24

The chemical composition of tiles from Bordesley: a case study in data treatment

M. N. Leese*

M. J. Hughes†

J. Stopford‡

24.1 Introduction

24.1.1 Aims of the paper

This paper describes a study of the chemical composition of decorated floor tiles found during the course of the excavations initiated at the Cistercian Abbey at Bordesley near Redditch, Worcestershire, by Philip Rahtz, and now directed by Susan Hirst, Susan Wright and Grenville Astill (Hirst *et al.* 1983). In order to extend the interpretation of the tiles, previously undertaken on the basis of manufacture and design, it was decided to analyse a subset of them by neutron activation analysis to determine their trace element composition. Also analysed were a few tiles from sites thought to be related to Bordesley (because tiles of similar appearance had been found there), as well as 13 clay samples from the immediate vicinity of the abbey. The methods of scientific examination used in a similar study of tile compositions are described in Leese *et al.* 1986.

This paper concentrates on the data analysis used to group together chemically similar tiles and to relate them to other types of information, such as the designs. While the specific findings relate strictly only to this data set, it was felt that the conclusions relating to data analysis techniques might have some wider relevance since the problems encountered, in particular the need to take account of dilution of clay by temper, are typical of many data sets concerned with ceramic composition.

The data analysis was aimed at finding groups of tiles of similar composition. This involved informal visual assessment of the data to eliminate unreliable variables and to get a general feel for the data, clustering the tiles together in order to reflect common sources and/or manufacturing techniques if possible, and finally examining the constituents of the resulting clusters. This paper does not cover clustering algorithms as such; this is because there is reasonable agreement that standard techniques such as average linkage and Ward's method seem to work well. It was felt that there were other, relatively neglected, aspects of data analysis which are equally important. These are:

1. What types of visual aids are the most helpful in the initial stages of the study, before clustering?

2. What measure of chemical similarity should be used in clustering items together, given that chemical concentrations may have been diluted?

* British Museum Research Laboratory
  London WC1B 3DG

† British Museum Research Laboratory
  London WC1B 3DG

‡ Centre for Mediaeval Studies
  University of York

3. Having found the clusters, what techniques are useful for interpreting them in relation to other information?

24.1.2 Background

The vast majority of the tiles are from Bordesley itself, with a few from related sites: Bristol, Coventry, Worcester, Hailes, Halesowen, Maxstoke and Kenilworth. The tiles date from about 1200–1550 AD, the dating being based partly on the floor levels in which some of them were found. Before starting the scientific study, the tiles had been divided by one of the authors (JS) into 14 'production groups' (PGs). These are groups of tiles made using similar manufacturing techniques; in the case of the tiles excavated from Bordesley, the *in situ* tiles in a particular PG came from the same stratigraphic layer. A PG is interpreted as probably comprising the products of a single industry operating during a specific time phase, possibly using a single clay source or group of sources. The archaeological problems which prompted the chemical analysis relate to the organization of the manufacture of different production groups, and in particular the need to distinguish between those tiles made and used locally, those made centrally and distributed to various sites and those made by itinerant tilers moving from site to site.

The data consist of concentrations of the following elements: Hf, Cs, Rb, Sc, (Fe), Co, Eu, U, Ta, La, (Na), (K), Sm, (Ba), Yb, (As), (Sb), Lu, Th, Cr, Tb and Ce; those in brackets were measured but were not used in the cluster analysis. About 170 tile samples and 13 clay samples were measured. The data were log-transformed: both measurement errors and the natural scatter in data from the few definite single sources observed tends to be expressed as a percentage of the level, and the data are positive. It is worth pointing out that these data (while relating to composition in the chemical sense) are not compositional in the strict statistical sense in that they do not sum to a constant; while forming a part of a such a fixed-sum composition, the constraints on them are quite weak because, as trace elements, they form only a small part of the whole. Nevertheless, because of possible dilution effects, they share with true compositional data a common feature, namely that the prime interest is in ratios rather than in raw data. This explains why the techniques relevant for compositional data turn out to be appropriate for this type of data also, as will be shown below.

