AGROCLIMATIC CLASSIFICATION: NUMERICAL-TAXONOMIC PROCEDURES-A REVIEW[1]

S. JEEVANANDA REDDY[2]

ABSTRACT - The paper catalogues the procedures and steps involved in agroclimatic classification. These vary from conventional descriptive methods to modern computer-based numerical techniques. There are three mutually independent numerical classification techniques, namely Ordination, Cluster analysis, and Minimum spanning tree; and under each technique there are several forms of grouping techniques existing. The choice of numerical classification procedure differs with the type of data set. In the case of numerical continuous data sets with both positive and negative values, the simple and least controversial procedures are unweighted pair group method (UPGMA) and weighted pair group method (WPGMA) under clustering techniques with similarity measure obtained either from Gower metric or standardized Euclidean metric. Where the number of attributes are large, these could be reduced to fewer new attributes defined by the principal components or coordinates by ordination technique. The first few components or coordinates explain the maximum variance in the data matrix. These revised attributes are less affected by noise in the data set. It is possible to check misclassifications using minimum spanning tree.

Index terms: graphical classification, ordination, cluster analysis, similarity measures.

CLASSIFICAÇÃO AGROCLIMÁTICA: MÉTODOS TAXINÔMICAS-UMA REVISÃO

RESUMO - Este trabalho classifica as sequências e procedimentos utilizados em classificação agroclimática. Estes variam de métodos convencionais descritivos a modernas técnicas numéricas baseadas em computador. Há três técnicas de classificação numérica mutuamente independente, chamada de ordenação, análises de clustes e diagramas de distância mínima; e sobre cada técnica há diversas formas de agrupamento das técnicas existentes. A escolha do tipo de classificação numérica difere com o tipo do conjunto de dados. No caso do conjunto de dados numéricos contínuos com valores positivos e negativos, os procedimentos simples e menos contestáveis são o método da média aritmética (UPGMA) e o método da média ponderada (WPGMA) sob técnicas agrupadas com medidas semelhantes obtidas das medidas de Gower ou das medidas padronizadas Euclidianas. Onde o número de características são grandes, essas poderiam ser reduzidas para poucos novos atributos definidos pelos componentes principais ou coordenados por técnicos de ordenação. Os primeiros poucos componentes ou coordenadores explicam a máxima variância na matriz dos dados. Estas características revisadas são menos afetadas por equívoco no conjunto de dados. É possível testar classificações equivocadas usando-se diagramas de distância mínima.

Termos para indexação: classificação gráfica, chamada de ordenação, análise de clustes, medidas semelhantes.

INTRODUCTION

Climate and its inherent processes form a continuum varying in time and space. Within the rather wide range of atmospheric conditions an infinite variety of combinations can appear. It is, therefore, natural to attempt grouping of kindred climates to obtain a classification that will permit the establishment of regional boundaries between areas of uniform climatic conditions. To buildup climatic categories is by no means an easy task. The least that can be achieved is a classification of climate for specific purposes rather than a climatic taxonomy comparable with that of plants. In each case, a specific set of limiting conditions will govern. Hence, the climatic classification of a place will change with the objective towards which the classification is directed.

The objective of the wider study of which this paper is only one component is to identify the semi-arid tropics and to divide these into agronomically relevant homogeneous zones that facilitate the transfer of location-specific dryland technology. Traditional crops, varieties and cropping systems often do not make full and efficient use of available soil and water resources. New techniques of resource management which more effectively con-

[1] Accepted for publication on March 19, 1983.

[2] Consultant (Agroclimatology, Centro de Pesquisa Agropecuária do Trópico Semi-Árido (CPATSA)/ EMBRAPA/IICA, Caixa Postal 23, CEP 56300 - Petrolina, PE - Brazil.

serve and utilize the rainfall and the soil are needed together with new crop production systems which increase productivity and minimise instability. Hundreds of experimental stations throughout the semi-arid tropics are involved in research to increase efficiency of food production. While the research output from a single station may not be large, the combined output of all of them must be considerable. It is also likely that the research results of any given station are relevant not only to immediately adjacent areas, but to widely dispersed regions in the world having similar physical environment. This involves the establishment of guiding parameters for the transfer of technology in terms of physical environmental characters to identify homoclimes or classification into zones of comparable climates.

Climatic classification procedures range from traditional descriptive (Köppen 1936, Thornthwaite 1948, Troll 1965, Cocheme & Franquim 1967, Hargreaves 1971, Papadakis 1975, Reddy, Prelo c), to modern computer based numerical techniques (Sokal & Sneath 1963, Moore & Russell 1967, Cormack 1971, Sneath & Sokal 1973). The entire range can be found in use for climatic as well as in soil, biological, ecological and geological classifications (Harbaugh & Merriam 1968, Arkley 1976, Nix 1975, Russel & Moore 1976, Austin & Nix 1978, Austin & Yapp 1978, Russel 1978). Sokal (1974) presented a classical treatise on purpose, principles, progress and prospects of classification.

The applicability of numerical taxonomic techniques in global climatic or bioclimatic or agroclimatic studies is not well known. Conventional descriptive methods utilize few attributes, while areas are grouped using arbitrary class intervals that can be presented relative to geo-coordinates as a continuum. Where many attributes are considered, numerical techniques confer advantages. Each location is placed in context relative to all others. The choice of numerical classification procedures differ with the type of data set. There are a number of mutually independent numerical classification techniques and under each technique there are several forms of grouping techniques existing. Any classification procedure involves a number of steps or strategies, from data collection through to interpretation of results. A comprehensive flow chart of these steps with alternative strategies and/or options are depicted in Fig. 1. Simplifying, the basic steps are:

i) identification of available raw data;
ii) derivation of attributes that define a particular character of interest;
iii) computation of similarity matrices, which integrate characters into a single entity;
iv) grouping or classification of the locations using these attributes of similarity matrices; and
v) interpretation of final results.

In this paper, an attempt is made to catalogue and discuss these different methods of classification as they apply to climate and to identify similarity metric that integrate the attributes of numerical, continuous data sets.

DATA MATRIX

The first and major item in classification is to identify available data sets. There are two problems associated with data collection, namely availability and accuracy. There are several forms of attributes namely binary, numerical etc., but the present discussion is restricted to numerical and continuous data sets only. The primary raw data set may be comprised of observed parameters like rainfall or temperature or derived parameters like potential evapotranspiration or radiation etc.

The second step involves the estimation of attributes from the raw data set. Choice of attributes used in the analysis is affected both by the purpose of the analysis and by the availability of data. Classifications are attribute dependent and therefore the choice of attributes will largely affect the classification obtained. According to Arkley (1976), to be both comprehensive and most effective, the differentiating characters or variable criteria used to form classes should contain the maximum possible information; the choice of attributes to be included in the classification should be such that the number of parameters are large and the general kinds of parameters included are well represented. The inclusion of large numbers of logically related properties should be avoided as they tend to create an inadvertent extra weight

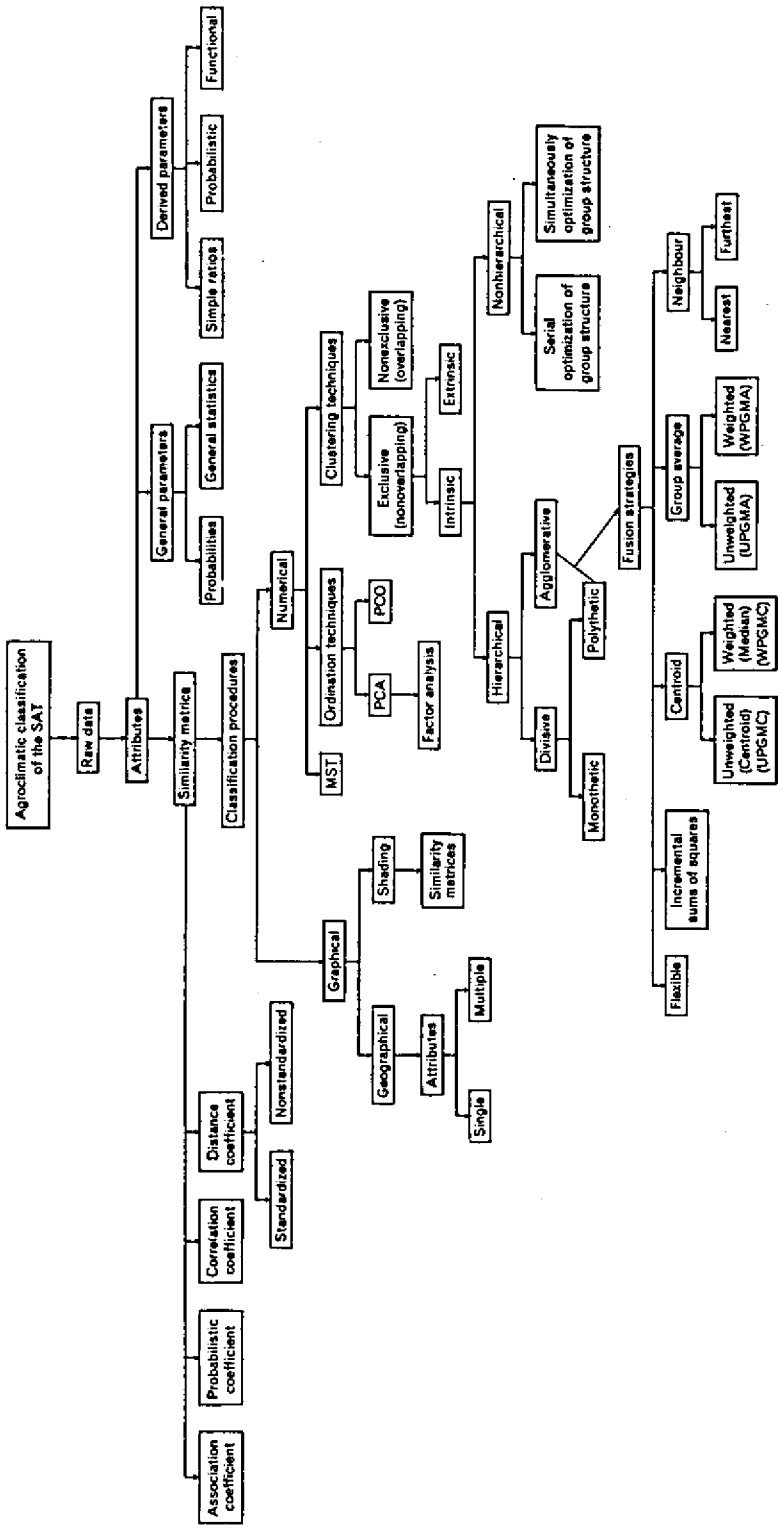


FIG. 1. Flow Chart of Agroclimatic Classification.

to such a group of properties in the classification.

