
**NEW ZEALAND JOURNAL OF ARCHAEOLOGY**

# Minimum Mahalanobis Distance Functions and Lithic Source Characterisation by Multi-element Analysis

**Foss Leach**

Anthropology Department, University of Otago

and

**Bryan Manly**

Biometrics Unit, University of Otago

**ABSTRACT**

Existing multivariate statistical techniques are liable to allocate artefacts to incorrect sources, when there are large numbers of potential lithic sources, and when significant multidimensional overlaps occur using trace-element data. A new algorithm is suggested which solves some of the problems. A power transformation is used to overcome difficulties with unequal variance, and a statistic is developed based on Mahalanobis $D^2$ which finds the minimum distance expressed in standardised units taking into account missing data. The probability of artefact allocation to a particular source is assessed in terms of the likelihood of getting an artefact as far or further from the group centroid, assuming that it comes from that source. Thus, there is no underlying assumption that the sum of probabilities equals unity. This feature makes the test very powerful at rejecting sources. Rejection will occur either when an artefact has been poorly analysed, or when it belongs to an as yet undefined source. Source overlap is evaluated by considering the number of misclassifications to each other source that are expected to be recorded for 10,000 random artefacts from each source. Test cases are performed on artefacts of known source, and the results compared with those obtained by normal discriminant function analysis. Neutron activation analysis data of New Zealand and Pacific obsidians is employed, and also that from X-Ray fluorescence analysis, to test the generality of the method.

*Keywords:* MULTIVARIATE STATISTICS, SOURCING, OBSIDIAN, NEUTRON ACTIVATION ANALYSIS, X-RAY FLUORESCENCE ANALYSIS, MAHALANOBIS D-SQUARED, DISCRIMINANT FUNCTION ANALYSIS, SOURCING ALGORITHM.

## INTRODUCTION — THE NATURE OF THE PROBLEM

When an archaeologist uncovers an artefact in a site, a series of fundamental questions arise, such as: "what is it made from; who made it; how was it made; and what was it used for?" During the past 15 years or so an additional question has been asked — "where did the material come from?" This recognises that archaeological objects can reveal not only the culture of the groups being directly studied by a site's contents, but also the geographic movements of people, and their communication with other groups. Raw materials and items of material culture are exchanged and traded across geographic boundaries; and identifying the original source of artefacts is therefore the key to the study of prehistoric communication. In the Pacific area, since archaeologists first focused attention on exchange networks, some impressive movements of raw materials over several thousand miles of ocean have been documented. It is hardly surprising, therefore, that "sourcing" is gaining a prominent place in the box of tricks which an archaeologist considers to be his tools of trade. Just how appropriate and reliable this box of tricks is for answering the question of

material origins is the subject of this paper. The particular concern is to evaluate the state of methodology in sourcing obsidian implements — have we a box of tools, or a box of toys?

Concern over the possible unreliability of obsidian sourcing methods arose when problems were encountered in identifying the origin of some artefacts from the Chatham Islands. Trace element characterisation revealed that these artefacts were rather similar to two sources in the Pacific, separated by 6300 km (Leach and Warren 1981). It became clear that commonly used mathematical techniques for linking artefacts to their source had some notable deficiencies. Although the problem was revealed by the Chathams artefacts, it is certainly not confined to them. This example raised the whole question of just how much confidence, if any at all, could be given to previous identifications of obsidian artefacts in the New Zealand region (Ward 1972, 1974c; Reeves and Ward 1976; Leach 1973; Leach and Anderson 1978).

Upon further examination, it quickly became clear that mathematical algorithms used in this field, in particular linear discriminant function methods based on Mahalanobis $D^2$, are in essence a rather questionable version of Ockham's razor — the most parsimonious solution which accounts for the greatest number of available facts. The most parsimonious solution, in this case, is the minimum Mahalanobis distance (MMD), a solution which is liable to give the wrong answers, as will be shown below. In our view, if *any* of the available facts cannot be accounted for, this is sufficient grounds for rejecting a theory (an hypothesised source), and re-examining one's assumptions. It is our purpose to develop a new algorithm along these lines. The approach comes closer to Popper's "conjecture and refutation" — there is nothing against wild theorising, as long as it is accompanied by the most rigorous attempt to refute the theory (reject the null hypothesis). Under these circumstances, only the most robust interpretations will survive, and it is hoped a more accurate view of prehistory will thereby result.

Two basic questions arise in any characterisation and identification procedure for obsidian implements; these are as follows:

Q1 Are the known sources distinguishable from each other?
Q2 Where does this artefact come from?

It may seem self-evident that in setting up a sourcing scheme, these questions would be answered in their correct order — that there would be published a thorough and critical review of the reliability of the method, preceding any artefact identifications. In our view, for the Pacific area at least, previous answers to Q1 are incomplete, unconvincing, and therefore unacceptable. There is a real possibility that unreliable answers to Q2 have already infiltrated archaeological literature, and will be very difficult to eradicate later.

Again, following Popper's recipe, an honest appraisal of whether a particular technique can answer these two questions satisfactorily, consists in specifying precisely the conditions under which the technique should be given up as useless (see Lakatos 1970:92). We should therefore identify what we consider to be the *Minimum Standards of Proof*. In our opinion, these are as follows:

(i) The identification method should be closely defined so that it is capable of being applied without personal bias or prejudice. Identifications based on trust, or appeal to an "authority" are not considered satisfactory. Thus, the verbal identification method: "I can see from the shape of the spectrum of this artefact that it comes from Mayor Island" — is an unacceptable algorithm. Since we largely have to deal with numeric information in sourcing obsidian

tools, ideally the method should be mathematical in nature, to avoid any possibilities of observer bias.

(ii) It is very doubtful that any characterisation method for obsidian can completely separate each source as thoroughly distinguishable. In view of this, it is only fair that any method developed should reveal to the archaeologists precisely and quantitatively what the overlaps are between each source and all others. It is very difficult to know how much to trust any given method unless this ingredient is fully exposed.

(iii) There is a tendency in this type of research to accept a conclusion about the source of an artefact without misgivings, although the strength of the supporting evidence is less than absolute. Therefore, in nominating the source of an artefact, the method should reveal precisely what problems there might be in its identification. A series of possibilities could be nominated, graded in likelihood according to some previously defined criteria.

(iv) In setting up a characterisation method, the power of the technique to reject wrong answers must be convincingly demonstrated. If an artefact came from a source other than those known about and characterised, but the method appeared to show that it belonged to one of the known alternatives, confidence in the method would be seriously undermined. If it is known that the technique would tell a lie under these circumstances, we could not possibly decide with real artefacts which identifications were correct and which were not! One way of revealing any weakness in refuting an incorrect theory would be to try out the method on a series of controlled cases of foreign artefacts, where the possibility that they came from any of the sources characterised can be ruled out.

This paper is an attempt to develop an algorithm for sourcing obsidian artefacts in the New Zealand and Pacific area which on the one hand is not cautious in making a positive identification when it is warranted, but on the other hand will ruthlessly reject cases where it is not. It is also designed to reveal the full extent of any weaknesses in the characterisation method it is applied to. As test cases for the algorithm, two sets of data will be examined. One series results from a programme of neutron activation analysis (NAA) on 32 sources of obsidian, which resulted in data on 23 elements. The other was a project of X-ray fluorescence (XRF) analysis on 18 sources of obsidian, resulting in information on five separate elements.

## SOME FUNDAMENTAL QUESTIONS

Before discussing possible features of a suitable algorithm for answering Q1 and Q2, there are several more basic issues which should be decided first:

Q3 Should we analyse artefacts to find the sources, or analyse sources first, and then try to identify the artefacts?

It could be suggested that since prehistoric people did not exploit all the known sources of obsidian, it would be sensible to define which quarries were used by analysis of the artefacts themselves, rather than by samples from the sources. Unfortunately, this alternative strategy loses one of the most valuable ingredients possessed by the more normal approach of working with source material first — that of an independent check on the progress of one's method of discrimination. Even working directly with source material, considerable problems are encountered in achieving reliable source separation; these problems are compounded when working blindfolded in a morass of data from artefacts. In studies of biological taxonomy, the difficulties of decomposing mixtures into separate populations without *a priori* knowledge are well known. Assuming common variance and normality, a mixed

series can be resolved into two Gaussian components by a method of moments using iteration by Newton's method of approximation. Small departures from these assumptions, however, introduce serious errors in the estimates. The method has been generalised to allow separation of a multivariate sample into two or more multivariate normal components by Wolfe (1970). As will be shown later, the assumption of common variance has little justification in sourcing studies, and it would be folly to rush in where angels fear to tread — Rao, with unusual reticence, comments on this whole problem: "a good deal of caution is needed in resolving a mixed series" (Rao 1952:304). In our view, it is far wiser to follow the Ward paradigm (1974a, 1974b) of defining as many of the known taxa as possible first, and face up to the formidable task of reconstructing the nature of unknown sources from artefacts from the vantage of experience with those which are well known. In short, it must be concluded that it is better to proceed from the known to the unknown, rather than attempt the reverse.

By and large, it is possible to obtain samples from most of the sources which might have been used in the past in a particular region. However, it is well known that a few artefacts are analysed which simply do not fit any of the known source patterns (Dixon 1976:307ff), and source reconstruction must be attempted. In these cases it is probably advisable to carry out normative analysis in order to obtain geological and geochemical clues as to precisely what kind of obsidian is involved, and approximately where it may have come from. This is the approach being taken in obsidian studies in the Mediterranean area. A special case of missing sources was recently reviewed by Ambrose *et al.* (1981). Sources in the Admiralty Islands area have been covered over by ash falls since their exploitation. By carrying out an unusual method of cluster analysis by iteration, a large number of artefacts were tentatively grouped into "sources".

Q4 Should we use multivariate methods to answer Q1 and Q2, or univariate methods?

The answer to this question depends entirely on the dispersion characteristics of the information we are confronted with. If we characterise the sources with two uncorrelated elements, then it is entirely appropriate to consider the question of possible source overlap, and artefact allocation to those sources, with univariate statistics. On the other hand, if the two elements are correlated, then these questions must be addressed with statistics appropriate to the bivariate distribution. This distinction can be generalised to the multivariate case. By assuming lack of correlation, in using univariate tests, we would effectively be claiming that each additional element adds 100 percent new information. That this is untrue is easily seen in the extreme case when each successive variable (element) is algebraically related to the last — the addition of such variables does not serve to further distinguish the sources one *iota*.

By using multivariate statistics (notably those based on Mahalanobis $D^2$), only the precise contribution of new knowledge by the addition of each new variable is considered, and only this fraction is taken into account in addressing questions of discrimination. Some of the pitfalls of using univariate statistics for this type of problem are illustrated in Figure 1.

In this example, the proposed correlation between elements is only moderate ($r \approx 0.6$); however, many element pairs in obsidian, especially amongst the rare-earth suite, are correlated to the order of $r \approx 0.97$. In cases like this, the problems illustrated in Figure 1 are greatly magnified.

Q5 Should we use element concentrations or ratios?

Exact replication of instrument conditions from one batch of samples to another

is extremely difficult — an electronic amplifier may have a slightly higher gain setting which could lead to element concentrations appearing to be somewhat higher than those of a previous batch. Also, if XRF analysis is performed on whole samples, it is well known that the uneven geometry will be the cause of variable diffraction effects. Even by running suitable standards as cross-checks at regular intervals through the series, variations in machine conditions cannot be ruled out. Some of the problems which arise can be partially overcome by turning the element concentrations (or even peak areas) into ratios with respect to one or more elements. There are several distinct advantages and disadvantages of this strategy which should be enumerated:

(i) Advantage: The problems of replicating experimental conditions are minimised by taking ratios.

(ii) Advantage: The dispersion matrix of ratios is likely to be a closer estimate of geochemical intra-source variation than that from concentrations. If there is variation due to change in machine conditions, taking ratios will reduce the experimental error in comparing samples; it should therefore result in improved discrimination between sources.

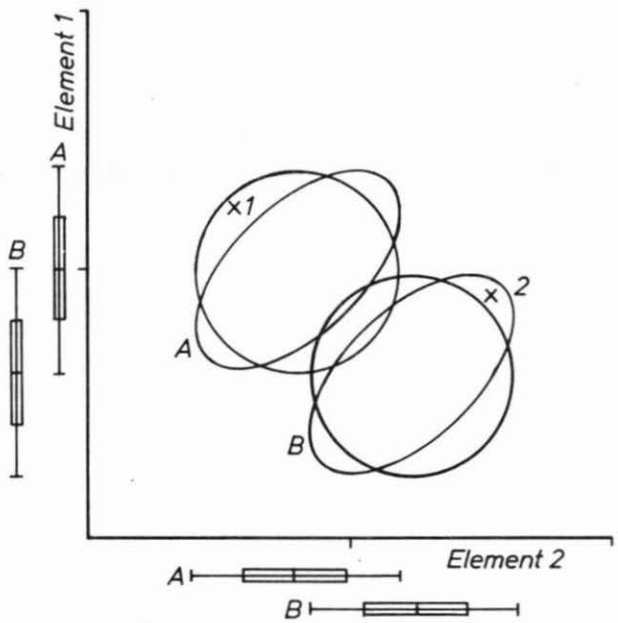


*Figure 1:* Artefact identification by trace element analysis should not be treated as a series of univariate problems, because many of the element pairs are highly correlated. In this simplified example, using univariate statistics we would have to assume that the 95% lines of equi-probability defined a series of circles around the sources A and B. In fact, because of element correlation, the 95% lines of equi-probability define two ellipses. By using univariate statistics we would conclude that the two sources overlap ($p = .05$), that artefact #1 is within the known distribution of source A ($p = .05$), and artefact #2 is significantly different from source B. Each of these propositions is manifestly untrue.