24.2 Data analysis

24.2.1 Visual aids

In the initial stages of the project it was required to identify both the overall structure of main groupings and also individual outliers among the tiles, as well as any specific features of interest among the variables. The latter point (often overlooked) is important from the point of view of archaeological and/or geological interpretation. An invaluable aid in this respect is the so-called *mountain plot*, or *variation diagram* as it is known in geology, based on the log-transformed values. This type of plot allows one to observe all the elemental concentrations both separately and simultaneously as shown in Fig. 1.1. A small number of elements have been omitted to avoid overcrowding the plot, and the scales for each element have been adjusted by removing a constant from each value; this ensures a sensible scale for all elements. The constant is the same for all items but differs from element to element, convenient values being the means over the complete data-set of the log-transformed values.

Fig. 1.1 immediately shows the high values for the heavy metals Cr, Fe and Co, and correspondingly lower values for the other elements, for three clays. These clays were from the vicinity of a 19th Century forge mill producing needles, and have been contaminated; they are therefore not typical of those used in the Middle
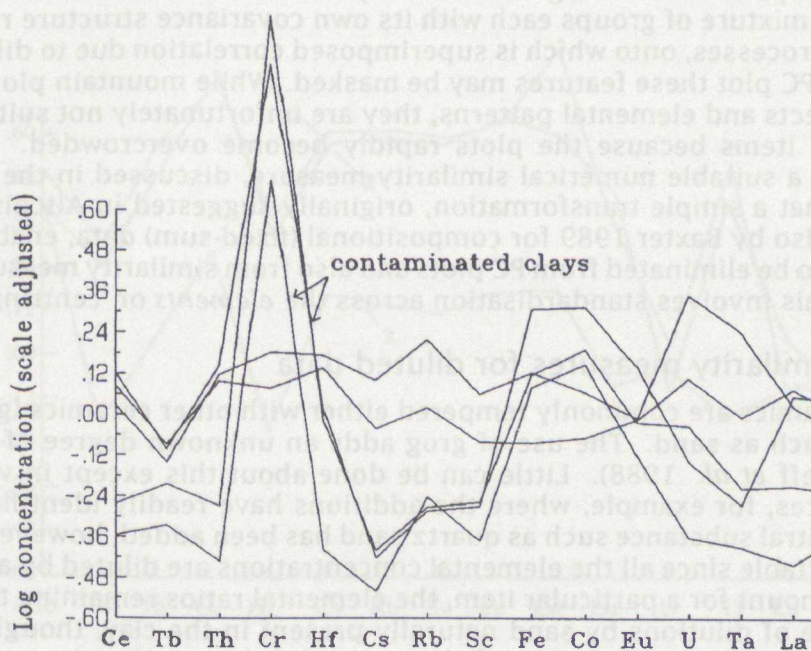
Figure 24.1: Mountain plot showing the effect of contamination on three clay samples. The vertical axis shows log-transformed concentrations, with scales adjusted by subtracting a constant (see text).

Ages. While this contamination effect is discernible from lists of the raw data and to some extent from principal components (PC) plots, it is not so obvious. Indeed, had only PCs based on the correlation matrix been plotted, these three contaminated clays would not have been unambiguously identified, because the three variable elements (Cr, Fe and Co) are given less weight in relation to the other less variable elements in such a plot. This example (and other similar ones encountered in this study) suggests that the almost universal practice of standardising to unit variance, which is what is involved in using the correlation matrix, may hide a real effect. Therefore no standardisation (other than the natural standardisation implicit in log-transformation) has been applied across the *tiles*.