Two types of attributes can be envisaged, namely, general or basic (commoly used) and derived (not so commonly used) (Fig. 1). Basic attributes which are commonly used are of two types: (i) statistical parameters, such as mean annual rainfall, mean monthly temperature, coefficient of variation (C.V.) of rainfall; (ii) probabilities, such as the probatibility of obtaining certain rainfall during specified or fixed amount probabilities (Robertson 1976), the rainfall expected at certain probability levels or fixed probabilities estimated by using incomplete gamma analysis (Hargreaves 1971). Both of these two types of basic attributes are generally derived by standard statistical procedures. Derived attributes represent those developed from concepts which vary according to the purpose of the study. These can be divided into three classes: (i) simple ratios such as the ratios of rainfall to potential evapotranspiration (Hargreaves 1971); (ii) probabilistic parameters such as the probabilities of derived attributes like mean growing season, wet and dry spells during the season, and (iii) functionally derived parameters. If the different derived parameters of basic attributes are interrelated, their relationship is first established. Then, using this established function a new attribute can be derived. This new attribute demonstrates the particular characteristic behaviour of that environment relative to others.

Table 1 presents a sample of data matrix representing 11 Indian locations, each with 11 agroclimatic attributes. At the bottom of this matrix is also presented the mean, standard deviation (hereafter referred as s.d.) and range of each attribute over these locations. Among these eleven attributes eight (δ, $\overline{G}$, C, $\overline{W}$, α, $\overline{D}$, β and A) are derived attributes (Reddy, Prelo a) and the remaining three (G', W', and D') are derived through a functional relationship (Reddy, Preloc). One can qualitatively distinguish two groups in Table 1, namely (i) locations 1 to 4 and (ii) locations 5 to 11. In group (i), location 1 is closer to 3; while 2 and 4 show anomalies with respect to certain attributes. In group (ii), 6 is closer to 7; and 5 is closer to 6-7 and 9 is closer to 5-7; 10 is equidistant from 8 and 11. It appears, however, that 10 is closer to 8 compared to 11 in the majority of the attributes.

SIMILARITY MEASURES

For better representation of a location, it may be important to use more items of information (attributes). The complexity of dealing with more than two attributes can be simplified by attribute integration using standard mathematical functions. Ideally, these produce summary coefficients representative of locational differences. The literature is abundant with such measures. Sneath & Sokal (1973) grouped these under four types, namely probability coefficients; association or matching coefficients); correlation coefficients and distance coefficients (or measure of distance or dissimilarity measure). The first two are not used with continuous (numerical) data but are commonly used with respect to binary or qualitative data, Association and correlation coefficients can usually be related to distances. The distance coefficients and correlation coefficient along with their geometric representation are presented below.

DISTANCE COEFFICIENTS

Distance coefficients are of two types: non-standardized (e.g. Euclidean metric, Mean character distance (MDC) and standardized (e.g. Canberra metric, Gower metric).

Non-standardized distances

Several distance coefficients have been proposed as measures of inter-individual relationships (Sneath & Sokal 1973). Coefficients chosen to represent the relationship between individuals are calculated for all pairs of individuals from the original data matrix. The choice of coefficient requires a knowledge of their relative merits and the kinds of taxonomic information produced. A geometric model is helpful in understanding the meaning of similarity coefficients. Individuals to be studied are thought of as points lying in a multidimensional space, the axes of which correspond to attributes. Let X_{hk} represent the data matrix with k attributes for h locations.

TABLE 1. Data matrix representing 11 locations with 11 attributes.

S.Nº	Location	Attributes*										
		δ	$\bar{G}$	C	$\bar{W}$	α	$\bar{D}$	β	G'	W'	D'	A
1	Indore	1.3	16.4	19	7.0	2.2	6.0	2.1	2.8	2.3	1.3	00
2	Ranchi	2.9	16.4	23	7.5	3.4	3.7	1.8	0.7	2,8	-1.0	00
3	Mahboobnagar	2.5	16.6	30	5.8	2.7	6.0	2.6	-2.1	1.1	1.3	05
4	Vishakhapatnam	4.4	16.7	50	5.3	3.3	7.1	4.2	-6.4	0.6	2.4	14
5	Hyderabad	2.9	12.9	45	4.2	2.5	5.0	2.5	-1.8	-0.5	0.3	13
6	Sholapur	4.0	11.3	57	3.6	2.0	5.1	3.0	-1.8	-1.1	0.4	24
7	Ongole	5.6	11.2	58	3.7	2.2	6.0	3.2	-1.8	-1.0	1.3	24
8	Ajmer	1.5	7.6	67	3.6	1.8	3.7	2.1	0.9	-1.1	-1.1	30
9	Chittoor	5.0	8.9	92	3.6	3.1	4.3	3.7	-2.0	-1.1	-0.4	44
10	Anantapur	4.8	5.2	104	2.7	1.7	3.7	2.5	1.2	-2.0	-1.0	52
11	Hissar	5.9	2.0	170	2.1	1.4	3.0	1.4	2.8	-2.6	-1.7	74
Mean		3.7	11.4	65	4.5	2.4	4.9	2.6	-0.7	-0.2	0.2	25
S.D.		1.5	3.3	42	1.6	0.6	1.2	0.8	2.5	1.6	1.2	19.5
Range		4.6	14.6	151	5.4	2.0	4.1	2.8	9.2	5.4	4.1	74

* δ = Standard deviation of Commencement of sowing rains, weeks

$\bar{G}$ = Mean effective rainy period, weeks

C = Coefficient of variation of G, %

$\bar{W}$ & $\bar{D}$ = Mean number of wet and dry weeks within G, weeks

α & β = Standard deviation of wet and dry weeks, weeks

G' = $\bar{G}$-G", G" is derived through a functional relation ($\bar{G}$ vs C) weeks

W' = $\bar{W}$-W", W" is derived through a functional relation ($\bar{G}$ vs $\bar{W}$), weeks

D' = $\bar{D}$-D", D" is derived through a functional relation (G vs $\bar{D}$), weeks

A = Percentage crop failure years or risk in crop production.

Fig. 2 presents a geometrical representation of locations A and B space defined by two axes. For simplicity it is assumed that each attribute is an orthogonal coordinate amenable to simple Pythagorean geometry. From the trigonometric relationship with ABC representing a right angled triangle, the distance between two locations (AB) is given as:

$$AB = \Delta = (BC^2 + CA^2)^{1/2} = ((x_{21} - x_{11})^2 + (x_{12} - x_{22})^2)^{1/2} = (\sum_{k=1}^{p} (x_{ik} - x_{jk})^2)^{1/2}$$

The taxonomic distance d_{ij} is related to the geometric distance by:

$$d_{ij} = (\Delta^2_{ij}/p)^{1/2}$$

This is also known as Euclidean or Pythagorean distance (Table 2, eq. 1 - refer to Table 2 only, hereafter). This represents the square root of the average of the squared differences between individuals over all attributes (p). d_{ij} measures the dissimilarity between the individuals i and j. Such a measure is sensitive to the magnitudes of the difference between the attributes; larger differences will contribute a relatively greater amount to the sum of squares of the differences. To prevent excessive dominance by attributes with large differences, prior data standardization is usually required. MCD is also known as Manhattan or City block metric (Cain & Harrison 1958) representing absolute average difference between individuals (eq. 2).

The above two measures could be standardized either by dividing each difference by the standard deviation of the locations $(s.d.)_k$ - eqs. 6 & 9 - or

by the range r_k of the respective attributes (k) - eqs. 8 & 10. The standardized dissimilarity metric can be expressed as similarity metric by $S_{ij} = 1 - d_{ij}$. Squared standardized Euclidean distance is known as Mahalanobis generalized distance. If the standardization is made using standard deviation then the squared Euclidean distance is also known as Burr standardized squared Euclidean distance (eq. 7).

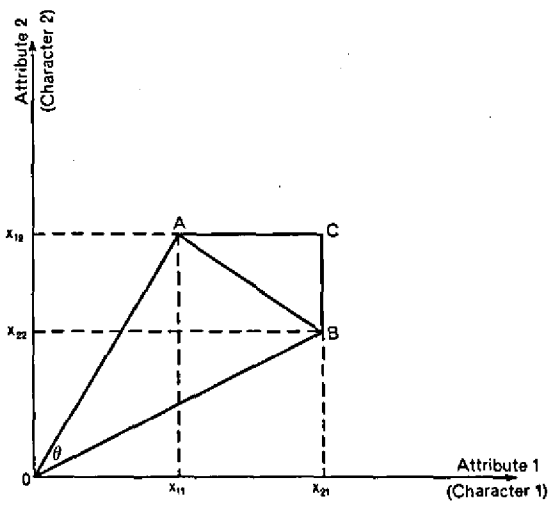


FIG. 2. Geometric presentation of similarity measures.

In the above two methods, the squared or absolute difference specifies the importance of magnitude rather than to the sign of the difference. However, the resultant magnitude in both cases differs substantially because it represents a second order difference in the former, and first order difference in the latter.

Standardized distances

The Canberra metric (Lance & Williams 1967b) is defined as the average of the ratio of absolute difference by the total of the two entities. Its use is restricted to positive values only, unless a correction to the denominator is made. Such a procedure was suggested by Gower (Sneath & Sokal 1973), and is applied as $(|x_{ik} + x_{jk}|)$. By

TABLE 2. Different forms of distance measures.

Measures of distance		d_{ij}	Eq. Nº
(a)	Non-standardized metric: Euclidean metric	$(\Sigma_{k=1}^{p} X^2)/p)^{1/2} = E$	1
	MCD	$(\Sigma_{k=1}^{p} \lvert X \rvert)/p = M$	2
(b)	Stardardized metric: Canberra metric	$(\Sigma_{k=1}^{p} (\lvert X \rvert/(x_{ik} + x_{jk})))/p$	3
	Bray-Curtis metric	$(\Sigma_{k=1}^{p} \lvert X \rvert)/(\Sigma_{k=1}^{p} (x_{ik} + x_{jk})$	4
	Gower*metric	$(\Sigma_{k=1}^{p} (1 - \frac{\lvert X \rvert}{r_k}))/p$	5
	Stardardized Euclidean metric	$E/s.d._k$	6
	Burr Standardized Euclidean metric	$(E/s.d_k)^2$	7
	Euclidean metric with range	E/r_k	8
	MCD with s.d.	$M/s.d._k$	9
	MCD with range	M/r_k	10

* Represents the similarity coefficient: $S_{ij} = 1 - d_{ij}$
r_k = Range of attribute k; $s.d._k$ = Standard deviation of attribute k;
$X = x_{ik} - x_{jk}$

using $|x_{ik}|$ instead of x_{ik} that the resulting distances change completely and thereby the whole final system. For example: let us consider four locations with an attribute - 4, 8, 12, 16. Then the corresponding distances in these two cases are: 3, 2, 5/3 and 1, 1, 1. In the former they are highly dissimilar while in the latter they are highly similar.