(iii) Disadvantage: If the main difference between two sources is the overall degree of trace-element enrichment, this difference will be obliterated by taking ratios.

(iv) Disadvantage: If an artefact happens to be particularly high or low for the normalising element (perhaps because of poor analysis, or its unique position in an obsidian flow), then *all* the element ratios will be scaled up or down by this amount. By using the concentration data, this artefact would appear only marginally different from its parent source; by taking ratios it will now appear to be substantially different.

In view of the disadvantages, it is difficult to be hard and fast on the subject of ratios, and each case must be considered on its own merits. If sacrificing the enrichment factor is not considered wasteful, then one of the objections is ruled out. The latter problem must be very carefully monitored though if ratios are to be used. The main normalising element should have a low coefficient of variation, and all the data should be thoroughly screened for rogue values. On the whole, it would be best to test for the degree of discrimination before and after ratios are taken, as an indication of the advisability of this approach.

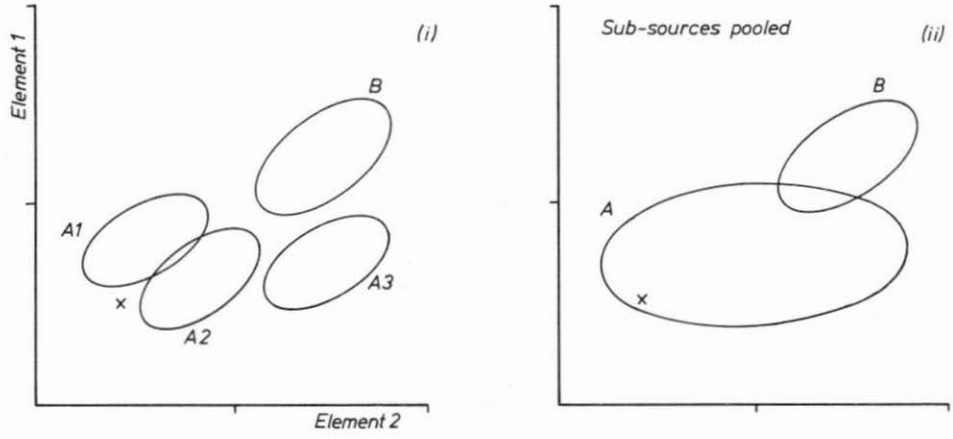


*Figure 2:* One of the adverse effects of clustering sub-sources into a single pooled "source" is that it can lead to incorrect artefact identifications and undue source overlap. In (i) source B is quite distinct from the sub-sources of A, and the artefact (x) appears not to belong to any known source. After clustering the sub-sources (ii), source A now overlaps with source B, and it has also captured the artefact.

Q6 Should we group sub-sources together before trying to source artefacts?

This is a difficult question to answer in theoretical terms. As will be seen from Figure 2, there are certain difficulties which could arise if sub-sources were clustered indiscriminantly prior to attempts to identify artefacts. From this example it could be suggested that if the multivariate means and variance matrix for two sub-sources are significantly different, then they should be left separated rather than pooled. If both these characteristics are identical within acceptable statistical limits, then it is probably wiser to pool the two series of results as a better overall estimate of the population values, particularly if there are other sources nearby. Suitable tests for

these two features are Hotelling's multivariate $T^2$ test, and Bartlett's $\chi^2$ test (see below). An alternative is to employ the procedure outlined by Rao (1952:362-3) where the change in average $D^2$ within a cluster is successively re-calculated after adding additional groups. If the change ($\Delta D^2$) is appreciable at any point, then the newly added group is considered to be outside the cluster. The technique requires an interactive computer programme (see Leach 1969). Another approach is to combine sub-sources together successively and re-calculate multivariate overlap between remaining sources. If this dramatically worsens, then the groupings were probably an unwise choice.

Q7 Should we ignore poorly resolved elements?

All analytical techniques produce results of variable accuracy through the periodic table; that is, the signal to noise ratio varies from one element to another depending on the method of analysis, and also in relation to whether the element has any discriminatory power in the first place. One or more elements could be considered extremely unreliable ($\pm$ 90%); yet for particular sources of obsidian the results may be far above or below those for all others. This would be a clear case when the data for the element should be retained. Unfortunately, this same element could be the cause of worsening the degree of overall overlap between other sources — the problem is: where does one draw the line? A suitable test for whether the inclusion of q extra elements significantly improves or worsens the discrimination between two populations which had been achieved with a smaller group of elements (p) is the ratio:

$$R = \frac{1 + x.D^2_{p+q}}{1 + x.D^2_{p}}$$

where $x = N_1.N_2/(N_1 + N_2) \cdot (N_1 + N_2 - 2)$

The significance of this ratio is assessed with the variance ratio:

$(R-1).(N_1 + N_2 - p - q - 1)/q$
which has q and $(N_1 + N_2 - p - q - 1)$ degrees of freedom (after Rao 1952:253).

A more pragmatic, and perhaps simpler test would be to evaluate the degree of multi-variate overlap between all sources before and after the addition of marginal elements.

## DESIGNING A PRACTICAL ALGORITHM

The concentrations of the various trace elements in obsidian are far from independent of each other; in fact, certain groups of elements, such as those of the rare-earth suite, are highly correlated (see Leach and Warren n.d.). In evaluating how distinctive one source of obsidian is from another, it is important not to exaggerate any differences which are found by simply multiplying by the number of correlated elements. If two elements are highly correlated, very little extra discrimination between sources is achieved by analysing for the second element as well as the first. With the advent of automated data logging equipment, of course, it is usually just as easy to accumulate data from many elements as only a few; and since each element will add a little bit of extra information ($r \neq 1.00$), then one judges it to be worthwhile to do just that. However, in setting up a source discrimination scheme based on such multi-element data, it is absolutely necessary to ensure that this exaggeration effect is avoided at all costs.

The only effective way of doing so is by following a method developed by Mahalanobis (1930, 1936). The technique on the one hand transforms the multi-

dimensional space so that the axes are orthogonal; that is, the transformed differences between groups are effectively between uncorrelated variables (r = 0.00). On the other hand, this transformation maximises the power of discrimination between individual specimens in previously constructed groups (see Constandse-Westermann 1972:48ff; Sneath and Sokal 1973:128). For the type of data we are dealing with in trace element characterisation of obsidian, the latter feature is highly desirable, and the former absolutely essential. The statistic used for this purpose is a squared multivariate distance function:

$$D_r^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} w^{ij} \quad (\bar{x}_i^1 - \bar{x}_i^2)(\bar{x}_j^1 - \bar{x}_j^2)$$

where $w^{ij}$ is the element of the inverse of the pooled dispersion matrix for variables i, j (= 1,2...,r), and $\bar{x}_i^1$ and $\bar{x}_i^2$ are the two sample means for the ith element.

CONFRONTING THE PROBLEM OF UNEQUAL VARIANCE

Multivariate statistical methods designed to discriminate between populations assume that the dispersion matrices of the different groups are the same (that is that the clusters have the same size, shape and orientation in n-dimensional hyperspace), and that the clusters have multivariate normal distributions. It would be extremely difficult, if not impossible, to design a suitable discriminant method which did not make these assumptions (Sneath and Sokal 1973:404). Although this problem is widely recognised, it is normally passed off with a shrug and a reference to the dubious panacea that:

> even for populations strongly divergent in their frequencies, i.e. having frequencies with rather diverging variances, the pooled dispersion matrix still works satisfactorily and may thus be designed and applied, although theoretically, the matrix does not represent a reality (Constandse-Westermann 1972:93; see also Sneath and Sokal 1973:127).

This is cold comfort indeed if we want to be reasonably certain about our artefact allocations; for while the significance of group separation may be largely unaffected by the falseness of these assumptions, we can expect them to play havoc when associating individual points (artefacts) with those groups.

In point of fact, it is rather unlikely that correlations will vary between different obsidian sources to any significant degree. Although the overall levels of different elements vary considerably, the geochemical factors which cause fractionation of elements must be very similar from one magma to another, ensuring that element correlations would be very similar too. It is difficult at this stage to test for inhomogeneity in correlation because the number of samples analysed for different obsidian sources is so small. It must be therefore left in abeyance at this stage, but closely examined later when more information is available.

Non-normality of the data could present a problem if this is verified. At the moment, the number of samples analysed from any one source is pitifully small, and it is not possible to examine thoroughly for non-normality. There is no obvious geochemical reason why variation within a source should not result in normal distributions; however, it would be wise to review this matter when more analyses have been carried out.

The really serious problem concerns the assumption that the variance from one group to another is the same. This is simply not true in the case of trace-element data for obsidian. The differences in variance from one source to another are dramatic, and a single example given in Table 1 will suffice to illustrate the point. With this amount of variation from one source to another, the effects of assuming a common

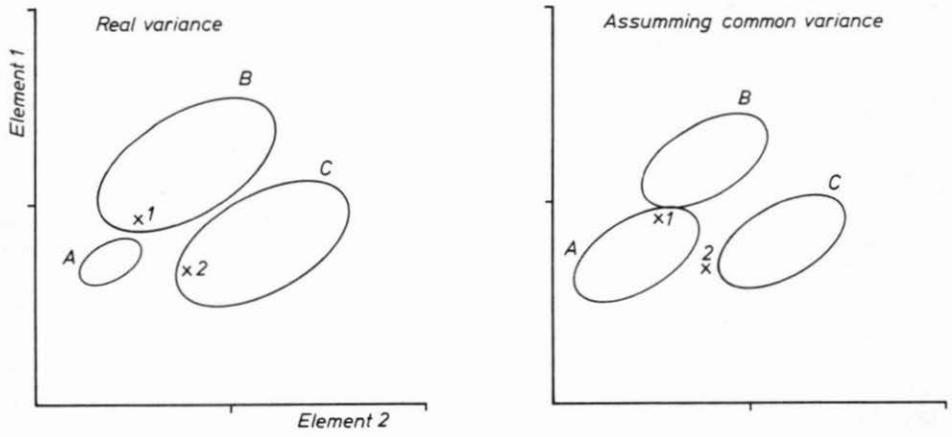


*Figure 3:* One of the adverse effects of assuming common variance is to promote the incorrect allocation of artefacts. In this example, artefact #1 actually belongs to source B, but after taking a pooled estimate of variance it now appears to belong to source A. Artefact #2, on the other hand, actually belongs to source C, but on assuming common variance now appears to define an unknown source.

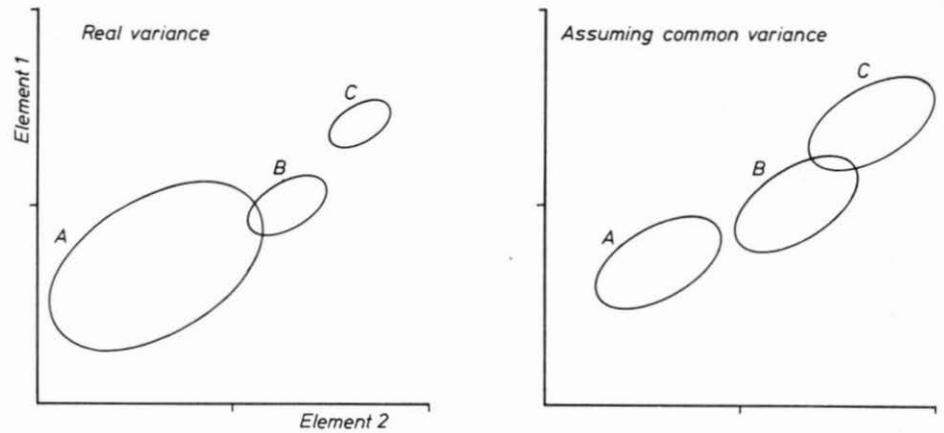


*Figure 4:* Another adverse effect of assuming common variance is to alter the structure of which sources are distinguishable from each other. In this example, sources A and B overlap, while sources B and C are highly distinctive. After taking pooled variance, sources A and B now appear to have completely separate distributions, while sources B and C now appear to have an overlapping region.

variance are bound to be dramatic. Two predictable examples are illustrated in Figures 3 and 4 — further comment is unnecessary.

Clearly, if at all possible, something should be done to try to alleviate this problem. An attractive possibility is to use a power transformation of the form $x' = (x^{\lambda} - 1)/\lambda$, $\lambda \neq 0$; $x' = \log_e x, \lambda = 0$ following Box and Cox (1964). The value of $\lambda$ can be chosen so that the transformation is as effective as possible in stabilising the within-group variance. A range of values for $\lambda$ was therefore tried for each element, and the value which minimises the Bartlett $\chi^2$ test statistic for equal variance is considered the most effective for that element. Rather conveniently, this turned out to be $\lambda \simeq 0.2$ for all the elements in this study. With this value, Bartlett's $\chi^2$ test is not significantly large for some 12 of the 23 elements studied by NAA, and only marginally so in the remainder. Without the transformation, all values are highly significant. The effects of this type of transformation may be seen in Figure 5, where it is clear that a value of $\lambda \simeq 0.2$ is a good overall choice.