A combination of mountain plots and PCs based on the covariance matrix seems to be a useful way of initially scanning the data. However, the validity of PCs on raw or even log-transformed data as an aid to identifying more detailed structure, for example groups of tiles having common clay sources, is doubtful. The data usually consist of a mixture of groups each with its own covariance structure resulting from geological processes, onto which is superimposed correlation due to dilution effects. In a simple PC plot these features may be masked. While mountain plots do indicate dilution effects and elemental patterns, they are unfortunately not suitable for large numbers of items because the plots rapidly become overcrowded. However, by considering a suitable numerical similarity measure, discussed in the next section, it is clear that a simple transformation, originally suggested in Aitchison 1986 and discussed also by Baxter 1989 for compositional (fixed-sum) data, enables the effect of dilution to be eliminated from PC plots and also from similarity measures in cluster analysis. This involves standardisation across the *elements* or 'centring').

24.2.2 Similarity measures for diluted data

Ancient ceramics are commonly tempered either with other ceramics (grog), or other materials such as sand. The use of grog adds an unknown degree of 'noise' to the data (see Neff *et al.* 1988). Little can be done about this except in very fortunate circumstances, for example, where the additions have readily identifiable features. Where a neutral substance such as quartz sand has been added, however, the effect is more predictable since all the elemental concentrations are diluted by approximately the same amount for a particular item, the elemental ratios remaining the same. The same is true of dilutions by sand naturally present in the clay, though in that case there may be some small degree of change in the elemental ratios as well because of differential settling of minerals in the original depositional environment of the sediment. Mountain plots of the Bordesley data indicate the possibility of natural or intentional sand tempering and therefore when seeking clusters having the same clay source, such dilution effects need to be eliminated from the measure of chemical similarity employed.

Mommsen *et al.* 1988 have suggested that instead of the commonly used Euclidean distance measure, chemical dissimilarity should be based on the spread (over the elements) of the factors required to multiply one composition to obtain the other. This can be visualised in Fig. 1.2: if one could apply exactly the same percentage increase or decrease to all elements to move one pattern on top of the other, in other words if the patterns were parallel, they would have the same clay composition apart from the fixed dilution factor. However, if the percentages varied from one element to another, there would be no common elemental pattern, and one would conclude that they had different basic clay compositions. In fact the upper two are almost identical in elemental ratios and also in absolute amounts; the very similar ratio pattern of the lower one suggests it is a diluted version of the upper two. The slight reverses in the elemental patterns for some of the elements at the RHS for this third sample suggest natural rather than intentional dilution.

Log concentration (scaled adjusted)

.60
.48
.36
.24
.12
.00
-.12
-.24
-.36
-.48
-.60

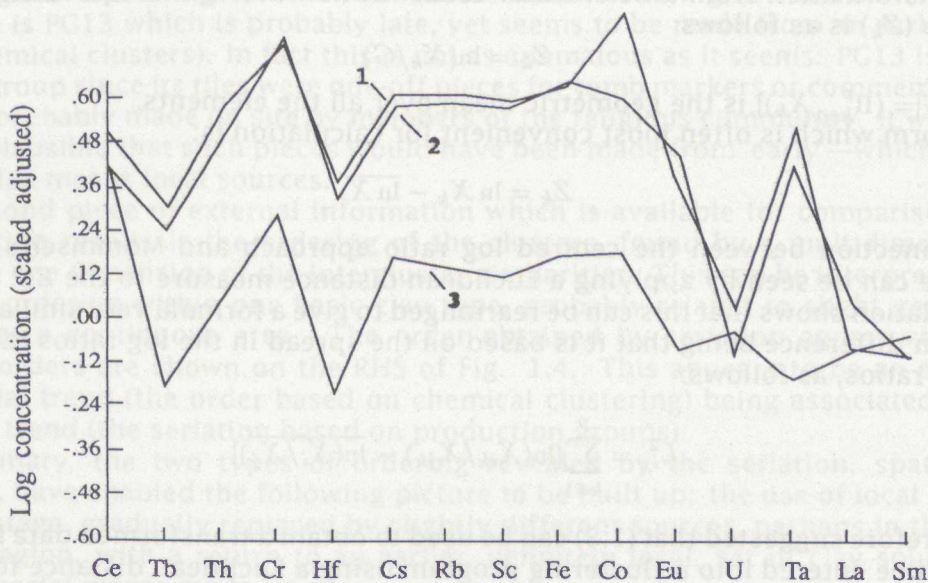Ce Tb Th Cr Hf Cs Rb Sc Fe Co Eu U Ta La Sm

Figure 24.2: Mountain plot showing two tile samples having almost identical compositions and a third having a similar elemental ratio pattern yet different absolute amounts, due to dilution.

