

On establishing ceramic chemical groups: exploring the influence of data analysis methods and the role of the elements chosen in analysis

Kostalena Michelaki,¹ Michael J. Hughes,² Ronald G.V. Hancock³

¹School of Human Evolution and Social Change, Arizona State University, Arizona, USA; ²School of Health and Bioscience, University of East London, Water Lane, London, UK; ³Department of Medical Physics and Applied Radiation Sciences and Department of Anthropology, McMaster University, Hamilton, Ontario, Canada

Abstract

Since the 1970s, archaeologists have increasingly depended on archaeometric rather than strictly stylistic data to explore questions of ceramic provenance and technology, and, by extension, trade, exchange, social networks and even identity. It is accepted as obvious by some archaeometrists and statisticians that the results of the analyses of compositional data may be dependent on the format of the data used, on the data exploration method employed and, in the case of multivariate analyses, even on the number of elements considered. However, this is rarely articulated clearly in publications, making it less obvious to archaeologists. In this short paper, we re-examine compositional data from a collection of bricks, tiles and ceramics from Hill Hall, near Epping in Essex, England, as a case study to show how the method of data exploration used and the number of elements considered in multivariate analyses of compositional data can affect the sorting of ceramic samples into chemical groups. We compare bivariate data splitting (BDS) with principal component analysis (PCA) and centered log ratio-principal component analysis (CLR-PCA) of different unstandardized data formats [original concentration data and logarithmically transformed (*i.e.* \log_{10} data)], using different numbers of elements. We confirm that PCA, in its various forms, is quite sensitive to the numbers and types of elements used in data analysis.

Introduction

Archaeologists have long acknowledged the importance of understanding the raw materi-

als that constitute the ceramic artifacts they find in the archaeological record. Knowing whether such materials were, or were not, local to the site where ceramics were found allows them to address questions of regional interactions. Knowing whether raw materials were mixed in different ways allows them to pursue questions related to technological traditions, communities of practice and, ultimately, of identity. Thus, archaeologists are consistently interested in the chemical groups that compositional analyses of ceramics suggest. Yet, it is not always clear that the definition of such chemical groups may be dependent on the way compositional data are transformed (Michelaki and Hancock, 2011), on the particular data exploration method selected (Baxter and Freestone, 2006) and even on the specific elements selected for consideration (Baxter and Jackson, 2001). Although some archaeometrists and statisticians consider such dependence obvious, the issue is rarely addressed clearly in the literature.

The purpose of our paper is to evaluate the influence that different data exploration methods [bivariate data splitting (BDS), principal component analysis (PCA) and centred log ratio-principal component analysis (CLR-PCA)] and varying numbers of chemical elements ($n=29, 25, 17, 10, 6$ and 4) will have on the same set of ceramic chemical data. As a case study, we have chosen 79 samples of tiles and bricks from Hill Hall, near Epping in Essex, England, that one of us (M.J.H.) analyzed previously, using inductively coupled plasma-atomic emission spectrometry (ICP-AES) (Hughes, 2009).

Materials and Methods

In a recent paper, Hughes (2009) published the results of his chemical analysis of 79 samples of ceramic building materials from Hill Hall, a house built originally by Sir Thomas Smith during the reign of Elizabeth I. Specifically, he examined 7 types of medieval brick (TB1, MB1-9, MB10, Med B1, Med B2, Med B3 and Med B4), one type of medieval roof tile (RT4) and one type of medieval floor tile (FT6). Moreover, he examined a set of post-medieval materials: three types of floor tiles (FT10-13, FT1 and FT2), one type of roof tile (RT3), one type of stove tile (glazed terracotta) and one type of terracotta, along with brick samples from the Chipping Ongar church.

The aim was to understand from where the bricks and tiles used in the construction of Hill Hall came. To achieve this goal, bivariate scatter plots of chemical elements were used to confirm that there were no sample outliers and that the data formed an analytically coherent set. As expected, some strong element correla-

Correspondence: Kostalena Michelaki, School of Human Evolution and Social Change, Arizona State University, 900 S. Cady Mall, Tempe, Arizona, 85287-2402, USA.
Tel. +1.480.965.6170 - Fax: +1.480.965.7671
E-mail: kmichela@asu.edu

Citation: Michelaki M, Hughes M, Hancock RG, 2013. On establishing ceramic chemical groups: exploring the influence of data analysis methods and the role of the elements chosen in analysis. In: RH Tykot (ed.) Proceedings of the 38th International Symposium on Archeometry – May 10th-14th 2010, Tampa, Florida. Open Journal of Archaeometry 1:e1.

