 from each period; units, $m \cdot s^{-1}$; separation shown for two and three types with covariances assumed equal and then unequal.

(a) Total and two types, equal covariance matrices.
(b) Three types, equal covariance matrices.
(c) Total and two types, unequal covariance matrices.
(d) Three types, unequal covariance matrices.

first type are 0.5887 versus 0.612 and the second type are 0.4113 versus 0.388. The comparisons indicate similarity. Under the assumption of unequal covariance matrices, the second type shown above (though losing a part to the first) then breaks down into two clusters which are proportionally 0.255 and 0.110.

Quite clearly the unequal covariance ensemble breaks down into at least three clusters, a south-southeast wind and an east-northeast wind, the early morning wind versus the noon wind, and the land breeze versus the sea breeze. The mean wind is from the east by southeast with components -2.2114 and -0.1054 $m \cdot s^{-1}$. This is the easterly trade though it is somewhat less than the average of 3.5 $m \cdot s^{-1}$ for that region [Crutcher, Wagner, and Arnett (1966)].

The land-sea breeze situation illustrates the output of the technique in tabular form and in two-dimensional illustrations with ellipses. Two assumptions are shown. First, there is the assumption that the underlying statistics have the same covariance matrix; i.e., the underlying physical bases are operating the same. Second, there is the assumption that the underlying statistics have differing covariance matrices; i.e., there is some reason to believe that the underlying physical bases operate differently. Whether the assumptions are correct in any particular case is not known for these are not tested here. Both assumptions are made. The results are presented. The reader can make his own assessment and choose the assumption he pleases. However, statistical tests are used to reach decisions as to whether there are two groups or less, three groups or two, etc. As with all other examples, the decision level chosen to work with is the probability level of 0.01. The null hypothesis is that the distribution of (k + 1) groups is not different from (k) groups; that is, rejection of the null hypothesis that there are (k + 1) groups rather than (k) groups is sought.

Output or input data in computer format and the intermediate steps are not shown in this example. For these the reader is referred to Wolfe (1971b) whose electronic computer program is adapted for use here. The mid-latitude tropospheric wind data set (paragraph 8.3.3) does contain some of the intermediate steps.

8.3.2 Tropical Stratospheric Wind Data Set

8.3.2.1 Input Information.

a. Canton Island, South Pacific, (U.S.A. and Great Britain)
b. The periods of record are the months of July and January, 1954-1964, and for 1957-1967.
c. The data are stratospheric winds, heights of pressure surfaces, and temperatures. The pressure levels are 50-, 30-, 20-, and 10-mb.
d. The number of variables is two for the first period and four for the second. These are for the first period, zonal and meridional components of the wind, positive from the west and south. For the second period, these are the winds, heights, and temperatures. The units are $m \cdot s^{-1}$, m, and °C.
e. The number in the samples vary from level to level because at

times the balloons failed to reach the higher altitudes. The numbers are 263, 244, and 162, respectively, for the first three levels above. Though the 10-mb data were processed, only one cluster was determined. The results are not shown.

f. The minimum number to be accepted into a cluster is three for the first period and five for the second.

g. The null hypotheses are made that (k + 1) clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of (k + 1) clusters.

h. The first 40 two-dimensional vector entries in each level are set up as the 40 means of 40 separate and individual clusters. These are 40 points in two dimensions.

i. Two assumptions are made. The first assumption is the equality of covariances. The second assumption is the non-equality of covariances.

8.3.2.2 Tables for Wind Configurations (1954-1964). Tables 3 through 10 provide the output data in tabular form provided by the Wolfe (1971b) NORMIX-NORMAP computer routine for the first period. An asterisk indicates the rejection of the null hypothesis that (k + 1) types are not significantly different from the (k) types. Thus, under the following assumption of equal then unequal covariances and at the 0.01 probability level, the greatest number of clusters (types) for the wind distributions 1954-1964 is indicated below:

	July		January	
	Equal	Unequal (Covariances)	Equal	Unequal
50-mb	2	3	2	2
30-mb	at least 4	3	1	1
20-mb	3	2	1	1
10-mb	1	1	1	1

The statistics for the single clusters are not shown.

Table 3 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 50-mb. Sample size is 263. The assumption is that the covariance matrices are the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind component, positive from the south.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a)	2 to 1 types 27.165587	(a) 53.61	(a) 0.00000000
(b)	3 to 2 types 61.844617	(b) 12.18	(b) 0.1605621*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample.

	1	2		1	2
Means	1.1251	-0.5502			
Std. Dev.	14.0700	3.5679			
Corr.	1.0000	-0.0300			
	-0.0300	1.0000			
<u>Within Group</u>	Two Clusters			Three Clusters	
Std. Dev.	7.9448	3.5109		7.4822	3.2166
Corr.	1.0000	-0.3185		1.0000	-0.2369
	-0.3185	1.0000		-0.2369	1.0000
Type Lambda	Means			Means	
1) 0.6361	9.9084	-0.0686	1) 0.6246	10.0418	0.0198
2) 0.3639	-14.2256	-1.3947	2) 0.0570	-1.6823	-6.7786
			3) 0.3184	-15.8612	-0.5528
Discriminant Functions					
Variable 1	0.1319			0.1370	-0.0126
Variable 2	0.1270			0.1041	0.3031

Table 4 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 30-mb. Sample size is 244. The assumption is that the covariance matrices are the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind component, positive from the south.

	Logarithm of the likelihood		chi-square with 4 d.f.		probability of null hypothesis.
(a)	2 to 1 types 71.531512	(a)	141.01	(a)	0.00000000
(b)	3 to 2 types 9.7739748	(b)	19.23	(b)	0.00070904
(c)	4 to 3 types 19.060173	(c)	37.42	(c)	0.00000015

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample.

	1	2		1	2		1	2
Means	-3.8760	-0.3253						
Std. Dev.	19.5398	4.0676						
Corr.	1.0000	0.0385						
	0.0385	1.0000						
Within Group	Two Clusters			Three Clusters			Four Clusters	
Std. Dev.	7.6585	4.0675		7.6400	3.6940		5.3554	3.6933
Corr.	1.0000	0.1121		1.0000	0.0937		1.0000	0.1113
	0.1121	1.0000		0.0937	1.0000		0.1113	1.0000
Type Lambda	Means			Means			Means	
1) 0.4827	-22.4853	-0.3004	1) 0.4828	-22.4797	-0.2876	1) 0.4394	-24.0466	-0.3111
2) 0.5173	13.4891	-0.3485	2) 0.0148	18.5728	13.4282	2) 0.1584	-0.5184	-0.5910
			3) 0.5024	13.3385	-0.7669	3) 0.0147	18.6547	13.5317
						4) 0.3875	16.7660	-0.7581
Discriminant Functions								
Variable 1	-0.1314			0.1315	0.0001		0.1879	-0.0003
Variable 2	-0.0283			-0.0257	0.2707		-0.0299	0.2708

Table 5 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 20-mb. Sample size is 162. The assumption is that the covariance matrices are the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind component, positive from the south.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a)	2 to 1 types 71.800990	(a) 140.50	(a) 0.00000000
(b)	3 to 2 types 13.764612	(b) 26.85	(b) 0.00002132
(c)	4 to 3 types -0.000039	(c) -0.00	(c) 1.000000*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample.

	1	2		1	2		1	2
Means	-10.0693	-0.1189						
Std. Dev.	19.9881	3.3763						
Corr.	1.0000	-0.0035						
	-0.0035	1.0000						
Within Group	Two Clusters			Three Clusters			Four Clusters	
Std. Dev.	6.6696	3.3737		5.1788	3.2879		5.1797	3.2877
Corr.	1.0000	0.1002		1.0000	0.2400		1.0000	0.2400
	0.1002	1.0000		0.2400	1.0000		0.2400	1.0000
Type Lambda	Means			Means			Means	
1) 0.4021	12.9050	-0.2799	1) 0.1455	2.5267	1.4483	1) 0.1453	2.5193	1.4514
2) 0.5979	-25.5230	-0.0106	2) 0.2780	16.7471	-1.0294	2) 0.2783	16.7613	-1.0289
			3) 0.5764	-26.1833	-0.0754	3) 0.5060	-26.1847	-0.0720
						4) 0.0705	-26.1924	-0.0956
Discriminant Functions								
Variable 1	0.1507			0.1989	0.0037		0.1988	0.0037
Variable 2	-0.0339			-0.0808	0.3027		-0.0808	0.3027

Table 6 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 50-mb. Sample size is 263. The assumption is that the covariance matrices are not the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind, positive from the south.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a)	2 to 1 types 30.173770	(a)	59.54	(a)	0.00000000
(b)	3 to 2 types 14.343754	(b)	28.25	(b)	0.00164537
(c)	4 to 3 types 7.0614291	(c)	13.88	(c)	0.17847916*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level.

Characteristics of the whole sample.

	1	2	1	2
Means	1.1251	-0.5502		
Std. Dev.	14.0700	3.5679		
Corr.	1.0000	-0.0300		
	-0.0300	1.0000		

Two Types With Proportions

	Type 1 (0.608)		Type 2 (0.392)	
Means	10.3910	0.2223	-13.2905	-1.4408
Std. Dev.	7.1747	3.4534	9.2004	3.5603
Corr.	1.0000	-0.2728	1.0000	-0.4408
	-0.2728	1.0000	-0.4408	1.0000

Eigen-Values	Vectors		Eigen-Values	Vectors	
52.6002	0.9865	0.1639	87.4360	0.9819	0.1896
10.8024	-0.1639	0.9865	9.8871	-0.1896	0.9819

Table 6 (continued)

Three Types With Proportions

	Type 1 (0.373)			Type 2 (0.338)	
	1	2		1	2
Means	10.7216	0.2663		4.9486	-2.2004
Std. Dev.	5.6926	3.5614		11.3340	3.9597
Corr.	1.0000	-0.6196		1.0000	0.4364
	-0.6196	1.0000		0.4364	1.0000
Eigen- Values	Vectors		Eigen- Values	Vectors	
38.5133	0.8993	0.4373	131.7641	0.9861	-0.1664
6.5754	-0.4373	0.8993	12.3750	0.1664	0.9861

	Type 3 (0.287)	
Means	-15.9766	0.3409
Std. Dev.	8.1386	2.0971
Corr.	1.0000	0.0347
	0.0347	1.0000
Eigen- Values	Vectors	
66.2424	1.0000	-0.0096
4.3924	0.0096	1.0000

Four Types With Proportions

	Type 1 (0.026)			Type 2 (0.177)	
Means	14.3882	6.6000		-7.4919	-4.1879
Std. Dev.	13.1061	2.4964		8.4065	3.5182
Corr.	1.0000	-0.9881		1.0000	-0.3777
	-0.9881	1.0000		-0.3777	1.0000
Eigen- Values	Vectors		Eigen- Values	Vectors	
177.8597	0.9827	0.1851	72.7376	0.9833	0.1820
0.1428	-0.1851	0.9827	10.3099	-0.1820	0.9833

Table 6 (continued)

Four Types With Proportions

	Type 3 (0.553)			Type 4 (0.245)	
Means	10.8686	-0.2486		-16.0757	0.6447
Std. Dev.	6.1573	3.1438		8.4811	1.8780
Corr.	1.0000	-0.3513		1.0000	0.1235
	-0.3513	1.0000		0.1235	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
39.4747	0.9746	0.2240	71.9849	0.9996	-0.0287
8.3209	-0.2240	0.9746	3.4705	0.0287	0.9996

Table 7 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 30-mb. Sample size is 244. The assumption is that the covariance matrices are not the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind, positive from the south.

Logarithm of the likelihood;		chi-square with 4 d.f.;		probability of null hypothesis.
(a)	2 to 1 types 85.675491	(a)	168.89	(a) 0.00000000
(b)	3 to 2 types 18.426554	(b)	36.25	(b) 0.00007627
(c)	4 to 3 types 9.2238620	(c)	18.11	(c) 0.05317713*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level.

Characteristics of the whole sample.

	1	2	1	2
Means	-3.8760	-0.3253		
Std. Dev.	19.5398	4.0676		
Corr.	1.0000	0.0385		
	0.0385	1.0000		

Two Types With Proportions

	Type 1 (0.435)		Type 2 (0.565)	
Means	-24.1471	-0.3279	11.7512	-0.3233
Std. Dev.	4.8449	3.5602	9.8518	4.4192
Corr.	1.0000	0.1552	1.0000	0.0754
	0.1552	1.0000	0.0754	1.0000

Eigen-Values	Vectors		Eigen-Values	Vectors	
24.1008	0.9736	-0.2281	97.1957	0.9991	-0.0422
12.0473	0.2281	0.9736	19.0754	0.0422	0.9991

Table 7 (continued)

Three Types With Proportions

	Type 1 (0.371)			Type 2 (0.282)	
	1	2		1	2
Means	-24.6160	-0.5386		-1.0247	0.4847
Std. Dev.	4.4515	2.4781		15.0486	6.0073
Corr.	1.0000	0.0685		1.0000	0.0738
	0.0685	1.0000		0.0738	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
19.8572	0.9985	-0.0550	226.6929	0.9984	-0.0350
6.0991	0.0550	0.9985	35.8548	0.0350	0.9994

	Type 3 (0.347)	
Means	15.9746	-0.7561
Std. Dev.	5.8907	3.3160
Corr.	1.0000	0.2317
	0.2317	1.0000
Eigen-Values	Vectors	
35.5352	0.9834	-0.1814
10.1015	0.1814	0.9834

Four Types With Proportions

	Type 1 (0.367)			Type 2 (0.043)	
	1	2		1	2
Means	-24.5988	-0.4851		11.1380	-7.6819
Std. Dev.	4.4655	2.4451		11.8059	1.1746
Corr.	1.0000	0.0648		1.0000	-0.9052
	0.0648	1.0000		-0.9052	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
19.9768	0.9987	-0.0505	140.5112	0.9960	0.0898
5.9429	0.0505	0.9987	0.2473	-0.0898	0.9960

Table 7 (continued)

Four Types With Proportions

	Type 3 (0.214)			Type 4 (0.376)	
Means	-5.0779	1.6368		15.3429	-0.4423
Std. Dev.	14.5772	6.1490		6.6104	2.8524
Corr.	1.0000	0.3582		1.0000	0.1930
	0.3582	1.0000		0.1930	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
218.2089	0.9845	-0.1752	44.0661	0.9949	-0.1007
32.0958	0.1752	0.9845	7.7677	0.1007	0.9949

Table 8 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of July during 1954-1964. The pressure level is 20-mb. Sample size is 162. The assumption is that the covariance matrices are not the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind, positive from the south.

Logarithm of the likelihood;	chi-square with 4 d.f.;	probability of null hypothesis.
(a) 2 to 1 types 84.858770	(a) 166.05	(a) 0.00000000
(b) 3 to 2 types 4.5207028	(b) 8.82	(b) 0.54944347*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level.

Characteristics of the whole sample.

	1	2	1	2
Means	-10.0693	-0.1189		
Std. Dev.	19.9881	3.3763		
Corr.	1.0000	-0.0035		
	-0.0035	1.0000		

Two Types With Proportions

	Type 1 (0.428)		Type 2 (0.572)	
Means	11.6050	-0.1364	-26.3059	-0.1058
Std. Dev.	9.1353	3.6246	4.5555	3.1775
Corr.	1.0000	-0.1093	1.0000	0.1934
	-0.1093	1.0000	0.1934	1.0000

Eigen-Values	Vectors		Eigen-Values	Vectors	
83.6403	0.9987	0.0513	21.4431	0.9709	-0.2395
12.9522	-0.0513	0.9987	9.4058	0.2395	0.9709

Table 8 (continued)

Three Types With Proportions

Type 1 (0.290)

	1	2
Means	8.0972	0.2056
Std. Dev.	19.1706	3.5413
Corr.	1.0000	-0.2219
	-0.2219	1.0000

Eigen-Values	Vectors	
84.8178	0.9951	0.0992
11.8221	-0.0992	0.9951

Type 2 (0.141)

	1	2
Means	18.5108	-0.7634
Std. Dev.	4.2761	3.7393
Corr.	1.0000	0.5415
	0.5415	1.0000

Eigen-Values	Vectors	
25.0548	0.7878	-0.6160
7.2129	0.6160	0.7878

Type 3 (0.569)

	1	2
Means	-26.3544	-0.1250
Std. Dev.	4.4932	3.1661
Corr.	1.0000	0.1812
	0.1812	1.0000

Eigen-Values	Vectors	
20.8047	0.9726	-0.2325
9.4078	0.2325	0.9726

Table 9 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of January 1954-1964. The pressure level is 50-mb. Sample size is 168. The assumption is that the covariance matrices are the same. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind component, positive from the south.

Logarithm of the likelihood: chi-square with 4 d.f.; probability of null hypothesis.

(a) 2 to 1 types 10.845283	(a) 21.24	(a) 0.00028396
(b) 3 to 2 types 5.498598	(b) 10.74	(b) 0.02970488*

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level.

Characteristics of the whole sample.

	1	2		1	2
Means	-2.4529	0.6009			
Std. Dev.	12.2183	3.2754			
Corr.	1.0000	0.0280			
	0.0280	1.0000			
<u>Within Group</u>					
Std. Dev.	6.9054	3.2747		6.8723	2.7151
Corr.	1.0000	0.0815		1.0000	0.0637
	0.0815	1.0000		0.0637	1.0000
<u>Type Lambda</u>					
1) 0.4538	-13.5119	0.6795	1) 0.4424	-13.7874	0.5320
2) 0.5462	6.7344	0.5356	2) 0.4393	6.2786	1.9283
			3) 0.1183	7.5054	-4.0685
Discriminant Functions					
Variable 1	0.1453			0.1458	0.0028
Variable 2	-0.0295			-0.0305	0.3678

Table 10 Upper wind statistics for Canton Island, U.S.A. and U.K. Units are $m \cdot s^{-1}$. The period of record is the months of January, 1954-1964. The pressure level is 50-mb. Sample size is 168. The assumption is that the covariance matrices are unequal. The variables are (1) the zonal wind component, positive from the west, and (2) the meridional wind component, positive from the south.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

	Logarithm of the likelihood		chi-square		probability
(a)	2 to 1 types 20.949167	(a)	41.03	(a)	0.00001117
(b)	3 to 2 types 3.286572	(b)	6.42	(b)	0.77912911*

Lambda is the proportional amount assigned to each type. An asterisk indicates the rejection of the null hypothesis at the 0.01 level.

Characteristics of the whole sample.

	1	2	1	2
Means	-2.4529	0.6009		
Std Dev.	12.2183	3.2754		
Corr.	1.0000	0.0280		
	0.0280	1.0000		

Two Types With Proportions

	Type 1 (0.627)			Type 2 (0.373)	
Means	-9.5312	1.0884		9.4278	-0.2174
Std. Dev.	9.6039	2.9303		4.4498	3.6407
Corr.	1.0000	0.3176		1.0000	0.2171
	0.3176	1.0000		0.2171	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
93.1791	0.9945	-0.1051	21.3319	0.9169	-0.3992
7.6426	0.1051	0.9945	11.7235	0.3992	0.9169

Table 10 (continued)

Three Types With Proportions

	Type 1 (0.581)			Type 2 (0.196)	
	1	2		1	2
Means	-10.7113	0.9144		8.9686	2.7487
Std. Dev.	8.8718	2.9316		4.8902	1.8624
Corr.	1.0000	0.2638		1.0000	-0.0712
	0.2638	1.0000		-0.0712	1.0000
Eigen-Values	Vectors		Eigen-Values	Vectors	
79.3732	0.9953	-0.0965	23.9343	0.9995	0.0317
7.9294	0.0965	0.9953	3.4478	-0.0317	0.9995

	Type 3 (0.223)	
Means	9.0077	-2.1027
Std. Dev.	4.4701	3.3315
Corr.	1.0000	0.2473
	0.2473	1.0000
Eigen-Values	Vectors	
21.3104	0.9407	-0.3393
9.7707	0.3393	0.9407

8.3.2.3 Figures and Discussion for Wind Configurations (1954-1964). Figures 9 through 19 show various combinations of the 0.50 probability ellipses as well as the wind plot diagrams to which these pertain. For example, figure 10a gives the 0.50 probability ellipse for the total group and a breakout into two groups. This 0.50 probability ellipse is the ellipse estimated to contain one-half of the winds from the cluster (type) to which it pertains. Figures 10b and 10c are similar illustrations for the 30- and 20-mb levels. Comparison between the results of the assumptions of equal and unequal covariance matrices are possible. Under the assumption of equal covariance matrices, the size, shape, and orientation of the cluster breakout are the same but not necessarily the same as the complete (total) distribution. That is, the flux indicated by the total may be quite different from that within the individual clusters though the individual clusters imply that the flux across the pertinent airstream in each cluster is the same as in another. Under the assumption of unequal covariance matrices, the various clusters may show different size, scale, shape, and orientation.

In addition, other figures in the above group permit comparison from level to level of the totals and of the various types. These tabular and graphical representations quite clearly indicate that elongated elliptical distributions computed from ordinary bivariate normal statistical routines should be backed up by plot or scatter diagrams. If a plot is not available, then a wind rose of some type should be available for examination. The wind rose usually referred to is the WBAN-120 (revised) available from the National Climatic Center (1958). This is based on the work by Crutcher (1957). A decision can then be made as to whether the computed bivariate statistics are valid. These assume a unimodal bivariate distribution. In the output statistics of WBAN-120 revised format, for example, a ratio of the major to minor axes in excess of four should indicate a need for study to see whether clustering is evident. See table 9 for such a comparison. Both plots and analytic procedures such as are available in discriminant function analysis, factor analysis, or principal component analysis or other clustering routines can be used.

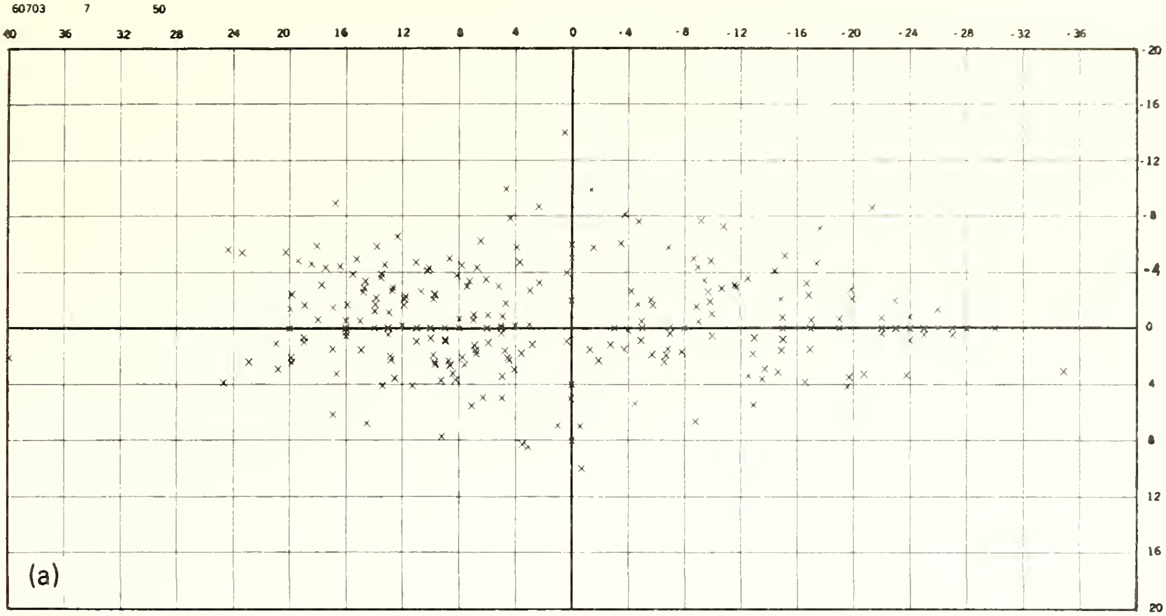
"A priori" considerations may also indicate that clustering techniques should be used. That is, previous research may indicate that the use of the total distribution as a unimodal bivariate (or multivariate) distribution is unwarranted and that probabilistic statements from such a model may be erroneous.

The previous example of the land-sea breeze effect at San Juan, Puerto Rico, and the present example of the stratospheric winds at Canton Island are "a priori" types.

The procedures also are helpful in establishing the quality assurance of the data. Clusters composed of only one or a few observations and far distant from the main cluster or clusters may be examined for validity. This feature will be discussed in a later section. In particular, in tables 6 and 7 and in figures 17a and 17b for the four-cluster breakout, the very low proportions in one of the clusters and the very large ratio of the major axis to minor axis indicate that these groups may be suspect for one or more reasons. The nonrejection of the null hypothesis, four clusters versus the three clusters, implies that the four-cluster breakout was not significantly different from the three-cluster breakout. Thus, the investigator can simply stay with the three-cluster configuration while examining the isolated questionable groups indicated in the four-cluster configuration.

The equatorial stratospheric winds situation illustrates the output of the technique in tabular form and in two-dimensional depiction with 0.50 ellipses. Two assumptions are used. First, there is the assumption that the underlying distributions are the same, i.e., that the statistics have the same covariance matrices (that the underlying bases are the same). Second, there is the alternative assumption that the statistics represent different physical and dynamic situations, i.e., the covariance matrices are not the same. There is some reason to believe that the underlying physical bases operate differently. The reader can select the assumption that best fits his knowledge and experience.

Under the assumption of equal covariance matrices, one of the outputs is a plot of discriminant function assignments, such as one versus two or one versus three, etc. This is a computer tabulation type of two-dimensional plotting with each individual data point carrying the number of the cluster type to which it is assigned by the classification (discrimination) procedure. Figures 13, 14, and 15 illustrate the output for Canton Island during July at the 50-mb and 30-mb levels. Figure 13 for the July 50-mb level shows the print plot of discriminant function 1 and 2 where three types are computed. Figures 14 and 15 for the July 30-mb level show the print plots of discriminant functions 1 and 2 for the first three-type dissection and then four-type dissection. It appears here that type 2 of 3 becomes type 3 of 4 while type 3 of 2 breaks down into types 2 and 4 of 4.



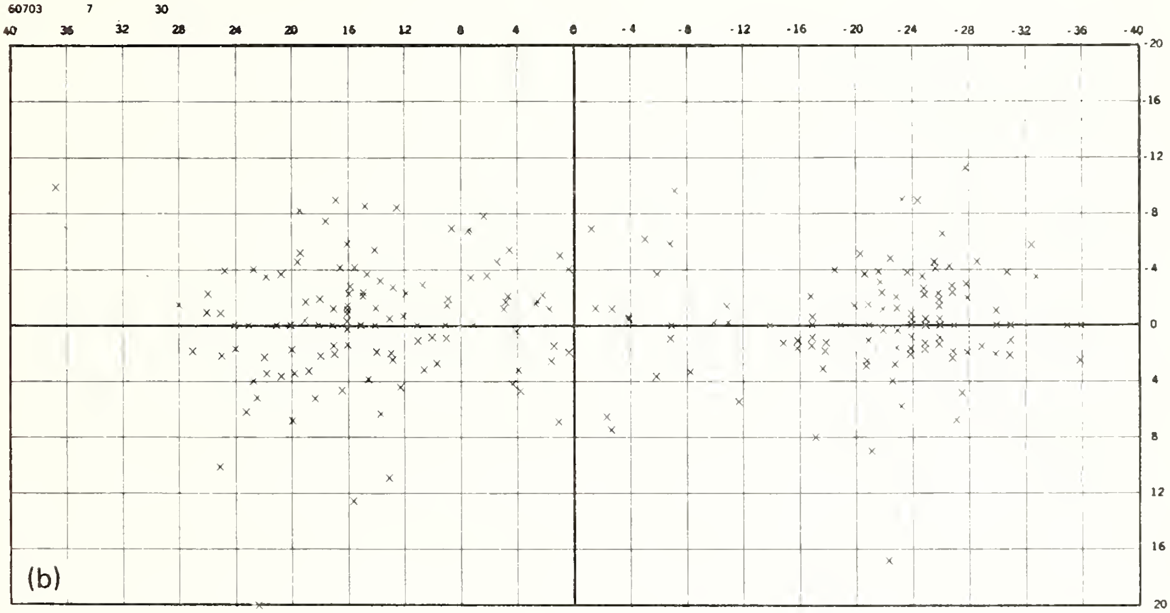


Figure 9 Canton Island, U.S.A. and U.K.; upper wind distribution plots; period of record, the months of July 1954-1964; pressure levels (a) 50-, (b) 30-, (c) 20-, and (d) 10-mb; units, m·s^{-1}.

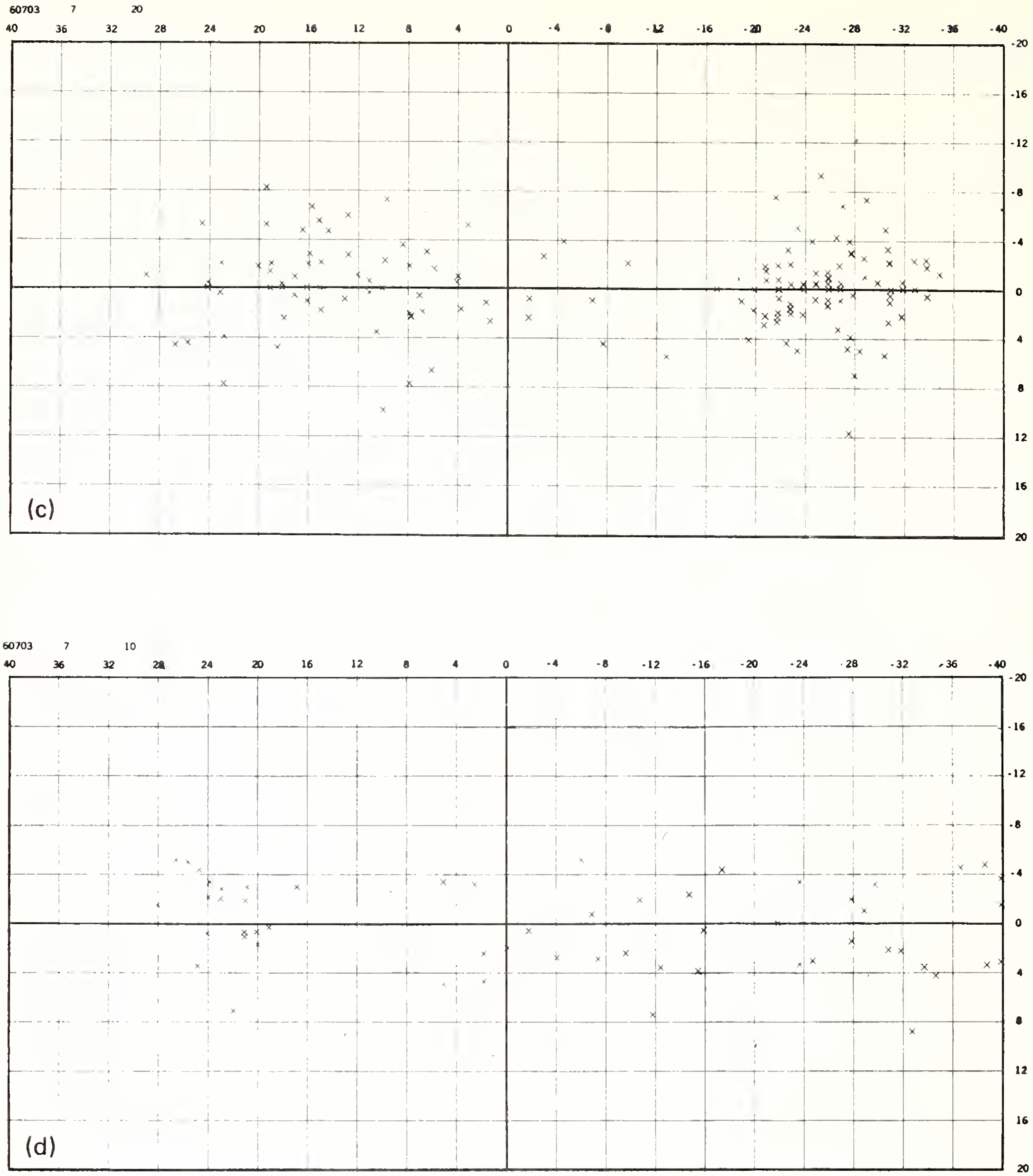


Figure 9 (continued)

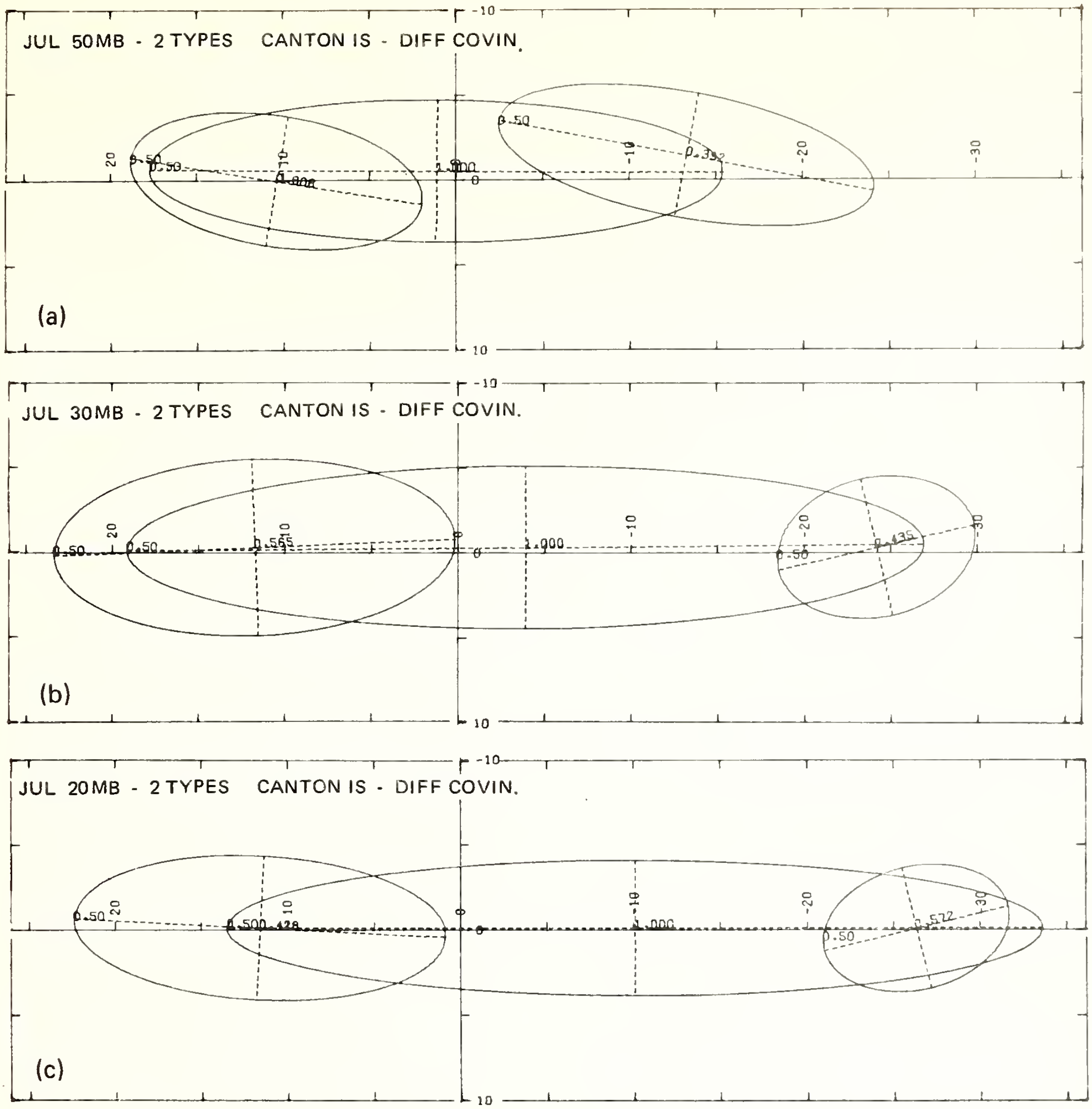


Figure 10 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50-, 30-, amd 20-mb; units, $m \cdot s^{-1}$; wind plot shown in figure 9; separation shown for two types, assumption of unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Total distribution and two types at 50-mb, n = 263.
(b) Total distribution and two types at 30-mb, n = 244.
(c) Total distribution and two types at 20-mb, n = 162.

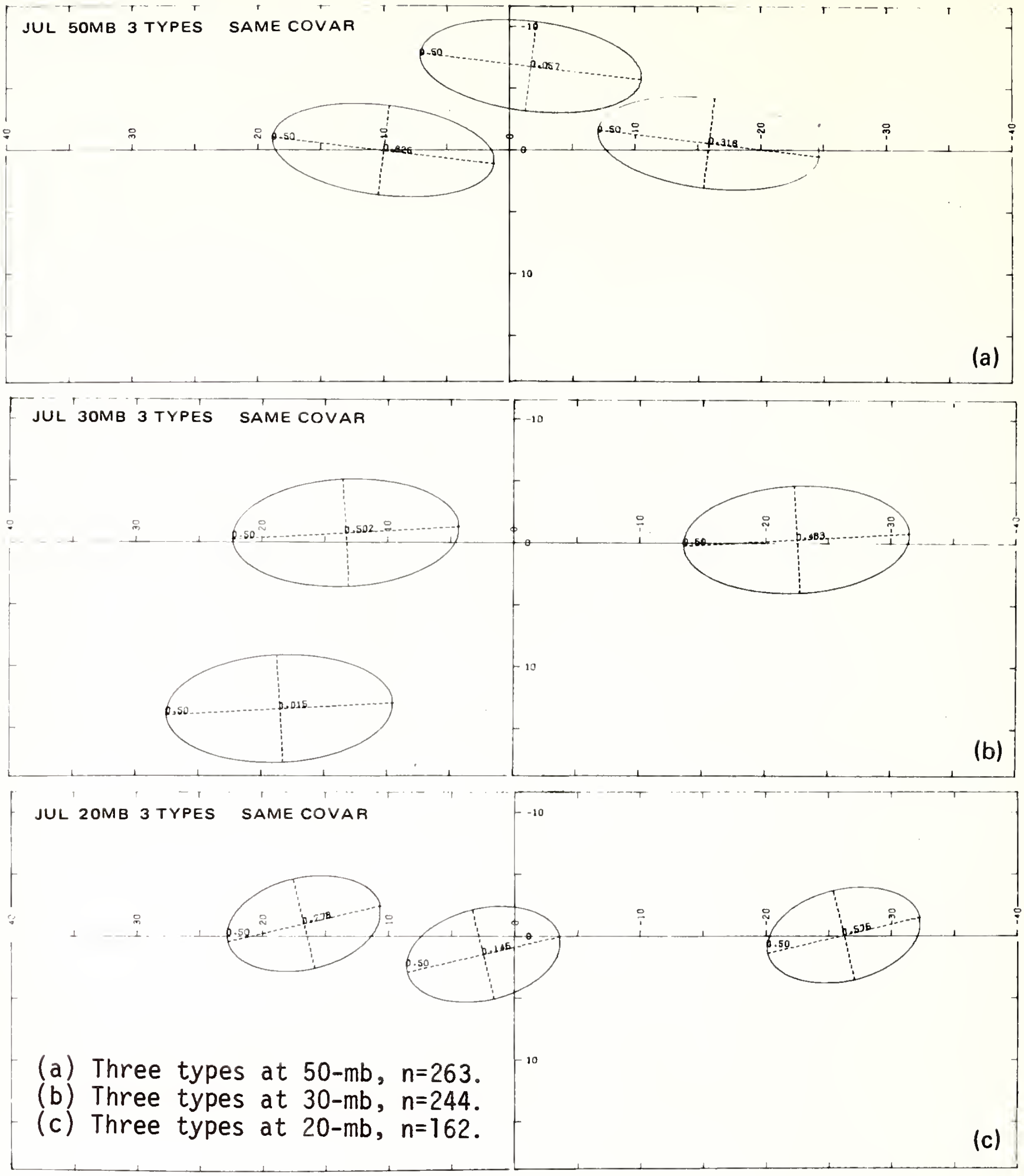


Figure 11 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50-, 30-, and 20-mb; units, m·s^{-1}; wind plot shown in figure 9; separation shown for three types, assumption of equal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Three types at 50-mb, n = 263.
(b) Three types at 30-mb, n = 244.
(c) Three types at 20-mb, n = 162.

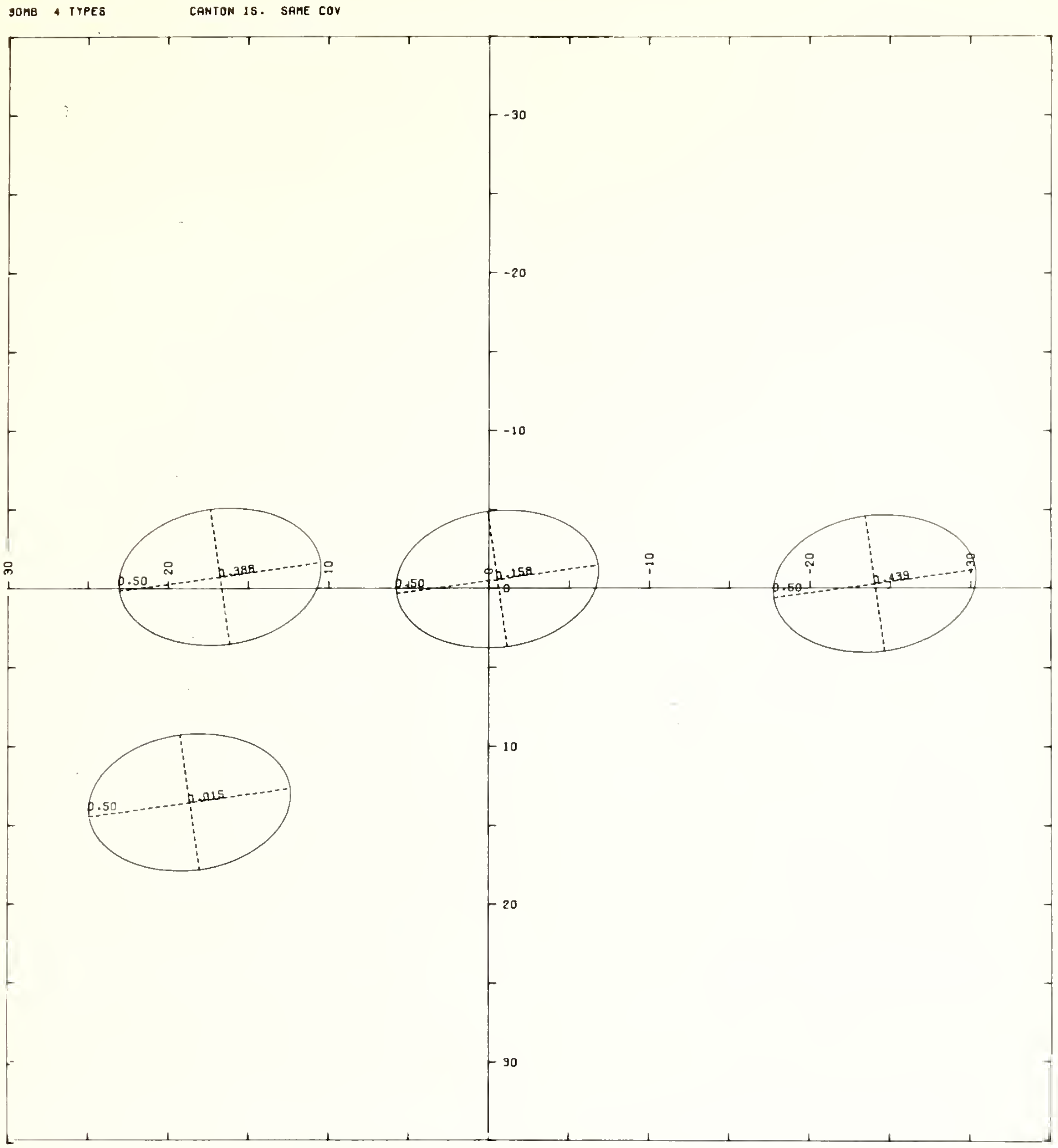


Figure 12 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure level, 30-mb; units, $m \cdot s^{-1}$; n = 244; wind plot shown in figure 9; separation shown for four types, assumption of equal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

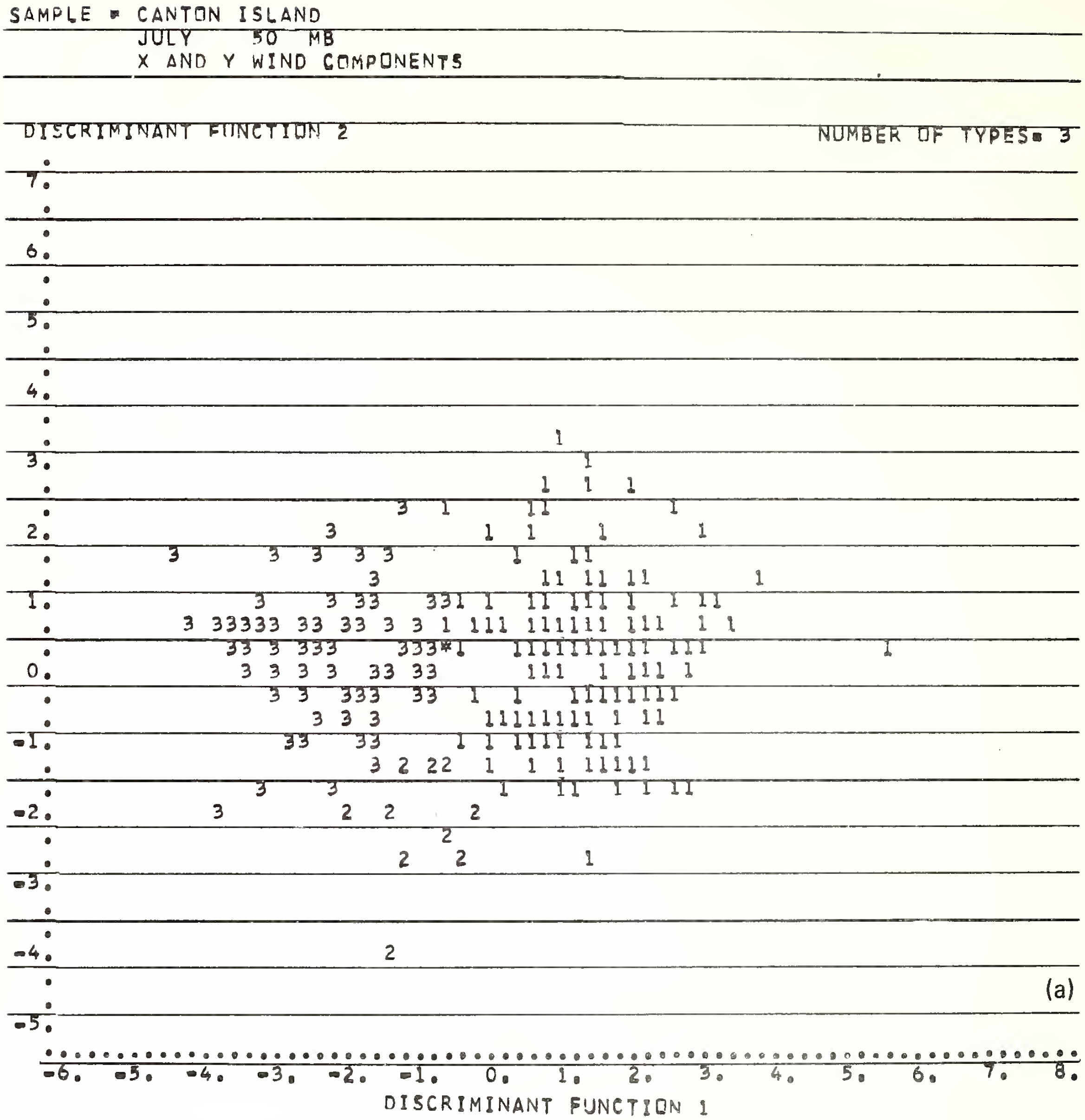


Figure 13 Plot of Discriminant Functions 1 versus 2 for Canton Island, U.S.A. and U.K.; July winds shown in figure 9 for winds shown in figure 11, based on assumption of equal covariance matrices; period of record 1954-1964; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Three types at 50-mb.
(b) Three types at 30-mb.
(c) Four types at 30-mb.
(d) Three types at 20-mb.