As might be expected, when element ratios are taken, the assumption of equal variance is even less justifiable than with the raw concentrations. Again, however, this power transformation results in a dramatic improvement, although in only one case amongst the 22 ratios is the resulting between-group variance rendered insignificantly different (see Fig. 6).

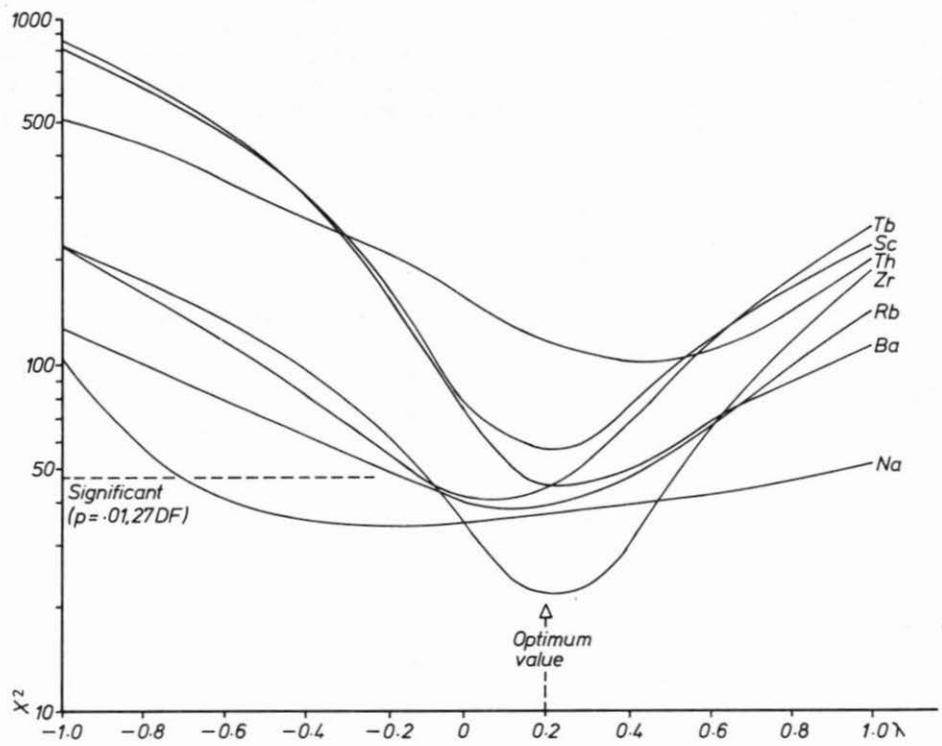


*Figure 5:* Bartlett's $\chi^2$ test for equal variance uniformly fails with trace element data ($\lambda = 1.0$); however, a power transformation of the form $x' = (x^{\lambda} - 1)/\lambda, \lambda \neq 0$; $x' = \log_e x, \lambda = 0$ results in a dramatic improvement, especially with $\lambda \simeq 0.2$. This figure illustrates the effects of the transformation for a selection of elements for New Zealand and Oceanic obsidians.

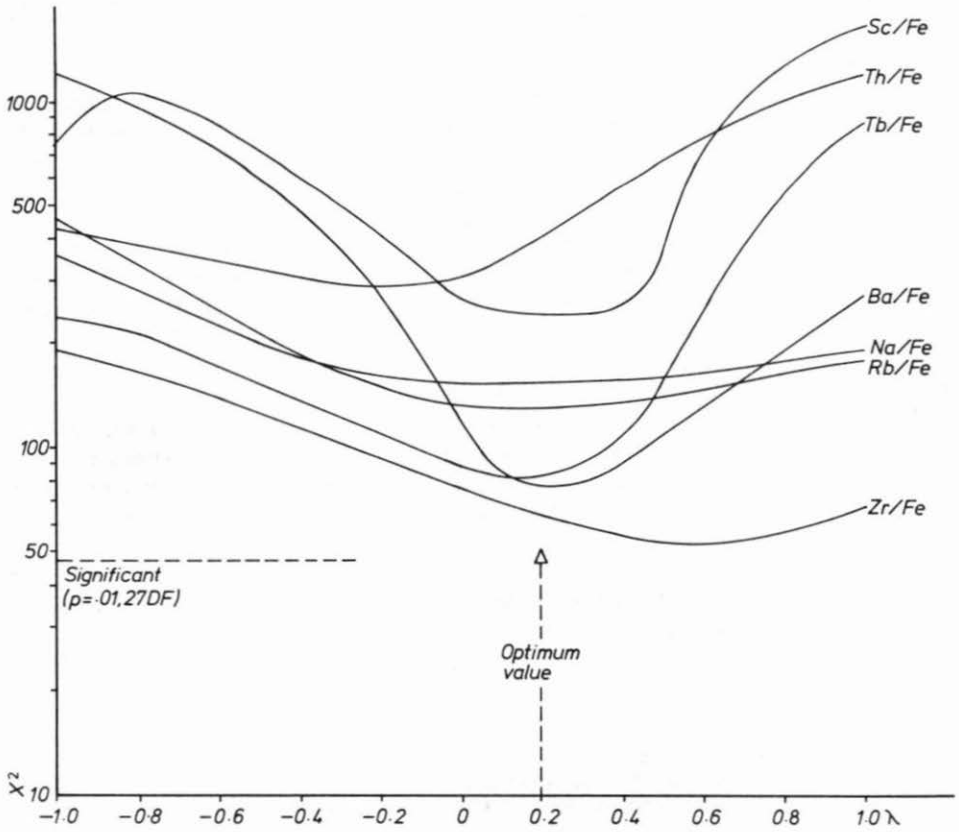


*Figure 6:* The assumption of equal variance is even less justifiable in the case of element ratios ($\lambda = 1.0$); but again a value of $\lambda = 0.2$ helps the situation considerably.

On the whole, this power transformation is not perfect by any means, but it does effect a very considerable improvement. It is interesting that a value of $\lambda \simeq 0.2$ has such a uniform effect across the different trace elements. It is not expected that this would necessarily be the best choice for other data sets; indeed, preliminary examination of PIXE-PIGME data on obsidian confirms this. Each case should therefore be examined on its own merits.

## THE PROBLEM OF ESTIMATING THE POOLED DISPERSION MATRIX IN THE FACE OF INCOMPLETE MULTIVARIATE DATA

Two problems commonly arise with trace element characterisation projects which are greeted with a mixture of astonishment and disbelief by statisticians. Firstly, for various reasons there are often many missing values which are neither genuine zeros nor below minimum detectable levels. Rogue analyses of various kinds occur even with the most sophisticated equipment and with due attention to surface decontamination. The importance of visual or manual inspection of results to delete

these incorrect data is consistently stressed by physicists who are often involved in the data collection stage. Secondly, amongst the ~100 sources of obsidian in the Pacific basin, quite a few are defined by the analysis of a single piece of rock. In some cases, such as where the quarry has been removed by modern urban or rural development, this one piece may be the sole remaining sample. In other cases, a source may be defined by the analysis of an unusual artefact without any knowledge of the original geographic provenance. For the most part, we are fortunate to have five samples analysed for each source, and overjoyed to have 15. In short, our sample sizes from which to estimate intra-source variation and other characteristics are generally pitifully small. Moreover, with a series of small samples, the calculated pooled dispersion matrix may have a number of illogical combinations. For instance, the condition: Na/Fe r = −ve, Na/Sc r = +ve, and Fe/Sc r = +ve, is strictly speaking impossible and could lead to negative eigenvalues. This type of result can easily occur with the small samples we are faced with in lithic sourcing.

Fortunately, a suitable algorithm has been developed to overcome these difficulties (Huseby *et al.* 1980; Schwertman and Allen 1979). This procedure estimates the variance-covariance matrix for each group using all the available data, obtains an estimate of the pooled dispersion matrix, and then "smoothes" it to ensure that it is a valid estimate with positive eigenvalues. A Moore-Penrose inverse matrix is also calculated, since this is needed for the calculation of Mahalanobis distances. This algorithm was used in the present study.

## ESTIMATING THE MINIMUM MAHALANOBIS DISTANCE WITH SOME PERSPECTIVE

One of the difficulties encountered in using distance statistics is that it is not normally very clear what the scale of the results is; and in fact, this varies from one

**TABLE 1**
**STANDARD DEVIATIONS OF SCANDIUM CONTENT (ppm)**
**IN NEW ZEALAND AND OCEANIC OBSIDIANS**
**ARRANGED IN DECREASING SIZE**

Although $\sigma$ clearly varies a great deal, it is largely a function of the mean value of the element in each case. The coefficient of variation, as might be expected, does not vary much from one source to another ($\bar{x}$ = 17.1%, $\sigma$ = 8.4).

| OBSIDIAN SOURCE | STANDARD DEVIATION ($\sigma$) | | STANDARD ERROR OF ($\sigma$) |
|---|---|---|---|
| #12 Tairua | 727.6 | ± | 162.7 |
| #24 Taupo | 709.9 | ± | 158.7 |
| # 9 Cooks Bay | 699.9 | ± | 142.9 |
| #15 Waihi black | 551.8 | ± | 225.3 |
| #22 Maraetai black | 526.7 | ± | 140.8 |
| #21 Maraetai red | 484.3 | ± | 153.1 |
| #23 Ongaroto | 400.2 | ± | 89.5 |
| #13 Maratoto | 349.9 | ± | 93.5 |
| #14 Waihi red | 304.6 | ± | 68.1 |
| #10 Purangi | 283.5 | ± | 75.8 |
| #11 Hahei | 274.6 | ± | 61.4 |
| #29 Motu Iti | 151.9 | ± | 62.0 |
| # 1 Weta | 140.3 | ± | 57.3 |
| #28 Maunga Orito | 100.8 | ± | 29.1 |
| # 6 Fanal Is. | 96.2 | ± | 34.0 |
| #30 Te Manavai | 87.1 | ± | 30.8 |
| # 3 Pungaere | 62.2 | ± | 15.6 |
| #16 Mayor Is. green | 55.8 | ± | 12.5 |
| # 2 Waiare | 42.1 | ± | 8.6 |
| #17 Mayor Is. honey | 31.9 | ± | 10.1 |

group to another because of missing values. To be sure, one can calculate the significance of distances using an F ratio test, or Hotelling's $T^2$ test, but this is a bit like introducing a ruler into one's field of view when observing various scattered objects with one eye closed. It does not really add an overall perspective, it merely enables individual inter-object distances to be evaluated with a familiar scale. No amount of measuring with the ruler will enable us to see the objects in three-dimensions. It would clearly be an advantage if we were able to relate the distance statistic to some familiar measure. This can be done as:

$$D^{2\prime} = D^2/r^*$$

where $r^*$ is the effective number of dimensions (variables = elements), taking into account the missing values. Calculating $r^*$ is somewhat complicated. The eigenvalues are computed for the variance-covariance matrix for measurements that are present. This must be done separately for each specimen because the missing measurements vary from one specimen to another. For a valid variance-covariance matrix, all the eigenvalues should be zero or positive. The number of positive eigenvalues equals the number of independent dimensions for the variables (cf. principal components), and this is the value used for $r^*$. An eigenvalue is judged to be positive if it is 1 percent or more of the sum of all eigenvalues. The 1 percent convention was used to avoid the effects of taking into account very small eigenvalues which correspond to rounding and sampling errors. Huseby *et al.* (1980) used a much smaller percentage, but for the present application it was found that their value resulted in many $D^2$ values that were dominated by very small eigenvalues. For the 23 element NAA study of obsidians, $r^*$ was ~7 in most cases, which implies that the measurements on specimens can be summarised in about 7 principal components. This clearly demonstrated the importance of intercorrelation in this type of data.

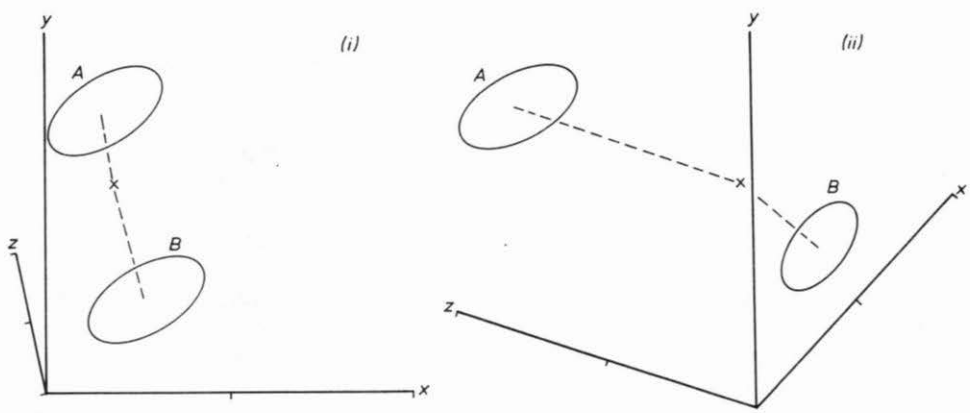


*Figure 7:* According to perspective (i), the artefact (x) appears to be closer to source A, yet in perspective (ii) it is clear that it is actually closer to B. This is because source A is defined in three dimensions, and B in only two. Clearly, apparent distance is not always the best guide to the source of an artefact. This figure is a simplification of a more complex case where the effective number of dimensions ($r^*$) varies.