Key words: ceramic compositional data, principal component analysis, bivariate data splitting, medieval and post-medieval England.

Acknowledgments: this paper owes a lot to the Fitch Laboratory at the British School at Athens and the School of Human Evolution and Social Change at Arizona State University for the support they provided to KM. Gary Schwartz read multiple drafts and helped us better shape and articulate our argument. The paper would not exist at all without RGVH's ideas and MH's original work.

Contributions: the authors contributed equally.

Conflict of interests: the authors declare no potential conflict of interests.

Presented at the 38th International Symposium on Archaeometry – May 10th-14th 2010, Tampa, Florida.

This work is licensed under a Creative Commons Attribution 3.0 License (by-nc 3.0).

©Copyright K. Michelaki et al., 2013

Licensee PAGEPress, Italy

Open Journal of Archaeometry 2013; 1:e1

doi:10.4081/arc.2013.e1

tions were found, as is common in the analysis of ceramics. For example, strong correlations were found between lanthanum (La), cerium (Ce) and neodymium (Nd) [rare earth elements (REEs)] and aluminium (Al) and titanium (Ti) (major elements of clay minerals). To explore the structure of the chemical data, while also correcting for temper dilution, the data were then submitted to a CLR-PCA (Aitchison, 1986; Aitchison *et al.*, 2002; Leese *et al.*, 1989; Pollard *et al.*, 2006). Seventeen elements that past experience had proven to be reliably diagnostic were included in the analysis. The results indicated the presence of two distinct chemical groups, separating 11 samples of brick and tile (*small* group) from 68 other ceramic building materials samples (*large* group). To check whether this separation of the 11 samples from the remaining 68 would persist despite variation in data explo-

ration methods, we initially assessed the data (in original concentration format) using BDS. Then we submitted the data to PCA (using both original concentration data and logarithmically transformed data) and to CLR-PCA, using progressively fewer elements, by eliminating elements of similar chemistry so that each periodic group was represented by increasingly fewer elements.

Results

Bivariate data splitting

Before the examination of bivariate plots commenced, the means, standard deviations, and coefficients of variation (CVs) of all measured element concentrations were calculated from each element. This provides a straightforward way to establish the relative heterogeneity of the element distributions.

The CVs varied quite radically:

<20% Al, Cr, K, La, Ti

20% - 25% Fe, Na, Sc, Zn, Zr, Ce, Nd

25% - 30% Mn, Ni, Li, V

30% - 40% Mg, Ni, Y, Sm, Eu, Dy, Yb

>40% P, Ba, Ca, Co, Sr, Pb

These broad distributions suggested that the samples consisted of relatively coarse wares, some with Ca-rich tempering, or that the different building material types may have been made in different places. To add to these possibilities, the use of sediments from chemically different local clay pits, or even from chemically different sediment horizons, at the same place of production, might also have contributed to an increase in the CVs. No matter what the reason, it was clear from the examination of concentrations of individual elements that sufficient variation in the data existed to identify distinct chemical groups.

Next, we addressed whether data exploration methods different from those originally used by Hughes (2009) would produce the same data sorting. First, a BDS approach was applied. This is often considered a time consuming process and is rarely favoured in the presentation of the results of compositional data analysis. Yet, bivariate plots illustrate the basic geochemical relationships within the data set and can, thus, provide chemical clarity. Calcium (Ca) and strontium (Sr) are positively correlated (Figure 1a), implying that the source of lime in selected samples, especially in the tin-glazed floor tiles (FT2), was from limestone associated with sea shells. Lead (Pb) did not correlate well with other elements, and highly elevated Pb levels were found only in glazed floor tiles. Barium (Ba) also did not correlate well with other elements, with elevated levels appearing in different sorts of unglazed floor tiles. Since zirconium

(Zr) was positively correlated with other silicate forming elements, the diluant of the ceramics (apart from limestone) is probably not zircon-rich quartz, but must be either silica-rich silicate mineral(s) or pure quartz. However, in the preparation of samples for ICP-AES an acid dissolution (hydrofluoric/perchloric acids) procedure was followed, which may not have completely dissolved all zirconia present. Thus, any interpretation of Zr data should be regarded with some caution.