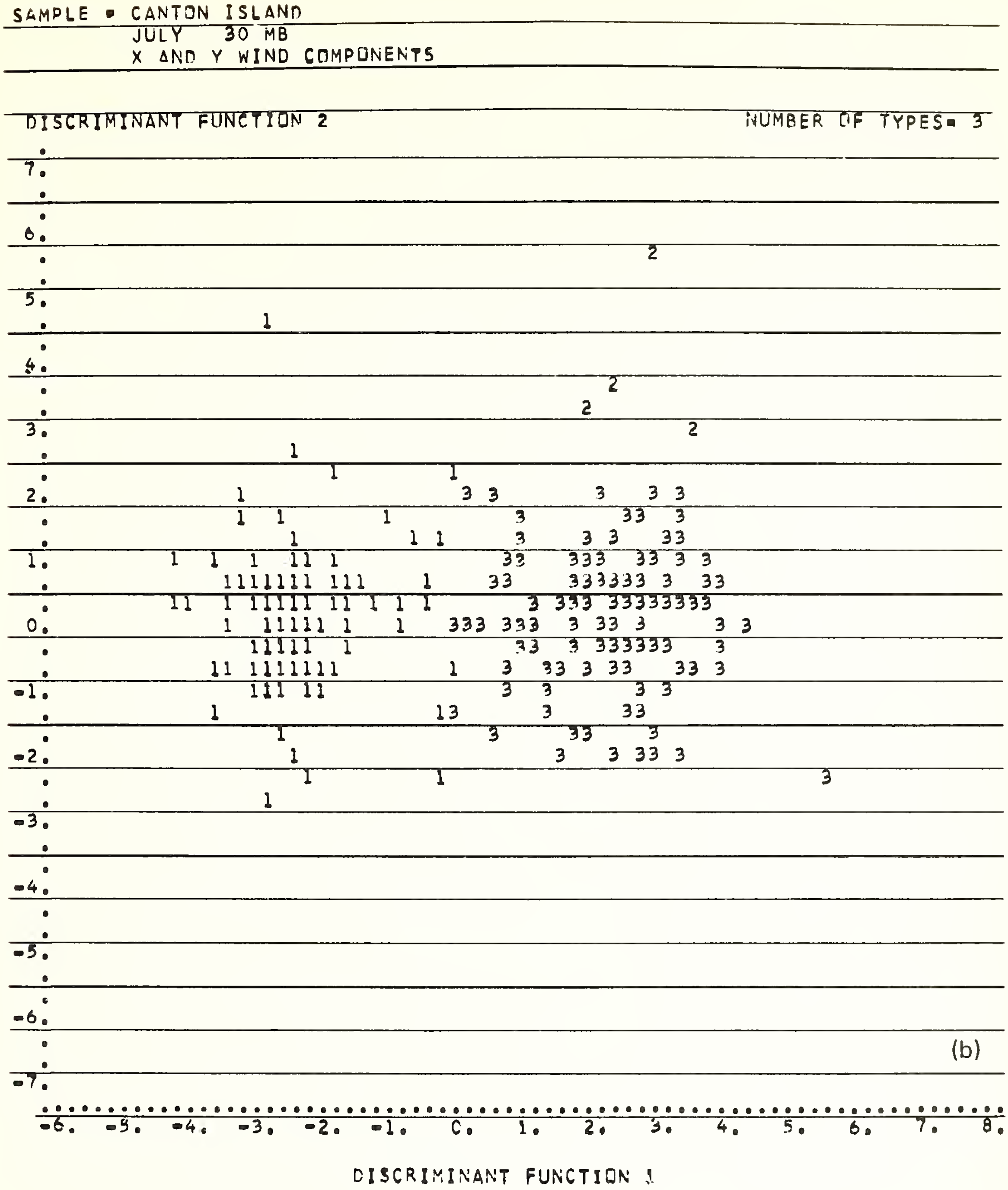


Figure 13 (continued)

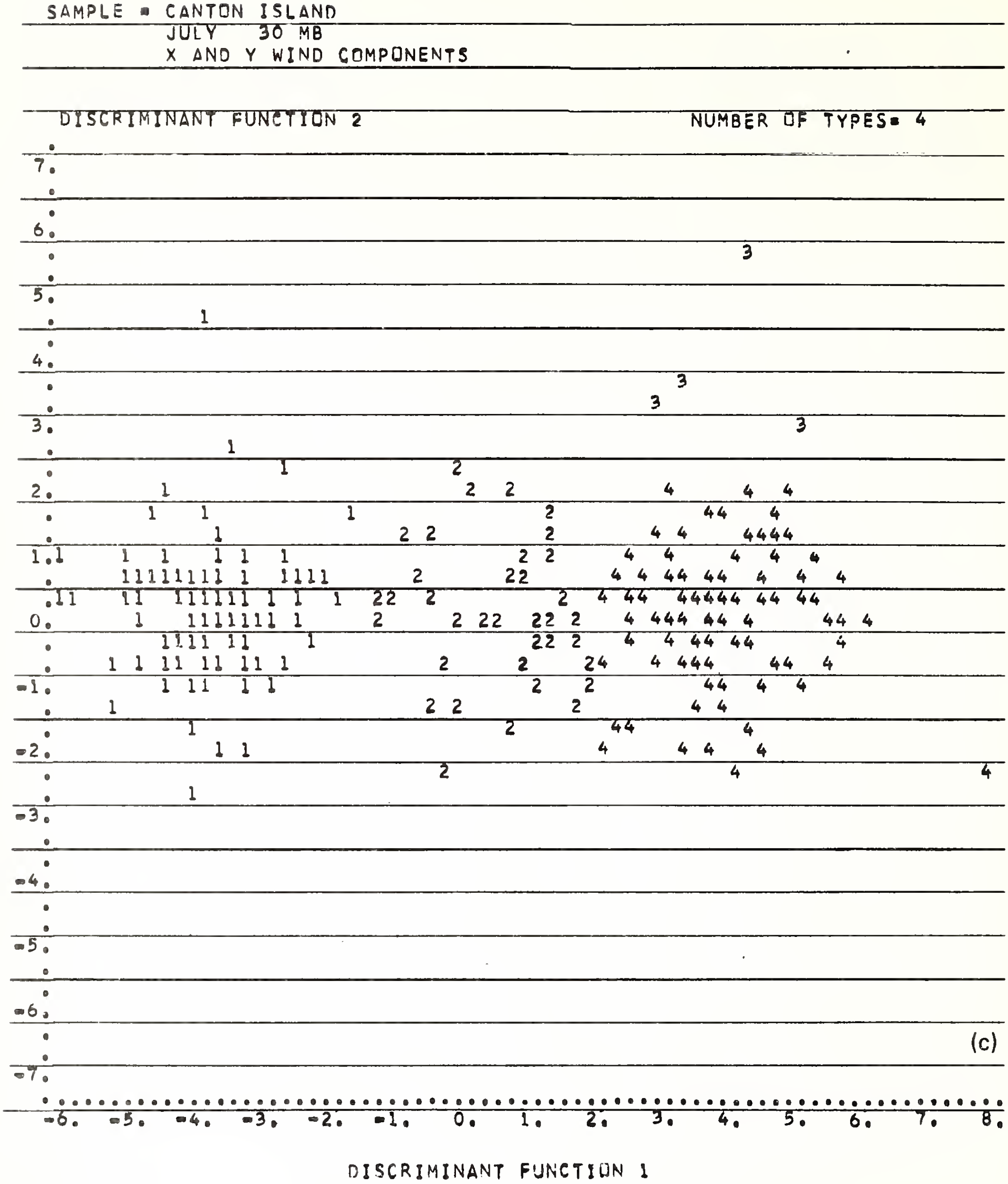


Figure 13 (continued)

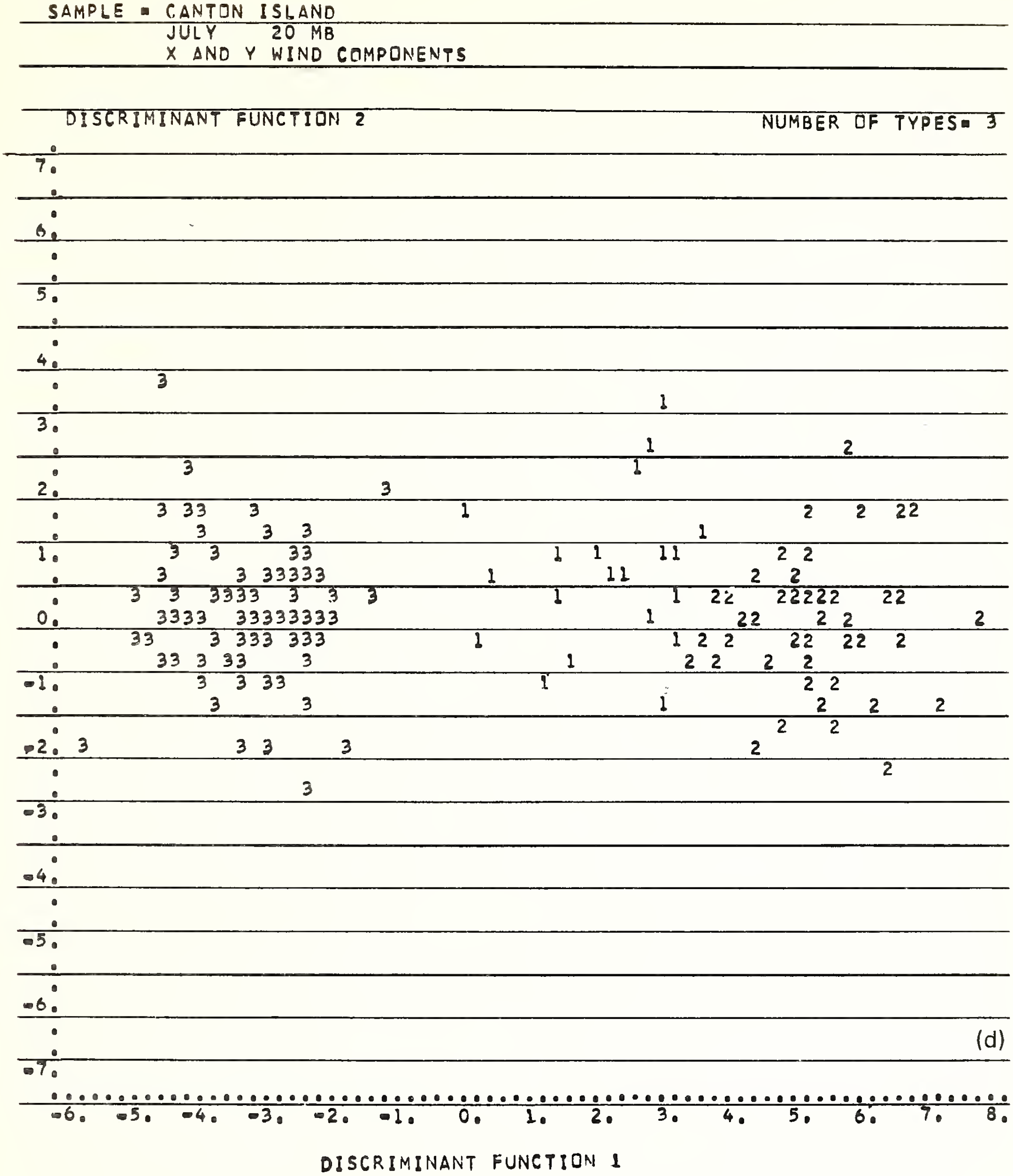


Figure 13 (continued)

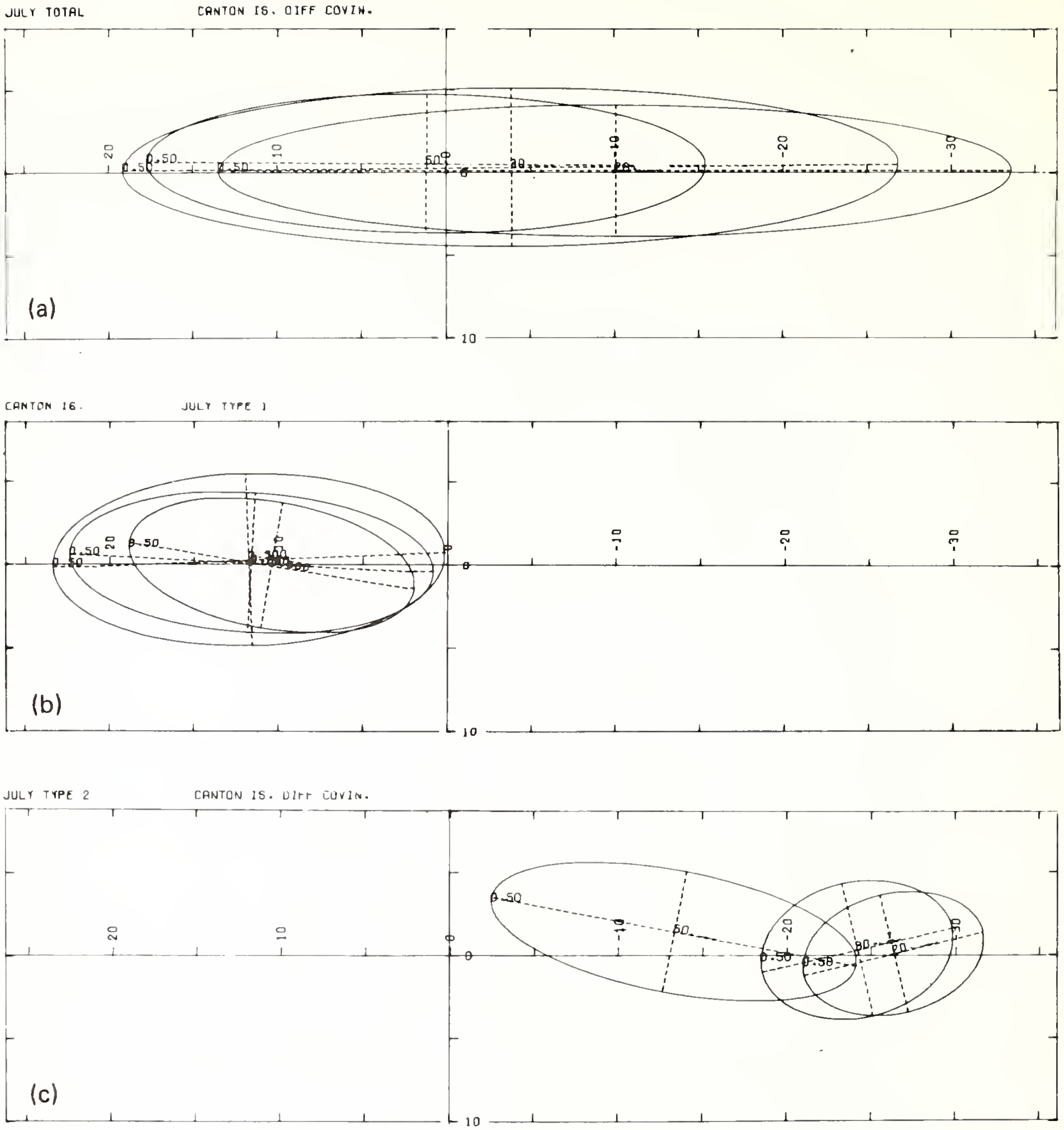


Figure 14 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50-, 30-, and 20-mb; units, m·s^{-1}; n = 263, 244, and 162, respectively, assumption of unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Total for the three levels.
(b) Type 1 for the three levels.
(c) Type 2 for the three levels.

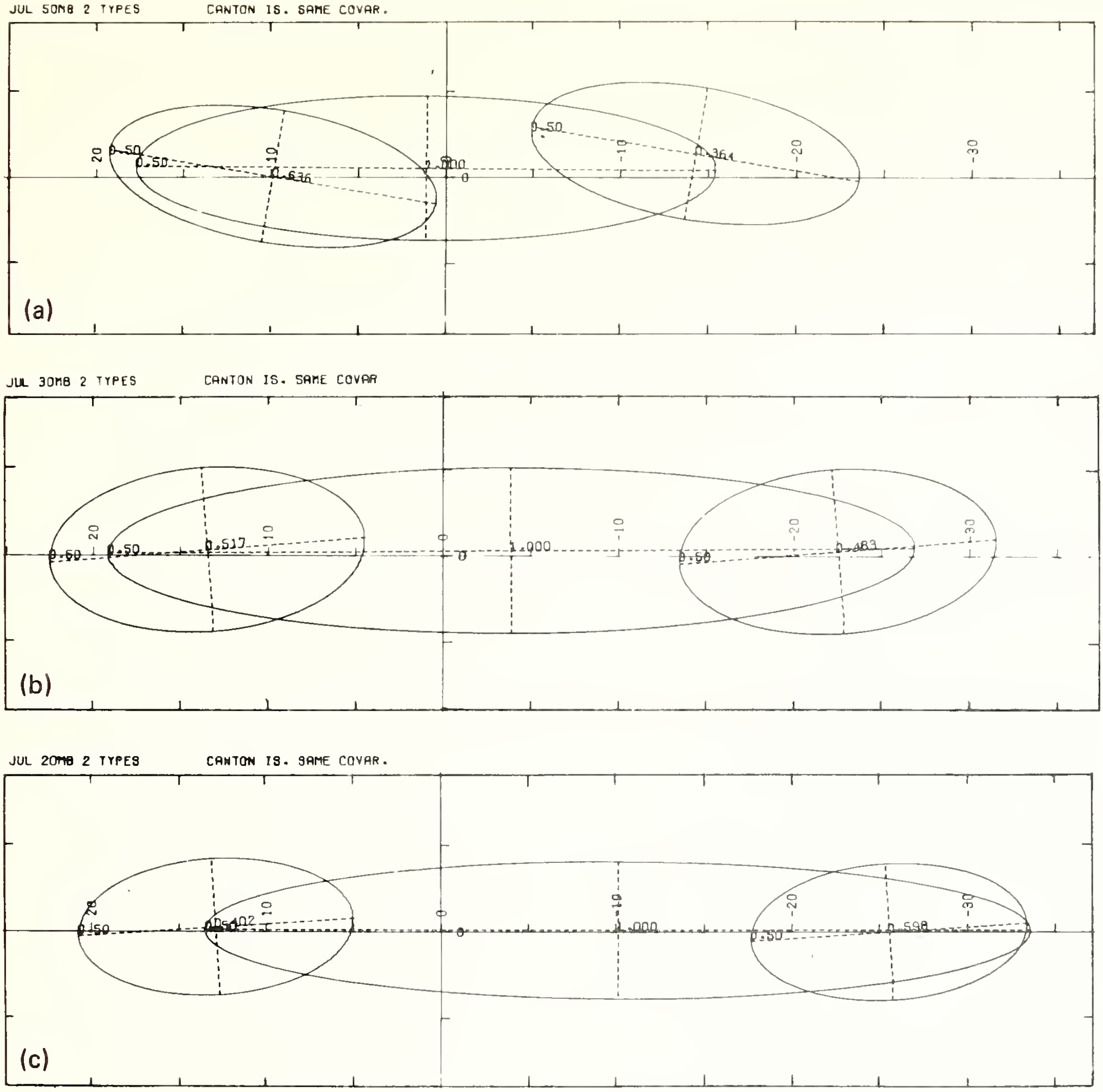


Figure 15 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50-, 30-, and 20-mb; units, $m \cdot s^{-1}$; n = 263, 244, and 162, respectively; separation shows two types; assumption of unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Total and 2 types at 50-mb.
(b) Total and 2 types at 30-mb.
(c) Total and 2 types at 20-mb.

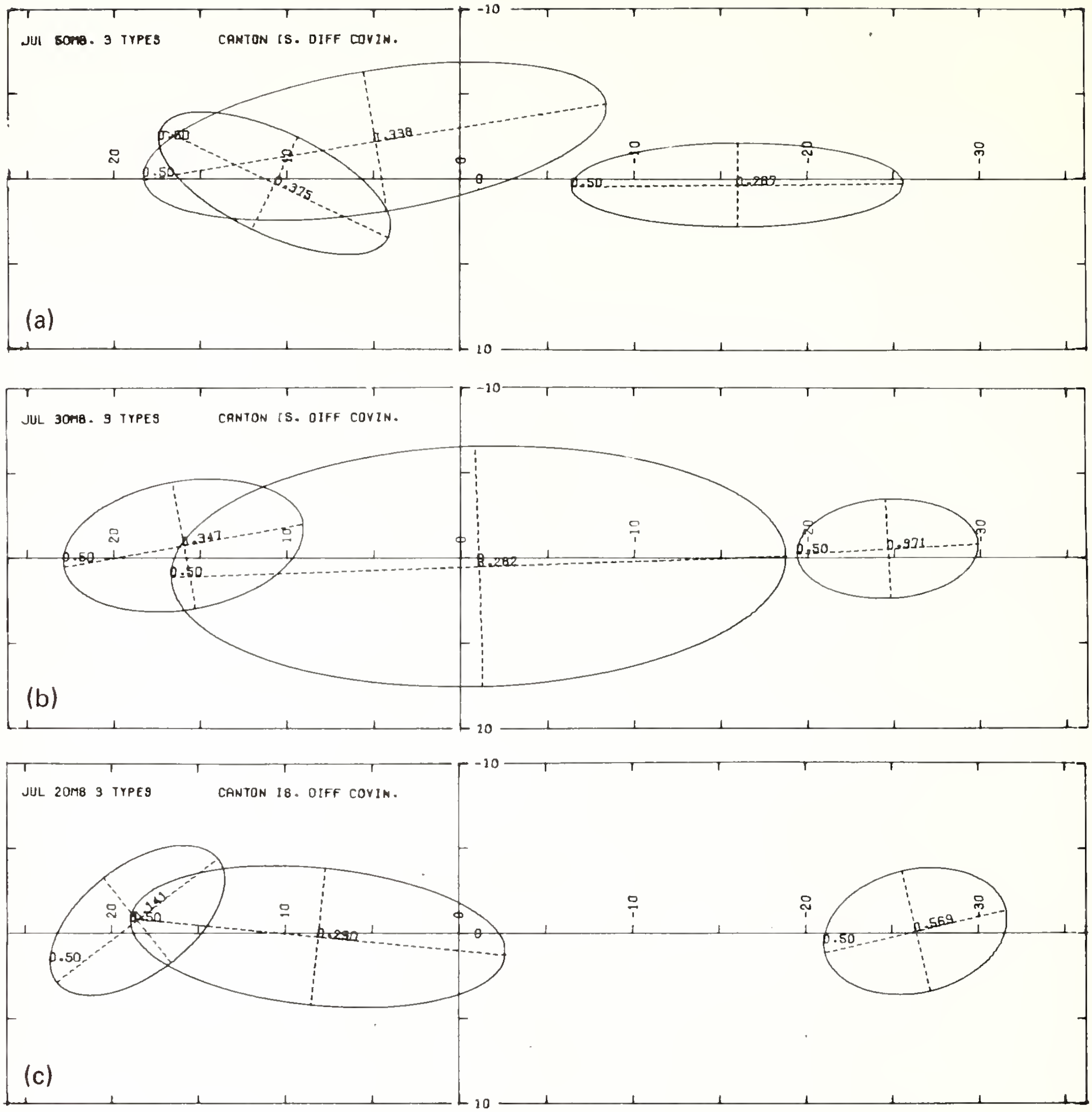


Figure 16 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50-, 30-, and 20-mb; units, $m \cdot s^{-1}$; n = 263, 244, and 162, respectively; separation shows three types, assumption of unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Three types at 50-mb.
(b) Three types at 30-mb.
(c) Three types at 20-mb.

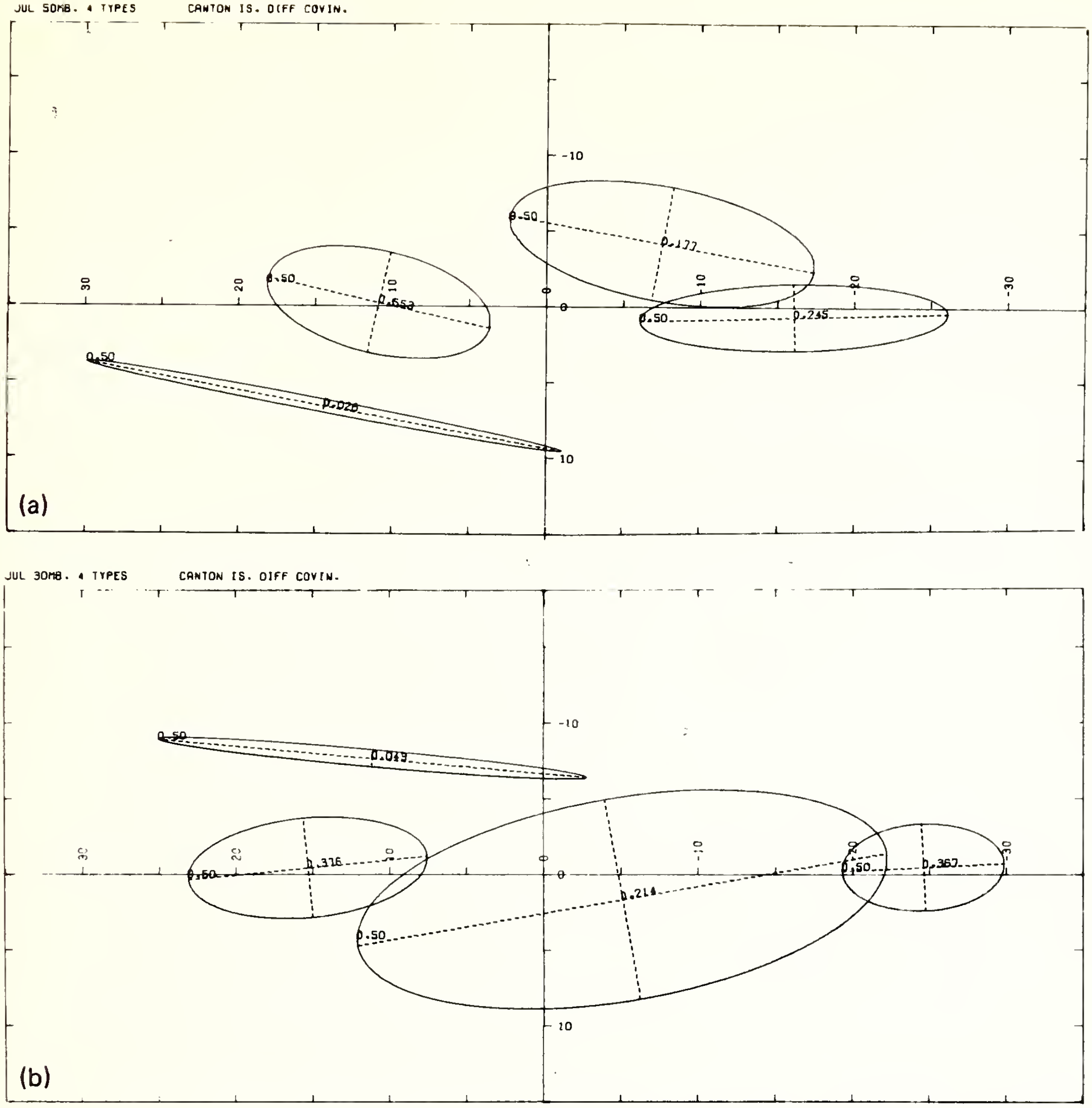


Figure 17 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record July 1954-1964; pressure levels, 50- and 30-mb; units, $m \cdot s^{-1}$; n = 263 and 244, respectively; separation shows four types, assumption of unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Four types at 50-mb.
(b) Four types at 30-mb.

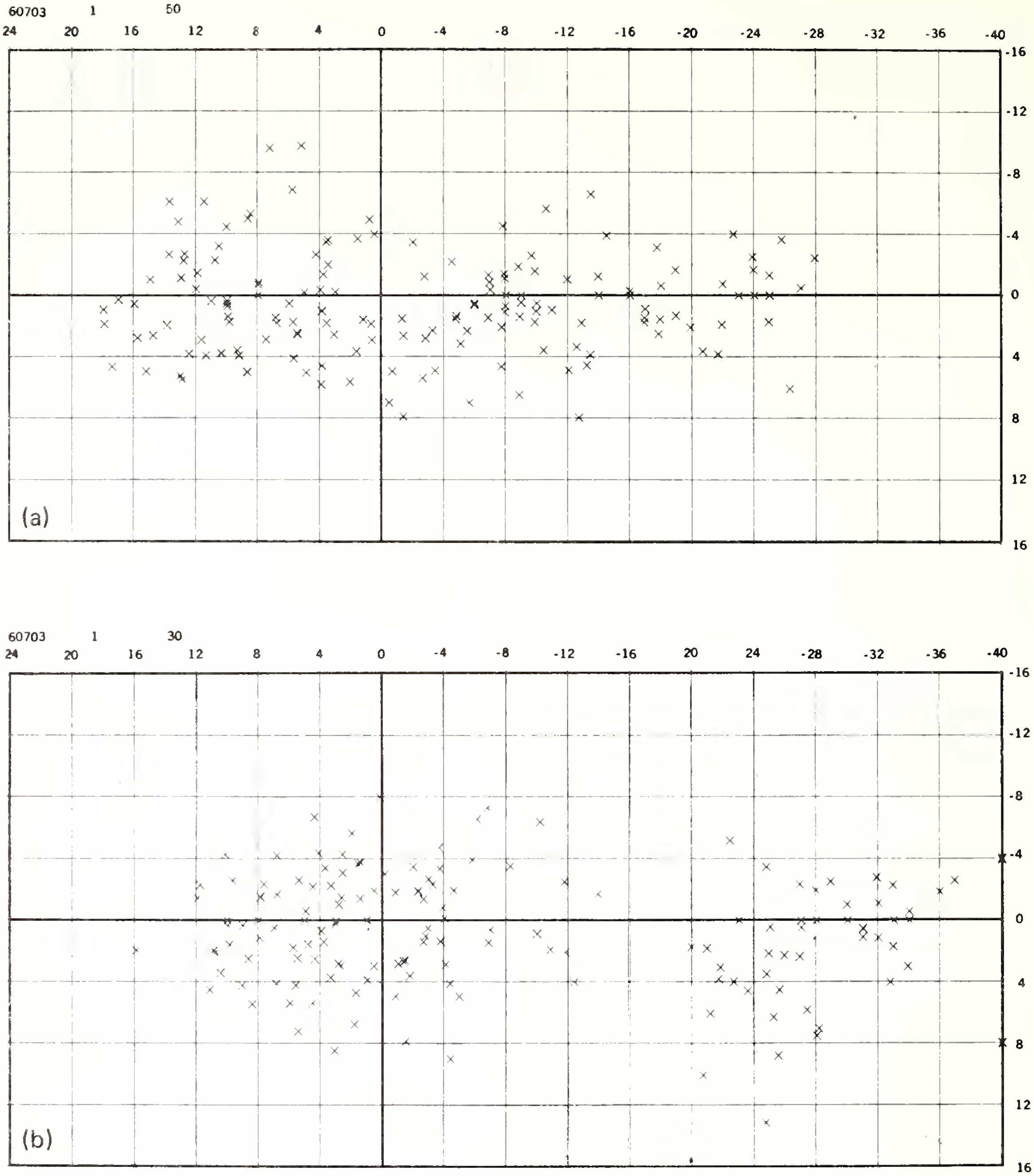
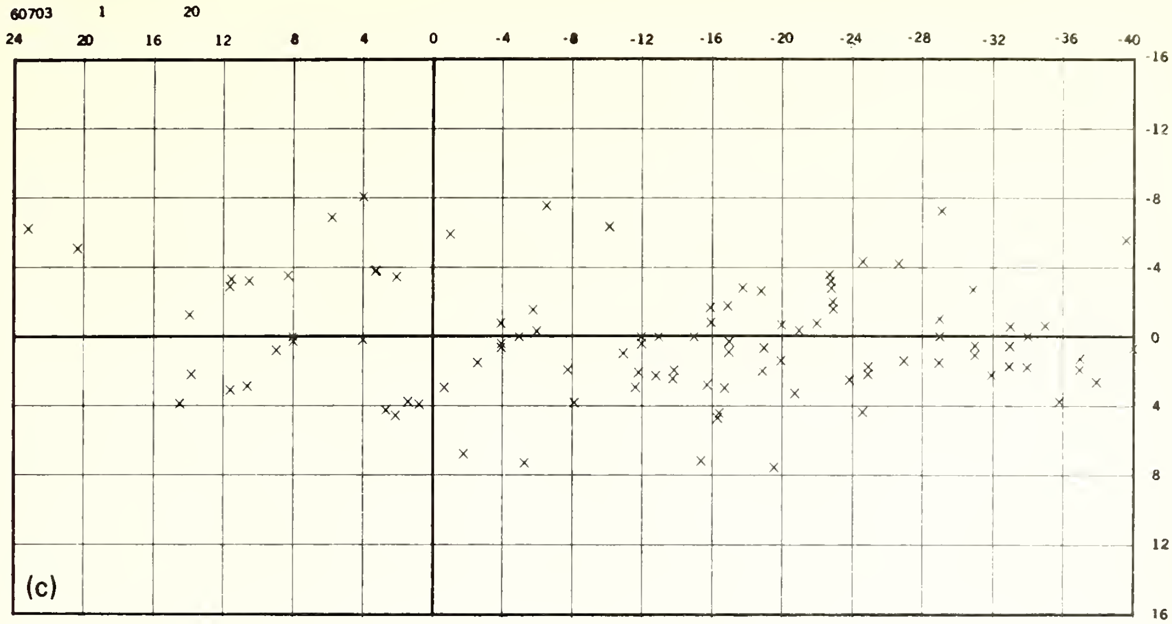


Figure 18 Canton Island, U.S.A. and U.K.; upper wind distribution plots; period of record January 1954-1964; pressure levels, (a) 50-, (b) 30-, (c) 20-, and (d) 10-mb; units, $m \cdot s^{-1}$.



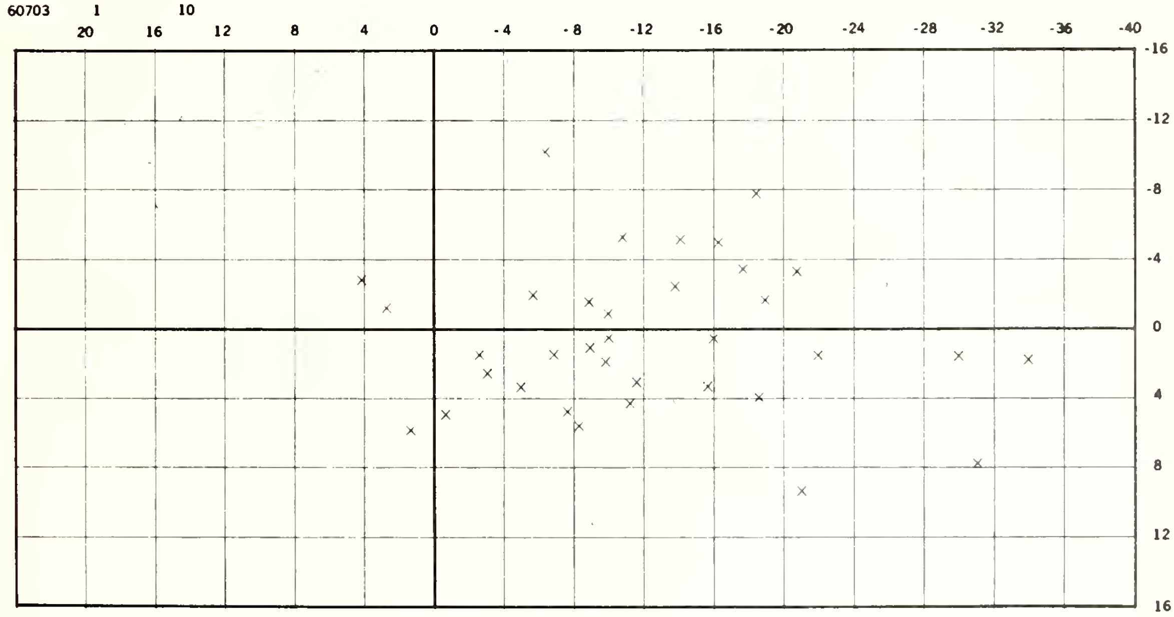


Figure 18 (continued)

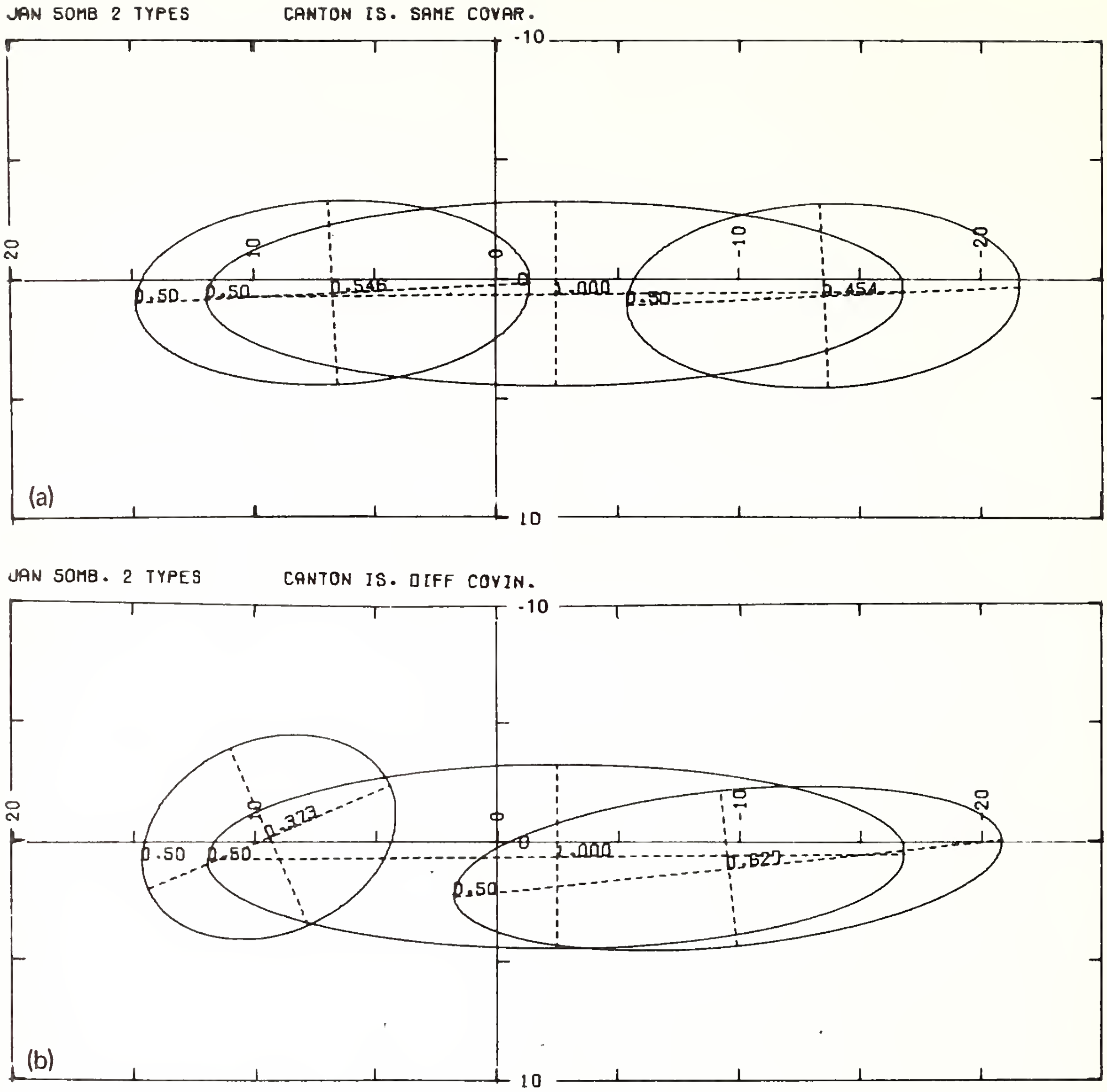


Figure 19 Canton Island, U.S.A. and U.K.; upper wind distributions; period of record January 1954-1964; pressure level, 50-mb; wind plot shown in figure 18; units, $m \cdot s^{-1}$; separation shows for two and three types with assumption of equal then unequal covariance matrices; elliptical error probable, (e.e.p.), 0.50 probability ellipses.

(a) Total and two types with equal covariances.
(b) Total and two types with unequal covariances.
(c) Three types with equal covariances.
(d) Three types with unequal covariances.

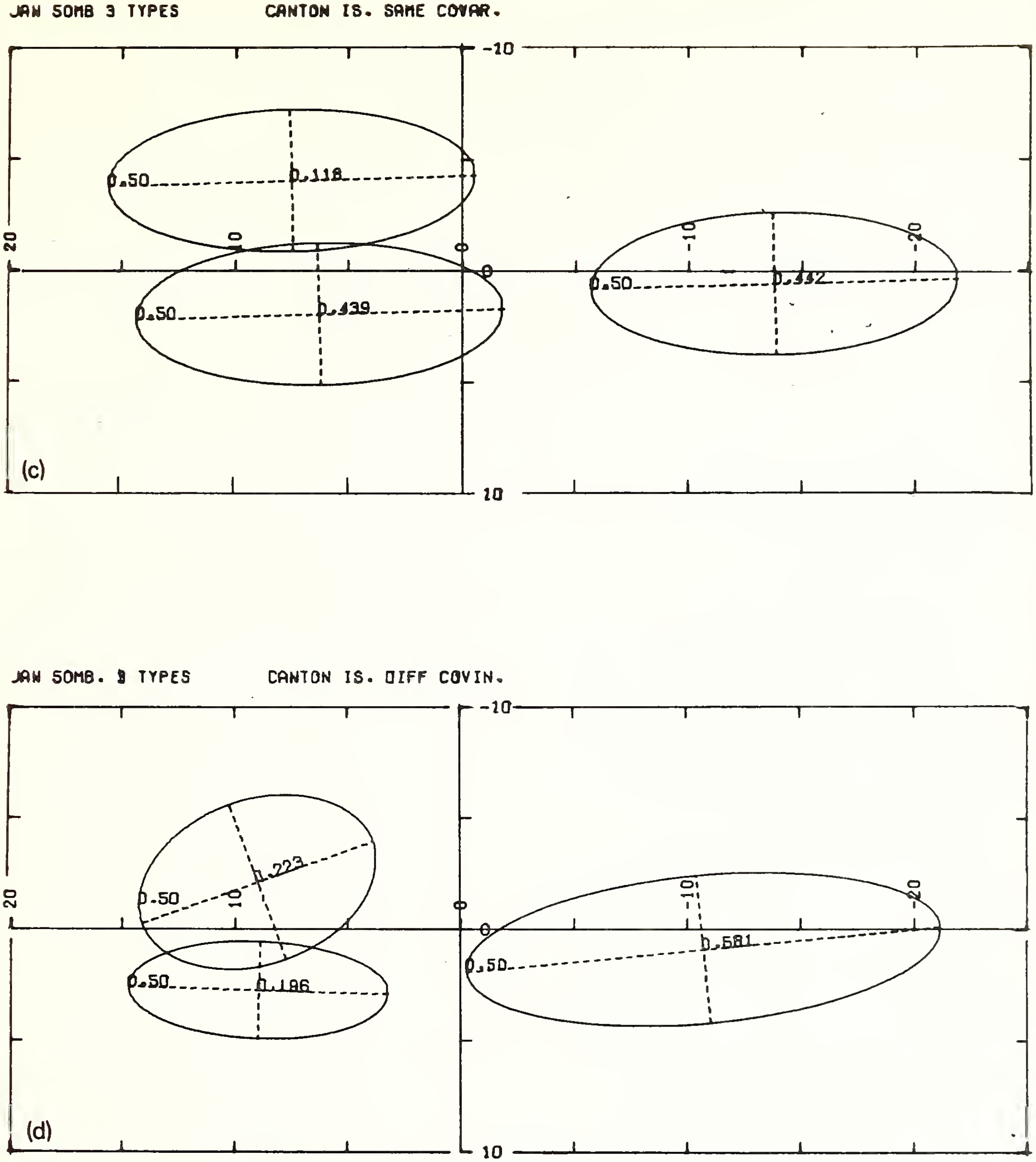


Figure 19 (continued)

8.3.2.4 Tables and Discussions for Height, Temperature and Wind Configuration (1957-1967). The foregoing sections discussed wind configurations only for the 50-, 30-, 20-, and 10-mb levels for January and July 1954-1964. The height, temperature, and wind configurations now are discussed for the 30-mb level during the months of January and July 1957-1967.

Veryard and Ebdon (1961b) in their study of the quasi-biennial oscillation with tropical stratospheres reported that the westerly regimes were warmer than the easterlies. There was a slight lag in the temperatures behind the wind changes. Most of the data examined were for the 80- to 50-mb levels. Substantiating this feature for July but not for January are the statistics presented in tables 11 and 13 for the 30-mb level. Here, the temperature for the westerlies and easterlies was -56.8 and -56.7 during January and -51.0 and -55.0 during July. In a further breakdown, not illustrated here, one cluster in January has a temperature of -59.1 though the westerly component for the cluster is only +4.2 $m \cdot s^{-1}$.

Tables 11 and 13 have been assembled differently from previous material. This permits a better reference for this discussion. For example, there are essentially only three clusters for the January clustering and only four clusters for the July clustering. In both months, except for a fourth cluster in July, the total sample breaks down into westerly regimes with slightly greater heights than the easterly regimes; but while there are no temperature differences in January, there are differences in July.

In January, the easterly cluster characteristics remain essentially fixed in all characteristics while the more variable westerly cluster breaks down into two clusters and then three. The last breakdown is not significant.

During July, the easterly cluster characteristics remain essentially fixed throughout subsequent breakdown of the distribution. The westerly group breaks into three significantly different groups. The difference between the most westerly group and the first easterly group is 41 $m \cdot s^{-1}$.

Examination of the standard deviations is revealing during both months. The ratios of the major axis to minor axis deviation ranges from about five for the total sample to two for the clusters.

A look at tables 12 and 14 is interesting. These tables present the correlation coefficients among the four variables, height, temperature, zonal wind component, and meridional wind component. It is difficult to assess the degrees of freedom in each case. From the QBO basis, a January or a July is expected to be consistent within itself. Therefore, a maximum of eleven points are available for the 1957-1967 period. The correlation coefficients are expected to be large, large enough to be significant. This appears to be the case for July but not for January except for the heights and the zonal winds.

The heights and the zonal winds appear to be significantly correlated for all cluster breakouts except one in each month where the correlation shifts to the meridional winds. During January there is no definite correlation between heights and temperatures at the 30-mb level. During July there is a definite correlation. This does not indicate that there is no correlation during January or July between the heights and the temperature structure between the surface and 30 mb. This has not been examined here.

The large correlation coefficients and the large ratio of standard deviations for the total samples may be considered to be a necessary indication but not a sufficient basis for the conclusion of a two-cluster existence. Conversely, small correlation coefficients and ratio of standard deviations of one are indications of only one cluster, i.e., the total sample is a homogeneous cluster.

Table 11 January upper air statistics for Canton Island, U.S.A and U.K. Units are m, °C and m·s^{-1}. The period of record is the months of January during 1957-1967. The pressure level is 30-mb. The sample size is 244. The assumption is that the covariance matrices are not the same. The variables are (1) height, (2) temperature, (3) the zonal wind component, positive from the west, and (4) the meridional wind component, positive from the south.

Logarithm of the likelihood;	chi-square with 28 d.f.;	probability of null hypothesis.
(a) 2 to 1 types 85.949209	(a) 168.02	(a) 0.00000000
(b) 3 to 2 types 39.701534	(b) 77.45	(b) 0.00000159
(c) 4 to 3 types 14.766845	(c) 28.75	(c) 0.42544651*

Cluster	Proportion	AVG (1) m	AVG (2) °C	AVG (3) m·s^{-1}	AVG (4) m·s^{-1}	Std. Dev. m	Std. Dev. °C	Std. Dev. m·s^{-1}	Std. Dev. m·s^{-1}
1	1.000	23,809.5	-56.7	-5.9	+0.0	83.7	2.5	17.1	3.7
1	0.662	23,842.5	-56.8	+5.3	-0.2	74.6	2.6	7.3	3.8
2	0.338	23,744.8	-56.7	-27.9	+0.4	61.4	2.2	5.6	3.5
1	0.305	23,782.2	-57.3	+0.6	+0.0	53.7	3.2	6.7	4.6
2	0.361	23,892.9	-56.3	+9.0	-0.4	46.7	1.8	5.7	3.0
3	0.334	23,744.1	-56.7	-28.0	+0.4	61.4	2.2	5.5	3.5
1	0.181	23,777.4	-56.2	-0.6	-1.0	31.1	3.4	7.1	4.4
2	0.313	23,894.4	-55.8	+9.0	-0.9	45.6	1.5	5.9	2.8
3	0.172	23,815.2	-59.1	+4.2	+1.9	80.9	1.6	6.1	3.9
4	0.333	23,744.0	-56.7	-28.1	+0.4	61.4	2.2	5.4	3.5

Table 12 January correlation coefficients for data shown in table 11. The variables are (1) height, (2) temperature, (3) zonal wind, and (4) meridional wind at 30 mb. The degrees of freedom (d.f.) are estimated as n/2 where n is the number of observations in each cluster except the total sample where the degrees of freedom are estimated as the number of Januarys. The significance level is 0.05 taken from Snedecor and Cochran (1967). An underlined number indicates significance. Correlation coefficients r_{ij}, i,j = 1,2,3,4. d.f. $r_{.05}$

Cluster	r_{11}	r_{12}	r_{13}	r_{14}	r_{22}	r_{23}	r_{24}	r_{33}	r_{34}	r_{44}	d.f.	$r_{.05}$
1	1.000	0.242	$\underline{+0.612}$	+0.104	1.000	-0.007	-0.199	1.000	-0.007	1.000	11	0.553
1	1.000	$\underline{+0.318}$	$\underline{+0.304}$	+0.142	1.000	-0.010	$\underline{-0.264}$	1.000	+0.148	1.000	81	0.216
2	1.000	+0.228	$\underline{+0.351}$	+0.281	1.000	+0.090	-0.041	1.000	+0.193	1.000	41	0.302
1	1.000	+0.338	-0.087	$\underline{+0.343}$	1.000	$\underline{-0.313}$	-0.286	1.000	+0.193	1.000	37	0.315
2	1.000	+0.174	$\underline{-0.335}$	+0.175	1.000	+0.100	-0.214	1.000	$\underline{+0.276}$	1.000	44	0.291
3	1.000	+0.219	$\underline{+0.341}$	$\underline{+0.294}$	1.000	+0.062	-0.027	1.000	+0.223	1.000	41	0.302
1	1.000	$\underline{+0.415}$	-0.001	+0.110	1.000	-0.304	-0.374	1.000	$\underline{+0.420}$	1.000	22	0.404
2	1.000	+0.050	-0.298	+0.214	1.000	+0.038	-0.071	1.000	$\underline{+0.417}$	1.000	33	0.334
3	1.000	$\underline{+0.543}$	+0.052	$\underline{+0.504}$	1.000	-0.018	$\underline{+0.411}$	1.000	-0.333	1.000	21	0.413
4	1.000	+0.222	$\underline{+0.338}$	$\underline{+0.294}$	1.000	+0.072	-0.016	1.000	+0.216	1.000	41	0.302

Table 13 July upper air statistics for Canton Island, U.S.A. and U.K. Units are m, °C, and m·s^{-1}. The period of record is the months of July during 1957-1967. The pressure level is 30 mb. Sample size is 244. The assumption is that the covariance matrices are not the same. The variables are the (1) height, (2) temperature, (3) the zonal wind component, positive from the west, and (4) the meridional component, positive from the south.

Logarithm of the likelihood; chi-square with 28 d.f.; probability of null hypothesis.

(a) 2 to 1 types 111.95569	(a) 219.32	(a) 0.00000000
(b) 3 to 2 types 43.53996	(b) 85.13	(b) 0.00000011
(c) 4 to 3 types 28.459438	(c) 55.54	(c) 0.00146371
(d) 5 to 4 types 16.638797	(d) 32.41	(d) 0.25821093*

Cluster	Proportion	AVG (1) m	AVG (2) °C	AVG (3) m·s^{-1}	AVG (4) m·s^{-1}	Std. Dev. m	Std. Dev. °C	Std. Dev. m·s^{-1}	Std. Dev. m·s^{-1}
1	1.000	23,997.3	-52.7	-2.6	0.1	69.7	3.2	19.0	3.7
1	0.581	24,038.9	-51.0	11.9	-0.2	45.5	2.9	9.7	4.2
2	0.419	23,939.5	-55.0	-22.7	0.6	57.0	1.9	5.7	2.9
1	0.473	24,041.2	-51.4	15.4	0.4	45.8	2.6	6.3	3.8
2	0.108	24,028.3	-49.3	-3.7	-3.0	42.3	3.2	7.2	4.8
3	0.419	23,939.7	-55.0	-22.6	0.7	57.3	1.8	5.9	2.9
1	0.201	24,019.2	-50.4	18.2	0.4	32.5	1.9	4.6	4.2
2	0.252	24,056.7	-52.1	14.0	0.5	48.1	2.7	6.1	3.4
3	0.153	24,030.7	-50.5	-5.0	-1.6	44.3	3.7	9.2	5.4
4	0.393	29,934.9	-55.1	-23.0	0.5	54.8	1.8	5.7	2.5
1	0.202	24,019.3	-50.4	18.2	0.4	32.5	1.9	4.5	4.2
2	0.248	24,056.0	-52.1	14.0	0.6	47.8	2.7	6.2	3.4
3	0.138	24,035.8	-50.0	-2.5	-2.6	45.6	3.6	8.1	4.6
4	0.144	23,971.3	-54.9	-19.2	0.5	47.2	1.3	5.9	3.9
5	0.268	23,920.4	-55.1	-24.8	0.8	52.7	2.1	4.4	2.1

Table 14 July correlation coefficients for data shown in table 13. The variables are (1) height, (2) temperature, (3) zonal wind, and (4) meridional wind at 30-mb. The degrees of freedom are estimated as n/2 where n is the number of observations in each cluster except the total sample where the degrees of freedom are estimated as the number of Julys. The significance level is 0.05 taken from Snedecor and Cochran (1967). An underlined number indicates significance.