Bray & Curtis (1957) suggested a slightly different similarity metric (eq. 4). The difference between the Canberra metric and the Bray & Curtis (1957) measure is that, in the former, the distance represents the sum of the average absolute differences of attributes divided by the sum totals. In eq. 3 both the numerator and denominator carry a summation symbol; the ratio tends to be greatly influenced by occasional outstanding values. By contrast in the Bray & Curtis (1957) measure (eq. 4) the outstanding differences can only contribute to one of the fractions and so does not come to dominate the index (Clifford & Stephenson 1975). It will be noted that both the Bray & Curtis (1957) and the Canberra measures of dissimilarity involve at each stage only the pair of entities under consideration.

The general similarity coefficient of Gower (1971) is similar to MCD but is divided by the range, taking into account both positive and negative values (eq. 5). In this there is also a provision to give weights or masking to different attributes. The MCD presents the dissimilarity measure (d_{ij}) while the Gower metric presents the similarity measure (S_{ij}). At each stage it considers the entire population in terms of the range (r_k) of a particular attribute k.

The basic differences among these distance measures stem from three factors: (i) use of absolute difference of squared difference between pairs in the numerator; (ii) use of population range or s.d. of an attribute or pair sum of attributes in the denominator with a summation on the numerator; (iii) use of single summation for both numerator and denominator with pair sums of individual attributes in the denominator. The latter two contribute to the major differences in similarity matrices. The similarity matrix obtained with population range or s.d. in the denominator does not change the original order obtained by the numerator. Therefore, it works as a true standardization procedure, retaining the original order shown by the data matrix.

CORRELATION COEFFICIENT

The Pearson product-moment correlation coefficient ranges between -1 and +1. Boyce (1969) presented the correlation coefficient in terms of the angular measure as (Fig. 2):

$$d^2_{ij} = 2 - 2 \cos \Theta;$$

if Θ is zero, then the two locations A and B lie on the same straight line passing through the origin 'O'. This means $x_{ik} = ax_{jk}$ for all values of k where x_{ik} and x_{jk} represent the values of the k^{th} attribute for locations i and j. 'a' is known as proportional constant, while in this case, the correlation coefficient is unity (+ve if both A and B lie on the same side of the origin and -ve if they lie on opposite sides of the origin). This suggests that angular measures or correlation coefficients are not correct measures to represent true distance between any two locations in terms of their attributes. The product moment correlation coefficient (c.c.), therefore, ignores the proportional differences being equal to the cosine of the angle between two locations when the attributes of the respective locations are expressed as deviates from the mean of all attributes. The new data matrix of the individual stations is represented by zero mean and unit variance. Therefore the c.c. is nonmetric. When converted to some simple complementary form, corresponding to distances, it does not obey the triangle inequality and it can also be shown that perfect correlation could occur between non-identical individuals. These properties of the correlation coefficient limit its applicability and it is therefore regarded as inappropriate (Webster 1979).

More appropriate and mathematically sound similarity measures for numerical (continuous) data appear to be the standardized Euclidean metric and Gower metric.

STANDARDIZATION AND TRANSFORMATION

Smith (1976) suggested several standardization procedures. The s.d. in the case of second order deviations (Euclidean metric), the equivalent of variance ($(s.d._k)^2$) in the case of squared Euclidean metric, and range in the case of first order metric such as the Gower metric represent mathematically appropriate standardization procedures. Using the data matrix of Table 1, the similarity measures were computed using eqs. 1 & 2 and standardized both by the range and s.d. (eqs. 6, 8-10). These results suggest that the magnitude of similarity measures (Table 2) obtained by using range standardization are lower than those obtained using s.d. When dispersion is more among the attributes of the two locations, the ratios are slightly more compared to the contrary situation. Sometimes these small variations of individuals may be sufficient to alter groups. Results emphasize the fact that the new way of standardization is no way superior to the conventional procedures: Euclidean metric by s.d. and MCD by range (the latter represent the Gower metric) - eqs. 5 and 6.

Smith (1976) also suggested data transformations. By transformation undue weight is often given to some attributes with square root or exponential transformation, the distortion in the original data is too large and tails off to one end which reduces the range of variation. This is a weakness in any classification procedure. This procedure is generally used to derive the relationship between two parameters if they are curvilinearly related by converting curvilinearity to linearity before regression. Ivimey-Cook (1969) states that it is difficult to produce an absolute justification for this course of action in every case, but, on the other hand, there is no special virtue in the conventionally used linear scale of measurement.

CLASSIFICATION PROCEDURES

Classification procedures can be divided into graphical and numerical. The former represents the traditional approach while the latter represents more modern computer techniques. Each has advantages and disadvantages in their application to climatic classification studies.

GRAPHICAL PROCEDURES

The general practice is to present the spatial distribution of an attribute in geo-coordinates. Zones are identified by dividing the attributes at discrete intervals. These studies are not only based on observed climatic parameters such as rainfall and temperature but also on derived parameters like potential evapotranspiration. Details on some of the graphical procedures were presented by Reddy (Prelo c). In these studies, climate is classified using attributes one at a time. Climatic boundaries were chosen arbitrarily, corresponding to certain critical values of vegetation types. However, since the limits are more subjective, they clearly reflect personal bias.

The second graphical procedure is the shading of areas of equal similarity measure. The widely used similarity measures are the correlation coefficient (Rao et al. 1972) and principal coordinates or components (Dyer 1975). The aims of such studies are twofold: identification of homoclimes, that can be used as a predictive measure. This approach, however, is limited to regional studies only.

The first graphical procedures are in wide use at both regional and global scale, the second is in use only in the regional scale studies. These are traditional descriptive approaches that are limited in the number of attributes while limits used in the demarcation of boundaries reflect the personal bias of the climatologist. The major advantage of these procedures is that they represent the continuum in geo-coordinates which facilitate interpretation and assist validation of results. Also, it is easy to fit new locations into these groups and also it is easy to remember these groups.

NUMERICAL PROCEDURES

Numerical methods have become feasible in recent years with the advent of computers. In general the human brain is unable to manipulate any considerable mass of data in an integrated fashion. The computer is no more efficient than

its program and may be as efficient as a highly trained taxonomist. Under the numerical procedures there are three mutually exclusive techniques; Ordination, Cluster analysis and Minimum spanning tree (MST).

Ordination

The two common procedures that are in wide use are principal component analysis and principal coordinate analysis.

Principal component analysis (PCA)

In the PCA first rows are standardized (unit variance, zero mean) to give a square matrix of moment correlation coefficients between pairs of rows. Computing the principal components of this matrix involves the computation of its Eigen values and Eigen vectors. The importance of these vectors is that they are orthogonal. In other words, a large proportion of the dispersion engendered by the n rows over the m columns may be accounted for by p dimensions. PCA can also be carried out on a variance-covariance matrix (Craddock & Flood 1969, Craddock 1973, Barnett 1977).

The p normalized vectors give the directions of a set of p orthogonal axes in p-dimensional space and are known as the principal axes. The linearly independent principal components are ranked in terms of the amount of the total variance each component explains. The first component explain the largest proportion of the data variance. The second component is orthogonal to the first and explains the second largest amount of variance and so on. Most of the variance in the original data matrix can be explained by a few new components; often as few as three principal axes will suffice.

PCA adheres strictly to the geometry of the original Euclidean model. Situations when principal components can be interpreted in any physical sense is largely fortuitous; principal components are mathematical constructs, and do not necessarily have any physical meaning. There have been numerous attempts to obtain meaningful variates from combinations of others using methods that are known collectively as factor analysis (Catell 1952). These are simple analytical rotation of principal components.

Principal coordinate analysis (PCO)

The PCO technique developed by Gower (1966) is an important advance in ordination techniques. He has shown that with a suitable measure of similarity or dissimilarity between individuals, coordinates can be found relative to principal axes. The first step in the analysis is to calculate a distance d_{ij} between every pair of rows, i and j or, from similarity indices, S_{ij}, by scaling in the range 0 (for maximum possible dissimilarity) to 1 (for identity) and $d_{ij} = (2(1 - S_{ij}))^{1/2}$. From these distances a matrix Q can be formed with elements $q_{ij} = 1/2\ d_{ij}^2$. The matrix Q is now adjusted by subtracting from each element the corresponding row mean (q_i) and column mean (q_h) and adding the grand mean (q). Thus, the new matrix F can be formed with elements:

$$f_{ih} = q_{ih} - q_i - q_h + q$$

The latent roots and vectors of F are found, and the vectors are arranged as columns in nxn matrix; the rows representing coordinates of the individuals. The vectors are normalized so that the sums of squares of their elements equal their corresponding latent roots. This transforms the matrix F into a new matrix G. Gower shows that when this transformation is made, and starting from the matrices Q & F defined above, the distance between any two points i and j, whose coordinates are the i^{th} and j^{th} rows of G, equals d_{ij}^2. The latent vectors scaled in this way represent exactly the distances between individuals and defines their positions relative to principal axes.

When the starting matrix consists of Euclidean distances, PCO gives results identical with those of PCA. This means mathematically that both are similar, but PCO is more flexible in terms of similarity measures. However, Webster (1979) states that although PCO is more versatile than classical PCA, the latter is preferable; while Sneath & Sokal (1973) identified many advantages of PCO over PCA. However, both suffer from difficulty in interpretation as coordinates or components do not contain the physical meaning. One important feature in these studies is the dimensional reduction. When p is considerably large, the dimensions can be used as new attributes with

less noise and may be used to represent the spatial variation in geo-coordinates as in the case of graphical presentation.

As an example, PCO was carried out using the data matrix presented in Table 1 with the squared standardized Euclidean metric. The results of the first three coordinates are depicted in Fig. 3. Locations in coordinates 1 and 2 have a concave while coordinates 1 and 3 have a convex distribution. In this diagram, the arrangement of locations into finite groups is subjective.

