BNL 44411
CONF-9004166--1

Received by OSTI

APR 1 0 1990

# "The Efficiencies and Error-rates of Euclidean and Mahalanobis Searches in Hypergeometries of Archaeological Ceramic Compositions

by

Garman Harbottle

Department of Chemistry

Brookhaven National Laboratory

Upton, NY 11973 USA

BNL--44411

DE90 009201

Abstract

Underlying all archeaological investigations where provenience is inferred from multivariate analytical characterization is the assumption that the group of objects analyzed form a unique set in multivariate hyperspace. In this paper large archaeological ceramic data sets are used to test this assumption. The results clearly demonstrate the dangers of using too few variates, and the power of multivariate matrix-inversion (Mahalanobis) techniques, in the formation of adequately unique groups.

**DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

---

## DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

## 1. Introduction

The idea, that specimens of natural raw materials and the archaeological artifacts made from them would display in their analytical compositions characteristics related to their origins, is hardly new. Through the 19th century scientists like Göbel (1842), Damour (1864, 1865, 1866) and Helm (cited in Caley (1976), Beck et al. (1971)) sensed the intimate connection between chemical analysis ans origin of archaeological artifacts. The work of Richards at Harvard (Richards, 1985) on the analysis of Athenian pottery is even more explicit: consistency in composition implies identity in origin.

One can sense that the idea is almost intuitive: and in provenience research today it is accepted almost on faith that a similarity in overall chemical composition suggests a possible geographic locational relationship. There has been in fact much fruitful research based on this hypothesis. In this paper a detailed study will be made of a number of test cases, and an attempt made to draw some mathematical and probabilisitic inferences to support this concept, which has been called the Provenience Postulate (Weigand, Harbottle and Sayre, 1977): "There exist differences in chemical composition between natural sources that exceed, in some recognizable way, the differences observed within a given source".

Over the past eighteen years the Brookhaven Archaeometry Project founded by R. W. Dodson and E. V. Sayre has carried out an extensive series of archaeological research studies, involving the analysis of many thousands of ancient ceramics pertaining to cultures in a great diversity of geographic settings in the New World, the Mediterranean and Middle East and China. Artifacts from many periods of human civilization have been sampled, and in some hundreds of cases, actual source clays located in or near archaeological zones were also analyzed. Typically, 20-24 chemical elements were determined for each specimen utilizing the technique of neutron activation analysis (Bieber et al. 1976).

Although attention heretofore has focussed on the properties and multivariate separation of individual pottery groups of particular research interest, the existence of a large data set covering many individual sources now gives us the opportunity to carry out some statistical studies of the compositional uniqueness of those specified archaeological groups; that is to say, of the degree of accidental overlap of ceramic compositions from the whole universe of ceramics with the sub-volume of that universe occupied by the specified group.

## 2. The Data Bases

The data bases on hand at Brookhaven are organized, for the sake of fitting into existing software, into packages of 4999 samples or less. These bases are as follows:

Old World

The data file consists of 4320 samples drawn from research projects on material from the Bronze Age onward. Analyzed ceramics from sites in southern France, especially around Marseilles, Greece (many regions), the Aegean, Cyprus, Turkey, Syria, Lebanon, Palestine (numerous regions and sites), Jordan, Mesopotamia, Iran, extending to Pakistan, Egypt, Sudan and North Africa are included.

New World

Two data files are employed, each containing the full 4999 analyzed specimens, or just under 10000 altogether. Pottery from many epochs includes specimens from numerous sites in the American Southwest, many regions of Mexico including heavy representation from the Valley of Mexico, Puebla, the Gulf Coast, Oaxaca, Yucatan, Chiapas extending into the Maya regions of Belize, Guatemala and Honduras. Some Caribbean material is also present.

China

A data file of 60 specimens of Neolithic pottery from the A. M. Sackler Collection is used. Although the number of analyses is small, it provides a compositional group that will allow a particularly interesting uniqueness test to be carried out.

Taxonomic manipulations such as clustering and search operations can be carried out in data bases organized in two broadly distinct types of procedural universes, or hyperspaces, the Eulidean and the Mahalanobis.

## 3. The Euclidean Hyperspace

We begin by constructing a Euclidean hyperspace of n dimensions (Harbottle, 1976; Sneath and Sokal, 1973). The n axes are all orthogonal, and often represent log (concentration) units, thus, a point in the hyperspace represents the log-transformed analytical data for n chemical elements in a particular specimen. A group of specimens having similar overall compositions would be represented by a cluster of points in the hyperspace. Two distinct archaeological ceramic groups are then two clouds or clusters that don't overlap. The Euclidean (i.e. "straight-line") distance between two specimen points A and B is

$$ED_{A,D} = \left[ \sum_{i=1}^{n} (A_i - B_i) \right]^{1/2} \qquad (1)$$

where the index i runs over the n elementary log concentration coordinates $A_i$ and $B_i$. In vector notation the squared Euclidean distance is given by (Solomon, 1971)

$$(ED_{A,B})^2 = (A - B)'(A - B) \qquad (2)$$

where A and B are column vectors each with n rows containing the log-transformed analytical data for individuals A and B: the difference row vector (A - B)' multiplies its transpose (A - B) to give the scalar squared distance.

If the data for a particular element in a specimen happens to be missing, then the $i^{th}$ term referring to that element would be meaningless, and the $(A_i - B_i)$ term for that element in Eqn. 1 would have to be dropped. This would, however, make the two samples A and B appear to be closer together than they deserve. The solution is to define a "mean Euclidean distance", $MED_{A,B}$, which simply averages all the m available $(A_i - B_i)$ terms, $(0 < m \leq n)$ (Sneath and Sokal, op. cit., p. 124).

$$MED_{A,B} = \left[\frac{1}{m} \sum_{i=1}^{n} (A_i - B_i)^2\right]^{1/2} \quad (3)$$

Note that with the employment of log concentration units $A_i - B_i$ equals the log of the ratio of concentrations of the $i^{th}$ element in samples A and B. If natural logs are used, $\log_e(X) \approx X - 1$, (for values of X not too different from 1.0) hence eqn. 3 in this case calculates approximately the root-mean-square average of the deviations from unity of the ratios of concentrations of the different elements taken one-by-one in samples A and B. Note also that, for a given n-hyperspace, the $ED = (MED)(n)^{1/2}$, if n = m, as usually happens. One of the goals of this study is to see in an actual case how effectively the use of Mean Euclidean Distance compensates for missing data (Fig. 1).

What we explore in this paper is to what extent a compositional cluster is unique, that is, in the whole hyperspace of archaeological ceramic compositions, are there other clusters that overlap this one? Are there stray samples that are found in this restricted volume of the hyperspace, and could thus be mistaken for a cluster member? In other words, how good is chemical characterization for establishing provenience clustering?

It will be helpful to define the center, of "centroid" of the cluster of designated compositions. The centroid is an artificial composition or point in hyperspace having for each chemical element the mean value of that element averaged over all the members of the set. Usually a geometric mean is taken. Thus for a cluster consisting of samples A, B,...P of which k have been analyzed for the $i^{th}$ element, the analytical values being $A_i$, $B_i$, etc., the geometric mean centroid value for the $i^{th}$ element, $Z_i$, is

$$Z_i = (\prod^{k}(A_i)(B_i)...(P_i)^{1/k} \quad (4)$$

and similarly for each of the elements i present in the data matrix. One may then consider distances from the centroid of the group outward to the different group members, speaking either of "real"

NOT MICROFILMED
PAGE

$$MED_{A,B} = \left[\frac{1}{m} \sum_{i=1}^{n} (A_i - B_i)^2\right]^{1/2} \quad (3)$$

Note that with the employment of log concentration units $A_i - B_i$ equals the log of the ratio of concentrations of the $i^{th}$ element in samples A and B. If natural logs are used, $\log_e(X) \approx X - 1$, (for values of X not too different from 1.0) hence eqn. 3 in this case calculates approximately the root-mean-square average of the deviations from unity of the ratios of concentrations of the different elements taken one-by-one in samples A and B. Note also that, for a given n-hyperspace, the $ED = (MED)(n)^{1/2}$, if n = m, as usually happens. One of the goals of this study is to see in an actual case how effectively the use of Mean Euclidean Distance compensates for missing data (Fig. 1).

What we explore in this paper is to what extent a compositional cluster is unique, that is, in the whole hyperspace of archaeological ceramic compositions, are there other clusters that overlap this one? Are there stray samples that are found in this restricted volume of the hyperspace, and could thus be mistaken for a cluster member? In other words, how good is chemical characterization for establishing provenience clustering?

It will be helpful to define the center, of "centroid" of the cluster of designated compositions. The centroid is an artificial composition or point in hyperspace having for each chemical element the mean value of that element averaged over all the members of the set. Usually a geometric mean is taken. Thus for a cluster consisting of samples A, B,...P of which k have been analyzed for the $i^{th}$ element, the analytical values being $A_i$, $B_i$, etc., the geometric mean centroid value for the $i^{th}$ element, $Z_i$, is

$$Z_i = (\prod^{k}(A_i)(B_i)...(P_i)^{1/k} \quad (4)$$

and similarly for each of the elements i present in the data matrix. One may then consider distances from the centroid of the group outward to the different group members, speaking either of "real"

DO NOT MICROFILM

$$MED_{A,B} = \left[\frac{1}{m} \sum_{i=1}^{n} (A_i - B_i)^2\right]^{1/2} \quad (3)$$

Note that with the employment of log concentration units $A_i - B_i$ equals the log of the ratio of concentrations of the $i^{th}$ element in samples A and B. If natural logs are used, $\log_e(X) \approx X - 1$, (for values of X not too different from 1.0) hence eqn. 3 in this case calculates approximately the root-mean-square average of the deviations from unity of the ratios of concentrations of the different elements taken one-by-one in samples A and B. Note also that, for a given n-hyperspace, the $ED = (MED)(n)^{1/2}$, if n = m, as usually happens. One of the goals of this study is to see in an actual case how effectively the use of Mean Euclidean Distance compensates for missing data (Fig. 1).

