References and Notes

1. P. J. Barrett, R. J. Baillie, E. H. Colbert, *Science* **161**, 460 (1968).
2. E.H.C., with a team of vertebrate paleontologists (W.J.B., J.A.J., and J.S.P.), joined the Institute's field party in order to search for vertebrate material.
3. D. H. Elliot, *Proceedings of the Second Columbia River Basalt Symposium* (1970).
4. G. W. Grindley, *N.Z. J. Geol. Geophys.* **6**, 307 (1963).
5. P. J. Barrett, *Inst. of Polar Studies Rep. No. 34* (1969).
6. J. F. Lindsay, *Inst. of Polar Studies Rep. No. 33* (1969).
7. P. J. Barrett, D. H. Elliot, J. D. Gunner, J. F. Lindsay, *Antarctic J. U.S.* **3**, 102 (1968).
8. Paleobotanical evidence reviewed by J. M. Schopf.
9. J. W. Collinson, J. S. Powell, D. H. Elliot, in preparation.
10. W. J. Gealy, personal communication.
11. D. A. Coates, personal communication.
12. In a recent paper Cosgriff infers that the *Lystrosaurus* zone of South Africa lies "just below the Permo-Triassic boundary." Except for this dissent, paleontological and geological opinion is essentially unanimous in regarding the *Lystrosaurus* zone as representing the lowest phase of Triassic history in South Africa. See J. W. Cosgriff, *J. Roy. Soc. W. Australia* **52**, 86 (1969).
13. Supported by grant No. GA-12315 from NSF to the Ohio State University Research Foundation (RF 2882). U.S. Naval Task Force 43 provided logistic support for the field work. Contribution No. 168 of the Institute of Polar Studies, Ohio State University.

8 June 1970

Cluster Analysis and Multidimensional Scaling of Archeological Sites in Northern Chile

Abstract. *Cluster analysis and multidimensional scaling procedures are used to test Chilean site relationships derived by traditional archeological methodologies. The results tend to confirm the intuitive evaluations but clearly indicate the value of computerized statistical analyses for this kind of comparative study.*

Archeological investigations in the Province of Tarapacá, northern Chile, resulted in the recovery of data from a number of preceramic sites. A traditional analysis of the associated artifacts (*1*) suggested the following site groupings: workshop sites (sites 9, 10, 11, and 28); seasonal campsites with some indication of vegetable food processing (sites 24, 25, 26, 29, and 31); campsites suggesting utilization of vegetable food resources and a minimum concern with hunting (sites 1, 2, 3, 4, 8, 12, 14, and 18); small campsites with distinctive knife and projectile point artifacts and some milling-stone elements (sites 2A and 14A); and a single component marked by separable projectile point types and the presence of maize (*2*).

The classification described above was fairly obvious from traditional archeological grouping procedures. However, within-group similarity tended to be based mainly on only one or two preselected diagnostic artifact classes and on such gross differences in site artifact collections as the relative abundance of scrapers and the presence or absence of thinned, bifacially flaked tools. These are valid observations, from which it was possible to make the above suggestions about intersite relationships and site groups. In order, however, to reduce possible bias in such traditional classification techniques, detailed comparative analyses that employ complete site inventories and artifact counts are planned for the future, by which groups of similar sites can hopefully be defined as polythetic units (*3*).

A careful preliminary testing of relationships, using formal statistical techniques, seemed to be in order, however, and the results of the pilot study are presented here. The preliminary analysis, it was hoped, would serve to support or refute the original intuitive, traditional classification of sites; it would provide a basis for planning and evaluating future statistical data manipulations; and it would help to determine the usefulness, generally, of studies based on partial, as opposed to complete, data inventories.

The pilot study was performed as follows. Seventy-four characters were selected for 20 preceramic sites along the lower reaches of the Arroyo Tarapacá, northern Chile. These characters were scored as present or absent for each archeological site. The characters are mainly bead and lithic artifact traits; no floral, faunal, or locational data and no midden characteristics were used, although they will enter into the final analysis planned for the future. The studied lithic artifacts account for more than 95 percent of the total artifact assemblages. In general, the characters are gross physical features of the specimens themselves and do not represent traditional artifact types. By using 74 characters, we hoped that the sample would be sufficiently large to minimize any distortion of site relationships that might follow from including possibly interdependent characters.