PCA was used by Dyer (1975) to forecast rainfall and to minimize rainfall collection network by identifying homogeneous zones in South Africa; Willimott (1977, 1978) to classify California into homogeneous zones; Gadgil & Joshi (1981) to classify India into homogeneous zones and Reddy & Virmani (1982) to classify the semi-arid tropical India and West Africa into homogeneous zones. In these studies climatic attributes differ; Gadgil & Joshi (1981) used pentad rainfall (72 attributes for 52 locations); Reddy & Virmani (1982) used three different attribute sets, namely (i) monthly rainfall (12 attributes); (ii) average weekly rainfall (52 attributes); and (iii) weekly probability of getting 10 mm/week or more (52 attributes) for 81 locations (43 Indian + 38 Niger). Their area of study

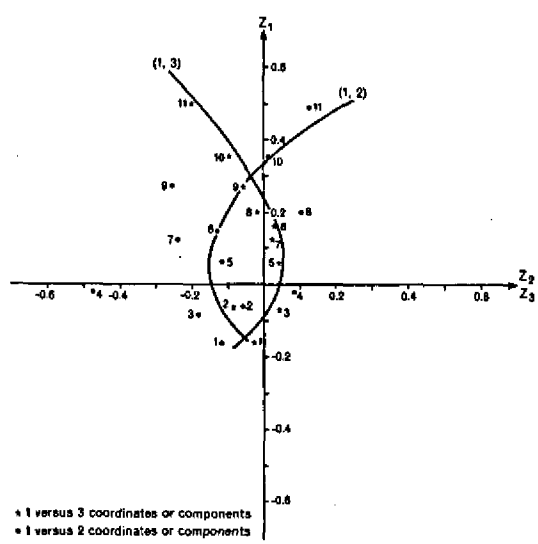


FIG. 3. Presentation of the 11 locations in the first three principal coordinates or components.

varied from local (Dyer 1975), to regional (Gadgil & Joshi 1981, Willimott 1977, 1978) and intercontinental (Reddy & Virmani 1982) scale. It is evident from these studies that if the proposed classification is only to subdivide a small region within a uniform general circulation pattern, the proposed classification looks quite satisfactory (Dyer 1975). In such studies one interest is to differentiate degree of local differences caused by orography, vegetation etc. Sometimes these differences are visually evident. If the interest is to group a nation or nations which have wide circulation patterns superposed on regional or local dissimilarities, then the proposed classification performance is less adequate, with many anomalies (Gadgil & Joshi 1981, Reddy & Virmani 1982). A problem associated with using both correlation or covariance, is that the mean of each station record does not influence the level of similarity between station records as these coefficients describe deviations about means. As a result, stations with highly different means could be identified as being similar when they are not (Reddy & Virmani 1982).

Clustering techniques

Clustering techniques seek to form 'clusters', 'groups' or 'classes' of individuals, such that individuals within a cluster are more similar in some sense than individuals from different clusters. Williams (1971) classifies clustering procedures (Fig. 1) into nonexclusive (overlapping) and exclusive (nonoverlapping). The overlapping procedure is of little use in the agroclimatic studies. Exclusive classifications are divided into extrinsic and intrinsic. Extrinsic procedures are monothetic divisive strategies used with qualitative data sets. These programs are not well developed (Clifford & Stephenson 1975). In an intrinsic classification all attributes used are regarded as equivalent. Fager & McGowan (1963) have initiated a non-hierarchical method of species classification where recurrent species groups with defined characteristics have been obtained. Techniques for non-hierarchical types are further divided into serial optimization of group structure, and simultaneously optimization of group structure (relatively undeveloped).

Hierarchical nonoverlapping classification produces groups whose relationship to one another are readily expressed in two dimensions, generally in the form of dendrogram. It is difficut to predict how many groups may be required. It seems this can best be decided by a process of trial and error. This reflects personal judgement or personal bias. Typically it appears best to generate an excess of groups and fuse some of these later. There are two basically different approaches of hierarchical classification procedures, monothetic divise and polythetic agglomerative. The first involves subdivision of the entities to be classified by one attribute after another considered in sequence (the classic climatic classification procedures). The second aggregates individuals into groups on the basis of their overall similarity with respect to all attributes considered simultaneously; a preferable approach. There are eight main fusion (classification) strategies that are non-overlapping, intrinsic, hierarchical, agglomerative-polythetic clustering techniques. FUSE (Turkey 1954) was designated as a package for the "exploratory analysis of data".

The basic procedures are similar. Beginning with the inter-individual similarity or distance matrix the methods fuse individuals or groups of individuals which are closest (or most similar), and proceed from the initial stage of n individuals to the final stage in which all individuals are in a single group. Differences between methods arise because of the different ways of defining distance (or similarity) between an individual and a group or between two groups. This suggests that the clustering techniques do not follow the hierarchy as presented above (Williams 1971) but they all represent different modes of fusion strategies and are based on the attributes state, type of groups reguired. All follow the same horizontal line rather than vertical lines as depicted in Fig. 1.

Using agglomerative-polythetic clustering, eight common strategies are available (Fig. 1). They are:

(i) NN - nearest neighbour or single linkage; (ii) FN - farthest neighbour or complete linkage; (iii) UPGMC (Centroid: unweighted pair group centroid method); (iv) WPGMC (Median: weighted pair group centroid method); (v) UPGMA (unweighted pair group method using arithmetic averages); (vi) WPGMA (weighted pair group method using arithmetic averages); (vii) IS - incremental sum of squares or minimum variance); and (viii) FB - flexible sorting, (Lance & Williams 1966, 1967a; Burr 1968, 1970). Lance & Williams (1966) generalised these under flexible fusion strategy. They are given as follows:

Fusion strategies

The generalised flexible strategy is expressed as:

$$d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} \beta_{dij} + \gamma \left| d_{hi} - d_{hj} \right|$$

where the parameters α_i, α_j, β and γ determine the nature of the sorting strategy; h, i and j are three groups containing n_h, n_i and n_j rows respectively and with intergroup distances d_{hi}, d_{hj} and d_{ij}. Here, d_{ij} is considered as the smallest of all distances, so that i and j fuse to form a new group k with $n_k = (n_i + n_j)$ elements.

Fig. 4 depicts the graphical representation of this equation. In the figure, if $d_{hi} < d_{hj}$ and h consists of one location ($n_h = 1$), i consists of three locations ($n_j = 3$) and j consists of two locations ($n_j = 2$ & $n_k = 5$), then new distance d_{hk} formed after the merger of i and j, differ under different fusion strategies (Table 4), for example:

NN: $d_{hk} = d_{hi}$

NF: $d_{hk} = d_{hj}$

WPGMA: $d_{hk} = 0.5\ (d_{hi} + d_{hj})$

UPGMA: $d_{hk} = (3/5)\ d_{hi} + (2/5)\ d_{hj}$

WPGMC: $d_{hk} = 0.5\ (d_{hi} + d_{hj}) - 0.25\ d_{ij}$

UPGMC: $d_{hk} = (3/5)\ d_{hi} + (2/5)\ d_{hj} - (3/5)(2/5) d_{ij}$

FB: $d_{hk} = 0.625\ (d_{hi} + d_{hj}) - 0.25\ d_{ij}$

IS: $d_{hk} = (((3 + 1)/(5 + 1))\ d_{hi} + ((2 + 1))/(5 + 1))\ d_{hj}) - (1/(5 + 1))\ d_{ij}$

This indicates that NN and FN do not give weight to the entire populations of the similarity matrix while computing the new distance matrix after each fusion. In the nearest neighbour strategy, a member enters a cluster at the similarity level equal to the highest similarity between the candi-

TABLE 3. Similarity matrices for different similarity measures with different standardization procedures using the data matrix presented in Table 1.

Locations	1	2	3	4	5	6	7	8	9	10	11
	I. M/K*										
1	0.00	0.33	0.22	0.50	0.31	0.41	0.46	0.44	0.56	0.60	0.79
2	0.24	0.00	0.32	0.55	0.38	0.46	0.51	0.47	0.49	0.60	0.77
3	0.16	0.27	0.00	0.30	0.19	0.29	0.32	0.42	0.41	0.54	0.76
4	0.42	0.45	0.24	0.00	0.38	0.39	0.33	0.62	0.42	0.62	0.86
5	0.28	0.33	0.16	0.29	0.00	0.14	0.23	0.26	0.29	0.36	0.58
6	0.38	0.42	0.26	0.35	0.12	0.00	0.15	0.27	0.24	0.28	0.51
7	0.36	0.49	0.25	0.30	0.17	0.09	0.00	0.41	0.23	0.35	0.56
8	0.38	0.36	0.39	0.56	0.24	0.22	0.31	0.00	0.49	0.26	0.38
9	0.55	0.44	0.38	0.36	0.24	0.19	0.17	0.28	0.00	0.30	0.50
10	0.55	0.49	0.51	0.60	0.33	0.27	0.33	0.18	0.24	0.00	0.23
11	0.72	0.66	0.74	0.83	0.56	0.49	0.52	0.38	0.45	0.21	0.00
	II. E/K										
1	0.00	0.89	0.57	1,40	1,04	1.32	1.27	1.37	1.83	2.05	2.56
2	1.29	0.00	0.91	1,53	1.09	1.48	1.71	1.34	1.57	1.75	2.30
3	0.79	1.80	0.00	0.86	0.59	0.93	0.79	1.33	1.38	1.48	2.43
4	1.88	1.88	1.04	0.00	1.13	1.23	1.08	1.98	1.45	2.07	2.96
5	1.15	1.23	0.69	1.32	0.00	0.41	0.62	0.89	0.89	1.24	2.03
6	1.44	1.59	1.02	1.39	0.48	0.00	0.30	0.75	0.62	0.96	1.78
7	1.56	1.78	1.02	1.20	0.77	0.47	0.00	1.05	0.74	1.16	1.88
8	1.61	1.67	1.44	2,19	0.97	0.94	1.35	0.00	0.95	0.72	1.33
9	1.90	1.76	1.49	1.60	1.02	0.82	0.91	1.25	0.00	0.84	1.62
10	2.20	2.15	1.74	2.23	1.36	1.05	1.27	0.97	1.05	0.00	0.80
11	2.85	2.69	2.61	3.10	2.15	1.90	2.05	1.56	1.78	0.89	0.00

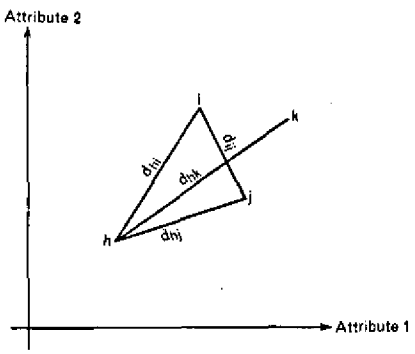
*M = MCD from Eq. 2 & E = Euclidean metric from Eq. 1
K = $s.d._k$ (upper triangle) ; here M/K represents Eq. 9
; & E/K represents Eq. 6
& K = r_k (lower triangle) ; here M/K represents Eq. 10
; & E/K represents Eq. 8
Eq. Nº are as referred in Table 2.

date and any member of the cluster; that is, a single link at a given similarity level is sufficient to allow entry to a cluster. The distance between a group and another individual is thus the distance between the individual and the nearest member of the group. The distance between groups is similarly the distance between their nearest members. The farthest neighbour is the exact antithesis of single linkage grouping; fusions are based on the distance between an entity and the most remote one in a group or between the most remote entities in two groups.