Scaled in this way, the distance $D^2{}'$ for an average individual to its own group would be $\sim 1.0$. Thus, the Mahalanobis $D^2{}'$ values are expressed in standardised units. If we find an artefact located at $D^2{}' = 6.0$ we can instantly appreciate that it is 6 units away from the centroid of the multivariate cloud defined for this particular obsidian source, whereas the average distance is only 1 unit. Although this statistic gives only an approximate idea of the relative location, the perspective which is added as a result is of great value.

### THE SOURCE WITH THE MINIMUM MAHALANOBIS DISTANCE TO AN ARTEFACT IS NOT NECESSARILY THE CORRECT SOURCE

There are two reasons why the statistically closest source to an artefact may not be the most plausible geographic source. Firstly, because of variations in the multivariate definition of each source due to missing values, an artefact could be closer to source A than source B, yet significantly different from A and insignificantly different from B. This peculiar situation arises as a direct result of variation in the effective number of variables (r*) and the effective sample sizes taking into account missing data (N*) of the sources. It is therefore necessary to ensure that these two terms are present in any statistical test of significance of $D^2{}'$ (see Figure 7).

Secondly, the artefact may actually come from an as yet undiscovered source, for which there has been no multivariate definition. It might be thought that this would be an easy condition to recognise, but this is not so. The commonest method for assessing group origin is by virtue of the significance associated with the discriminant functions based on Mahalanobis $D^2$. There is a serious pitfall in this approach, which in view of the widespread use of this model, should be pointed out in some detail. The probability assigned to the Kth source is as follows:

$$P_{ak} = \frac{e^{(-D^2_{xk}/2)}}{\sum_{i=1}^{q} e^{(-D^2_{xi}/2)}}$$

where $i = 1, \ldots .q$ (groups)
and $D^2_{xk}$ = Mahalanobis distance from the artefact (x) to the source k.

It will be readily seen from this function that $\Sigma P = 1.00$. What this implies in effect is that one must have *a priori* knowledge that the source of the artefact is actually amongst those previously defined. If the artefact is poorly analysed for any reason (and therefore lies outside the defined sources), or its true source is not present, it will be assigned to a source with almost any discriminant score (Sneath and Sokal 1973:404). It should be abundantly clear that this statistic is not a *test* in the sense that it has any power to reject incorrect sources. If there is any reason to suspect that there could be more sources than those already defined, then the results of this type of analysis are thrown into disarray. There is simply no *post hoc* means of evaluating which answers are correct, and which are not. If we were using this technique to distinguish male and female from skeletal data, this problem would not arise; but when additional alternatives are possible, it is a serious objection which must be taken into account (see Rao 1952:290).

Fortunately, there are several ways in which this problem can be approached, and the most attractive is probably the multivariate version of the students t test, that is, Hotelling's $T^2$. This tests whether the difference between the multivariate means is zero ($Ho = \mu_1 - \mu_2 = 0$). Since the artefact constitutes a sample with $N = 1$, the

significance of our scaled Mahalanobis distance $D^{2\prime}$ being $>1.0$ simplifies to the following:

$$T^{2\prime} = \frac{r^*.N^*}{(N^*+1)}.D^{2\prime}$$

where $N^*$ is the average sample size of the source being considered, taking into account missing values. $T^{2\prime}$ is approximately distributed as $\chi^2$ with $r^*$ degrees of freedom.

It will be noticed that there is no pre-condition that $\Sigma p = 1.0$, and this test may be thought of as:

Giving, through the $\chi^2$ distribution, the probability of obtaining an artefact as unusual as this, given that it comes from this source.

A small probability will lead to a source being rejected. This test is very powerful at rejecting sources, and it can be applied to as many sources nearby or far away as one wishes. If the test fails to yield an insignificant result amongst any of the known sources, then one has to accept one of the following possible explanations:

(i) The artefact is from a new source
(ii) The artefact comes from one of the known sources, but the dispersion matrix for this source is very different from the estimated pooled dispersion matrix.
(iii) The analysis of the artefact involved an error.

The difference between the two types of statistic mentioned is illustrated in Figure 8. Finally, it must be noted that the $T^{2\prime}$ statistic cannot be applied to test an individual with its own parent group.

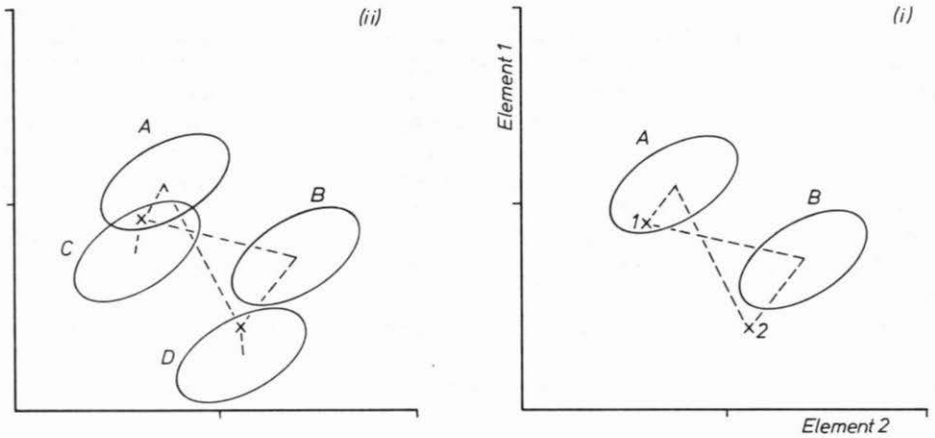


*Figure 8:* The hypothetical result of characterising some known obsidian sources is shown in (i). However, two sources were available to prehistoric people (sources C and D) which (for whatever reason) we have not included in our analysis of sources. Their true dispersion characteristics are shown in (ii). Based on the information given in (i), the $P_{ak}$ statistic would assess artefact #1 as belonging to A with a high probability, and artefact #2 will be allocated to B, but with a probability which is difficult to predict. The test $T^{2\prime}$, on the other hand, will reject B as the source of artefact #2. Note that the centroid of source C is closer to artefact #1 than that of source A. There is no way of predicting this eventuality.

## HOW GOOD IS A CHARACTERISATION METHOD AT DISTINGUISHING THE SOURCES?

This, of course, is one of the central questions in any attempt to set up an identification scheme; yet strange as it may seem, there are few techniques at our disposal for providing a satisfactory and convincing answer about the success or otherwise of any system. If we had only one element to consider, the problem is relatively simple — we can calculate the mean and standard deviation for each source, and plot them out to show which sources are very distinctive even at $3\sigma$ or $4\sigma$, and also reveal which ones overlap at $1\sigma$. Similarly, if we had two elements to consider, we can construct equi-probability ellipses, including, say, 95 percent of the distribution, and again illustrate which sources do and do not overlap, and by how much. What should we do for the multivariate case though? In practice, what people have done is to pick out selections of element combinations and illustrate as many as possible — a ratio of two elements on one axis against another pair on the second, and even triads of element ratios on ternary plots, and so on. These techniques of illustrating the success of one's method may be persuasive, but they are anything but entirely convincing, particularly since so many element pairs are highly correlated. What is needed is a quantifiable statement for each defined source of the form:

> This method would allow me to identify an artefact from source A correctly 99 percent of the time, but 1 percent of artefacts from this source may be incorrectly assigned to source B.

After all, a lot can turn on the results of individual artefacts. For instance, we could find an artefact in New Caledonia which analysis suggests may have originated in New Britain. Such an identification could not conceivably be an absolute; and in proportion to the importance of such a conclusion, we should know what the possibility is of having made an error. For instance, if the situation is: *"99 percent probable from New Britain, and 1 percent possible from Vanuatu"* this is one thing; but if it is *"65 percent versus 35 percent"* we are likely to be more circumspect in jumping to conclusions on the basis of one artefact. There is an unfortunate tendency in archaeology for a probable fact to become a certain fact once it has been repeated several times in the literature. If a confidence level was explicit in formulations about artefact sourcing, it would go some way towards minimising this tendency. The first thing which must be noted is that it is simply not possible to calculate the probability that an artefact comes from this or that source — that is, statements like those above in italics could not be yielded by statistical analysis. It is the old story — a significance test does not give the probability that the null hypothesis is true! It is aimed at refutation, not persuasion. How, then, could we ascertain how reliable our identifications are?

One possible way of approaching this problem is to carry out some controlled tests of the method — jumble up a random selection of pieces from known sources, and then treat them as artefacts to see how often the correct answer is obtained. A novel version of this type of test was carried out by Ward (1972, 1974a, 1974b), who used the $P_{ak}$ statistic cited above on all source material analysed to see how much confusion between sources was inherent in his characterisation method. Apart from the deficiencies of the particular statistic already mentioned, this is an entirely reasonable approach, even if it does come close to a circular argument. The main problem with the controlled test approach is the large number of analyses which would be required to reveal the extent of any overlaps. For instance, if we wanted to know the extent of possible mis-classifications to, say, a level of 1 percent, and we had 30 sources to consider, we would need to test an absolute minimum of 3000 pseudo-artefacts. This is out of the question — the adequate characterisation of

sources is a progressive and long-term project, beginning with, say, five pieces from each source, and adding knowledge both from artefacts and additional source material over a period of years. However, in the interests of making use of artefact identifications here and now, we need to know what the possibility of misidentification is, even though we have defined our sources with only five pieces each at this stage. Over a period of years, as our knowledge of the true dispersion characteristics of each source improves, we should expect the confidence of previous artefact identifications to improve in some cases, and worsen in others.

The first step in this process is to show that there is a significant difference between the multivariate means for each pair of sources; that is, that the difference between the centroids is significant. This can be done with a Hotelling's $T^2$ test as:

$$T^2 = \frac{N_1^* . N_2^*}{N_1^* + N_2^*} . D_r^{2*}$$

where $N_1^*$ and $N_2^*$ are the effective sample sizes of the two sources, and $r^*$ the effective number of elements (after Constandse-Westermann 1972:53).

This is an important procedure, but is only the first step towards an adequate statement of the degree of success of the characterisation method. Two source centroids could be significantly separated in hyperspace according to this test, and yet still share 60 percent common space. If $H_0$ is not rejected, then this would be good grounds for clustering the two sources into a single group — hopefully, this will only occur within single geographic regions.

The second step would be to assess the multivariate overlap directly once the various sources have been defined. It has been shown that between two multivariate pairs this is:

$$\alpha = 1\sqrt{2\pi} \int_{D/2}^{\infty} e^{(-y^2/2)}\, dy$$

where D = the Mahalanobis distance between the two group centres (after Rao 1952:355; see also Anderson 1957:§6.4)

This expression can be used to see how often an artefact from source A is closer to source B than it is to A, for each other known source B. If 1 percent of the artefacts from source A are closer to source B, then it is immediately clear that the two sources share a common region of about 1 percent of their distributions in hyperspace. For the archaeologist, this expression gives a clearer picture of source overlaps than a statement of the significance of intergroup centroid distances.

## THE POPPERS/RAZOR ALGORITHM

A series of three computer programs were written in Fortran to carry out the various tasks mentioned above. These accept a matrix of grouped source data (either as concentrations or ratios) with missing values, and a string of artefact data at the end for identification. The first program — POPPERS/RAZOR 1 — calculates variances, following the technique of Schwertman and Allen (1979) and Huseby *et al.* (1980), and presents the values of the Bartlett $\chi^2$ test for equal variance for values of $\lambda$, covering any desired range for the power transformation described by Box and Cox (1964), for each element. Inspection of these results will suggest an optimum value of $\lambda$ for each element in subsequent operations. We suggest $-1$ to $+1$ as a reasonable range for $\lambda$ in general.

The second program — POPPERS/RAZOR 2 — takes the original data, and using the chosen optimum values of $\lambda$ carries out the power transformations. Having stabilised within-group variance, the pooled dispersion matrix is estimated and smoothed. The effective sample sizes (taking missing data into account) are calculated. This is followed by an evaluation of each individual in each group. The five closest sources according to the $D^{2\prime}$ statistic are found, and the significance of departure from each centroid according to $T^{2\prime}$ calculated. This should reveal that most if not all individuals in a certain group can be correctly allocated to that group. The mean distances to the closest five groups are also given. A matrix then follows of the percentage frequencies with which members of each group have been incorrectly allocated to each other group based on distance, not taking into account whether the distance is significantly large or not. This is followed by a matrix which gives the overlap ($\alpha$) for each group, multiplied by 10,000 for convenience. This shows the degree of success achieved by the characterisation method, by presenting a model of the overlap of each source with all others in terms of what is expected from the identification of a random sample of 10,000 artefacts from each source.

