

# CLUSTER ANALYSIS FOR ASSESSING THE AGRONOMIC ADAPTATION OF *PANICUM MAXIMUM* JACQ. ACCESSIONS<sup>1</sup>

MOACYR B. DIAS-FILHO, MIGUEL SIMÃO NETO and EMANUEL A. S. SERRÃO<sup>2</sup>

**ABSTRACT** - Cluster analysis techniques were conducted using the *degree of agronomic adaptation* — an integrated measure of susceptibility to diseases and insects, growth of vigor, recovery from simulated herbivory, soil cover, and potential for dry matter production — of 118 accessions of the forage grass *Panicum maximum* Jacq., evaluated during a two-year period, in eastern Brazilian Amazonia. The objective was to determine the feasibility of clustering procedures to classify the studied accessions according to their agronomic adaptation. Hierarchical (complete linkage) cluster and nonhierarchical (*k*-means) cluster analysis were found to be suitable statistical tools for assigning accessions to discrete groups. The 118 *P. maximum* accessions were classified into four distinct clusters of 21, 40, 41, and 16 accessions with, respectively, excellent, good, regular, and poor degree of agronomic adaptation.

Index terms: multivariate analysis, Amazonia, forage grass, statistical analysis.

## ANÁLISE DE AGRUPAMENTO NA AVALIAÇÃO DA ADAPTAÇÃO AGRONÔMICA DE ACESSOS DE *PANICUM MAXIMUM* JACQ.

**RESUMO** - Técnicas de análise de agrupamento foram aplicadas utilizando o "grau de adaptação agronômica" — uma medida integrada relacionada à suscetibilidade a pragas e doenças, vigor de crescimento, recuperação em relação à herbivoria simulada, capacidade de rebrota e cobertura do solo — de 118 acessos da gramínea forrageira *Panicum maximum* Jacq., avaliados durante um período de dois anos, na Amazônia oriental brasileira. O objetivo foi determinar o potencial de técnicas de análise de agrupamento e classificar os acessos estudados de acordo com o grau de sua adaptação agronômica. Técnicas hierárquicas (técnica do vizinho mais distante) e não hierárquicas ("*k*-means") de análise de agrupamento foram consideradas procedimentos estatísticos bastante adequados para agrupar os acessos estudados em classes distintas. Os 118 acessos de *P. maximum* foram classificados em quatro grupos formados, respectivamente, por 21, 40, 41, e 16 acessos, que apresentaram, através do período de avaliação, um excelente, bom, regular e mau grau de adaptação agronômica.

Termos para indexação: análise multivariada, Amazônia, gramínea forrageira, análise estatística.

## INTRODUCTION

Multivariate methods such as classification techniques are considered effective methodological tools in the study of agronomic adaptation of plant germplasm (Crossa, 1990). During the early phases of plant germplasm evaluation, when the number of individuals being studied and the variables being measured are relatively large, numerical classification techniques such as cluster analy-

sis can become valuable statistical tools to classify individuals according to their similarities (i.e., to separate individuals into relatively homogeneous groups). Examples of the use of cluster techniques with the previous or similar objectives are becoming increasingly common in the plant breeding and agronomic literature (Murphy et al., 1986; Jolliffe et al., 1989; Brown, 1991; Dias Filho et al., 1992; White, 1993).

Cluster analysis methods fall into two distinct groups: hierarchical, and nonhierarchical methods. Hierarchical clustering may be performed both by agglomerative and divisive methods. The results of both hierarchical procedures can be displayed

<sup>1</sup> Accepted for publication on May 10, 1994.

<sup>2</sup> Eng.-Agr., Ph.D., EMBRAPA-Centro de Pesquisa Agroflorestal da Amazônia Oriental (CPATU), Caixa Postal 48, CEP 66.095-100, Belém, PA, Brazil.

in a two-dimensional diagram known as a dendrogram or tree diagram (Du Toit et al., 1986) in which the objects (e.g., plant genotype) are joined according to their similarities, into successively larger clusters. Nonhierarchical classification groups objects into a predetermined number of  $k$  clusters or classes with the objective of minimizing the within-cluster variability while maximizing the between-cluster variability. A more in-depth description of these clustering techniques can be found in specialized texts like Sneath & Sokal (1973), Johnson & Wichern (1992) and Everitt (1993).