The spread in the factors over the elements is given by the following:

$$D_{ij}^2 = \sum_{R=1}^{p} [X_{ik}/X_{ij} - \overline{X_i/X_j}],\qquad(24.1)$$

where X_{ik} is the value of the kth variable (elemental concentration) for the ith unit (tile) and the bar indicates the average over the p elements. While Mommsen *et al* have elaborated this in various ways, the various formulae they propose are all based on the fact that (1.1) reflects the ability of one pattern to fit on top of another by adjusting the general level.

Because this measure requires the calculation of a dissimilarity matrix as an initial step, and because there is no simply derived plot which can accompany the results of the calculations, it was decided to investigate the *centred log ratio* to ascertain if this would provide a more convenient way of taking account of dilution effects. The transformation from an elemental concentration (X) to the equivalent centred log ratio (Z_k) is as follows:

$$Z_k = \ln(X_k/G)\qquad(24.2)$$

where $G[= (\Pi_{k=1}^{p} X_k)]$ is the geometric mean over all the elements.

The form which is often most convenient for calculation is:

$$Z_k = \ln X_k - \overline{\ln X}.\qquad(24.3)$$

The connection between the centred log ratio approach and Mommsen's distance measure can be seen by applying a Euclidean distance measure to the Zs; algebraic manipulation shows that this can be rearranged to give a formula very similar to (1.1), the main difference being that it is based on the spread in the log ratios rather than the raw ratios, as follows:

$$L_{ij}^2 = \sum_{k=1}^{p} [\ln(X_{ik}/X_{jk}) - \overline{\ln(X_i/X_j)}]\qquad(24.4)$$

It is therefore suggested that (1.3) can be used to obtain a transformed data file which can then be entered into a clustering program using a Euclidean distance formula in the standard way. This process has the advantage that one can work on the data file directly, perform the centring, and only at the final stage enter the data to a clustering program; the intermediate step of computing a special distance matrix is not necessary. Furthermore, having determined the transformation from X to Z which will lead to a sensible similarity measure, it is a logical step to display PCs based on the Z values since the resulting plots are are unaffected by dilution effects whose presence might be hidden in a plot based on the raw data.

24.2.3 Interpretation of results

Fig. 1.3 shows a PC plot based on the centred log ratios (the Zs) of all the individual tiles and clays, with an indication of the relative spreads of each chemical cluster given by the sizes of the 68% concentration ellipses. The main structure is very easy to discern on this plot. For example, clusters 4, 10 and 11, which contain mainly imports and tiles found at the related sites, are distinct. So also are clusters 2 and 9, containing tiles from St Stephen's church in the Abbey grounds, and late tiles of production groups 1, 9, and 11 from the Abbey itself. Clusters 16 and 17 contain mainly clay samples. The details of the clusters will be the subject of further archaeological interpretation at a later stage. However in the present context it is appropriate to concentrate on the eight 'main Bordesley group' of overlapping

clusters in the centre, since this is where further analysis of the data can be of most benefit.

It was decided to investigate seriation as a way of arranging the the production groups and chemical clusters and their inter-relationships. While this is not a classical seriation application, as described for instance by Doran & Hodson 1975, the approach nevertheless makes sense if it can be assumed that the manufacturers of each production group used a set of several different clays whose constituents were gradually phased in and out over time. To perform the seriation, each tile in the main Bordesley group was classified in a two-way table according to its PG, and its cluster based on chemical composition. The seriation algorithm produced a rearranged the table as shown in Fig. 1.4. While seriation is by no means perfect, the resulting orders did fit in very well with two independent pieces of information.