The third program — POPPERS/RAZOR 3 — takes the original matrix and estimates missing values according to the multiple regression technique developed by Higham *et al.* (1980). The completed matrix is given as output on to a disk file which can be used for additional operations such as canonical analysis and cluster analysis of the sources, factor analysis of the elements, and so on. In addition, the matrix of

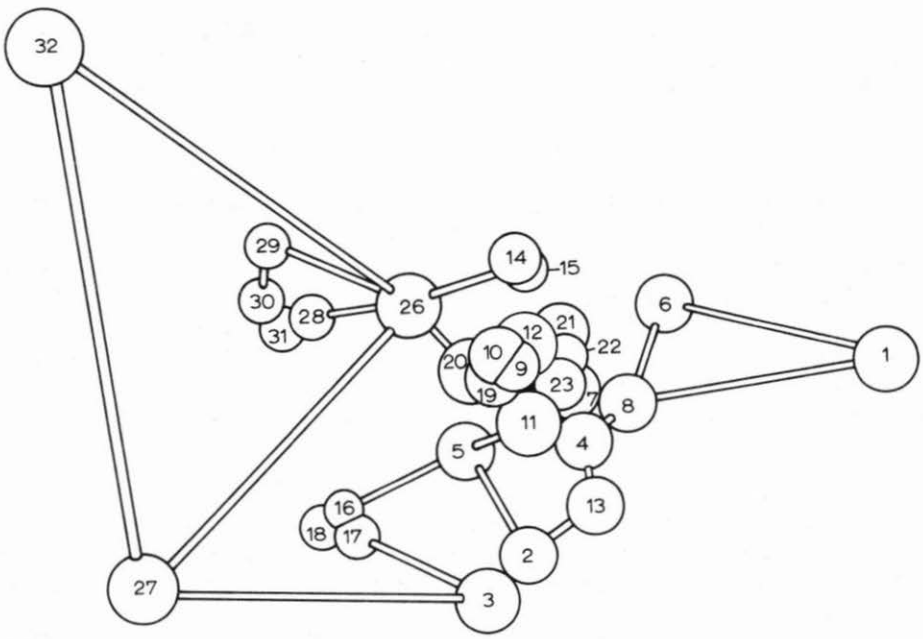


*Figure 9:* Pictogram model showing the relationships between the various Oceanic and New Zealand obsidian sources. This was produced after canonical analysis (see Leach and Warren n.d.). The source numbers follow Table 2.

raw Mahalanobis $D^2$ values resulting from the power transformation is also placed on to an output file. This can be used for developing a pictogram model showing the relationships between the sources using Gower's method of latent root vector analysis (1966).

## APPLICATION OF THE ALGORITHM TO NEUTRON ACTIVATION ANALYSIS DATA FOR NEW ZEALAND AND OCEANIC OBSIDIANS

### BACKGROUND

Some 32 obsidian sources in New Zealand and the Pacific have been recently characterised by neutron activation analysis (NAA) yielding concentration data for 23 elements across 203 samples and 29 artefacts (Leach and Warren n.d.). This is an ideal set of results with which to test out the algorithm. The source names and numbers are given in Table 2. A pictogram of the relationship between each source is given in Figure 9. This results from carrying out canonical analysis on the data.

The first task is to assess the relative merits of using as data the raw concentrations or element ratios. Iron is a suitable element of obsidian ratios with a mean coefficient of variation in this data of $14.4 \pm 1.3\%(\sigma = 7.0)$. Although some elements studied have a marginally lower coefficient of variation, Iron has been used as a normalising element on obsidian previously (McCallum *et al.* 1979) and for the sake of consistency was chosen for this study also. POPPERS/RAZOR was run using the raw concentration data and that of ratios against Iron, and the results compared. In the evaluation test of each source, it was found that correct allocation to the parent group varied from about 50 percent to 100 percent in both cases. By taking ratios, the situation deteriorated in only one case (where the change, $\Delta$, equals 14%) of Awana samples being confused with Te Ahumata (both on Great Barrier Island), and improved markedly (as high as $\Delta 25\%$) in nine cases. The average improvement was $\Delta 14.2\%$ ($\sigma = 6.4$). On this basis, taking ratios appears to be very worthwhile. An even more telling test is with the probability of misclassification ($\alpha$). Multivariate overlap with another group varies from nothing to $\sim 30\%$ in both cases. Six sources deteriorated by taking ratios (as high as $\Delta 8.4\%$) with a mean of $\Delta 3.1\%$ ($\sigma = 3.0$). On the other hand, 17 sources improved by taking ratios (as high as $\Delta 23.4\%$), with a mean of $\Delta 7.7\%(\sigma = 8.6)$. There is little doubt that by taking ratios a significant and worthwhile overall improvement in discrimination is achieved, and this was therefore carried out.

Next, the question of grouping sub-sources was considered. For the New Zealand sources, this subject has already been examined in great detail by Ward (1972) using the interactive $\Delta D^2$ clustering procedure described above. This resulted in a pooling of about 30 sub-sources down to 18 distinctive groups. The NAA research was carried out on samples from these 18 groups, but different hand specimen varieties were kept separate. These, and some Pacific Island sub-sources, should be considered for possible pooling. These are the different coloured varieties from Waihi, Rotorua, Maraetai, and Mayor Island; and four sub-sources from Rapanui.

The Hotelling's $T^2$ test showed that the red and black varieties from Waihi, Rotorua, and Maraetai are all significantly different from each other ($p < .001$). This is sufficient grounds for not pooling these pairs, as positive disadvantages could result (see above). Again, the honey and green varieties from Mayor Island are significantly different ($p < .001$), but the yellow Mayor Island obsidian cannot be easily separated from either of the other two ($p > .5$). This is a rather awkward result, because it could be pooled with either or both of the other two with equal justification. In point of fact, $H_0$ cannot be rejected, largely because $N^* = 0.9$. Given that all the Mayor Island sources were found to be so highly distinct from all

others considered, there is little advantage to be gained by pooling, and they were left separated.

Of the four sources on Rapanui, none of the pairs are significantly different in element composition ($p > .5$); however, they are rather different in hand specimen characteristics. Again, these sources are highly distinct from all others studied, and little benefit would result from pooling. The more conservative approach of leaving them separated was taken.

Three of the elements analysed were only poorly resolved for a number of reasons. These are: Arsenic, Bromine, and Antimony. The question arises — should

**TABLE 2**
**NEW ZEALAND AND PACIFIC OBSIDIAN SOURCES USED IN THE NEUTRON ACTIVATION ANALYSIS STUDY**

| | SOURCE | N |
|---|---|---|
| NORTHLAND | | |
| 1 | Weta | 3 |
| 2 | Waiare | 12 |
| 3 | Pungaere | 9 |
| 4 | Huruiki | 12 |
| GREAT BARRIER AREA | | |
| 5 | Burgess Island | 2 |
| 6 | Fanal Island | 4 |
| 7 | Awana | 7 |
| 8 | Te Ahumata | 12 |
| COROMANDEL-BAY OF PLENTY | | |
| 9 | Cooks Bay | 12 |
| 10 | Purangi | 7 |
| 11 | Hahei and Poikeke Island | 10 |
| 12 | Tairua | 10 |
| 13 | Maratoto | 7 |
| 14 | Waihi red | 10 |
| 15 | Waihi black | 3 |
| 16 | Mayor Island green | 10 |
| 17 | Mayor Island honey | 5 |
| 18 | Mayor Island yellow | 1 |
| INLAND NORTH ISLAND | | |
| 19 | Rotorua red | 6 |
| 20 | Rotorua black | 8 |
| 21 | Maraetai red | 5 |
| 22 | Maraetai black | 7 |
| 23 | Ongaroto | 10 |
| 24 | Taupo | 10 |
| SOUTH ISLAND | | |
| 25 | Canterbury Peninsula | 4 |
| PACIFIC ISLANDS | | |
| 26 | Admiralty Islands | 1 |
| 27 | Pu'u Wa'a Wa'a, Hawaii | 1 |
| 28 | Maunga orito, Rapanui | 6 |
| 29 | Motu iti, Rapanui | 3 |
| 30 | Te Manavai, Rapanui | 4 |
| 31 | Rano kau, Rapanui | 1 |
| 32 | Pitcairn Island | 1 |
| | TOTAL | 203 |

these be simply ignored in the characterisation process? This was assessed using the multivariate overlap statistic ($\alpha$). It was found that by taking out these three elements, 10 sources improved in their characteristics, in so far as source overlap was reduced by up to $\Delta 7.3\%$, with a mean of $\Delta 2.3\%$ ($\sigma = 3.0$); while seven deteriorated with the reduced set of elements. Although a smaller number of sources is involved in those which deteriorated, the change is more marked — up to $\Delta 19.6\%$, with a mean of $\Delta 6.3\%$ ($\sigma = 7.8$). This amount of deterioration was judged to be intolerable, and the elements were therefore left in.

HOW GOOD IS THE CHARACTERISATION METHOD?

We are now in a position to provide a satisfactory answer to Q1, posed at the beginning of this paper — just how distinctive are the sources by this method of characterisation; in other words, how successful is the method in its primary objective? The answer to this is given in Table 3 which identifies the amount of overlap which each source has with each other. It also tells us how often we could expect to identify an artefact from any individual source incorrectly. As might be expected, the largest errors will occur in correctly identifying the exact source in cases where there are several in one general locality. Thus, artefacts from Maunga orito on Rapanui might easily be mis-allocated to Motu iti 20 percent of the time, to Te Manavai on 34 percent of occasions, and to Ranokau some 14 percent. However, the characterisation of Rapanui sources is effective to the extent that less than 1 in 10,000 artefacts from this island would be confused with any other known source analysed. The sources on Mayor Island cluster together in a similar manner, and are again highly distinct from other materials.

Close inspection of the Table will show that sources on the Coromandel Peninsula overlap to some extent with those in the central North Island, and artefacts judged as being from these sources would have to be suitably qualified. For example, about 11 percent of all artefacts made from Rotorua black obsidian could be incorrectly assigned to the source at Purangi on the Coromandel.

Of greater interest perhaps is the fact that obsidian from the Admiralty Islands is not as distinct from the New Zealand sources as we might wish. Fortunately the overlaps are minor, but it is a sobering thought that 2 percent of artefacts from this area of the western Pacific could be incorrectly identified as from Hahei, and a further 2 percent as from Taupo. In cases where results for artefacts suggest long range transportation in the Pacific, it would be wise to view this against the possibilities of mis-classification presented in this Table. For instance, 2 percent source overlap between Taupo and the Admiralty Islands obsidian might be invoked if an artefact of apparent origin in the Admiralties was found in the Cook Islands. Conversely, the identification of an artefact found in New Zealand as of Rapanui origin would gain credibility by consulting this Table, because there is evidently no overlap between the Rapanui sources and those in New Zealand.

On the bottom of the Table is given the cumulative percentage mis-classifications for each source. Note that in cases of extreme source overlap this figure could exceed 100 percent, if artefacts were often closer to several wrong sources than to the true source. On the whole, these figures seem reasonably satisfactory ($\bar{x} = 29\%$, $\sigma = 23\%$), but the characterisation method could not by any means be considered 100 percent effective. The great advantage of this Table is knowing what confidence can be placed in this method; it also adds a dimension of perspective to the canonical analysis model (Figure 9). A slight difficulty is that we have no yardstick for comparison as to how effective this method is against another. To this end, POPPERS/

**TABLE 3**

**MATRIX OF PERCENTAGE INCORRECT ALLOCATIONS EXPECTED USING POPPERS/RAZOR ON 10,000 RANDOM ARTEFACTS FROM EACH SOURCE**

Within-group variance was stabilised with $\lambda = 0.2$. Values between .05-.49 have been rounded to the nearest decimal place (.1-.5), and values >.5 have been rounded to the nearest integer.

| | SOURCE | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| 1 | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | | 13 | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | | | | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | | | | 2 | * | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | | | | .04 | | * | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | | | | 1 | .4 | | * | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | | | | 3 | .2 | .05 | 8 | * | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | | | | 1 | 4 | | .1 | .04 | * | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | | | | .4 | 4 | | .1 | .03 | 31 | * | | | | | | | | | | | | | | | | | | | | | | |
| 11 | | | | .6 | 10 | | .1 | .04 | 6 | 5 | * | | | | | | | | | | | | | | | | | | | | | |
| 12 | | | | .3 | 1 | .01 | .02 | .01 | 2 | 2 | 4 | * | | | | | | | | | | | | | | | | | | | | |
| 13 | | | | | | | .01 | .4 | | | | | * | | | | | | | | | | | | | | | | | | | |
| 14 | | | | | | .01 | | | | | | .1 | | * | | | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | .1 | | 10 | * | | | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | | * | | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | | 25 | * | | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | | | 20 | 19 | * | | | | | | | | | | | | | | |
| 19 | | | | .2 | 1 | | .1 | .1 | 9 | 5 | 1 | 1 | | | | | | | * | | | | | | | | | | | | | |
| 20 | | | | .1 | 1 | | .1 | .01 | 6 | 11 | 2 | 3 | | | | | | | 20 | * | | | | | | | | | | | | |
| 21 | | | | .04 | .5 | .2 | .6 | .1 | .1 | .1 | .4 | 2 | .01 | .02 | | | | | 1 | 1 | * | | | | | | | | | | | |
| 22 | | | | .03 | .5 | .03 | .4 | .04 | .02 | .04 | .1 | 1 | .01 | | | | | | .3 | .4 | 28 | * | | | | | | | | | | |
| 23 | | | | 1 | 1 | .13 | .2 | .1 | .4 | .2 | 1 | 6 | .01 | .04 | .02 | | | | 1 | 1 | 7 | 4 | * | | | | | | | | | |
| 24 | | | | .1 | 1 | .03 | .5 | .05 | .3 | 1 | 1 | 3 | | | | | | | 1 | 3 | 11 | 9 | 3 | * | | | | | | | | |
| 25 | | | | | | | | | | | | | | | | | | | | | | | | | * | | | | | | | |
| 26 | | | | .2 | .2 | | .2 | .01 | .3 | 1 | 2 | 1 | | | | | | | .1 | 1 | .1 | .1 | .1 | 2 | | * | | | | | | |
| 27 | | | | | | | | | | | | | | | | | | | | | | | | | | | * | | | | | |
| 28 | | | | | | | | | | | | | | | | | | | | | | | | | | | | * | | | | |
| 29 | | | | | | | | | | | | | | | | | | | | | | | | | | | | 20 | * | | | |
| 30 | | | | | | | | | | | | | | | | | | | | | | | | | | | | 34 | 25 | * | | |
| 31 | | | | | | | | | | | | | | | | | | | | | | | | | | | | 14 | 8 | 18 | * | |
| 32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | * |
| Σ | 0 | 13 | 13 | 10 | 27 | .5 | 12 | 12 | 60 | 61 | 33 | 27 | .4 | 10 | 10 | 45 | 44 | 39 | 41 | 50 | 52 | 44 | 26 | 36 | 0 | 8 | 0 | 68 | 53 | 77 | 40 | 0 |