Statistical analysis in the initial stages of evaluation of the agronomic adaptation of forage germplasm to a particular environment can be considered an ideal situation for applying cluster analysis techniques. The number of accessions being evaluated is usually very large and many characteristics are simultaneously measured on each individual, usually by visual ratings, throughout a certain period. In such situations, the pasture scientist is usually not only interested in the *best* accessions, but rather, in distinct classes or *clusters* of similar accessions that could be sequentially tested at later evaluation phases. Also, as no assumptions such as normality, equal variances, or equal sample sizes need to be made to apply cluster analysis, this technique is fitted to analyze data from initial stages of agronomic adaptation trials, which usually do not completely fit such assumptions.

The objective of this work was to assess the feasibility of cluster analysis to summarize and interpret data from a two-year study, conducted in the eastern Brazilian Amazonia, on the agronomic adaptation of 118 accessions of the forage grass *Panicum maximum* Jacq. to the local biotic and abiotic conditions.

## MATERIALS AND METHODS

Data were from a two-year trial (Feb. 1988 to May 1990) conducted on 118 accessions of the forage grass *Panicum maximum* Jacq. in Paragominas (3° 05' S), eastern Brazilian Amazonia. The edaphic and climatic characteristics of the area have been described by Dias

Filho et al. (1992). Each accession was planted singly by seed in 1.8 m by 1.2 m plots, arranged in randomized complete blocks with three replications. Seeds were supplied by the Centro Nacional de Pesquisa de Gado de Corte (CNPGC), Campo Grande, Brazil. Table 1 lists the studied accessions according to their coding from SCPA (Sistema Cooperativo de Pesquisa Agropecuária) and ORSTON (Institut Français de Recherche Scientifique pour le Développement en Coopération).

At ten occasions, during a two-year period, a visual rating of each accession was made based on its *degree of agronomic adaptation*. This integrated measure of agronomic traits included: susceptibility to diseases and insects, recovery from simulated herbivory, soil cover, and potential for dry matter production (based on a high leaf:stem ratio). This overall rating was based on a zero to 6 scale; a rating of zero indicated very poor adaptation and a 6 indicated excellent adaptation. This procedure allowed for measurements on a common scale with an overall level and variation comparable across measurements.

Hierarchical clustering was carried out on the matrix of the ten evaluations of the *degree of agronomic adaptation* using complete linkage clustering method as the amalgamation rule. This agglomerative method starts by computing a matrix of Euclidean distances among objects' (i.e., accessions') means, finding the minimum entry in this matrix, and, subsequently, merging the corresponding objects to get specific clusters (Johnson & Wichern, 1992). The complete linkage amalgamation rule was preferred after comparing its results with those of single linkage and weighted pair-group average (WPGMA) methods. Subsequently, a nonhierarchical  $k$ -means type of clustering was done on the data set. This procedure partitions items into  $k$  initial clusters, assigning each object to the cluster whose mean (centroid), on the basis of Euclidean distances, is nearest. Recalculation of the centroid is then made for the cluster receiving a new object, and for the cluster losing this object (Sneath & Sokal, 1973). The ultimate goal of  $k$ -means clustering is to minimize the within-cluster variability while maximizing the between-cluster variability.

All computations and graphical output were performed using the computer program STATISTICA for Windows release 4.5 (Statistica, 1994).