What we explore in this paper is to what extent a compositional cluster is unique, that is, in the whole hyperspace of archaeological ceramic compositions, are there other clusters that overlap this one? Are there stray samples that are found in this restricted volume of the hyperspace, and could thus be mistaken for a cluster member? In other words, how good is chemical characterization for establishing provenience clustering?

It will be helpful to define the center, of "centroid" of the cluster of designated compositions. The centroid is an artificial composition or point in hyperspace having for each chemical element the mean value of that element averaged over all the members of the set. Usually a geometric mean is taken. Thus for a cluster consisting of samples A, B,...P of which k have been analyzed for the $i^{th}$ element, the analytical values being $A_i$, $B_i$, etc., the geometric mean centroid value for the $i^{th}$ element, $Z_i$, is

$$Z_i = (\prod^{k}(A_t)(B_i)...(P_i)^{1/k} \quad (4)$$

and similarly for each of the elements i present in the data matrix. One may then consider distances from the centroid of the group outward to the different group members, speaking either of "real"

Eulidean distances ED (Eqn. 1) or, the equivalent mean Eucildean distances (Eqn. 3), which are smaller by $m^{1/2}$.

The advantage of using the mean Euclidean distance is that we can compare distance results obtained in two different hyperspaces having n and m dimensions, provided the distribution of mean distances is not much different in the two. To be more concrete: suppose we have calculated an MED between two specimens A and B in an n-dimensional hyperspace. Will we get a much different MED if we drop the number of elements considered from n down to m; let us say from 17 down to 13?

In Figure 1 is shown a series of distributions of number of cases per MED interval vs MED, for differreing numbers of variates or dimensions n. To obtain these curves, a tightly-bound, well-defined group of ceramic compositions, Attic Black-glaze (Fillieres et al., 1983) was taken. The centroid was calculated, and for each hyperspace of n dimensions (n = 17,10,6 and 3) the distance matrix of MED's from the centroid was calculated and the distribution of distances plotted. It will be seen that the larger number of variates (17,10) all show, within the rather limited statistical precision, similar distribution curves, justifying the use of MED to cover cases where the analysis is missing in a few elements. A second feature is that for the larger numbers of variates, the distributions seem to resemble a normal distribution (Sneath and Sokal, 1973: 197). Another important feature is that for the number of variates or dimension n>6 there is a region near the origin in which there are no cases, the size of the region (in MED units) tending to increase with increasing n. This can be explained as a hypergeometric volume effect, to be discussed below. Finally, for n=3, it will be seen (Fig 1) that the distribution curve extends right down to the origin. This effect, also coming from hypergeometric volume relationships, will be important when we discuss the lack of compositional uniqueness in low-dimensional hyperspaces and the (bad) effect this one has on search and identification procedures. It bears directly on the oft-repeated question, "How many elements should I use in provenience studies involving chemical analysis?"

The effects seen in Fig. 1 are commonly found in other multivariate systems in nature. A case given in Sneath and Sokal (op. cit., p. 286) gives similar distribution curves in 49 and 3 dimensions in the taxonomy of certain strains of bacteria. The distribution curves shown in Fig. 1 are an example of Heincke's Law (Heincke, 1898; Sneath and Sokal, op. cit., p. 305-6) which may be paraphrased as "Every individual of a natural group has about the same distance from the centroid" in multivariate space.

## 4. Volumes of hyperspheres, and skewed distributions

We have spoken about "distances" in hypergeometric space of n dimension (Eqn. 1): such a distance measured radially from a centroid defines a hyperspherical surface containing a volume. If the radius from the centroid goes out to the most distant member of the group, then we may think of a hypervolume "containing" the n-dimensional group. Obviously, for an infinitely numerous group, the containment surface could be that surface which contains 95%, 99% etc., of the group members.

The volume of a hypersphere of n dimensions is given by the general expression:

$$V(n) = (2\pi^{n/2}/n\Gamma(n/2))r^n \qquad (5)$$

where r is the radius, n is the dimensionality and  is the gamma-function.

For high-dimensional hypergeometrics of the type encountered in the multivariate analytical data base, the very rapid increase in volume, going as $r^n$, fully explains the lack of short distances in distribution curves such as Fig. 1. The reason there are no specimens close to the centroid (Fig. 1) is that there is almost no volume there. For low dimensionalities, n=3 for example, the power dependence no longer excludes the possibility of short distances.

This effect of skewing the MED distribution toward the origin for low-dimensional populations has an important consequence in archaeological provenience investigation. If we are searching through a data base for specimens similar to a particular archaeological group, like the Attic Black-glazed ware (Fig. 1), and we use as a criterion Euclidean distance from the Attic

centroid, then we may visualize the following. Starting at the centroid in let us say 17-d space and moving radially in equal increments of MED, there are at first no samples because the volume is so small, even though the density of speciment points--the number per unit volume, is at a peak. As we go outward the number of cases (= density times volume) rises rapidly, then falls off because, with dramatically increasing volume, the edge of the group is reached and no more specimens are encountered. At some futher distance, numbers of other groups begin to be picked up (Sneath and Sokal, 1973: 286). Now if the MED values of these other group distributions are based on low dimensionality distributions, for example n=3, and are thus skewed toward the origin, there may easily occur false attributions of non Attic group members owing to the overlap of the distributions. This overlapping effect will become more evident when we investigate (below) the error-rates in some typical Euclidean searches as a function of dimensionality. These searches will test the quality of uniqueness of the specified ceramic groups.

It often happens that in nature the compositional groups themselves are not spherical in shape, but elongated into (hyper) ellipsoids, which stretch out preferentially along certain directions. This leads us to an entirely different way of describing multivariate space, of defining the uniqueness of archaeological groups and of carrying out search procedures. And it puts in the hands of archaeological ceramists a characterization method of remarkable strength, as will be seen.

## 5. The Mahalanobis Hyperspace

As mentioned above, it is quite common for pairs of chemical elements to be correlated in a suite of natural specimens. For example, in Fig. 2, are plotted the analytical data for the elements iron and scandium in three Middle Eastern pottery goups (Brooks et al., 1974; Harbottle, 1976). The correlation of iron and scandium (both trivalent ions, of nearly equal size) in nature is well-known (Turekian et al., 1973). In fact the correlation coefficient r in the red field clays in Fig. 2 is greater than 0.99. In the bivariate plot the red and yellow clays/ceramics emerge as

distinct, separable compositional groups, which demonstrates the need for multivariate data treament that takes into account the phenomenon of correlation.

A circle centered at the intersection of the major axis A and the minor axis B of the red field clay group, if large enough to contain the specimen points in that group, will greatly overlap a similar circle with its center at the centroid of the limestone (yellow) hill clays. For the red field clays the ellipse of containment (it is actually drawn to contain 95% of the points) has only 1/15 the area of the containing circle, and furthermore readily splits those points away from the similar 95% confidence ellipse of the limestone hill clays. Obviously, the spatial efficiency and discriminating ability become greater, the closer the correlation coefficient gets to unity.

With higher-dimensional groups that have many bivariate correlations, the improvement becomes even more dramatic.

Thus far, the cigar shapes of these hyperellipsoids have been described only in the mormal Euclidean multivariate space. But one can define a new space in which at least one of the elongated groups can be made entirely spherical. The procedure is simply to divide the standardized distance from the centroid to any point in space by the standard deviation of the group in that direction: note that in the red field clays group (Fig. 2) as in all groups having bivariate correlations, the standard deviation is a continously varying function of direction relative to the group axes A and B. Sayre (1976) has shown that the square of a transformed Euclidean distance, (the Mahalanobis distance MD), can be calculated from

$$MD = \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - \overline{X}_i) I_{ij} (X_j - \overline{X}_j) \qquad (6)$$

where the $X_i$ and $X_j$ are the specimen positions on the $i^{th}$ and $j^{th}$ of the n (standardized) coordinates while the $\overline{X}_i$ and $\overline{X}_j$ are the corresponding positions of the centroid of the population of specimens in the group. The quantity $I_{ij}$ is the $i^{th}$ - $j^{th}$ element of the inverse variance-covariance matrix for the group. If in the group there is no correlation of any one of the n dimensions with another, then all the matrix elements $I_{ij}$ for $i \neq j$ are zero, while the diagonal

elements are 1 and the Mahalanobis distance MD becomes simply the squared Euclidean distance, $(ED)^2$. Compare Equation (1). In vector notation, the Mahalanobis distance calculation becomes

$$(MD_{A,B}) = (A - B)'(W)^{-1}(A - B) \quad (7)$$

where the column vectors in n rows are defined as above, Eqn. (2), while $(W)^{-1}$ is the inverse variance-covariance matrix. Other references to this transformation will be found in Sayre (1976) who has shown that the axes A and B (Fig. 2) of the ellipse (or in multidimensional space, hyperellipsoid) containing the group (with stardarized coordinates) are in fact the characteristic vectors, or eigenvectors, of the variance-covariance matrix W of that group. The characteristics vectors provide a new coordinate system for a transformed space in which at least for the basic group (here the red field clays) all correlation has been removed. The basic, or core group has thus become spherical: other groups in the transformed space can become more-or-less spherical as well depending on the extent to which their total bivariate correlational structure resembles that of the basic group. A plot of the specimen points in the transformed vector 1 - vector 2 space is shown in Fig. 3. In the new space the groups are separated without overlap. The fall-off of the density of points from the red clay group centroid in the transformed space (Fig. 3) can be described by the distribution function

$$f = Ke^{(-MD/2)} \quad (8)$$

which is analogous to the univariate normal distribution function

$$f = Ke^{-((X-\overline{X})/\sigma)^2/2} \quad (9)$$

where $(X-\overline{X}/\sigma$ is the standardized "distance" from the centroid ($\sigma$ the standard deviation) and MD the Mahalanobis distance (Sayre 1976). MD, as noted above already has the characteristics of being a squared distance. The quantity K is constant dependent on n, the number of variates. By analogy to Eqn. (9), Eqn. (8) assumes multivariate normality.