Cluster	r_{11}	r_{12}	r_{13}	r_{14}	r_{22}	r_{23}	r_{24}	r_{33}	r_{34}	r_{44}	d.f.	$r_{.05}$
1	1.000	$\underline{+0.602}$	$\underline{+0.696}$	-0.009	1.000	$\underline{0.538}$	-0.034	1.000	-0.078	1.000	11	0.553
1	1.000	$\underline{+0.279}$	+0.158	+0.116	1.000	-0.044	+0.033	1.000	+0.075	1.000	70	0.232
2	1.000	$\underline{+0.359}$	$\underline{+0.316}$	+0.066	1.000	-0.020	+0.068	1.000	-0.058	1.000	52	0.269
1	1.000	$\underline{+0.420}$	+0.015	+0.136	1.000	$\underline{+0.349}$	+0.036	1.000	-0.214	1.000	57	0.257
2	1.000	-0.705	$\underline{+0.605}$	-0.057	1.000	-0.174	+0.425	1.000	-0.314	1.000	13	0.514
3	1.000	$\underline{+0.351}$	$\underline{+0.325}$	+0.074	1.000	-0.047	+0.086	1.000	-0.037	1.000	52	0.269
1	1.000	+0.180	-0.085	$\underline{+0.379}$	1.000	-0.277	+0.289	1.000	+0.070	1.000	25	0.381
2	1.000	$\underline{+0.832}$	$\underline{+0.366}$	-0.005	1.000	$\underline{+0.446}$	-0.137	1.000	$\underline{-0.525}$	1.000	30	0.349
3	1.000	+0.037	+0.536	+0.034	1.000	+0.130	+0.049	1.000	$\underline{+0.431}$	1.000	19	0.433
4	1.000	$\underline{+0.337}$	$\underline{+0.279}$	-0.065	1.000	+0.049	+0.030	1.000	-0.084	1.000	48	0.276
1	1.000	+0.177	-0.090	$\underline{+0.375}$	1.000	-0.275	+0.289	1.000	+0.072	1.000	25	0.381
2	1.000	$\underline{+0.838}$	$\underline{+0.366}$	-0.004	1.000	$\underline{+0.461}$	-0.136	1.000	$\underline{-0.530}$	1.000	30	0.349
3	1.000	-0.012	$\underline{+0.637}$	+0.076	1.000	-0.034	+0.281	1.000	-0.181	1.000	17	0.456
4	1.000	+0.257	-0.377	+0.244	1.000	-0.553	-0.026	1.000	+0.218	1.000	17	0.456
5	1.000	$\underline{+0.404}$	$\underline{+0.421}$	+0.047	1.000	+0.128	+0.197	1.000	-0.293	1.000	33	0.335

8.3.3 Mid-Latitude Tropospheric Wind Data Set

8.3.3.1 Input Information.

a. Rantoul, Illinois, U.S.A., latitude 40°18' north, longitude 88°09' west, elevation 227 m.
b. The period of record is the month of October for the years 1950-1955.
c. The data are the zonal and meridional components at the 700-, 500-, and 300-mb levels, $m \cdot s^{-1}$.
d. The number of variables is six.
e. The number in the sample is 503.
f. The minimum number to be accepted into a cluster is seven, one more than the number of variates.
g. The null hypotheses are made that $(k + 1)$ clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of $(k + 1)$ clusters.
h. The first 40 six-dimensional vector entries are set up as the 40 means of 40 separate and individual clusters. These are 40 points in six dimensions.
i. The assumption is made that the covariance matrices are not equal.

8.3.3.2 Tables. Table 15 shows the sequential input data with six variates. The variates are the zonal and meridional components of the 700-mb wind, the zonal and meridional components of the 500-mb wind, and the zonal and meridional components of the 300-mb wind in $m \cdot s^{-1}$.

Table 16 shows the computer output after the hierarchical grouping and the subsequent grouping into two clusters after 39 iterations. The two clusters with proportions of 0.171 and 0.829 comprise the total set (1.000). First, the characteristics of the total group are shown. This would be the assumption of a unimodal six-variate model. Then the characteristics of the two clusters are provided. Then, through the use of the discriminant function, assignments of each datum to the clusters may be made as shown. The probability of assignment to each cluster is printed. For example, the first datum of table 15 is assigned to cluster 2 with a probability of 0.935 versus 0.065 for cluster 1. The previous section illustrated computer plots for discriminant functions.

Table 17 provides, from the same input data, the output for three clusters rather than two. The number of iterations required in this case is 101. The probability of the null hypothesis being rejected is 0.00000349. The three-cluster configuration is judged to be better than a two-cluster configuration. Therefore, the program continues further to test a four-cluster configuration versus the three-cluster form.

Table 18 provides, again from the same input data, the output for the four-cluster configuration. The initial computing output, i.e., the first estimate of the learning process, is shown as the zero iteration. This set then becomes the initial estimation in the iteration process leading to 101 iterations before the results converged. Please note the changes from the initial estimates to the final estimates. For example, for the zeroth iteration, the estimate for the 700-mb zonal component in the first cluster is 11.2762 and after 101 iterations the estimate is 11.1041. The respective standard deviations are 5.5488 and 7.2047. Though the means do not appear to be very different, this is almost a forty percent reduction in variance in the cluster.

The probability of rejection of the null hypothesis in the above case is 0.00980137. This is near the decision level of 0.01. Therefore, no further calculations are shown here.

Table 15 A multivariate (6) set of Rantoul, Illinois, October 1950-55, upper wind components, zonal and meridional, at the 700-, 500-, and 300-mb levels. For example, input variables 1 and 2 are the zonal and meridional components, respectively, at the 700-mb level. The units are m·s^{-1}.

```
           SAMPLE = RANTOUL
                    OCTOBER
                    700,500,AND 300MB
                    X AND Y COMPONENTS


       NVBLES=  6  NSAMPL= 503  DIFE COV MATRIX
       MIN CLUSTER SIZE=  7  HYPOTHESIS TEST=0.000
       INITIAL KMEANS= 40  TIME LIMIT=*0  ITER LIMIT=100
       NO. OF TYPES=  1  2  3  4  5  6  0  0  0  0  0  0  0
       STORAGE REQUIREMENT=  111932 DECIMAL BYTES    .
       DIMENSION A(  3991)


       ANTICIPATED EXECUTION TIME =   21.51 MINUTES

                   PROGRAM NORMIX
     WOLFE NORMAL MIXTURE ANALYSIS PROCEDURE(1974 REVISION)

    INPUT VARIABLES
SEQ         1          2          3          4          5          6
  1     7.071      7.071      2.121     -2.121      5.563      2.248
  2     4.298     10.126      1.954      4.603     10.000      0.000
  3    -0.000     14.000      4.243      4.243      5.523     -2.344
  4    -0.000     10.000     20.398      8.241      8.485     -8.485
  5    -0.000     16.000      1.954      4.603     11.046     -4.689
  6     5.470     12.887     -0.000     11.000      5.657      5.657
  7     4.636      1.873     15.000      0.000     24.107      9.740
  8     7.000      0.000     19.331     -8.205     20.229    -50.068
  9    11.314    -11.314     21.172     -8.987     31.297    -13.285
 10    12.887     -5.470     14.142    -14.142     16.108    -39.869
 11     9.192      9.192     -3.126     -7.364      8.285     -3.517
 12     2.828      2.828      4.636      1.873      2.828      2.828
 13     3.536      3.536      2.735      6.444      7.778      7.778
 14    -1.124      2.782      2.735      6.444      7.778      7.778
 15     1.954      4.603      3.126      7.364     34.306     13.860
 16     0.000      5.000     12.053      4.870     13.285     31.297
 17     4.243     -4.243      4.243     -4.243     29.000      0.000
 18     5.994    -14.835      7.778     -7.778     21.000      0.000
 19    -3.126     -7.364     11.314    -11.314     17.490     -7.424
 20     2.622     -6.490     19.092    -19.092     23.180      9.365
 21     0.000      2.000      8.000      0.000     21.000      0.000
 22     2.000      0.000      3.000      0.000     18.385     18.385
 23     6.490      2.622     11.046     -4.689     30.597     12.362
 24     5.523     -2.344      3.371     -8.345     11.046     -4.689
 25     7.000      0.000     12.000      0.000     21.000      0.000
 26    11.314    -11.314     14.728     -6.252    -36.000      0.000
 27    15.762      6.368     20.000      0.000     20.251     -8.596
 28     4.636      1.873     12.887     -5.470     28.536    -12.113
 29     3.536     -3.536     17.490     -7.424     21.213    -21.213
 30     8.285     -3.517      0.000    -14.000      9.740    -24.107
 31     9.272      3.746     11.046     -4.689      4.870    -12.053
 32    17.000      0.000     11.967     -5.079      6.364     -6.364
 33    11.000      0.000     12.000      0.000     13.908      5.619
 34    16.000      0.000     13.808     -5.861      9.000      0.000
 35    16.544      7.492     15.649     -6.642      5.523     -2.344
 36     9.000      0.000     11.000      0.000      7.364     -3.126
 37     4.636      1.873     12.000      0.000     13.000      0.000
 38     5.657      5.657      8.000      0.000     13.808     -5.861
 39     6.490      2.622      8.345      3.371     18.000      0.000
 40     1.954      4.603      9.272      3.746     19.000      0.000
 41    12.728     12.728     10.000      0.000     20.398      8.241
 42     6.364      6.364      8.485      8.485     18.544      7.492
 43    12.981      5.245     17.617      7.118     12.021     12.021
 44    11.000      0.000     12.053      4.870     19.471      7.867
 45    16.689      6.743     14.835      5.994     22.252      8.991
 46    16.000      0.000     18.000      0.000     32.451     13.111
 47    13.908      5.619     15.762      6.368     26.888     10.864
 48     0.000     -3.000      7.118    -17.617     18.000      0.000
 49     7.071     -7.071      9.900     -9.899      9.192     -9.192
 50     6.364     -6.364     10.607    -10.607     12.887     -5.470
 51     2.248     -5.563      5.619    -13.908      7.118    -17.617
 52     3.371     -8.345     -6.252    -14.728      0.000    -26.000
 53     0.000     -9.000      0.000    -17.000      0.000    -26.000
 54    -4.950     -4.950    -20.506    -20.506     -7.815    -18.410
 55     0.000     -9.000      0.000    -13.000      5.994    -14.835
 56     2.622     -6.490      4.495    -11.126      6.364     -6.364
 57     3.371     -8.345      0.000     -8.000      0.000     -4.000
 58     4.243      4.243      0.000     -5.000      0.000     -5.000
 59     1.124     -2.782      1.498     -3.709    -11.126     -4.495
 60    -1.854     -0.749      0.000     -2.000     -7.417     -2.997
```

Table 15 (continued)

	INPUT VARIABLES					
SEQ	1	2	3	4	5	6
61	-0.921	0.391	-0.921	0.391	-8.345	-3.371
62	0.000	5.000	-2.997	7.417	-0.749	1.854
63	0.000	5.000	2.344	5.523	2.735	6.444
64	0.000	3.000	0.000	2.000	1.414	1.414
65	12.000	0.000	10.000	0.000	15.649	-6.642
66	7.071	7.071	11.000	0.000	11.046	-4.689
67	6.364	6.364	6.000	0.000	5.657	-5.657
68	8.345	3.371	8.345	3.371	11.000	0.000
69	7.417	2.997	9.000	0.000	13.000	0.000
70	8.345	3.371	8.345	3.371	12.053	4.870
71	3.907	9.205	7.417	2.997	12.981	5.245
72	3.907	9.205	7.071	7.071	15.762	6.368
73	3.126	7.364	12.053	4.870	15.000	0.000
74	5.657	5.657	7.071	7.071	13.908	5.619
75	8.345	3.371	9.899	9.900	16.689	6.743
76	8.485	8.485	10.607	10.607	13.435	13.435
77	7.778	-7.778	13.000	0.000	19.331	-8.205
78	11.000	0.000	16.569	-7.033	26.000	0.000
79	12.981	5.245	18.544	7.492	25.000	0.000
80	11.314	11.314	16.689	6.743	27.816	11.238
81	10.607	10.607	13.908	5.619	14.142	14.142
82	9.192	9.192	11.314	11.314	20.398	8.241
83	8.485	8.485	16.263	16.263	30.597	12.362
84	8.485	-8.485	22.092	-9.378	31.297	-13.285
85	4.603	-1.954	17.490	-7.424	21.172	-8.987
86	11.126	4.495	12.000	0.000	22.000	0.000
87	12.981	5.245	13.908	5.619	25.961	10.489
88	8.485	8.485	11.314	11.314	19.471	7.867
89	11.000	0.000	21.000	0.000	35.000	0.000
90	17.000	0.000	25.000	0.000	49.000	0.000
91	9.205	-3.907	20.251	-8.596	38.000	0.000
92	9.000	0.000	17.000	0.000	28.743	11.613
93	-2.622	6.490	13.435	13.435	9.899	9.900
94	6.444	-2.735	4.636	1.873	24.000	0.000
95	4.000	0.000	12.887	-5.470	28.000	0.000
96	13.000	0.000	4.243	-4.243	18.000	0.000
97	16.689	6.743	13.000	0.000	10.126	-4.298
98	11.314	-11.314	18.410	-7.815	24.854	-10.550
99	9.205	-3.907	14.849	-14.849	23.933	-10.159
100	13.808	-5.861	15.649	-6.642	17.490	-7.424
101	13.808	-5.861	14.728	-6.252	13.808	-5.861
102	8.485	-8.485	21.000	0.000	27.816	11.238
103	7.071	-7.071	15.649	-6.642	11.987	-29.670
104	8.485	-8.485	9.192	-9.192	15.649	-6.642
105	6.444	-2.735	13.000	0.000	17.000	0.000
106	6.364	-6.364	7.364	-3.126	0.000	-12.000
107	2.762	-1.172	0.000	-9.000	-14.848	-34.979
108	2.828	-2.828	0.000	-6.000	-7.424	-17.490
109	0.000	-4.000	0.000	-3.000	-4.243	4.243
110	1.498	-3.709	2.762	-1.172	-3.000	0.000
111	0.000	-4.000	4.950	-4.950	10.199	4.121
112	-3.126	-7.364	0.000	-6.000	0.000	-5.000
113	-7.778	-7.778	-4.689	-11.046	-6.364	-6.364
114	-10.607	-10.607	-12.021	-12.021	4.495	-11.126
115	-3.536	-3.536	0.000	-6.000	0.000	-24.000
116	1.498	-3.709	3.746	-9.272	11.314	-11.314
117	3.536	-3.536	11.000	0.000	13.808	-5.861
118	6.444	-2.735	2.762	-1.172	21.000	0.000
119	11.046	-4.689	10.126	-4.298	20.000	0.000
120	6.444	-2.735	9.000	0.000	11.046	-4.689
121	10.126	-4.298	19.000	0.000	33.000	0.000
122	18.000	0.000	18.410	-7.815	17.490	-7.424
123	18.000	0.000	22.000	0.000	28.000	0.000
124	18.000	0.000	27.000	0.000	28.000	0.000
125	13.000	0.000	30.000	0.000	30.000	0.000
126	16.000	0.000	22.092	-9.378	37.000	0.000
127	5.657	-5.657	22.000	0.000	21.000	0.000
128	10.126	-4.298	17.490	-7.424	0.000	-22.000
129	13.808	-5.861	14.728	-6.252	0.000	-28.000
130	0.000	-10.000	5.619	-13.908	15.556	-15.556
131	4.243	-4.243	2.622	-6.490	0.000	-5.000
132	3.682	-1.563	3.000	0.000	-10.000	0.000
133	2.828	-2.828	3.000	0.000	-8.485	8.485
134	1.873	-4.636	4.603	-1.954	6.490	2.622
135	3.000	0.000	9.205	-3.907	9.272	3.746
136	3.000	0.000	3.536	-3.536	7.417	2.997
137	8.285	-3.517	8.285	-3.517	16.689	6.743
138	8.285	-3.517	11.046	-4.689	4.950	-4.950
139	11.967	-5.079	11.314	-11.314	4.495	-11.126
140	0.000	-9.000	0.000	-13.000	-16.411	-38.661
141	0.000	-8.000	0.000	-16.000	-33.941	-33.941
142	9.205	-3.907	5.245	-12.981	4.870	-12.053
143	8.485	-8.485	6.743	-16.689	14.235	-35.233
144	12.887	-5.470	7.778	-7.778	11.314	-11.314
145	12.887	-5.470	16.000	0.000	15.649	-6.642
146	7.778	-7.778	16.569	-7.033	27.615	-11.722
147	5.657	-5.657	11.314	-11.314	26.695	-11.331
148	4.243	-4.243	14.728	-6.252	23.013	-9.768
149	8.285	-3.517	10.126	-4.298	26.000	0.000
150	10.199	4.121	14.000	0.000	24.107	9.740
151	17.000	0.000	20.251	-8.596	22.000	0.000

Table 15 (continued)

	INPUT VARIABLES					
SEQ	1	2	3	4	5	6
152	17.000	0.000	14.728	-6.252	30.000	0.000
153	12.728	-12.728	12.021	-12.021	27.615	-11.722
154	15.649	-6.642	8.616	-21.325	10.864	-26.888
155	14.728	-6.252	13.435	-13.435	21.213	-21.213
156	13.808	-5.861	17.490	-7.424	21.172	-8.987
157	13.908	5.619	15.762	6.368	26.000	0.000
158	14.000	0.000	14.000	0.000	19.000	0.000
159	18.410	-7.815	17.490	-7.424	17.000	0.000
160	2.622	-6.490	6.444	-2.735	15.649	-6.642
161	-2.344	-5.523	4.950	-4.950	13.808	-5.861
162	-6.490	-2.622	3.536	-3.536	18.410	-7.815
163	-7.364	3.126	3.000	0.000	9.900	-9.899
164	-1.873	4.636	6.490	2.622	11.967	-5.079
165	1.172	2.762	3.536	-3.536	11.046	-4.689
166	1.854	0.749	3.709	1.498	12.887	-5.470
167	4.243	4.243	4.243	4.243	17.490	-7.424
168	2.344	5.523	4.603	-1.954	16.569	-7.033
169	3.536	3.536	6.490	2.622	11.967	-5.079
170	5.657	5.657	10.199	4.121	20.000	0.000
171	7.071	7.071	11.126	4.495	24.000	0.000
172	7.071	7.071	5.079	11.967	16.000	0.000
173	7.071	7.071	16.689	6.743	24.107	9.740
174	12.000	0.000	15.000	0.000	31.000	0.000
175	3.000	0.000	10.000	0.000	25.961	10.489
176	6.000	0.000	13.908	5.619	31.524	12.737
177	8.000	0.000	20.000	0.000	31.524	12.737
178	8.285	-3.517	9.205	-3.907	18.000	0.000
179	8.485	-8.485	0.707	-0.707	7.778	7.778
180	13.435	-13.435	26.000	0.000	29.456	-12.503
181	12.021	-12.021	29.456	-12.503	28.991	-28.991
182	0.000	-15.000	10.114	-25.034	20.229	-50.068
183	0.000	-14.000	11.238	-27.815	16.483	-40.796
184	5.245	-12.981	10.489	-25.961	22.627	-22.627
185	4.495	-11.126	8.485	-8.485	19.799	-19.799
186	3.746	-9.272	6.743	-16.689	10.114	-25.034
187	8.285	-3.517	12.887	-5.470	20.251	-8.596
188	7.364	-3.126	12.000	0.000	21.000	0.000
189	7.000	0.000	13.000	0.000	19.000	0.000
190	6.490	2.622	9.272	3.746	9.192	9.192
191	3.536	-3.536	11.314	-11.314	19.799	-19.799
192	13.808	-5.861	12.728	-12.728	10.114	-25.034
193	7.000	0.000	8.000	0.000	7.364	-3.126
194	5.563	2.248	10.000	0.000	9.205	-3.907
195	7.364	-3.126	11.967	-5.079	9.205	-3.907
196	4.950	-4.950	0.000	-10.000	-11.331	-26.695
197	5.000	0.000	0.000	-14.000	0.000	-22.000
198	0.000	-5.000	4.870	-12.053	0.000	-14.000
199	7.364	-3.126	10.607	-10.607	9.900	-9.899
200	4.950	-4.950	11.314	-11.314	16.263	-16.263
201	13.808	-5.861	19.331	-8.205	8.991	-22.252
202	7.778	-7.778	23.335	-23.334	24.749	-24.749
203	5.619	-13.908	9.365	-23.180	8.991	-22.252
204	0.000	-17.000	17.000	0.000	13.435	-13.435
205	-5.080	-11.967	0.000	-12.000	13.808	-5.861
206	0.000	-7.000	7.364	-3.126	16.000	0.000
207	-2.344	-5.523	2.997	-7.417	13.000	0.000
208	-3.907	-9.205	2.622	-6.490	14.000	0.000
209	-9.192	-9.192	0.000	-13.000	8.285	-3.517
210	-6.364	-6.364	0.000	-7.000	7.364	-3.126
211	-5.657	-5.657	-8.485	-8.485	7.778	-7.778
212	-6.364	-6.364	-3.517	-8.285	9.205	-3.907
213	-7.071	-7.071	-3.536	3.536	6.364	-6.364
214	2.622	-6.490	1.873	-4.636	-8.205	-19.331
215	2.622	-6.490	0.000	-9.000	-8.987	-21.172
216	7.778	-7.778	-4.298	-10.126	-16.971	-16.971
217	3.746	-9.272	-3.907	-9.205	-13.435	-13.435
218	4.495	-11.126	0.000	-6.000	-12.021	-12.021
219	2.997	-7.417	-3.517	-8.285	-4.298	-10.126
220	0.000	-10.000	0.000	-5.000	-3.517	-8.285
221	-1.841	0.781	1.498	-3.709	-1.124	2.782
222	-1.414	-1.414	1.414	-1.414	2.735	6.444
223	2.121	-2.121	3.682	-1.563	3.126	7.364
224	3.536	-3.536	5.523	-2.344	5.657	5.657
225	4.603	-1.954	7.000	0.000	4.243	4.243
226	5.000	0.000	8.000	0.000	9.272	3.746
227	6.000	0.000	3.709	1.498	7.778	7.778
228	2.344	5.523	8.345	3.371	9.000	0.000
229	4.243	4.243	4.950	4.950	9.000	0.000
230	6.490	2.622	5.523	-2.344	20.000	0.000
231	7.417	2.997	0.921	-0.391	16.000	0.000
232	-0.391	-0.920	3.000	0.000	18.410	-7.815
233	0.000	3.000	-2.000	0.000	15.649	-6.642
234	2.735	6.444	0.000	3.000	12.021	-12.021
235	0.000	2.000	3.000	0.000	10.607	-10.607
236	-2.622	6.490	1.414	-1.414	4.870	-12.053
237	0.000	7.000	-1.000	0.000	-4.689	-11.046
238	-0.000	9.000	-1.873	4.636	0.000	-12.000
239	-3.746	9.272	-1.873	4.636	0.000	-10.000
240	-2.997	7.417	-0.921	0.391	10.607	-10.607
241	0.391	0.921	1.414	1.414	12.728	-12.728
242	0.000	5.000	4.495	-11.126	23.933	-10.159

Table 15 (continued)

INPUT VARIABLES

SEQ	1	2	3	4	5	6
243	4.950	4.950	12.000	0.000	19.331	-8.205
244	11.126	4.495	14.000	0.000	26.000	0.000
245	10.000	0.000	21.000	0.000	24.000	0.000
246	15.000	0.000	19.000	0.000	22.092	-9.378
247	10.607	-10.607	23.335	-23.334	53.389	-22.662
248	7.492	-18.544	10.489	-25.961	20.506	-20.506
249	0.000	-14.000	7.492	-18.544	16.971	-16.971
250	0.000	-12.000	7.118	-17.617	15.556	-15.556
251	3.746	-9.272	8.485	-8.485	18.410	-7.815
252	3.000	0.000	8.000	0.000	15.000	0.000
253	4.636	1.873	11.126	4.495	20.000	0.000
254	3.536	3.536	4.950	4.950	11.314	11.314
255	3.517	8.285	14.142	14.142	16.971	16.971
256	8.485	8.485	8.205	19.331	21.920	21.920
257	8.485	8.485	14.142	14.142	23.335	23.335
258	4.243	4.243	11.314	11.314	11.722	27.615
259	0.000	6.000	9.899	9.900	8.987	21.172
260	1.563	3.682	3.907	9.205	8.987	21.172
261	-3.000	0.000	2.622	-6.490	0.000	-3.000
262	5.657	-5.657	8.000	0.000	12.887	-5.470
263	9.000	0.000	9.272	3.746	10.199	4.121
264	5.657	-5.657	6.490	2.622	4.950	4.950
265	2.997	-7.417	4.603	-1.954	0.000	-7.000
266	-5.080	-11.967	-5.080	-11.967	-10.550	-24.854
267	-6.642	-15.649	-10.550	-24.854	-16.411	-38.661
268	-4.689	-11.046	-10.159	-23.933	-13.285	-31.297
269	-1.563	-3.682	-8.205	-19.331	-14.848	-34.979
270	0.000	-9.000	0.000	-14.000	0.000	-24.000
271	4.950	-4.950	4.121	-10.199	0.000	-18.000
272	2.000	0.000	4.603	-1.954	3.746	-9.272
273	8.000	0.000	8.285	-3.517	6.000	0.000
274	7.000	0.000	4.950	4.950	13.908	5.619
275	9.272	3.746	13.000	0.000	13.908	5.619
276	8.345	3.371	12.053	4.870	17.678	17.678
277	12.000	0.000	14.835	5.994	16.263	16.263
278	13.000	0.000	17.000	0.000	35.233	14.235
279	9.272	3.746	16.000	0.000	38.000	0.000
280	11.126	4.495	14.835	5.994	34.000	0.000
281	7.778	7.778	15.762	6.368	26.000	0.000
282	7.417	2.997	18.000	0.000	28.000	0.000
283	2.828	2.828	15.000	0.000	33.000	0.000
284	9.000	0.000	16.000	0.000	24.854	-10.550
285	7.000	0.000	15.000	0.000	32.218	-13.676
286	11.000	0.000	15.762	6.368	27.615	-11.722
287	14.835	5.994	12.887	-5.470	31.297	-13.285
288	16.000	0.000	15.649	-6.642	28.000	0.000
289	21.325	8.616	21.000	0.000	23.013	-9.768
290	12.887	-5.470	19.000	0.000	26.000	0.000
291	22.000	0.000	25.000	0.000	14.835	5.994
292	14.000	0.000	15.649	-6.642	21.000	0.000
293	13.908	5.619	20.251	-8.596	24.107	9.740
294	14.000	0.000	12.053	4.870	29.000	0.000
295	12.981	5.245	16.689	6.743	32.000	0.000
296	18.000	0.000	25.000	0.000	29.670	11.987
297	17.000	0.000	23.013	-9.768	43.000	0.000
298	21.000	0.000	26.000	0.000	27.000	0.000
299	10.607	-10.607	19.000	0.000	34.059	-14.457
300	7.071	-7.071	19.331	-8.205	36.820	-15.629
301	4.495	-11.126	19.331	-8.205	38.661	-16.411
302	0.000	-4.000	5.619	-13.908	22.627	-22.627
303	2.000	0.000	11.046	-4.689	18.385	-18.385
304	6.000	0.000	12.000	0.000	19.331	-8.205
305	7.417	2.997	21.000	0.000	24.000	0.000
306	18.000	0.000	17.490	-7.424	18.385	-18.385
307	15.649	-6.642	14.728	-6.252	15.556	-15.556
308	11.046	-4.689	12.887	-5.470	21.172	-8.987
309	9.272	3.746	13.000	0.000	20.251	-8.596
310	16.000	0.000	16.000	0.000	12.000	0.000
311	14.728	-6.252	11.967	-5.079	21.000	0.000
312	12.981	5.245	17.000	0.000	17.490	-7.424
313	15.762	6.368	21.000	0.000	18.410	-7.815
314	18.544	7.492	22.000	0.000	16.000	0.000
315	18.544	7.492	24.000	0.000	24.000	0.000
316	24.107	9.740	15.556	15.556	32.451	13.111
317	20.398	8.241	28.743	11.613	20.506	20.506
318	16.689	6.743	20.398	8.241	25.456	25.456
319	21.213	21.213	21.213	21.213	18.385	18.385
320	15.762	6.368	13.000	0.000	32.451	13.111
321	12.981	5.245	26.000	0.000	26.000	0.000
322	9.192	9.192	34.000	0.000	21.325	8.616
323	5.079	11.967	16.689	6.743	21.325	8.616
324	15.556	15.556	5.470	12.887	21.325	8.616
325	13.435	13.435	5.079	11.967	10.159	23.933
326	16.971	16.971	20.506	20.506	12.113	28.536
327	18.385	18.385	29.698	29.699	21.881	51.548
328	10.199	4.121	14.457	34.059	-0.000	78.000
329	9.272	3.746	13.000	0.000	-0.000	72.000
330	7.364	-3.126	5.563	2.248	-23.226	57.485
331	5.657	-5.657	3.371	-8.345	-4.689	-11.046
332	3.371	-8.345	0.000	-32.000	0.000	-41.000

Table 15 (continued)

	INPUT VARIABLES					
SEQ	1	2	3	4	5	6
333	4.870	-12.053	16.857	-41.723	20.229	-50.068
334	12.728	-12.728	24.749	-24.749	38.184	-38.184
335	19.331	-8.205	32.218	-13.676	52.469	-22.272
336	14.728	-6.252	28.536	-12.113	38.184	-38.184
337	12.887	-5.470	23.933	-10.159	49.707	-21.099
338	7.778	-7.778	19.799	-19.799	26.870	-26.870
339	7.778	-7.778	18.385	-18.385	15.734	-38.942
340	-6.364	-6.364	0.000	-15.000	7.867	-19.471
341	0.000	-8.000	0.000	-15.000	13.435	-13.435
342	-4.243	-4.243	0.000	-12.000	6.368	-15.762
343	-5.657	-5.657	-3.517	-8.285	-15.762	-6.368
344	-3.536	-3.536	-4.950	-4.950	-9.899	-9.899
345	0.000	-6.000	0.000	-9.000	-8.345	-3.371
346	0.000	-8.000	-3.126	-7.364	-7.778	-7.778
347	-6.000	0.000	-5.657	-5.657	-5.657	-5.657
348	-1.954	-4.603	-1.563	-3.682	-7.071	-7.071
349	-4.243	-4.243	-6.364	-6.364	-17.000	0.000
350	-1.954	-4.603	-7.417	-2.997	-13.908	-5.619
351	-5.000	0.000	-2.828	-2.828	-12.021	-12.021
352	1.172	2.762	-4.603	1.954	-22.252	-8.991
353	0.000	3.000	-0.375	0.927	-12.053	-4.870
354	1.172	2.762	2.344	5.523	-6.490	-2.622
355	4.243	4.243	4.689	11.046	-3.682	1.563
356	1.563	3.682	5.079	11.967	-3.371	8.345
357	2.782	1.124	3.907	9.205	-0.000	15.000
358	2.121	2.121	4.298	10.126	-0.000	19.000
359	1.563	3.682	6.364	6.364	7.815	18.410
360	15.000	0.000	9.192	-9.192	30.377	-12.894
361	10.126	-4.298	13.000	0.000	28.536	-12.113
362	11.000	0.000	22.000	0.000	23.180	9.365
363	22.000	0.000	25.000	0.000	25.961	10.489
364	18.410	-7.815	23.013	-9.768	44.000	0.000
365	14.849	-14.849	18.385	-18.385	9.740	-24.107
366	3.746	-9.272	7.071	-7.071	12.887	-5.470
367	10.126	-4.298	11.126	4.495	23.000	0.000
368	7.778	7.778	9.192	9.192	31.524	12.737
369	12.887	-5.470	14.728	-6.252	24.854	-10.550
370	21.000	0.000	30.597	12.362	54.704	22.102
371	19.000	0.000	22.252	8.991	28.743	11.613
372	12.887	-5.470	8.241	-20.398	12.728	-12.728
373	9.192	-9.192	17.678	-17.678	23.335	-23.334
374	11.046	-4.689	21.172	-8.987	33.138	-14.066
375	6.364	-6.364	13.808	-5.861	31.297	-13.285
376	6.444	-2.735	10.126	-4.298	15.762	6.368
377	8.285	-3.517	8.485	-8.485	12.887	-5.470
378	8.485	-8.485	7.071	-7.071	11.967	-5.079
379	5.657	-5.657	9.192	-9.192	11.046	-4.689
380	9.192	-9.192	12.021	-12.021	15.556	-15.556
381	14.835	5.994	26.000	0.000	31.000	0.000
382	17.000	0.000	21.000	0.000	25.000	0.000
383	12.887	-5.470	19.000	0.000	23.000	0.000
384	8.485	-8.485	18.000	0.000	25.000	0.000
385	7.778	-7.778	15.000	0.000	23.000	0.000
386	7.417	2.997	15.000	0.000	23.933	-10.159
387	9.272	3.746	17.000	0.000	24.854	-10.550
388	5.523	-2.344	18.000	0.000	26.695	-11.331
389	3.682	-1.563	10.126	-4.298	24.854	-10.550
390	3.709	1.498	14.000	0.000	23.013	-9.768
391	8.485	8.485	14.000	0.000	22.000	0.000
392	7.778	7.778	10.000	0.000	26.000	0.000
393	6.364	6.364	3.536	3.536	22.000	0.000
394	9.272	3.746	10.000	0.000	24.000	0.000
395	3.907	9.205	6.000	0.000	28.000	0.000
396	7.071	7.071	7.417	2.997	20.000	0.000
397	5.657	5.657	13.908	5.619	21.000	0.000
398	9.192	9.192	16.000	0.000	26.000	0.000
399	12.981	5.245	12.981	5.245	16.689	6.743
400	14.835	5.994	18.410	-7.815	22.000	0.000
401	13.908	5.619	19.000	0.000	16.000	0.000
402	16.000	0.000	12.981	5.245	14.000	0.000
403	17.617	7.118	19.471	7.867	21.325	8.616
404	17.000	0.000	16.971	16.971	23.335	23.335
405	22.252	8.991	15.238	35.900	16.411	38.661
406	14.849	14.849	17.192	40.502	35.355	35.355
407	17.617	7.118	21.881	51.548	22.662	53.389
408	21.000	0.000	44.000	0.000	38.184	38.184
409	11.046	-4.689	56.000	0.000	54.704	22.102
410	14.142	-14.142	27.615	-11.722	53.000	0.000
411	8.485	-8.485	31.297	-13.285	35.900	-15.238
412	7.364	-3.126	21.920	-21.920	22.627	-22.627
413	4.243	-4.243	7.071	-7.071	22.627	-22.627
414	6.000	0.000	11.967	-5.079	28.536	-12.113
415	8.000	0.000	14.728	-6.252	12.728	-12.728
416	7.000	0.000	12.887	-5.470	8.485	-8.485
417	4.000	0.000	11.046	-4.689	8.485	-8.485
418	6.490	2.622	4.603	-1.954	5.657	-5.657
419	9.000	0.000	9.000	0.000	4.636	1.873
420	7.364	-3.126	9.000	0.000	10.607	10.607
421	5.563	2.248	7.000	0.000	10.199	4.121
422	9.272	3.746	9.000	0.000	10.000	0.000

Table 15 (continued)

	INPUT VARIABLES					
SEQ	1	2	3	4	5	6
423	14.835	5.994	9.899	9.900	16.689	6.743
424	14.142	14.142	17.617	7.118	23.000	0.000
425	15.762	6.368	10.607	10.607	20.398	8.241
426	22.252	8.991	21.920	21.920	18.544	7.492
427	11.000	0.000	15.762	6.368	19.799	19.799
428	12.981	5.245	20.398	8.241	25.456	25.456
429	11.046	-4.689	18.544	7.492	15.238	35.900
430	7.364	-3.126	8.000	0.000	12.000	0.000
431	7.071	-7.071	8.000	0.000	16.000	0.000
432	4.000	0.000	10.000	0.000	14.835	5.994
433	1.124	-2.782	4.603	-1.954	12.981	5.245
434	-1.954	-4.603	2.000	0.000	1.873	-4.636
435	-0.927	-0.375	1.841	-0.781	8.485	-8.485
436	0.000	5.000	7.364	-3.126	10.126	-4.298
437	0.000	6.000	5.000	0.000	10.607	-10.607
438	1.854	0.749	6.444	-2.735	14.142	-14.142
439	5.523	-2.344	5.657	-5.657	15.359	-38.015
440	-2.828	-2.828	5.245	-12.981	31.113	-31.113
441	-5.657	-5.657	0.000	-12.000	22.627	-22.627
442	0.000	-9.000	0.000	-13.000	0.000	-14.000
443	0.000	-12.000	0.000	-13.000	-10.199	-4.121
444	0.000	-17.000	-5.861	-13.808	-37.087	-14.984
445	0.000	-13.000	-5.361	-13.808	10.489	-25.961
446	0.000	-10.000	0.000	-16.000	0.000	-22.000
447	0.000	-13.000	-7.033	-16.569	5.619	-13.908
448	3.371	-8.345	4.495	-11.126	11.046	-4.689
449	0.000	-9.000	0.000	-6.000	-4.000	0.000
450	2.997	-7.417	0.000	-12.000	0.000	-21.000
451	3.371	-8.345	4.870	-12.053	10.489	-25.961
452	4.870	-12.053	0.000	-16.000	0.000	-43.000
453	5.994	-14.835	0.000	-26.000	11.238	-27.815
454	5.619	-13.908	9.365	-23.180	11.238	-27.815
455	7.071	-7.071	11.314	-11.314	9.740	-24.107
456	7.778	-7.778	7.118	-17.617	10.489	-25.961
457	11.967	-5.079	10.607	-10.607	16.263	-16.263
458	6.444	-2.735	10.607	-10.607	16.569	-7.033
459	13.000	0.000	19.000	0.000	22.000	0.000
460	12.000	0.000	19.000	0.000	28.000	0.000
461	18.000	0.000	23.000	0.000	33.000	0.000
462	13.808	-5.861	27.000	0.000	33.000	0.000
463	6.364	-6.364	6.364	-6.364	26.695	-11.331
464	6.444	-2.735	10.000	0.000	11.046	-4.689
465	6.000	0.000	6.000	0.000	13.435	-13.435
466	7.000	0.000	10.126	-4.298	15.649	-6.642
467	9.000	0.000	13.808	-5.861	21.920	-21.920
468	13.908	5.619	11.967	-5.079	27.615	-11.722
469	12.728	12.728	18.544	7.492	32.000	0.000
470	20.398	8.241	25.034	10.114	43.000	0.000
471	25.034	10.114	41.723	16.857	46.000	0.000
472	32.451	13.111	22.627	22.627	48.214	19.480
473	25.000	0.000	45.432	18.356	60.104	60.104
474	16.263	-16.263	34.648	-34.648	41.719	-41.719
475	21.213	-21.213	27.577	-27.577	28.991	-28.991
476	15.556	-15.556	22.627	-22.627	28.284	-28.284
477	15.649	-6.642	19.799	-19.799	22.627	-22.627
478	12.728	-12.728	16.263	-16.263	24.854	-10.550
479	15.649	-6.642	20.506	-20.506	22.627	-22.627
480	16.569	-7.033	16.263	-16.263	21.213	-21.213
481	9.192	-9.192	20.251	-8.596	19.799	-19.799
482	9.192	-9.192	14.849	-14.849	14.142	-14.142
483	7.364	-3.126	10.607	-10.607	13.000	0.000
484	9.205	-3.907	11.967	-5.079	17.490	-7.424
485	5.657	5.657	15.762	6.368	20.000	0.000
486	9.899	9.900	15.556	15.556	25.961	10.489
487	14.849	14.849	17.678	17.678	21.325	8.616
488	6.364	6.364	17.617	7.118	28.743	11.613
489	5.470	12.887	8.205	19.331	26.163	26.163
490	4.298	10.126	8.987	21.172	8.596	20.251
491	7.071	7.071	12.894	30.377	28.991	28.991
492	8.596	20.251	-0.000	25.000	-0.000	58.000
493	9.192	9.192	5.861	13.808	5.861	13.808
494	14.000	0.000	9.899	9.900	5.079	11.967
495	12.053	4.870	17.617	7.118	5.470	12.887
496	11.126	4.495	8.485	8.485	5.470	12.887
497	16.689	6.743	5.657	5.657	1.414	1.414
498	10.199	4.121	6.364	-6.364	7.778	-7.778
499	12.000	0.000	9.900	-9.899	12.021	-12.021
500	14.728	-6.252	14.142	-14.142	18.385	-18.385
501	18.410	-7.815	20.506	-20.506	37.477	-37.477
502	14.728	-6.252	18.385	-18.385	36.820	-15.629
503	17.490	-7.424	17.000	0.000	35.900	-15.238

Table 16 Separation of a multivariate (6) set of Rantoul, Illinois, October 1950-55, upper wind components, zonal and meridional mixed distribution, at the 700-, 500-, and 300-mb levels into two separate distributions. The units are $m \cdot s^{-1}$.

```
                              PROGRAM NORMIX
                 WOLFE NORMAL MIXTURE ANALYSIS PROCEDURE(1974 REVISION)

  SAMPLE = RANTOUL
           OCTOBER
           700,500,AND 300MB
           X AND Y COMPONENTS

        SAMPLE SIZE =          503
        NUMBER OF VARIABLES =   6
        NUMBER OF TYPES =       2

                         ITERATION NUMBER  39

LIKELIHOOD OF  2 TYPES IN THIS SAMPLE =   -0.73297066D 04

                    CHARACTERISTICS OF THE WHOLE SAMPLE

                                  MEANS
     1            2            3            4            5            6
  7.3252      -0.9486      10.6700      -2.5704      15.4049      -3.6999

                           STANDARD DEVIATIONS
     1            2            3            4            5            6
  6.6262       6.8921       8.9594      10.3149      13.7082      16.0058

                              CORRELATIONS
     1            2            3            4            5            6
  1.0000       0.2215       0.7525       0.2313       0.5589       0.2311
  0.2215       1.0000       0.1138       0.7581       0.1265       0.5757
  0.7525       0.1138       1.0000       0.1587       0.7513       0.1977
  0.2313       0.7581       0.1587       1.0000       0.1200       0.7856
  0.5589       0.1265       0.7513       0.1200       1.0000       0.1266
  0.2311       0.5757       0.1977       0.7856       0.1266       1.0000


                    CHARACTERISTICS OF TYPE    1

              THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.171

                                  MEANS
     1            2            3            4            5            6
 10.4545      -2.0123      16.2238      -2.6858      18.7634      -3.3254

                           STANDARD DEVIATIONS
     1            2            3            4            5            6
  7.4317       9.0069      12.5099      17.4952      20.1227      29.7834

                              CORRELATIONS
     1            2            3            4            5            6
  1.0000       0.3578       0.6076       0.4401       0.5639       0.3522
  0.3578       1.0000       0.0297       0.7469       0.0730       0.6333
  0.6076       0.0297       1.0000       0.1637       0.7420       0.1966
  0.4401       0.7469       0.1637       1.0000       0.1148       0.7884
  0.5639       0.0730       0.7420       0.1148       1.0000       0.0799
  0.3522       0.6333       0.1966       0.7884       0.0799       1.0000

EIGENVALUES     EIGENVECTORS
                   1            2            3            4            5            6
1165.8027       0.1014       0.1798       0.1382       0.2729       0.4299       0.8240
 507.7779       0.1864      -0.0377       0.3840      -0.1749       0.7770      -0.4266
 108.4213       0.1060       0.4467      -0.2043       0.7783       0.0692      -0.3701
  56.7679       0.4498      -0.0490       0.7497       0.1994      -0.4398      -0.0070
  30.0842       0.1181       0.8592       0.0471      -0.4824      -0.1135       0.0091
  22.0357       0.8529      -0.1618      -0.4768      -0.1293       0.0181       0.0437


                    CHARACTERISTICS OF TYPE    2

              THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.829

                                  MEANS
     1            2            3            4            5            6
  6.6783      -0.7287       9.5218      -2.5466      14.7107      -3.7773

                           STANDARD DEVIATIONS
     1            2            3            4            5            6
  6.2548       6.3457       7.5377       8.0693      11.8422      11.2131
```

Table 16 (continued)

```
                                  CORRELATIONS
      1             2             3             4             5             6
   1.0000        0.2039        0.8038        0.1455        0.5563        0.1878
   0.2039        1.0000        0.1937        0.7943        0.1673        0.5849
   0.8038        0.1937        1.0000        0.1721        0.7626        0.2195
   0.1455        0.7943        0.1721        1.0000        0.1280        0.7914
   0.5563        0.1673        0.7626        0.1280        1.0000        0.1756
   0.1878        0.5849        0.2195        0.7914        0.1756        1.0000

EIGENVALUES     EIGENVECTORS
                   1             2             3             4             5             6
 236.2003       0.2483       -0.2052        0.6858       -0.0776        0.3252        0.5605
 155.9890       0.2529        0.2350        0.1020        0.6736        0.5317       -0.3660
  31.5369       0.3590       -0.3140        0.4636       -0.0919       -0.4232       -0.6083
  28.5496       0.3591        0.4036        0.0176        0.4196       -0.6195        0.3846
   8.3730       0.5713       -0.5859       -0.5400        0.0860        0.0879        0.1540
   6.7008       0.5387        0.5457       -0.1120       -0.5902        0.2024       -0.1008
```

```
SAMPLE = RANTOUL
         OCTOBER
         700,500,AND 300MB
         X AND Y COMPONENTS
```

```
   PROBABILITIES
OF TYPE MEMBERSHIP
      1       2

  1 0.065 0.935
  2 0.011 0.989
  3 0.015 0.985
  4 0.996 0.004
  5 0.022 0.978
  6 0.037 0.963
  7 0.026 0.974
  8 1.000 0.000
  9 0.119 0.881
 10 0.851 0.149
 11 0.981 0.019
 12 0.005 0.995
 13 0.007 0.993
 14 0.003 0.997
 15 0.122 0.878
 16 0.976 0.024
 17 0.034 0.966
 18 0.054 0.946
 19 0.088 0.912
 20 1.000 0.000
 21 0.003 0.997
 22 0.125 0.875
 23 0.269 0.731
 24 0.011 0.989
 25 0.005 0.995
 26 1.000 0.000
 27 0.028 0.972
 28 0.013 0.987
 29 0.294 0.706
 30 0.218 0.782
 31 0.029 0.971
 32 0.034 0.966
 33 0.011 0.989
 34 0.037 0.963
 35 0.407 0.593
 36 0.011 0.989
 37 0.008 0.992
 38 0.005 0.995
 39 0.005 0.995
 40 0.003 0.997
 41 0.323 0.677
 42 0.006 0.994
 43 0.043 0.957
 44 0.017 0.983
 45 0.025 0.975
 46 0.071 0.929
 47 0.016 0.984
 48 0.916 0.084
 49 0.010 0.990
 50 0.013 0.987
 51 0.013 0.987
 52 0.181 0.819
 53 0.020 0.980
 54 0.996 0.004
 55 0.005 0.995
 56 0.009 0.991
 57 0.012 0.988
 58 0.030 0.970
 59 0.028 0.972
 60 0.008 0.992
 61 0.008 0.992
 62 0.008 0.992
 63 0.004 0.996
 64 0.003 0.997
 65 0.017 0.983
```

```
   PROBABILITIES
OF TYPE MEMBERSHIP
      1       2

 66 0.010 0.990
 67 0.009 0.991
 68 0.007 0.993
 69 0.005 0.995
 70 0.007 0.993
 71 0.007 0.993
 72 0.004 0.996
 73 0.007 0.993
 74 0.005 0.995
 75 0.021 0.979
 76 0.010 0.990
 77 0.093 0.907
 78 0.024 0.976
 79 0.025 0.975
 80 0.014 0.986
 81 0.038 0.962
 82 0.010 0.990
 83 0.050 0.950
 84 0.163 0.837
 85 0.066 0.934
 86 0.008 0.992
 87 0.016 0.984
 88 0.009 0.991
 89 0.022 0.978
 90 0.114 0.886
 91 0.081 0.919
 92 0.032 0.968
 93 0.234 0.766
 94 0.048 0.952
 95 0.014 0.986
 96 0.205 0.795
 97 0.031 0.969
 98 0.058 0.942
 99 0.055 0.945
100 0.013 0.987
101 0.015 0.985
102 0.403 0.597
103 0.878 0.122
104 0.009 0.991
105 0.009 0.991
106 0.045 0.955
107 0.875 0.125
108 0.028 0.972
109 0.014 0.986
110 0.010 0.990
111 0.008 0.992
112 0.004 0.996
113 0.009 0.991
114 0.005 0.995
115 0.026 0.974
116 0.004 0.996
117 0.013 0.987
118 0.037 0.963
119 0.013 0.987
120 0.009 0.991
121 0.034 0.966
122 0.023 0.977
123 0.019 0.981
124 0.069 0.931
125 0.685 0.315
126 0.106 0.894
127 0.615 0.385
128 0.811 0.189
129 0.892 0.108
130 0.006 0.994
```

```
   PROBABILITIES
OF TYPE MEMBERSHIP
      1       2

131 0.009 0.991
132 0.037 0.963
133 0.075 0.925
134 0.006 0.994
135 0.019 0.981
136 0.008 0.992
137 0.018 0.982
138 0.016 0.984
139 0.027 0.973
140 0.926 0.074
141 0.998 0.002
142 0.030 0.970
143 0.240 0.760
144 0.023 0.977
145 0.069 0.931
146 0.027 0.973
147 0.011 0.989
148 0.016 0.984
149 0.010 0.990
150 0.020 0.980
151 0.061 0.939
152 0.056 0.944
153 0.047 0.953
154 0.272 0.728
155 0.033 0.967
156 0.014 0.986
157 0.019 0.981
158 0.011 0.989
159 0.038 0.962
160 0.006 0.994
161 0.003 0.997
162 0.002 0.998
163 0.004 0.996
164 0.003 0.997
165 0.004 0.996
166 0.004 0.996
167 0.015 0.985
168 0.005 0.995
169 0.004 0.996
170 0.004 0.996
171 0.005 0.995
172 0.092 0.908
173 0.011 0.989
174 0.013 0.987
175 0.013 0.987
176 0.023 0.977
177 0.098 0.902
178 0.007 0.993
179 0.311 0.689
180 0.998 0.002
181 0.997 0.003
182 0.977 0.023
183 0.857 0.143
184 0.204 0.796
185 0.043 0.957
186 0.020 0.980
187 0.007 0.993
188 0.008 0.992
189 0.006 0.994
190 0.010 0.990
191 0.020 0.980
192 0.063 0.937
193 0.007 0.993
194 0.006 0.994
195 0.012 0.988
```