A Burroughs 5500 computer was used to compute a matrix of similarity scores and to carry out two different analyses of this matrix. Jaccard's coefficient of association was calculated to depict the similarity between all pairs of the 20 sites. This coefficient computes similarity on the basis of positive matches and on the basis of the characters unique to each site in a pair; characters that are absent from both members of a pair are not included in the calculation. First, the sites were clustered by using the average linkage, unweighted pair-group method (*4*), the

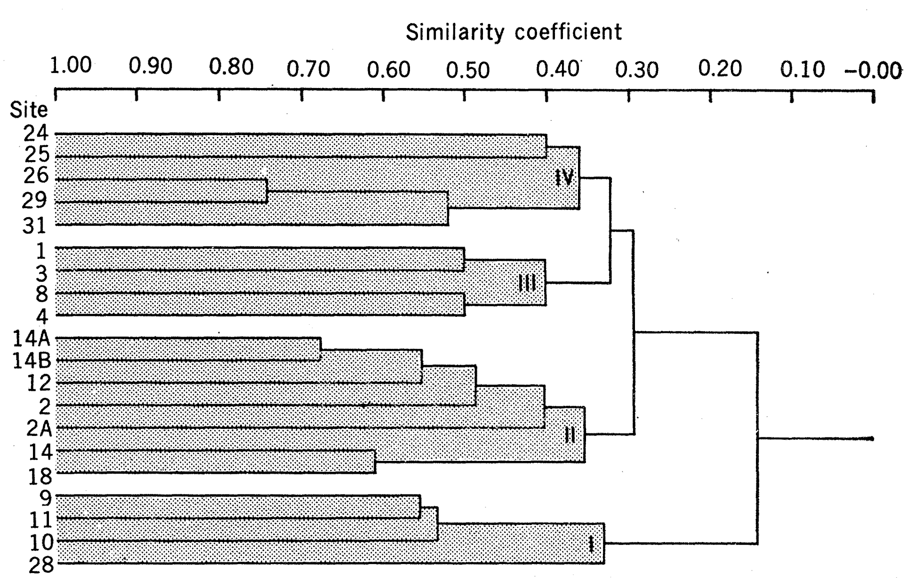


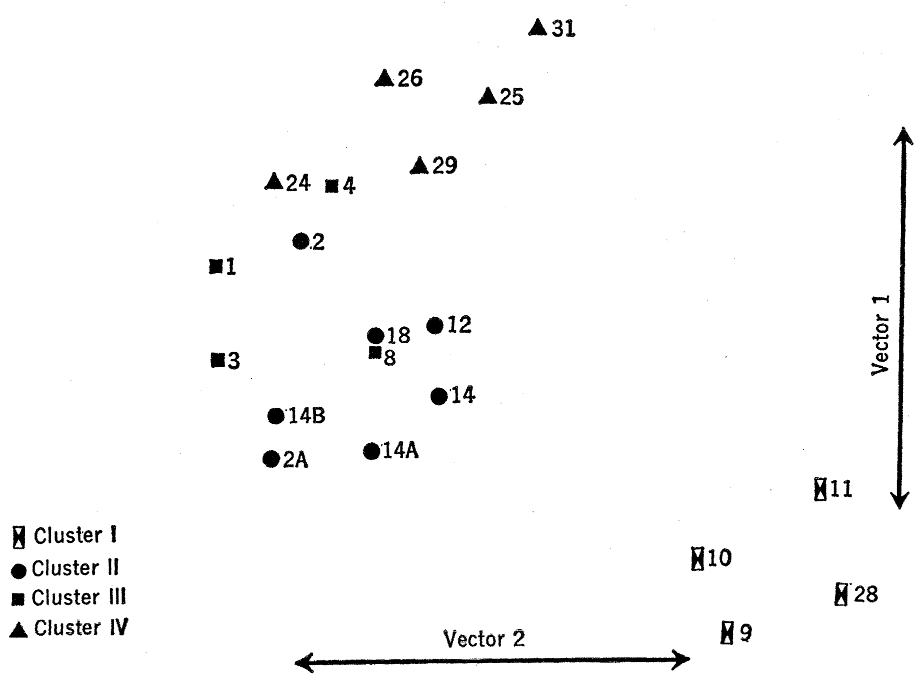
Fig. 1. Dendrogram showing arrangement of sites derived by cluster analysis.