The multivariate distribution function in Eqn. (8) must not be confused with the distributions of multivariate distances in actual groups plotted in Fig. 1. The former refers to the

radial fall-off in the density of points in the hyperspace, while the latter refers to radial distributions of actual distances from centroids to points.

Having found the characteristic vectors of a particular focal group and constructed a new space having those vectors as axes, the inverse variance-covariance matrix can be ued to transfrom the coordinates of all points to the new system. The focal group is now spherical, and Eqn. 8 can be used, if one assumes multivariate normality, to calculate a probability of group membership, for every point, in the focal group.

## 6. Multivariate Computer Programs

Although there are numerous "package" multivariate programs available (Hill, 1984; Wishart, 1978; Nie et al., 1975) in this research the following in-house programs were used (Bishop, 1985).

SEARCH. This program (written by Bishop in 1983) calculates the mean Euclidean distance from a specified sample (the "reference vector"), using any or all of up to 36 chemical elements, to every one of the remaining samples in the data base. The program then lists all specimens whose mean Euclidean distances do not exceed an operator-specified value, gives their distance values, and also the distances after adjustment for a proportionate change or dilution factor ("best relative fit" (Sayre, unpublished)).

ADCORR. Written by Sayre, this fully multivariate program calculates the variance-covariance matrix and its inverse for a core group of up to 300 members, 36 elements, log concentration data. The eigenvalues and eigenvectors (A and B in Fig. 2) can also be calculated, as well as the correlation matrix. Following the methods outlined in Section 5, the probability of group membership for each member of the core group is then calculated. The program will also consider non-core specimens (in a "comparison group") and calculate the Mahalanobis distance and probability of group membership of each one of these as well.

ADSEARCH. This program, also written by Sayre and modified by Sayre and Bishop (1985), operates in much the same way as ADCORR, except that the "comparison group" can now be very large, in fact, the whole data base. The inverse variance-covariance matrix of the core group is calculated, and then the Mahalanobis distance to every other specimen in the data file. The printout does not give characteristic vector scores but lists only the bare sample identification number and the calculated probability of group membership.

## 7. Uniqueness of Archaeological Ceramic Groups in Euclidean and Mahalanobis Hyperspaces. Search Efficiencies and Error Rates

We can now measure the uniqueness of multivariate archaeological ceramic groups, that quality whose untested assumption underlies the widespread use of analytical characterization in provenience research.

As suggested in Section 1, we can use the following procedure. We consider a coherent, archaeologically well-characterized ceramic group, Attic Black-glaze. We test its multivariate coherence and demand that its members show an appropriate distribution of probabilities P of group membership, as determined by ADCORR: mostly in the range 5-100%, not more than ca. 1 in 20 with a $P < 5\%$, not more than ca. 1 in 100 with $P < 1\%$.

To test for the occurence of non-Attic specimens that would have distances placing them near the centroid of the Attic group (Section 6) we search from the Attic centroid outwards in increasing MED units in a hyperspace that is full of archaeological ceramics of every description, but contains no Attic Black-glaze.

The test hyperspace chosen is the 10000 diverse samples in the New World data base: since Attic Black-glaze is not expected here we would expect, that the New World specimens encountered in this Euclidean search are genuine accidentals, real errors in provenience attribution, hence a measure of the error rate. Similarly, a definitive, well-studied Mesoamerican

compositional group like Thin Orange of Teotihuacan can be searched against the Old World data base. And a pottery group from Neolithic and Shang-dynasy China can be searched against both.

As with the Euclidean, so with the Mahalanobis space, where ADSEARCH is used to find specimens in the opposite hemisphere that accidentally fall into the multivariate transformed group with a certain probability of containment. And in both types of search, both tests of uniqueness, we can explore the error rate as a function of the number and kind of variates taken. In this way we can test the Provenience Postuilate. We can see how well analyses of archaeological ceramics for only a few elements can serve to characterize the archeaological compositional groups.

## 8. Results of Searches

### A. Euclidean Space

i. Attic Black-glaze. The tighly-bound group is described in Fillieres et al. (1983) and has 107 members. The Euclidean results are plotted in Fig. 4. Here, and in all the Euclidean search plots, the open squares represent the distribution of distances from the centroid of the focal group to its own members, which in this case are seen to fall within 0.095 MED units (Eqn. 3). The open triangles are the number from the centroid to any of the 10,000 New World specimens. The open circles are the same, except that the sample in the Attic group that lay farthest from the centroid, i.e., at the very edge of the Attic group, was taken as the point from which distances to the New World specimens was calculated. For the latter search only one file of 4999 specimens was considered.

The plot shows that the Euclidean signature, or fingerprint of the Attic Black-glaze, is essentially unique if 17 variates are taken: there is no overlap with the error curve at the level of 1 in 10,000. In other words, the Attic group is isolated in 17-d space--at least the New World portion of that space. To demonstrate this isolation, we note that at an MED of .150 (Fig. 4), there is a 1% error rate -- i.e., 1% of the data base is retrieved in error. At this point, however, the

volume of 17-d space encompassed is $(.150/.095)^{17} = 2350.$ times larger than the volume needed to enclose the entire Attic Black-glaze group.

Similar plots, but for 10 and 6 variates, are given in Fig. 5 and 6. At 10 variates, there is a small error rate, about 71 cases per 10,000 (less than 1%) in the envelope of distances that totally encloses the focal group. For 6 variates the situation is much worse, and uniqueness has begun to be lost in such a low-dimensional space. It shoud be noted that a large number of analyses of Greek and Cypriote pottery have utilized a method that determines 8 variates. In Fig. 1 where a number of plots of the focal group alone are given, it can be seen that for 3 variates (solid triangles) the Attic distribution tends to skew toward the origin: one would expect other groups in 3-d space to do the same, with a very large resultant error rate coming from the overlap. Compare Sneath and Sokal (1973) Fig. 5-32.

One odd result that comes out of the data plotted in Fig. 5 (and others) concerns the behavior of the error curve (open triangles). If the population density of ceramic points in 10-d space in the vicinity of the Attic centroid were roughly uniform in all directions one would expect the volume effect to cause the error curve to rise by a factor of ca. $2^{10}$ or 1000 in going from MED = 0.075 to MED = 0.155. Instead it increases by less than a factor of 2. This must mean that in 10-d and other high-dimensional spaces there are "clumps" of compositional points encountered with empty spaces between, rather than a uniform "soup" of points.

ii. Nile Alluvium. This group of 85 specimens was analyzed by the late Joan Huntoon (unpublished), and represents late Bronze pottery made of Nile alluvium at or near Tell D'aba, in the delta. The Nile centroid was "searched" against the New World files: the results gave much the same pattern seen with the Attic group.

iii. Red Field Clay. This group of 92 specimens (Brooks et al., 1974) from the Palestinian coastal plain is interesting in that it is the most highly multiply-correlated group of clays and ceramics we have investigated. Fig. 2 shows one of these high bivariate correlations, iron-

scadium, and in Table 2 where general characteristics of the groups are presented it can be seen that the "correlation fraction", i.e., the number of bivariate correlation coefficients exceeding 0.80 divided by the total number, is 0.27, which is high. The error rate plot shows a focal group very stretched-out on the MED axis (open squares) in strong contrast to thte Attic Black-glaze (Fig. 4). Perhaps the difference is to be sought in the great correlation of the red clay, or in the narrower geographic source of the Black-glaze, which may have come from a single clay pit.

In this error-rate plot the accidentals begin to rise sharply when only about 2/3 of the focal group has been contained. Thus, although a potential group member with a short MED to centroid, say 0.05 to 0.10, might be favorably considered to be a group member, one with a longer distance, 0.20-0.24, would have to be regarded with caution, even though there are several actual group members in that range.

iv. Chinese Neolithic. These specimens, analyzed for the A. M. Sackler Foundation, form a relatively compact core group, all samples being contained within an envelope of MED = 0.115 from the centroid for 17 variates. Here we can use both the Old and New World files for error-rate estimation, actually some 13,200 comparisons in all. The results are plotted in Fig. 7 as solid triangles, summing Old and New World files. The point at which the cumulative error rate equals 1% (132 cases) is just at the MED value for containment of the focal group. Thus, the Euclidean signature of this group is more unique than several of the other groups but not so unique as Black-glaze.

vi. Thin Orange: A New World Example. As with the red field clay described in iii above the correlation fraction is high (f = 0.22) and the basic group is very much stretched-out and somewhat displaced from the origin. The projection for error-rate estimation is now against the Old World: 4320 target cases used. As with red clay, the 1% cumulative error is in the midst of the distribution of the focal group. Alhtough it was not tested, one can be pretty sure that the

distributions for 10 and 6 of fewer variates would show higher error rates relative to the focal group, and could not yield worthwhile provenience identification.