In the rest of the strategies, the whole population is taken into account; however, in the case of WPGMC, UPGMC, FB and IS, weight is given to the distance of a group that is currently formed a separate group while WPGMA and UPGMA considers the population left in the similarity matrix after the new group. There is no need to consider the distance which has already formed a new group while computing the new distance, as for example d_{ij} which relates to i and j is already taken into account in the formation of group ij.

With UPGMA a candidate for entry to a cluster is admitted at a similarity level equal to the average similarity between the candidate and the existing measure of the cluster. As the similarity levels are lowered remaining entities join one or another of the clusters. These procedures give an equal influence throughout the clustering process

to each individual. In the case of UPGMA (centroid), fusion of an entity into a group, or fusion of pairs of groups depends on the coordinates of the centroid. Groups are fused on minimal distance between centroids. Gower's (1967) centroid method is perhaps the most attractive fusion strategy from a geometric point of view taking into account the position of all members of each group in determining fusion. However, its exact geometric representation is still not entirely satisfactory (Webster 1979). In centroid sorting, if a small group fuses with a large one, it loses its identity and new centroid may come to lie entirely within the confines of the larger group. To indicate the individuality of the smaller group, it is desirable that group obtained after fusion should be intermediate in position. This is effected in WPGMC (or FB) sorting by regarding the groups as of unit size and obtaining a weighted median position after fusion (Clifford & Stephenson 1975). This strategy was apparently first suggested by Gower (1966) with a view to preventing large groups from dominating classifications to the exclusion of smaller groups. In WPGMA, like WPGMC equal weights are given to both groups irrespective of the number of entities in the individual groups.



Graphical representation of different fusion strategies of clustering
Figure 4

FIG. 4. Graphical representation of different fusion strategies of clustering.

IS has been proposed by several workers: Ward (1963) described it as an "error sum of squares" strategy; Anderson (1966) proposed it

TABLE 4. Hierarchical agglomerative-polythetic fusion strategies expressed as flexible strategy of Lance & Williams.

Fusion strategy	Flexible strategy parameters				Reference(s)
	α_i	α_j	β	γ	
NN$^{\$}$	0.5	0.5	0.0	-0.5	Sokal & Sneath (1963), Lance & Williams (1967a)
FN$^{\$}$	0.5	0.5	0.0	+0.5	Sorensen (1948), Sokal & Sneath (1963), McQuitty (1964), Lance & Williams (1967a).
UPGMC	n_i/n_k*	n_j/n_k	$-n_in_j/n_k$	0.0	Sokal & Michener (1958), Gower (1967), Lance & Williams (1967a)
WPGMC	0.5	0.5	-0.25	0.0	Gower (1966), Lance & Williams (1967a)
UPGMA	n_i/n_k	n_j/n_k	0.0	0.0	Sokal & Michener (1958), McQuitty (1964)
WPGMA	0.5	0.5	0.0	0.0	Lance & Williams (1967a), McQuitty (1966, 1967)
FB$^{\$}$	0.625	0.625	-0.25	0.0	Lance & Williams (1967a)
IS$^{\$}$	$\frac{n_h+n_i}{n_h+n_k}$	$\frac{n_h+n_j}{n_h+n_k}$	$\frac{-n_h}{n_h+h_k}$	0.0	Ward (1963), Anderson (1966), Orloci (1967), Burr (1968, 1970)

$d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} + \gamma \left| d_{hi} - d_{hj} \right|$

$ NN = Nearest neighbour; FN = Farthest neighbour; FB = Flexible β = -.25;
IS = Incremental sums of squares.

* $n_k = n_i + n_j$

under the name of "minimum variance clustering"; Orloci (1967) also developed the strategy under the " sum of squares method"; and, finally Burr (1968, 1970) coined the term "incremental sums of squares". Squares of Euclidean distance is used as a distance measure and after uniting the pair of elements whose squared distance is minimum, subsequent entities are fused such that the sum of squared distances within a cluster increases by a minimum. Because the total sum of squares is constant, if the sum of squared distances within a cluster increases minimally, then it follows that the squared distance between clusters is increased maximally.

Ward (1963) and Burr (1970) point out clustering could be based on the minimum sum of squares within clusters resulting from each fusion than on minimal increase of this value. Such a procedure frequently leads to absurd results and is not recommended (Clifford & Stephenson 1975). This clustering method may also be applied with other dissimilarity measures. A method of clustering allied to that just described is one in which there is a minimal increase in the variance (Wishart 1969, Anderson 1971) rather than the sum of squares within a cluster at each step in the fusion cycle. Its formulation is given by Burr (1970); however, its properties are not well known.

Comparative analysis

Clusters were determined using the eight fusion strategies for the data matrix presented in Table 1 with three similarity measures obtained from (i) GM - Gower metric with 11-attribute data matrix, (ii) SEM - standardized Euclidean metric with 11-attribute data matrix and (iii) EM - Euclidean metric with 7- attribute data matrix representing 7 principal coordinates from Gower's principal coordinate analysis. The results are presented in Fig. 5.

Using NN strategy the grouping under EM is poor. This is not improved much with the other two measures but the clarity is slightly better with GM. Using FN, groups formed under GM & SEM are similar to those under NN. Groups formed under EM appear to be more reasonable. Groups fomed under FB are anomalous while the groups formed under UPGMA appear to be acceptable. Groups formed under WPGMA with GM are similar to UPGMA, but the groups formed under other two measures show misclassifications. Groups formed under UPGMC & WPGMC show some misclassifications. Groups formed under IS with SEM & EM are similar and good. The groups formed under GM show poor clusters.

The above results suggest that the clusters formed under no two measures similar even under similar fusion strategies. The clusters formed under no two fusion strategies are similar. It generally appears that the first order metric (Gower metric) with first order fusion strategy (UPGMA & WPGMA) is the best while second order metric (Euclidean metric) with second order fusion strategy (IS) is the second best.

Tests of singnificance of results

Belbin (1982) suggested a simple test to determine distortional effects (Lance & Williams 1967 a e b, Williams et al. 1970) called the space distortion coefficient (SDC) defined as the ratio of level of last fusion (the maximum dissimilarity as suggested by dendrogram - D_1) to large dissimilarity in the association matrix - D_m i.e., D_1/D_m. Values around 0.6 indicate space-conservation; while values less than 0.4 suggest strong-contraction and values greater than 0.9 indicate space-dilation. For the example presented in Fig. 5, these estimates are presented in Table 5. This table suggests that NN and UPGMC are space-contracting; UPGMA, WPGMA & WPGMC are space-conserving and FN, FB and IS come under space-dilating strategies. However, according to Belbin (1982), UPGMC & WPGMC are space-dilating strategies, and Williams (1976a) and Sneath & Sokal (1973) observed UPGMC as space-conserving strategy. A reasonable rule with regard to choice of strategy is to utilize only space-conserving strategy unless data suggests specific effects may assist interpretation of structure (Belbin 1982). Therefore, in terms of space--conservation, UPGMA, WPGMA & WPGMC are the more reasonable fusion strategies. These criteria however, do not specify the significance of clusters misclassifications.

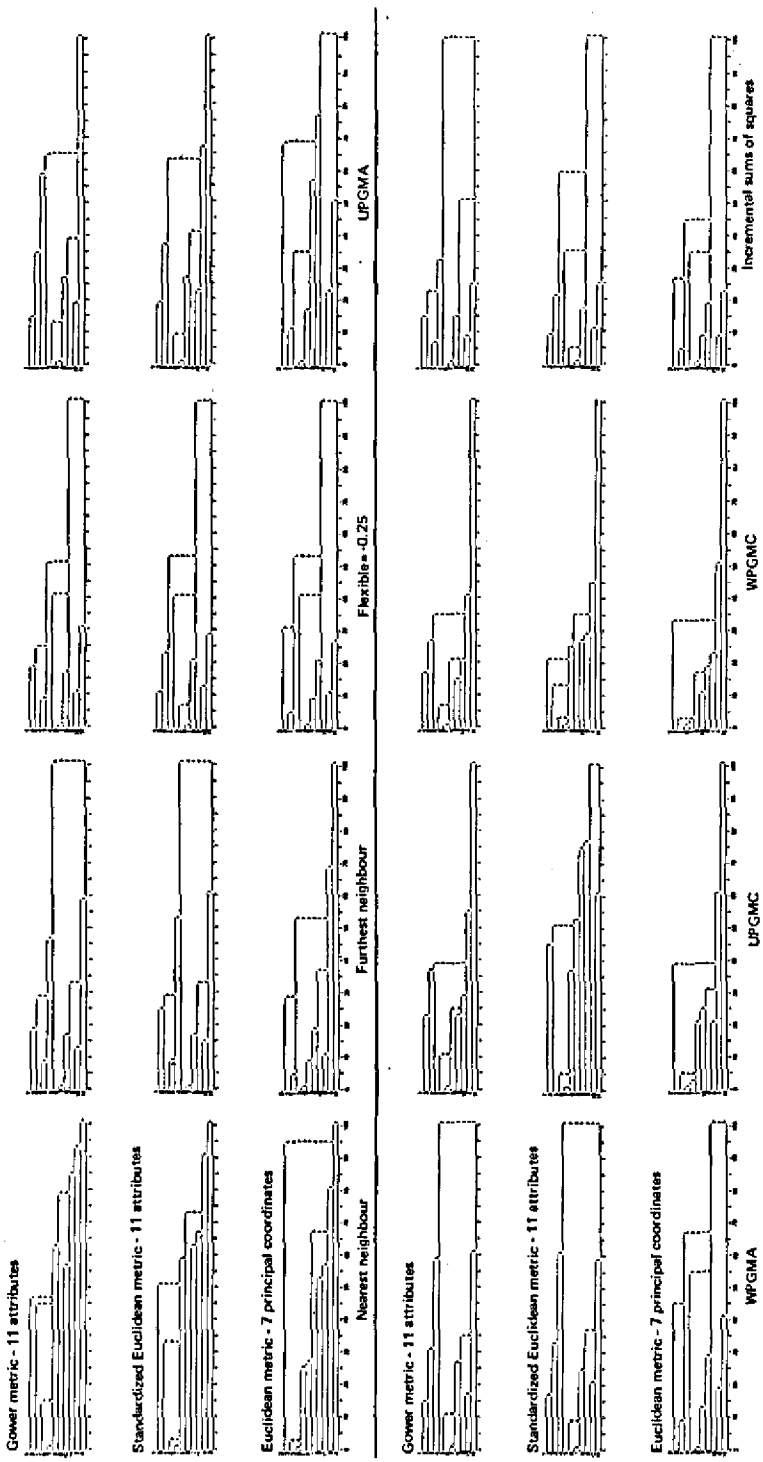


FIG. 5. Dendogram of different fusion strategies with different similarity metric and data matrices.