Table 16 (continued)

	PROBABILITIES OF TYPE MEMBERSHIP	
	1	2
196	0.208	0.792
197	0.121	0.879
198	0.023	0.977
199	0.015	0.985
200	0.013	0.987
201	0.397	0.603
202	0.983	0.017
203	0.073	0.927
204	1.000	0.000
205	0.005	0.995
206	0.006	0.994
207	0.005	0.995
208	0.005	0.995
209	0.021	0.979
210	0.003	0.997
211	0.008	0.992
212	0.003	0.997
213	0.022	0.978
214	0.107	0.893
215	0.045	0.955
216	0.371	0.629
217	0.066	0.934
218	0.131	0.869
219	0.019	0.981
220	0.014	0.986
221	0.014	0.986
222	0.007	0.993
223	0.013	0.987
224	0.011	0.989
225	0.010	0.990
226	0.006	0.994
227	0.013	0.987
228	0.005	0.995
229	0.005	0.995
230	0.011	0.989
231	0.062	0.938
232	0.004	0.996
233	0.011	0.989
234	0.035	0.965
235	0.004	0.996
236	0.006	0.994
237	0.015	0.985
238	0.025	0.975
239	0.007	0.993
240	0.005	0.995
241	0.012	0.988
242	0.181	0.819
243	0.007	0.993
244	0.008	0.992
245	0.037	0.963
246	0.034	0.966
247	0.922	0.078
248	0.084	0.916
249	0.020	0.980
250	0.017	0.983
251	0.006	0.994
252	0.003	0.997
253	0.006	0.994
254	0.006	0.994
255	0.026	0.974
256	0.213	0.787
257	0.033	0.967
258	0.211	0.789
259	0.063	0.937
260	0.025	0.975
261	0.017	0.983
262	0.016	0.984
263	0.012	0.988
264	0.033	0.967
265	0.023	0.977
266	0.028	0.972
267	0.200	0.800
268	0.197	0.803
269	0.649	0.351
270	0.013	0.987
271	0.015	0.985
272	0.005	0.995
273	0.011	0.989
274	0.018	0.982
275	0.012	0.988
276	0.031	0.969
277	0.057	0.943
278	0.087	0.913
279	0.017	0.983
280	0.029	0.971
281	0.009	0.991
282	0.013	0.987
283	0.012	0.988
284	0.023	0.977
285	0.058	0.942

	PROBABILITIES OF TYPE MEMBERSHIP	
	1	2
286	0.566	0.434
287	0.145	0.855
288	0.032	0.968
289	0.056	0.944
290	0.041	0.959
291	0.132	0.868
292	0.020	0.980
293	0.928	0.072
294	0.145	0.855
295	0.028	0.972
296	0.088	0.912
297	0.195	0.805
298	0.039	0.961
299	0.933	0.067
300	0.103	0.897
301	0.415	0.585
302	0.023	0.977
303	0.024	0.976
304	0.008	0.992
305	0.069	0.931
306	0.046	0.954
307	0.051	0.949
308	0.010	0.990
309	0.009	0.991
310	0.020	0.980
311	0.030	0.970
312	0.016	0.984
313	0.038	0.962
314	0.050	0.950
315	0.035	0.965
316	0.964	0.036
317	0.486	0.514
318	0.224	0.776
319	0.171	0.829
320	0.364	0.636
321	0.166	0.834
322	1.000	0.000
323	0.028	0.972
324	0.811	0.189
325	0.643	0.357
326	0.367	0.633
327	0.999	0.001
328	1.000	0.000
329	1.000	0.000
330	1.000	0.000
331	0.018	0.982
332	0.980	0.020
333	1.000	0.000
334	0.931	0.069
335	0.796	0.204
336	0.998	0.002
337	0.535	0.465
338	0.476	0.524
339	0.940	0.060
340	0.009	0.991
341	0.011	0.989
342	0.005	0.995
343	0.037	0.963
344	0.006	0.994
345	0.028	0.972
346	0.011	0.989
347	0.005	0.995
348	0.007	0.993
349	0.047	0.953
350	0.019	0.981
351	0.013	0.987
352	0.208	0.792
353	0.021	0.979
354	0.020	0.980
355	0.080	0.920
356	0.053	0.947
357	0.036	0.964
358	0.055	0.945
359	0.024	0.976
360	0.310	0.690
361	0.130	0.870
362	0.111	0.889
363	0.069	0.931
364	0.139	0.861
365	0.193	0.807
366	0.007	0.993
367	0.101	0.899
368	0.031	0.969
369	0.015	0.985
370	0.954	0.046
371	0.188	0.812
372	0.388	0.612
373	0.090	0.910
374	0.048	0.952
375	0.026	0.974

	PROBABILITIES OF TYPE MEMBERSHIP	
	1	2
376	0.016	0.984
377	0.008	0.992
378	0.011	0.989
379	0.010	0.990
380	0.012	0.988
381	0.064	0.936
382	0.017	0.983
383	0.044	0.956
384	0.119	0.881
385	0.047	0.953
386	0.014	0.986
387	0.020	0.980
388	0.107	0.893
389	0.006	0.994
390	0.015	0.985
391	0.012	0.988
392	0.013	0.987
393	0.027	0.973
394	0.008	0.992
395	0.028	0.972
396	0.007	0.993
397	0.007	0.993
398	0.017	0.983
399	0.011	0.989
400	0.207	0.793
401	0.024	0.976
402	0.087	0.913
403	0.021	0.979
404	0.942	0.058
405	1.000	0.000
406	1.000	0.000
407	1.000	0.000
408	1.000	0.000
409	1.000	0.000
410	0.816	0.184
411	0.996	0.004
412	0.991	0.009
413	0.045	0.955
414	0.010	0.990
415	0.025	0.975
416	0.020	0.980
417	0.014	0.986
418	0.007	0.993
419	0.012	0.988
420	0.024	0.976
421	0.006	0.994
422	0.008	0.992
423	0.099	0.901
424	0.021	0.979
425	0.157	0.843
426	0.992	0.008
427	0.094	0.906
428	0.238	0.762
429	0.997	0.003
430	0.008	0.992
431	0.023	0.977
432	0.007	0.993
433	0.005	0.995
434	0.006	0.994
435	0.003	0.997
436	0.010	0.990
437	0.005	0.995
438	0.006	0.994
439	0.980	0.020
440	0.292	0.708
441	0.015	0.985
442	0.007	0.993
443	0.082	0.918
444	0.967	0.033
445	0.161	0.839
446	0.011	0.989
447	0.055	0.945
448	0.009	0.991
449	0.015	0.985
450	0.015	0.985
451	0.025	0.975
452	0.914	0.086
453	0.183	0.817
454	0.061	0.939
455	0.048	0.952
456	0.033	0.967
457	0.015	0.985
458	0.014	0.986
459	0.014	0.986
460	0.012	0.988
461	0.023	0.977
462	0.374	0.626
463	0.028	0.972
464	0.010	0.990
465	0.021	0.979

Table 16 (continued)

```
   PROBABILITIES
OF TYPE MEMBERSHIP
       1      2

466 0.005 0.995
467 0.055 0.945
468 0.071 0.929
469 0.019 0.981
470 0.265 0.735
471 0.999 0.001
472 1.000 0.000
473 1.000 0.000
474 1.000 0.000
475 0.556 0.444
476 0.241 0.759
477 0.132 0.868
478 0.030 0.970
```

```
   PROBABILITIES
OF TYPE MEMBERSHIP
       1      2

479 0.194 0.806
480 0.037 0.963
481 0.362 0.638
482 0.032 0.968
483 0.051 0.949
484 0.007 0.993
485 0.015 0.985
486 0.029 0.971
487 0.067 0.933
488 0.015 0.985
489 0.054 0.946
490 0.087 0.913
491 0.998 0.002
```

```
   PROBABILITIES
OF TYPE MEMBERSHIP
       1      2

492 1.000 0.000
493 0.045 0.955
494 0.312 0.688
495 0.171 0.829
496 0.041 0.959
497 0.332 0.668
498 0.035 0.965
499 0.023 0.977
500 0.023 0.977
501 0.785 0.215
502 0.149 0.851
503 0.955 0.045
```

```
LOGARITHM OF LIKELIHOOD RATIO OF  2 TO  1 TYPES =    0.18925055D 03
CHI-SQUARE WITH 54 DEGREES OF FREEDOM=    372.86
PROBABILITY OF NULL HYPOTHESIS= 0.00000000
```

Table 17 Separation of a multivariate (6) set of Rantoul, Illinois, October 1950-55, upper wind components, zonal and meridional mixed distribution, at the 700-, 500-, and 300-mb levels into three distinct distributions. The units are $m \cdot s^{-1}$.

```
                              PROGRAM NORMIX
                  WOLFE NORMAL MIXTURE ANALYSIS PROCEDURE (1974 REVISION)

   SAMPLE = RANTOUL
            OCTOBER
            700,500,AND 300MB
            X AND Y COMPONENTS


        SAMPLE SIZE =          503
        NUMBER OF VARIABLES =   6
        NUMBER OF TYPES =       3

                       ITERATION NUMBER 101

  LIKELIHOOD OF  3 TYPES IN THIS SAMPLE =   -0.72717729D 04

                    CHARACTERISTICS OF THE WHOLE SAMPLE


                                  MEANS
     1           2           3           4           5           6
  7.3252     -0.9486     10.6700     -2.5704     15.4049     -3.6999

                           STANDARD DEVIATIONS
     1           2           3           4           5           6
  6.6262      6.8921      8.9594     10.3149     13.7082     16.0058

                               CORRELATIONS
     1           2           3           4           5           6
  1.0000      0.2215      0.7525      0.2313      0.5589      0.2311
  0.2215      1.0000      0.1138      0.7581      0.1265      0.5757
  0.7525      0.1138      1.0000      0.1587      0.7513      0.1977
  0.2313      0.7581      0.1587      1.0000      0.1200      0.7856
  0.5589      0.1265      0.7513      0.1200      1.0000      0.1266
  0.2311      0.5757      0.1977      0.7856      0.1266      1.0000


                    CHARACTERISTICS OF TYPE    1

           THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.182

                                  MEANS
     1           2           3           4           5           6
 10.5173     -1.7616     16.0647     -2.8166     19.4572     -3.5912

                           STANDARD DEVIATIONS
     1           2           3           4           5           6
  7.4715      9.1043     12.1651     17.3130     19.5470     29.2311

                               CORRELATIONS
     1           2           3           4           5           6
  1.0000      0.2932      0.6046      0.3978      0.5377      0.3250
  0.2932      1.0000     -0.0233      0.7480      0.0540      0.6252
  0.6046     -0.0233      1.0000      0.1336      0.7194      0.1719
  0.3978      0.7480      0.1336      1.0000      0.0917      0.7906
  0.5377      0.0540      0.7194      0.0917      1.0000      0.0573
  0.3250      0.6252      0.1719      0.7906      0.0573      1.0000

 EIGENVALUES      EIGENVECTORS
                 1           2           3           4           5           6
 1122.2211    0.0936      0.1893      0.0998      0.3370      0.5134      0.7539
  481.4084    0.1910     -0.0373      0.4161     -0.1832      0.7241     -0.4807
  107.0029    0.0873      0.4475     -0.2426      0.7334      0.0923     -0.4409
   59.8619    0.4557     -0.0371      0.7255      0.2552     -0.4444      0.0453
   29.3601    0.0895      0.8629      0.0824     -0.4746     -0.1131      0.0505
   23.1239    0.8553     -0.1286     -0.4742     -0.1571      0.0274      0.0405


                    CHARACTERISTICS OF TYPE    2

           THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.183

                                  MEANS
     1           2           3           4           5           6
  5.9298      1.1741     11.3642      0.4841     19.9322     -4.8802

                           STANDARD DEVIATIONS
     1           2           3           4           5           6
  4.0452      6.0597      6.8744      4.7102      8.3419      7.2647
```

Table 17 (continued)

```
                              CORRELA7IONS
  1             2             3             4             5             6
1.0000       -0.3849        0.7454       -0.0248        0.7052        0.3055
-0.3849       1.0000       -0.5189        0.6927       -0.5375        0.4247
0.7454       -0.5189        1.0000       -0.3651        0.8079        0.0263
-0.0248       0.6927       -0.3651        1.0000       -0.3992        0.7287
0.7052       -0.5375        0.8079       -0.3992        1.0000       -0.1160
0.3055        0.4247        0.0263        0.7287       -0.1160        1.0000

EIGENVALUES      EIGENVECTORS
                    1             2             3             4             5             6
137.9714         0.2345        0.2674       -0.1184       -0.0908        0.5475        0.7426
 71.7321        -0.3902        0.2424        0.7711        0.3460       -0.0810        0.2609
 16.9294         0.5133        0.2587       -0.0625        0.7630        0.1223       -0.2621
  9.8691        -0.2481        0.3673        0.1214       -0.2410        0.6562       -0.5478
  5.1576         0.6598        0.2014        0.4947       -0.4725       -0.2137       -0.1023
  3.2296        -0.1801        0.7921       -0.3577       -0.0925       -0.4499        0.0351

                     CHARACTERISTICS OF TYPE   3

           THE PROPORTION OF THE POPULA7ION FROM 7HIS TYPE= 0.635

                                  MEANS
    1             2             3             4             5             6
  6.8098       -1.3264        8.9194       -3.3796       12.9361       -3.3911

                           STANOARO OEVIATIONS
    1             2             3             4             5             6
  6.6829        6.2294        7.6649        8.4692       12.2260       11.9304

                               CORRELA7IONS
    1             2             3             4             5             6
  1.0000        0.3441        0.8404        0.1994        0.5799        0.1907
  0.3441        1.0000        0.3814        0.8169        0.2775        0.6420
  0.8404        0.3814        1.0000        0.2586        0.7613        0.2987
  0.1994        0.8169        0.2586        1.0000        0.1608        0.8260
  0.5799        0.2775        0.7613        0.1608        1.0000        0.2643
  0.1907        0.6420        0.2987        0.8260        0.2643        1.0000

EIGENVALUES      EIGENVEC7ORS
                    1             2             3             4             5             6
278.7235         0.2436       -0.2469        0.6157       -0.2932        0.2506        0.5932
154.5198         0.2676        0.1768        0.2977        0.5868        0.6060       -0.3113
 37.4431         0.3437       -0.3103        0.4015       -0.2494       -0.3559       -0.6599
 22.2485         0.3694        0.4038        0.1879        0.4234       -0.6307        0.2970
  6.6393         0.5399       -0.6093       -0.4085        0.2521        0.0212        0.1577
  6.1806         0.5683        0.5266       -0.2955       -0.5145        0.2124       -0.0515


                                PROGRAM NORMIX
                  WOLFE NORMAL MIX7URE ANALYSIS PROCEOURE(1974 REVISION)


SAMPLE = RANTOUL
         OCTOBER
         700,500,ANO 300MB
         X ANO Y COMPONEN7S
```

```
    PROBABILITIES
  OF TYPE MEMBERSHIP
     1     2     3

 1 0.249 0.001 0.750
 2 0.004 0.944 0.052
 3 0.001 0.993 0.005
 4 1.000 0.000 0.000
 5 0.001 0.998 0.001
 6 0.015 0.942 0.043
 7 0.032 0.009 0.959
 8 1.000 0.000 0.000
 9 0.027 0.960 0.013
10 0.906 0.000 0.094
11 1.000 0.000 0.000
12 0.007 0.062 0.932
13 0.008 0.196 0.796
14 0.005 0.006 0.990
15 0.241 0.000 0.759
16 0.960 0.000 0.040
17 0.041 0.000 0.959
18 0.116 0.001 0.883
19 0.198 0.001 0.802
20 1.000 0.000 0.000
21 0.008 0.052 0.941
22 0.102 0.000 0.898
23 0.219 0.000 0.781
24 0.014 0.000 0.986
25 0.003 0.614 0.383
26 1.000 0.000 0.000
27 0.034 0.000 0.966
28 0.013 0.768 0.219
29 0.087 0.900 0.012
30 0.267 0.000 0.733
```

```
    PROBABILITIES
  OF TYPE MEMBERSHIP
     1     2     3

31 0.033 0.000 0.967
32 0.028 0.000 0.972
33 0.012 0.011 0.978
34 0.039 0.000 0.961
35 0.495 0.000 0.505
36 0.011 0.006 0.983
37 0.006 0.535 0.459
38 0.005 0.675 0.321
39 0.002 0.772 0.225
40 0.004 0.618 0.377
41 0.740 0.000 0.260
42 0.006 0.536 0.458
43 0.037 0.009 0.954
44 0.014 0.379 0.607
45 0.028 0.005 0.967
46 0.065 0.001 0.934
47 0.016 0.125 0.858
48 0.907 0.000 0.093
49 0.011 0.000 0.989
50 0.014 0.000 0.986
51 0.014 0.000 0.986
52 0.164 0.000 0.836
53 0.015 0.000 0.985
54 1.000 0.000 0.000
55 0.004 0.000 0.996
56 0.008 0.000 0.992
57 0.011 0.000 0.989
58 0.089 0.000 0.911
59 0.019 0.000 0.981
60 0.008 0.000 0.991
```

```
    PROBABILITIES
  OF TYPE MEMBERSHIP
     1     2     3

61 0.009 0.001 0.991
62 0.013 0.379 0.607
63 0.006 0.081 0.913
64 0.006 0.088 0.907
65 0.023 0.014 0.963
66 0.017 0.205 0.778
67 0.019 0.038 0.943
68 0.007 0.224 0.769
69 0.006 0.183 0.811
70 0.007 0.190 0.803
71 0.007 0.716 0.277
72 0.003 0.722 0.275
73 0.004 0.864 0.133
74 0.004 0.636 0.361
75 0.019 0.504 0.477
76 0.011 0.227 0.762
77 0.018 0.955 0.027
78 0.024 0.003 0.973
79 0.050 0.206 0.743
80 0.021 0.166 0.813
81 0.053 0.067 0.881
82 0.013 0.438 0.549
83 0.203 0.000 0.796
84 0.022 0.969 0.009
85 0.031 0.844 0.125
86 0.011 0.106 0.883
87 0.017 0.138 0.845
88 0.013 0.424 0.563
89 0.017 0.754 0.229
90 0.273 0.029 0.698
```

Table 17 (continued)

	PROBABILITIES OF TYPE MEMBERSHIP		
	1	2	3
91	0.108	0.004	0.888
92	0.032	0.008	0.960
93	0.663	0.000	0.337
94	0.045	0.536	0.419
95	0.020	0.011	0.969
96	0.340	0.000	0.659
97	0.030	0.000	0.970
98	0.060	0.811	0.129
99	0.053	0.000	0.947
100	0.020	0.000	0.980
101	0.020	0.000	0.980
102	0.773	0.038	0.189
103	0.989	0.000	0.011
104	0.010	0.001	0.989
105	0.007	0.629	0.364
106	0.048	0.000	0.952
107	0.777	0.000	0.223
108	0.021	0.000	0.979
109	0.012	0.000	0.988
110	0.008	0.000	0.991
111	0.007	0.000	0.993
112	0.003	0.000	0.997
113	0.006	0.000	0.994
114	0.003	0.000	0.997
115	0.071	0.052	0.877
116	0.005	0.006	0.988
117	0.009	0.781	0.210
118	0.052	0.107	0.841
119	0.016	0.009	0.976
120	0.009	0.340	0.652
121	0.014	0.901	0.084
122	0.019	0.000	0.981
123	0.028	0.006	0.966
124	0.149	0.018	0.833
125	0.284	0.702	0.014
126	0.082	0.000	0.918
127	0.360	0.625	0.015
128	0.855	0.000	0.145
129	0.868	0.000	0.132
130	0.009	0.001	0.991
131	0.008	0.000	0.992
132	0.025	0.000	0.975
133	0.059	0.000	0.941
134	0.006	0.002	0.992
135	0.021	0.005	0.974
136	0.010	0.001	0.989
137	0.018	0.001	0.981
138	0.016	0.000	0.984
139	0.023	0.000	0.977
140	0.835	0.000	0.165
141	0.985	0.000	0.015
142	0.030	0.000	0.970
143	0.272	0.000	0.728
144	0.024	0.000	0.976
145	0.166	0.050	0.784
146	0.008	0.943	0.049
147	0.015	0.013	0.971
148	0.007	0.870	0.122
149	0.012	0.030	0.958
150	0.021	0.018	0.962
151	0.058	0.000	0.942
152	0.054	0.000	0.946
153	0.097	0.001	0.902
154	0.179	0.000	0.821
155	0.033	0.000	0.967
156	0.025	0.001	0.973
157	0.032	0.158	0.810
158	0.013	0.010	0.977
159	0.077	0.000	0.923
160	0.008	0.448	0.544
161	0.005	0.007	0.988
162	0.011	0.000	0.988
163	0.039	0.123	0.838
164	0.006	0.723	0.271
165	0.007	0.241	0.752
166	0.002	0.774	0.224
167	0.004	0.947	0.049
168	0.004	0.831	0.164
169	0.002	0.837	0.161
170	0.002	0.838	0.160
171	0.003	0.809	0.187
172	0.106	0.672	0.222
173	0.015	0.292	0.692
174	0.013	0.399	0.588
175	0.012	0.000	0.987
176	0.036	0.000	0.963
177	0.115	0.003	0.881
178	0.008	0.025	0.967
179	0.435	0.000	0.564
180	0.938	0.062	0.000

	PROBABILITIES OF TYPE MEMBERSHIP		
	1	2	3
181	1.000	0.000	0.000
182	0.999	0.000	0.001
183	0.954	0.000	0.046
184	0.149	0.000	0.851
185	0.033	0.824	0.143
186	0.023	0.000	0.977
187	0.008	0.292	0.700
188	0.005	0.672	0.323
189	0.004	0.635	0.362
190	0.010	0.062	0.929
191	0.042	0.262	0.695
192	0.049	0.000	0.951
193	0.007	0.042	0.951
194	0.007	0.271	0.722
195	0.014	0.005	0.981
196	0.100	0.000	0.900
197	0.197	0.000	0.803
198	0.024	0.000	0.976
199	0.016	0.000	0.984
200	0.021	0.025	0.954
201	0.556	0.000	0.444
202	0.996	0.000	0.004
203	0.070	0.000	0.930
204	1.000	0.000	0.000
205	0.004	0.000	0.996
206	0.010	0.003	0.987
207	0.005	0.000	0.995
208	0.004	0.000	0.996
209	0.013	0.000	0.987
210	0.003	0.000	0.997
211	0.010	0.000	0.990
212	0.002	0.000	0.998
213	0.073	0.000	0.927
214	0.078	0.000	0.922
215	0.026	0.000	0.974
216	0.255	0.000	0.745
217	0.039	0.000	0.961
218	0.084	0.000	0.916
219	0.017	0.000	0.983
220	0.012	0.000	0.988
221	0.019	0.000	0.981
222	0.007	0.000	0.993
223	0.012	0.000	0.988
224	0.010	0.000	0.989
225	0.010	0.006	0.985
226	0.006	0.051	0.943
227	0.015	0.010	0.975
228	0.003	0.708	0.289
229	0.003	0.671	0.326
230	0.019	0.070	0.911
231	0.139	0.124	0.737
232	0.009	0.485	0.506
233	0.015	0.725	0.260
234	0.009	0.965	0.026
235	0.002	0.872	0.126
236	0.003	0.942	0.055
237	0.056	0.112	0.831
238	0.032	0.839	0.129
239	0.004	0.955	0.040
240	0.002	0.977	0.022
241	0.005	0.927	0.069
242	0.563	0.001	0.436
243	0.002	0.923	0.075
244	0.010	0.206	0.784
245	0.013	0.870	0.117
246	0.067	0.003	0.930
247	0.957	0.000	0.043
248	0.093	0.000	0.907
249	0.026	0.000	0.974
250	0.019	0.000	0.981
251	0.010	0.034	0.956
252	0.003	0.336	0.660
253	0.004	0.747	0.249
254	0.008	0.027	0.964
255	0.050	0.000	0.949
256	0.340	0.000	0.660
257	0.035	0.001	0.964
258	0.149	0.000	0.851
259	0.062	0.000	0.938
260	0.023	0.000	0.977
261	0.025	0.001	0.974
262	0.012	0.653	0.335
263	0.012	0.108	0.880
264	0.039	0.015	0.946
265	0.026	0.003	0.971
266	0.017	0.000	0.983
267	0.057	0.000	0.943
268	0.115	0.000	0.885
269	0.686	0.000	0.314
270	0.011	0.000	0.989

	PROBABILITIES OF TYPE MEMBERSHIP		
	1	2	3
271	0.012	0.000	0.988
272	0.007	0.122	0.871
273	0.011	0.000	0.988
274	0.015	0.417	0.568
275	0.014	0.038	0.948
276	0.029	0.006	0.965
277	0.060	0.032	0.908
278	0.068	0.000	0.932
279	0.036	0.070	0.893
280	0.055	0.464	0.481
281	0.010	0.725	0.265
282	0.008	0.784	0.208
283	0.051	0.025	0.925
284	0.012	0.879	0.108
285	0.042	0.907	0.051
286	0.885	0.070	0.045
287	0.269	0.000	0.731
288	0.029	0.000	0.971
289	0.041	0.000	0.959
290	0.016	0.888	0.096
291	0.150	0.000	0.850
292	0.020	0.000	0.980
293	0.941	0.000	0.059
294	0.105	0.646	0.249
295	0.053	0.366	0.582
296	0.107	0.008	0.885
297	0.132	0.000	0.868
298	0.059	0.000	0.941
299	0.310	0.690	0.000
300	0.034	0.945	0.021
301	0.292	0.702	0.006
302	0.065	0.028	0.907
303	0.008	0.947	0.044
304	0.003	0.902	0.095
305	0.016	0.919	0.065
306	0.037	0.000	0.963
307	0.078	0.000	0.922
308	0.015	0.058	0.927
309	0.012	0.406	0.582
310	0.019	0.000	0.981
311	0.043	0.000	0.957
312	0.022	0.001	0.977
313	0.043	0.000	0.957
314	0.036	0.000	0.964
315	0.030	0.000	0.970
316	0.980	0.000	0.020
317	0.395	0.000	0.604
318	0.175	0.001	0.823
319	0.133	0.000	0.867
320	0.432	0.000	0.568
321	0.207	0.338	0.455
322	1.000	0.000	0.000
323	0.033	0.664	0.303
324	0.977	0.003	0.020
325	0.926	0.002	0.072
326	0.227	0.000	0.773
327	0.997	0.000	0.003
328	1.000	0.000	0.000
329	1.000	0.000	0.000
330	1.000	0.000	0.000
331	0.013	0.000	0.987
332	0.968	0.000	0.032
333	1.000	0.000	0.000
334	0.993	0.000	0.007
335	0.989	0.000	0.011
336	1.000	0.000	0.000
337	0.928	0.021	0.050
338	0.773	0.000	0.227
339	0.990	0.000	0.010
340	0.015	0.000	0.985
341	0.009	0.000	0.991
342	0.009	0.001	0.991
343	0.024	0.000	0.976
344	0.006	0.000	0.994
345	0.022	0.000	0.978
346	0.008	0.000	0.992
347	0.010	0.000	0.990
348	0.005	0.000	0.995
349	0.041	0.000	0.959
350	0.014	0.000	0.986
351	0.017	0.000	0.982
352	0.142	0.000	0.858
353	0.023	0.000	0.977
354	0.021	0.004	0.975
355	0.079	0.002	0.919
356	0.050	0.001	0.949
357	0.029	0.001	0.970
358	0.041	0.000	0.959
359	0.023	0.000	0.977
360	0.380	0.000	0.620

Table 17 (continued)

	PROBABILITIES OF TYPE MEMBERSHIP		
	1	2	3
361	0.032	0.937	0.031
362	0.117	0.295	0.588
363	0.084	0.000	0.916
364	0.199	0.000	0.801
365	0.430	0.000	0.570
366	0.009	0.011	0.980
367	0.020	0.926	0.053
368	0.072	0.005	0.922
369	0.027	0.032	0.940
370	0.991	0.000	0.009
371	0.314	0.120	0.565
372	0.429	0.000	0.571
373	0.198	0.000	0.802
374	0.072	0.611	0.317
375	0.011	0.926	0.063
376	0.015	0.001	0.984
377	0.009	0.000	0.991
378	0.014	0.001	0.986
379	0.010	0.000	0.990
380	0.018	0.000	0.982
381	0.100	0.102	0.798
382	0.024	0.006	0.970
383	0.031	0.785	0.184
384	0.027	0.951	0.022
385	0.019	0.905	0.076
386	0.007	0.888	0.105
387	0.025	0.684	0.291
388	0.027	0.961	0.012
389	0.005	0.786	0.209
390	0.005	0.957	0.038
391	0.014	0.514	0.472
392	0.025	0.357	0.618
393	0.022	0.801	0.177
394	0.010	0.295	0.695
395	0.149	0.138	0.712
396	0.005	0.784	0.212
397	0.005	0.813	0.182
398	0.023	0.481	0.495
399	0.012	0.048	0.939
400	0.238	0.000	0.762
401	0.024	0.001	0.975
402	0.106	0.001	0.894
403	0.019	0.001	0.980
404	0.974	0.001	0.024
405	1.000	0.000	0.000
406	1.000	0.000	0.000
407	1.000	0.000	0.000
408	1.000	0.000	0.000
409	1.000	0.000	0.000
410	0.981	0.000	0.019
411	0.702	0.298	0.000
412	0.997	0.000	0.003
413	0.044	0.772	0.184
414	0.010	0.719	0.271
415	0.041	0.010	0.949
416	0.026	0.005	0.968
417	0.021	0.134	0.845
418	0.011	0.008	0.981
419	0.012	0.001	0.988
420	0.024	0.002	0.975
421	0.007	0.058	0.935
422	0.010	0.012	0.978
423	0.136	0.039	0.825
424	0.053	0.004	0.943
425	0.222	0.058	0.720
426	0.993	0.000	0.007
427	0.096	0.010	0.894
428	0.183	0.002	0.816
429	0.999	0.000	0.001
430	0.009	0.158	0.833
431	0.022	0.473	0.505
432	0.008	0.047	0.945
433	0.005	0.001	0.995
434	0.008	0.013	0.980
435	0.003	0.514	0.483
436	0.012	0.652	0.336
437	0.002	0.956	0.043
438	0.003	0.878	0.119
439	0.998	0.000	0.002
440	0.891	0.022	0.086
441	0.072	0.002	0.925
442	0.005	0.000	0.995
443	0.064	0.000	0.936
444	0.896	0.000	0.104
445	0.185	0.000	0.814
446	0.008	0.000	0.992
447	0.038	0.000	0.962
448	0.007	0.000	0.993
449	0.011	0.000	0.989
450	0.012	0.000	0.988
451	0.043	0.002	0.955
452	0.894	0.000	0.106
453	0.101	0.000	0.899
454	0.062	0.000	0.938
455	0.082	0.000	0.918
456	0.027	0.000	0.973
457	0.015	0.000	0.985
458	0.015	0.001	0.984
459	0.018	0.197	0.785
460	0.008	0.656	0.335
461	0.038	0.032	0.930
462	0.039	0.957	0.004
463	0.039	0.304	0.658
464	0.010	0.353	0.637
465	0.023	0.583	0.394
466	0.006	0.126	0.868
467	0.177	0.029	0.794
468	0.135	0.000	0.865
469	0.075	0.063	0.861
470	0.577	0.000	0.423
471	1.000	0.000	0.000
472	1.000	0.000	0.000
473	1.000	0.000	0.000
474	1.000	0.000	0.000
475	0.967	0.000	0.033
476	0.636	0.000	0.364
477	0.122	0.000	0.878
478	0.056	0.000	0.944
479	0.183	0.000	0.817
480	0.032	0.000	0.968
481	0.704	0.206	0.090
482	0.048	0.000	0.952
483	0.052	0.000	0.948
484	0.009	0.070	0.921
485	0.015	0.766	0.220
486	0.091	0.015	0.894
487	0.129	0.000	0.871
488	0.032	0.024	0.943
489	0.123	0.000	0.877
490	0.119	0.000	0.881
491	1.000	0.000	0.000
492	1.000	0.000	0.000
493	0.061	0.089	0.850
494	0.323	0.001	0.576
495	0.136	0.001	0.864
496	0.041	0.006	0.952
497	0.419	0.000	0.581
498	0.063	0.000	0.937
499	0.023	0.000	0.977
500	0.020	0.000	0.980
501	0.847	0.000	0.153
502	0.113	0.000	0.887
503	0.989	0.008	0.003

```
LOGARITHM OF LIKELIHOOD RATIO OF  3 TO  2 TYPES =     0.57933643D 02
CHI-SQUARE WITH 54 DFGREES OF FREEDOM=     114.02
PROBABILITY OF NULL HYPOTHESIS= 0.00000349
```

Table 18 Separation of a multivariate (6) set of Rantoul, Illinois, October 1950-55, upper wind components, zonal and meridional mixed distribution, at the 700-, 500-, and 300-mb levels into four distributions. The units are $m \cdot s^{-1}$.

```
                         PROGRAM NORMIX
          WOLFE NORMAL MIXTURE ANALYSIS PROCEDURE(1974 REVISION)


 SAMPLE = RANTOUL
          OCTOBER
          700,500,AND 300MB
          X AND Y COMPONENTS


       SAMPLE SIZE =          503
       NUMBER OF VARIABLES =  6
       NUMBER OF TYPES =      4

                     ITERATION NUMBER    0

LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =   -0.727177290 04

                  CHARACTERISTICS OF THE WHOLE SAMPLE


                               MEANS
    1             2             3            4             5            6
 7.3252       -0.9486       10.6700      -2.5704       15.4049      -3.6999

                       STANDARD DEVIATIONS
    1             2             3            4             5            6
 6.6262        6.8921        8.9594      10.3149       13.7082      16.0058

                          CORRELATIONS
    1             2             3            4             5            6
 1.0000        0.2215        0.7525       0.2313        0.5589       0.2311
 0.2215        1.0000        0.1138       0.7581        0.1265       0.5757
 0.7525        0.1138        1.0000       0.1587        0.7513       0.1977
 0.2313        0.7581        0.1587       1.0000        0.1200       0.7856
 0.5589        0.1265        0.7513       0.1200        1.0000       0.1266
 0.2311        0.5757        0.1977       0.7856        0.1266       1.0000

                  CHARACTERISTICS OF TYPE   1

            THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.406

                               MEANS
    1             2             3            4             5            6
11.2762        2.4484       14.1460       1.2900       20.3731       5.8172

                       STANDARD DEVIATIONS
    1             2             3            4             5            6
 5.5488        4.8968        7.8421       8.8618       10.7092      13.4027

                          CORRELATIONS
    1             2             3            4             5            6
 1.0000        0.1304        0.6960       0.1793        0.4565       0.0233
 0.1304        1.0000        0.0569       0.6547       -0.0423       0.4006
 0.6960        0.0569        1.0000       0.1498        0.6851       0.0924
 0.1793        0.6547        0.1498       1.0000        0.0492       0.7451
 0.4565       -0.0423        0.6851       0.0492        1.0000      -0.0354
 0.0233        0.4006        0.0924       0.7451       -0.0354       1.0000

                  CHARACTERISTICS OF TYPE   2

            THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.189

                               MEANS
    1             2             3            4             5            6
 3.7585        4.1860        7.5600       2.5486       15.1635      -3.0406

                       STANDARD DEVIATIONS
    1             2             3            4             5            6
 5.5488        4.8968        7.8421       8.8618       10.7092      13.4027

                          CORRELATIONS
    1             2             3            4             5            6
 1.0000        0.1304        0.6960       0.1793        0.4565       0.0233
 0.1304        1.0000        0.0569       0.6547       -0.0423       0.4006
 0.6960        0.0569        1.0000       0.1498        0.6851       0.0924
 0.1793        0.6547        0.1498       1.0000        0.0492       0.7451
 0.4565       -0.0423        0.6851       0.0492        1.0000      -0.0354
 0.0233        0.4006        0.0924       0.7451       -0.0354       1.0000

                  CHARACTERISTICS OF TYPE   3

            THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.310

                               MEANS
    1             2             3            4             5            6
 6.3438       -7.1940       11.4481      -9.7104       16.9286     -14.1947
```

Table 18 (continued)

```
                                  STANDARD DEVIATIONS
      1              2              3              4              5              6
   5.5488         4.8968         7.8421         8.8618        10.7092        13.4027

                                     CORRELATIONS
      1              2              3              4              5              6
   1.0000         0.1304         0.6960         0.1793         0.4565         0.0233
   0.1304         1.0000         0.0569         0.6547        -0.0423         0.4006
   0.6960         0.0569         1.0000         0.1498         0.6851         0.0924
   0.1793         0.6547         0.1498         1.0000         0.0492         0.7451
   0.4565        -0.0423         0.6851         0.0492         1.0000        -0.0354
   0.0233         0.4006         0.0924         0.7451        -0.0354         1.0000

                            CHARACTERISTICS OF TYPE    4

                  THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.095

                                         MEANS
      1              2              3              4              5              6
   0.7826        -5.2503        -0.4771        -5.9032       -10.1840       -11.3442

                                  STANDARD DEVIATIONS
      1              2              3              4              5              6
   5.5488         4.8968         7.8421         8.8618        10.7092        13.4027

                                     CORRELATIONS
      1              2              3              4              5              6
   1.0000         0.1304         0.6960         0.1793         0.4565         0.0233
   0.1304         1.0000         0.0569         0.6547        -0.0423         0.4006
   0.6960         0.0569         1.0000         0.1498         0.6851         0.0924
   0.1793         0.6547         0.1498         1.0000         0.0492         0.7451
   0.4565        -0.0423         0.6851         0.0492         1.0000        -0.0354
   0.0233         0.4006         0.0924         0.7451        -0.0354         1.0000
```

```
ITERATION    1          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.75273606D 04
ITERATION    2          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.742158890 04
ITERATION    3          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.73726356D 04
ITERATION    4          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.73415493D 04
ITERATION    5          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.731713750 04
ITERATION    6          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72971605D 04
ITERATION    7          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72839862D 04
ITERATION    8          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.727597600 04
ITERATION    9          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72705389D 04
ITERATION   10          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72664532D 04
ITERATION   11          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72632691D 04
ITERATION   12          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72607145D 04
ITERATION   13          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72585786D 04
ITERATION   14          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72567352D 04
ITERATION   15          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72551401D 04
ITERATION   16          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72537853D 04
ITERATION   17          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72526475D 04
ITERATION   18          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72640087D 04
ITERATION   19          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72492749D 04
ITERATION   20          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724790960 04
ITERATION   21          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72557666D 04
ITERATION   22          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72464581D 04
ITERATION   23          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724571510 04
ITERATION   24          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72482515D 04
ITERATION   25          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724469450 04
ITERATION   26          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72442351D 04
ITERATION   27          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72488781D 04
ITERATION   28          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72429321D 04
ITERATION   29          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72421681D 04
ITERATION   30          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724726580 04
ITERATION   31          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724015590 04
ITERATION   32          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72393329D 04
ITERATION   33          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.724444610 04
ITERATION   34          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723802560 04
ITERATION   35          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723748540 04
ITERATION   36          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723803760 04
ITERATION   37          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72370254D 04
ITERATION   38          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723682500 04
ITERATION   39          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72393245D 04
ITERATION   40          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723641500 04
ITERATION   41          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723614200 04
ITERATION   42          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72376909D 04
ITERATION   43          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72358764D 04
ITERATION   44          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72356825D 04
ITERATION   45          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723977760 04
ITERATION   46          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72354601D 04
ITERATION   47          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723520430 04
ITERATION   48          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72354146D 04
ITERATION   49          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723494060 04
ITERATION   50          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72348226D 04
ITERATION   51          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723499140 04
ITERATION   52          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723428970 04
ITERATION   53          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72339124D 04
ITERATION   54          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72365962D 04
ITERATION   55          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723365670 04
ITERATION   56          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72330784D 04
ITERATION   57          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.723659940 04
```

Table 18 (continued)

```
ITERATION  58          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72327417D 04
ITERATION  59          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72317961D 04
ITERATION  60          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72391156D 04
ITERATION  61          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72329895D 04
ITERATION  62          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72317396D 04
ITERATION  63          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72312275D 04
ITERATION  64          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72373290D 04
ITERATION  65          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72325621D 04
ITERATION  66          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72315565D 04
ITERATION  67          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72310672D 04
ITERATION  68          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72308061D 04
ITERATION  69          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72341678D 04
ITERATION  70          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72314630D 04
ITERATION  71          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72308762D 04
ITERATION  72          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72306692D 04
ITERATION  73          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305889D 04
ITERATION  74          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305538D 04
ITERATION  75          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305298D 04
ITERATION  76          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305978D 04
ITERATION  77          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305359D 04
ITERATION  78          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305249D 04
ITERATION  79          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305211D 04
ITERATION  80          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305414D 04
ITERATION  81          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305229D 04
ITERATION  82          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305195D 04
ITERATION  83          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305183D 04
ITERATION  84          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305197D 04
ITERATION  85          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305173D 04
ITERATION  86          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305169D 04
ITERATION  87          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305172D 04
ITERATION  88          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305164D 04
ITERATION  89          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305160D 04
ITERATION  90          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305161D 04
ITERATION  91          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305152D 04
ITERATION  92          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -D.72305147D 04
ITERATION  93          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305180D 04
ITERATION  94          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305142D 04
ITERATION  95          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305128D 04
ITERATION  96          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305156D 04
ITERATION  97          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305112D 04
ITERATION  98          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305094D 04
ITERATION  99          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305095D 04
ITERATION 100          LOG LIKELIHOOD OF  4 TYPES IN THIS SAMPLE =    -0.72305057D 04

                       ITERATION TIME =      5343.00 SECONDS.
                       TOTAL TIME USED =    11562.01 SECONDS.

                             ITERATION NUMBER 101

          LIKELIHOOD OF   4 TYPES IN THIS SAMPLE =    -0.72305057D 04

                       CHARACTERISTICS OF THE WHOLE SAMPLE

                                   MEANS
      1             2             3             4             5             6
   7.3252       -0.9486       10.6700       -2.5704       15.4049       -3.6999

                            STANDARD DEVIATIONS
      1             2             3             4             5             6
   6.6262        6.8921        8.9594       10.3149       13.7082       16.0058

                               CORRELATIONS
      1             2             3             4             5             6
   1.0000        0.2215        0.7525        0.2313        0.5589        0.2311
   0.2215        1.0000        0.1138        0.7581        0.1265        0.5757
   0.7525        0.1138        1.0000        0.1587        0.7513        0.1977
   0.2313        0.7581        0.1587        1.0000        0.1200        0.7856
   0.5589        0.1265        0.7513        0.1200        1.0000        0.1266
   0.2311        0.5757        0.1977        0.7856        0.1266        1.0000

                       CHARACTERISTICS OF TYPE   1

               THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.193

                                   MEANS
      1             2             3             4             5             6
  11.1041       -0.3525       16.6560       -1.3465       22.8856       -0.6482

                            STANDARD DEVIATIONS
      1             2             3             4             5             6
   7.2047        9.2886       11.3528       17.2976       14.8718       27.5962

                               CORRELATIONS
      1             2             3             4             5             6
   1.0000        0.2647        0.5390        0.3324        0.5169        0.2525
   0.2647        1.0000       -0.1035        0.7443       -0.1470        0.6119
   0.5390       -0.1035        1.0000        0.0730        0.7250        0.0848
   0.3324        0.7443        0.0730        1.0000       -0.0241        0.7911
   0.5169       -0.1470        0.7250       -0.0241        1.0000       -0.1123
   0.2525        0.6119        0.0848        0.7911       -0.1123        1.0000
```

Table 18 (continued)

```
EIGENVALUES    EIGENVECTORS
                  1          2          3          4          5          6
1021.7463      0.0666     0.2519     0.1199     0.1879     0.5660     0.7497
 326.0660      0.2057    -0.0519     0.4355    -0.0397     0.6673    -0.5643
 104.5090      0.0270     0.5524    -0.1959     0.7483    -0.0513    -0.3054
  40.0002      0.4817     0.0627     0.7053     0.1472    -0.4735     0.1439
  33.4545     -0.0490     0.7903     0.0619    -0.5999    -0.0621    -0.0739
  23.2217      0.8474    -0.0148    -0.5063    -0.1474     0.0608     0.0017

                 CHARACTERISTICS OF TYPE   2

          THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.293

                                 MEANS
    1            2            3            4            5            6
 6.7197       1.7879      11.3394       0.7926      20.0058      -3.5688

                          STANDARD DEVIATIONS
    1            2            3            4            5            6
 5.2062       5.7219       6.7337       5.1816       8.3725       8.0912

                              CORRELATIONS
    1            2            3            4            5            6
 1.0000      -0.0490       0.7322       0.2324       0.5236       0.4435
-0.0490       1.0000      -0.2795       0.6794      -0.3196       0.4835
 0.7322      -0.2795       1.0000      -0.1278       0.7437       0.1412
 0.2324       0.6794      -0.1278       1.0000      -0.2774       0.7870
 0.5236      -0.3196       0.7437      -0.2774       1.0000      -0.0414
 0.4435       0.4835       0.1412       0.7870      -0.0414       1.0000

EIGENVALUES    EIGENVECTORS
                  1          2          3          4          5          6
 120.3379      0.3038     0.2911    -0.2839     0.3728     0.6914    -0.3541
 102.2809     -0.2543     0.3305     0.7201     0.4626    -0.1243    -0.2796
  20.2525      0.5440     0.1697    -0.2136     0.5162    -0.5472     0.2525
  13.1203     -0.1758     0.4315     0.0912     0.0344     0.3317     0.8145
   6.8706      0.7153     0.0421     0.5396    -0.4197     0.1338     0.0349
   4.7389     -0.0663     0.7677    -0.2360    -0.4509    -0.2815    -0.2609


                 CHARACTERISTICS OF TYPE   3

          THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.421

                                 MEANS
    1            2            3            4            5            6
 6.3830      -2.6212       8.3100      -5.2661      10.5901      -4.5639

                          STANDARD DEVIATIONS
    1            2            3            4            5            6
 6.6569       5.5893       7.5718       7.9899      11.9976      12.0984

                              CORRELATIONS
    1            2            3            4            5            6
 1.0000       0.3045       0.8419       0.1154       0.6065       0.1489
 0.3045       1.0000       0.3415       0.7889       0.1293       0.6263
 0.8419       0.3415       1.0000       0.1965       0.8310       0.3252
 0.1154       0.7889       0.1965       1.0000       0.0264       0.8370
 0.6065       0.1293       0.8310       0.0264       1.0000       0.3023
 0.1489       0.6263       0.3252       0.8370       0.3023       1.0000

EIGENVALUES    EIGENVECTORS
                  1          2          3          4          5          6
 266.8513      0.2422     0.2551     0.6124    -0.4722    -0.1661     0.5007
 159.8561      0.2087    -0.1989     0.4059     0.6391    -0.5800    -0.0858
  37.9134      0.3663     0.2998     0.3293    -0.1248     0.2361    -0.7721
  12.6521      0.3106    -0.4448     0.2560     0.2736     0.7070     0.2558
   5.5168      0.5554     0.5782    -0.4125     0.3347     0.0417     0.2707
   4.2504      0.5989    -0.5228    -0.3407    -0.4076    -0.2805    -0.0841


                 CHARACTERISTICS OF TYPE   4

          THE PROPORTION OF THE POPULATION FROM THIS TYPE= 0.094

                                 MEANS
    1            2            3            4            5            6
 5.6571      -3.2181       6.8328      -3.4952       7.2228      -6.5235

                          STANDARD DEVIATIONS
    1            2            3            4            5            6
 6.3812       6.6990       8.7872       8.7555      18.2393      16.5218

                              CORRELATIONS
    1            2            3            4            5            6
 1.0000       0.3470       0.9558       0.3597       0.4762       0.1332
 0.3470       1.0000       0.4550       0.9402       0.6151       0.6943
 0.9558       0.4550       1.0000       0.4338       0.4895       0.2214
 0.3597       0.9402       0.4338       1.0000       0.5464       0.8140
 0.4762       0.6151       0.4895       0.5464       1.0000       0.3738
 0.1332       0.6943       0.2214       0.8140       0.3738       1.0000
```

Table 18 (continued)

EIGENVALUES	EIGENVECTORS 1	2	3	4	5	6
529.7700	0.1370	-0.1791	0.5272	-0.1521	0.5265	0.6091
204.9783	0.2462	0.0753	0.0656	0.6255	-0.5210	0.5164
80.2752	0.2120	-0.2016	0.7391	-0.[illegible]347	-0.4152	-0.4215
25.7264	0.3267	0.1956	0.1260	0.6207	0.5243	-0.4232
3.3614	0.6787	-0.6084	-0.3929	-0.1170	0.0120	-0.0310
1.0011	0.5552	0.7164	-0.0352	-0.4105	-0.0634	0.0685