## RESULTS AND DISCUSSION

The dendrogram from the hierarchical cluster analysis is presented in Fig. 1. Although the inter-

**TABLE 1. Codes of the 118 *Panicum maximum* accessions evaluated in the cluster analysis.**

SCPA <sup>1</sup>	ORSTON <sup>2</sup>	SCPA	ORSTON	SCPA	ORSTON
BRA-003581	15	BRA-004766	G85	BRA-007099	T18
BRA-003611	60	BRA-004821	G93	BRA-007102	T21
BRA-003638	63	BRA-004944	KK6	BRA-007111	T23
BRA-003689	73	BRA-005134	K194	BRA-007129	T24
BRA-003727	79	BRA-005436	K124	BRA-007145	T29
BRA-003735	80	BRA-005444	K125	BRA-007170	T46
BRA-003743	82	BRA-005461	K130	BRA-007218	T58
BRA-003751	83	BRA-005541	K143	BRA-007234	T60
BRA-003808	89	BRA-005657	K155	BRA-007251	T62
BRA-003891	101	BRA-005681	K159	BRA-007269	T65
BRA-003948	106	BRA-005690	K160	BRA-007285	T68
BRA-003999	115	BRA-005703	K162	BRA-007293	T71
BRA-004014	117	BRA-005711	K163	BRA-007307	T72
BRA-004022	118	BRA-005738	K165	BRA-007315	T74
BRA-004049	173	BRA-005746	K166	BRA-007331	T77
BRA-004120	354	BRA-005754	K168	BRA-007366	T84
BRA-004154	G3	BRA-005860	K181	BRA-007407	T90
BRA-004162	G4	BRA-005878	K182	BRA-007412	T91
BRA-004197	G7	BRA-005886	K183	BRA-007439	T95
BRA-004201	G8	BRA-005894	K184	BRA-007447	T96
BRA-004219	G9	BRA-006041	K2	BRA-007455	T97
BRA-004260	G14	BRA-006238	K51	BRA-007480	T104
BRA-004278	G15	BRA-006254	K55	BRA-007498	T106
BRA-004341	G22	BRA-006262	K58	BRA-007501	T108
BRA-004359	G23	BRA-006271	K59	BRA-007528	T110
BRA-004383	G27	BRA-006319	K64	BRA-007609	KK8
BRA-004413	G31	BRA-006343	K68	BRA-007617	KK10
BRA-004421	G32	BRA-006360	K71	BRA-007633	KK14
BRA-004448	G34	BRA-006645	K190A	BRA-007641	KK15
BRA-004464	G36	BRA-006653	K190B	BRA-007650	KK16
BRA-004511	G42	BRA-006661	K191	BRA-007668	KK17
BRA-004537	G45	BRA-006670	K193	BRA-007676	KK18
BRA-004545	G46	BRA-006777	K213	BRA-007731	KK33
BRA-004553	G47	BRA-006785	K214	BRA-008087	97
BRA-004561	G48	BRA-006815	K217	BRA-008117	G54
BRA-004588	G51	BRA-006823	K219	BRA-008141	G109
BRA-004600	G56	BRA-006891	K237	BRA-008150	K5
BRA-004669	G68	BRA-006980	K248	BRA-008184	K192
BRA-004693	G71	BRA-006998	K249		
BRA-004707	G73	BRA-007081	T17		

<sup>1</sup> Code from the Sistema Cooperativo de Pesquisa Agropecuária.

<sup>2</sup> Code from the Institut Français de Recherche Scientifique pour le Developpement