It must be admitted that one of the biggest deficiencies of cluster analysis is the lack of rigorous tests for the presence of clusters and for testing for the significance of clusters that are found (Lennington & Flake 1974, Ling 1971, Sneath & Sokal 1973). Although some criteria have been proposed (Goodall 1966a, b), the main deficiences are the specification of suitable null hypothesis, the determination of the sampling distribution of distance (or similarity) between data points and the development of flexible test procedure.

Rohlf (1974) summarizes a number of different measures of comparing two dissimilarity matrices, however most are either difficult to interpret or rarely used or both (Belbin 1982). One measure listed by Rohlf (1974) that is in common usage and simple to interpret is the Cophenetic correlation coefficient (Sokal & Rohlf 1962). This measure compares the dissimilarities implied between all individuals from the fusion table or dendrogram with those of the original measures of association. This is the Pearson's Product Moment correlation coefficient for observed (original) and expected (dendrogram) dissimilarities. As might be expected, the space-conserving strategies would, on average, produce the best correlation coefficient, because the correlation utilizes only half its range (inverse relationship should be non-existent). According to Belbin (1982), an alternative and simpler approach to this problem is to use Bray & Curtis (1957) measure and expressed as:

$$\text{Fedelity} = \Sigma_{k=1}^{n} \left| d_{ik} - d_{jk} \right| / \Sigma_{k=1}^{n} (d_{ik} + d_{ij})$$

where Fidelity = 0 perfect match and 1 for complete mismatch, d_{ik} = value of kth comparison of original dissimilarity and d_{jk} = value of kth comparison of dendrogram. A disadvantage of this type of measure is that it fails to detect the difference between different structures; markedly different dendrograms may produce the same fidelity value.

Table 5 presents these two coefficients for all the cases presented in Fig. 5. From this table it is seen that in terms of Bray & Curtis (1957) value, UPGMA is the best fusion strategy with all the three similarity measures and for the Cophenetic correlation coefficient, it is the best out of three for two measures using the 11-attributes and the second best using 7-attribute matrix (in this case UPGMC shows slightly higher value). In terms of the Bray & Curtis (1957) value, the second best method is WPGMA with all the three measures; however, in terms of the Cophenetic correlation coefficient, UPGMC & WPGMC appear preferable to WPGMA. Even in the case of WPGMA, it is relatively high. The Bray & Curtis (1957) value in WPGMC appears to be superior to UPGMC. The Bray & Curtis (1957) value suggests that IS is the poorest strategy. Cophenetic correlation coefficient suggests that NN is the poorest strategy. Even though FB and WPGMC are quite similar functionally, they differ substantially. These results suggest, therefore, that first preference could be given to UPGMA followed in order by WPGMA, WPGMC, UPGMC, FN, FB, NN, and finally IS.

Harbaugh & Merriam (1968) did not find any difference between the results obtained from standardized correlation coefficient or Euclidean distance using either UPGMA or NN in terms of structure in geological studies. Boyce (1969) states that the overall patterns of relationship produced by the UPGMA, WPGMA, WPGMC with measures of correlation are very similar and there are no topological differences between the dendrograms based on averages although the levels at which corresponding stems join do differ. In agroclimatic classification studies, however, the level at which the groups are found are very important. Russell (1978) used Canberra metric with FB fusion strategy to classify global climates. He used 16 monthly measured and derived attributes. The classification, however, does not distinguish locations with very different climatic regions. For example, Bellary, a very dry location, is grouped with Hyderabad, Sholapur, and Vishakhaptnam, wetter locations. Similarly Poona with Jabalpur & Raipur; Bikener & Jodhpur with Allahabad; Dwaraka with Bombay. These results may reflect inappropriate attribute data as much as they do the classificatory method.

Minimum spanning tree

The minimum spanning tree (MST) of a set of

TABLE 5. Coefficients of comparison between different fusion strategies using different similarity matrices with different types of attributes.

Similarity* metric	Coefficient$	Fusion strategies							
		NN	FN	FB	UPGMA	WPGMA	UPGMC	WPGMC	IC
GM	C	.426	.540	.512	.744	.606	.672	.671	.508
	B	.336	.244	.272	.110	.150	.246	.199	.357
	S	.302	1.000	1.070	.676	.728	.497	.658	1.320
SEM	C	.407	.521	.567	.731	.592	.596	.635	.568
	B	.269	.222	.262	.102	.140	.259	.186	.340
	S	.356	1.000	1.122	.703	.760	.375	.646	1.318
EM	C	.218	.498	.455	.504	.463	.529	.513	.452
	B	.216	.158	.231	.116	.136	.175	.156	.285
	S	.363	1.000	1.039	.576	.810	.491	.685	1.246

* GM = Gower metric; SEM = Standardizec (with S.D.) Euclidean metric;
EM = Euclidean metric (In the case of GM & SEM, the similarity matrices are computed from the data matrix in Table 1; while in the case of EM this is obtained from the 7 principal coordinate data matrix)

$ C = Cophenetic correlation coefficient; B = Bray-Curtis coefficient;
S = Space distorsion coefficient.

NN = Nearest neighbour; FN = Farthest neighbour; FB = Flexible β = -0.25; IS = Incremental sums of squares.

points is the network of minimum total length such that every point is joined by some path to every other point, and no closed loops occur. Because of this character MST was treated as a separate classificatory procedure. Eventually n-1 links are required to connect n points. Several methods of computing the MST are known, of which the algorithm of Prim (1957) is the most efficient (Ross 1969). Wroclaw Taxonomy (Florek et al. 1951) also uses the MST. The MST uses the similarity matrix. From the MST, the single linkage cluster analysis (NN) of Sneath can be computed directly (Gower & Ross 1969).

MST can be derived more efficiently than the corresponding dendrogram, and MST reveals not only which pair or pairs of individuals are most alike, but also which pairs of individuals in different branches of the tree are most similar (Webster 1979). MST is thus a useful way of exploring the distribution of individuals in character space and complements ordination analysis. The disadvantages of the MST is that it provides no information about how the various branches of the tree should lie relative to each other. This can be overcome for small trees by drawing the tree on the vector diagram provided by ordination results. Groups so defined are not clusters in any taxonomic sense but are purely a device to lessen computation. MST will be unique if the input does not contain any identical similarities. MST may also be used to check the groups produced by an intensely clustering strategy for misclassification.

DISCUSSION

From the above presentation, the most efficient way of grouping appears to be cluster analysis; the results are generally presented in the form of dendrogram (Mayer et al. 1953 introduced this term). Sneath & Sokal (1973) state that "there are as yet no satisfactory methods for testing from the similarity matrix itself whether clustering or ordination is most appropriate, although a high Cophenetic correlation may suggest that dendrogram is a reasonable representation of well clustered distribution". It is sometimes possible to discard a method completely because the results appear nonsensical, but in others the choice of which is best can not readily be made. In taxonomy, an experienced worker can generally detect an entity

which appears to have been misclassified. He can also judge good and bad classification. He makes these judgements on the basis of experience and intuition which may not be easy to quantify or even verbalize. In the clustering technique, angles between branches are of no importance, but points of origin of branches are very important. Williams & Lance (1969) believe that inadvertent chopping of continuous variation into somewhat arbitrary clusters does not usually damage the analysis irretrievably, because the continuity is generally fairly evident. Clifford & Stephenson (1975) states that it is not always desirable to truncate each branch of a dendrogram at the same level. In contrast, ordination may not describe sharp discontinuities if they can not be displayed in the first few dimensions. The major disadvantage with the ordination technique is the difficulty in interpretation; the components or coordinates do not contain the physical meaning even though each component indicates the attributes that contributed significantly. MST provides no information about how the various branches of the tree should lie relative to each other, therefore no clusters are defined in any taxonomic sense.

SUMMARY

Based on the above discussions the summary on the suitability of different methods for the classification of climate is presented below:

Merits and demerits of graphical and numerical procedures

Diversity is not only confined to data sets, but is a feature also of procedures that are involved in classification. These vary from the traditional descriptive to the more modern computer based numerical techniques. They differ in many respects. For example, in the descriptive procedures, it is not possible to handle many attributes simultaneously. The limit for a class or group in terms of attributes is prespecified at a discrete interval; therefore, addition or removal of locations will not alter the position of the location, while in the numerical techniques, this is not so. In the numerical procedures, no two methods give identical results while in the descriptive procedures the attributes that define the class or group differ and therefore so do associated groups unless the differentiating attributes are linearly correlated. The internal homogeneity is low in the descriptive procedures and relatively high in the numerical procedures. In the descriptive procedures the area presents a continuum of an attribute or group of attributes or group or class and in the numerical procedures it presents discrete or discontinuous. In the descriptive procedures the personal bias is more than numerical procedures. In both techniques, the differentiating characteristics or criterion variables (attributes) used to form classes should contain the maximum possible information for better groups, i.e., choice of attributes is critical for better classification. Because of these characteristics, in the broader zonation of world climates, the former is more useful and the latter is more useful in the finite grouping of these zones or in the agroclimatic classification. The major advantage of the numerical techniques over the descriptive methods is the ease with which the attributes can be integrated and group locations with least bias. The major weakness of the numerical methods is that no two methods give identical results and there is no established procedure for choice of optimal method. Also, with the change of data type (i.e., qualitative or quantitative) the choice of methods differ substantially and hence in each case one has to try all possible methods and check which method is suitable for his data. This process is not only time consuming but costly. Finally, the groups formed are to be validated subjectively since there is no formal test of homogeneity of misclassification.

Similarity measures for numerical (continuous) data set

Among the several similarity measures (that are used in the integration of attributes), the two that are commonly used in numerical (continuous) data are distance measures and correlation coefficient. Under the standardized and non-standardized distance measures Bray & Curtis (1957) and Canberra measures under the former involve at each stage only the pair of entities; while in the case of Euclidean metric standardized by population s.d. and mean character distance (MCD)

standardized by population range (Gower metric) considers entire population at each stage. Because of this, in the former group the similarity measures of some pairs gain undue weightage; a disadvantage - the purpose of standardization is to bring the differences into a uniform scale which is not achieved and as a result some groups get undue weight. Also applicability of the Canberra metric is limited to positive values. Some of the suggested modifications to extend this procedure to both positive and negative values appears to be invalid. Even though both MCD standardized by range and Euclidean metric standardized by s.d. are mathematically sound (obey the triangle inequality), their magnitudes differ. This is because the former represents the first order absolute difference while the latter presents the second order squared (and its square root) difference. Correlation coefficient is not a correct measure to represent the true distance between any two locations in terms of their attributes. It does not obey the triangle inequality and perfect correlation could occur between non-identical attributes. This tendency of correlation limits its applicability when the extremes are highly correlated.