PROBABILITIES
OF TYPE MEMBERSHIP

	1	2	3	4
1	0.622	0.051	0.327	0.000
2	0.007	0.992	0.001	0.000
3	0.008	0.991	0.002	0.000
4	1.000	0.000	0.000	0.000
5	0.004	0.996	0.000	0.000
6	0.022	0.978	0.000	0.000
7	0.036	0.031	0.933	0.001
8	1.000	0.000	0.000	0.000
9	0.060	0.935	0.005	0.000
10	0.699	0.000	0.008	0.293
11	1.000	0.000	0.000	0.000
12	0.006	0.072	0.837	0.085
13	0.018	0.469	0.321	0.191
14	0.015	0.127	0.839	0.019
15	0.002	0.000	0.000	0.998
16	0.991	0.000	0.008	0.000
17	0.017	0.014	0.004	0.965
18	0.223	0.013	0.764	0.000
19	0.366	0.046	0.589	0.000
20	1.000	0.000	0.000	0.000
21	0.011	0.841	0.147	0.001
22	0.059	0.000	0.158	0.783
23	0.307	0.000	0.692	0.001
24	0.030	0.002	0.967	0.000
25	0.006	0.758	0.195	0.041
26	0.000	0.000	0.000	1.000
27	0.132	0.124	0.743	0.001
28	0.011	0.984	0.005	0.001
29	0.155	0.844	0.001	0.000
30	0.820	0.000	0.180	0.000
31	0.025	0.003	0.972	0.000
32	0.015	0.000	0.985	0.000
33	0.013	0.061	0.894	0.032
34	0.052	0.000	0.948	0.000
35	0.729	0.000	0.271	0.000
36	0.008	0.085	0.724	0.184
37	0.011	0.497	0.478	0.014
38	0.007	0.937	0.054	0.001
39	0.005	0.881	0.024	0.091
40	0.005	0.968	0.024	0.003
41	0.944	0.044	0.012	0.000
42	0.012	0.817	0.025	0.146
43	0.038	0.153	0.580	0.229
44	0.021	0.638	0.236	0.106
45	0.056	0.720	0.224	0.000
46	0.096	0.022	0.869	0.013
47	0.032	0.757	0.186	0.025
48	0.962	0.000	0.038	0.000
49	0.009	0.001	0.876	0.113
50	0.016	0.001	0.956	0.027
51	0.012	0.000	0.980	0.007
52	0.424	0.000	0.576	0.000
53	0.007	0.000	0.943	0.050
54	1.000	0.000	0.000	0.000
55	0.004	0.000	0.863	0.132
56	0.008	0.000	0.946	0.046
57	0.008	0.000	0.780	0.212
58	0.147	0.007	0.846	0.000
59	0.004	0.000	0.451	0.545
60	0.004	0.001	0.842	0.153
61	0.003	0.003	0.859	0.135
62	0.025	0.826	0.103	0.046
63	0.009	0.151	0.721	0.119
64	0.009	0.176	0.704	0.111
65	0.045	0.660	0.271	0.024
66	0.020	0.651	0.329	0.000
67	0.027	0.399	0.574	0.000
68	0.009	0.694	0.270	0.027
69	0.009	0.424	0.555	0.012
70	0.011	0.488	0.480	0.020
71	0.015	0.866	0.118	0.000
72	0.007	0.960	0.021	0.013
73	0.010	0.950	0.038	0.002
74	0.008	0.855	0.059	0.078
75	0.038	0.867	0.045	0.050
76	0.025	0.546	0.370	0.059
77	0.035	0.961	0.004	0.000
78	0.028	0.017	0.955	[illegible]
79	0.051	0.696	0.019	[illegible]
80	0.041	0.834	0.124	0.001

PROBABILITIES
OF TYPE MEMBERSHIP

	1	2	3	4
81	0.100	0.255	0.645	0.000
82	0.027	0.872	0.012	0.088
83	0.974	0.020	0.004	0.003
84	0.064	0.933	0.003	0.000
85	0.077	0.838	0.084	0.000
86	0.020	0.786	0.193	0.000
87	0.036	0.786	0.163	0.015
88	0.028	0.803	0.019	0.150
89	0.024	0.932	0.043	0.000
90	0.375	0.406	0.016	0.203
91	0.079	0.029	0.893	0.000
92	0.029	0.019	0.942	0.010
93	0.973	0.001	0.027	0.000
94	0.044	0.671	0.001	0.284
95	0.032	0.169	0.797	0.001
96	0.946	0.012	0.042	0.000
97	0.040	0.003	0.957	0.000
98	0.064	0.893	0.043	0.000
99	0.056	0.000	0.943	0.000
100	0.018	0.026	0.845	0.111
101	0.016	0.008	0.881	0.094
102	0.847	0.047	0.106	0.000
103	0.996	0.003	0.001	0.000
104	0.011	0.005	0.884	0.100
105	0.012	0.659	0.328	0.001
106	0.016	0.009	0.689	0.286
107	0.097	0.000	0.882	0.021
108	0.004	0.000	0.957	0.039
109	0.007	0.000	0.529	0.464
110	0.004	0.002	0.833	0.161
111	0.008	0.000	0.978	0.013
112	0.003	0.001	0.995	0.002
113	0.006	0.000	0.978	0.015
114	0.023	0.000	0.403	0.574
115	0.128	0.414	0.400	0.059
116	0.009	0.035	0.899	0.057
117	0.014	0.901	0.085	0.000
118	0.083	0.447	0.011	0.458
119	0.031	0.078	0.815	0.076
120	0.010	0.568	0.352	0.070
121	0.028	0.950	0.022	0.000
122	0.019	0.000	0.981	0.000
123	0.032	0.483	0.415	0.071
124	0.093	0.681	0.142	0.084
125	0.416	0.583	0.001	0.000
126	0.073	0.001	0.926	0.000
127	0.549	0.450	0.001	0.000
128	0.031	0.000	0.002	0.967
129	0.005	0.000	0.002	0.993
130	0.014	0.012	0.973	0.000
131	0.005	0.000	0.949	0.046
132	0.005	0.000	0.544	0.450
133	0.012	0.000	0.109	0.879
134	0.006	0.007	0.958	0.030
135	0.030	0.007	0.938	0.026
136	0.014	0.005	0.956	0.024
137	0.026	0.004	0.875	0.095
138	0.009	0.005	0.653	0.333
139	0.019	0.000	0.981	0.000
140	0.003	0.000	0.015	0.983
141	0.000	0.000	0.000	1.000
142	0.024	0.000	0.976	0.000
143	0.540	0.000	0.257	0.203
144	0.030	0.001	0.969	0.000
145	0.071	0.758	0.143	0.028
146	0.017	0.964	0.018	0.000
147	0.037	0.143	0.808	0.012
148	0.013	0.931	0.055	0.000
149	0.029	0.202	0.458	0.312
150	0.037	0.074	0.887	0.003
151	0.096	0.000	0.904	0.000
152	0.236	0.008	0.756	0.000
153	0.258	0.035	0.564	0.143
154	0.103	0.000	0.897	0.000
155	0.064	0.000	0.936	0.000
156	0.028	0.073	0.777	0.122
157	0.037	0.871	0.011	0.082
158	0.018	0.353	0.615	0.014
159	0.114	0.001	0.880	0.005
160	0.012	0.837	0.151	0.000

PROBABILITIES
OF TYPE MEMBERSHIP

	1	2	3	4
161	0.011	0.228	0.761	0.000
162	0.021	0.891	0.088	0.000
163	0.009	0.984	0.007	0.000
164	0.006	0.978	0.016	0.000
165	0.015	0.593	0.383	0.010
166	0.004	0.919	0.015	0.062
167	0.011	0.879	0.000	0.109
168	0.005	0.989	0.005	0.000
169	0.004	0.919	0.027	0.050
170	0.004	0.952	0.012	0.032
171	0.006	0.965	0.003	0.027
172	0.232	0.707	0.000	0.062
173	0.037	0.619	0.332	0.012
174	0.017	0.778	0.041	0.164
175	0.022	0.006	0.960	0.013
176	0.171	0.080	0.749	0.000
177	0.087	0.010	0.902	0.000
178	0.012	0.078	0.774	0.135
179	0.796	0.005	0.188	0.010
180	0.949	0.051	0.000	0.000
181	0.997	0.003	0.000	0.000
182	1.000	0.000	0.000	0.000
183	0.988	0.000	0.012	0.000
184	0.157	0.000	0.843	0.000
185	0.054	0.934	0.012	0.000
186	0.021	0.000	0.935	0.044
187	0.012	0.533	0.390	0.065
188	0.010	0.803	0.180	0.007
189	0.007	0.688	0.284	0.022
190	0.010	0.066	0.778	0.146
191	0.072	0.668	0.258	0.002
192	0.024	0.000	0.975	0.001
193	0.006	0.158	0.697	0.139
194	0.008	0.411	0.502	0.079
195	0.013	0.024	0.815	0.148
196	0.005	0.000	0.992	0.003
197	0.153	0.000	0.847	0.000
198	0.023	0.000	0.951	0.026
199	0.016	0.001	0.981	0.002
200	0.032	0.095	0.861	0.011
201	0.026	0.000	0.011	0.963
202	1.000	0.000	0.000	0.000
203	0.139	0.000	0.855	0.006
204	1.000	0.000	0.000	0.000
205	0.006	0.000	0.994	0.000
206	0.017	0.066	0.917	0.000
207	0.006	0.000	0.992	0.002
208	0.007	0.000	0.993	0.000
209	0.018	0.000	0.982	0.000
210	0.004	0.001	0.996	0.000
211	0.006	0.001	0.017	0.976
212	0.007	0.001	0.977	0.016
213	0.510	0.420	0.070	0.000
214	0.008	0.000	0.543	0.448
215	0.002	0.000	0.734	0.263
216	0.018	0.000	0.982	0.000
217	0.006	0.000	0.982	0.012
218	0.009	0.000	0.692	0.298
219	0.011	0.000	0.973	0.016
220	0.007	0.001	0.988	0.004
221	0.023	0.001	0.924	0.052
222	0.007	0.000	0.875	0.118
223	0.010	0.000	0.699	0.291
224	0.009	0.001	0.770	0.220
225	0.007	0.010	0.786	0.196
226	0.006	0.054	0.814	0.126
227	0.028	0.042	0.863	0.067
228	0.008	0.758	0.217	0.018
229	0.006	0.851	0.077	0.067
230	0.068	0.765	0.165	0.002
231	0.240	0.754	0.006	0.000
232	0.005	0.987	0.002	0.006
233	0.008	0.917	0.000	0.074
234	0.029	0.969	0.000	0.002
235	0.004	0.959	0.008	0.030
236	0.007	0.983	0.010	0.001
237	0.063	0.397	0.540	0.000
238	0.053	0.944	0.002	0.000
239	0.011	0.985	0.002	0.002
240	0.002	0.997	0.000	0.001

Table 18 (continued)

	PROBABILITIES OF TYPE MEMBERSHIP			
	1	2	3	4
241	0.010	0.923	0.000	0.067
242	0.799	0.177	0.024	0.000
243	0.005	0.982	0.009	0.004
244	0.016	0.863	0.119	0.003
245	0.028	0.901	0.071	0.000
246	0.055	0.316	0.111	0.518
247	0.972	0.000	0.028	0.000
248	0.155	0.000	0.844	0.001
249	0.036	0.000	0.964	0.000
250	0.025	0.000	0.975	0.000
251	0.017	0.124	0.858	0.001
252	0.006	0.552	0.415	0.026
253	0.009	0.962	0.028	0.001
254	0.016	0.085	0.604	0.295
255	0.236	0.017	0.747	0.000
256	0.933	0.012	0.003	0.052
257	0.104	0.012	0.548	0.337
258	0.183	0.000	0.803	0.014
259	0.096	0.000	0.904	0.000
260	0.034	0.000	0.711	0.254
261	0.032	0.002	0.958	0.008
262	0.017	0.863	0.119	0.001
263	0.011	0.314	0.551	0.123
264	0.035	0.063	0.898	0.005
265	0.017	0.033	0.945	0.005
266	0.004	0.000	0.979	0.016
267	0.004	0.000	0.674	0.323
268	0.032	0.000	0.968	0.000
269	0.143	0.000	0.857	0.000
270	0.005	0.000	0.766	0.229
271	0.004	0.000	0.966	0.029
272	0.007	0.217	0.622	0.154
273	0.010	0.005	0.982	0.003
274	0.029	0.762	0.088	0.121
275	0.018	0.095	0.883	0.005
276	0.033	0.009	0.767	0.191
277	0.060	0.061	0.699	0.180
278	0.083	0.006	0.786	0.125
279	0.043	0.798	0.013	0.146
280	0.039	0.420	0.000	0.540
281	0.019	0.950	0.006	0.026
282	0.013	0.916	0.070	0.000
283	0.043	0.914	0.043	0.000
284	0.022	0.965	0.005	0.008
285	0.091	0.909	0.000	0.001
286	0.947	0.053	0.000	0.001
287	0.961	0.035	0.004	0.000
288	0.067	0.005	0.929	0.000
289	0.272	0.001	0.727	0.000
290	0.022	0.936	0.042	0.000
291	0.412	0.015	0.525	0.049
292	0.025	0.003	0.972	0.000
293	0.989	0.000	0.011	0.000
294	0.109	0.832	0.000	0.059
295	0.049	0.523	0.001	0.427
296	0.122	0.081	0.751	0.046
297	0.124	0.001	0.875	0.000
298	0.078	0.238	0.561	0.123
299	0.806	0.194	0.000	0.000
300	0.042	0.958	0.001	0.000
301	0.125	0.875	0.000	0.000
302	0.162	0.737	0.093	0.008
303	0.018	0.980	0.003	0.000
304	0.006	0.972	0.016	0.006
305	0.052	0.907	0.040	0.000
306	0.059	0.000	0.941	0.000
307	0.074	0.009	0.654	0.263
308	0.020	0.357	0.420	0.203
309	0.014	0.908	0.036	0.042
310	0.016	0.026	0.953	0.005
311	0.083	0.032	0.878	0.007
312	0.051	0.344	0.599	0.006
313	0.126	0.080	0.792	0.002
314	0.070	0.008	0.922	0.000
315	0.075	0.066	0.859	0.000
316	0.995	0.005	0.000	0.000
317	0.299	0.037	0.070	0.593
318	0.245	0.010	0.735	0.010
319	0.887	0.017	0.096	0.000
320	0.943	0.009	0.049	0.000
321	0.180	0.761	0.057	0.001
322	1.000	0.000	0.000	0.000
323	0.073	0.841	0.085	0.000
324	0.806	0.194	0.000	0.000
325	0.934	0.065	0.001	0.000
326	0.545	0.014	0.441	0.000
327	0.984	0.000	0.000	0.016
328	1.000	0.000	0.000	0.000
329	1.000	0.000	0.000	0.000
330	1.000	0.000	0.000	0.000
331	0.004	0.000	0.950	0.046
332	0.919	0.000	0.081	0.000
333	1.000	0.000	0.000	0.000
334	0.999	0.000	0.001	0.000
335	0.994	0.006	0.000	0.000
336	1.000	0.000	0.000	0.000
337	0.890	0.110	0.000	0.000
338	0.925	0.004	0.071	0.000
339	0.999	0.000	0.001	0.000
340	0.027	0.009	0.964	0.000
341	0.024	0.000	0.911	0.065
342	0.014	0.014	0.966	0.007
343	0.011	0.000	0.433	0.556
344	0.002	0.000	0.820	0.178
345	0.014	0.000	0.402	0.585
346	0.002	0.000	0.674	0.324
347	0.010	0.002	0.949	0.040
348	0.002	0.001	0.883	0.114
349	0.014	0.000	0.258	0.729
350	0.004	0.000	0.882	0.114
351	0.005	0.002	0.921	0.072
352	0.005	0.000	0.995	0.000
353	0.005	0.001	0.957	0.036
354	0.009	0.027	0.854	0.110
355	0.074	0.060	0.824	0.041
356	0.046	0.012	0.941	0.001
357	0.026	0.003	0.944	0.027
358	0.039	0.001	0.924	0.036
359	0.026	0.001	0.836	0.137
360	0.994	0.002	0.004	0.000
361	0.099	0.894	0.000	0.006
362	0.243	0.253	0.500	0.004
363	0.129	0.040	0.829	0.002
364	0.186	0.004	0.805	0.005
365	0.120	0.000	0.043	0.836
366	0.011	0.034	0.951	0.003
367	0.049	0.948	0.003	0.000
368	0.037	0.068	0.000	0.895
369	0.042	0.392	0.328	0.238
370	0.999	0.000	0.001	0.000
371	0.158	0.772	0.058	0.011
372	0.661	0.000	0.339	0.000
373	0.338	0.006	0.655	0.001
374	0.076	0.870	0.054	0.000
375	0.016	0.982	0.002	0.000
376	0.017	0.002	0.913	0.068
377	0.009	0.002	0.987	0.002
378	0.014	0.007	0.829	0.150
379	0.010	0.001	0.942	0.046
380	0.018	0.005	0.870	0.107
381	0.101	0.736	0.161	0.002
382	0.025	0.424	0.469	0.082
383	0.026	0.910	0.064	0.000
384	0.067	0.921	0.012	0.000
385	0.038	0.920	0.042	0.000
386	0.013	0.971	0.005	0.011
387	0.028	0.933	0.009	0.029
388	0.066	0.934	0.000	0.000
389	0.005	0.984	0.008	0.003
390	0.010	0.988	0.002	0.000
391	0.018	0.893	0.089	0.000
392	0.022	0.970	0.009	0.000
393	0.021	0.979	0.000	0.000
394	0.017	0.932	0.047	0.004
395	0.041	0.959	0.000	0.000
396	0.006	0.990	0.003	0.000
397	0.011	0.967	0.018	0.004
398	0.024	0.922	0.054	0.000
399	0.016	0.616	0.366	0.001
400	0.403	0.001	0.596	0.000
401	0.039	0.139	0.822	0.001
402	0.116	0.583	0.294	0.007
403	0.033	0.501	0.464	0.002
404	0.986	0.011	0.003	0.000
405	1.000	0.000	0.000	0.000
406	1.000	0.000	0.000	0.000
407	1.000	0.000	0.000	0.000
408	1.000	0.000	0.000	0.000
409	1.000	0.000	0.000	0.000
410	0.944	0.004	0.052	0.000
411	0.941	0.059	0.000	0.000
412	1.000	0.000	0.000	0.000
413	0.065	0.686	0.001	0.248
414	0.011	0.967	0.005	0.017
415	0.056	0.162	0.735	0.046
416	0.031	0.054	0.863	0.052
417	0.026	0.234	0.713	0.026
418	0.017	0.117	0.866	0.000
419	0.008	0.017	0.934	0.041
420	0.022	0.003	0.714	0.261
421	0.010	0.097	0.864	0.028
422	0.013	0.159	0.827	0.000
423	0.098	0.891	0.011	0.000
424	0.136	0.830	0.034	0.000
425	0.151	0.846	0.003	0.000
426	1.000	0.000	0.000	0.000
427	0.100	0.013	0.797	0.089
428	0.162	0.002	0.633	0.203
429	1.000	0.000	0.000	0.000
430	0.010	0.288	0.550	0.152
431	0.037	0.730	0.230	0.002
432	0.010	0.061	0.906	0.024
433	0.007	0.007	0.934	0.053
434	0.010	0.097	0.893	0.000
435	0.005	0.886	0.075	0.035
436	0.025	0.694	0.281	0.000
437	0.004	0.990	0.005	0.002
438	0.006	0.963	0.011	0.020
439	0.432	0.000	0.000	0.568
440	0.319	0.680	0.000	0.000
441	0.041	0.958	0.001	0.001
442	0.003	0.000	0.867	0.130
443	0.009	0.000	0.016	0.975
444	0.000	0.000	0.000	1.000
445	0.049	0.001	0.001	0.949
446	0.004	0.000	0.896	0.100
447	0.158	0.000	0.515	0.327
448	0.008	0.000	0.915	0.076
449	0.009	0.000	0.767	0.224
450	0.006	0.000	0.924	0.071
451	0.048	0.021	0.306	0.624
452	0.058	0.000	0.005	0.937
453	0.180	0.000	0.820	0.000
454	0.065	0.000	0.929	0.006
455	0.056	0.005	0.491	0.449
456	0.019	0.000	0.981	0.000
457	0.022	0.001	0.975	0.001
458	0.017	0.003	0.979	0.001
459	0.015	0.599	0.308	0.079
460	0.012	0.846	0.122	0.020
461	0.038	0.657	0.215	0.090
462	0.111	0.887	0.001	0.000
463	0.034	0.448	0.004	0.514
464	0.011	0.583	0.367	0.039
465	0.024	0.650	0.008	0.319
466	0.010	0.330	0.624	0.036
467	0.228	0.288	0.022	0.461
468	0.833	0.133	0.034	0.000
469	0.151	0.842	0.002	0.005
470	0.315	0.002	0.000	0.682
471	1.000	0.000	0.000	0.000
472	1.000	0.000	0.000	0.000
473	1.000	0.000	0.000	0.000
474	1.000	0.000	0.000	0.000
475	0.998	0.000	0.002	0.000
476	0.822	0.000	0.177	0.001
477	0.134	0.000	0.866	0.000
478	0.066	0.000	0.927	0.007
479	0.231	0.000	0.769	0.000
480	0.028	0.000	0.972	0.000
481	0.413	0.582	0.005	0.000
482	0.080	0.000	0.887	0.033
483	0.093	0.000	0.906	0.001
484	0.011	0.238	0.607	0.144
485	0.032	0.942	0.025	0.000
486	0.478	0.188	0.011	0.324
487	0.859	0.076	0.014	0.052
488	0.105	0.335	0.559	0.001
489	0.115	0.000	0.001	0.884
490	0.637	0.002	0.361	0.000
491	1.000	0.000	0.000	0.000
492	1.000	0.000	0.000	0.000
493	0.097	0.723	0.180	0.000
494	0.245	0.222	0.531	0.002
495	0.063	0.017	0.129	0.791
496	0.046	0.156	0.796	0.002
497	0.601	0.012	0.387	0.000
498	0.106	0.002	0.893	0.000
499	0.024	0.000	0.976	0.000
500	0.018	0.000	0.982	0.000
501	0.995	0.000	0.005	0.000
502	0.145	0.000	0.855	0.000
503	0.984	0.015	0.000	0.001

```
LOGARITHM OF LIKELIHOOD RATIO OF  4 TO  3 TYPES =     0.412672840 02
CHI-SQUARE WITH 54 DEGREES OF FREEDOM=      81.14
PROBABILITY OF NULL HYPOTHESIS= 0.00986137
```

8.3.3.3 Figures and Discussion. Figures 20(a) through (f) illustrate the tabular output of table 16. In (a), under the assumption of only one cluster at each of three levels, i.e., a unimodal bivariate distribution, the ellipse shows the relative size of the distribution at the 700-, 500-, and 300-mb levels. The 0.50 probability ellipses all have more or less the same orientations. In (b), the 700-mb unimodal assumption is illustrated along with the two clusters in the assumed bimodal bivariate distribution. The same procedure is followed in (c) and (d) for the 500- and 300-mb levels. For further comparison, (e) and (f) show the level-to-level comparison for the cluster type 1 and the cluster type 2.

Figures 21(a) through (g) illustrate the tabular output of table 17. Figure 21(a) is the same as figure 20(a). Figure 21(b) shows the 700-mb unimodal distribution of figure 21(a) with the added three-cluster breakout at the 700-mb level. The same procedure follows through figures 21(c) and 21(d). Figures 21(e), (f), and (g) show the comparison of type distributions through the three levels. For example, figure 21(e) shows roughly the same orientation but greatly increased variance with altitudes. The same is true for figures 21(f) and (g).

Figures 22(a) through (h) follow the same procedural pattern as the previous figures. Its comparable table is table 18. Here there are four clusters. Again the initial unimodal distribution is shown in figure 22(a) while (b), (c), and (d) show the four clusters at each level. The following four show the cluster types at the three levels. Though the small numbers may be difficult to read, they appear also in the tables. Of major importance here is the pictorial display of the orientations and sizes of the ellipses or clusters.

The change in orientation with altitude is noted. Some strong change in orientation between types is also noted.

Further work will be done with these distributions to determine the relationship among these clusters, their orientation and dispersion and concurrent weather.

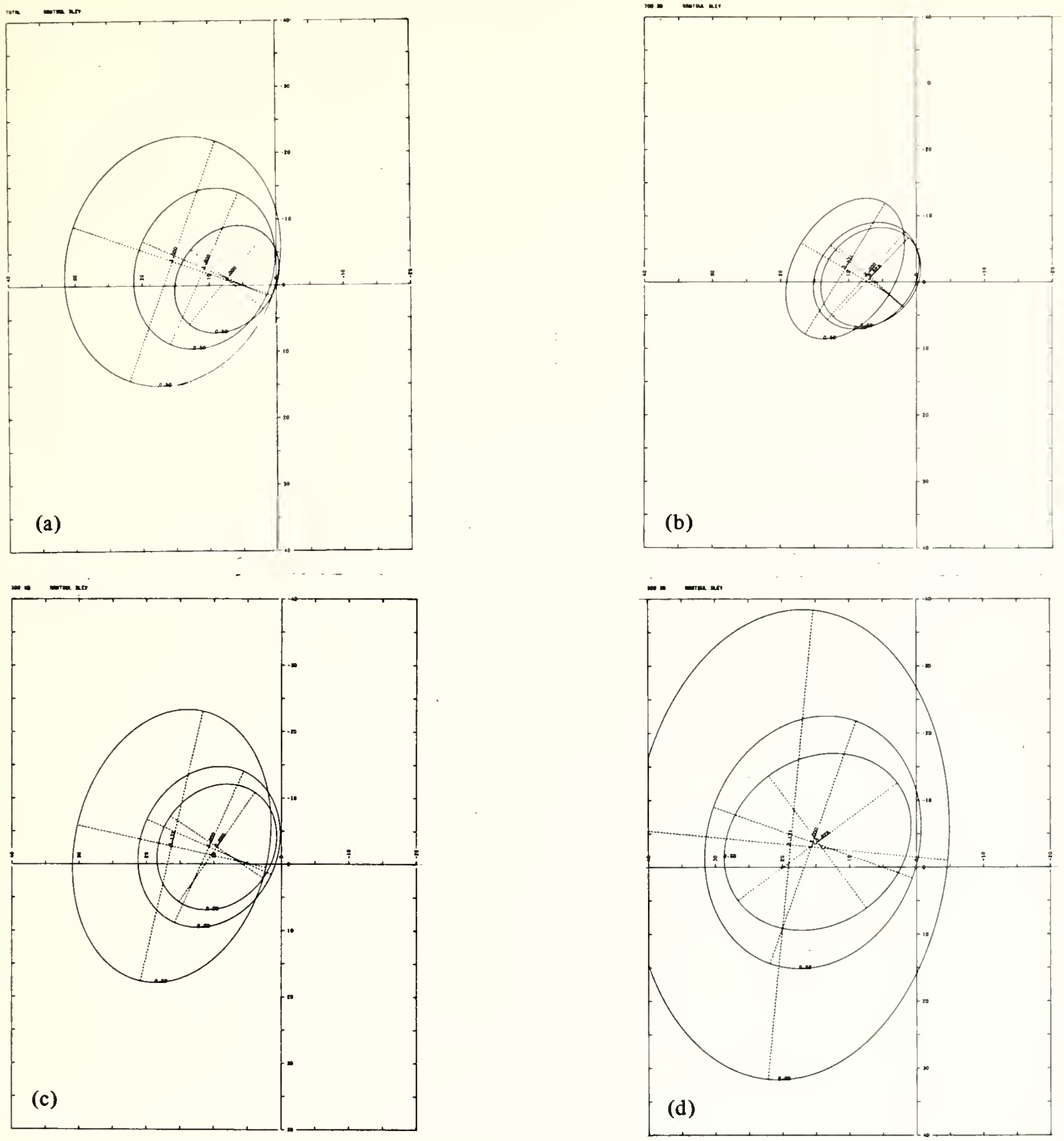


Figure 20 Bivariate distributions of winds in m·s^{-1} at Rantoul, Illinois, October 1950-1955 at the 700-, 500-, and 300-mb levels. Two cluster types (1 and 2) are assumed in the total mixed observed distribution (0.171 + 0.829 = 1.000). (a) Total distribution, (b) 700-mb mixed, (c) 500-mb mixed, (d) 300-mb mixed, (e) Type 1, and (f) Type 2.

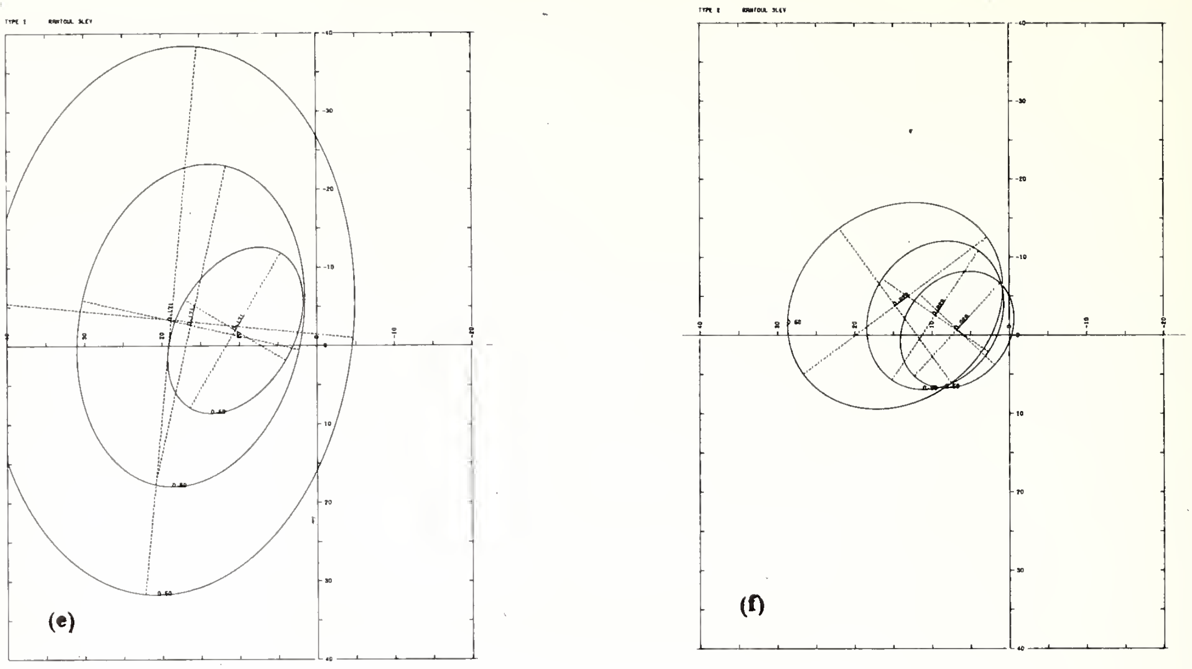


Figure 20 (continued)

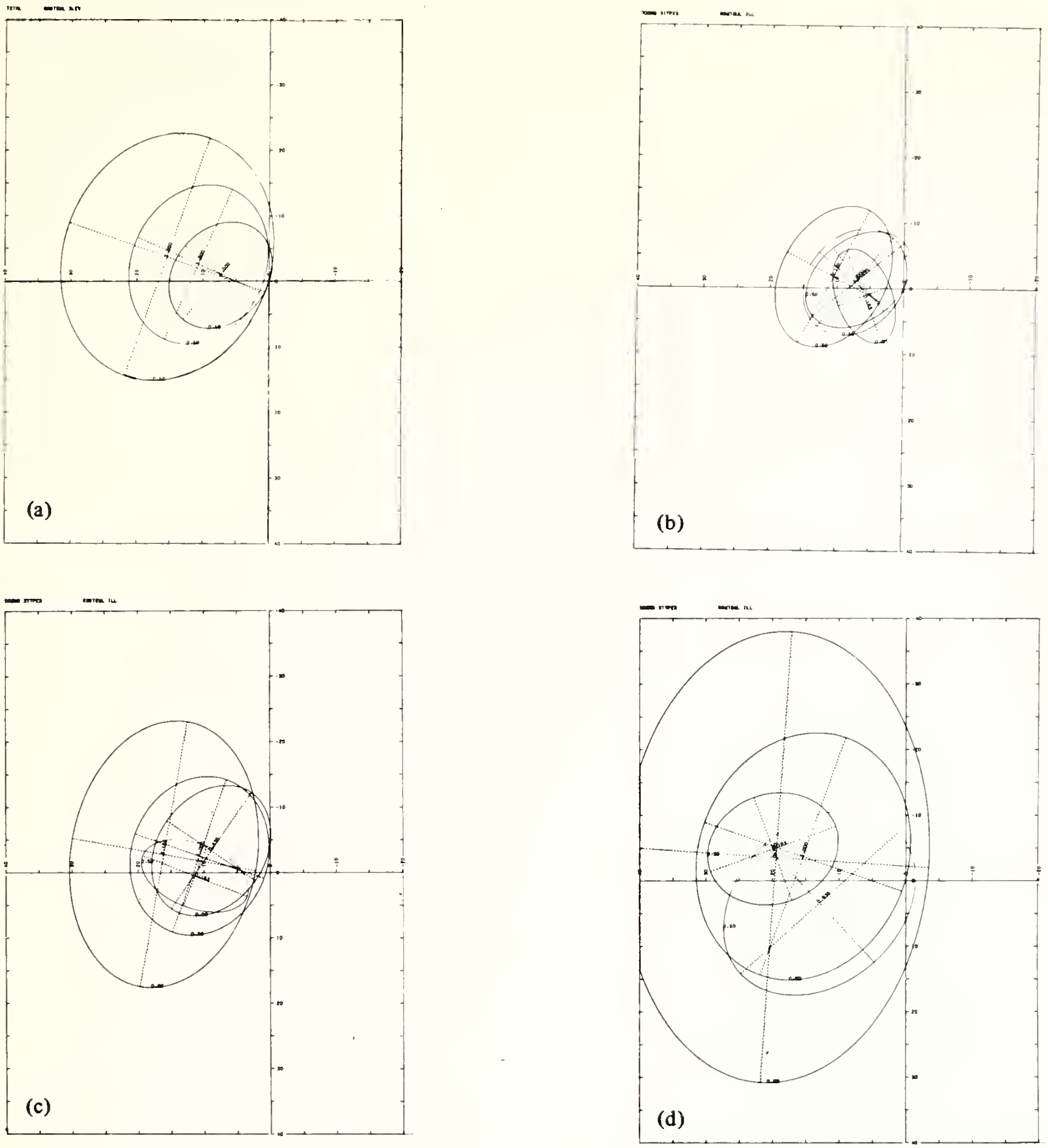


Figure 21 Bivariate distributions of winds in $m \cdot s^{-1}$ at Rantoul, Illinois, October 1950-1955 at the 700-, 500-, and 300-mb levels. Three cluster types (1, 2, and 3) are assumed in the total mixed observed distribution (0.182 + 0.183 + 0.635 = 1.000). (a) Total distribution, (b) 700-mb mixed, (c) 500-mb mixed, (d) 300-mb mixed, (e) Type 1, (f) Type 2, and (g) Type 3.

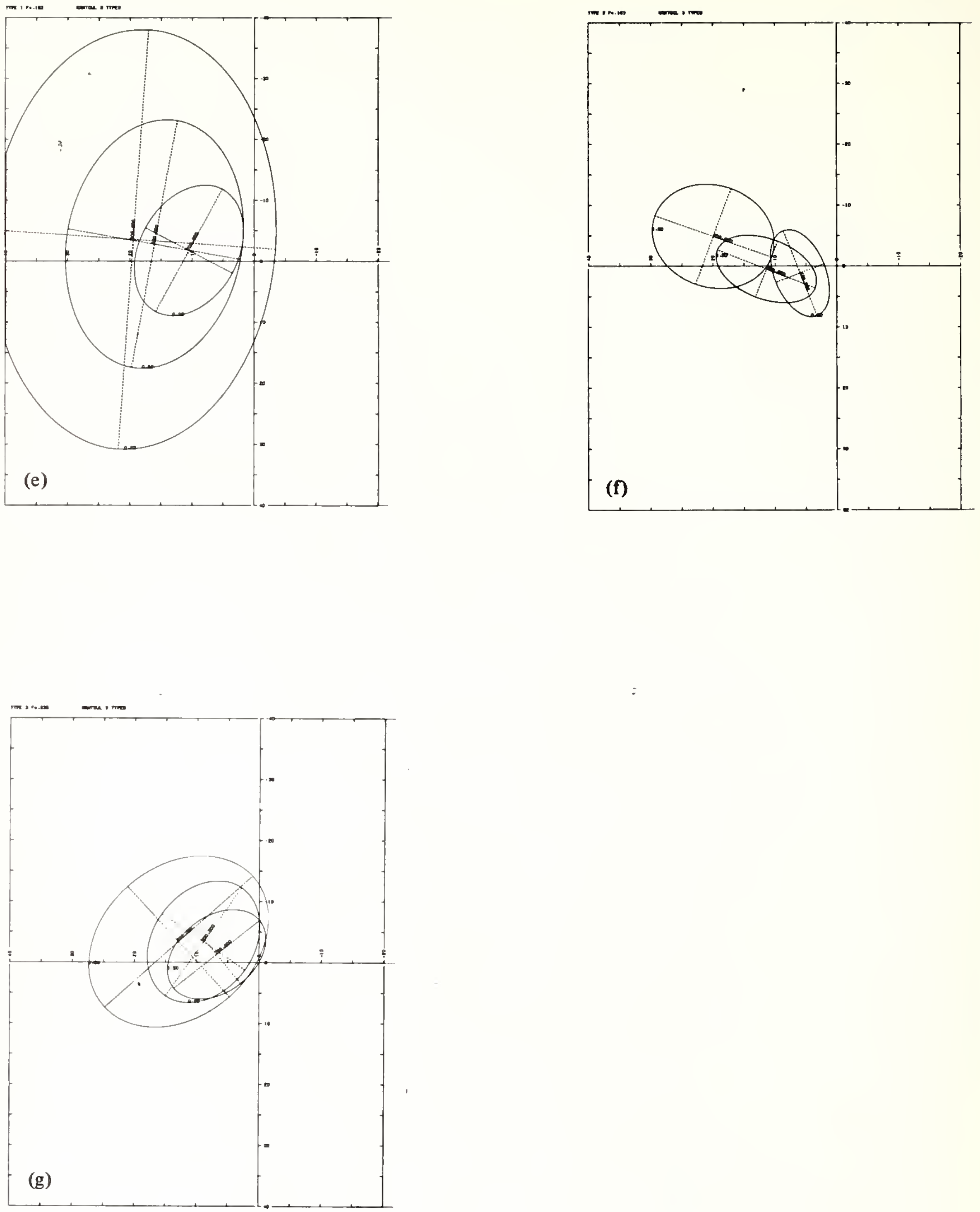


Figure 21 (continued)

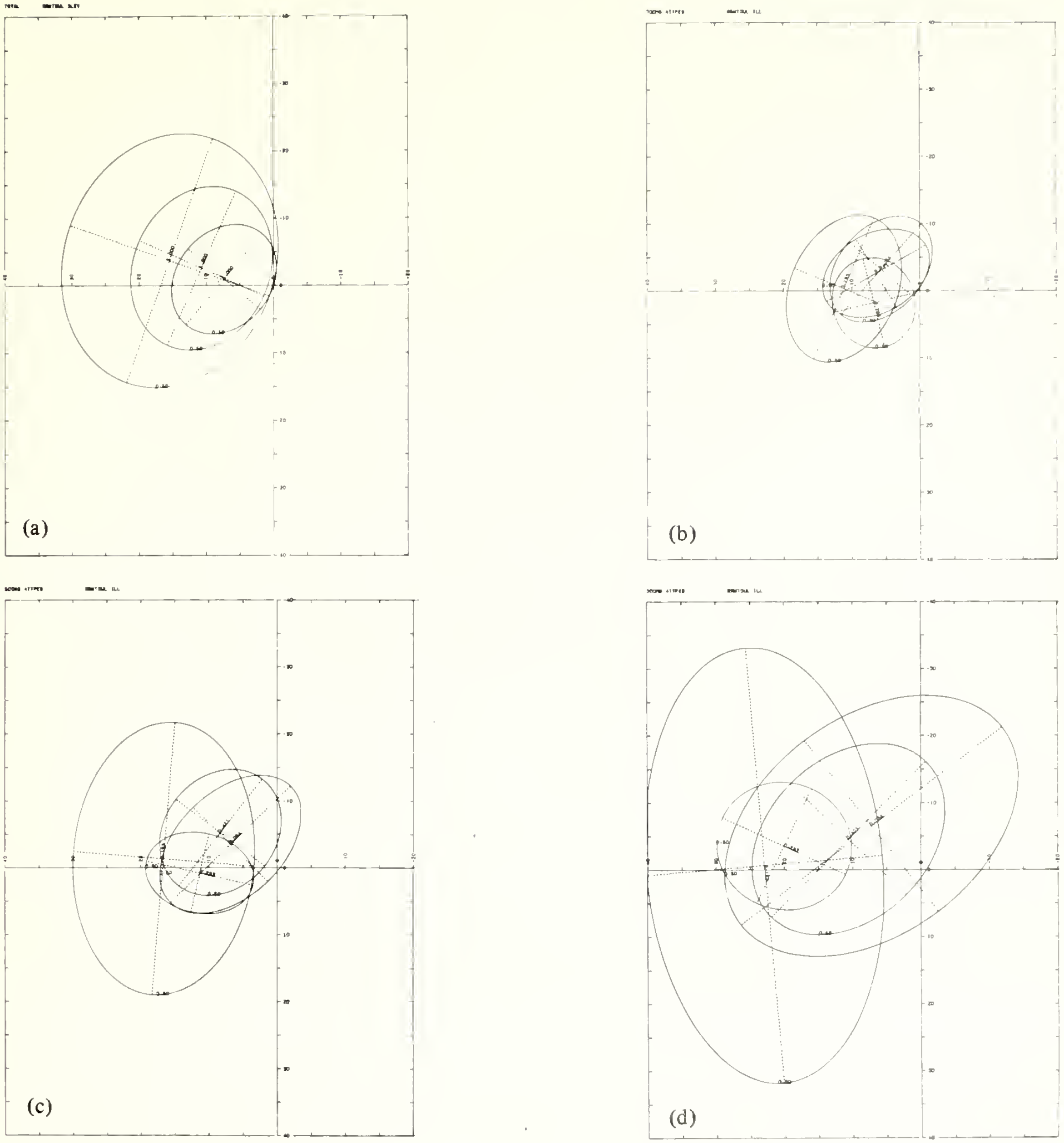


Figure 22 Bivariate distributions of winds in $m \cdot s^{-1}$ at Rantoul, Illinois, October 1950-1955 at the 700-, 500-, and 300-mb levels. Four cluster types (1, 2, 3, and 4) are assumed in the total mixed observed distribution ($0.094 + 0.193 + 0.293 + 0.421 \approx 1.000$). (a) Total distribution, (b) 700-mb mixed, (c) 500-mb mixed, (d) 300-mb mixed, (e) Type 1, (f) Type 2, (g) Type 3, and (h) Type 4.

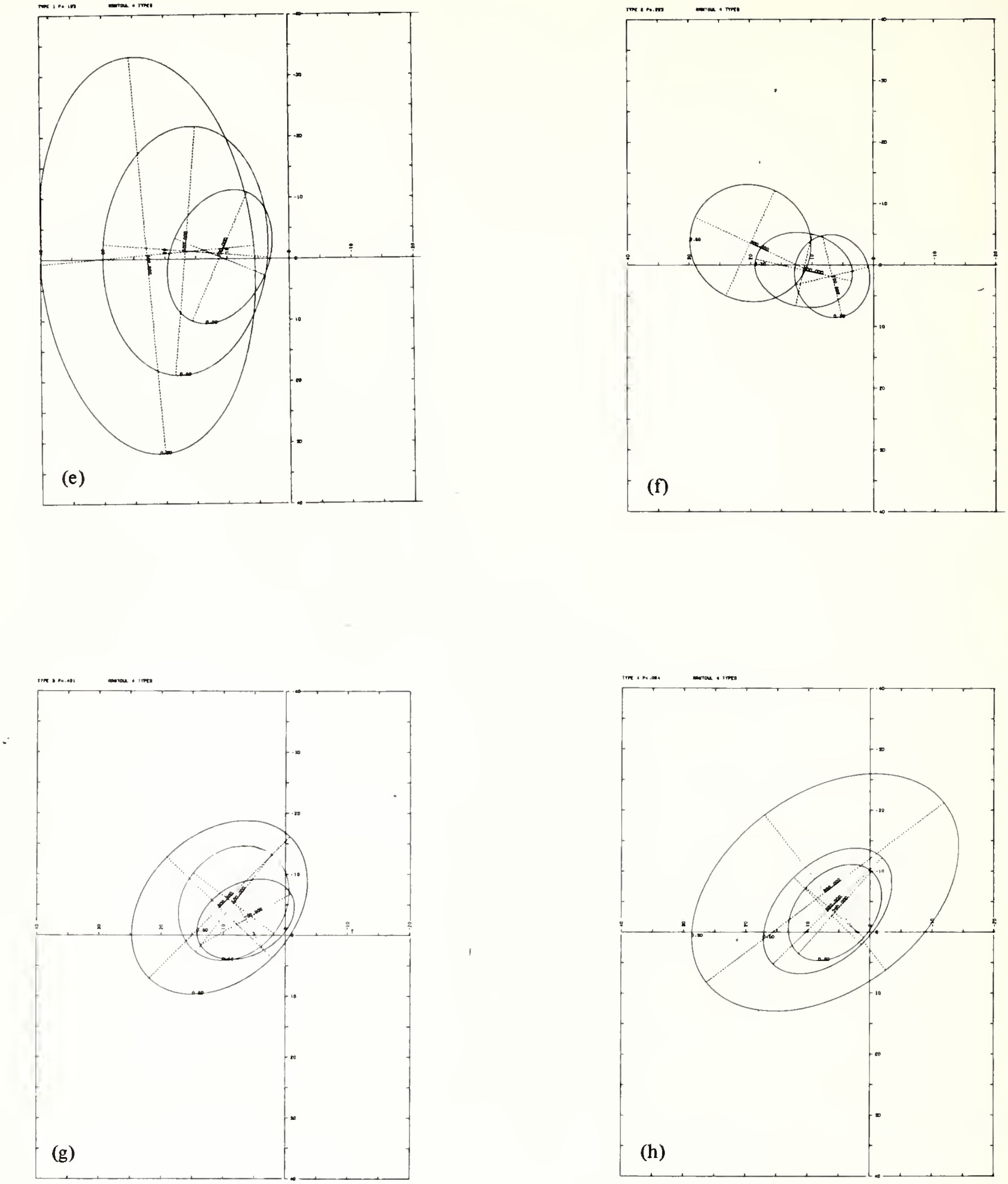


Figure 22 (continued)

8.3.4 Mountain Pass Wind Data Set

8.3.4.1 Input Information.

a. Stampede Pass, Easton, WA, U.S.A. - latitude 47°17' north, longitude 121°20' west, elevation 1206 m.
b. The period of record is the month of December for the years 1966-1970. The local standard time hours are 0700 and 1300.
c. The data are surface winds ($m \cdot s^{-1}$) and temperatures °C.
d. The number of variables is two then three. The first two are the zonal and meridional components of the wind, positive from the west and south, then third is the temperature in °C.
e. The number in the sample is 310, 155 from 0700 and 155 from 1300 l.s.t.
f. The minimum number to be accepted into any cluster is one more than the number of variates.
g. The null hypotheses are made that (k + 1) clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of (k + 1) clusters.
h. The first 40 two-dimensional vector entries are set up as the 40 means of 40 separate and individual clusters. These are 40 points in two or three dimensions.
i. The assumption is made that the covariance matrices are not equal.

8.3.4.2 Tables. Table 19 provides the output data for the two-variable wind component distributions taken in tabular form from the Wolfe (1971b) NORMIX-NORMAP computer routine. Table 20 provides the output data for the three-variable temperature and wind components.

In both tables the mixture proportions, by cluster type, the means, standard deviations, correlation matrices, and the eigenvalue-eigenvector matrices are given. An asterisk indicates the rejection of the null hypothesis that the (k + 1) type is not significantly different from the (k) types.

Table 19 Surface wind statistics for Stampede Pass, Easton, WA, U.S.A. Units are $m \cdot s^{-1}$. The period of record is the month of December, 1966-1970. The sample size is 310 taken 155 from each of the local standard time hours 0700 and 1300. The assumption is that the covariance matrices are not the same. The variables are the zonal and meridional components of the wind positive from the west and south, respectively.

Logarithm of the likelihood; chi-square with 10 d.f.; probability of null hypothesis.

	Logarithm of the likelihood	chi-square	probability
(a)	2 to 1 types 50.327187	(a) 99.52	(a) 0.00000000
(b)	3 to 2 types had no final solution		

Lambda is the proportional amount assigned to each type. An asterisk indicates rejection of the null hypothesis at the 0.01 level. Minus signs indicate the wind components are from the north or east.

Characteristics of the whole sample.

	1	2	1	2
Means	-0.4915	1.1879		
Std. Dev.	5.7123	1.6058		
Corr.	1.0000	0.1667		
	0.1667	1.0000		

Two Types With Proportions

	Type 1 (0.516)		Type 2 (0.484)	
Means	-4.8276	0.9075	4.1357	1.4865
Std. Dev.	3.6052	1.6646	3.4644	1.4954
Corr.	1.0000	-0.4170	1.0000	0.6036
	-0.4170	1.0000	0.6036	1.0000

Eigen-Values	Vectors		Eigen-Values	Vectors	
13.5690	0.9746	0.2239	12.9175	0.9597	-0.2810
2.1665	-0.2239	0.9746	1.3209	0.2810	0.9597

Table 20 Surface temperature and wind statistics for Stampede Pass, Easton, WA, U.S.A. Units are °C and $m \cdot s^{-1}$. The period of record is the month of December 1966-1970. The sample size is 310 taken 155 from each of the local standard time hour 0700 and 1300. The assumption is that the covariance matrices are not the same. Variable 1 is the temperature. The wind components are positive from the west and south, respectively, as variables 2 and 3.

Logarithm of the likelihood; chi-square with 10 d.f.; probability of null hypothesis.

(a)	2 to 1 types 71.240940	(a)	140.41	(a)	0.00000000
(b)	3 to 2 types 50.982624	(b)	100.32	(b)	0.00000000
(c)	4 to 3 types 30.190515	(c)	59.31	(c)	0.00000264

Characteristics of the whole sample.