B. Mahalanobis Space

It was shown in Section 5 above that the transformation of the Euclidean space to a hyperspace of the same number of dimensions, using the inverse variance-covarience matrix of the analytical data of a particular focal group, can sharply improve the discrimination of that group away from its neighbors.

A Mahalanobis search is carried out in the transformed space, and looks for specimens ("targets") in the whole hyperspace of archaeological ceramic compositions that are similar to the focal group, i..e. would fall within the transformed "sphere" of, for example, 95% confidence. Such targets will, of course, have small Mahalanobis distances (Eqns. 6 and 7). The program ADSEARCH, which carries out these searches expresses the distances as probabilities of group membership (Eqn. 8) similar to the output of discriminant analysis package programs like BMDP7M (Hill, 1984). These results, as in 8A above, are given under the headings of the focal groups employed in ADSEARCH.

i. Attic Black-glaze

In Table 2 is given the results of a Mahalanobis search of the Attic core (after transformation of the n-dimensional hyperspace) as focal group against the New World files. In short, in 9998 New World specimens we see how many would accidentally have exactly the right elementary concentrations to have a finite probability of being contained within the multivariate Attic group, taking account of correlations in that group. Three target containment probability ranges are considered separately, 0.1 to 1%, 1 to 10% and 10% to 100%. If the search is first applied to the Attic group itself, 101 of the samples have probabilities ranging from 10 to 100% and 6 from 4-10% out of 107 total. The result of the Mahalanobis search of the New World data base is that there were no acidental matches whatever in 9998 tries. In other words, to the level of

0.1% probability in a Mahalanobis space base on 15 variates, Attic Black-glaze is a unique, isolated composition. This result alone demonstrates the extraordinary power of the matrix-inversion procedures.

The number and identity of variates employed was then varied to see what the effect would be of reducing the number of elements determined: the results are given in Table 3. We see that for 15, 12, and one of the 7-variate cases, there were no hits out of 9998 tries: the Attic group surely appears to be unique in these hyperspaces. When 7 variates were deliberately chosen to be correlated or uncorrelated, differences were seen. In general the pattern is , increase in number of correlated pairs of elements decreases the number of chance hits as would be expected. For example, in the case of merely 4 highly correlated variates, Sc, Fe, Th, Co, if we disregard the events of < 1% probability, there are still only 5 accidents in 9998. When, however,4 uncorrelated variates were taken (Na, K, Rb, Mn) there were hundreds of targets hit.

ii. Nile Alluvium

No errors were seen in the New World files, i.e. the error rate was less than one in ten thousand even for a probability of group membership as low as 0.1%. In Mahalanobis hyperspace, this composition also appears to be highly unique (Table 2).

iii. Red Filed Clay

Only 13 variates were considered, and these were 7 accidentals with P < 1% and 1 hit with P = 37% out of 9998 (Table 2). When the search was carried out in a Mahalanobis 4-d space with correlated variates, the error rate was very much higher, 120 with P > 5% out of 4999. This rate is many times higher than with the analogous Attic Black-glaze search.

iv. Chinese Neolithic

No targets hit in 9998 New World trials, only 1 with P > 1% in the Old World file (3200 specimens).

v. Marseilles Amphorae

Because the Marseilles group (Fillieres, 1978), amphorae from the Greek period, was one of the largest we have assembled, numbering 155 specimens, it was used to test the effect of the gradual increase in variances and covariances that occurs when the core group size is diminished. In Table 1 the results of these searches are given. It can readily be seen that for a ratio of 3.3 in samples to variates with 15 variates, very good results are obtained: you have analyzed only 50 specimens, and the group can be so well characterized that the extension in hyperpspace is small, and the error rate low. But when the same ratio drops to 2.3 uniqueness begins to be lost. In the final row of Table 1 the ratio has been increased to 7 by the strategy of reducing the number of variates. The results speak for themselves and underscore once again Sneath and Sokal's first neo-Adansonian principle (op. cit., p. 5): "the greater the content of information in the taxa of a classification and the more characters on which it is based, the better a given classification will be".

vi. Thin Orange and Plumbate: Two New World Examples

In 17-d Euclidean space the Thin Orange (of Teotihuacan) group, because of high correlation, was spread out on the MED axis and had a significant error rate, but in the Mahalanobis search against the Old World there were only 3 error hits recorded, all of less than 0.2% probability, out of 4320 tries.

The same Mahalanobis search was made through the New World files. A few hits with probabililties from 15-70% were recorded, of specimens that were not in the original core group. It was found that all had been labelled "Thin Orange" by the responsible archaeologist.

San Juan plumbate pottery is one of the most distinctive, technologically unique and widely-traded ceramics of late pre-columbian Mesoamerica (Dutton, 1943: Neff, 1984). It gets its name from the vitrified gray surface, and its heartland is along the rivers draining the Pacific slope of Guatemala. Again, the results of searches in Mahalanobis 15 and 6-vector spaces demonstrate a high degree of uniqueness of composition (Table 2).

## 9. Conclusions

In this research we have first compiled extensive files or data bases of archaeological ceramic compositions representing a diverse selection of archaeological periods and contexts, uniform only to the extent that within each file only specimens from one of the two great geographic areas, "Old World" and "New World", are included. Well-characterized archaeological, compositional groups in either the "Old World" or the "New World" are then assembled, their centroid compositions calculated, and the Euclidean or Mahalanobis distances of group members from the centroids determined. This information chacacterizes the group in terms of its multivariate properties -- i.e. the compactness, the correlational qualities, and the degree to which one must extend a search from the centroid in order to encompass the group members. The groups were then projected against the data files of the opposite hemisphere -- e.g. the New World files for Old World groups and vice versa. In this way the "error rates" due to accidental matches in composition could be estimated.

One knows, of course, that despite the fact that each file contains a geochemically very diverse group of archaeological ceramic compositions, there is no assurance that the error rate in that file, let us say the New World file, for an Old World group, let us day Attic Black-glaze, would be the same as the error rate for the same group in an Old World file. Where ceramics are distinctive enough, e.g., Black-glaze or The Orange, error rates can be estimated in the files corresponding to the same hemisphere as the group's location, and there does not seem to be any great increase in error rate observable under these conditions.

What is being tested here is the uniqueness f the compositional "fingerprint" or characterization by analytical composition of an archaeological ceramic group. The assumption here tested, that the fingerprint is unique, underlies all provenience research based on matching of chemical compositions, for all kinds of materials. It also underlies the forensic use of pattern recognition to demonstrate common origins.

The data presented here permits a number of conclusions to be drawn:

1) In several well-characterized, tightly-knit groups (Attic Black-glaze, Nile Alluvium), the use of 17 elements results, in Euclidean space, in essentially unique fingerprints. The error-rates are no more than one or two per ten thousand. Even when the elements are reduced to 10, there is still in these cases excellent characterization. When only 6 elements are employed, uniqueness is seriously compromised. Results based on less than 6 elements are essentially meaningless.

2) In several other cases, the Marseilles Amphorae, red field clay of Palestine, Thin Orange etc. there is a small error rate, typically 1-4%, in a Euclidean search utilizing even 17 elements. Reduction of the number of elements will seriously compromise the uniqueness of the compositional signature in groups similar to these. Euclidean methodology employing 7 or 8 elements, or might be generated by an x-ray fluorescence analytical scheme, will be inadequate as a basis for provenience determination.

3) It goes without saying the an increase in the number of variates employed can, in general, only increase the uniqueness of the fingerprint.

The results of the present investigation ought to give great encouragement to the use of other multielement analytical techniques such as Inductively Coupled Plasma Emission in generating data for the numerical taxonomy of archaeological materials. In some recent studies (Gilber and Harbottle, unpublished) the use of commercially available ICP has produced low-cost analyses with useful quantitative determination of 25 elements in ceramics (Colonial-period and 19th Century bricks). All indications are that use of 25 variates ought to permit excellent discrimination of groups in Euclidean space.

4) With the growth of comprehensive data banks of archaeological ceramic analyses -- i.e. files that encompass great numbers of archaeologically-interesting ceramic sites and ware-groups, the question of sourcing a single intrusive specimen becomes important. Such a problem can obviously involve only Euclidean search procedures. The results presented in the Figures related

to typical searches suggest the the single-specimen search may, given a sufficiently large number of variates, produce archaeologically-interesting pairings. Short Euclidean distances cannot prove archaeological affiliation but, when combined with concordance of stylistic criteria, similarity in paste petrography and other attributes may now give strong additional evidence as to long-distance contacts.

5) It was sugested in Section 5. "The Mahalanobis Hyperspace" (above) that the widespread occurrence of bivariate correlation in natural clays and archaeological ceramics could considerably sharpen the uniqueness of the characterization of archaeological groups if one first transformed the space in such a way as to remove correlations from the focal group.

The results of searches in these transformed spaces have been given in Tables 1-3. These methods require a substantial body of analytical data -- at least 30-50 specimens for the core group, to permit construction of a useful variance-covariance matrix. It may readily be seen in Table 2 that at the level of 15 variates, even with only moderate correlation present (e.g., Attic Black-glaze, Nile Alluvium, China, Plumbate) there are almost zero errors observed, even at the level of 0.1% probability. (It should be noted that in actual research projects, a much less stringent level of $P = 5\%$ is often employed). This astonishing sharpness of characterization using matrix-inversion techniques means to the archaeologist that the characterization of an archaeological ceramic group in Mahalanobis space appears to be essentially unique: the fingerprint is one-of-a-kind, to a far greater degree than in the Euclidean space.