New modes of standardization are in no way superior to the conventional procedures, i.e., first order differences (MCD) by population range and second order differences (Euclidean metric) by s.d. of population.

A weakness in the transformation of data to linearity is that this not only reduces the range of variation, but as in the Bray & Curtis (1957) and Canberra measures, undue weight is acquired by some pairs of measures.

Therefore, in the case of numerical (continuous) data the two more appropriate similarity measures are standardized (with s.d.) Euclidean metric of the second order differences, and Gower metric (MCD standardized by range) of the first order differences.

Applicability of numerical techniques for agroclimatic classification

Among the three numerical classification procedures, namely ordination, minimum spanning tree (MST) and clustering MST could be used as a check rather than as a separate classification procedure.

Ordination

Both principal component analysis (PCA) and principal coordinate analysis (PCO) under ordination are mathematically sound techniques. When the starting matrix consists of Euclidean distances, both give identical results. This means mathematically that both are similar, but PCO is more flexible in terms of similarity measures. But both suffer from the same weakness; that is the difficulty in interpretation, as coordinates or components are difficult to interpret in physical terms. A problem associated with ordination (PCA or PCO) using both correlation or covariance is that the mean of each station record does not influence the level of similarity between station records as these coefficients describe deviations about means. As a result, stations with highly different means could be seen as identical. Therefore, when the selected attributes of any pair of locations are highly correlated, irrespective of their magnitude, ordination (particularly PCA) is less suitable. Therefore, ordination is an exploratory technique rather than a technique for grouping or to obtain reasonable classes. Ordination can be used to generate new standardized attributes that are fewer in number and contain less noise than the original attributes. Also these explain the maximum variance in the data set. These new attributes could be used in the computation of similarity matrix and then calculation of clusters. The new attributes can be used to describe the spatial distribution and to identify homogeneous zones with respect to first few coordinates.

Cluster techniques

Under clustering there are several procedures existing in the literature. The most appropriate procedures for numerical (continuous) data set are hierarchical-nonoverlapping-agglomerative-polythetic techniques. Under these procedures there are eight fusion strategies, namely: NN, FN, UPGMC, WPGMC, UPGMA, WPGMA, IS, FB. The basic steps are similar in all of these, beginning

with the inter-individual similarity or distance matrix the methods fuse individuals or groups of individuals which are most similar and proceed from the initial stage of all individuals under individual groups to the final stage in which all individuals are in a single group. Out of these eight fusion strategies, two, namely NN and FN do not give weight to the entire population of similarity matrix, and these are respectively categorised as space-contracting and space-dilating strategies. WPGMC, UPGMC, FB and IS are biased by the distance of a group that is currently formed. UPGMA is mathematically simple and sound; and gives equal weight to all the individuals in a group. In UPGMC if small group fuses with a large one, the small group loses its identity. While FB and WPGMC are mathematically similar, FB is space-dilating strategy and on the contrary WPGMC is space-conserving strategy. IS and UPGMC are respectively space-dilating and space-contracting strategies. In terms of space-conservation, UPGMA, WPGMA and WPGMC are the more acceptable fusion strategies. According to the Cophenetic correlation coefficient, NN is the least acceptable strategy. IS is the least acceptable strategy according to the Bray & Curtis (1957) value while UPGMA is the most acceptable fusion strategy irrespective of similarity metric with WPGMA the second best. This is also true for the Cophenetic correlation coefficient under the majority of similarity metric. The Cophenetic correlation coefficient suggests that UPGMC are superior to WPGMA while WPGMC is still better than UPGMC. Therefore, according to these tests, UPGMA is superior consistently over others. Next in order comes WPGMA and WPGMC.

All these tests emphasize the mathematically soundness of different fusion strategies, but do not address problems of the level of misclassification in the clusters as such. Sometimes it is possible to discard a method completely because the results appear nonsensical. This type of subjective test also suggested that UPGMA, then WPGMA are the two fusion strategies with least misclassifications. Surprisingly, IS with Euclidean metric also produced acceptable clusters. This also emphasizes the fact that the above mentioned test procedures are not in fact tests for the testing of clusters. However, in IS the results are not consistent with other similarity metric but also IS is not as simple a procedure mathematically as that of UPGMA.

Therefore, both mathematically and practically the preferred fusion strategy for numerical (continous) data sets is UPGMA, followed with WPGMA the second preference.

To make resulting groups more meaningful for the interpretation of results as well as to facilitate the fitting of new location into these groups, some level of subjective judgement seems to be necessary.

APPENDIX

Terminology

It is necessary to address some of the confusing terminology that exists in the literature.

According to Simpson (1961), **systematics** is the scientific study of the kinds and diversity of objects and of any and all relationships among them; **taxonomy** is the theoretical study of classification, including its bases, principles, procedures and rules; and **classification** is the ordering of objects into groups (or sets) on the basis of their relationships, that is, of their associations by contiguity, similarity, or both. Therefore, taxonomy is a part of systematics, and classification is a part of taxonomy. Systematics cover wider aspects while the term classification is used in a restricted sense. Here the objects refer to climatic stations.

Individuals or locations or entities (Sneath uses OTU = operational taxonomic units) are the elements to be ordered or classified. Each individual has a number of items of information called **attributes** (Clifford & Stephenson 1975 and Williams 1976a present the details on the types of attributes that are in common use in taxonomic studies). Some arrange these into two categories, namely quantitative (continuous) and qualitative for the sake of simplicity.

The term **similarity measure** or **similarity coefficient** or **similarity metric** are synonymous. They involve the integration of different attributes through a mathematical function to provide a similarity or dissimilarity parameter. With the correlation coefficient the highest value indicates close similarity while in the case of distance

measure the lowest distance represents the most similar, because it is inappropriate to compare differences in attributes with a range of 0.0 to 1.0 with those with a range of 100 to 1000. The importance of bringing all these to a single range of 0 to 1 by a suitable method is emphasized. This process is known as **standardization.**

Exclusive refers to a given element occurring in one class and one class only. **Non-exclusive** refers to a given element that may appear simultaneously in more than one sub-class. Under **intrinsic** all attributes are regarded as equivalent while in the **extrinsic** an external attribute is declared in advance, i.e. specification is given in advance about an attribute. **Agglomerative** refers to a type of clustering algorithm which operates by successive grouping together of objects. Under **monothetic**, a class is defined by a single attribute while in the polythetic a class is defined by more than one attribute. Monothetic classifications are those in which the classes established differ by at least one property while in polythetic classification groups of individuals share a large proportion of their properties, but do not necessarily agree in any one property. **Hierarchical** refers to the process of optimization of a route between the entire population and the set of individuals of which it is composed while under **non-hierarchical** systems, the structure of the individual groups are optimized. **Clustering** is the formation of groups defined by hierarchical or non-hierarchical methods. A method of cluster analysis is said to be stable if small changes in the data lead to commensurately small changes in the results. A **dendrogram** is the diagramatic ilustration of relationships based on the degree of similarity. A nested-hierarchy permits grouping of a large number of taxonomic groups into fewer groups of higher rank. It is only when these groupings are mutually exclusive that optimum results can be achieved (for example, a given class at a level X can belong to only one class A' at level X-1, and this class A' to only one class A'' at level X-2, and so on). **Ordination** refers to the disposition of individuals in a reduced space defined by fewer axes than the original number of properties studied for those individuals.

If the distance from other objects contracts as the number of individuals in a group increases, this is known as a clustering strategy. **Space-dilating** strategies produce the opposite effects; as groups grow in size, they appear to recede from all other objects, and the chance of more individuals joining that group diminishes. **Space-conservation** refers to a situation where contraction and dilation effects are not evident.

Eigenvalue refers to the latent root of the data matrix (a scalar) and Eigenvector refers to the latent vector of the data matrix (a vector). Some of the terms like **R- and Q- techniques; A- and I-space** can be simplified by using rows as characters, the pairs for which association is to be examined and columns by the attributes.

ACKNOWLEDGEMENTS

The work was carried out during the stay of the author at the Australian National University, Canberra.

The author is thankful to The Australian National University for financial support and Drs. N.S. McDonald, H.A. Nix and Lee Belbin for discussions and comments on the draft. The author also expresses his thanks to Mr. Kewin A. Cowen for the cartographic assistance.

REFERENCES

ANDERSON, A.J.B. Numerical examination of multivariate soil samples. J. Int. Assoc. Math. Geol., 3: 1-15, 1971.

ANDERSON, A.J.B. A review of some recent developments in numerical taxonomy. Devana, University of Aberdeen, Scotland, 1966. Tese Mestrado.

ARKLEY, R.J. Statistical methods in soil classification research. Adv. agron., 28:36-70, 1976.

AUSTIN, M.P. & NIX, H.A. Regional classification of climate and its relation to Australian range land, In: STUDIES of the Australian Arid Zone. III. Water in range lands. Melbourne, CSIRO, 1978. p.9-17.

AUSTIN, M.P. & YAPP, G.A. Definition of rainfall regions of South-Eastern Australia by numerical classification methods. Arch. Met. Geogh. Biokl., Ser. B, 26:121-42, 1978.

BARNETT, T.P. The principal time and space scales of the pacific Trade wind Fields. J. Atmos. Sci., 34: 221-36, 1977.

BELBIN, L. Fuse: A FORTRAN 4 program for cluster analysis fusion for mini-computers. Canberra, CSIRO, 1982.

BOYCE, A.J. Mapping diversity: a comparative study of some numerical methods. In: COLE, A.J., ed. Numerical taxonomy. London, Academic Press, 1969. p.1-30.

BRAY, J.R. & CURTIS, J.T. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr., 27:325-49, 1957.

BURR, E.J. Cluster sorting with mixed character types: I. Standardization of character values. Aust. comput. J., 1:97-9, 1968.

BURR, E.J. Cluster sorting with mixed character types: II. Fusion strategies. Aust. comput. J., 2:98-103, 1970.

CAIN, A.J. & HARRISON, G.A. An analysis of the taxonomist judgement of affinity. Proc. Zool. Soc., London, 131:85-98, 1958.

CATELL, B. Factor analysis. Herper, New York, 1952.

CLIFFORD, H.T. & STEPHENSON, W. An introduction to numerical classification. London, Academic Press, 1975.