	1	2	3		1	2	3
Means	-3.7276	-0.4915	1.1880				
Std. Dev.	4.9093	5.7122	1.6057				
Corr.	1.0000	0.3966	0.1963				
	0.3966	1.0000	0.1667				
	0.1963	0.1667	1.0000				

Two Types With Proportions

		Type 1 (0.361)				Type 2 (0.639)	
Means	-5.7450	-6.3469	1.2950		-2.5879	2.8146	1.1276
Std. Dev.	6.0549	2.6497	1.6848		3.6615	4.0974	1.5561
Corr.	1.0000	0.3404	0.2216		1.0000	0.2585	0.2343
	0.3404	1.0000	-0.2746		0.2585	1.0000	0.5703
	0.2616	-0.2746	1.0000		0.2343	0.5703	1.0000
Eigen- Values		Vectors		Eigen- Values		Vectors	
37.7515	0.9883	-0.1390	-0.1172	20.1353	0.5191	0.8539	-0.0367
6.6923	0.1724	0.9172	0.3591	10.9421	0.8287	-0.5133	-0.2229
2.0772	0.0576	-0.3733	0.9259	1.5390	0.2092	-0.0853	0.9741

Table 20 (continued)

Three Types With Proportions

Type 1 (0.484)

	1	2	3
Means	-4.3942	-4.7886	0.9025
Std. Dev.	3.4814	3.3110	1.6817
Corr.	1.0000	0.1548	0.2493
	0.1548	1.0000	-0.4559
	0.2493	-0.4559	1.0000
Eigen-Values		Vectors	
13.4287	0.7910	-0.5840	-0.1826
10.7331	0.6107	0.7353	0.2938
1.7490	-0.0373	-0.3439	0.9383

Type 2 (0.019)

	1	2	3
Means	-26.2038	-10.3733	1.0727
Std. Dev.	1.3766	3.2741	0.8621
Corr.	1.0000	0.0819	-0.6418
	-0.0819	1.0000	0.1371
	-0.6418	-0.1371	1.0000
Eigen-Values		Vectors	
10.7529	-0.0452	0.8963	0.4412
2.2441	0.9981	0.0591	-0.0178
0.3610	0.0420	-0.4395	0.8972

Type 3 (0.497)

	1	2	3
Means	-2.2015	4.0838	1.4708
Std. Dev.	3.7677	3.4445	1.4989
Corr.	1.0000	0.1728	0.1781
	0.1728	1.0000	0.5822
	0.1781	0.5822	1.0000
Eigen-Values		Vectors	
16.0077	0.8031	-0.5952	-0.0283
10.9239	0.5672	0.7782	-0.2697
1.3748	0.1826	0.2006	0.9625

Table 20 (continued)

Four Types With Proportions

	Type 1 (0.243)				Type 2 (0.326)		
	1	2	3		1	2	3
Means	-5.2142	-6.0801	1.4122		-3.3637	-2.5027	0.1948
Std. Dev.	3.8526	2.1293	2.0588		2.8396	3.7895	0.8571
Corr.	1.0000	-0.0634	0.5280		1.0000	0.1975	-0.1654
	-0.0634	1.0000	-0.3357		0.1975	1.0000	-0.7010
	0.5280	-0.3357	1.0000		-0.1654	-0.7010	1.0000
Eigen- Values		Vectors		Eigen- Values		Vectors	
16.3777	0.9390	0.2252	-0.2601	15.3705	0.2837	0.9589	0.0076
5.0933	-0.0827	0.8816	0.4646	7.4239	0.9463	-0.2813	0.1594
2.1440	0.3339	-0.4147	0.8464	0.3639	-0.1550	0.0380	0.9872
	Type 3 (0.019)				Type 4 (0.412)		
Means	-26.2038	-10.3733	1.0727		-2.0772	4.8709	1.8429
Std. Dev.	1.3766	3.2741	0.8621		3.9258	3.1108	1.3780
Corr.	1.0000	-0.0819	-0.6418		1.0000	0.1651	0.1695
	-0.0819	1.0000	0.1371		0.1651	1.0000	0.4586
	-0.6418	0.1371	1.0000		0.1695	0.4586	1.0000
Eigen- Values		Vectors		Eigen- Values		Vectors	
10.7529	-0.0452	0.8963	0.4412	16.2040	0.9409	-0.3372	-0.0315
2.2441	0.9981	0.0591	-0.0178	9.3669	0.3222	0.9198	-0.2241
0.3610	0.0420	-0.4395	0.8972	1.4169	0.1046	0.2007	0.9740

8.3.4.3 Figures and Discussion. Figure 23a shows the single cluster distribution versus the two-cluster breakout, each cluster assumed to be bivariate normal. In essence, these assumptions are the unimodal versus the bimodal bivariate distributions. Table 19 provides the statistics for this figure. Clearly, it is seen that the total distribution is not well represented by the single unimodal elliptical bivariate normal distribution with the mean at (-0.4915, 1.1879) $m \cdot s^{-1}$ with east-west and north-south component standard deviations of 5.7123 and 1.6058 $m \cdot s^{-1}$, respectively. The ratio is almost four to one.

The total distribution breaks out into two separate distributions, each assumed to be unimodal. The attempt to determine whether the distribution might actually be trimodal rather than bimodal met with no success. Therefore, it is assumed that the bimodal bivariate distribution is a better representation than a unimodal or trimodal bivariate representation. The two modes are east-southeast and west-southwest.

Let us now discuss the location and terrain features of Stampede Pass. Stampede Pass is located in mountainous terrain on the Main Cascade Divide at latitude 47°17' north and longitude 121°20' west. The elevation of the ground at the station is about 1206 m. The wind instruments are approximately 11 m higher.

East of the station the ground drops abruptly into the Yakima River Valley, about 600 m down and a little over three km distant. This land and valley fall towards the southeast. The lowest part of Stampede Pass is 5/8 km north and 30 m lower. There is a ridge 1-1/4 km south of the station and about 200 m higher. West of the station the land drops rapidly about 600 m over a distance of 6.4 km to the Green River Valley. This land and valley fall toward the southwest. To the north, from west northwest through east, there are ridges and peaks which rise to 1.2- and 2.7-km.

General winds from the west will be channeled from the west southwest up and around the ridge nose then turning to the east southeast and thence southeast. General winds from the east will traverse this same channel but in the opposite sense. It would seem then that a wind distribution through the pass would have to be elliptical. In addition, it would seem that traveling weather systems would create one distribution from the west southwest and one from the east southeast. This agrees with the tabular values (table 19) and the illustration (figure 23a).

Though not presented in either the table or the figure, it should be mentioned that the computer routine did attempt to converge to a solution for the trimodal assumption. The last estimates did indicate a tendency for the east southeast mode to break down into two east southeast modes located on the major axis of the mode shown but centered, one further to the east southeast and one to the west northwest.

Table 20 and figures 23b, c, and d present the situation when the temperature arguments are added to those of the wind. The singularity problem noted above is resolved and the computations indicate definitely that four cluster types are present. The null hypothesis would have been rejected at the 0.00000264 level. The selected decision level was 0.01. Therefore, there is a good probability that five or even six clusters would have been isolated if the computer had been allowed to continue. The option selected was to examine the structure through only four groups.

Figure 23b provides the two-cluster breakout with an illustration of the 0.25 error ellipses of the wind. The mixture proportion as well as the mean temperature (°C) associated with each group is shown. As progress is made through the next two subfigures, note that the western group changes only slightly moving farther away as the breakdown continues. Please note the breakout of the eastern group into two groups with the easternmost group being extremely cold (≈-26°C). Note that it then remains relatively fixed while the central group of (b) breaks down into two clusters.

Note again that the easternmost group which comprises only two percent of the total is extremely cold. This leads to two conjectures:

(1) This group is an outlier and the data are bad.

(2) This group is an outlier but is a valid cluster. The output of the data permits identification of the individual datum. The two-percent cluster occurred in the same period of time and actually composed a string of data. The records were checked. The data are correct. Figure 24 is a copy of a National Meteorological Center surface analysis chart for 1200 Greenwich time on December 30. A star marks the approximate location of Stampede Pass. Note the incursion of the cold air mass from Canada. With its increasing cold and speed as seen on prior maps, it is worthwhile to read the comments by Phillips (1969) published in the Climatological Data publication of the National Oceanic and Atmospheric Administration. The temperatures mentioned are in degrees Fahrenheit.

"Washington - December 1968

"Special Weather Summary

> "Until near the end of the month, weather systems from over the Pacific moved across the State at frequent intervals. In western Washington, this resulted in measurable precipitation on 23 to 28 days and on 12 to 18 days in eastern Washington.
>
> "West of the Cascades, rather heavy precipitation was recorded on several days, falling as rain in lowlands during the first half of the month, and as snow and rain the latter half. In the mountains, most precipitation fell as snow, with near record depths for December on the ground at the end of the month.
>
> "East of the Cascades, snow began accumulating on the ground the first of the month. After the middle of the month, most agricultural areas in southern counties were covered with 1 to 3 inches of snow. Temperatures were near or above normal for the first half of the month and slightly below normal from the 15th to the 25th.

> "An outbreak of very cold arctic air accompanied by strong northerly and northeasterly winds began moving into northern valleys of eastern Washington, and other localities near the Canadian State on the 26th, spreading over most of the State on the 27th. Temperatures continued to fall for 3 days, with the lowest occurring on the 30th. In many respects, this was the most severe outbreak of cold air since the winter of 1949-50. Minimum temperatures dropped below previous records at several stations in eastern Washington and in the Cascades. The -48° recorded at Mazama and Winthrop 1 SW is a new record for the State and -43° at Chesaw 4 NNW is also below previous record of -42° at Deer Park on January 20, 1937. Other stations where minimums dropped below previously recorded low temperatures were: Anatone -32°, Chelan -18°, Colfax 1 NW - 33°, Colville Airport -33°, Dayton 1 SW -25°, Holden Village -32°, Lacrosse 3 ESE -34°, Leavenworth 3 S -36°, Methow -37°, Pomeroy -27°, Pullman 2 NW -32°, Republic -38°, Rosalia -29°, Snoqualmie Pass -19°, Stampede Pass -21°, Stevens Pass -25°, Stehekin 3 NW -21°, and Waterville -33°.
>
> "On the 30th, a warmer moist airmass from over the Pacific began moving inland over the colder air near the surface. Snow began falling during the day, becoming heavy at night and continuing through the 31st. West of the Cascades, snow depths in the lowlands ranged from 8 to 15 inches and 24 to 36 inches or more in foothills. In numerous localities, highway traffic was at a near standstill on the 31st and many offices and businesses remained closed.
>
> "Preliminary reports from fruit producing areas indicate the low temperatures caused extensive damage to stone fruits and perhaps some damage to other fruit trees. Most of the winter wheat section was covered with 1 to 3 inches of snow, thus very little freeze damage is expected."

Note the fourth paragraph. Ludlum (1969) disucsses this particular feature in Weatherwise on pages 36-37 of the February issue. Dickey and Wing (1963) also discuss the problem of arctic air flowing into the Pacific Northwest.

The above example and discussion point out the usefulness of a program such as this for editing and quality control of multivariate data; i.e., data groups other than one element at a time. Here, an outlier group was isolated but it was a valid group. The authors did not realize that this particular outlier group was embedded in the data set used. Simply, a location and a period were selected where it was thought that the program would successfully and pointedly demonstrate its capability.

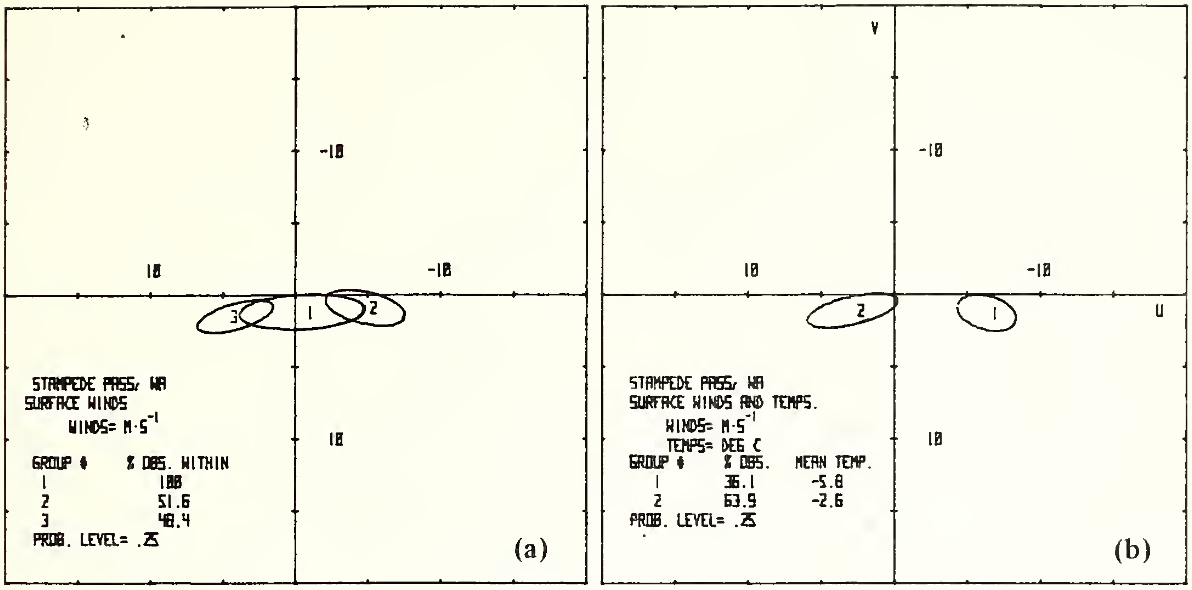


Figure 23 Stampede Pass, Easton, Washington, U.S.A.; winds ($m \cdot s^{-1}$) and temperatures (°C), December 1966-1970, showing breakdown of winds only into groups 2 and 3 from group 1 (a) and breakdown of wind-temperature combination into 2, 3, and 4 groups (b, c, and d).

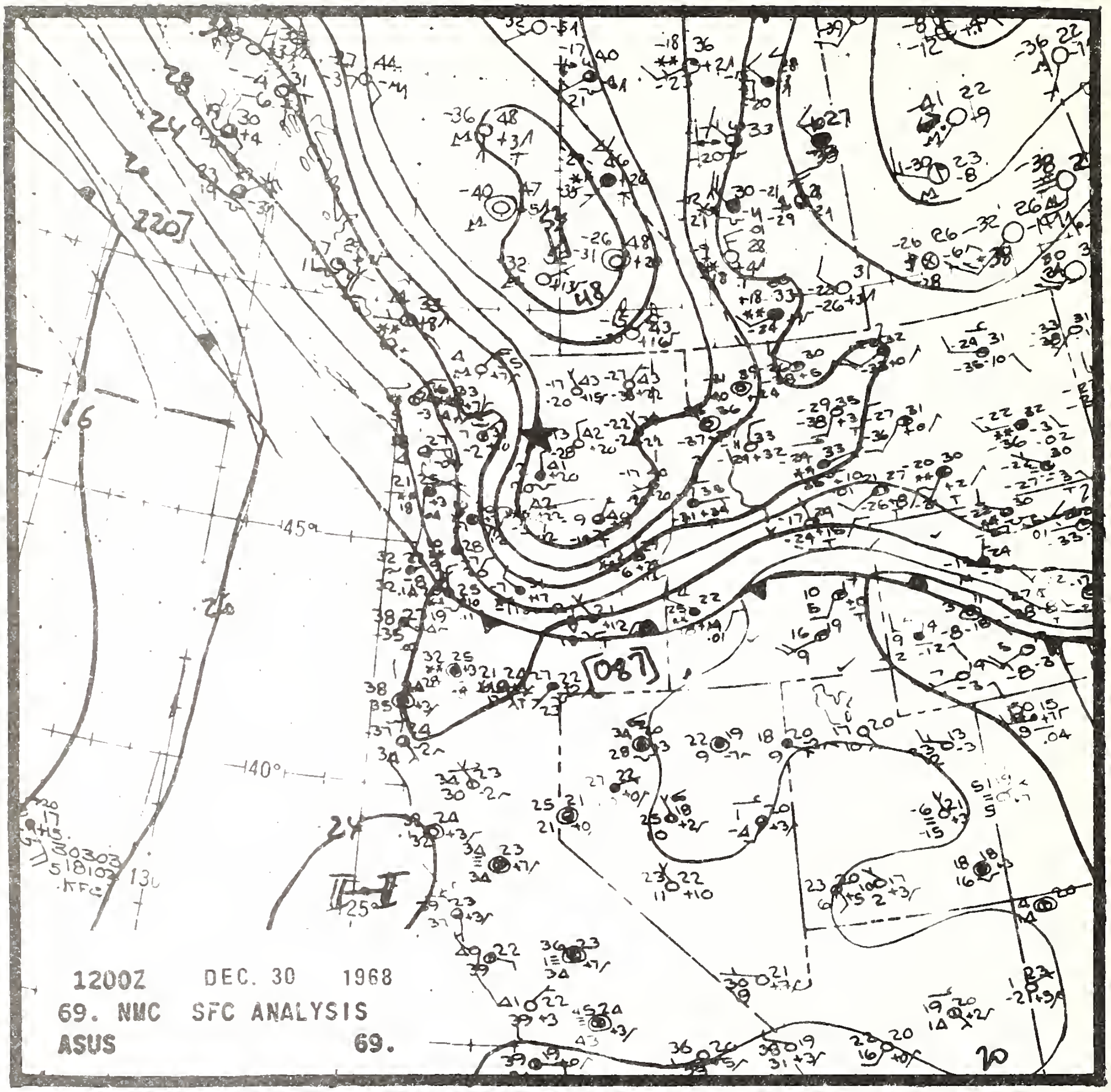


Figure 24 Selected area of North American chart, 1200Z, Monday, December 3, 1968, NMC analysis. The star represents the approximate position of Stampede Pass, Easton, Washington.

8.3.5 Marine Surface Data Set

8.3.5.1 Input Information.

a. OSV "C," 52°45' north latitude, 35°30' north longitude
b. The period of record is the month of February for the years 1964 through 1972.
c. The data are the 1200 G.C.T., pressures, temperatures, dew points, and surface winds.
d. The number of variables is five. The wind is broken into the zonal and meridional components. The units are mb, °C, and $m \cdot s^{-1}$.
e. The number in the sample is 251.
f. The minimum number to be accepted into a sample is six.
g. The null hypotheses are made that (k + 1) clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of (k + 1) clusters.
h. The first 40 five-dimensional vector entries are set up as the 40 means of 40 separate and individual clusters. These are 40 points in five dimensions.
i. Non-equality of covariances is assumed.

8.3.5.2 Tables - Output Information. The program computes the necessary statistics for two clusters versus one cluster, three clusters versus two clusters, and four clusters versus three clusters. These are taken from the tabular outform of the Wolfe (1971b) NORMIX computer routine.

The output statistics are now shown for the above in tables 21, 22, and 23 even though the null hypothesis is rejected for the last case.

The first variable is the surface atmospheric pressure in mb, the second variable is the dry-bulb air temperature in °C, the third variable is the dew point temperature in °C, while the fourth and fifth variables are the zonal and meridional components of the wind in $m \cdot s^{-1}$.

Table 21 1200 GCT Marine observations from Ocean Station C. for the month of February 1964 through 1972. Surface pressure (mb), temperature (°C), dew point (°C), and zonal and meridional components ($m \cdot s^{-1}$). Sample size is 251. The assumption is that the covariance matrices are not the same. Number of variables is 5. Number of types is 2.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a) 2 to 1 types 56.0886786 (a) 109.27 (a) 0.00000002

Characteristics of the whole sample.

	1	2	3	4	5
Means	1005.9072	5.4825	1.8749	2.0414	1.0315
Std. Dev.	18.1051	2.3600	3.9201	10.3810	8.2983
Corr.	1.0000	0.0494	-0.0379	-0.1026	-0.0258
	0.0494	1.0000	0.8656	-0.6136	0.3518
	-0.0379	0.8656	1.0000	-0.5179	0.3649
	-0.1026	-0.6136	-0.5179	1.0000	-0.1667
	-0.0258	0.3518	0.3649	-0.1667	1.0000

<u>Characteristics of Type 1</u> <u>Proportion 0.391</u>

Means	1003.7997	3.5541	-1.8868	9.3676	0.6684
Std. Dev.	17.0627	1.8933	2.7773	9.4358	7.1053
Corr.	1.0000	-0.0872	-0.3678	-0.2149	-0.3972
	-0.0872	1.0000	0.6632	-0.4550	0.2966
	-0.3678	0.6632	1.0000	-0.0368	0.4421
	-0.2149	-0.4550	-0.0368	1.0000	0.0659
	-0.3972	0.2966	0.4421	0.0659	1.0000

Eigen-Values			Eigen-Vectors		
306.6460	0.9698	0.0935	0.2220	0.0364	-0.0103
87.8493	-0.0081	-0.1100	0.0435	0.4788	0.8699
39.2844	-0.0615	-0.0649	0.1325	0.8582	-0.4878
7.1223	-0.1499	0.9526	0.2469	0.0639	0.0716
1.0497	-0.1819	-0.2600	0.9329	-0.1698	0.0122

Table 21 (continued)

Characteristics of Type 2 Proportion 0.609

	1	2	3	4	5
Means	1007.2661	6.7228	4.2951	-2.6718	1.2650
Std. Dev.	18.6248	1.7062	2.3078	7.9436	8.9748
Corr.	1.0000	0.0316	-0.0360	0.0461	0.1425
	0.0316	1.0000	0.8233	-0.3347	0.5239
	-0.0360	0.8233	1.0000	-0.2660	0.6019
	0.0461	-0.3347	-0.2660	1.0000	-0.2521
	0.1425	0.5239	0.6019	-0.2521	1.0000
Eigen-Values			Eigen-Vectors		
349.0990	0.9960	-0.0624	-0.0619	0.0136	-0.0064
93.7420	0.0047	0.1040	0.0119	0.5748	0.8115
50.8501	-0.0015	0.1504	0.0583	0.7957	-0.5838
4.4547	0.0182	-0.5355	0.8425	0.0531	0.0186
0.6223	0.0872	0.8221	0.5319	-0.1827	0.0137

Table 22 Marine observations from Ocean Station C. February's 12Z 1964 through 1972. Surface pressure (mb), temperature (°C), dew point (°C), and u and v components ($m \cdot s^{-1}$). Sample size is 251. Number of variables is 5. Number of types is 3.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a) 3 to 2 types 34.213746 (a) 66.52 (a) 0.00529292

Characteristics of the whole sample.

	1	2	3	4	5
Means	1005.9072	5.4825	1.8749	2.0414	1.0315
Std. Dev.	18.1051	2.3600	3.9201	10.3810	8.2983
Corr.	1.0000	0.0494	-0.0379	-0.1026	-0.0258
	0.0494	1.0000	0.8656	-0.6136	0.3518
	-0.0379	0.8656	1.0000	-0.5179	0.3649
	-0.1026	-0.6136	0.5179	1.0000	0.1667
	-0.0258	0.3518	0.3649	-0.1667	1.0000

<u>Characteristics of Type 1</u> <u>Proportion 0.404</u>

Means	1004.3747	3.5415	-1.7947	9.2681	0.5814
Std. Dev.	16.7221	1.8331	2.8160	9.2252	6.6616
Corr.	1.0000	-0.0051	-0.2979	-0.2291	-0.3637
	-0.0051	1.0000	0.6631	-0.4344	0.2463
	-0.2979	0.6631	1.0000	-0.0466	0.4377
	-0.2291	-0.4344	-0.0466	1.0000	-0.0184
	-0.3637	0.2463	0.4377	-0.0184	1.0000

Eigen-Values			Eigen-Vectors		
292.9837	0.9721	0.1281	0.1947	0.0230	-0.0123
81.4946	0.0014	-0.1013	0.0607	0.4627	0.8786
37.4986	-0.0517	-0.0665	0.1702	0.8606	-0.4726
7.3821	-0.1641	0.9654	0.1807	0.0647	0.0650
1.0418	-0.1594	-0.1921	0.9470	-0.2014	0.0187

Table 22 (continued)

Characteristics of Type 2	Proportion 0.293				
	1	2	3	4	5
Means	1013.0465	7.6058	5.5791	-6.0191	8.3698
Std. Dev.	16.5701	1.7053	2.0416	6.4452	5.5672
Corr.	1.0000	-0.0380	0.0040	0.5042	-0.3574
	-0.0380	1.0000	0.7820	-0.0466	0.4095
	0.0040	0.7820	1.0000	0.0123	0.2695
	0.5042	-0.0466	0.0123	1.0000	0.1064
	-0.3574	0.4095	0.2695	0.1064	1.0000
Eigen-Values			Eigen-Vectors		
290.2247	0.9706	-0.0742	0.2232	-0.0503	-0.0099
39.5450	-0.0056	0.0677	0.2053	0.5424	0.8118
18.6790	-0.0008	0.0662	0.1783	0.7922	-0.5799
5.0617	0.2083	0.7341	-0.6370	0.1050	0.0311
0.6686	-0.1205	0.6683	0.6859	-0.2543	-0.0602

Characteristics of Type 3	Proportion 0.303				
Means	1001.0589	6.0145	3.1819	0.2071	-5.4395
Std. Dev.	19.1798	1.2040	1.7959	8.1453	6.6050
Corr.	1.0000	-0.3396	-0.4957	-0.0446	-0.0783
	-0.3396	1.0000	0.6806	-0.3227	0.2792
	-0.4957	0.6806	1.0000	-0.1058	0.4472
	-0.0446	-0.3227	-0.1058	1.0000	-0.0009
	-0.0783	0.2792	0.4472	-0.0009	1.0000
Eigen-Values			Eigen-Vectors		
369.3045	0.9979	0.0191	0.0381	0.0481	-0.0049
66.4312	-0.0215	-0.0524	0.0466	0.5165	0.8532
43.9556	-0.0470	-0.0328	0.1137	0.8443	-0.5205
2.2771	-0.0225	0.9976	0.0286	0.0510	0.0283
0.5416	-0.0313	-0.0232	0.9913	-0.1244	-0.0190

Table 23 Marine observations from Ocean Station C. February's 12Z 1964 through 1972. Surface pressure (mb), temperature (°C), dew point (°C), and u and v components ($m \cdot s^{-1}$). Sample size is 251. The assumption is that the covariance matrices are not the same. Number of variables is 5. Number of types is 4.

Logarithm of the likelihood; chi-square with 4 d.f.; probability of null hypothesis.

(a) 4 to 3 types 26.048976 (a) 50.54 (a) 0.12269608*

Characteristics of the whole sample.

	1	2	3	4	5
Means	1005.9072	5.4825	1.8749	2.0414	1.0315
Std. Dev.	18.1051	2.3600	3.9201	10.3810	8.2983
Corr.	1.0000	0.0494	-0.0379	-0.1026	-0.0250
	0.0494	1.0000	0.8656	-0.6136	0.3518
	-0.0379	0.8656	1.0000	-0.5179	0.3649
	-0.1026	-0.6136	-0.5179	1.0000	-0.1667
	-0.0258	0.3518	0.3649	-0.1667	1.0000

Characteristics of Type 1 Proportion 0.415

	1	2	3	4	5
Means	1004.6928	3.6113	-1.6696	9.1991	0.6093
Std. Dev.	16.4198	1.8662	2.8892	9.1816	6.5831
Corr.	1.0000	0.0279	-0.2572	-0.2530	-0.3379
	0.0279	1.0000	0.6812	-0.4327	0.2445
	-0.2572	0.6812	1.0000	-0.0553	0.4315
	-0.2530	-0.4327	-0.0553	1.0000	-0.0167
	-0.3379	0.2445	0.4315	-0.0167	1.0000

Eigen-Values			Eigen-Vectors		
283.0719	0.9703	0.1492	0.1893	0.0180	-0.0123
79.4998	0.0057	-0.1035	0.0660	0.4599	0.8794
37.6133	-0.0465	-0.0741	0.1842	0.8582	-0.4711
7.8690	-0.1853	0.9594	0.1914	0.0672	0.0646
1.0262	-0.1485	-0.2029	0.9430	-0.2171	0.0198

Table 23 (continued)

Characteristics of Type 2 Proportion 0.103

	1	2	3	4	5
Means	993.1830	6.9100	4.7324	-11.1843	9.5660
Std. Dev.	15.4344	2.2137	2.1577	5.5325	5.7345
Corr.	1.0000	-0.3363	-0.1864	-0.2038	-0.3080
	-0.3363	1.0000	0.8070	-0.3760	0.3751
	-0.1864	0.8070	1.0000	-0.3469	0.1786
	-0.2038	-0.3760	-0.3469	1.0000	0.3500
	-0.3080	0.3751	0.1786	0.3500	1.0000
Eigen-Values			Eigen-Vectors		
244.1472	0.9858	0.1475	0.0645	0.0430	0.0182
38.1681	-0.0488	-0.0649	0.3284	0.5138	0.7884
23.8818	-0.0261	-0.0802	0.2516	0.7522	-0.6032
4.3911	-0.0857	0.7232	-0.5913	0.3368	0.0810
0.6828	-0.1331	0.6668	0.6893	-0.2344	-0.0877

Characteristics of Type 3 Proportion 0.205

	1	2	3	4	5
Means	1021.2997	7.8968	5.9650	-3.1858	7.5392
Std. Dev.	10.7461	1.1947	1.8153	4.9641	5.6589
Corr.	1.0000	-0.2843	-0.3873	0.1972	-0.3949
	-0.2843	1.0000	0.7539	-0.2296	0.6340
	-0.3873	0.7539	1.0000	-0.1704	0.4406
	0.1972	-0.2296	-0.1704	1.0000	0.1491
	-0.3949	0.6340	0.4406	0.1491	1.0000
Eigen-Values			Eigen-Vectors		
123.5411	0.9595	0.1429	-0.2409	0.0214	-0.0199
31.6815	-0.0395	0.0634	-0.1636	0.3614	0.9149
18.7018	-0.0715	0.0511	-0.1420	0.9056	-0.3898
2.6280	0.0930	0.6378	0.7547	0.1115	0.0507
0.3140	-0.2529	0.7524	-0.5706	-0.1906	-0.0898

Table 23 (continued)

Characteristics of Type 4 Proportion 0.278

	1	2	3	4	5
Means	1001.1050	5.9713	3.0994	0.0923	-6.2485
Std. Dev.	18.5631	1.2205	1.7767	8.2187	5.9934
Corr.	1.0000	-0.3701	-0.5146	-0.0858	-0.0108
	-0.3701	1.0000	0.6727	-0.3130	0.2402
	-0.5146	0.6727	1.0000	-0.0737	0.4216
	-0.0858	-0.3130	-0.0737	1.0000	-0.0087
	-0.0108	0.2402	0.4216	-0.0087	1.0000
Eigen- Values			Eigen- Vectors		
346.2454	0.9974	0.0434	0.0134	0.0557	-0.0046
67.1930	-0.0241	-0.0545	0.0491	0.5311	0.8438
36.5567	-0.0494	-0.0311	0.1275	0.8330	-0.5352
2.1563	-0.0464	0.9968	0.0287	0.0495	0.0302
0.5521	-0.0046	-0.0228	0.9901	-0.1358	0.0262

8.3.5.3 Figures and Discussion. Figure 25 is an attempt to illustrate a part of the output of table 21. The basic two-dimensional representation shows the decomposition of the assumed unimodal distribution into two modes or groups. The probability ellipses represent the area of the central 25 percent of the wind vector origins if the assumption which they represent is valid. As shown in table 21, the probability of the null hypothesis for two groups versus one group not being rejected is very low, i.e., 0.00000002. Therefore, the more valid assumption is that two groups better represent the data set than does one group. The single group centered at (2.0414, 1.0315) is shown for reference.

At each point, the following concurrent values are printed. the mixture proportion, the mean pressure, the mean temperature, and the mean dew point. The two groups appear to have not too different mean pressures but considerably different temperature and dew points.

Figure 26 prepared from table 22 similarly portrays the decomposition of the total group into three groups. It is quite apparent that group 1 is not much different from group 1 of figure 25. It is just as apparent that group 2 of figure 25 really breaks down into groups 2 and 3 as shown in figure 26. The pressures in groups 2 and 3 are quite different though the temperatures and dew points are not too different. The appearance here implies, as does the probability of non-rejection of the null hypothesis being small, that the trimodal representation is better than the bimodal which is in turn better than the unimodal representation.

Figure 27 prepared from table 23 depicts the further decomposition of the data set into four groups. Again, group 1 retains essentially the same mixture proportions and characteristics. The northerly group (group 4) remains almost the same as in the last decomposition. It is group 2 of figure 26 that now breaks down into groups 2 and 3 of figure 27. The pressure differences are remarkable, providing the maximum and minimum pressures for the entire four-cluster configuration. Group 3 has the highest temperatures and dew points of the entire ensemble. The 10-percent mixture proportion of group 2 leaves the impression that this is a real group though the null hypothesis is not rejected. This non-rejection probability of 0.12 implies that at our decision level of 0.01 probability, the trimodal representation is a better representation than the four-cluster configuration. However, it is interesting to conjecture that the differences shown are the difference between storm and non-storm situations with winds from the southeast quadrant.

No other representations are made though any pairing or triplets could be graphed in two and three space, respectively, and could be labeled with the fourth and fifth variable mean values.

The authors consider this to be a good representation of the clustering techniques of the Wolfe NORMIX (1971b) computer routine. It was hoped that this routine would isolate a cluster of six or more data which might be considered to be an outlier cluster which could be examined. There were 251 data. Ten percent is essentially 25 or 26 data in the clusters. The instructions provided to the computer were to collect no less than six data in a cluster. This is a default option which sets the minimum number in any cluster to be one more than the number of variables. In this case, five plus one is six.

Therefore, (a) single, doublet or triplet outlier(s) would not be isolated. In this case, if the mixture proportion ran as low as 2.5 percent, the group would be considered as an outlier group and would be examined to see whether the data were bad because of bad sensors, bad recording, bad entry of data into the archives, or simply a group of valid observations out of a very rare weather situation.

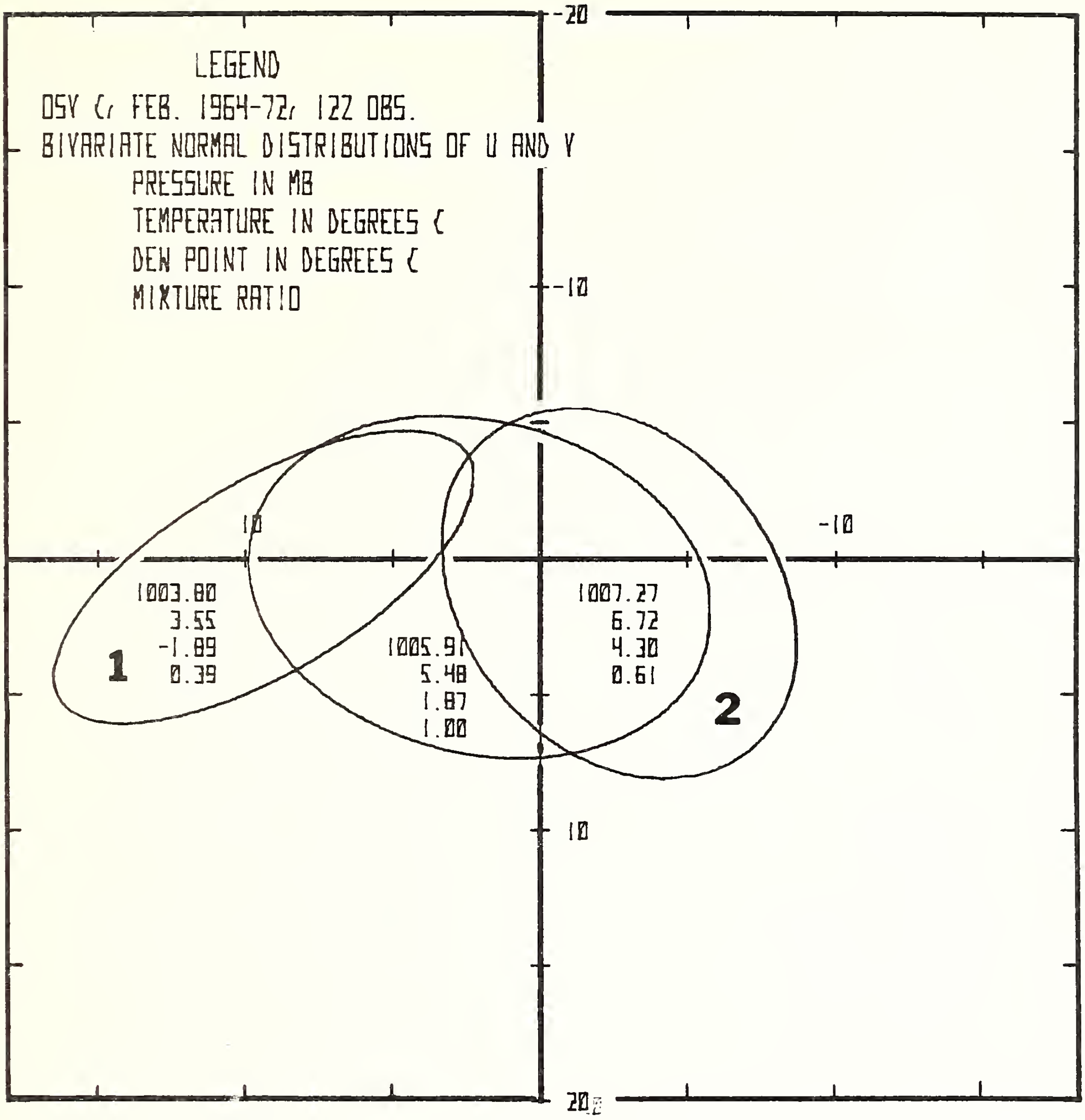


Figure 25 OSV "C" surface distribution of pressure (mb), temperature (°C), dew point (°C), and wind components ($m \cdot s^{-1}$), February, 1200 G.C.T., 1964 through 1972; n = 251. Covariances are assumed to be unequal. The 0.25 probability ellipses are shown for the wind distribution. The total distribution and the breakout into two clusters are shown. The mixture proportion and the averages, the pressure, the temperature, and the dew point data within each cluster are shown.

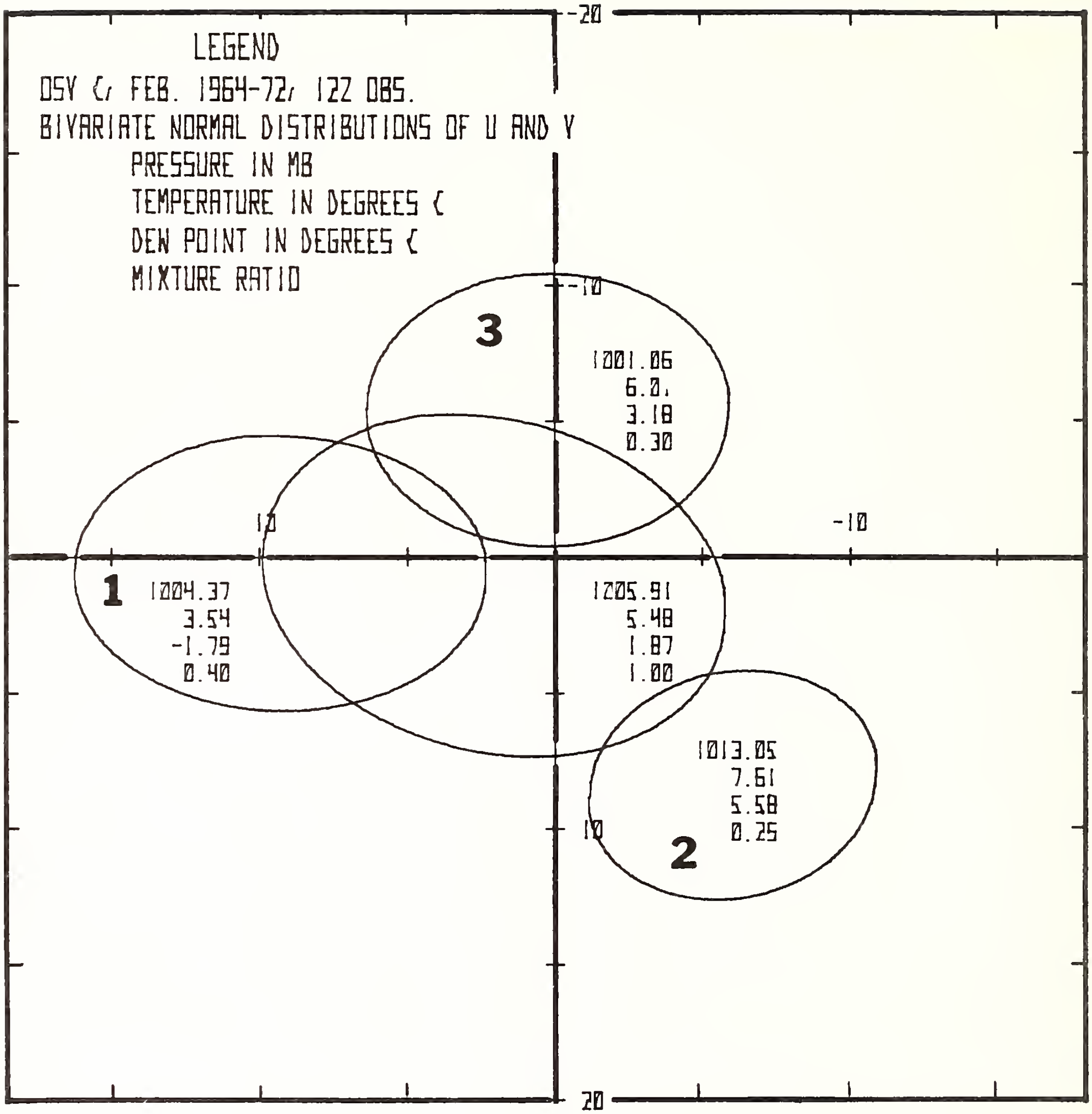


Figure 26 OSV "C" surface distribution of pressure (mb), temperature (°C), dew point (°C), and wind components ($m \cdot s^{-1}$), February, 1200 G.C.T., 1964 through 1972; n = 251. Covariances are assumed to be unequal. The 0.25 probability ellipses are shown for the wind distribution. The total distribution and the breakout into three clusters are shown. The mixture proportion and the averages, the pressure, the temperature, and the dew point data within each cluster are shown.

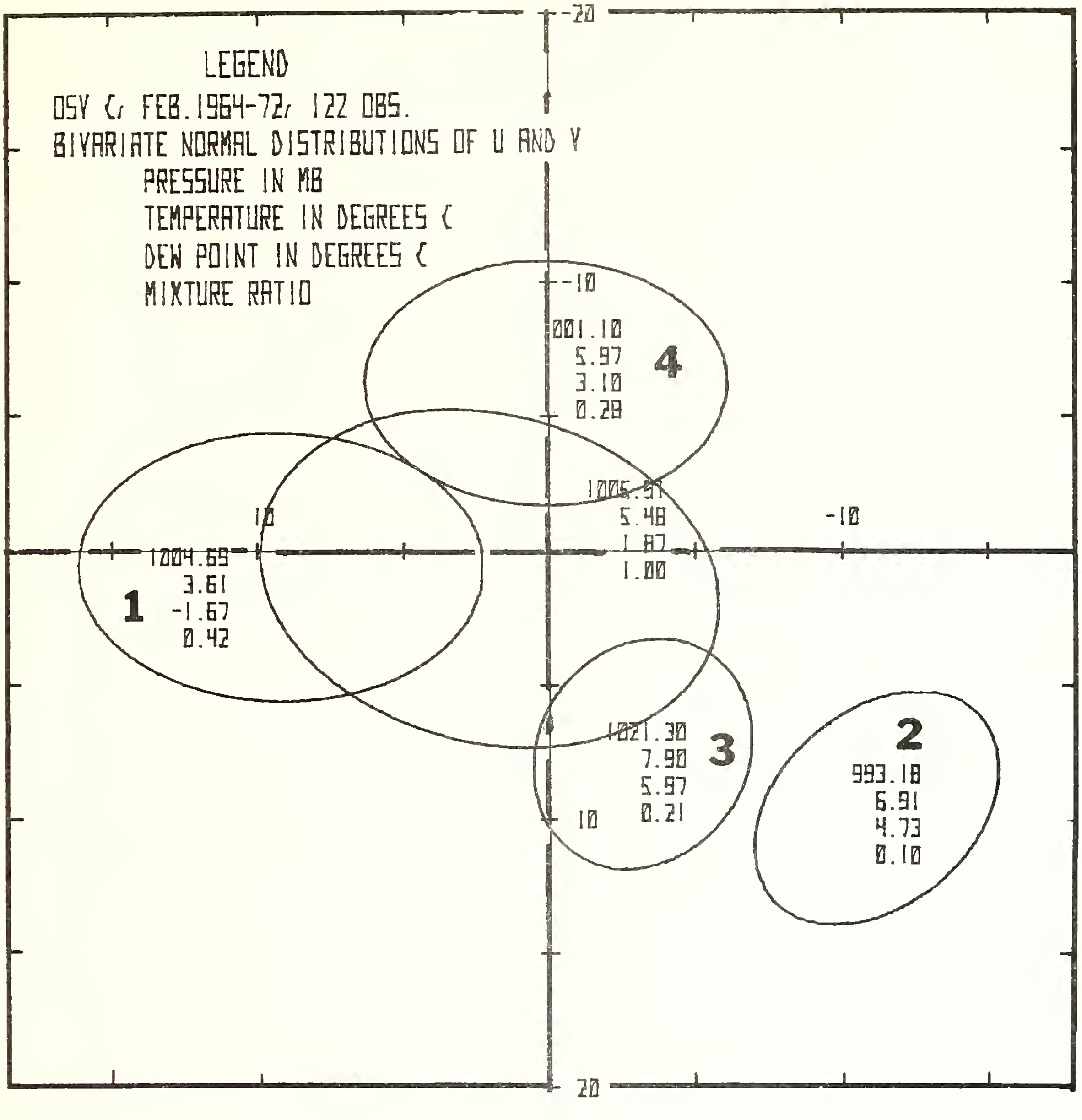


Figure 27 OSV "C" surface distribution of pressure (mb), temperature (°C), dew point (°C), and wind components ($m \cdot s^{-1}$), February, 1200 G.C.T., 1964 through 1972; n = 251. Covariances are assumed to be unequal. The 0.25 probability ellipses are shown for the wind distribution. The total distribution and the breakout into four clusters are shown. The mixture proportion and the averages, the pressure, the temperature, and the dew point data within each cluster are shown.

8.3.6 Radiosonde and Rawinsonde Data Set

8.3.6.1 Input Information.

a. Balboa (Albrook Field), Canal Zone
b. The period of record is the month of July 1961-1970.
c. The data are pressures or height (mb or m) temperatures (°C), dew points (°C), east-west (u) wind components, and north-south (v) wind components, $m \cdot s^{-1}$ positive from the west and south.
d. The number of variables are 20; the above elements for the surface and the 850-, 700-, and 500-mb levels.
e. The number in the sample is 259.
f. The minimum number to be accepted into a sample is 21.
g. The null hypotheses are made that (k + 1) clusters are not significantly different from the k clusters. The decision probability level selected is 0.01. Rejection of the hypothesis then permits the assumption of (k + 1) clusters.
h. The first 40 twenty-dimensional vector entries are set up as the 40 vector means of 40 separate and individual clusters. These are 40 points in twenty dimensions.
i. Two assumptions are made. The first assumption is the equality of covariance matrices. The second assumption is the non-equality of the covariance matrices.

8.3.6.2 Tables and Discussion. Table 24 provides selected output data of the Wolfe (NORMIX) computer routine. The logarithm of the likelihood ratio of 2 to 1 types is 676.30485, the chi-square with 40 degrees of freedom is 1240.33, and the probability of the null hypothesis not being rejected is to seven decimals, 0.0000000. Therefore, the two-cluster configuration is not rejected. The computation failed to converge for the three-cluster versus the two-cluster configuration. Convergence under the assumption of unequal covariance matrices also failed. No figures are provided here as the number of variables is too high. A number of two-dimensional figures, however, could be made from the statistics provided.

In order to determine the capability of the NCC computer (a Univac Series 70) and in view of the previous work, a multivariate problem with 40 element vectors was chosen for Balboa, C.Z. There were eight levels of each radiosonde with five elements, pressure (or height), temperature, dew points, and the east-west and north-south components of the winds. The levels were the surface and the 950-, 900-, 850-, 800-, 700-, 600-, and 500-mb levels. The period chosen was the months of July during 1961-1970. The assumption was made that the covariance matrices were different. The computation failed to converge for a two-group separation. The number of vector elements was reduced to 20. Again the computation failed to converge.

The assumption of different covariance matrices was then replaced by the assumption of equal covariance matrices. The computation converged for a breakout of two groups but failed on three groups when the vectors were composed of 20 elements, five elements from each of four of the levels above, namely the surface, and the 850-, 700-, and 500-mb levels as these were the only elements believed to have moisture measurements in sufficient quantities to assure enough input data.

Examination of table 24 shows that though there does not appear to be too much difference, there is some. About 95 percent of the data comprise one (Type 1) cluster while 5 percent of the data comprise the other (Type 2) cluster. Cluster 1 is cooler than cluster 2 at all levels through 500 mb, an altitude of roughly 5854 m. Cluster 1 is more moist than cluster 2 at altitudes above about 3,100 m and probably above 2000 m. Cluster 1 exhibits lower wind speeds than does cluster 2 from the surface upwards. The difference ranges from 1.2 $m \cdot s^{-1}$ at 1500 m to 1.6 $m \cdot s^{-1}$ at 5854 m. Cluster 1 winds shift from east northeast to east by southeast at the highest level while the winds in cluster 2 remain east by northeast throughout the layer. The surface pressure of cluster 1 is about 0.6 mb higher than that of cluster 2.

The types of weather that accompany these groups have not been investigated here. It would be interesting to do so. Balboa, C.Z., data were selected (1) to provide an insight into the lower level atmospheric characteristics, (2) to further understanding of the capability of this program, adopted for use on the Univac 70 at the NCC, to handle multidimensional problems, and (3) to demonstrate the utilization of such a program to consider each radiosonde observation as a point in multidimensional space.

From the above experience it appears that the program, restricted by the present Univac 70 configuration, can handle 300 input 20-dimensional data problems. The first assumption made should be the assumption of equal covariance matrices.

Use of the program for data at other stations where greater differences may be expected may permit the use of more input data, greater dimensions, and the assumption of unequal covariance matrices.

The minimum number to be accepted into a cluster is always one more than the number of dimensions. In this case the minimum number is 21. This is about 8.1 percent. Cluster Type 2 has only about 5.4 percent. Therefore, it would be advisable to check those observations assigned to Type 2 by the discrimination function for the possibility of all or some of these being outliers. This is not done here as this is the basis for work beyond the scope of this paper.

Table 24 Means and standard deviations for the total set and clusters 1 and 2 of the data for Balboa, C.Z. These data are the pressure or height (mb or m), temperature (°C), dew point (°C), and the u and v components of the wind ($m \cdot s^{-1}$) at the surface, 850-, 700-, and 500-mb levels. The dimensions are 20. Equal covariance matrices are assumed for types 1 and 2.