end result of which may be seen as an indication of taxonomic distance between entities (sites). The clustering procedure first selects the pairs of sites that are most similar to each other, then progressively builds up clusters by adding other sites to the nearest existing clusters. The results appear in the form of a dendrogram (Fig. 1). There is no necessary implication of "genetic" relationship in the dendrogram; it should be interpreted merely as an indicator of taxonomic distance, not as a family tree.

The dendrogram produced by the first analysis exhibits the following patterns (Fig. 1). Four main clusters are indicated: (cluster I) sites 28, 10, 11, and 9; (cluster II) sites 18, 14, 2A, 2, 12, 14B, and 14A; (cluster III) sites 4, 8, 3, and 1; and (cluster IV) sites 31, 29, 26, 25, and 24. Cluster I represents the workshops proposed in the original, traditional evaluation. The grouping is consistent with that evaluation and with other data not included in the program input. Cluster II includes the sites believed to be part of an occupation concerned with the exploitation of vegetable resources rather than hunting. All its sites share large numbers of traits characterizing heavy core tools, and most sites share milling-stone characters. These features are consistent with a cultural emphasis on processing plant resources. However, there are several internal relationships that must be considered within the cluster. Sites 14 and 18 are very similar to one another and are set apart somewhat because they lack thinned, bifacially flaked tools. Sites 14A and 14B are likewise paired and set apart. Both share milling stones, a variety of heavy core tools, and some similar knife and projectile point elements. The dendrogram shows that site 12 is the most similar to the 14A–14B pair. This similarity reflects shared core tool elements and milling stones but overrides some obvious differences in projectile point forms. Site 2 next joins the 14A–14B–12 cluster. This relationship is consistent with the intuitive placement of site 2 in the local sequence. Finally, site 2A joins with the other four sites at a rather low level of linkage. As originally proposed in the intuitive evaluation, sites 2A and 14A would not have been included in this grouping but would have represented a separate, though related, cluster on the basis of similar projectile points. Otherwise, agreement within cluster II is consistent with the initial, traditional grouping.

Cluster III is set aside primarily because of minimum input data. All included sites have incomplete inventories. All sites in cluster IV reflect a pattern of seasonal site occupation with a probable hunting emphasis. Sites 26 and 29 are very similar. Site 31 is closely related to this pair and would perhaps have been even closer if its inventory were more complete. Sites 24 and 25 group with these three sites at a rather low level of similarity, and their relationships are consistent with the intuitive classification and with additional data not included in the present study.

As a second method of analysis, multidimensional scaling was carried out (*5*). This technique attempts to fit configurations based on the rank order of the similarity scores into some *n*-dimensional space. The space with the