COCHEME, J. & FRANQUIM, P. A study of the agroclimatology of semi-arid area South of the Sahara in West Africa. Paris, FAO/UNESCO/WMO Interagency Project Tech. Rept., 1967. p.117-29.

CORMACK, R.M. A review of classification. Roy Stat. Soc. A, 134:321-67, 1971.

CRADDOCK, J.M. Problems and prospects for Eigenvector analysis in meteorology. Statistician, 22:133-45, 1973.

CRADDOCK, J.M. & FLOOD, C.R. Eigenvector for representing the 500 mb geopotential surface over the northern hemisphere. Quart. J. Roy. Met. Soc., 95:576-93, 1969.

DYER, T.G.J. The assignment of rainfall stations into homogeneous groups: an application of principal components analysis. Quart. J. Roy. Met. Soc., 101:1005-13, 1975.

FAGER, E.W. & MCGOWAN, J.A. Zooplankton Species groups in the North Pacific. Science, 140:453-60, 1963.

FLOREK, K.; LUKACZEWICZ, J.; PERKAL, J.; STEINHAUS, H. & ZUBRZYCKI, S. Taksonomia Wraclawska. Przegl. Antropol., 17:193-207, 1951.

GADGIL, S. & JOSHI, N.V. Use of principal component analysis in rational classification of climates. In: INTERNATIONAL CROPS RESEARCH INSTITUT FOR THE SEMI-ARID TROPICS. Climatic classification: a consultan's meeting. Pantancheru, India, 1981. p.17-26.

GOODALL, D.W. A new similarity index based on probabilities. Biometrics, 22:882-907, 1966a.

GOODALL, D.W. Numerical Taxonomy of bacteria - some published data re-examined. J. Gen. Microbiol, 42:25-37, 1966b.

GOWER, J.C. A comparison of some methods of cluster analysis. Biometrics, 23:623-8, 1967.

GOWER, J.C. A general coefficient of similarity and some of its properties. Biometrics, 27:857-71, 1971.

GOWER, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53:325-38, 1966.

GOWER, J.C. & ROSS, G.J.S. Minimum spanning trees and single linkage cluster analysis. Appl. Statist, 18:54-64, 1969.

HARBAUGH, J.W. & MERRIAM, D.F. Computer applications in stratigraphic analysis. London, John Wiley, 1968.

HARGREAVES, G.H. Precipitation, dependability and potential for agricultural production in North East Brazil. Brasília, EMBRAPA/Utah State University, 1971. 123p. (Publication, 74-D159).

IVIMEY-COOK, R.B. The phenetic relationships between species of Ononis. In: COLE, A.J., ed. Numerical taxonomy. London, Academic Press, 1969. p.69-90.

KÖPPEN, W. Das geographische system der klimate. In: KÖPPEN, W. & GINER, R. ed. Handbuch der Klimatologie. Berlin, Gerbrender, Borntrager, 1936. v.1 Part. c.

LANCE, G.N. & WILLIAMS, W.T. A general theory of classification sorting strategies. I. Hierarchical systems. Comput. J., 1:373-80, 1967a.

LANCE, G.N. & WILLIAMS, W.T. A generalized sorting stategy for computer classifications. Nature, London, 212:218, 1966.

LANCE, G.N. & WILLIAMS, W.T. Mixed-data classification programs. I. Agglomerative systems. Aust. Comput. J., 1:15-20, 1967b.

LENNINGTON, R.K. & FLAKE, R.H. Statistical evaluation of a family of clustering methods. In: ESTABLOOK, G.F. Numerical taxonomy. San Francisco, W.H. Freeman, 1974. p.1-37.

LING, R.F. Cluster analysis. s.l, Department of Statistics, Yale University, 1971. Tese Doutorado.

MCQUITTY, L.L. Capabilities and improvements of linkage analysis as a clustering method. Educ. Psycol. Measure., 24:441-56, 1964.

MCQUITTY, L.L. Expansion of similarity analysis by reciprocal pairs for discrete and continuous data. Educ. Psycol. Measure., 27:253-5, 1967.

MCQUITTY, L.L. Similarity analysis by reciprocal pairs for discrete and continuous data. Educ. Psycol. Measure., 26:825-31-1966.

MAYER, E., LINSLEY, E.G. & USINGER, R.L. Methods and principles of systematic zoology. New York, Mc Graw-Hill, 1953.

MOORE, A.W. & RUSSELL, J.S. Comparison of coefficients and grouping procedures with numerical analysis of soil trace elements data. Geoderama., 1: 139-58, 1967.

NIX, H.A. The Australian climate and its effects on grain yield and quality in Australian field crops. I. In: LAZENBY, A. & MATHERSON, E.M. eds.

Wheat and other temperate cereals. Sydney, Angus and Robertson, 1975.

ORLOCI, L. An agglomerative method for classification of plant communities. J. Ecol., 55:193-206, 1967.

PAPADAKIS, J. Climates of the world and their agricultural potentialities. Cordoba, Papadabis, 1975.

PRIM, R.C. Shortest connection networks and some generalisations. Bull. system. Tech. J., 36:1389-401, 1957.

RAO, K.N.; JAYANTHI, S. & BHARGAVA, V.K. Indian Monsoon correlations - Part I: Monthly intercorrelation for all the Meteorological sub-divisions of India. New Delhi, IMD, 1972. (Met. monograph climatology, 4).

REDDY, S.J. Agroclimatic classification: I. A method for the computation of attributes. Agric. Meteorol. Prelo a.

REDDY, S.J. Agroclimatic classification: II. Identification of attributes. Agric. Meteorol. Prelo b.

REDDY, S.J. Climatic classification: the semi-arid tropics and its environment - a review. Pesq. agropec. bras. Prelo c.

REDDY, S.J. & VIRMANI, S.M. Grouping of climates of India and West Africa: using principal component analysis. In: INTERNATIONAL CROPS RESEARCH INSTITUTE FOR THE SEMI-ARID TROPICS, Hyderabab. Annual report 1980-1981. Agroclimatology. Patancheru, 1982. p.21-5. (ICRISAT. Progress rept., 5).

ROBERTSON, G.W. Dry and wet spells. Teskam,UNDP/FAO, Tun Razak. Agric. Res. Center. Sungh 1976. (Project Field Report, Agrometeorology: A-6).

ROHLF, F.J. Methods of comparing classifications. Ann. Rev. Ecology & systemtics, 14:101-13, 1974.

ROSS, G.J.S. Classification techniques for large sets of data. In: COLE, A.J., ed. Numerical taxonomy. London, Acadamic Press, 1969. p. 224-33.

RUSSELL, J.S. Classification of climate and the potential usefulness of pattern analysis techniques in agroclimatological Research. In: PROC. Agroclimatological research needs of the semi-arid tropics. Patancheru, ICRISAT, 1978. p.47-58.

RUSSELL, J.S. & MOORE, A.W. Classification of climate: Pattern analysis with Australian and Southern African data as an example. Agric. Meteorol., 16: 45-70, 1976.

SIMPSON, G.G. Principles of animal taxonomy. New York, Columbia Univ. Press, 1961.

SMITH, R.W. Numerical analysis of ecolocical survey of data. s.l. University of Southern California, 1976. Tese Doutorado.

SNEATH, P.H.A. & SOKAL, R.R. Numerical taxonomy. San Francisco, W.H. Freeman, 1973.

SOKAL, R.R. Classification: purpose, principles, progress, prospects; clustering and other new techniques have changed classificatory principles and practice in many sciences. Science, 185:1115-23, 1974.

SOKAL, R.R. & MICHENER, C.D. A statistical method for evaluating systematic relationships. Kans. Univ. Sci. Bull. (38):1409-38, 1958.

SOKAL, R. R. & ROHLF, F. J. The comparison of dendrograms by objective methods. Taxon., 11: 33-40, 1962.

SOKAL, R.R. & SNEATH, P.H.A. Principles of numerical taxonomy. London, W.H. Freeman, 1963.

SORENSEN, T. A method of establishing groups of equal amplitude plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. Biol. Skr., 5: 1-12, 1948.

THORNTHWAITE, C. W. An approach towards a rational classification of climate. Geogr. Rev., 38:55-64, 1948.

TROLL, C. Seasonal climates of the Easth. In: RODENWALDT, E. & JUSATZ, H. ed. World maps of climatology. Berlin, Springer - Verlag, 1965. p.28.

TURKEY, J.M. Unsolved problems of experimental statistics. J. Am. Stat. Ass., 49:706-31, 1954.

WARD, J.H. Hierarchical grouping to optimizing an objective function. J. Am. Stat. Ass., 58:236-44, 1963.

WEBSTER, R. Quantitative and numerical methods in soil classification and survey: monographs on soil survey. Oxford, Clarendon Press, 1979.

WILLIAMS, W.T. Attributes. In: WILLIAMS, W.T. ed. Pattern analysis in agricultural science. Melbourne, CSIRO, Elsevier, 1976a. 331.

WILLIAMS, W.T. Hierarchical agglomerative strategies. In: WILLIAMS, W.T. ed. Pattern analysis in agricultural science. Melbourne, CSIRO, Elsevier, 1976b. 331p.

WILLIAMS, W.T. Principles of clustering. Annu. Rev. Syst. 2:303-26, 1971.

WILLIAMS, W.T.; CLIFFORD, H.T. & LANCE, G.N. Group-size dependance: a rationale for choice between numerical classifications. Computer J., 14:157-62, 1970.

WILLIAMS, W. T. & LANCE, G. N. Application of computer classification techniques to problems in land survey. Bull. Int. Statist. Inst., 42:345-55, 1969.

WILLIMOTT, C.J.A. A component analytic approach to precipitation regionalization in California. Arch. Met. Geophys. Biokl. Ser. B, 24:269-81, 1977.

WILLIMOTT, C.J.A. P-mode principal component analysis grouping and precipitation regions in California. Arch. Met. Geophys. Biokl., Ser. B, 26:277-95, 1978.

WISHART, D. An algorithm for hierarchical classifications. Biometrics, 25:165-70, 1969.

CONVITE AOS PESQUISADORES BRASILEIROS

O jornal brasileiro de ciências — SPECTRUM — convida os pesquisadores brasileiros a publicar nele trabalhos científicos nas áreas de:

Melhoramento de plantas
Conservação de alimentos
Fertilizantes
Controle de pragas
Plantas oleaginosas e óleos vegetais, e
Produtos veterinários

O Editor desta publicação, Engº Eduardo Subacius, sugere artigos voltados para o aspecto de técnica laboratorial (testes, análises, controle de qualidade, estatística, processos laboratoriais etc.).

Os interessados podem dirigir sua correspondência para o seguinte endereço:

SPECTRUM
Av. Santa Inês, 836 - Sala 2
CEP 02415 - São Paulo, SP.