Logarithm of the likelihood;	chi-square with 4 d.f.;	probability of null hypothesis.
(a) 2 to 1 types		(a) 0.054

Characteristics of the whole sample (units are $m \cdot s^{-1}$)

	Type 1		Whole Sample		Type 2	
	Proportion 0.9459		1.000		Proportion 0.054	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Surface-pressure (1)	1002.20	12.18	1002.18	12.24	1001.66	12.18
temperature (2)	23.91	0.92	23.92	0.92	24.07	0.92
dew point (3)	23.14	1.16	23.15	1.16	23.33	1.16
u (4)	1.22	1.20	1.23	1.20	1.27	1.20
v (5)	-1.47	1.01	-1.45	1.02	-1.11	1.01
Height (1)	1500.84	9.39	1499.24	11.53	1471.21	9.39
temperature (2)	17.78	0.96	17.81	0.97	18.35	0.96
dew point (3)	14.76	1.95	14.80	1.95	15.36	1.95
u (4)	-3.96	2.61	-3.99	2.62	-4.55	2.61
v (5)	-2.68	2.39	-2.71	2.40	-3.32	2.39
Height (1)	3141.02	11.05	3139.29	13.21	3109.00	11.05
temperature (2)	9.46	1.02	9.49	1.03	10.08	1.02
dew point (3)	2.82	2.74	2.81	2.74	2.52	2.74
u (4)	-6.28	3.01	-6.34	3.02	-7.37	3.01
v (5)	0.24	2.36	0.20	2.36	-0.52	2.36
Height (1)	5852.69	14.42	5852.75	14.43	5853.93	14.42
temperature (2)	-6.67	0.90	-6.79	0.91	-6.29	0.90
dew point (3)	-14.27	4.75	-14.38	4.77	-16.36	4.75
u (4)	-6.16	4.00	-6.23	4.01	-7.54	4.00
v (5)	1.40	3.05	1.36	3.06	0.59	3.05

9. MULTIVARIATE QUALITY ASSURANCE AND CONTROL

The capability of the NORMIX program permits a quick elementary view of principal component analysis and of multivariate quality assurance and control. The techniques of assurance and control are the same. Assurance is for incoming data or material, while control is for data or material processing. Essentially, these are filtering techniques.

Here, we look at the separation of homogeneous subsets out of mixed sets which permits the isolation of "outliers" for examination.

Figure 28 (Crutcher, 1966) is an example of a two-tailed Gaussian filter operating on a set of univariate heterogeneous data to isolate outlying data. This essentially sets up two or more groups or subsets. The progress of the sub-figures (a) through (e) schematically shows how the procedure works on a univariate distribution. The alpha level of rejection is 0.05. In this illustration the main group is isolated and the statistics estimated. The other groups have been isolated for further study as to whether they are valid data.

The NORMIX program clustering techniques essentially go through the same technique. However, the decision (alpha) level chosen for the previous examples was 0.01 probability rather than 0.05.

Use can be made of the univariate Gaussian filter or the NORMIX filter can operate on any univariate set of mixed normal distributions. Refer to the Canton Island, July, 30-mb data set. Figure 15 illustrates by means of 0.50 probability ellipses the breakdown of the total set of data into two subsets with quite different characteristics. (See section 8.3.2.3.) A univariate Gaussian filter could be used on the zonal or meridional components. To demonstrate the application of the NORMIX clustering technique as well as principal component techniques, zonal and meridional components are transformed to uncorrelated pairwise data along the major and minor axes, the two principal axes of the distribution. Prior to and after the rotation, the components are standardized to a zero mean and a variance of 1. See table 25.

Figures 29a and 29b show the frequency distribution computer printer plot of the components along the two principal axes. Figure 29a clearly delineates the two clusters. In the latter case, separation would be most difficult on a one by one basis, but at least the mean and variance can be estimated. The existence of groups with equal mean and variance would not be indicated. Clearly indicated in Figures 29a and 29b are the potential invalid outliers at +1.95 on the major axis and beyond 4.16 on the minor axis. These are simply pointed out to the reader but are not examined for validity.

Figures 30a, 30b, and 30c (Crutcher, 1966) illustrate a Gaussian filter technique in two dimensions. Figure 30a illustrates a sample drawn from a homogeneous bivariate distribution contaminated by three subsets of data. One of the subsets consisting of only a datum is first isolated, eliminated from further immediate consideration, but set aside for investigation as to its validity.

The clear separation along the major axis exhibited by the frequency diagram of figure 30a is shown by the above statistics. The separation along the minor axis is not so clearly demonstrated.

Refer to figure 29a and tables 25 and 26. Table 25 provides the statistics for the components along the major axis of the distribution of winds at Canton Island, U.S.A. and U.K., for July 1954-1964. These are standardized transformed variables and as such are dimensionless in the terms of units as in all figures in this section. The mean of the total distribution is zero and the variance is about 1.0006 to four decimals. As this is a standardized set of data, the mean should be zero and the variance one. The error in the fourth place is a numerical rounding error. The set breaks down into two clusters. The mean of the first group is -1.0420 with a standard deviation of 0.2442. The respective values for the second group are +0.7923 and 0.5127.

The output from the Cohen-Falls (1967) program provides the data in table 26. Note the output from ungrouped data as used in the NORMIX program, table 25. There are slight differences which are attributed to the use of ungrouped data and then grouped data.

The one outlier near 1.95 in figure 30a would not be set aside as a cluster or outlier by the NORMIX program. As indicated elsewhere, the minimum cluster size is always one more than the number of variates.

Table 26 also shows the output of the Cohen-Falls program for the dimensionless wind components along the minor axis as illustrated in figure 29b. It is recognized that, with the exception of the outliers, the mixed distribution has degenerated into a single standardized univariate unimodal normal distribution with a zero mean and variance of one. Note the proportion of two groups 0.998 and 0.002 with only 244 observations. Note the mean of -0.009617 in the first group with the extraordinarily large mean of 7.410603 for the second group. Remember the standard deviation of the entire set is only one. Then look at the untenable negative variance obtained which is printed as x.xxxxx. However, this negative variance indicates the difficulties in arriving at a solution and clearly implies that it is the simple case of a symmetric mixed or compound distribution with equal means and equal variances with proportions equal. Therefore, the decision will be to use the estimates for the total sample as estimates of the two groups (a mean of zero and a variance of one in each case). This is the trivial degenerate case which does provide some computing difficulty but no interpretative difficulty.

The elimination of the outliers along the minor axis and the re-standardization of the data will provide a mean of zero and a variance of one rather than the values given in table 26.

Figures 31a through 31c show the frequency distributions of the components along three principal axes of the Stampede Pass, Easton, Washington, wind and temperature standardized data. These refer to the figure 23. The existence of outliers is clearly demonstrated. Outliers lie beyond -2.56 on the first principal axis, beyond -1.68 on the second principal axis, and in both tails of the third principal axis. Undoubtedly, some of these belong to the extreme low temperature cluster isolated and discussed in section 8.3.4. The flatness of the distribution on all three axes and the more extreme bimodality exhibited on the third axis show the existence of the groups or clusters already isolated.

Extension of the Gaussian filter to more than one dimension at a time essentially is the standardization of all components along the major axes so that the distribution may be assumed to be spherical (Crutcher, 1966). The distribution of the standardized vectors in n-dimensions then is chi-square with n degrees of freedom. The appropriate decision probability levels then are obtained from any standard text book containing chi-square tables. Alternatively, there are computer routines to compute the required values.

For example, the magnitude of the standardized vector in the one-dimensional case for the rejection of the top and bottom 0.5 percent (0.005 probability) implying non-rejection of the central 99 percent (0.99 probability) will be the square root of the 0.99 chi-square value divided by 1 (the degrees of freedom). This is 2.576 and is, of course, the t-distribution value for a large sample. In the case of a two-dimensional distribution, the magnitude of the standardized bivariate vector is the square root of 9.21/2 or 2.146. For the three-dimensional case, the vector radius would be 1.944.

The point made here is that a datum lying on the main principal axis, the major axis, has a better chance of being valid than an outlier on only one of the original coordinate axes. The chance for an observation to be incorrect in two variables is much less than the chance for an error in the observation of one variable. An outlying datum on one axis is, of course, projected as zero on the other transformed axes so that an outlier on one axis will not be associated with an outlier on another axis. In fact, its effect is to increase the frequency count at the mean or zero of the other axes. Another way of viewing this is that the rejection ellipse has a better chance of enclosing a possible outlier than the rejection bounds on either of the two original correlated axes.

Figure 30b shows the effect of computing a new filter ellipse once the outlier shown in figure 30a has been eliminated. Now the two small contaminating subsets undetected in the first operation are isolated. Figure 30c then illustrates the last step where all data points remaining are inside the filter ellipse. All rejected data are set aside for investigation as to their validity. Before this check is made, however, these data may be processed to see whether they constitute one or more groups or clusters.

The procedure, of course, brings up the possible censoring or truncation problem. Adequate adjustments and better estimates of the parameters can be made.

There is no discussion here of the applications of higher moments to the problem of outlier detection, isolation, and removal.

Figures 32a through 32t illustrate the frequency distributions of the radiosonde data discussed in section 8.3.6. This discussion indicated only a slight separation of clusters as contrasted with the stratospheric data of section 8.3.2. The frequency diagrams of figures 32a through 32t also show only a slight separation potential in the flat of platykurtic curves.

Figures 32a through 32t are shown more to demonstrate the outlier problem than the clustering problem. The mathematical requirements of the clustering techniques always specify that the minimum cluster size is one more than the number of variates. Thus, with twenty variables the smallest cluster size is twenty-one. A cluster of this size would really indicate a true valid cluster created by weather conditions or a continuing bias in the procedures or operating conditions. The lone outlier(s) in multi-space might go undetected. However, the frequency distribution along a principal axis or a clustering technique along each of the axes would isolate, identify, and permit removal of the questionable data. Therefore, the frequency distributions of figures 29a and b and 32a through t can be used to look at the few individual outliers. Clustering techniques would have a better chance to show this outlier as there is only one variable or a cluster size of two.

Alt et al. (1973) discuss quality assurance and control in their paper on the use of control charts for multivariable data. Crutcher and Falls (1976) discuss the testing of data sets for multivariate normality. The above two reports and this present report add to the rather meager literature on the subject of multivariate quality assurance and control.

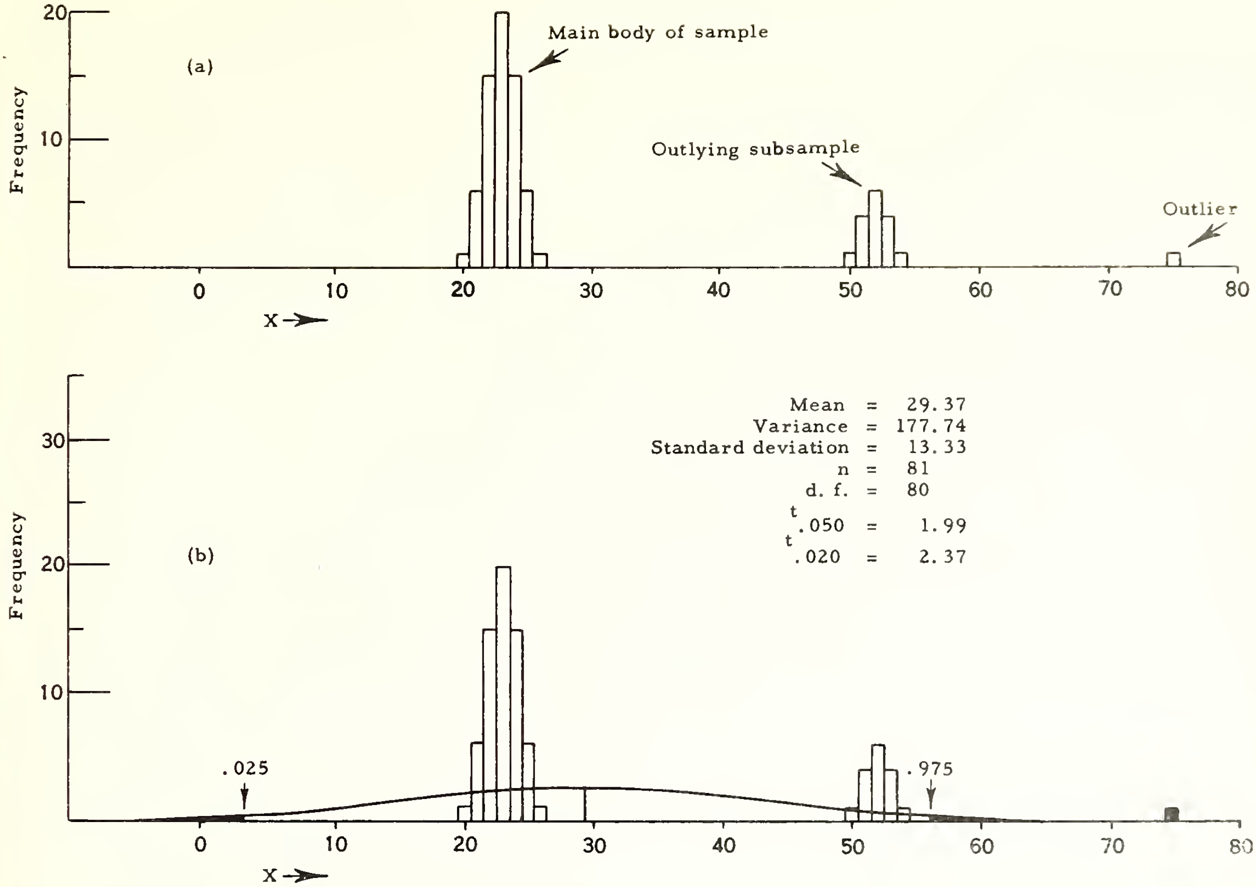


Figure 28 Example of a two-tailed Gaussian filter operating on a set of heterogeneous data to isolate, set aside, and eliminate outlying data. The dark areas show the rejection level, 0.05 (0.025 in each tail). Crutcher (1966).

(a) Here are the data in histogram form.
(b) Here are the data with the theoretical fitting curve under the assumption of normality and independence of data. The rejection areas under each tail are shown beyond the 0.025 and 0.975 points. The observation at 75 is rejected.
(c) Here is the theoretical fitting curve of the data which passed the filter in (b). Again the rejection areas under each tail beyond the 0.025 and 0.975 points are shown. Data beyond 52 are rejected.
(d) Here is the theoretical fitting curve of the data which passed the filter in (c). Again the rejection areas under each tail beyond the 0.025 and 0.975 points are shown. Data beyond 47 are rejected.
(e) Here is the theoretical fitting curve for the data which passed the filter in (d). Although the rejection areas are not shown, the singleton counts below 21 and above 25 would be rejected. This rejection is an unwanted rejection but is a penalty which must be accepted at this point. The next filtering step would result in no rejection.

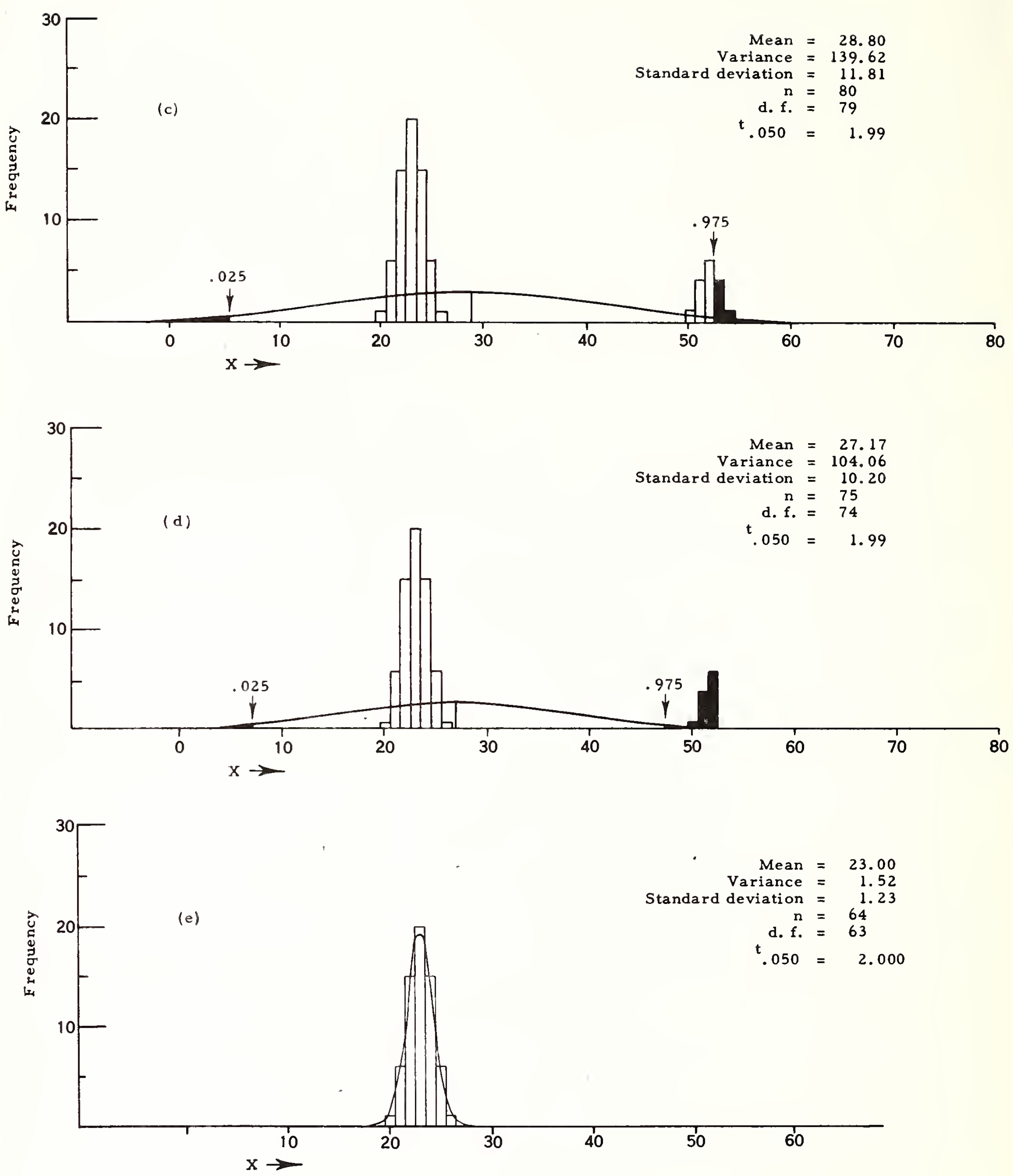
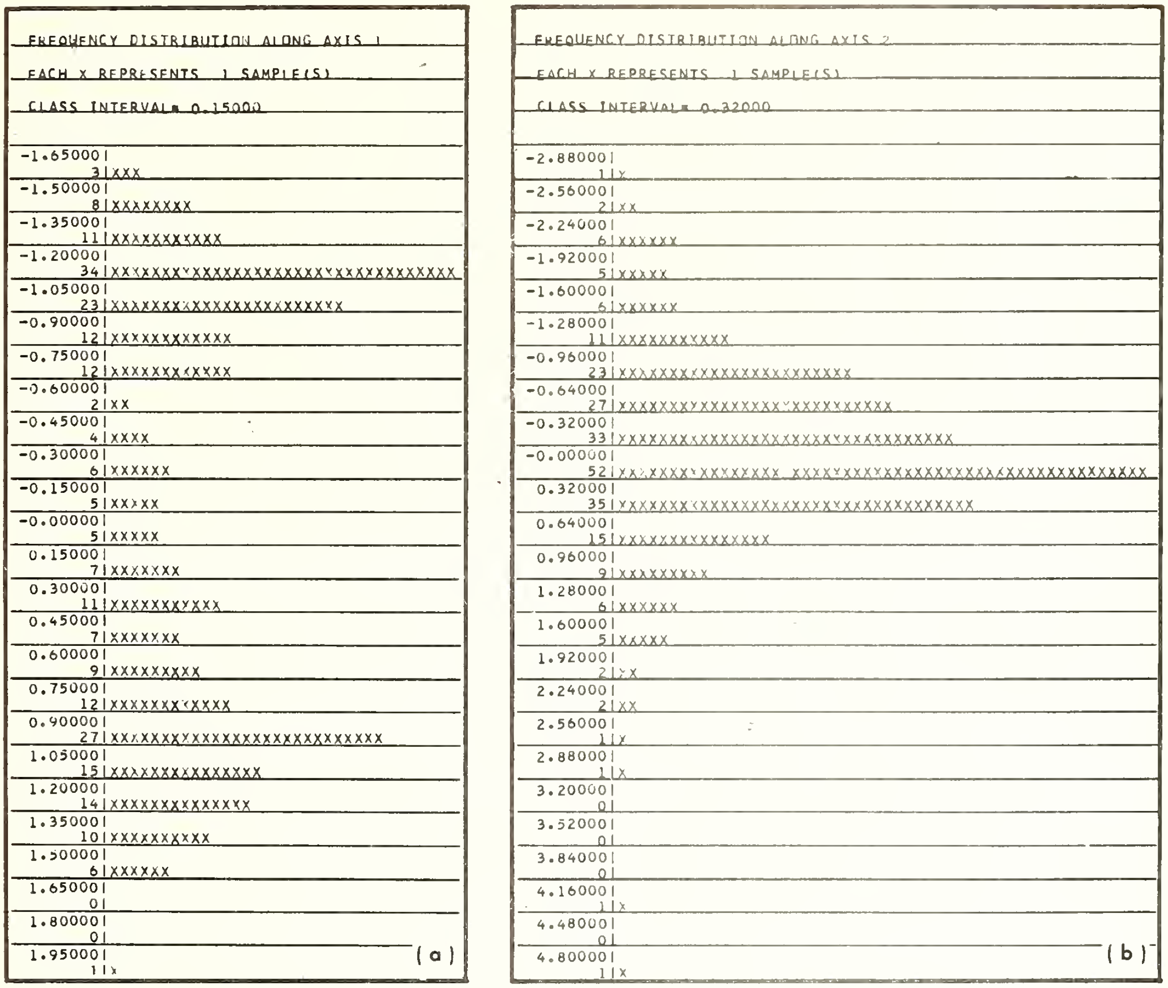


Figure 28 (continued)

SAMPLE = CANTON ISLAND JULY 30 MB X AND Y WIND COMPONENTS



```
FREQUENCY DISTRIBUTION ALONG AXIS 1
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.15000

-1.65000I
       3IXXX
-1.50000I
       8IXXXXXXXX
-1.35000I
      11IXXXXXXXXXXX
-1.20000I
      34IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-1.05000I
      23IXXXXXXXXXXXXXXXXXXXXXXX
-0.90000I
      12IXXXXXXXXXXXX
-0.75000I
      12IXXXXXXXXXXXX
-0.60000I
       2IXX
-0.45000I
       4IXXXX
-0.30000I
       6IXXXXXX
-0.15000I
       5IXXXXX
-0.00000I
       5IXXXXX
 0.15000I
       7IXXXXXXX
 0.30000I
      11IXXXXXXXXXXX
 0.45000I
       7IXXXXXXX
 0.60000I
       9IXXXXXXXXX
 0.75000I
      12IXXXXXXXXXXXX
 0.90000I
      27IXXXXXXXXXXXXXXXXXXXXXXXXXXX
 1.05000I
      15IXXXXXXXXXXXXXXX
 1.20000I
      14IXXXXXXXXXXXXXX
 1.35000I
      10IXXXXXXXXXX
 1.50000I
       6IXXXXXX
 1.65000I
       0I
 1.80000I
       0I
 1.95000I
       1IX                    (a)
```

```
FREQUENCY DISTRIBUTION ALONG AXIS 2
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.32000

-2.88000I
       1IX
-2.56000I
       2IXX
-2.24000I
       6IXXXXXX
-1.92000I
       5IXXXXX
-1.60000I
       6IXXXXXX
-1.28000I
      11IXXXXXXXXXXX
-0.96000I
      23IXXXXXXXXXXXXXXXXXXXXXXX
-0.64000I
      27IXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.32000I
      33IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000I
      52IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.32000I
      35IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.64000I
      15IXXXXXXXXXXXXXXX
 0.96000I
       9IXXXXXXXXX
 1.28000I
       6IXXXXXX
 1.60000I
       5IXXXXX
 1.92000I
       2IXX
 2.24000I
       2IXX
 2.56000I
       1IX
 2.88000I
       1IX
 3.20000I
       0I
 3.52000I
       0I
 3.84000I
       0I
 4.16000I
       1IX
 4.48000I
       0I
 4.80000I
       1IX                    (b)
```

Figure 29 Distribution of wind standardized components along the two principal axes of the Canton Island, U.S.A. and U.K., July, 30 mb.

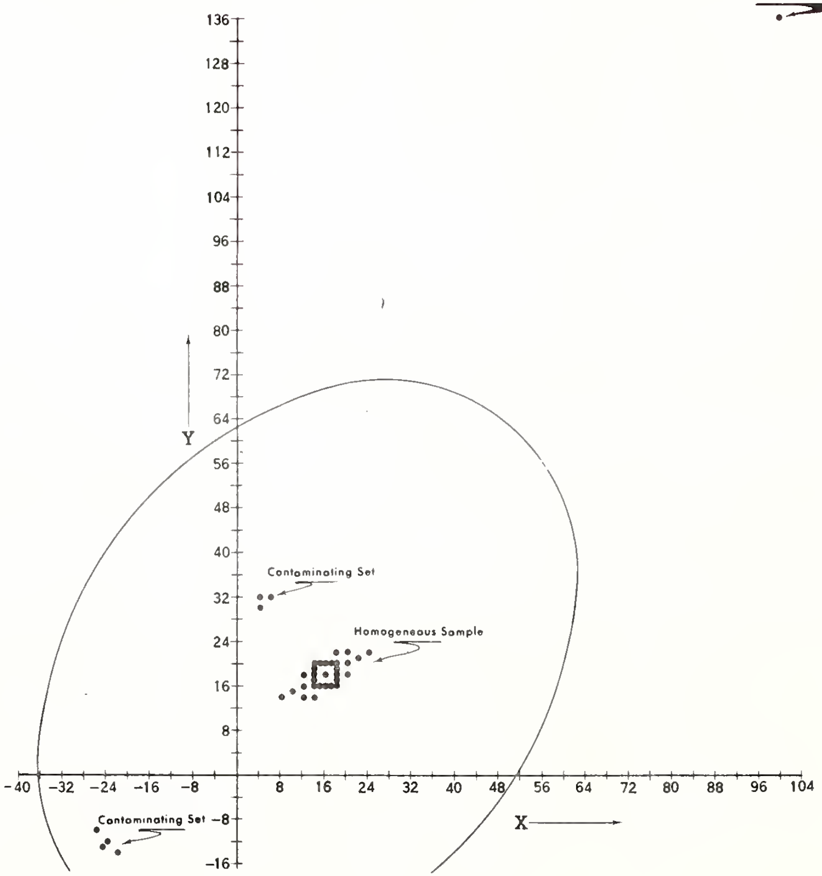


Figure 30a Schematic illustration of a sample drawn from a homogeneous bivariate distribution contaminated by a lone outlier at (100, 136) and two groups of data at approximately (-24, -12) and (4, 32). The result is a heterogeneous distribution. The ellipse shown is a theoretical 0.95 probability ellipse. The lone outlier will be rejected as not being part of the homogeneous distribution. Crutcher (1966).

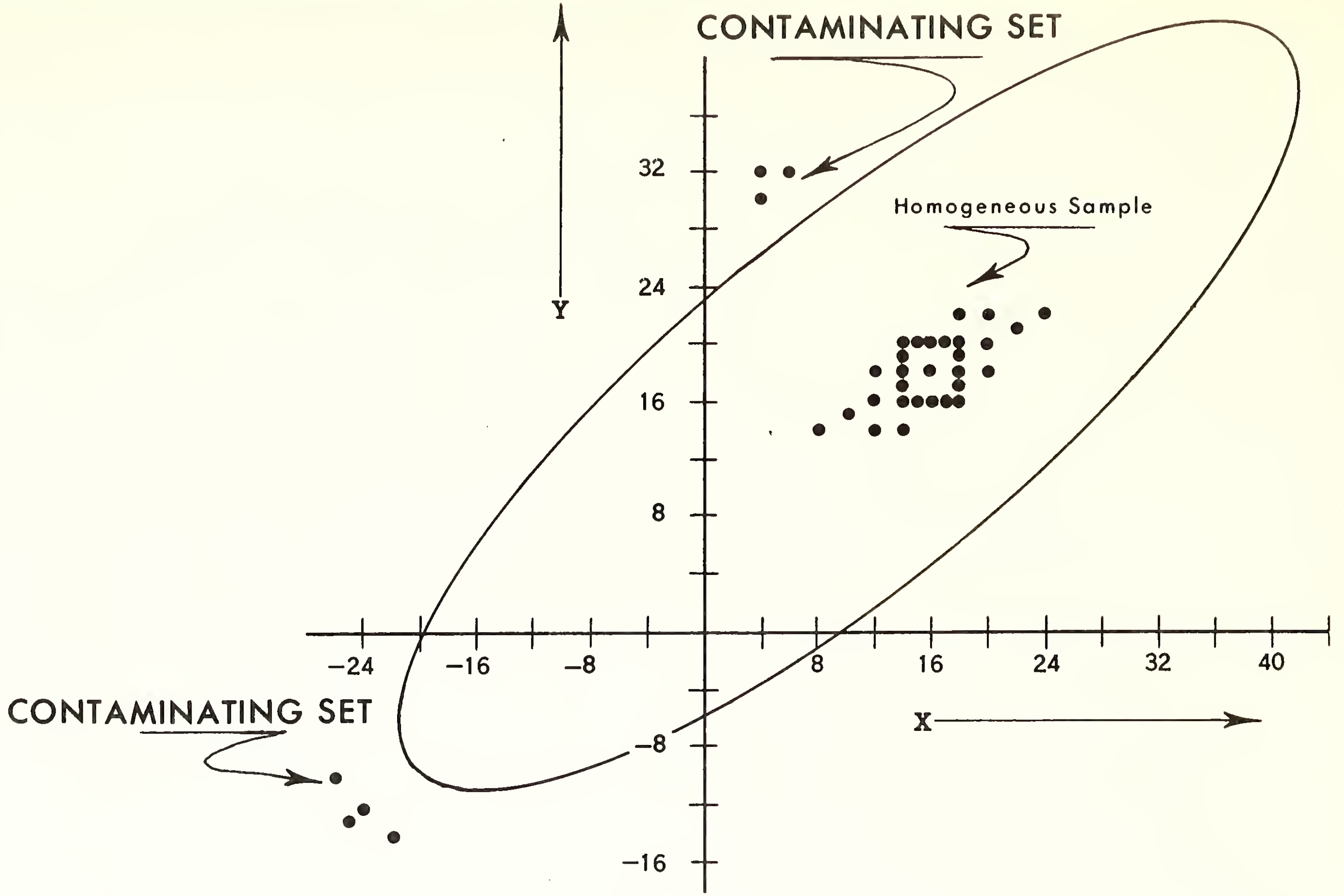


Figure 30b Schematic illustration of a sample drawn from a homogeneous bivariate distribution contaminated by two small sets. The result is a heterogeneous sample. The lone outlier of figure 30a has been eliminated as it did not appear within the 0.95 probability ellipse. Here, the two contaminating sets exist outside the 0.95 probability ellipse of this figure and will be eliminated in figure 30c. Crutcher (1966).

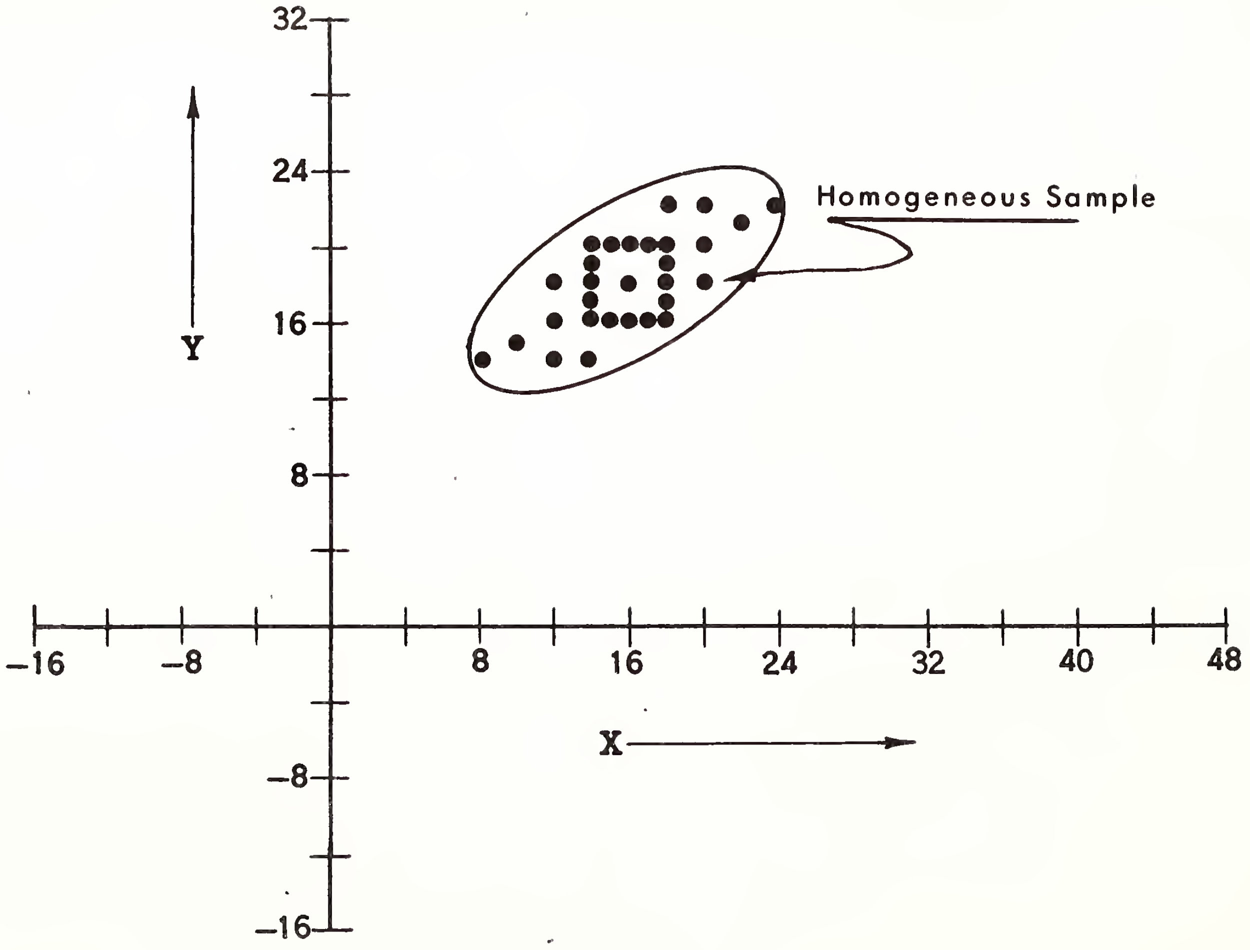


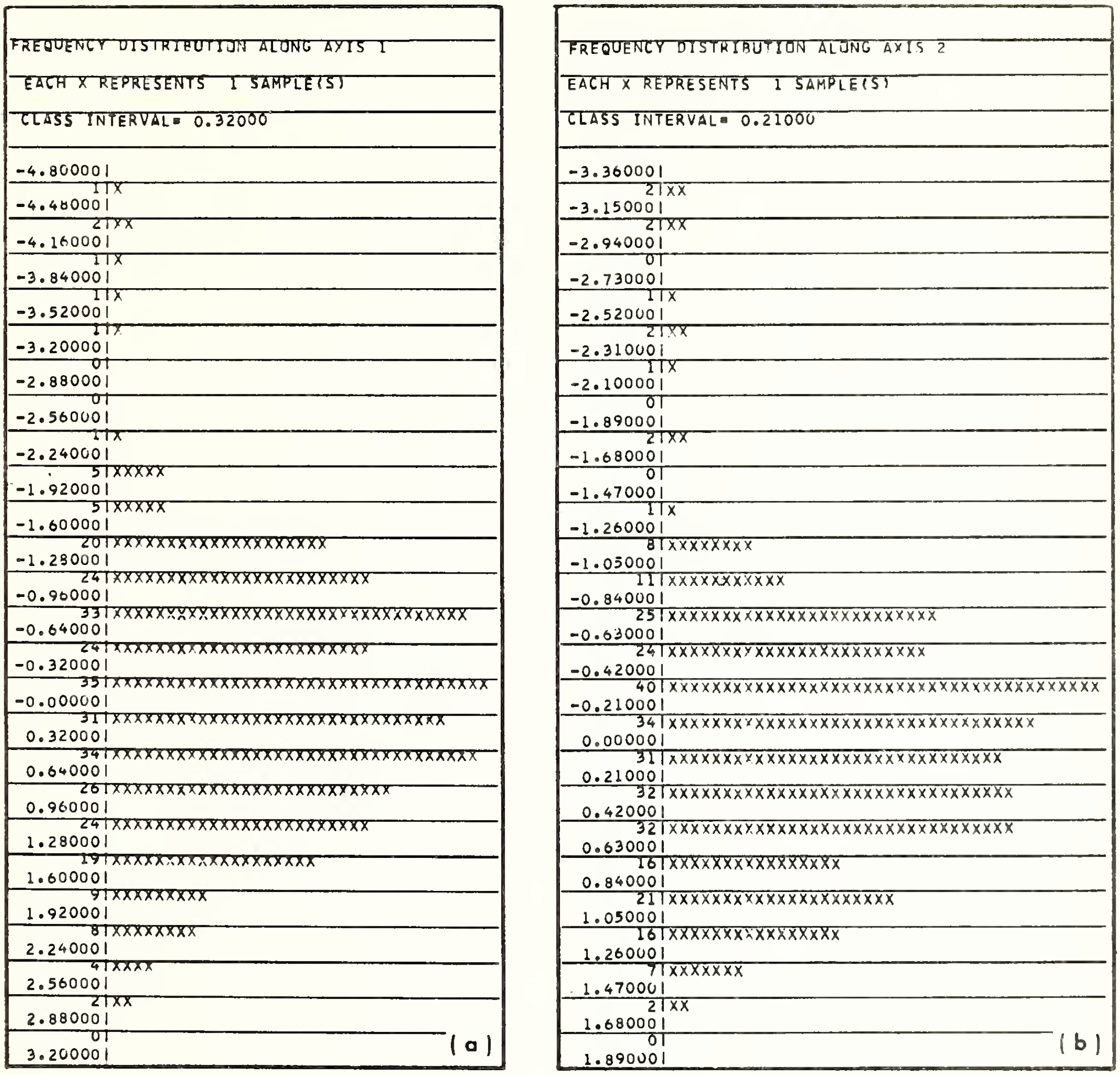
Figure 30c Schematic illustration of a sample drawn from a homogeneous bivariate distribution. It exists as a result of the filtering action of the 0.95 probability ellipse illustrated in figures 30a and 30b. Here, the 0.95 probability ellipse contains all sample data points of the remaining group. Crutcher (1966).

STAMPEDE PASS, WASHINGTON

SURFACE TEMPERATURE (DEGREE C), U AND V WIND COMPONENTS (MPS)

DECEMBER 1966, 1967, 1968, 1969, 1970

HOURS 07 AND 13 LOCAL TIME (1ST 155 INPUTS ARE HR 07)



```
FREQUENCY DISTRIBUTION ALONG AXIS 1
 EACH X REPRESENTS   1 SAMPLE(S)
 CLASS INTERVAL= 0.32000

-4.80000I
       1IX
-4.48000I
       2IXX
-4.16000I
       1IX
-3.84000I
       1IX
-3.52000I
       1IX
-3.20000I
       0I
-2.88000I
       0I
-2.56000I
       1IX
-2.24000I
       5IXXXXX
-1.92000I
       5IXXXXX
-1.60000I
      20IXXXXXXXXXXXXXXXXXXXX
-1.28000I
      24IXXXXXXXXXXXXXXXXXXXXXXXX
-0.96000I
      33IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.64000I
      24IXXXXXXXXXXXXXXXXXXXXXXXX
-0.32000I
      35IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000I
      31IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.32000I
      34IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.64000I
      26IXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.96000I
      24IXXXXXXXXXXXXXXXXXXXXXXXX
 1.28000I
      19IXXXXXXXXXXXXXXXXXXX
 1.60000I
       9IXXXXXXXXX
 1.92000I
       8IXXXXXXXX
 2.24000I
       4IXXXX
 2.56000I
       2IXX
 2.88000I
       0I
 3.20000I
```

(a)

```
FREQUENCY DISTRIBUTION ALONG AXIS 2
EACH X REPRESENTS   1 SAMPLE(S)
CLASS INTERVAL= 0.21000

-3.36000I
       2IXX
-3.15000I
       2IXX
-2.94000I
       0I
-2.73000I
       1IX
-2.52000I
       2IXX
-2.31000I
       1IX
-2.10000I
       0I
-1.89000I
       2IXX
-1.68000I
       0I
-1.47000I
       1IX
-1.26000I
       8IXXXXXXXX
-1.05000I
      11IXXXXXXXXXXX
-0.84000I
      25IXXXXXXXXXXXXXXXXXXXXXXXXX
-0.63000I
      24IXXXXXXXXXXXXXXXXXXXXXXXX
-0.42000I
      40IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.21000I
      34IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.00000I
      31IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.21000I
      32IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.42000I
      32IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.63000I
      16IXXXXXXXXXXXXXXXX
 0.84000I
      21IXXXXXXXXXXXXXXXXXXXXX
 1.05000I
      16IXXXXXXXXXXXXXXXX
 1.26000I
       7IXXXXXXX
 1.47000I
       2IXX
 1.68000I
       0I
 1.89000I
```

(b)

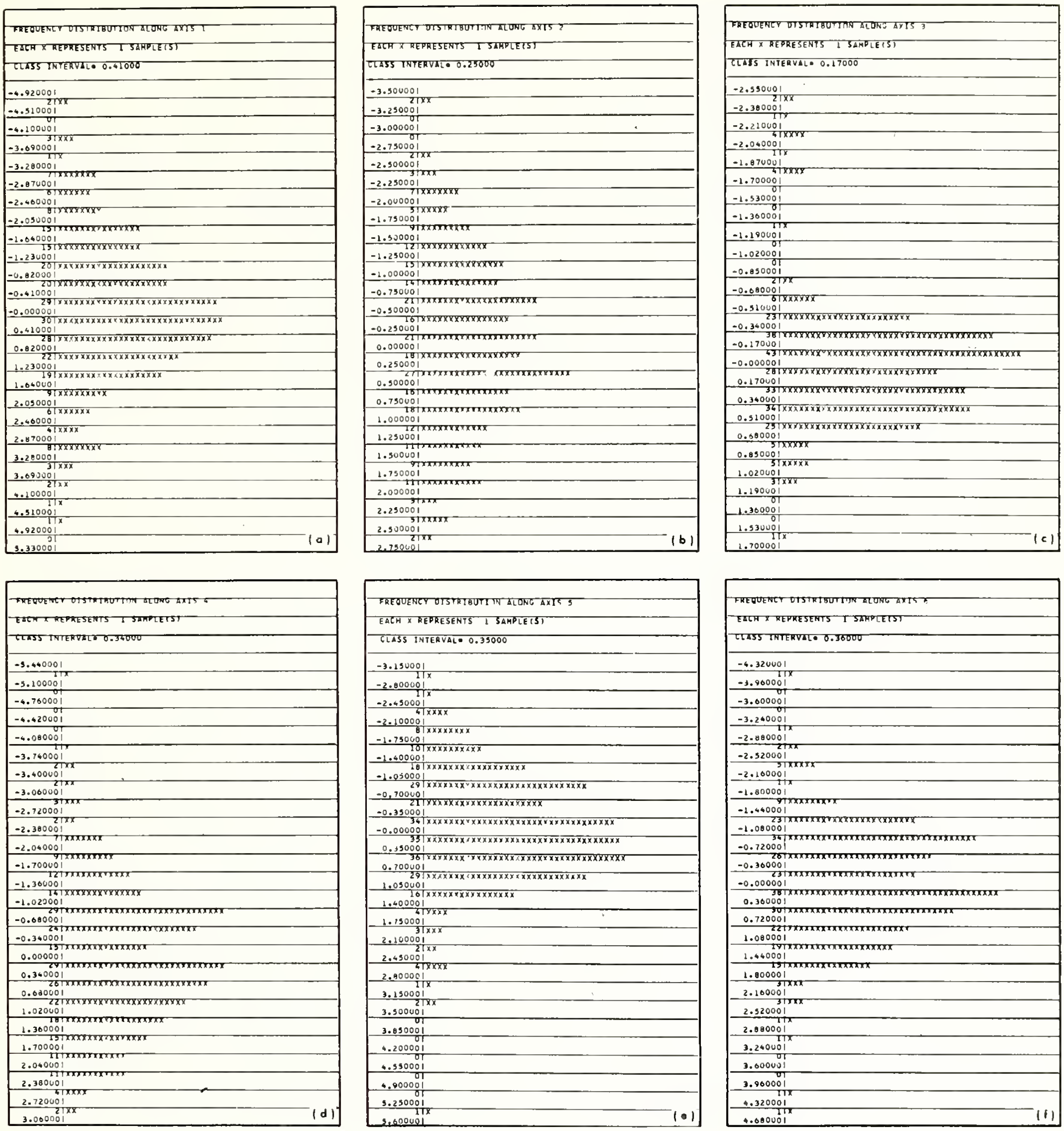
Figure 31 Distribution of wind and temperature standardized components along the three principal axes of the Stampede Pass, Easton, Washington, December 1968-1970 data.

```
FREQUENCY DISTRIBUTION ALONG AXIS 3

EACH X REPRESENTS  1 SAMPLE(S)

CLASS INTERVAL= 0.32000

-3.84000I
       1IX
-3.52000I
       0I
-3.20000I
       0I
-2.88000I
       1IX
-2.56000I
       0I
-2.24000I
       5IXXXXX
-1.92000I
       3IXXX
-1.60000I
       9IXXXXXXXXX
-1.28000I
       7IXXXXXXX
-0.96000I
      84IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.64000I
      20IXXXXXXXXXXXXXXXXXXXX
-0.32000I
      31IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000I
      51IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.32000I
      19IXXXXXXXXXXXXXXXXXXX
 0.64000I
      23IXXXXXXXXXXXXXXXXXXXXXXX
 0.96000I
      25IXXXXXXXXXXXXXXXXXXXXXXXXX
 1.28000I
      12IXXXXXXXXXXXX
 1.60000I
       8IXXXXXXXX
 1.92000I
       5IXXXXX
 2.24000I
       4IXXXX
 2.56000I
       1IX
 2.88000I
       0I
 3.20000I
       0I
 3.52000I
       0I
 3.84000I
       1IX
 4.16000I
```

(c)

Figure 31 (continued)

ALBROOK AFB CANAL ZONE   RADIOSO
HEIGHT (PRESSURE AT SURFACE), TEMPERATURE, DEW POINT, U AND V WIND COMP
AT SURFACE, 850MB, 700MB, 500MB
JULY 12Z 1961 THRU 1970



```
FREQUENCY DISTRIBUTION ALONG AXIS 1
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.41000

-4.92000|
       2|XX
-4.51000|
       0|
-4.10000|
       3|XXX
-3.69000|
       1|X
-3.28000|
       7|XXXXXXX
-2.87000|
       6|XXXXXX
-2.46000|
       8|XXXXXXXX
-2.05000|
      15|XXXXXXXXXXXXXXX
-1.64000|
      15|XXXXXXXXXXXXXXX
-1.23000|
      20|XXXXXXXXXXXXXXXXXXXX
-0.82000|
      20|XXXXXXXXXXXXXXXXXXXX
-0.41000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      30|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.41000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.82000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 1.23000|
      19|XXXXXXXXXXXXXXXXXXX
 1.64000|
       9|XXXXXXXXX
 2.05000|
       6|XXXXXX
 2.46000|
       4|XXXX
 2.87000|
       8|XXXXXXXX
 3.28000|
       3|XXX
 3.69000|
       2|XX
 4.10000|
       1|X
 4.51000|
       1|X
 4.92000|
       0|
 5.33000|
```

(a)

```
FREQUENCY DISTRIBUTION ALONG AXIS 2
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.25000

-3.50000|
       2|XX
-3.25000|
       0|
-3.00000|
       0|
-2.75000|
       2|XX
-2.50000|
       3|XXX
-2.25000|
       7|XXXXXXX
-2.00000|
       5|XXXXX
-1.75000|
       9|XXXXXXXXX
-1.50000|
      12|XXXXXXXXXXXX
-1.25000|
      15|XXXXXXXXXXXXXXX
-1.00000|
      14|XXXXXXXXXXXXXX
-0.75000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.50000|
      16|XXXXXXXXXXXXXXXX
-0.25000|
      21|XXXXXXXXXXXXXXXXXXXXX
 0.00000|
      18|XXXXXXXXXXXXXXXXXX
 0.25000|
      27|XXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.50000|
      16|XXXXXXXXXXXXXXXX
 0.75000|
      18|XXXXXXXXXXXXXXXXXX
 1.00000|
      12|XXXXXXXXXXXX
 1.25000|
      11|XXXXXXXXXXX
 1.50000|
       9|XXXXXXXXX
 1.75000|
      11|XXXXXXXXXXX
 2.00000|
       3|XXX
 2.25000|
       5|XXXXX
 2.50000|
       2|XX
 2.75000|
```

(b)

```
FREQUENCY DISTRIBUTION ALONG AXIS 3
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.17000

-2.55000|
       2|XX
-2.38000|
       1|X
-2.21000|
       4|XXXX
-2.04000|
       1|X
-1.87000|
       4|XXXX
-1.70000|
       0|
-1.53000|
       0|
-1.36000|
       1|X
-1.19000|
       0|
-1.02000|
       0|
-0.85000|
       2|XX
-0.68000|
       6|XXXXXX
-0.51000|
      23|XXXXXXXXXXXXXXXXXXXXXXX
-0.34000|
      38|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.17000|
      43|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.17000|
      33|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.34000|
      34|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.51000|
      25|XXXXXXXXXXXXXXXXXXXXXXXXX
 0.68000|
       5|XXXXX
 0.85000|
       5|XXXXX
 1.02000|
       3|XXX
 1.19000|
       0|
 1.36000|
       0|
 1.53000|
       1|X
 1.70000|
```

(c)

```
FREQUENCY DISTRIBUTION ALONG AXIS 4
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.34000

-5.44000|
       1|X
-5.10000|
       0|
-4.76000|
       0|
-4.42000|
       0|
-4.08000|
       1|X
-3.74000|
       2|XX
-3.40000|
       2|XX
-3.06000|
       3|XXX
-2.72000|
       2|XX
-2.38000|
       7|XXXXXXX
-2.04000|
       9|XXXXXXXXX
-1.70000|
      12|XXXXXXXXXXXX
-1.36000|
      14|XXXXXXXXXXXXXX
-1.02000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.68000|
      24|XXXXXXXXXXXXXXXXXXXXXXXX
-0.34000|
      15|XXXXXXXXXXXXXXX
 0.00000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.34000|
      26|XXXXXXXXXXXXXXXXXXXXXXXXXX
 0.68000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 1.02000|
      18|XXXXXXXXXXXXXXXXXX
 1.36000|
      15|XXXXXXXXXXXXXXX
 1.70000|
      11|XXXXXXXXXXX
 2.04000|
      11|XXXXXXXXXXX
 2.38000|
       4|XXXX
 2.72000|
       2|XX
 3.06000|
```