6) When the number of variates drops, the uniqueness of the Mahalanobis signature is compromised. With red clay, 4 variates Fe, Sc, Eu, Co, there were about 120 errors with $P > 5\%$ out of 4999 tries, and many more in the range of $1\% < P < 5\%$ (Table 2). To explore the dependence of error in Mahalanobis space on the number and type of variates, the Attic Black-glaze was taken as the core, and variates varied. The results (Table 3) show a general trend up in error rate with decreasing number of variates as might be expected. But there were exceptions, for

example, a case with 9 variates had more error one with 7, which had no clear pattern of correlation. On the other hand, when 7 variates were deliberately chosen to be uncorrelated, the error rate (mis-attributions per 9998 tries) went up. Only when the variates were reduced to 4 or 3 did the rate approach or exceed 1%.

7) It should be noted that among the ceramic groups chosen as test cases (Table 2) several are fine-paste (Attic Black-glaze, Plumbate); others have natural aplastics present in greater or less degree (Nile Alluvium, red clay) one is moderately tempered (Marseilles) and at least one, Thin Orange, is highly tempered.

The presence of temper can in some cases have a diluent effect on many of the trace elements present, and this can produce correlation and a stretching-out of the group in Euclidean space (Fig. 9). However, in the characterization of such groups in Mahalanobis space, the presence of temper seems to have little effect on the error rate (Table 2).

Scientists involved in the characterization of archeaological ceramics through their analytical compositional signatures have had the intuitive feeling that the method worked fairly well; that one could distinguish sherds that lay in or out of particular clusters with some assurance, provided one employed a sufficient number of elements. The present paper demonstrates just how good that characterization can become, and roughly how many elements need be considered, to achieve a reliable result.

## ACKNOWLEDGMENT

This research was carried out at Brookhaven National Laboratory under contract DE-AC02-76CH-00016 with the U.S. Department of Energy. The author thanks H. Neff for helpful comments and R. L. Bishop and E. V. Sayre for extensive help with problems of multivariate

data handling, and R. B. Marr (Applied Math Dept) for much help with the geometries of multidimensional space.

## Literature Cited

Beck, C. W.; Adams, A. B.; Southhard, G. C.; Fellows, C. (1971) in Science and Archaeology, edited by R. H. Brill, pp. 235-240, Cambridge: MIT Press.

Bieber, A. M., Jr.; Brooks D. W.; Harbottle G.; Sayre, E. V. (1976) Archaeometry **18**, 59-74.

Bishop, R. L. (1985) Introdution to Data Analysis with the Smithsonian Archaeometric Research Collections and Records (SARCAR) Facility. Conservation Analytical Laboratory, Smithsonian Institution, Washington, DC.

Brooks D.; Bieber, A. M., Jr.; Harbottle, G.; Sayre, E. V. (1974) "Biblical Studies through Activation Analysis of Ancient Pottery", in Archaeological Chemistry, edited by C. W. Beck, pp. 49, American Chemical Society, Washington, DC.

Caley, E. R. (1951) J. Chem. Educ. **28**, 64.

Caley, E. R. (1967) J. Chem. Educ. **44**, 120.

Damour, M. A. (1864) Comptes rendus **61**, 313.

Damour, M. A. (1864) ibid. **61**, 357.

Damour, M. A. (1866) ibid. **63**, 1038.

Dutton, B. P. (1943) A History of Plumbate Ware. Papers of the School of American Research, No. 31, Santa Fe.

Fillieres, D. (1978) Thesis, University of Paris I Pantheon-Sorbonne, Doctorate 3$^{e}$ Cycle.

Fillieres, D.; Harbottle, G.; Sayre, E. V. (1983) J. Field Archaeology **10**, 55.

Gilbert, A. and Harbottle, G. Research project on the provenience of Colonial Bricks from Eastern U.S.A., in progress, unpublished.

Göbel, F. (1842) Erlangen Ueber den Einfluss der Chemie auf die Ermittelung der Völker der Vorzeit oder Resultate der chemischen Untersuchug metallischer Altertümer insbesondere der in den Ostseegouvernements Volkommenden, Behufs der Ermittelung der Völker, von welchen sie abstammen.

Harbottle, G. (1976) "Activation Analysis in Archaeology", in Radiochemistry, Vol. 3 A Specialist Periodical Report. p. 33, edited by G. W. A. Newton, London, The Chemical Society.

Heincke, F. (1898) Abh. Deutsch. Seefishereivereins **2**, 1-223.

Hill, M. (1984) BMDP User's Digest, BMDP Statistical Software Inc., Los Angeles, CA.

Neff, H, (1984) Thesis, University of California at Santa Barbara, Doctorate.

Nie, N, H,; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K.; Bent, D. H. (1975) SPSS Statistical Package for the Social Sciences, New York, McGraw Hill.

Richards, T. W. (1895) J. Am. Chem. Soc. **17**, 152.

Sayre, E. V. (1976) Brookhaven Procedures for Statistical Analysis of Multivariate Archeometric Data. Brookhaven Report, BNL 21693.

Sayre, E. V (unpublished) E. V. Sayre may be contacted through the Conservation-Analytical Laboratory, Museum Support Center, Smithsonian Institution, Washington, DC 20560.

Sneath, P. H. A.; Sokal, R. R. (1973) Numerical Taxonomy, Chapter 4, San Francisco, W. H. Freeman.

Solomon, H. (1971) "Cluster Anaylsis", in Mathematics in the Archaeological and Historical Sciences, edited by F. R. Hodson, D. G. Kendall, and P. Tautu, Edinburgh, Edinburgh University, p. 62.

Turekian, K. K.; Katz. A.; Chen, A. (1973) Limnology and Oceanography **18**, 240.

Weigand, P. C.; Harbottle, G.; Sayre, E. V. (1977) "Turquoise sources and source analysis", in Exchange Systems in Prehistory, edited by T. K. Earle, and J. E. Ericson, pp. 15-34. New York: Academic Press.

Wishart, D. (1978) CLUSTAN User manual, program Library Unit, Edinburgh, Edinburgh University.

Table 1

Marseilles Amphorae vs New World Data File*

| Size of core, No. of Specimens | P/n≠ | Number of Targets Hit in Error, With a Probability of Group Membership Equal To: | | |
|---|---|---|---|---|
| | | 0.1 - 1% | 1 - 10% | 10 - 100% |
| 155 | 10.3 | 6 | 0 | 0 |
| 100 | 6.7 | 8 | 0 | 0 |
| 75 | 5.0 | 6 | 1 | 0 |
| 50 | 3.3 | 7 | 1 | 0 |
| 35 | 2.3 | 52 | 17 | 1 |
| 25 | 1.7 | 200 | 35 | 4 |
| 35 | 7.0** | 274 | 106 | 48 |

* 9998 Specimens, 15 variates taken except for final entry in Table.

≠ The ratio of the number of specimens P to the number of variates n.

** P/n increased by reducing variates from n = 15 to n = 5. Elements chosen were Ce, Eu, La, Cr, Th, all highly intercorrelated.

**Table 2**

Result of Searches in Mahalanobis Space

| Archaeological Group | Number of Members | Numbers of Variates | Correlation Fraction ≠ | Searched Against: Old World M = 4320 | Searched Against: New World M = 9998 | Number of Error Hits with Containment Probability: 0.1-1% | 1-10% | 10-100% |
|---|---|---|---|---|---|---|---|---|
| Attic Black-glaze | 107 | 15 | 0.04 | | X | 0 | 0 | 0 |
| Nile Alluvium | 85 | 15 | 0.06 | | X | 0 | 0 | 0 |
| Red Clay | 92 | 13 | 0.27 | | X | 7 | 0 | 1 |
| Red Clay | 92 | 4 | 1.00 | | X | about 120 with p > 5% of 4999 | | |
| Marseilles | 155 | 15 | 0.08 | | X | 6 | 0 | 0 |
| Thin Orange | 50 | 15 | 0.22 | X | | 3 | 0 | 0 |
| Plumbate | 44 | 15 | 0.01 | X | | 0 | 0 | 0 |
| Plumbate | 57 | 6 | | X | | 1 | 0 | 0 |
| China | 57 | 15 | 0.02 | X* | | 10 | 1 | 0 |
| China | 57 | 15 | | | X | 0 | 0 | 0 |

≠ Fraction of correlation coefficients exceeding 0.8.

* Old world file had 3200 entries in this search.

## Figure Captions

Figure 1. Attic Black-glaze. Core group distances from centroid plotted as number of cases per 0.01 MED unit vs MED.

Figure 2. Iron-scandium plot of three groups of Middle Eastern ceramics and clays. A = Axis of greatest variance, B = axis of least variance of red field clay group.

Figure 3. Euclidean search of Attic Black-glaze vs New World. In Fig. 3, and also in the other Figures open squares are MED distances of the focal group, open triangles are error hits in files of the opposite hemisphere: see text. In Fig. 3, the open circles are distances of error hits from a sample at the outer edge of the focal cluster (vs. 4999 specimens).

Figure 4. Same as Fig. 4, but 10 variates.

Figure 5. Same as Fig. 4, but 6 variates.

Figure 6. Euclidean Search, a group of Greek-style Amphores made at Marseilles, 17 variates. Searched against New World file.

Figure 7. Euclidean search, a group of Neolithic and Shang-dynasty ceramics from China, searched against both Old and New World files. Filled triangles, total error hits, both hemispheres.