The first piece of information is the basic classification of the PGs into early and late (based on appearance and stratigraphy), as shown along the top line of Fig. 1.4. There is good overall agreement with the seriation ordering of the PGs. The one exception is PG13 which is probably late, yet seems to be made from an early set of clays (chemical clusters). In fact this is not as anomalous as it seems: PG13 is a very unusual group since its tiles were one-off pieces for tomb markers or commemorative plaques, probably made on site by members of the religious community. It would be entirely plausible that such pieces would have been made from 'early'—which in this context also means local sources.

The second piece of external information which is available for comparison with the seriation results is the ordering of the clusters, found by a multidimensional scaling to one dimension of the intercluster similarities. This can be interpreted as a chemical ordering within one basic clay type, probably related to slight geological shifts over a continuous area. The order obtained by seriation agrees very well (the two orders are shown on the RHS of Fig. 1.4. This appears to be an example of a spatial trend (the order based on chemical clustering) being associated with a temporal trend (the seriation based on production groups).

In summary, the two types of ordering revealed by the seriation, spatial and temporal, have enabled the following picture to be built up: the use of local clays in an early stage, gradually replaced by slightly different sources, perhaps in the same general region, with a return to an earlier, definitely local, set of clay sources for certain special pieces made on site.

24.3   Conclusions

A reliable set of techniques for displaying and clustering trace element data from ceramics subject to simple dilution effects are mountain plots in the initial stages and then the centred log ratio transformation, with the resulting data used in PC plots and in a Euclidean distance measure in cluster analysis. Seriation has been a useful way of summarising the resulting relationships between production groups and chemical clusters and has enabled external information, in the form of independently derived orderings, to be compared with the cluster results. This has provided validation for, and assists in, the interpretation of the clusters.
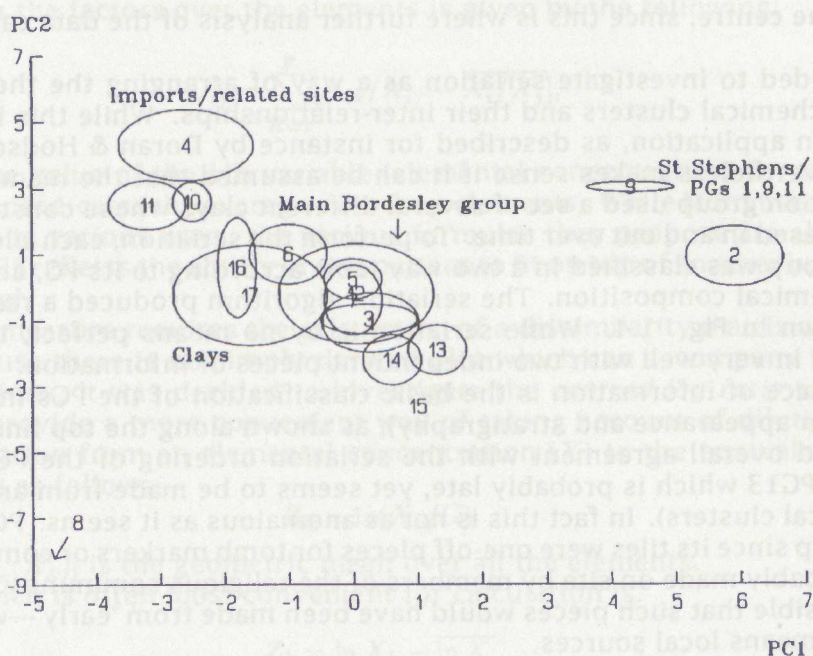
Acknowledgements