**TABLE 4**

**MATRIX OF PERCENTAGE INCORRECT ALLOCATIONS EXPECTED USING POPPERS/RAZOR ON 10,000 RANDOM ARTEFACTS FROM EACH SOURCE**

In this case $\lambda = 1.00$, equivalent to assuming common variance between the sources. The rounding conventions are given in Table 3.

| | SOURCE | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| 1 | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | | 38 | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | 2 | | | * | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | | | | .2 | * | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | .2 | | | 1 | | * | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | 2 | | | 26 | .2 | 1 | * | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | 2 | | | 46 | .2 | 1 | 29 | * | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | .1 | | | 12 | 3 | .1 | 10 | 13 | * | | | | | | | | | | | | | | | | | | | | | | |
| 10 | .1 | | | 10 | 4 | .1 | 10 | 11 | 40 | * | | | | | | | | | | | | | | | | | | | | | |
| 11 | 1 | | | 24 | 1 | .2 | 27 | 27 | 22 | 21 | * | | | | | | | | | | | | | | | | | | | | |
| 12 | .1 | | | 9 | 2 | 1 | 11 | 10 | 14 | 17 | 11 | * | | | | | | | | | | | | | | | | | | | | |
| 13 | 1 | | | 20 | .1 | .2 | 21 | 22 | 11 | 9 | 25 | 5 | * | | | | | | | | | | | | | | | | | | |
| 14 | .01 | | | 1 | | 12 | 1 | 1 | .1 | .1 | .2 | 1 | .1 | * | | | | | | | | | | | | | | | | | |
| 15 | .02 | | | 2 | .01 | 11 | 1 | 2 | .3 | .4 | .5 | 3 | .4 | 35 | * | | | | | | | | | | | | | | | | |
| 16 | .3 | 2 | .4 | | | | | | | | | | | | | * | | | | | | | | | | | | | | | |
| 17 | .2 | 4 | .1 | | | | | | | | | | | | | 31 | * | | | | | | | | | | | | | | |
| 18 | .4 | 2 | 2 | | | | | | | | | | | | | 36 | 36 | * | | | | | | | | | | | | | |
| 19 | .1 | | | 8 | 3 | .1 | 11 | 9 | 25 | 32 | 18 | 22 | 8 | .2 | 1 | | | | * | | | | | | | | | | | | |
| 20 | .04 | | | 8 | 3 | .2 | 11 | 9 | 21 | 27 | 15 | 31 | 6 | .4 | 1 | | | | 37 | * | | | | | | | | | | | |
| 21 | 1 | | | 5 | .02 | 6 | 12 | 6 | 1 | 1 | 5 | 3 | 6 | 1 | 1 | | | | 3 | 3 | * | | | | | | | | | | |
| 22 | 2 | | | 5 | .01 | 4 | 12 | 6 | 1 | 1 | 4 | 2 | 8 | .4 | 1 | | | | 2 | 2 | 35 | * | | | | | | | | | |
| 23 | 2 | | | 20 | .02 | 4 | 20 | 20 | 3 | 3 | 10 | 5 | 14 | 2 | 3 | | | | 3 | 4 | 14 | 16 | * | | | | | | | | |
| 24 | .5 | | | 10 | .1 | 4 | 20 | 11 | 4 | 5 | 10 | 12 | 9 | 2 | 3 | | | | 8 | 10 | 25 | 17 | 17 | * | | | | | | | |
| 25 | | | | | | | | | | | | | | | | | | | | | | | | | * | | | | | | |
| 26 | 1 | | | 13 | .1 | 7 | 20 | 13 | 2 | 3 | 8 | 7 | 6 | 2 | 3 | | | | 3 | 4 | 16 | 12 | 18 | 19 | | * | | | | | |
| 27 | | | | | 3 | | | | .1 | .1 | .01 | .01 | | | | | | | .02 | .02 | | | | | | | * | | | | |
| 28 | .03 | 5 | 3 | | | | | | | | | | | | | 1 | .2 | 1 | | | | | | | | | * | | | | |
| 29 | .2 | 3 | 2 | | | | | | | | | | | | | 1 | .3 | 1 | | | | | | | | | 31 | * | | | |
| 30 | .01 | 4 | 2 | | | | | | | | | | | | | .3 | .1 | .2 | | | | | | | | | 31 | 19 | * | | |
| 31 | .02 | 6 | 3 | | | | | | | | | | | | | 1 | .2 | 1 | | | | | | | | | 41 | 29 | 28 | * | |
| 32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | * |
| Σ | 16 | 64 | 51 | 222 | 20 | 53 | 245 | 238 | 183 | 195 | 230 | 166 | 172 | 60 | 69 | 73 | 72 | 80 | 193 | 193 | 144 | 130 | 178 | 187 | 0 | 157 | 3 | 113 | 87 | 85 | 109 | 0 |

RAZOR was also run on the same data, but with $\lambda = 1.00$, and the elements as concentrations rather than as ratios. This should be a suitable test of the increased effectiveness, if any, of the algorithm developed. Under these conditions, POPPERS/RAZOR reverts to something similar to the old and more familiar algorithm for discriminant analysis, although superior in some respects. The results are given in Table 4. It does not require much comparison of the two Tables to realise that the degree of improvement rendered by the new algorithm is massive. In the case of the older technique, source overlaps are far greater, and misidentifications probable rather than possible. The mean cumulative percentage of mis-classification is 118% ($\sigma = 76\%$). It is hard to imagine that the NAA data could be used at all for artefact identifications under these circumstances.

## THE IDENTIFICATION OF ARTEFACTS

We can now turn our attention to the central issue, that of Q2 — where does this or that artefact come from? It could be thought that with a characterisation method involving 23 elements, there could not possibly be any ambiguity about the origin of an artefact. This is manifestly untrue, and developing a statistic which will give reliable answers is very important. It has been argued on theoretical grounds (see above) that the minimum Mahalanobis distance will not necessarily designate the correct source of an artefact; that failure to take account of unequal variance between sources will lead to incorrect identifications; and that methods which assume that the correct source is present amongst the alternatives offered will give untrustworthy results. It remains to be seen whether the particular characterisation method examined here is so robust that the differences between the various identification statistics are trivial. The proof of the pudding, after all, is in the eating — the alternatives may all give the same answer anyway!

A suitable test would be to try out various algorithms on some artefacts where we have some archaeological clues on possible origins already. The artefacts in this semi-controlled test are in three groups as follows:

| | |
|---|---|
| *Group A:* | Source fairly certain to be Mayor Island (sources #16, #17, #18). |
| GW362 | A large block of obsidian collected by Ward from one of the obsidian sources in New Zealand and in the comparative collection of the Otago Anthropology Department. Hand specimen characteristics are consistent with an origin from Mayor Island, but the piece had lost its accession number. |
| D75.595 | A large block on display in the Otago Museum, long assumed to be of Mayor Island origin. |
| D75.207 | A large block in the Otago Museum collected at Long Beach, and again assumed to be of Mayor Island origin. |
| *Group B:* | A source, or sources, likely to be in New Zealand. If so, Mayor Island is fairly certain for most pieces. |
| AA518-523 | Six obsidian artefacts from the Waihora site in the Chatham Islands. |
| AA524-531 | Eight further artefacts from Pitt Island in the Chathams. |
| | All these samples are green in transmitted light, one of the characteristics of (but not confined to) Mayor Island. Two exceptions to a vitreous texture are AA526, and AA528 which have a distinctly matt surface. Some of the Rapanui obsidian is a very similar matt green. There are good archaeological grounds for supposing that some type of connection between early people in New Zealand and the Chathams existed. A New Zealand source for these artefacts would be consistent with this view. |
| *Group C:* | A source or sources outside the Pacific area, and not amongst those characterised. |
| 314 | Ras Shamra |
| 337 | Sarab |
| 341 | Jarmo |
| 348 | Tell Shemsharah |
| 354 | Jarmo |
| 368 | Bougras |

| | |
|---|---|
| AG16 | Meyden |
| AG36 | Nemrut |
| AG79 | Harman Tepe |
| AG38 | Taskin |
| H1 | Forgia Vecchia |
| N/04 | Rocche Rosse |

These samples are all from the Middle East, and we can rule out an origin from any of the Pacific or New Zealand sources studied here.

Several separate attempts were made to identify the origin of these artefacts. Firstly, linear discriminant analysis was performed on all the sources and the largest discriminant function (LDF) and associated probability ($P_{ak}$) found for each artefact using a program called DISCOTEK (Leach 1969). The results are presented in Table 5. According to this method, 22 of the artefacts can be allocated to a known source with high probability, and seven must be rejected as from unknown sources (asterisked). Apart from two artefacts, all those from Groups A and B appear to be from Mayor Island as expected. Curiously, however, amongst Group C, seven of the 12 artefacts are also allocated to known sources with high probability. Who would have thought that an artefact from Tell Shemsharah had its origin at Te Ahumata on Great Barrier Island?

The main difficulty with this method is that we have no way of knowing when we can trust it, and when we can't — it is simply erratic. Other than on *a priori* grounds, there is no way of distinguishing whether the result for a piece from the Chathams is any more reliable than that for a piece from Jarmo. In short, this controlled test shows that the discriminant function method is logically unsuited to the problem of sourcing artefacts, and should be abandoned.

Secondly, as already mentioned, when $\lambda = 1.00$ POPPERS/RAZOR reverts to a superior form of discriminant analysis where the troublesome assumption of $\Sigma p = 1.00$ does not have to be made. In cases where the MMD is significant, a possible source can be searched for amongst the remainder. This type of identification was also performed on the artefacts, and the results may be compared with those previously obtained in Table 5. The answers by this method initially appear to be more satisfactory in that a source can be found for every artefact without problems of significance. Other than the sample from Rocche Rosse in the Middle East, all distances are less than 2 standardised units from the centroid of the closest source; and even for this sample ($D^{2\prime} = 2.4$) is not judged to be significant by the $T^{2\prime}$ test. Once again, it is clear that there is difficulty in rejecting sources when the true source of origin of an artefact is not amongst those offered. This was predicted on theoretical grounds above. The artefacts in Groups A and B, for which we have some clues as to origin, seem to have been identified approximately in line with our expectations. The first three are from Mayor Island as predicted, and most of the Chathams artefacts are also designated as from this source. The remaining six of the Chathams artefacts have been allocated to Rapanui sources. Such a conclusion would be of great importance archaeologically, and the hand specimen similarity of two of the artefacts to the Rapanui material has already been noted. Can these results be trusted though? In our view, the failure to demonstrate any power to reject the null hypothesis in known cases where it is false undermines any confidence in the method. There are no good grounds for thinking that the conclusion that some of the Chathams artefacts are from Mayor Island is any more reliable than the conclusion that our Tell Shemsharah artefact now appears to derive from Fanal Island in the Great Barrier area!

Thirdly, POPPERS/RAZOR was attempted on the artefacts with $\lambda = 0.2$, and the results are also given in Table 5. This time, it is hoped, the results can be relied upon.