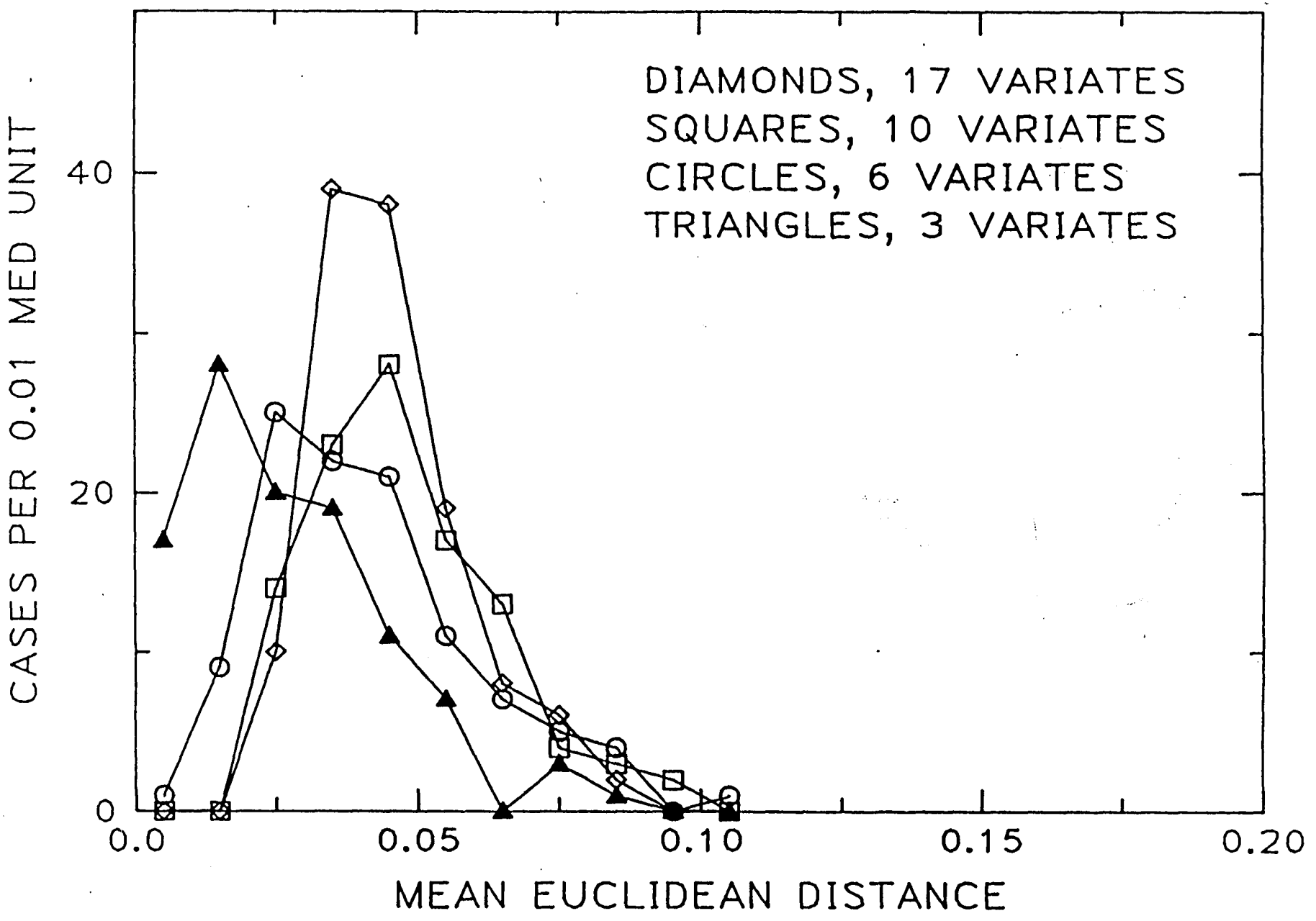
DIAMONDS, 17 VARIATES
SQUARES, 10 VARIATES
CIRCLES, 6 VARIATES
TRIANGLES, 3 VARIATES
CASES PER 0.01 MED UNIT
0
20
40
0.0
0.05
0.10
0.15
0.20
MEAN EUCLIDEAN DISTANCE

Figure 1

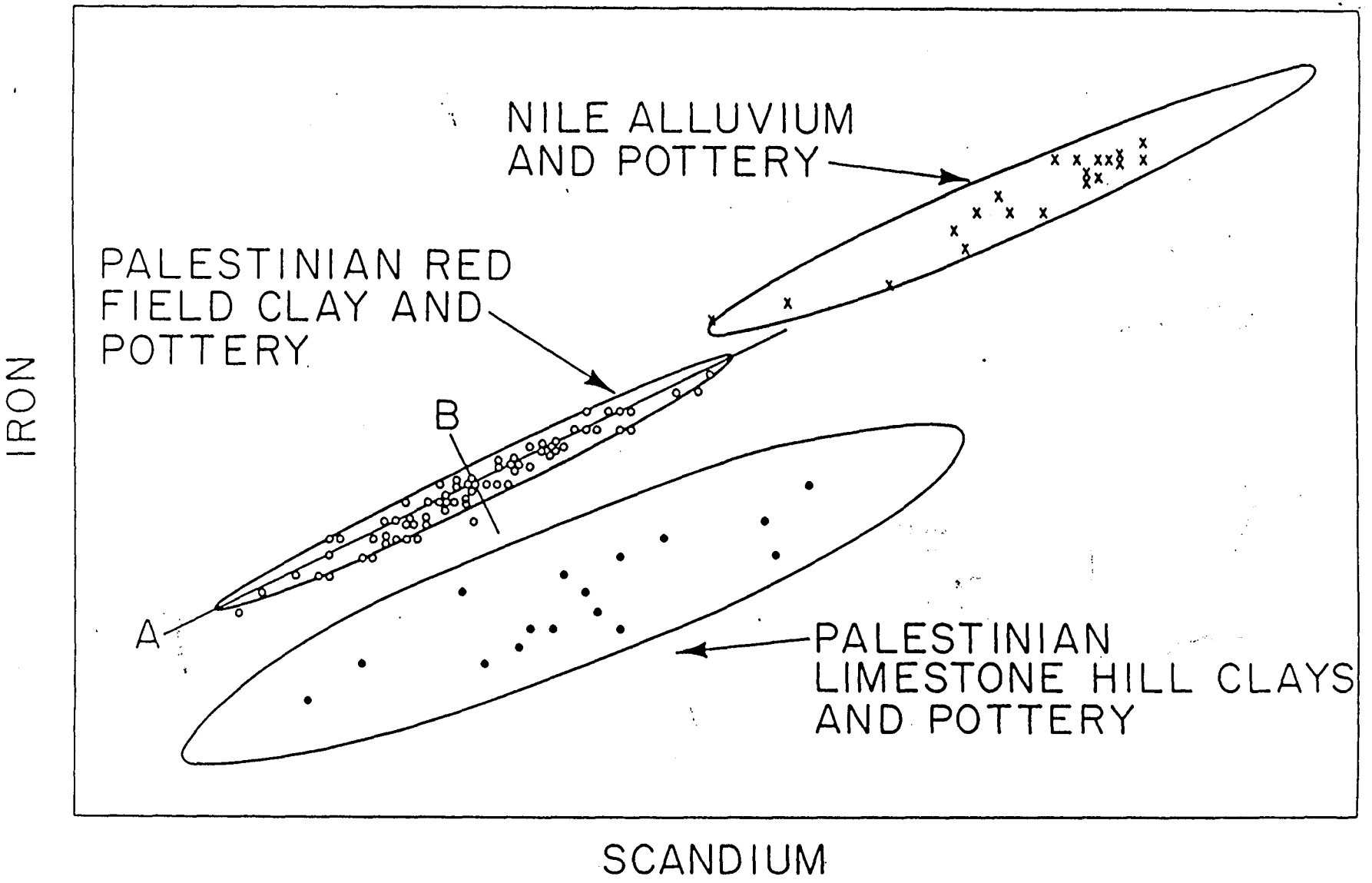
NILE ALLUVIUM
AND POTTERY
PALESTINIAN RED
FIELD CLAY AND
POTTERY
B
A
PALESTINIAN
LIMESTONE HILL CLAYS
AND POTTERY
IRON
SCANDIUM