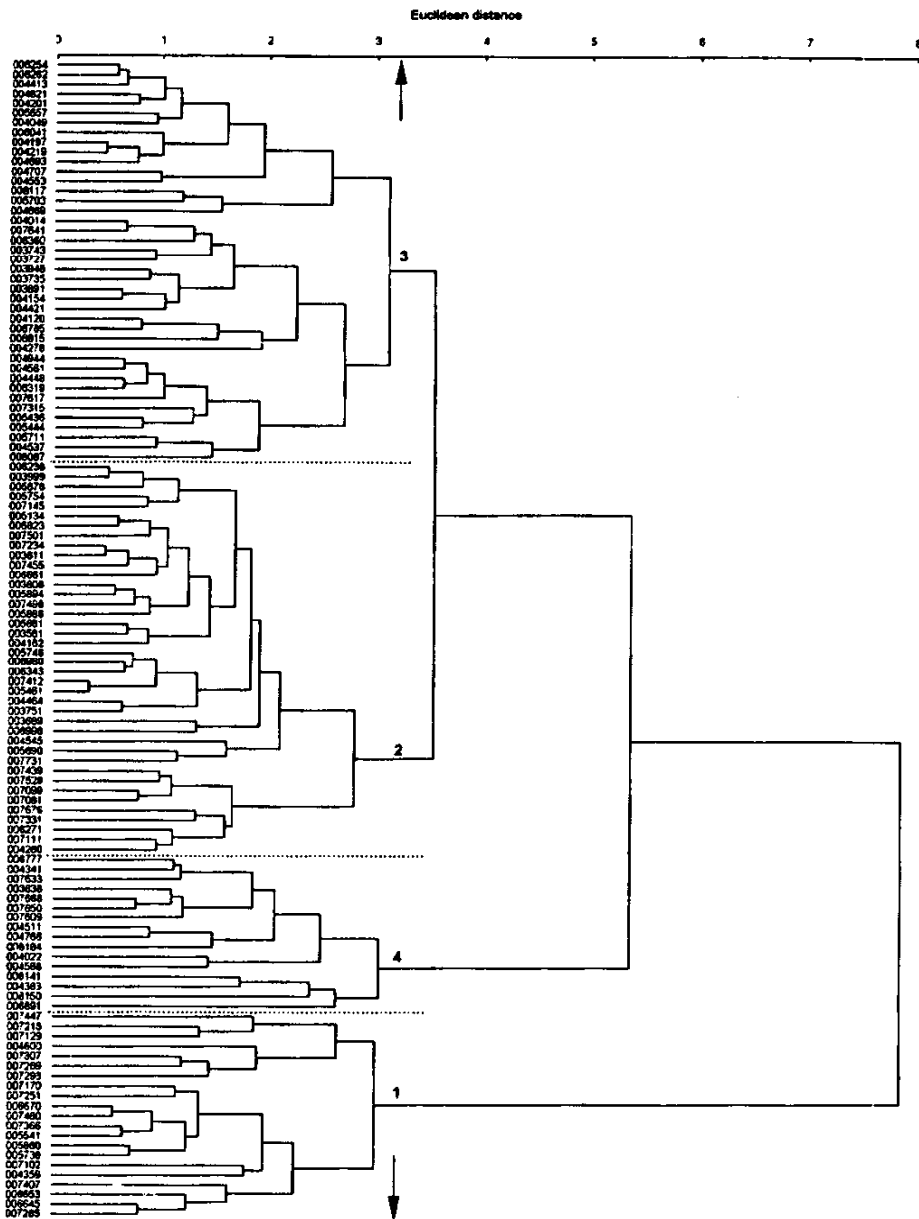


FIG. 1. Dendrogram based on the agronomic adaptation of 118 accessions of *Panicum maximum*. Accessions are identified by their SCPA coding (BRA-#). The clustering method was complete linkage and the distance was Euclidean. Clusters are limited by dotted lines and identified by numbers. Arrows indicate the selected cutoff distance to retain clusters.

pretation of a dendrogram can be subjective, as no definitive methods exist for determining the number of clusters to be retained (Pielou, 1984), a plot of linkage distances (i.e., Euclidean distances) at successive clustering steps can be useful to decide the number of clusters, by suggesting a cutoff for the dendrogram. If a clear plateau can be identified in this plot, it means that many clusters were formed at the same linkage distance; this distance may be an optimal cutoff in determining the number of clusters to be retained (Aldenderfer & Blashfield, 1984). In Fig. 2, a plot of linkage distances across steps (i.e., amalgamation schedule) shows a plateau up to the Euclidean distance of, approximately, 3.2 (indicated by an arrow).