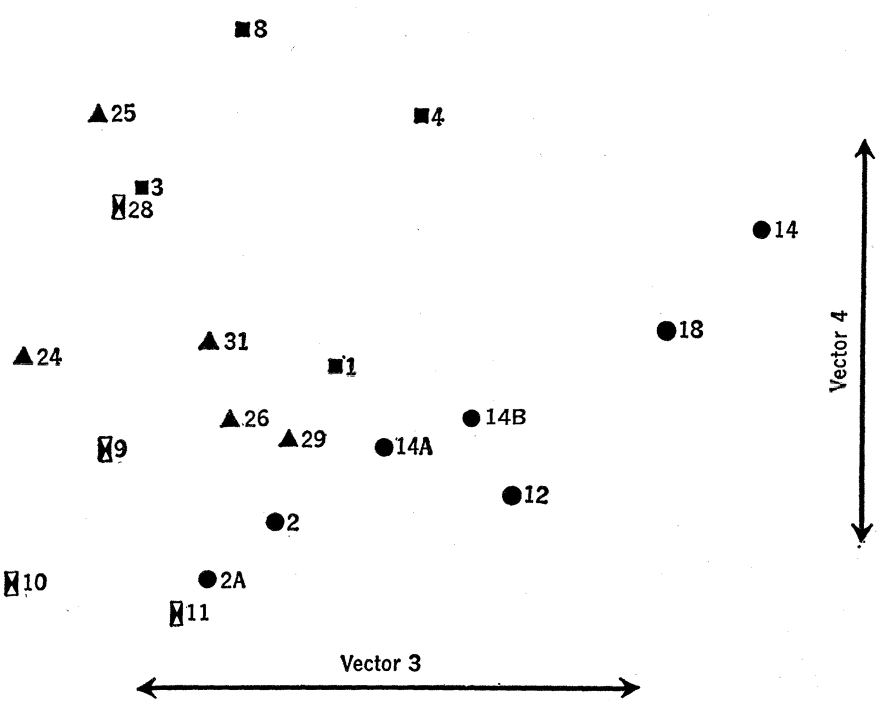


Fig. 2 (top). Plot of site distributions obtained by multidimensional scaling (vectors 1 and 2). Fig. 3 (bottom). Plot of site distributions obtained by multidimensional scaling (vectors 3 and 4).

fewest dimensions into which the configuration will fit, as indicated by a measure of strain or stress, is seen as the simplest and best "explanation" of that configuration. The resulting dimensions may be treated as vectors or factors determining the configuration. Since the technique used here is sensitive to initial input configurations, a varimax rotation of the factor analysis of the principal axes of the matrix was used for the initial data configuration (*6*).

The resulting configuration seemed to fit reasonably well into a four-dimensional space (7.4 percent stress) (*7*). The clusters produced by dendrograms (Fig. 1) are so indicated on the multidimensional plots (Figs. 2 and 3) by appropriate symbols. Figure 2 is a plot of the configuration on the most important vectors (first and second dimensions). Except for cluster III (low input sites), the patterning of the sites on these two vectors agrees with the clusters represented by the dendrogram. In general, the vertical dimension (vector 1) can be seen as a shift in subsistence orientation occurring through time. Starting at the bottom of Fig. 2, there is a major shift from a probable hunting economy based on the use of large leaf-shaped projectile points and knives (cluster I) (*8*) to an economy associated with more refined leaf-shaped projectile points and milling-stone elements (sites 2A and 14A). These, in turn, are followed by the remainder of the sites in clusters II and III, where milling-stone elements are a common item along with large numbers of core tools (choppers, hammers, planes, and heavy flake scrapers). Typically, the members of clusters II and III (except 14A and 2A) include only occasional leaf-shaped knives. In contrast, triangular knife forms are found with increasing frequency. Cluster IV is located at the top of vector 1 and includes sites that have large numbers of triangular knives (some with concave bases). Milling-stone elements, many well-made scrapers, and numerous projectile points are found in this grouping. Cluster IV sites have the most diverse inventories of the sites included in the study.

Because the input data emphasize the total artifact inventory, in which probable time-sensitive forms are in the minority, the placement of sites *within* each cluster is not in complete agreement with their suggested chronology. However, the order of the different site groups themselves definitely agrees with the chronology as it is now defined. Cluster I is almost certainly part of a larger Andean hunting pattern generally accepted as being some 8000 years old (*9*). Sites 14A and 2A of cluster II have been dated at 6000 to 7000 years of age (*10*), and sites included within the remainder of cluster II and in cluster III have radiocarbon dates indicating occupations from 4800 to 3700 years ago (*11*). The sites of cluster IV have not yet been dated exactly, but there is general agreement among Chilean archeologists that they represent a more recent preceramic period (*12*).

The second dimension (horizontal axis, Fig. 2) separates cluster I markedly from all other clusters. Indeed, the artifacts of cluster I sites suggest a substantially different cultural tradition; all these sites are also workshop sites characterized by the by-products of lithic artifact manufacture. The third and fourth dimensions (Fig. 3), although not so informative or important as the first two vectors, do separate the sites on the basis of "core-toolness" and "amount of data," respectively. The horizontal direction (dimension 3) has a Spearman rank-order correlation of 0.87 (significant at less than 0.001), with the ranking based on the ratio of core tools to the total lithic assemblage. Sites 14 and 18 contain heavy core tools almost exclusively, whereas sites 9, 10, and 11 are dominated by thinned bifaces and have few core tools. The vertical placement of the sites on Fig. 3 (dimension 4) has a rank-order correlation of 0.74 (significant at less than 0.001), with the ranking determined by the number of absent, or no comparison, input characters. Thus less is known about sites 4, 8, 3, 28, and 25 than about the remaining sites.