Figure 2

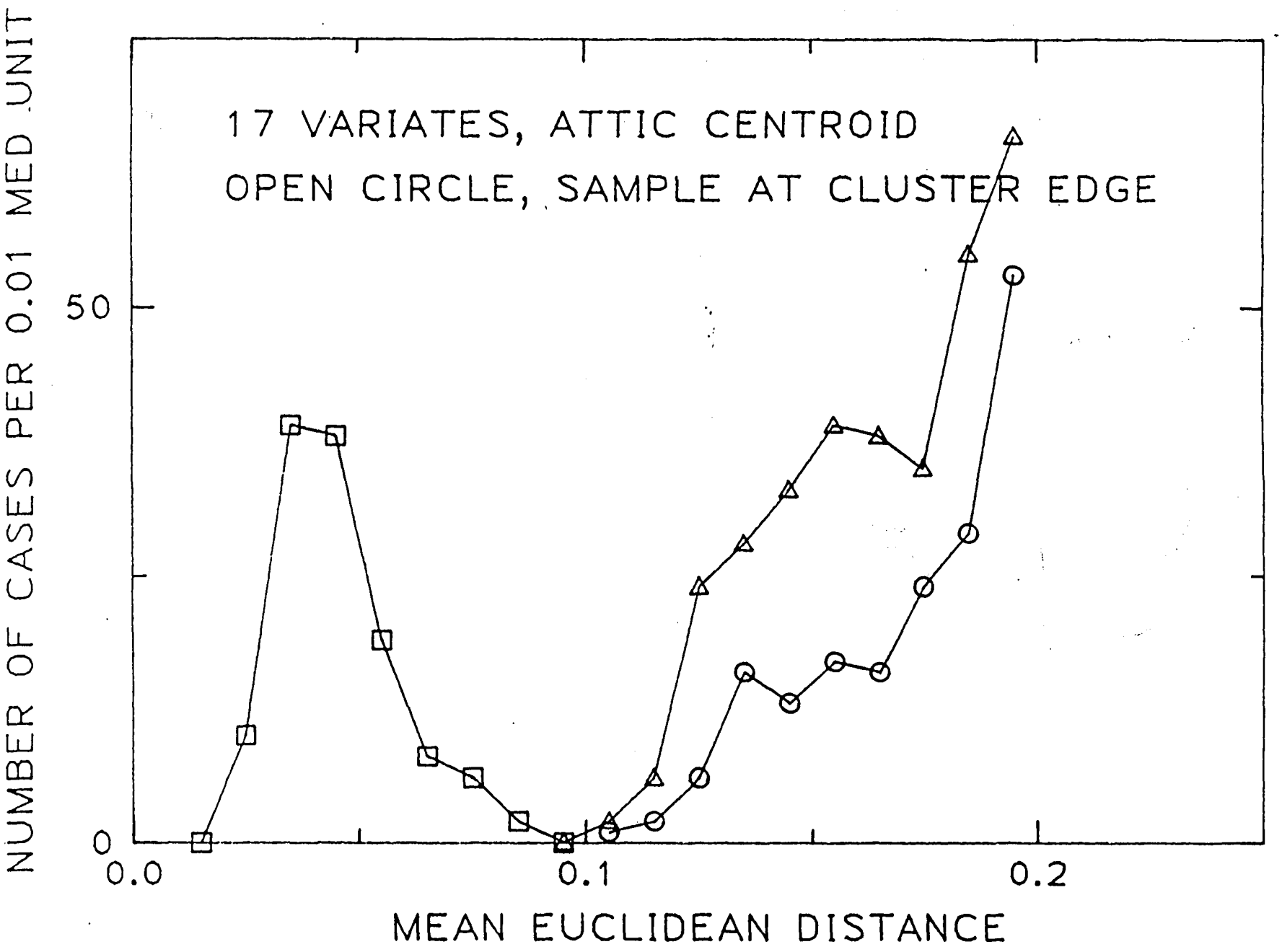
17 VARIATES, ATTIC CENTROID
OPEN CIRCLE, SAMPLE AT CLUSTER EDGE
NUMBER OF CASES PER 0.01 MED UNIT
50
0
0.0
0.1
0.2
MEAN EUCLIDEAN DISTANCE

Figure 13

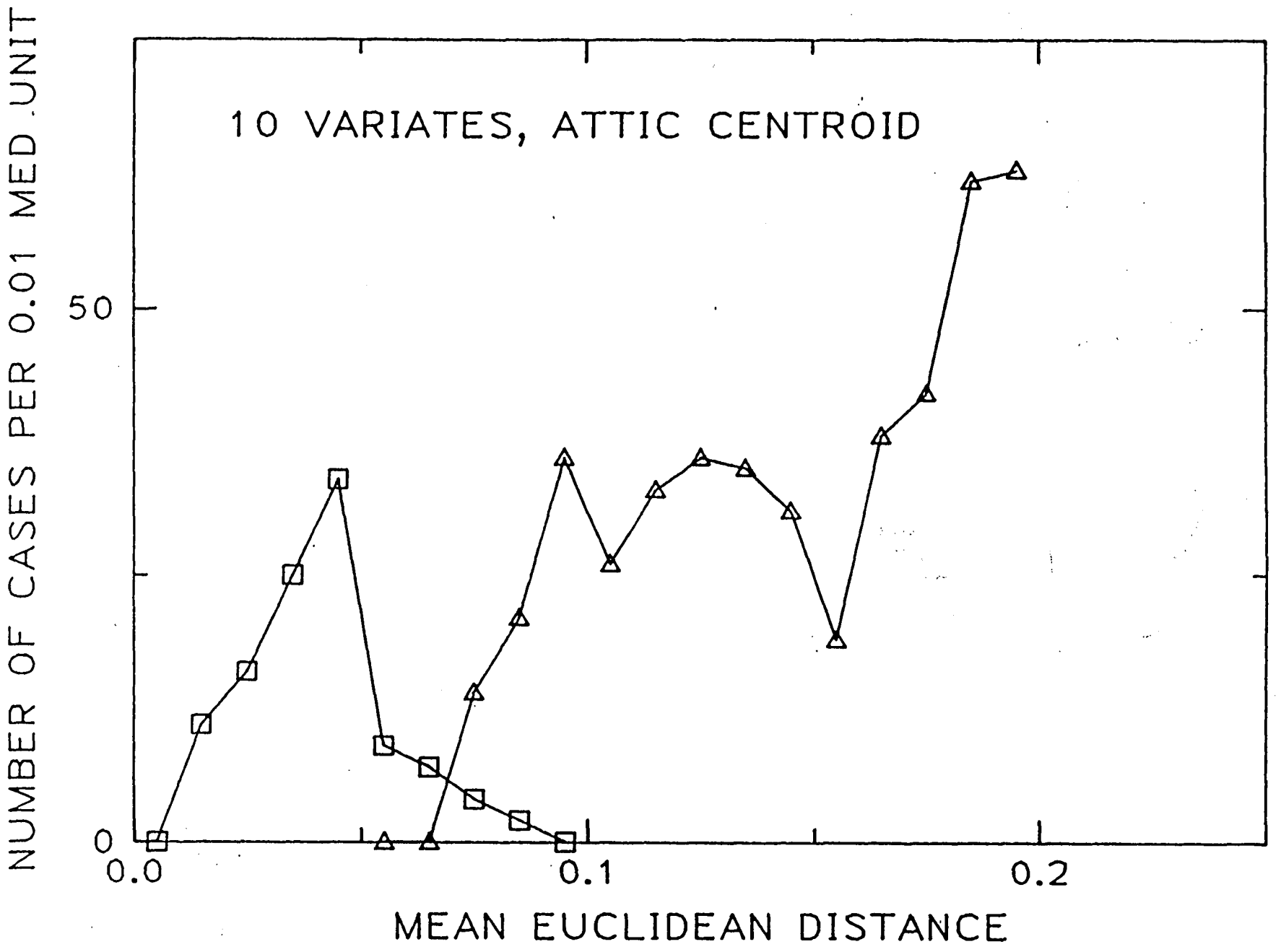
10 VARIATES, ATTIC CENTROID
NUMBER OF CASES PER 0.01 MED UNIT
50
0
0.0
0.1
0.2
MEAN EUCLIDEAN DISTANCE

Figure 54

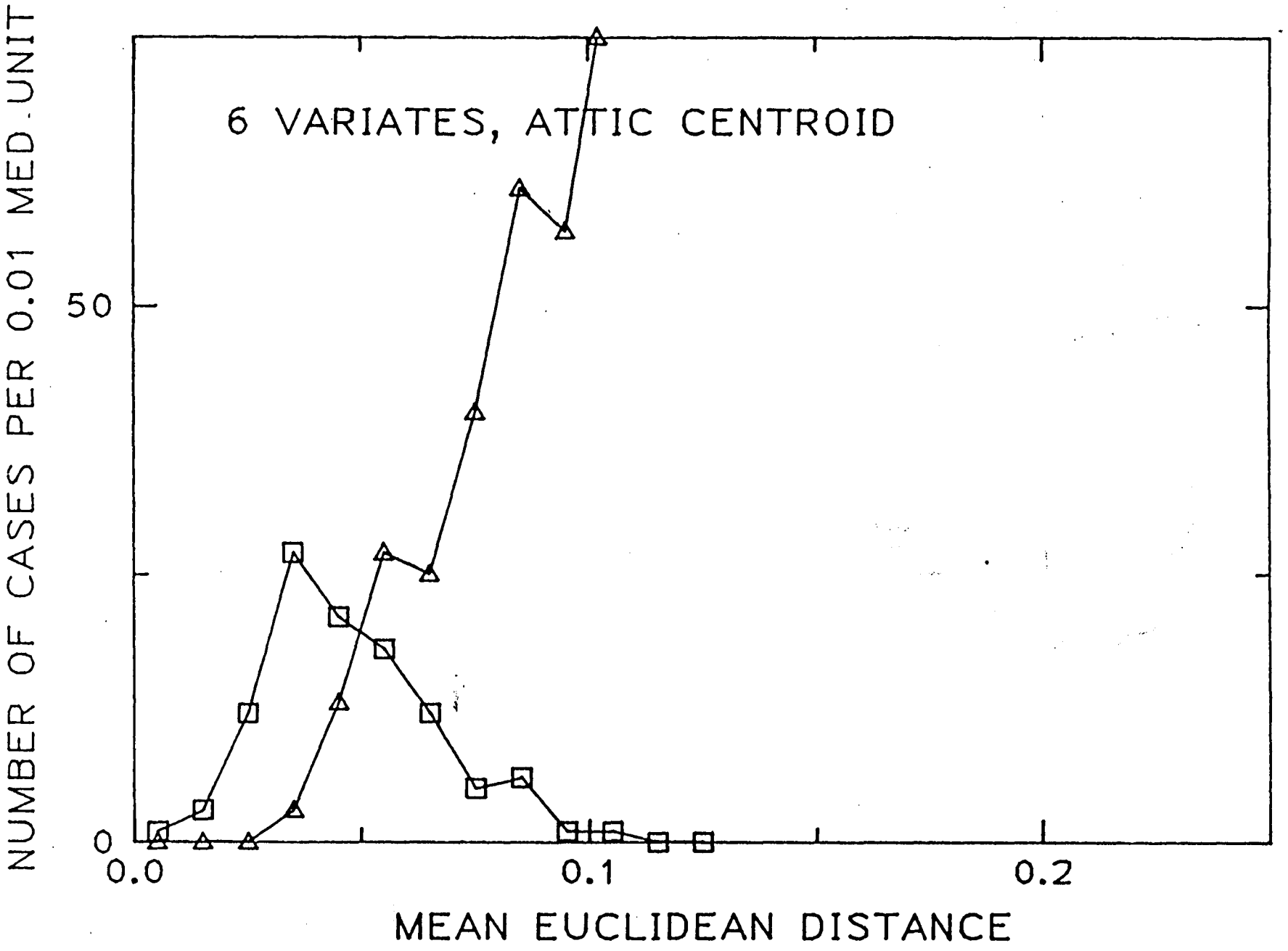
6 VARIATES, ATTIC CENTROID
NUMBER OF CASES PER 0.01 MED.UNIT
50
0
0.0
0.1
0.2
MEAN EUCLIDEAN DISTANCE