Examining the dendrogram (Fig. 1), and determining 3.2 as a cutoff point (as suggested by the amalgamation schedule in Fig. 2), gives us four distinct clusters of *Panicum maximum* accessions. These clusters are labeled in Fig. 1 from 1 to 4, in descending order of the agronomic adaptation of their members. Cluster 1, containing 21 accessions, corresponds to the accessions with *excellent* agronomic adaptation. Cluster 2, consisting of 40 accessions, aggregates accessions that had

*good* agronomic adaptation, however, inferior to the accessions in Cluster 1. Clusters 3 and 4, with, respectively, 41 and 16 accessions, group the accessions that had *marginal* and *poor* agronomic adaptations. The classification of each cluster into the different agronomic adaptation classes was aided by examining the descriptive statistics (mean and standard deviation of the degree of agronomic adaptation) for the members of each cluster (data not shown).

The nonhierarchical *k*-means clustering was done so that the number of prespecified clusters corresponded to the number of clusters assigned from the hierarchical clustering procedure (i.e., four clusters—*excellent*, *good*, *marginal*, and *poor* agronomic adaptations). The results of the *k*-means clustering almost entirely matched the clusters found in the hierarchical clustering procedure. Cluster 1 consisted of the same 21 accessions assigned by the dendrogram in Fig. 1. Clusters 2, 3, and 4, however, with, respectively, 39, 40, and 18 accessions, had a slightly different membership from the clusters assigned by the hierarchical clustering procedure. Accessions BRA 003948, 004197, 004707, 004821, 004944,

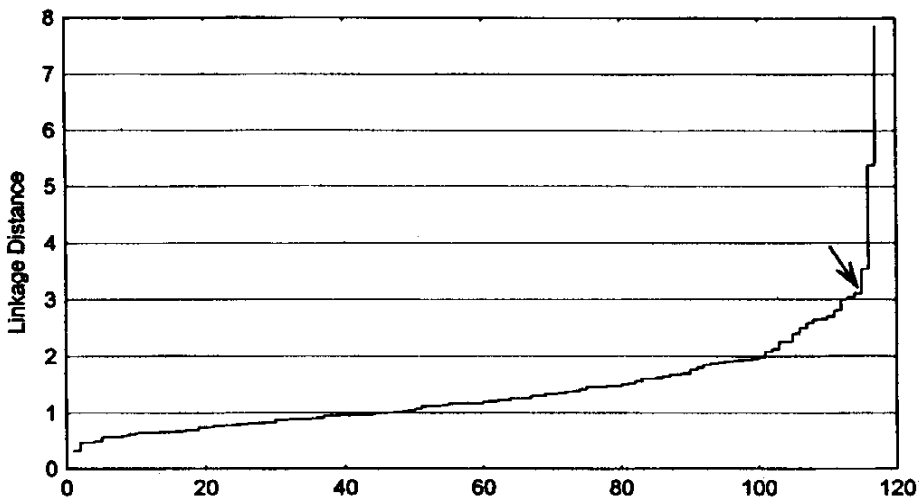


FIG. 2. Stepchart of the linkage distances at successive clustering steps. The clustering method was complete linkage and the distance was Euclidean. The arrow indicates the chosen cutoff distance for the dendrogram in Fig. 1.

005711, 006041, 006262, 006254, and 007315, which in the dendrogram from the hierarchical clustering were part of Cluster 3 (Fig. 1), in the *k*-means cluster were moved to Cluster 2 (which groups accessions with *good* agronomic adaptation). Accessions BRA 007111, 007331, and 007676, in *k*-means clustering were members of Cluster 3 (which groups accessions with *marginal* agronomic adaptation), whereas in the hierarchical clustering, these accessions were members of Cluster 2 (Fig. 1). Finally, *k*-means procedure added the accessions BRA 004014 and 004553 to Cluster 4 (which groups accessions with *poor* agronomic adaptation), these accessions were part of Cluster 3 in the dendrogram from the hierarchical clustering procedure (Fig. 1). So, the little inconsistencies observed in the cluster memberships from the two clustering methods performed on the *Panicum maximum* accessions were concentrated in the adjacent Clusters 2 and 3, which probably share some mutual characteristics.