Figure 24.3: Principal components plot based on the centred log ratio data for the Bordesley tiles; 68% concentration ellipses for the clusters are shown, along with the cluster identification numbers (8 and 15 are singleton clusters). Four main subgroups are indicated. Production groups ocurring in the central ('main Bordesley group') have been seriated (see Fig. 1.4).
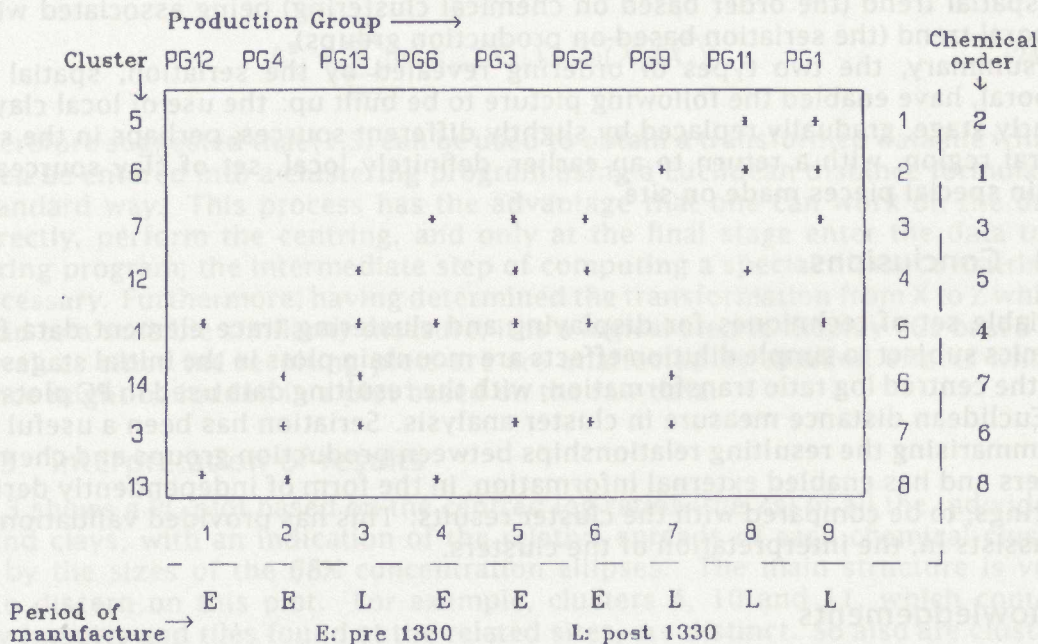

Figure 24.4: The result of seriating PGs represented among the main Bordesley group of clusters (see Fig. 1.3). Percentages were used but presence/absence only is indicated. Compare the external estimate of the date (L:Late, E:early) along the bottom with the order implied by the seriation (1–9). A 'chemical' order of clusters obtained from the intercluster similarities (interpreted as a spatial trend) is shown on the outer right, in bold; cf the order implied by the seriation (1–8, inner right).

Bibliography

AITCHISON, J. A. 1986. *The Statistical Analysis of Compositional Data.* Chapman and Hall, London.

BAXTER, M. J. 1989. "Multivariate analysis of data on glass compositions: a methodological note", *Archaeometry*, 31 (1): 45–53.

DORAN, J. E. & F. R. HODSON 1975. *Mathematics and Computers in Archaeology.* Edinburgh University Press, Edinburgh.

HIRST, S. M., D. A. WALSH, & S. M. WRIGHT 1983. *Bordesley Abbey II.* British Series 111. British Archaeological Reports.

LEESE, M. N., M. J. HUGHES, & J. CHERRY 1986. "A scientific study of N. Midlands medieval tile production", *Oxford Journal of Archaeology*, 5 (3): 355–370.

MOMMSEN, H., A. KREUSER, & J WEBER 1988. "A method for grouping pottery by chemical composition", *Archaeometry*, 30 (1): 47–57.

NEFF, H., R. L. BISHOP, & D. E. ARNOLD 1988. "Reconstructing ceramic production from ceramic compositional data: an example from Guatemala", *Journal of Field Archaeology*, 15: 339–348.