Figure 6-5

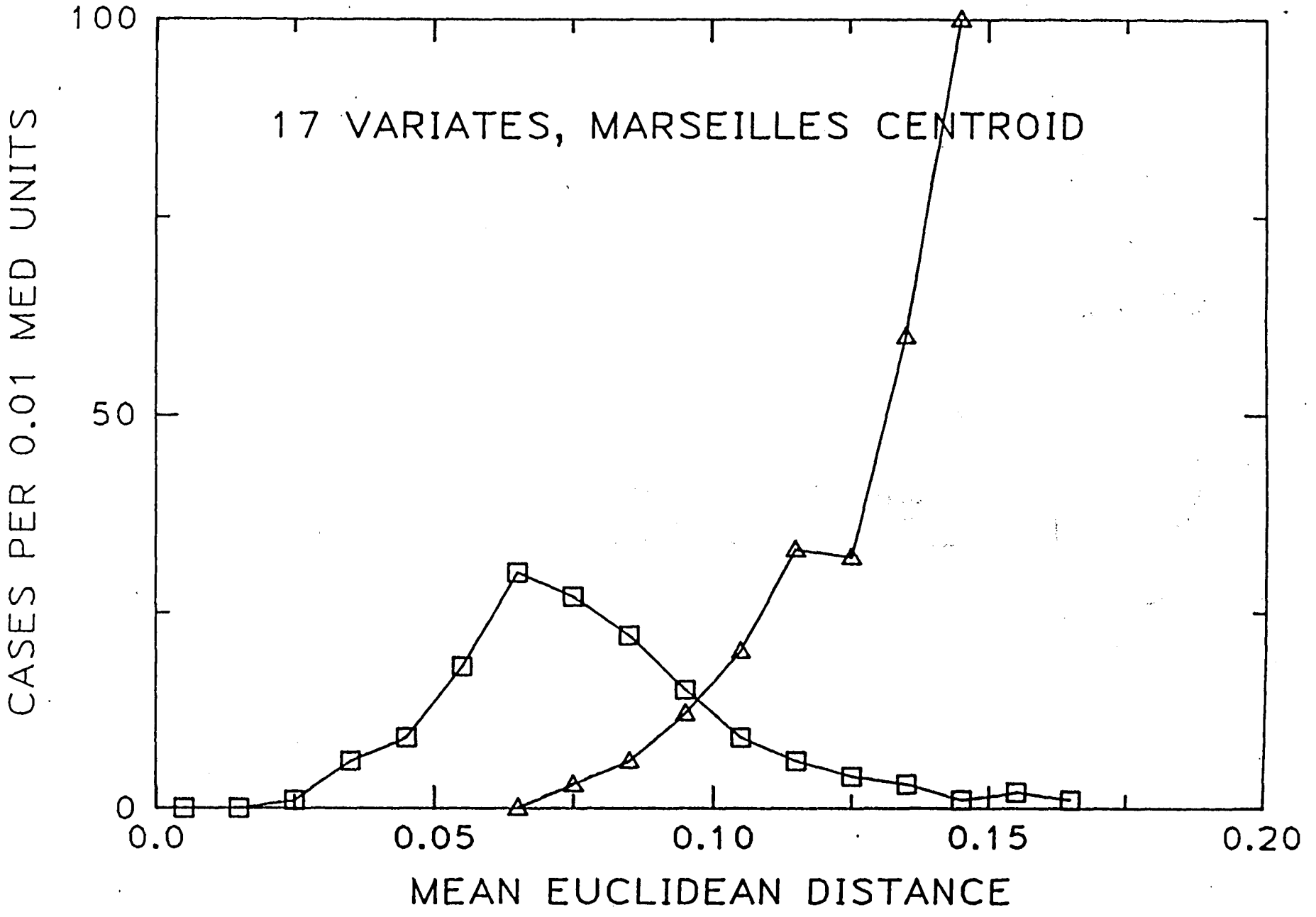
17 VARIATES, MARSEILLES CENTROID
CASES PER 0.01 MED UNITS
100
50
0
0.0
0.05
0.10
0.15
0.20
MEAN EUCLIDEAN DISTANCE

Figure 30 b

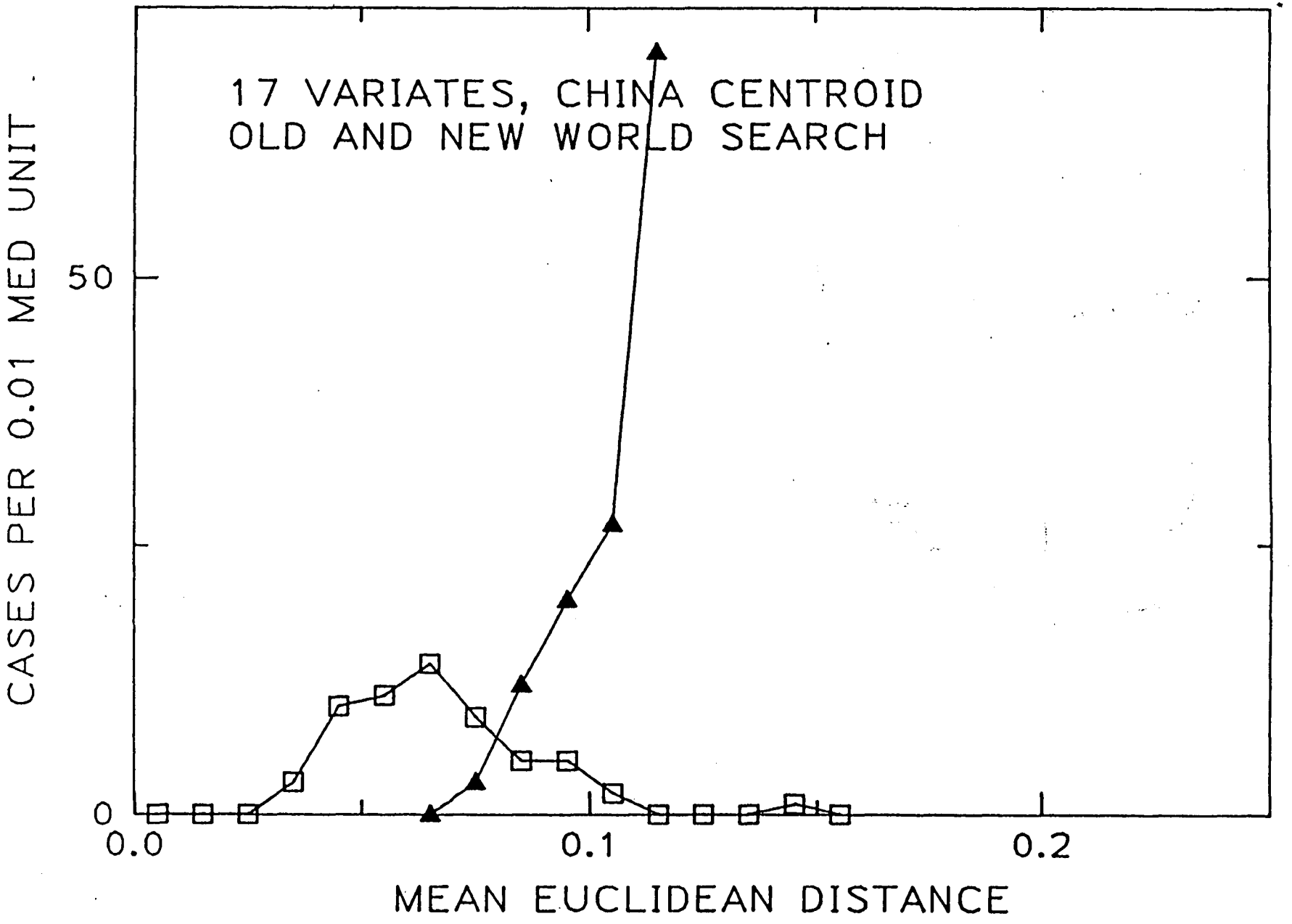
17 VARIATES, CHINA CENTROID
OLD AND NEW WORLD SEARCH
CASES PER 0.01 MED UNIT
50
0
0.0
0.1
0.2
MEAN EUCLIDEAN DISTANCE

Figure 17