To evaluate the appropriateness of the *k*-means clustering procedure with the prespecified number of clusters (i.e., four), a standard between-groups (i.e., clusters) analysis of variance for each dimension (i.e., variable) was performed. The result of this analysis is shown in Table 2. The large magnitude of the *F* values suggested a small within-cluster variability and a large between-cluster variability; this result strengthens the appropriateness of the *k*-means procedure. In addition, a plot of the means of each cluster (Fig. 3) shows how distinct the four prespecified clusters were at each period of evaluation (dimension), further supporting the appropriateness of the *k*-means procedure.

Although nonhierarchical clustering has been suggested for early stages of data analysis (e.g., Crossa, 1990), the use of *k*-means, as a subsequent procedure after hierarchical clustering, may also be considered. In this situation, the prespecified number of clusters in *k*-means should be based on the number of clusters retained by a hierarchical clustering method such as complete linkage. Comparing the kinds of solutions suggested by the two procedures could further reinforce the final grouping membership decision and reduce some of the subjectivity in assigning accessions to dis-

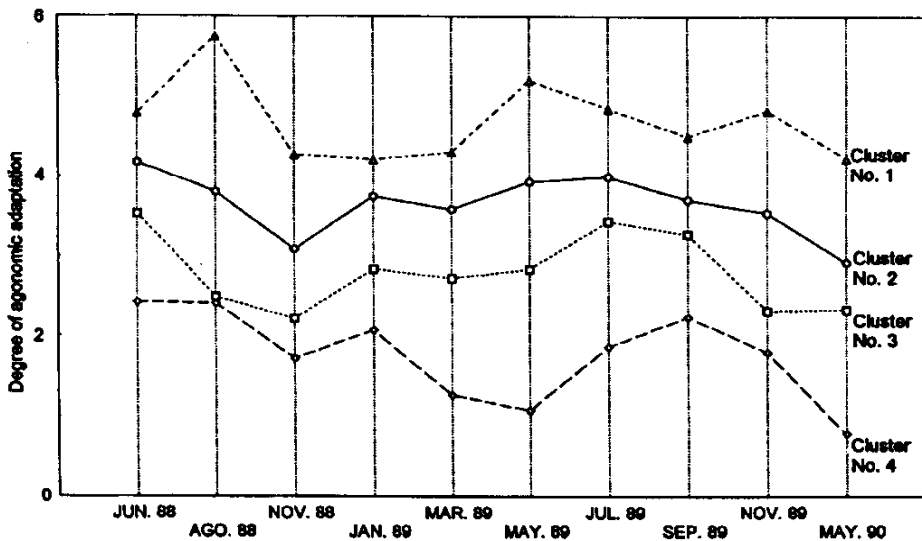
tinct clusters. As no great differences could be detected in the cluster memberships assigned by the two clustering procedures for this study, the memberships suggested for each group in Fig. 1 were accepted for further investigation.

Each group or cluster of accessions produced by the hierarchical clustering procedure (Figure 1) could be further investigated by looking for sub-clusters within each retained cluster. For example, within the group of accessions which had excellent agronomic adaptation (Cluster 1 in Fig. 1) a sub-cluster is formed by the accessions BRA 007293, 007269, 007307, and 004600 at the Euclidean distance of approximately 2.0. These accessions had the highest agronomic adaptation. At the same distance, two additional sub-clusters can be detected within Cluster 1 (Fig. 1), one containing BRA 007285, 006645, 006653, and 007407, the other containing BRA 007129, 007218, and 007447; these two sub-clusters group the accessions that had intermediate performance within the group of accessions with excellent agronomic performance. Finally, at the same distance, a sub cluster including BRA 004359, 007102, 005738, 005860, 005541, 007366, 007480, 006670, 007251, and 007170 is formed (Fig. 1); these accessions had inferior performance relative to the other accessions within Cluster 1. Similar investigations could be performed for each of the remaining three main clusters identified in Fig. 1. The use of other multivariate techniques such as multidimensional scaling or principal factor analysis could also be employed to graphically depict the structural nature within each cluster. In this case, however, a correlation matrix for the members of each examined cluster would have to be calculated.