(d)

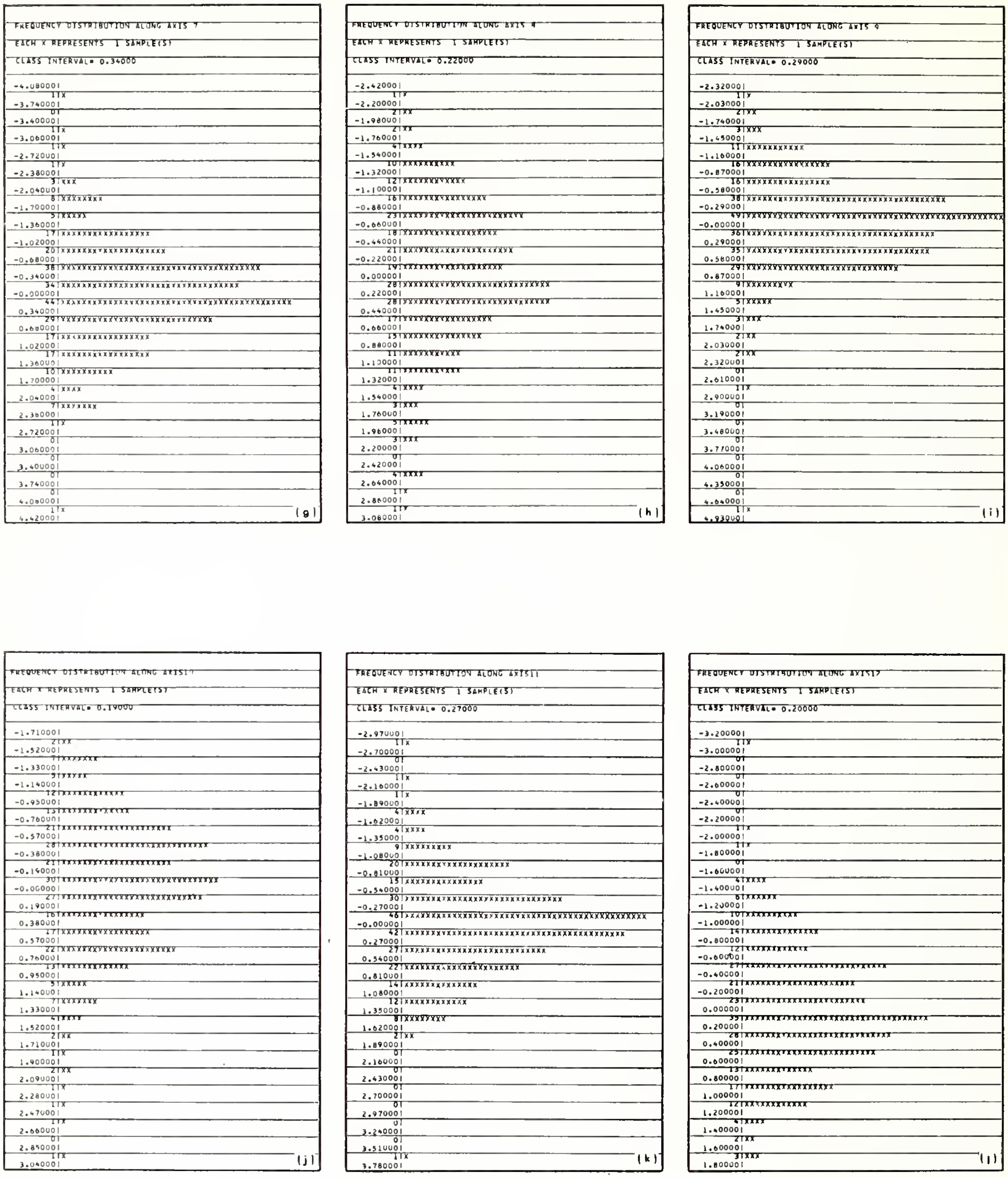
```
FREQUENCY DISTRIBUTION ALONG AXIS 5
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.35000

-3.15000|
       1|X
-2.80000|
       1|X
-2.45000|
       4|XXXX
-2.10000|
       8|XXXXXXXX
-1.75000|
      10|XXXXXXXXXX
-1.40000|
      18|XXXXXXXXXXXXXXXXXX
-1.05000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.70000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.35000|
      34|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      35|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.35000|
      36|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.70000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 1.05000|
      16|XXXXXXXXXXXXXXXX
 1.40000|
       4|XXXX
 1.75000|
       3|XXX
 2.10000|
       2|XX
 2.45000|
       4|XXXX
 2.80000|
       1|X
 3.15000|
       2|XX
 3.50000|
       0|
 3.85000|
       0|
 4.20000|
       0|
 4.55000|
       0|
 4.90000|
       0|
 5.25000|
       1|X
 5.60000|
```

(e)

```
FREQUENCY DISTRIBUTION ALONG AXIS 6
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.36000

-4.32000|
       1|X
-3.96000|
       0|
-3.60000|
       0|
-3.24000|
       1|X
-2.88000|
       2|XX
-2.52000|
       5|XXXXX
-2.16000|
       1|X
-1.80000|
       9|XXXXXXXXX
-1.44000|
      23|XXXXXXXXXXXXXXXXXXXXXXX
-1.08000|
      34|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.72000|
      26|XXXXXXXXXXXXXXXXXXXXXXXXXX
-0.36000|
      23|XXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      38|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.36000|
      30|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.72000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 1.08000|
      19|XXXXXXXXXXXXXXXXXXX
 1.44000|
      15|XXXXXXXXXXXXXXX
 1.80000|
       3|XXX
 2.16000|
       3|XXX
 2.52000|
       1|X
 2.88000|
       1|X
 3.24000|
       0|
 3.60000|
       0|
 3.96000|
       1|X
 4.32000|
       1|X
 4.68000|
```

(f)

Figure 32 Distribution of wind, temperature, height, and dew point standardized components along the 20 principal axes of the Balboa, C.Z., July data. Four levels are involved, surface, 850-, 700-, and 500-mb.



```
FREQUENCY DISTRIBUTION ALONG AXIS 7
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.34000

-4.08000|
       1|X
-3.74000|
       0|
-3.40000|
       1|X
-3.06000|
       1|X
-2.72000|
       1|X
-2.38000|
       3|XXX
-2.04000|
       8|XXXXXXXX
-1.70000|
       5|XXXXX
-1.36000|
      17|XXXXXXXXXXXXXXXXX
-1.02000|
      20|XXXXXXXXXXXXXXXXXXXX
-0.68000|
      38|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.34000|
      34|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      44|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.34000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.68000|
      17|XXXXXXXXXXXXXXXXX
 1.02000|
      17|XXXXXXXXXXXXXXXXX
 1.36000|
      10|XXXXXXXXXX
 1.70000|
       4|XXXX
 2.04000|
       7|XXXXXXX
 2.38000|
       1|X
 2.72000|
       0|
 3.06000|
       0|
 3.40000|
       0|
 3.74000|
       0|
 4.08000|
       1|X
 4.42000|
```

(g)

```
FREQUENCY DISTRIBUTION ALONG AXIS 8
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.22000

-2.42000|
       1|X
-2.20000|
       2|XX
-1.98000|
       2|XX
-1.76000|
       4|XXXX
-1.54000|
      10|XXXXXXXXXX
-1.32000|
      12|XXXXXXXXXXXX
-1.10000|
      16|XXXXXXXXXXXXXXXX
-0.88000|
      23|XXXXXXXXXXXXXXXXXXXXXXX
-0.66000|
      18|XXXXXXXXXXXXXXXXXX
-0.44000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.22000|
      19|XXXXXXXXXXXXXXXXXXX
 0.00000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.22000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.44000|
      17|XXXXXXXXXXXXXXXXX
 0.66000|
      15|XXXXXXXXXXXXXXX
 0.88000|
      11|XXXXXXXXXXX
 1.10000|
      11|XXXXXXXXXXX
 1.32000|
       4|XXXX
 1.54000|
       3|XXX
 1.76000|
       5|XXXXX
 1.96000|
       3|XXX
 2.20000|
       0|
 2.42000|
       4|XXXX
 2.64000|
       1|X
 2.86000|
       1|X
 3.08000|
```

(h)

```
FREQUENCY DISTRIBUTION ALONG AXIS 9
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.29000

-2.32000|
       1|X
-2.03000|
       2|XX
-1.74000|
       3|XXX
-1.45000|
      11|XXXXXXXXXXX
-1.16000|
      16|XXXXXXXXXXXXXXXX
-0.87000|
      16|XXXXXXXXXXXXXXXX
-0.58000|
      38|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.29000|
      49|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      36|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.29000|
      35|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.58000|
      29|XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.87000|
       9|XXXXXXXXX
 1.16000|
       5|XXXXX
 1.45000|
       3|XXX
 1.74000|
       2|XX
 2.03000|
       2|XX
 2.32000|
       0|
 2.61000|
       1|X
 2.90000|
       0|
 3.19000|
       0|
 3.48000|
       0|
 3.77000|
       0|
 4.06000|
       0|
 4.35000|
       0|
 4.64000|
       1|X
 4.93000|
```

(i)

```
FREQUENCY DISTRIBUTION ALONG AXIS10
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.19000

-1.71000|
       2|XX
-1.52000|
       7|XXXXXXX
-1.33000|
       5|XXXXX
-1.14000|
      12|XXXXXXXXXXXX
-0.95000|
      13|XXXXXXXXXXXXX
-0.76000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.57000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.38000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.19000|
      30|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      27|XXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.19000|
      16|XXXXXXXXXXXXXXXX
 0.38000|
      17|XXXXXXXXXXXXXXXXX
 0.57000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 0.76000|
      13|XXXXXXXXXXXXX
 0.95000|
       5|XXXXX
 1.14000|
       7|XXXXXXX
 1.33000|
       4|XXXX
 1.52000|
       2|XX
 1.71000|
       1|X
 1.90000|
       2|XX
 2.09000|
       1|X
 2.28000|
       1|X
 2.47000|
       1|X
 2.66000|
       0|
 2.85000|
       1|X
 3.04000|
```

(j)

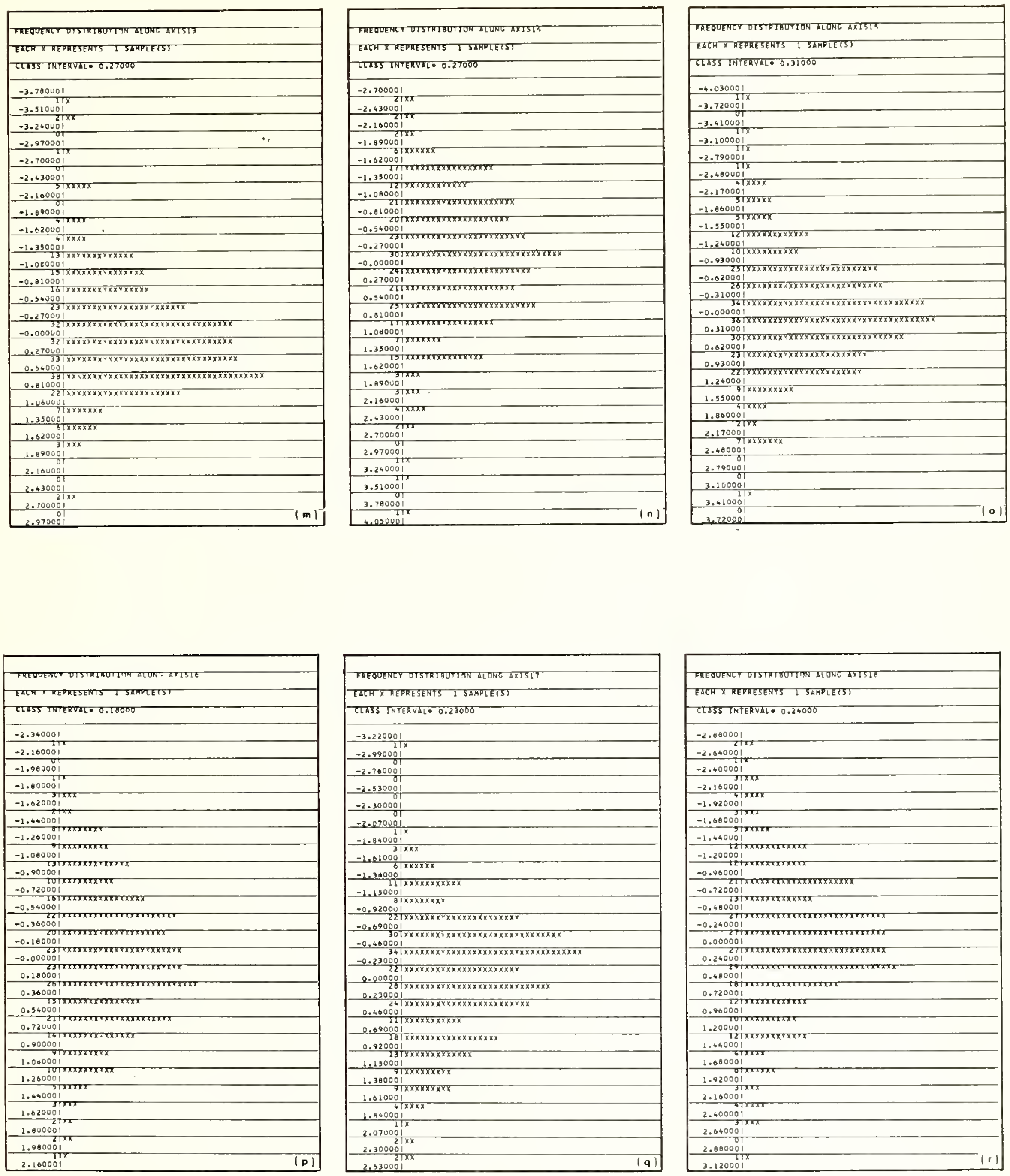
```
FREQUENCY DISTRIBUTION ALONG AXIS11
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.27000

-2.97000|
       1|X
-2.70000|
       0|
-2.43000|
       1|X
-2.16000|
       1|X
-1.89000|
       4|XXXX
-1.62000|
       4|XXXX
-1.35000|
       9|XXXXXXXXX
-1.08000|
      20|XXXXXXXXXXXXXXXXXXXX
-0.81000|
      15|XXXXXXXXXXXXXXX
-0.54000|
      30|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.27000|
      46|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      42|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.27000|
      27|XXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.54000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 0.81000|
      14|XXXXXXXXXXXXXX
 1.08000|
      12|XXXXXXXXXXXX
 1.35000|
       8|XXXXXXXX
 1.62000|
       2|XX
 1.89000|
       0|
 2.16000|
       0|
 2.43000|
       0|
 2.70000|
       0|
 2.97000|
       0|
 3.24000|
       0|
 3.51000|
       1|X
 3.78000|
```

(k)

```
FREQUENCY DISTRIBUTION ALONG AXIS12
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.20000

-3.20000|
       1|X
-3.00000|
       0|
-2.80000|
       0|
-2.60000|
       0|
-2.40000|
       0|
-2.20000|
       1|X
-2.00000|
       1|X
-1.80000|
       0|
-1.60000|
       4|XXXX
-1.40000|
       6|XXXXXX
-1.20000|
      10|XXXXXXXXXX
-1.00000|
      14|XXXXXXXXXXXXXX
-0.80000|
      12|XXXXXXXXXXXX
-0.60000|
      27|XXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.40000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.20000|
      23|XXXXXXXXXXXXXXXXXXXXXXX
 0.00000|
      35|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.20000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.40000|
      25|XXXXXXXXXXXXXXXXXXXXXXXXX
 0.60000|
      13|XXXXXXXXXXXXX
 0.80000|
      17|XXXXXXXXXXXXXXXXX
 1.00000|
      12|XXXXXXXXXXXX
 1.20000|
       4|XXXX
 1.40000|
       2|XX
 1.60000|
       3|XXX
 1.80000|
```

(l)

Figure 32 (continued)



```
FREQUENCY DISTRIBUTION ALONG AXIS13
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.27000

 -3.78000I
         1IX
 -3.51000I
         2IXX
 -3.24000I
         0I
 -2.97000I
         1IX
 -2.70000I
         0I
 -2.43000I
         5IXXXXX
 -2.16000I
         0I
 -1.89000I
         4IXXXX
 -1.62000I
         4IXXXX
 -1.35000I
        13IXXXXXXXXXXXXX
 -1.08000I
        15IXXXXXXXXXXXXXXX
 -0.81000I
        16IXXXXXXXXXXXXXXXX
 -0.54000I
        23IXXXXXXXXXXXXXXXXXXXXXXX
 -0.27000I
        32IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.00000I
        32IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.27000I
        33IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.54000I
        38IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.81000I
        22IXXXXXXXXXXXXXXXXXXXXXX
  1.08000I
         7IXXXXXXX
  1.35000I
         6IXXXXXX
  1.62000I
         3IXXX
  1.89000I
         0I
  2.16000I
         0I
  2.43000I
         2IXX
  2.70000I
         0I
  2.97000I
```

(m)

```
FREQUENCY DISTRIBUTION ALONG AXIS14
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.27000

 -2.70000I
         2IXX
 -2.43000I
         2IXX
 -2.16000I
         2IXX
 -1.89000I
         6IXXXXXX
 -1.62000I
        17IXXXXXXXXXXXXXXXXX
 -1.35000I
        12IXXXXXXXXXXXX
 -1.08000I
        21IXXXXXXXXXXXXXXXXXXXXX
 -0.81000I
        20IXXXXXXXXXXXXXXXXXXXX
 -0.54000I
        23IXXXXXXXXXXXXXXXXXXXXXXX
 -0.27000I
        30IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.00000I
        24IXXXXXXXXXXXXXXXXXXXXXXXX
  0.27000I
        21IXXXXXXXXXXXXXXXXXXXXX
  0.54000I
        25IXXXXXXXXXXXXXXXXXXXXXXXXX
  0.81000I
        17IXXXXXXXXXXXXXXXXX
  1.08000I
         7IXXXXXXX
  1.35000I
        15IXXXXXXXXXXXXXXX
  1.62000I
         3IXXX
  1.89000I
         3IXXX
  2.16000I
         4IXXXX
  2.43000I
         2IXX
  2.70000I
         0I
  2.97000I
         1IX
  3.24000I
         1IX
  3.51000I
         0I
  3.78000I
         1IX
  4.05000I
```

(n)

```
FREQUENCY DISTRIBUTION ALONG AXIS15
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.31000

 -4.03000I
         1IX
 -3.72000I
         0I
 -3.41000I
         1IX
 -3.10000I
         1IX
 -2.79000I
         1IX
 -2.48000I
         4IXXXX
 -2.17000I
         5IXXXXX
 -1.86000I
         5IXXXXX
 -1.55000I
        12IXXXXXXXXXXXX
 -1.24000I
        10IXXXXXXXXXX
 -0.93000I
        25IXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.62000I
        26IXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.31000I
        34IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.00000I
        36IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.31000I
        30IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.62000I
        23IXXXXXXXXXXXXXXXXXXXXXXX
  0.93000I
        22IXXXXXXXXXXXXXXXXXXXXXX
  1.24000I
         9IXXXXXXXXX
  1.55000I
         4IXXXX
  1.86000I
         2IXX
  2.17000I
         7IXXXXXXX
  2.48000I
         0I
  2.79000I
         0I
  3.10000I
         1IX
  3.41000I
         0I
  3.72000I
```

(o)

```
FREQUENCY DISTRIBUTION ALONG AXIS16
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.18000

 -2.34000I
         1IX
 -2.16000I
         0I
 -1.98000I
         1IX
 -1.80000I
         3IXXX
 -1.62000I
         2IXX
 -1.44000I
         8IXXXXXXXX
 -1.26000I
         9IXXXXXXXXX
 -1.08000I
        13IXXXXXXXXXXXXX
 -0.90000I
        10IXXXXXXXXXX
 -0.72000I
        16IXXXXXXXXXXXXXXXX
 -0.54000I
        22IXXXXXXXXXXXXXXXXXXXXXX
 -0.36000I
        20IXXXXXXXXXXXXXXXXXXXX
 -0.18000I
        23IXXXXXXXXXXXXXXXXXXXXXXX
 -0.00000I
        23IXXXXXXXXXXXXXXXXXXXXXXX
  0.18000I
        26IXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.36000I
        15IXXXXXXXXXXXXXXX
  0.54000I
        21IXXXXXXXXXXXXXXXXXXXXX
  0.72000I
        14IXXXXXXXXXXXXXX
  0.90000I
         9IXXXXXXXXX
  1.08000I
        10IXXXXXXXXXX
  1.26000I
         5IXXXXX
  1.44000I
         3IXXX
  1.62000I
         2IXX
  1.80000I
         2IXX
  1.98000I
         1IX
  2.16000I
```

(p)

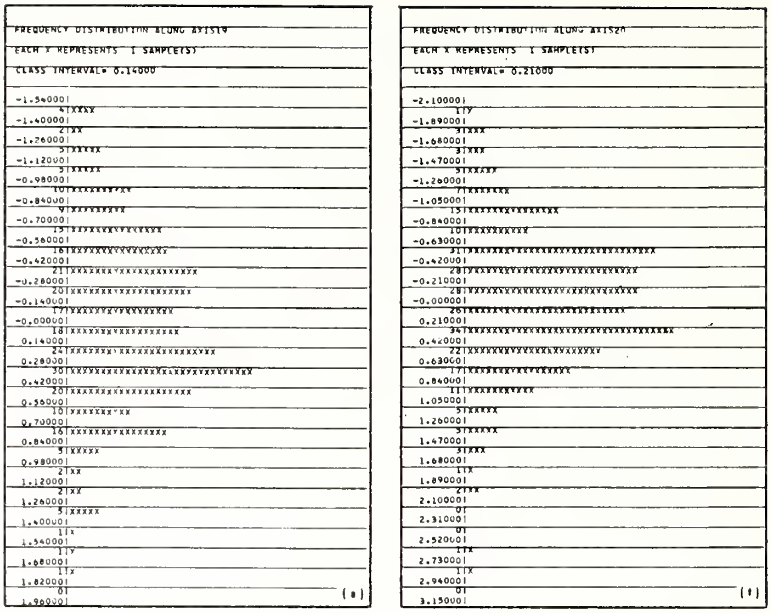
```
FREQUENCY DISTRIBUTION ALONG AXIS17
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.23000

 -3.22000I
         1IX
 -2.99000I
         0I
 -2.76000I
         0I
 -2.53000I
         0I
 -2.30000I
         0I
 -2.07000I
         1IX
 -1.84000I
         3IXXX
 -1.61000I
         6IXXXXXX
 -1.38000I
        11IXXXXXXXXXXX
 -1.15000I
         8IXXXXXXXX
 -0.92000I
        22IXXXXXXXXXXXXXXXXXXXXXX
 -0.69000I
        30IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.46000I
        34IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.23000I
        22IXXXXXXXXXXXXXXXXXXXXXX
  0.00000I
        28IXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.23000I
        24IXXXXXXXXXXXXXXXXXXXXXXXX
  0.46000I
        11IXXXXXXXXXXX
  0.69000I
        18IXXXXXXXXXXXXXXXXXX
  0.92000I
        13IXXXXXXXXXXXXX
  1.15000I
         9IXXXXXXXXX
  1.38000I
         9IXXXXXXXXX
  1.61000I
         4IXXXX
  1.84000I
         1IX
  2.07000I
         2IXX
  2.30000I
         2IXX
  2.53000I
```

(q)

```
FREQUENCY DISTRIBUTION ALONG AXIS18
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.24000

 -2.88000I
         2IXX
 -2.64000I
         1IX
 -2.40000I
         3IXXX
 -2.16000I
         4IXXXX
 -1.92000I
         3IXXX
 -1.68000I
         5IXXXXX
 -1.44000I
        12IXXXXXXXXXXXX
 -1.20000I
        12IXXXXXXXXXXXX
 -0.96000I
        21IXXXXXXXXXXXXXXXXXXXXX
 -0.72000I
        13IXXXXXXXXXXXXX
 -0.48000I
        27IXXXXXXXXXXXXXXXXXXXXXXXXXXX
 -0.24000I
        27IXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.00000I
        27IXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.24000I
        29IXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  0.48000I
        18IXXXXXXXXXXXXXXXXXX
  0.72000I
        12IXXXXXXXXXXXX
  0.96000I
        10IXXXXXXXXXX
  1.20000I
        12IXXXXXXXXXXXX
  1.44000I
         4IXXXX
  1.68000I
         6IXXXXXX
  1.92000I
         3IXXX
  2.16000I
         4IXXXX
  2.40000I
         3IXXX
  2.64000I
         0I
  2.88000I
         1IX
  3.12000I
```

(r)

Figure 32 (continued)



```
FREQUENCY DISTRIBUTION ALONG AXIS19
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.14000
-1.54000|
       4|XXXX
-1.40000|
       2|XX
-1.26000|
       5|XXXXX
-1.12000|
       5|XXXXX
-0.98000|
      10|XXXXXXXXXX
-0.84000|
       9|XXXXXXXXX
-0.70000|
      15|XXXXXXXXXXXXXXX
-0.56000|
      16|XXXXXXXXXXXXXXXX
-0.42000|
      21|XXXXXXXXXXXXXXXXXXXXX
-0.28000|
      20|XXXXXXXXXXXXXXXXXXXX
-0.14000|
      17|XXXXXXXXXXXXXXXXX
-0.00000|
      18|XXXXXXXXXXXXXXXXXX
 0.14000|
      24|XXXXXXXXXXXXXXXXXXXXXXXX
 0.28000|
      30|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.42000|
      20|XXXXXXXXXXXXXXXXXXXX
 0.56000|
      10|XXXXXXXXXX
 0.70000|
      16|XXXXXXXXXXXXXXXX
 0.84000|
       5|XXXXX
 0.98000|
       2|XX
 1.12000|
       2|XX
 1.26000|
       5|XXXXX
 1.40000|
       1|X
 1.54000|
       1|X
 1.68000|
       1|X
 1.82000|
       0|                                        (e)
 1.96000|
```

```
FREQUENCY DISTRIBUTION ALONG AXIS20
EACH X REPRESENTS  1 SAMPLE(S)
CLASS INTERVAL= 0.21000
-2.10000|
       1|X
-1.89000|
       3|XXX
-1.68000|
       3|XXX
-1.47000|
       5|XXXXX
-1.26000|
       7|XXXXXXX
-1.05000|
      15|XXXXXXXXXXXXXXX
-0.84000|
      10|XXXXXXXXXX
-0.63000|
      31|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.42000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.21000|
      28|XXXXXXXXXXXXXXXXXXXXXXXXXXXX
-0.00000|
      26|XXXXXXXXXXXXXXXXXXXXXXXXXX
 0.21000|
      34|XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 0.42000|
      22|XXXXXXXXXXXXXXXXXXXXXX
 0.63000|
      17|XXXXXXXXXXXXXXXXX
 0.84000|
      11|XXXXXXXXXXX
 1.05000|
       5|XXXXX
 1.26000|
       5|XXXXX
 1.47000|
       3|XXX
 1.68000|
       1|X
 1.89000|
       2|XX
 2.10000|
       0|
 2.31000|
       0|
 2.52000|
       1|X
 2.73000|
       1|X
 2.94000|
       0|                                        (f)
 3.15000|
```

Figure 32 (continued)

Table 25 Separation of standardized transformed components along the major axis of the distribution of the Canton Island, U.S.A. and U.K., winds at the 30-mb level during the Julys 1954-1964. The sample size is 244. There are no dimensions in terms of units. The assumption is that the variances are not the same. NORMIX Clustering - John Wolfe Program, 1971.

Logarithm of the likelihood;	chi-square with 4 d.f.;	probability of null hypothesis.
(a) 2 to 1 types 81.046327	(a) 160.43.	(a) 0.00000000

Characteristics of the whole sample (dimensionless).

Mean	0.0000
Std. Dev.	1.0003
Corr.	1.0000

	Type 1 Proportion 0.432		Type 2 Proportion 0.568
Mean	-1.0420		0.7923
Std. Dev.	0.2442		0.5127
Corr.	1.0000		1.0000
Eigen-Values	Vectors	Eigen-Values	Vectors
0.0597	1.0000	0.2628	1.0000

Table 26 Separation of standardized transformed components along the major and minor axis of the distribution of the Canton Island, U.S.A. and U.K., winds at the 30-mb level during the Julys 1954-1964. The sample size is 244. There are no dimensions in terms of units. The assumption is that the variances are not the same. Estimates in Mixed Normal Distribution, Cohen and Falls Program, 1967.

Major Axis

Characteristics of the whole sample (dimensionless) - grouped data.

Mean	-0.000000
Variance	+1.000637
Std. Dev.	1.000252

	Type 1 Proportion 0.493	Type 2 Proportion 0.507
Mean	-0.932133	0.948055
Variance	+0.111975	0.121955
Std. Dev.	+0.334614	0.349221

Minor Axis

Characteristics of the whole sample (dimensionless) - grouped data.

Mean	-0.009180
Variance	+1.002371
Std. Dev.	+1.001184

	Type 1 Proportion 0.998	Type 2 Proportion 0.002
Mean	-0.009617	+7.410603
Variance	+0.875908	-4.123101
Std. Dev.	+0.93589	X.XXXXXX

10. PREDICTION

In the multimodal multivariate case, prediction first should be to the modes or clusters. Once a mixed multivariate data set has been separated into its various homogeneous parts, prediction within the cluster can be made. The characteristics of the group are known and the appropriate regression equation for each homogeneous group can be developed.

In a time series which may be composed of two or more periodic or aperiodic parts, it will be necessary to determine which components are active at the moment. Prediction is then again straightforward.

The problem still remains as to the deterministic regime operating at the moment. This may require prediction into the appropriate cluster and then prediction within the cluster.

The problem of prediction is an important one considered to be beyond the scope of the present paper. It is one to which the authors intend to return and to provide the necessary procedures to forecast to the cluster then within the cluster. Briefly, this has been touched on in section 3 on Discriminant Techniques.

SUMMARY

Clustering techniques to separate mixed distributions of meteorological data are presented. The techniques and data sets are restricted to multimodal multivariate distributions. Dichotomous or polychotomous and non-normal distributions have not been discussed.

The major element considered is wind. Three specific examples include the situations of the land and sea breeze, the quasi-biennial oscillation, and winds in a pass. Three other examples are continental tropospheric winds, a location on the ocean, and a tropical troposphere. One of the examples does include temperature; another includes temperatures, dew points, and heights of pressure surfaces; another includes pressure, temperature, and dew point; while another includes heights of the pressure surfaces and temperatures.

The electronic computer program available to make the separation of mixtures into their homogeneous parts for weather data effectively does the job.

The technique will be useful in establishing weather or climate groups which can be set aside for study and for use in guidance and forecasting.

REFERENCES

Alt, F. B., 1973: Variable control charts for multivariable data. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, 42 pp.

Anderson, Edgar, 1935: The irises of the Gaspe Pennisula. *Bulletin of the American Iris Society*, 59, 2-5.

Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analyses*. John Wiley and Sons, New York, 272-279.

Anderson, T. W., and Rubin, H., 1956: Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 5, 111-150.

Andrews, D. F., Gnanadesikan, R., and Warner, J. L., 1971: Transformations of multivariate data. *Biometrics*, 27, 825-840.

Angell, J. K., and Korshover, J., 1962: The biennial wind and temperature oscillations of the equatorial stratosphere and their possible extension to higher latitudes. *Monthly Weather Review*, 90, 127-132.

Aversen, J. N., and McCabe, G. P., Jr., 1975: Subset selection problems for variances with applications to regression analysis. *Journal of the American Statistical Association*, 70 (349), 166-170.

Baker, Frank B., and Hubert, Lawrence J., 1975: Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, 70 (349), 31-38.

Bannerjee, K. S., and Carr, R. N., 1971: A comment on ridge regression, biased estimates for non-orthogonal problems. *Technometrics*, 15 (4), 895-898.

Barnard, M. M., 1935: The secular variations of skull characters in four series of Egyptian skulls. *Annals of Eugenics*, London, 6, 352-371.

Bartlett, M. S., 1950: Tests of significance in factor analysis. *British Journal of Psychology*, 3, 77-85.

Bartlett, M. S., 1951: A further test note on tests of significance in factor analysis. *British Journal of Psychology*, Statistical Section, 4, 1-2.

Bartlett, M. S., 1953: Factor analysis in psychology as a statistician sees it. *Uppsala Symposium on Psychological Factor Analysis*, *Uppsala*, *Almqvist Wiksell*, 23-24.

Berlage, H. P., 1956: The Southern Oscillation, a 2-3 year fundamental oscillation of worldwide significance. *IUGG 10th General Assembly*, *Rome*, *1954*, *Scientific Proceedings International Association of Meteorology*, London, 336-345.

Boehm, Albert, 1974: Modeled conditional climatology using multiple predictors. Paper presented at the *American Meteorological Society, Climatology Conference and Workshop at Asheville, North Carolina, on October 8, 1974*, 36 pp.

Bose, R. C., and Roy, S. N., 1938: The distribution of the Studentized D^2 Statistic. *Sankhyä*, 4 (Part 1), 19-38.

Box, G. E. P., and Tidwell, P. W., 1962: Transformation of the independent variables. *Technometrics*, 4, 531-550.

Brooks, C. E. P., and Carruthers, N., 1950: Upper winds over the world. Great Britain Meteorological Office, Air Ministry, *Geophysical Memoirs No. 85*, Her Majesty's Stationery Office, London, 150 pp.

Brown, George W., 1947: Discriminant functions. *Annals of Mathematical Statistics*, 18, 514-528.

Burrau, C., 1934: The half-invariants of two typical laws of errors with an application to the problem of dissecting a frequency curve into components. *Skandinavisk Aktuarietidskrift*, 17, 1-6.

Cattell, Raymond B., 1952: *Factor Analysis*. Harper and Brothers, New York, 462 pp.

Charlier, C. V. L., 1906: Researches into the theory of probability. Meddelanden fran Lunds Astronomiska Observatorium, Lunds Universitets Arsskrift, N.S., Asd2, Band 1, nr. 5, Series 2, Band 1, No. 4.

Charlier, C. V. L., and Wicksell, S. D., 1924: On the dissection of frequency functions. Arkiv för Matematik, Astronomi och Fysik, Bd. 18, No. 6.

Clayton, H. H., 1885: A lately discovered meteorological cycle. *American Meteorological Journal*, 1, pp. 130 and 528.

Cohen, A. C., 1965: Estimation in mixtures of two normal distributions. *University of Georgia Technical Report*, No. 13, 38 pp.

Cohen, A. C., and Falls, L. W., 1967: Estimation in mixed frequency distributions. *NASA TN D-4033*, National Aeronautics and Space Administration, George C. Marshall Space Flight Center, Huntsville, Alabama, 90 pp.

Cox, Gertrude M., and Martin, William P., 1937: Use of a discriminant function for differentiating soils with different azotobacter populations. Iowa State College, *Journal of Science*, 11 (3), 323-331.

Crutcher, Harold L., 1957: On the standard vector deviation wind rose. *Journal of Meteorology*, 14 (1), 28-33.

Crutcher, Harold L., 1958: Upper wind distribution statistical parameter estimates. *Technical Paper No. 34*, U.S. Department of Commerce, Weather Bureau, Asheville, North Carolina, 59 pp.

Crutcher, Harold L., 1959: Upper wind statistics of the Northern Hemisphere. Issued by the Office of the Chief of Naval Operations, *NAVAER 50-1C-535*, Vol. I, 35 pp. with 168 charts and 54 overlays, Vol. II, 35 pp. with 168 charts and 54 overlays, and Vol. III, 23 pp. with 56 charts.

Crutcher, Harold L., 1960: Statistical grouping of climates and the statistical discrimination among climate groups. A dissertation in the Department of Meteorology and Oceanography submitted to the faculty of the Graduate School of Arts and Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University, University Microfilms, Ann Arbor, Michigan, 462 pp.

Crutcher, Harold L., 1966: Gaussian filters for quality control of vector data. (Unpublished), 42 pp.

Crutcher, Harold L., and Clutter, Jerome L., 1962: Heterogeneous wind distributions. North Carolina Operations Analysis Standby Unit, Chapel Hill, North Carolina, (unpublished manuscript), 56 pp. and appendices.

Crutcher, H. L., Wagner, A. C., and Arnett, J. S., 1966: Components of the 1000-mb winds of the Northern Hemisphere. Issued by the Office of the Chief of Naval Operations, *NAVAIR 50-1C-51*, 4 pp. with 74 charts.

Crutcher, Harold L., and Falls, Lee W., 1976: Multivariate normality. *NASA TN D-8226*, National Aeronautics and Space Administration, George C. Marshall Space Flight Center, Huntsville, Alabama, 273 pp.

Davies, Owen L., 1956: *The Design and Analysis of Industrial Experiments*. Hafner Publishing Company, New York, 637 pp.

Dempster, A. P., 1973: *Aspects of the Multinomial Logit Model*. Edited by P. R. Krishnaiah, Academic Press, Inc., New York, 129-142.

Dickey, Woodrow W., and Wing, Robert, N., 1963: The unusual: Arctic air into the Pacific Northwest. *Weatherwise*, 16 (6), 259-263.

Draper, N. R., 1963: Ridge analysis of response surfaces. *Technometrics*, 5 (4), 469-479.

Duda, Richard O., and Hart, Peter E., 1973: *Pattern Classification and Scene Analysis*. John Wiley and Sons, Inc., New York, 481 pp.

Ebdon, R. A., 1960: Notes on the wind flow of 50 mb in tropical and subtropical regions in January 1957 and January 1958. *Quarterly Journal of the Royal Meteorological Society*, 86, 540-542.

Essenwanger, Oskar, 1954: Neue Methode der Zerlegung von Haufigkeitsverteilungen in Gausische Normalkurven und ihre Anwendung in der Meteorologie. Berichte des Deutschen Wetterdienstes Nr. 10, DK 551.501.45, Bad Kissengen, 1-11.

Finney, D. J., 1952: *Probit Analysis, A Statistical Treatment of the Sigmoid Response Curve*, Second Edition, Cambridge University Press, New York, 318 pp.

Fisher, R. A., 1928: The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society*, Vol. A121, 654 pp.

Fisher, R. A., 1936: The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 (2), 179-188.

Fix, E., and Hodges, J. L., 1952: Discriminatory analysis - non-parametric discrimination: consistency properties. *Report No. 4*, USAF School of Aviation Medicine, Randolph Field, Texas, 21 pp.

Friedman, H. P., and Rubin J., 1967: On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62 (320), 1159-1178.

Fruchter, B., 1954: *Introduction to Factor Analysis*. D. Van Nostrand Co., New York, 280 pp.

Galton, Francis, 1889: *Natural Inheritance*. MacMillan and Company, London, 259 pp.

Girschik, M. A., 1936: Principal components, *Journal of the American Statistical Association*, 31, 519-528.

Good, I. J., 1965: Categorization of classification. *Mathematics and Computer Science in Biology and Medicine*, Her Majesty's Stationery Office, London, 115-125.

Graybill, Franklin A., 1961: *An Introduction to Linear Statistical Models, Vol. 1*. McGraw-Hill, New York, 463 pp.

Graystone, P., 1959: Tropical meteorology. Meteorological Office Discussion, *Meteorology Magazine*, 88, 113-119.

Gupta, S. S., and Sobel, M., 1962: On selecting a subset containing the population with the smallest variance. *Biometrika*, 49 (3, 6), 495-507.

Hald, A., 1952: *Statistical Theory With Engineering Applications*, John Wiley and Sons, Inc., New York, 602-606.

Harman, Harry H., 1967: *Modern Factor Analysis*, The University of Chicago Press, Chicago, Illinois, Library of Congress Card 67-20572, 1967, 474 pp.

Hartigan, J. A., 1975: *Clustering Algorithms*, John Wiley and Sons, Inc., New York, 351 pp.

Hartley, Herman O., 1959: Two-component mixtures of circularly symmetric wind roses. *Preliminary Technical Report No. 1*, Statistical Laboratory, Iowa State College, Ames, Iowa, 25 pp.

Heastie, H., and Stephenson, P. M., 1960: Upper winds over the world, Parts 1 and 2. Great Britain Meteorological Office, Air Ministry, *Geophysical Memoirs No. 103*, Third Number, Vol. XIII, M.O. 631c, Her Majesty's Stationery Office, London, 217 pp.

Hoerl, A. E., 1962: Applications of ridge analysis to regression problems. *Chemical Engineering Progress*, 58, 54-59.

Hoerl, A. E., and Kennard, R. W., 1970(a): Ridge regression: biased estimation for non-orthogonal problems. *Technometrics*, 12 (1), 55-67.

Hoerl, A. E., and Kennard, R. W., 1970(b): Ridge regression: application to non-orthogonal problems. *Technometrics*, 12 (1), 68-82.

Hotelling, Harold, 1931: The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2, 360-378.

Hotelling, Harold, 1933: Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, pp. 417-441 and 498-520.

Hotelling, Harold, 1936(a): Relations between two sets of variates. *Biometrika*, 28, 321-377.

Hotelling, Harold, 1936(b): Simplified calculation of principal components. *Psychometrika*, 1, 27-35.

Hotelling, Harold, 1954: Multivariate analysis. In *Statistics and Mathematics in Biology*, (edited by O. Kempthorne et al.), The Iowa State College Press, Ames, Iowa, 67-80.

Hotelling, Harold, 1957: The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, 10, 69-79.

Hsu, P. L., 1938: Notes on Hotelling's generalized T. *Annals of Mathematical Statistics*, 9, 231-243.

Kendall, M. G., and Smith, B. Babington, 1950: Factor analysis. *Journal of the Royal Statistical Society*, B12, 60-94.

Kendall, Maurice G., and Stuart, Alan, 1968: *The Advanced Theory of Statistics*, Volume 3, Design and Analysis, And Time-Series, Second Edition, Hafner Publishing Company, New York, 557 pp.

Kullback, Solomon, 1968: *Information Theory and Statistics*. Dover Publications, Inc., New York, 199 pp.

Landsberg, H. E., Mitchell, J. M., Crutcher, Harold L., and Quinlan, F. T., 1963: Surface signs of the biennial atmospheric pulse. *Monthly Weather Review*, 91, Contributions in memory of Harry Wexler, U.S. Department of Commerce, Weather Bureau, 549-556.

Lawley, D. N., and Maxwell, A. E., 1963: *Factor Analysis as a Statistical Method*. Butterworths, London, 117 pp.

Ludlum, David, 1969: The snowfall season of 1967-68. *Weatherwise*, 22 (1), 26-31.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium of Mathematical Statistics and Probability*, University of California Press, Vol. 1, 281-297.

McCreary, F. E., Jr., 1959: Stratospheric winds over the tropical Pacific Ocean. Presented at the *Conference on Stratospheric Meteorology*, *American Meteorological Society*, *Minneapolis*, *Minnesota*, *September*, 10 pp.

Mahalanobis, P. C., 1936: On the generalized distance in statistics. *Proceedings of the National Institute of Science*, *India*, 12, 45-55.

Miller, Robert G., 1962: Statistical prediction by discriminant analysis. *Meteorological Monographs*, 4 (25), American Meteorological Society, Boston, 54 pp.

Moore, P. G., 1957: Transformations to normality using fractional powers of the variable. *Journal of the American Statistical Association*, 52 (278), 237-246.

Mulaik, Stanley A., 1972: *The Foundation of Factor Analysis*, McGraw-Hill Book Company, New York, 453 pp.

Myers, Raymond H., 1971: *Response Surface Methodology*. Allyn and Bacon, Inc., 246 pp.

Newell, Reginald E., Kidson, J. W., Vincent, Dayton G., and Boer, George J., 1974: *The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes*, MIT Press, Cambridge, Massachusetts, 2, 371 pp.

Pearson, Karl, 1894: Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, Series A, 185, 71-110 + 5 plates.

Pearson, Karl, 1901: On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6, 559-572.

Pearson, Karl, and Lee, Alice, 1903: On the laws of inheritance of man. *Biometrika*, 2, 357-462.

Phillips, Earl L., 1969: Washington-December 1968, special weather summary, climatological data. Environmental Data Service, Environmental Science Services Administration, U.S. Department of Commerce, 72 (12), 198 pp.

Rao, C. R., 1950: A note on the distribution of D^2-D^2 and some computational aspects of $p + q$ the D statistic and discriminant function. *Sankhyä*, 10, 257-268.

Rao, C. R., 1952: *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, Inc., New York, 390 pp.

Reed, R. J., 1960: The structure and dynamics of the 26-month oscillation. Presented at the *40th Anniversary Meeting of the American Meteorological Society*, *Boston*, 393-402.

Reed, R. J., and Rogers, D. G., 1962: The circulation of the tropical stratosphere in the years 1954-1960. *Journal of the Atmospheric Sciences*, 19 (2), 127-135.

Reed, R. J., Campbell, W. J., Rasmussen, L. A., and Rogers, D. J., 1961: Evidence of a downward-propagating, annual wind reversal in the equatorial stratosphere. *Journal of Geophysical Research*, 66, 813-818.

Roy, S. N., 1939: p-statistics of some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyä*, 4, 381-396.

Roy, S. N., 1942(a): The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis. *Sankhyä*, 6, 15-34.

Roy, S. N., 1942(b): Analysis of variance for multivariate normal populations, the sampling distributions of the requisite p-statistics on the null and non-null hypothesis. *Sankhyä*, 6, 35-40.

Schneider-Carius, K., and Essenwanger, O., 1955: Eigentümlichkeiten der Niederschlagsverhältnisse im Norden und Süden der Schweizer Alpen, dargestellt durch die Niederschlagswahrscheinlichkeit von Basel, St. Gotthard und Lugano. *Archiv für Meteorologie, Geophysik und Bioklimatologie*, Serie B, Allegemeine und Biologische Klimatologie, Band 7, 1. Heft, 32-48.

Smith, C. A. B., 1947: Some examples of discrimination. *Annals of Eugenics*, 14, 272-282.

Snedecor, George W., and Cochran, William G., 1967: *Statistical Methods - Applied to Experiments in Agriculture and Biology*. Iowa State College Collegiate Press, Ames, Iowa, 534 pp.

Sokal, Robert R., and Sneath, P. H. A., 1963: *Principles of Numerical Taxomony*. W. H. Freeman and Company, San Francisco, Library of Congress Card Number 63-13914, 359 pp.

Spearman, C., 1904: General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Spearman, C., 1926: *The Abilities of Man*. MacMillan Company, New York, 416 pp.

Strömgren, B., 1934: Tables and diagrams for dissecting a frequency curve into components by the half-invariant method. Skandinavisk Aktuarietidskrift, 17, 7-54.

Student (W. S. Gosset), 1925: New tables for testing the significance of observations. Metrometrics, 5 (3), 105-120.

Tatsuoka, M. M., 1971: Multivariate Analysis. John Wiley and Sons, Inc., New York, 300 pp.

Thurstone, L. L., 1947: Multiple Factor Analysis. The University of Chicago Press, Chicago, Illinois, 535 pp.

Tryon, Robert C., 1959: Domain sampling formulation of cluster and factor analysis. Psychometrika, 24, 113-135.

Tryon, Robert C., and Bailey, Daniel E., 1970: Cluster Analysis. McGraw-Hill Book Company, New York, 347 pp.

Tucker, G. B., 1960: Upper winds over the world, Part 3. Great Britain Meteorological Office, Air Ministry, Geophysical Memoirs No. 105, Fifth Number, Volume XIII, M.O. 631e, Her Majesty's Stationery Office, London, 101 pp.

Tukey, John W., 1967: On the comparative anatomy of transformations. Annals of Mathematical Statistics, 28, 602-632.

Veryard, R. G., 1960: Contribution to "Discussion of synoptic and dynamic aspects of climatic changes," by E. B. Kraus. Quarterly Journal of the Royal Meteorological Society, 86, 569-570.

Veryard, R. G., and Ebdon, R. A., 1961(a): Fluctuations in equatorial stratospheric winds. Nature, 189, 791-793.

Veryard, R. G., and Ebdon, R. A., 1961(b): Fluctuations in tropical stratospheric winds. Meteorological Magazine, 90 (1066), 125-143.

Veryard, R. G., and Ebdon, R. A., 1963: The 26-month tropical stratospheric wind oscillation and possible causes. Meteorologische Abhandlungen, Freien Universität Berlin, 36, 225-244.

Ward, Joe H., Jr., Hall, K., and Buchhorn, J., 1967: PERSUB Reference Manual PRL-TR-67-3 (II). Personnel Research Laboratory, Lackland Air Force Base, Texas, 60 pp.

Wilks, S. S., 1932: Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

Wolfe, John H., 1965: A computer program for the maximum likelihood analysis of types. Technical Bulletin 65-15, U.S. Naval Personnel Research Activity, San Diego, (Defense Documentation Center, AD 620 020), 52 pp.

Wolfe, John H., 1967(a): NORMIX cluster and pattern analysis of normal mixtures. FOCUS-G7-USNP-NORMIX, FORTRAN COOP-COOP Monitor, U.S. Naval Personnel Research Activity, San Diego, 82 pp.

Wolfe, John H., 1967(b): NORMIX: computational methods for estimating the parameters of multivariate normal mixtures of distributions. Research Memorandum SRM-2, U.S. Naval Personnel Research Activity, San Diego, (Defense Documentation Center, AD 656 588), 30 pp.

Wolfe, John H., 1968: NORMAP: cluster and pattern analysis of normal mixtures with common covariances. FORTRAN 63 COOP-COOP MONITOR-1604, U.S. Naval Personnel Research Activity, San Diego, 88 pp.

Wolfe, John H., 1970: Pattern clustering by multivariate analysis. Multivariate Behavioral Research, 5, 329-350.

Wolfe, John H., 1971(a): A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical Bulletin STB-72-2, Naval Personnel and Training Research Laboratory, San Diego, (Defense Documentation Center, AD 731 039), 7 pp.

Wolfe, John H., 1971(b): NORMIX 360 computer program. Research Memorandum SRM 72-4, Naval Personnel and Training Research Laboratory, San Diego, 125 pp.

AUTHOR INDEX

☆U.S. GOVERNMENT PRINTING OFFICE: 1977-240-848/117

(Continued from inside front cover)

EDS 16 NGSDC 1 - Data Description and Quality Assessment of Ionospheric Electron Density Profiles for ARPA Modeling Project. Raymond O. Conkright, in press, 1976.

EDS 17 GATE Convection Subprogram Data Center: Analysis of Ship Surface Meteorological Data Obtained During GATE Intercomparison Periods. Fredric A. Godshall, Ward R. Seguin, and Paul Sabol, October 1976.

EDS 18 GATE Convection Subprogram Data Center: Shipboard Precipitation Data. Ward R. Seguin and Paul Sabol, November 1976.

NOAA--S/T 76-1912