**TABLE 5**

**ARTEFACT IDENTIFICATIONS ACCORDING TO THREE DIFFERENT METHODS**

See the text for detailed discussion. The asterisks (*) indicate values of $P_{ak}$ which are significant ($p = .05$). LDF = Largest discriminant function. MMD = Minimum Mahalanobis distance. † = particularly noteworthy results (see text). Significance levels throughout this paper are as follows: NS (Not significant) $p > .05$, PS (Possibly significant) $p \leqq .05 > .01$, S (Significant) $p \leqq .01 > .001$, and HS (Highly significant) $p \leqq .001$.

| | Discriminant Analysis | | POPPERS/RAZOR ($\lambda = 1.00$) | | | POPPERS/RAZOR ($\lambda = 0.20$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample | LDF | $P_{ak}$ | MMD | $D^{2\prime}$ | Sig | MMD | $D^{2\prime}$ | Sig | Other close sources |
| GW362 | #18 | .982 | #18 | 1.0 | NS | #16 | 1.4 | NS | #17 NS, #18 NS |
| D75.595 | #18 | .995 | #18 | 0.7 | NS | #16 | 2.4 | PS | #18 NS |
| D75.207 | #18 | .998 | #18 | 0.6 | NS | #16 | 2.8 | S | #18 NS |
| AA518 | #18 | .999 | #31 | 1.0 | NS | #16 | 2.3 | PS | #18 NS |
| AA519 | #18 | 1.000 | #18 | 0.5 | NS | #16 | 3.0 | S | #18 NS |
| AA520 | #18 | .999 | #18 | 1.4 | NS | #16 | 2.9 | S | #18 PS |
| AA521 | #18 | 1.000 | #18 | 0.5 | NS | #16 | 2.9 | S | #18 NS |
| AA522 | #18 | .999 | #18 | 1.1 | NS | #16 | 2.4 | PS | #18 NS |
| AA523 | #18 | .999 | #31 | 0.6 | NS | #16 | 2.3 | PS | #18 NS |
| AA524 | #18 | .945* | #31 | 0.5 | NS | #16 | 2.3 | PS | #18 NS |
| AA525 | #18 | .999 | #18 | 0.4 | NS | #16 | 2.7 | PS | #18 NS |
| AA526 | #18 | .999 | #18 | 0.7 | NS | #16 | 5.7 | HS | #31, $D^{2\prime} = 23.6$, HS |
| AA527 | #18 | .999 | #18 | 0.7 | NS | #16 | 3.4 | S | #18 PS |
| AA528 | #18 | .985 | #18 | 0.7 | NS | #17 | 7.0 | HS | #27, $D^{2\prime} = 30.2$, HS |
| AA529 | #18 | .997 | #30 | 1.5 | NS | #16 | 3.4 | S | #18 PS |
| AA530 | #18 | .992 | #31 | 0.7 | NS | #16 | 2.5 | PS | #18 NS |
| AA531 | #18 | .723* | #31 | 0.8 | NS | #16 | 1.8 | NS | #18 NS |
| 314 | #23 | .599* | #22 | 1.6 | NS | # 6 | 7.5 | HS | |
| 337 | #23 | .906* | # 6 | 0.4 | NS | # 6 | 5.9 | HS | |
| 341 | # 7 | .567* | # 6 | 0.5 | NS | # 6 | 6.4 | HS | |
| 348 | # 8 | .980 | # 6 | 1.0 | NS | # 6 | 6.4 | HS | |
| 354 | # 8 | .952* | #23 | 1.8 | NS | # 6 | 7.0 | HS | |
| 368 | # 8 | .391* | #22 | 1.6 | NS | # 6 | 6.7 | HS | |
| AG16 | # 8 | 1.000 | # 1 | 0.7 | NS | # 1 | 12.0 | HS | |
| AG36 | #18 | 1.000 | #16 | 1.2 | NS | #18 | 10.6 | HS | |
| AG79 | #11 | 1.000 | # 4 | 0.3 | NS | # 8 | 2.0 | NS† | # 4, $D^{2\prime} = 2.5$, PS |
| AG38 | #11 | 1.000 | # 4 | 0.2 | NS | # 4 | 1.9 | NS† | # 8, $D^{2\prime} = 2.1$, NS |
| H1 | # 8 | 1.000 | # 1 | 1.2 | NS | # 1 | 9.7 | HS | |
| N/04 | # 8 | 1.000 | # 1 | 2.4 | NS | # 1 | 16.5 | HS | |

**TABLE 6**

**COMPARISON OF TWO EXCEPTIONAL SAMPLES WITH TWO SOURCES JUDGED TO BE SIMILAR BY POPPERS/RAZOR**

The concentrations are wt% for NA and FE; the remainder are ppm. The $\Delta\sigma$ values indicate the number of standard deviations the sample is away from the source compared. NB: These values were calculated using trailing digits not shown in this Table. Asterisked (*) values are $>3\sigma$ from the source mean. See text for discussion.

| | #4 Huruiki | | #8 Te Ahumata | | AG38 | | | AG79 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ | $\sigma$ | $\bar{x}$ | $\sigma$ | Conc | $\Delta\sigma$#4 | $\Delta\sigma$#8 | Conc | $\Delta\sigma$#4 | $\Delta\sigma$#8 |
| Na | 3.4 | ± 0.3 | 2.9 | ± 0.5 | 3.0 | *1.33* | *0.20* | 3.9 | *1.67* | *2.00* |
| Sc | 4.2 | ± 1.0 | 4.3 | ± 0.7 | 3.7 | *0.52* | *0.78* | 4.5 | *0.26* | *0.30* |
| Fe | 1.0 | ± 0.2 | 0.9 | ± 0.1 | 1.0 | *0.00* | *1.00* | 1.2 | *1.00* | *3.00** |
| Co | 0.8 | ± 0.2 | 0.8 | ± 0.2 | 0.6 | *0.73* | *0.67* | 0.6 | *1.03* | *1.05* |
| As | 13 | ± 10 | 6 | ± – | – | — | — | – | — | — |
| Br | 1 | ± 1 | 1 | ± 1 | 0.2 | *0.69* | *1.95* | 0.2 | *0.77* | *2.09* |
| Rb | 148.2 | ± 18.9 | 211.5 | ± 45.8 | 142 | *0.32* | *1.52* | 187 | *2.05* | *0.54* |
| Zr | 121.9 | ± 33.5 | 121.9 | ± 25.8 | 139 | *0.51* | *0.66* | 185 | *1.88* | *2.43* |
| Sb | 0.3 | ± – | – | ± – | 0.5 | — | — | 0.6 | — | — |
| Cs | 9.6 | ± 1.0 | 11.4 | ± 1.5 | 6 | *3.60** | *3.60** | 8 | *1.60* | *2.27* |
| Ba | 566 | ± 170 | 524 | ± 85 | 539 | *0.16* | *0.18* | 686 | *0.71* | *1.91* |
| La | 33.4 | ± 3.2 | 39.1 | ± 6.4 | 37 | *1.25* | *0.33* | 44 | *3.44** | *0.77* |
| Ce | 57.4 | ± 5.7 | 69.0 | ± 8.1 | 62 | *0.88* | *0.86* | 75 | *3.16** | *0.74* |
| Nd | 30.8 | ± 5.5 | 34.3 | ± 6.3 | 32 | *0.18* | *0.37* | 34 | *0.55* | *0.05* |
| Sm | 6.1 | ± 0.7 | 6.8 | ± 1.2 | 6 | *0.00* | *0.67* | 7 | *1.43* | *0.17* |
| Eu | 0.5 | ± 0.1 | 0.4 | ± 0.1 | 0.5 | *0.67* | *1.55* | 0.7 | *1.85* | *3.67** |
| Tb | 1.0 | ± 0.2 | 1.1 | ± 0.2 | 1.0 | *0.14* | *0.42* | 1.3 | *1.30* | *1.22* |
| Yb | 8.3 | ± 1.0 | 8.1 | ± 1.5 | 7 | *1.00* | *0.73* | 9 | *1.00* | *0.60* |
| Lu | 0.5 | ± 0.1 | 0.6 | ± 0.1 | 0.4 | *1.45* | *1.20* | – | — | — |
| Hf | 4.6 | ± 0.4 | 4.1 | ± 0.6 | 4 | *1.50* | *0.17* | 5 | *1.00* | *1.50* |
| Ta | 1.1 | ± 0.2 | 0.9 | ± 0.2 | 1 | *0.50* | *0.50* | 1 | *0.50* | *0.50* |
| Th | 12.4 | ± 1.3 | 21.9 | ± 3.1 | 17 | *3.54** | *1.58* | 21 | *6.62** | *0.29* |
| U | 3.4 | ± 0.7 | 4.8 | ± 0.8 | 4 | *0.86* | *1.00* | 5 | *2.29* | *0.25* |
| | | | | | $\bar{x}$ | *0.94* | *0.95* | | *1.71* | *1.27* |
| | | | | | $\sigma$ | *0.98* | *0.79* | | *1.43* | *1.04* |
| | | | | | $D^{2\prime}$ | *1.9* | *2.1* | | *2.5* | *2.0* |
| | | | | | $T^{2\prime}$ | *NS* | *NS* | | *PS* | *NS* |

The first test is: has it any power to refute incorrect guesses? The results for the foreign pieces (Group C) suggests that it does have this ability. It will be seen that with two important exceptions (†), these artefacts are located at highly significant distances of 5.9 to 16.5 standardised units ($D^{2\prime}$) from the nearest known sources.

The two exceptions, artefacts #AG79 and #AG38 from Harman Tepe and Taskin, are found to be both rather similar to sources #14:Huruiki, and #8:Te Ahumata in New Zealand. It is not possible that there was any prehistoric connection between the Middle East and New Zealand, and yet it is scarcely credible that obsidian sources could be found in each area which could be so similar. These two cases are examined in detail in Table 6. This gives the element composition, together with the known variation for each of the two sources concerned, and the composition of the two samples. Inspection of the figures will reveal that the two samples really are remarkably similar to the two New Zealand sources mentioned. A few values are $>3\sigma$ from the source means, but this alone does not necessarily rule out an origin from a particular source. For instance, #AG79 has 1.2% Fe, and is $3\sigma$ from the mean of Te Ahumata. Amongst the 12 samples which define this source, one specimen has 1.3% Fe (GS110Z). It is interesting to note that the average values for $\Delta\sigma$ given in the Table closely follow those of the $D^{2\prime}$ statistic. On the whole, the $D^{2\prime}$ statistic, and the $T^{2\prime}$ significance test appear, even in these two cases, to be functioning precisely in the manner desired. We are forced to conclude that on the basis of these elements studied at least, there is an obsidian source in the Middle East area with astonishing resemblance to two in New Zealand. It will be noted, incidentally, that overlap between these two New Zealand sources is of the order of about 3 percent (see Table 3).

Having survived this controlled test a little battered, but more or less intact, we can now view the results of the remaining artefacts with rather more confidence. The first thing which is notable is that, as predicted, the MMD alone is not the only criterion to consider in identifying artefacts. For instance, it will be observed that several of the artefacts are significantly different from the closest source, and not significantly different from another source slightly further away. The computer program is designed to test the five nearest sources to any artefact, and often finds several which are insignificantly different. Coming to a balanced conclusion about the origin of each artefact requires taking into account several aspects of the results. Sufficient of this is given in Table 5 for a summary conclusion that most of the artefacts in Groups A and B can be classified as Mayor Island in origin with conviction. Two examples from the Chathams are separated by 5.7 and 7.0 standardised distances respectively from Mayor Island. These are the two samples #AA526 and #AA528 which were previously mentioned as having hand specimen characteristics unfamiliar at Mayor Island. The next closest sources for these two artefacts are on Rapanui, but $D^{2\prime}$ is 23.6 and 30.2 units respectively — this source is therefore out of the question. These two artefacts, therefore, must be classed as from an as yet "unknown" obsidian source sharing features in common with both Mayor Island and Rapanui.

It is interesting that none of these artefacts from the Chatham Islands are especially close to the centroids of the Mayor Island sources. The coefficients of variation for the elements for this assemblage of 14 artefacts was looked at. Compositional variability is markedly less than the known source variation at Mayor Island. This raises the distinct possibility that only one large block of this obsidian ever found its way to the Chatham Islands from New Zealand. A single biased sample would explain both the atypical character of the assemblage, and its low compositional variability.

## APPLICATION OF THE ALGORITHM TO XRF DATA OF NEW ZEALAND OBSIDIANS

In 1972 Ward carried out a project of characterisation of the New Zealand obsidian sources, by wavelength-dispersive XRF analysis. He analysed Zirconium, Manganese, Titanium, Rubidium and Strontium for 216 powdered samples from about 30 localities and grouped them into 18 distinctive sources. Since then, some 213 artefacts have been similarly analysed, and linear discriminant analysis used to identify their source of origin. As an additional test of the algorithm developed here, it was considered worthwhile to apply it to this XRF data. Non-destructive energy-dispersive XRF has largely superseded this type of analysis of artefacts now, but the question arises as to how reliable these early source identifications really were.