For the data examined in this study, cluster analysis techniques proved to be a suitable way to categorize the 118 evaluated accessions of *Panicum maximum*. We were able to conveniently divide this relatively large number of accessions into four distinct hierarchical groups and to further classify these groups or clusters into hierarchical sub-clusters. These procedures can greatly help the pasture scientist in summarizing and interpreting data in the early stages of studies conducted on the agronomic adaptation of a large number of forage accessions.

**TABLE 2. Analysis of variance for each dimension (agronomic adaptation at each period of evaluation) generated by the k-means procedure.**

Variable	Between sum of squares	Degrees of freedom	Within sum of squares	Degrees of freedom	F
June 88	15.75	3	15.54	114	38.51
August 88	42.61	3	10.13	114	159.80
November 88	20.83	3	21.55	114	36.73
January 89	15.30	3	13.93	114	41.73
March 89	26.58	3	11.61	114	87.02
May 89	48.01	3	18.92	114	96.43
July 89	23.67	3	16.05	114	55.04
September 89	13.48	3	12.14	114	42.17
November 89	31.07	3	13.77	114	85.76
May 90	31.10	3	23.92	114	49.41

**FIG. 3. Plot of the agronomic adaptation means, at each measurement period, for the four clusters detected by the k-means clustering procedure.**

## CONCLUSIONS

1. Cluster analysis was found to be a useful statistical tool for summarizing and interpreting data on the agronomic adaptation (based on visual ratings of a number of agronomic traits) of a large number of forage accessions.

2. The application of the nonhierarchical procedure *k*-means clustering, following hierarchical clustering, proved to be useful to further strengthen the final clustering membership decision, as well as to reduce some of the subjectivity in allocating accessions to distinct clusters.

### ACKNOWLEDGEMENTS

To Sana Gardescu and Peter Marks (Cornell University) for reviewing an early draft of this manuscript.

### REFERENCES

- ALDENDERFER, M.S.; BLASHFIELD, R.K. **Cluster analysis**. Newbury Park: Sage Pubns., 1984. 87p. (Sage University Paper series on Quantitative Applications in the Social Sciences, 44).
- BROWN, J.S. Principal component and cluster analysis of cotton cultivar variability across the U.S. cotton belt. **Crop Science**, v.31, p.915-922, 1991.
- CROSSA, J. Statistical analysis of multilocation trials. **Advances in Agronomy**, v.44, p.55-85, 1990.
- DIAS FILHO, M.B.; SIMÃO NETO, M; SERRÃO, E.A.S. Adaptação de acessos de *Centrosema acutifolium*, *Desmodium ovalifolium* e *Pueraria phaseoloides* na Amazonia oriental do Brasil. **Pasturas Tropicales**, v.14, p.18-23, 1992.
- DU TOIT, S.H.C.; STEYN, A.G.W.; STUMPF, R.H. **Graphical exploratory data analysis**. New York: Springer-Verlag, 1986. 314 p.
- EVERITT, B.S. **Cluster analysis**. 3rd ed. New York: Halsted Press, 1993. 170 p.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 3rd ed. New Jersey: Prentice Hall, 1992. 642 p.
- JOLLIFFE, I.T.; ALLEN, O.B.; CHRISTIE, B.R. Comparison of variety means using cluster analysis and dendrograms. **Experimental Agriculture**, v.25, p.259-269, 1989.
- MURPHY, J.P.; COX, T.S.; RODGERS, D.M. Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. **Crop Science**, v.26, p.672-676, 1986.
- PIELOU, A.C. **The interpretation of ecological data**. New York: John Wiley and Sons, 1984. 263 p.
- SNEATH, P.H.A.; SOKAL, R.R. **Numerical taxonomy: the principles and practice of numerical classification**. San Francisco: W.H. Freeman, 1973. 573 p.
- STATISTICA FOR WINDOWS **General conventions and statistics I**. Tulsa: StatSoft, 1994. v.1, 717p.
- WHITE, W.H. Cluster analysis for assessing sugarcane borer resistance in sugarcane line trials. **Field Crops Research**, v.33, p.159-168, 1993.