In sum, the distribution of the 20 Chilean sites by both the clustering and multidimensional scaling techniques agrees remarkably well with the intuitive evaluations of the same site relationships. The results of this pilot study can be viewed as a verification of the nonstatistical classification attempted beforehand and as an illustration of the utility of multivariate analysis. In spite of the general agreement, our conclusion is that computerized statistical analysis of archeological assemblages is definitely preferable to traditional intuitive techniques of site classification. By weighting many characters equally, the pilot study avoids possible bias in choosing traits for cluster definition. The approach allowed us to determine inductively which traits cause groupings and are therefore "key" attributes for the data set, thus avoiding the pitfalls of a priori assumptions about trait importance. Further, the usefulness of the pilot study points to the feasibility of doing a future comparative analysis with complete site inventories and artifact counts. The more complete study will help assess the reliability of pilot studies, such as the present one, made with only partial data inventories.

D. L. TRUE
R. G. MATSON

*Department of Anthropology,
University of California,
Davis 95616*

References and Notes

1. D. L. True, L. Núñez, P. Núñez, *Amer. Antiquity* **35**, 170 (1970).
2. A description and detailed discussion of this component is in preparation.
3. M. Beckner, *The Biological Way of Thought* (Columbia Univ. Press, New York, 1959).
4. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy* (Freeman, San Francisco, 1963), p. 182.
5. P. Shepart, *Psychometrica* **27**, 125 (1962); *ibid.*, p. 219; J. B. Kruskal, *ibid.* **29**, 1 (1964).
6. H. H. Harman, *Modern Factor Analysis* (Univ. of Chicago Press, Chicago, ed. 2, 1967).
7. J. B. Kruskal, *Psychometrika* **29**, 115 (1964). Stress is seen as a measure of goodness of fit. Low stress refers to the fact that the actual rank order of the interpoint distances of the configuration is very close to the reverse rank order of the similarity coefficients.
8. Artifacts typical of these complexes or industries are described in the literature as Ayampitin [A. Gonzales, *La Historiografía de la Gruta de Intihuasi i Sus Relaciones con Otros Sitios Precerámicos de Sudamérica* (Revista del Instituto de Antropología, Universidad Nacional de Cordova, Cordova, 1960), vol. 1, pp. 1–292] or as Puripica or Tulan [M. Orellano, *El Precerámico en el Desierto de Atacama* (Chile) (University of Madrid, Madrid, 1963), vol. 9, pp. 6–35].
9. E. Lanning, *Quaternaria* **8**, 133 (1966); A. Gonzales, *Congresso Internacional de Americanistas XXXVI, Actas i Memorías* (1966), pp. 15–41.
10. Sample WSU-987, 6430 ± 130 years ago; sample GAK-2432, 6830 ± 270 years ago; sample GAK-2205, 5970 ± 120 years ago.
11. Sample GAK-2529, 4780 ± 130 years ago; sample GAK-2433, 3910 ± 170 years ago; sample UCLA-1293, 4690 ± 80 years ago.
12. The stratigraphy at Intihuasi Cave (northern Argentina) confirms this trend for the region at large and also has a transition from early leaf-shaped projectile point forms to more recent levels with triangular knives and projectile point types [A. Gonzales, *La Historiografía de la Gruta de Intihuasi i Sus Relaciones con Otros Sitios Precerámico de Sudamérica* (Revista del Instituto de Antropología, Universidad Nacional de Cordova, Cordova, 1960), vol. 1, pp. 1–292].
13. This work in northern Chile was done under the auspices of the University of California/University of Chile Cooperative Program in 1966–67. Financial support for the field research was provided by the Ford Foundation, and the analyses of the recovered material were supported by a faculty research grant from the University of California, Davis. We thank L. Johnson, Jr., for reading the original version of this report and making helpful suggestions.

22 May 1970