### HOW GOOD WAS THE METHOD OF SOURCE CHARACTERISATION?

The five elements analysed have relatively low correlation co-efficients, and are therefore a good choice for maximising group discrimination. It was decided to normalise the data by taking ratios to Zirconium, in order to cut down on any problems of machine variability. This element has a rather higher co-efficient of variation than Iron (approximately 22%), but was a reasonable choice amongst those available. Within-group variance was again stabilised by the Box and Cox transformation ($\lambda = 0.2$). The multivariate overlaps between sources are shown in Table 7. On the whole, these are very good results considering the small number of

**TABLE 7**
**MATRIX OF PERCENTAGE MULTIVARIATE OVERLAP USING POPPERS/RAZOR ON THE XRF DATA ON NEW ZEALAND OBSIDIAN SOURCES**

In this case $\lambda = 0.2$, and the elements were turned into ratios against Zirconium. The source numbers follow those in Table 2. NB: [1] = sources #14 & #15; [2] = sources #16, #17 & #18; [3] = sources #19 & #20; [4] = sources #21 & #22.

| | SOURCE | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14[1] | 16[2] | 19[3] | 21[4] | 23 | 24 |
| 1 | . | | | | | | | | | | | | | | | | | |
| 2 | | . | | | | | | | | | | | | | | | | |
| 3 | | 19 | . | | | | | | | | | | | | | | | |
| 4 | | | | . | | | | | | | | | | | | | | |
| 6 | | | | .1 | . | | | | | | | | | | | | | |
| 7 | | | | 4 | | . | | | | | | | | | | | | |
| 8 | | | | 1 | .02 | 7 | . | | | | | | | | | | | |
| 9 | | | | .03 | | | | . | | | | | | | | | | |
| 10 | | | | .1 | | .01 | | 6 | . | | | | | | | | | |
| 11 | | | | .02 | | | | 4 | 15 | . | | | | | | | | |
| 12 | | | | .01 | | | | .1 | .01 | | . | | | | | | | |
| 13 | | | | | | .1 | | | | | | . | | | | | | |
| 14[1] | | | | | | | | | | | 2 | | . | | | | | |
| 16[2] | | | | | | | | | | | | | | . | | | | |
| 19[3] | | | | 1 | | .1 | | 9 | 4 | 3 | .1 | | | | . | | | |
| 21[4] | | | | .5 | | .01 | | 1 | .1 | .1 | 2 | | .04 | | 6 | . | | |
| 23 | | | | 1 | | .01 | | 2 | 1 | .2 | 4 | | .03 | | 9 | 21 | . | |
| 24 | | | | .1 | | | | .3 | .03 | .01 | 14 | | 1 | | 1 | 12 | 14 | . |
| Σ | 0 | 19 | 19 | 8 | .1 | 11 | 8 | 22 | 26 | 22 | 22 | .1 | 3 | 0 | 33 | 43 | 52 | 42 |

elements analysed. Relatively large overlaps occur between Waiare/Pungaere 19% (cf. 13% by NAA), Purangi/Hahei 15% (cf. 5%), and the Maraetai/Ongaroto/Taupo cluster average about 16% (cf. 6%).

## THE IDENTIFICATION OF ARTEFACTS

Despite the reasonably satisfactory discrimination between sources by this method, when it comes to identifying artefacts some difficulty is encountered. This is because the between-group multivariate distances established by this technique are fairly small — considerably lower than is the case with the large number of elements in the NAA research, for example. These distances are as important as the matter of whether the sources actually overlap. This can be made clear by considering a specific example. The scaled $D^2{}'$ values for the source pair Huruiki/Awana are about the same for both the XRF analysis and the NAA — 5.0 and 3.5 respectively. However, the Mahalanobis distances are $D^2_4 = 20$ and $D^2_{20} = 70$ respectively. In the XRF case, because r* is only 4, the region of confidence around an artefact (N = 1) which actually belongs to Huruiki (for argument sake) may well overlap with Awana as well, despite the fact that these two sources only share 4 percent common space (sources #4 and #7 see Table 7). In the case of the NAA results, r* is at least 20, and the region of confidence of a Huruiki artefact will be much further away from the Awana source, which this time shares about 1 percent common space (sources #4 and #7 see Table 3). The statistic used for rejecting the null hypothesis (Hotelling's $T^2$) is both powerful and conservative; $T^2$ is proportional to $D^2$, and the test is likely to fail with $D^2_4 = 20$, where it succeeds with $D^2_{20} = 70$. Thus, it is a great advantage in identifying artefacts not only to have shown that the sources do not overlap, but that they are effectively so far in the distance, they are out of sight as well.

Using the $T^2{}'$ statistic, the 213 artefacts previously analysed were re-identified with the following results:

(i) Mayor Island: 111 artefacts (including 25 PS)

(ii) Cooks Bay: 22 artefacts
Only one of these could be identified with full confidence to this source alone. The remainder are insignificantly different from an average of a further three sources. Commonly, these include Hahei, Rotorua, Purangi and Taupo; and more rarely Ongaroto and Maraetai.

(iii) Hahei: 5 artefacts
These are variously confused with Cooks Bay, Rotorua, Taupo, Ongaroto, Waihi, Tairua and Purangi.

(iv) Ongaroto: 4 artefacts
Again, the identifications are usually confused with Taupo, Maraetai, Hahei and Rotorua.

(v) Rotorua: 3 artefacts
Here there is confusion with Maraetai, Taupo, Hahei and Cooks Bay.

(vi) Purangi: 1 artefact
This could also belong to Hahei, Taupo, Cooks Bay, or Ongaroto.

(vii) Maraetai: 1 artefact
This could also belong to Ongaroto, Taupo, Rotorua, or Hahei.

(viii) No Known Source: 66 artefacts

An embarrassing feature of these identifications is that they are considerably different from those established by the $P_{ak}$ statistic — results of which are published and widely cited (for example, see Reeves and Ward 1976:281; Davidson 1979:243; 1981:18). This cautious reappraisal of the basic assumptions involved in sourcing statistics has revealed more than was bargained for; and shows just how careful we should be in this whole field.

The large number of artefacts for which a source could not be found (31 percent) is remarkable. The $D^2{}'$ values for these were carefully looked at, and it is clear that they cluster in several regions of the hyperspace distribution. It is a large enough series to attempt decomposing into multivariate components which would help to define the several unknown sources (after Wolfe 1970), although this is beyond the scope of this present study. Some of the artefacts may well have been poorly analysed, and their oddness would reflect this. Equally, there are clear patterns in the element data too, and this is suggestive of more than one genuine missing source. Fifty of the artefacts are nearest to Mayor Island with a mean $D^2{}'$ value of 15.2 standardised units from this centroid. Of these 50, a group of 28 are conspicuous in having very low Rubidium concentrations of from 1 to 100ppm ($\bar{x}=66.6$, $\sigma=29.6$). By contrast, the closest source (Mayor Island) has a mean of 190ppm ($\sigma=8.0$).

## CONCLUSIONS

Techniques designed to identify the original source of artefacts on the basis of multi-element characteristics have been largely accepted on trust by archaeologists in the past. It is argued that these are far less precise than is commonly thought, and this places an extra burden of proof on the designers of these schemes to test them out more fully before identifying artefacts. Some suggestions have been made as to what we consider to be the minimum standards of proof for a method to be classed as acceptable. These criteria are essentially aimed at ruthlessly revealing the weaknesses of a scheme rather than its strengths, and this follows Karl Popper's approach of "conjectures and refutations" (Popper 1972).

An algorithm is developed which thoroughly assesses the degree of discrimination achieved by any particular method of source characterisation.[1] This assessment is a suitable background against which an archaeologist can gauge the reliability of particular artefact identifications which are made by this method.

This algorithm was applied to two sets of multi-element data on about 30 sources of New Zealand and Oceanic obsidian as test cases of the method. In one case (NAA), information was available for 23 elements, and in the other (XRF), for only 5. Although discrimination between sources in both cases was fairly satisfactory, the reliability with which artefacts could be unequivocally matched to them is far from ideal in either case. On the whole, neither scheme is as robust as first impressions suggested.

It would not have been possible to carry out the re-evaluation of the XRF system of artefact identifications by this algorithm had the data not been readily available to other researchers. This highlights the need for a permanent record file of such data (perhaps in microfiche), so that it can be re-examined at later dates, as our knowledge of obsidian sources improves. There is now a great deal of such information being generated by archaeologists and their associates, and any action on this matter would need to be taken soon.

Finally, it should be pointed out that the problems highlighted in this paper are not confined to obsidian artefact identification, or even lithic sourcing in general. Multivariate discrimination is a very widespread method of analysis, applied in fields ranging from sediment identification to the study of hominid evolution. One of the most questionable and at the same time most hazardous assumptions is the *a priori* judgement that a particular subject does belong to one of the previously defined alternative groups. The implications of this for studies of fossil hominids were recently touched upon by Wilson (1979:32). In the case of obsidian sourcing, a close look at such basic assumptions has shown where serious errors could result unless

counter measures are taken. It may be wise for others to examine their assumptions too.

Note

1. Copies of the program set POPPERS/RAZOR may be obtained from the senior author.

ACKNOWLEDGEMENTS

The ideas behind this paper have been fermenting for nearly a decade; during this time, exchanges with a number of colleagues have been most profitable. Thanks are particularly due to Graeme Ward and Sir Karl Popper. Exposing the weaknesses of one's work, as well as its strengths, is far from typical in archaeology — in this respect, we are especially grateful to Graeme Ward. Without full access to his original data, the re-examination presented here would not have been possible. The illustrations in this paper were re-drawn by Martin Fisher, and the manuscript typed by Ann Trappitt; for this assistance, our thanks are due.

REFERENCES

Ambrose, W. R., Bird, J. R. and Duerden, P. 1981. The impermanence of obsidian sources in Melanesia. *In* Leach, B. F. and Davidson, J. M. (Eds) *Archaeological studies of Pacific stone resources*: 1-19. British Archaeological Reports 104.

Anderson, T. W. 1957. *An introduction to multivariate statistics.* Ching Sheng, Formosa.

Box, G. E. P. and Cox, D. R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society* B26:211-252.

Constandse-Westermann, T. S. 1972. *Coefficients of biological distance.* Humanities Press, New York.

Davidson, J. M. 1979. New Zealand. *In* Jennings, J. D. (Ed.) *The Prehistory of Polynesia*: 222-248. Australian National University Press, Canberra.

Davidson, J. M. 1981. The Polynesian foundation. *In* Oliver, W. H. and Williams, B. R. (Eds) *The Oxford History of New Zealand*: 3-27. Oxford University Press, Wellington.

Dixon, J. E. 1976. Obsidian characterization studies in the Mediterranean and Near East. *In* Taylor, R. E. (Ed.) *Advances in obsidian glass studies: archaeological and geochemical perspectives*: 288-333. Noyes Press, Park Ridge, New Jersey.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325-338.

Higham, C. F. W., Kijngam, A. and Manly, B. F. J. 1980. An analysis of prehistoric canid remains from Thailand. *Journal of Archaeological Science* 7:149-165.

Huseby, J. R., Schwertman, N. C. and Allen, D. M. 1980. Computation of the mean vector and dispersion matrix for incomplete multivariate data. *Communications in Statistics — computation and simulation* B9(3):301-309.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. *In* Lakatos, I. and Musgrave, A. (Eds) *Criticism and the growth of knowledge*: 91-196. Cambridge University Press.

Leach, B. F. 1969. *The concept of similarity in prehistoric studies.* Studies in Prehistoric Anthropology 1. Anthropology Department, University of Otago.

Leach, B. F. 1973. Obsidian in the Chatham Islands. *New Zealand Archaeological Association Newsletter* 16(3):104-106.

Leach, B. F. and Anderson, A. J. 1978. The prehistoric sources of Palliser Bay obsidian. *Journal of Archaeological Science* 5:301-307.

Leach, B. F. and Warren, S. E. n.d. The chemical composition of New Zealand and Oceanic obsidians from neutron activation analysis. Manuscript.

Leach, B. F. and Warren, S. E. 1981. Neutron activation analysis of New Zealand and Oceanic obsidians: towards a simple screening technique. *In* Leach, B. F. and Davidson, J. M. (Eds) *Archaeological Studies of Pacific Stone Resources*: 151-166. British Archaeological Reports 104.

McCallum, G. J., McFadgen, B. G., Morre, B. G., Richardson, L. M. and Sheppard, R. A. 1979. Characterization of New Zealand obsidian by energy dispersive X-ray fluorescence. INS MS-1001, Institute of Nuclear Science, Wellington.

Mahalanobis, P. 1930. On tests and measures of group divergence. *Journal of the Asiatic Society of Bengal* 26:541-588.

Mahalanobis, P. 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Science, India* 2:49-55.

Popper, K. R. 1972. *Conjectures and refutations: the growth of scientific knowledge.* Routledge and Kegan Paul, London.

Rao, C. R. 1952. *Advanced statistics methods in biometric research.* John Wiley and Sons, New York.

Reeves, R. D. and Ward, G. K. 1976. Characterization studies of New Zealand obsidians: towards a regional prehistory. *In* Taylor, R. E. (Ed.) *Advances in obsidian glass studies*: 259-287. Noyes Press, Park Ridge, New Jersey.

Schwertman, N. C. and Allen, D. M. 1979. Smoothing an indefinite variance-covariance matrix. *Journal of Statistical computation and simulation* 9:183-194.

Sneath, P. H. A. and Sokal, R. R. 1973. *Numerical Taxonomy.* Freeman and Co., San Francisco.

Ward, G. K. 1972. *Obsidian and New Zealand archaeology.* MA Thesis, Anthropology Department, University of Otago.

Ward, G. K. 1974a. A paradigm for sourcing New Zealand archaeological obsidians. *Journal of the Royal Society of New Zealand* 4:47-62.

Ward, G. K. 1974b. A systematic approach to the definition of sources of raw material. *Archaeometry* 16:41-53.

Ward, G. K. 1974c. Source of obsidians from the Motutapu undefended site N38/37. *Records of the Auckland Institute and Museum* 11:13-14.

Wilson, S. R. 1979. Comments on: Rightmire, G. P. 1979. Implications of Border Cave skeletal remains for later Pleistocene human evolution. *Current Anthropology* 20(1):23-35.

Wolfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. *Multivariate Behaviour Research* 5:329-350.

*Received 24 December 1981*