All other reported elements correlated positively, especially the REEs (e.g. Figure 1b), as previously reported (Hughes, 2009), but in some cases the correlations were *lumpy*, especially in the K₂O and Na₂O plots with other elements (e.g. Figure 1c,d). This finding allowed the data set to be split into three groups: Group 1 (30 samples; Table 1) with higher elemental concentrations; Group 2 (29 samples; Table 1)

with lower elemental concentrations; and Group 3 (20 samples; Table 1) with samples that had relatively higher K₂O and Na₂O contents than was deemed to be *normal* compared to their closer chemical neighbour, Group 2. From Figure 1c,d, it appears that Group 1 and Group 2 might be two distinct groups, or may be the high and low ends of a single, highly variable chemical group. This is why they are marked in all figures as different shades of the same symbol (x). For the 3 possible chemical groups, means and standard deviations are presented in Table 1. In this table, the elements are ordered alphabetically by major and minor element oxide, and alphabetically by trace element, with the REEs in order of atomic number. Also, the elements with the most diverse scatter are listed at the bottom of the table, since these were not included in the calculations of inter-group ratios (see below).

Table 1. Summary statistics for the three chemical groups, including group ratios performed on group means.

Elements	G1 30 samples	G2 29 samples	G3 20 samples	G1/G2	G1/G3	G2/G3
Al ₂ O ₃ %	16.2±1.4	12.5±1.8	11.7±1.6	1.3	1.4	1.1
Fe ₂ O ₃ %	7.7±0.6	5.9±1.4	5.8±0.9	1.3	1.3	1.0
MgO%	2.1±0.3	1.6±0.5	0.9±0.2	1.3	2.2	1.7
Na ₂ O%	0.47±0.05	0.36±0.09	0.55±0.07	1.3	0.9	0.7
K ₂ O%	3.38±0.22	2.43±0.50	2.78±0.30	1.4	1.2	0.9
TiO ₂ %	0.95±0.09	0.70±0.11	0.85±0.10	1.3	1.1	0.8
Co	35±14	24±21	16±4	1.4	2.2	1.6
Cr	138±14	104±17	102±19	1.3	1.4	1.0
Cu	29±5	28±8	19±4	1.0	1.5	1.5
Li	82±14	61±12	54±18	1.3	1.5	1.1
Ni	63±7	46±10	26±5	1.4	2.4	1.8
Sc	18±2	14±2	12±2	1.3	1.6	1.2
V	156±18	115±32	100±21	1.4	1.6	1.1
Y	32±5	24±5	15±4	1.3	2.1	1.6
Zn	93±8	76±19	64±8	1.2	1.4	1.2
Zr	100±20	81±12	64±10	1.2	1.6	1.3
La	46±5	35±5	33±4	1.3	1.4	1.1
Ce	96±14	69±10	70±8	1.4	1.4	1.0
Nd	49±5	37±5	33±4	1.3	1.5	1.1
Sm	8.2±1.6	5.3±1.2	4.8±1.0	1.6	1.7	1.1
Eu	1.80±0.29	1.22±0.23	0.95±0.20	1.5	1.9	1.3
Dy	5.4±0.9	3.9±0.8	2.6±0.7	1.4	2.1	1.5
Yb	2.8±0.4	2.0±0.4	1.3±0.3	1.4	2.1	1.5
		Mean±SD		1.3±0.1	1.6±0.4	1.2±0.3
CaO%	0.69±0.24	6.00±8.82	0.55±0.74	0.1	1.3	10.9
MnO%	0.038±0.008	0.032±0.009	0.039±0.013	1.2	1.0	0.8
P ₂ O ₅ %	0.11±0.02	0.13±0.08	0.09±0.03	0.8	1.3	1.5
Ba	540±290	330±60	440±30	1.6	1.2	0.8
Pb	68±76	182±240	23±10	0.4	2.9	7.8
Sr	110±10	170±130	93±10	0.6	1.2	1.9

G1, group 1; G2, group 2; G3, group 3; SD, standard deviation.

Table 2 shows that Group 3 samples include all the samples from the *small* group of Hughes (2009) plus four samples from MB1-9; three samples from TB1; and one sample from MB3.

Even though bivariate plots (see K_2O vs MgO ; Figure 1d) suggest separations among three, or perhaps between two, chemical groups, the final step in the BDS approach was to identify the elements that showed the most pronounced differences [*e.g.* nickel (Ni), lithium (Li), potassium oxide (K_2O) and sodium oxide (Na_2O)] and to plot their inter-elemental ratios (*e.g.* Figure 1e,f). These showed clearly that there were only two distinct, non-overlapping chemical groups: 59 samples (30 plus 29) in the first group, and 20 samples in the second. By uniting the original first and second groups of samples, these figures also confirmed that the two groups were part of an associated continuum of data. This is confirmed by the group ratios presented at the right hand side of Table 1, after the summary

elemental concentration data for the three groups, where mean concentration ratios are much closer between Groups 1 and 2 (1.3 ± 0.1) than between Groups 1 and 3 (1.6 ± 0.4) and between Groups 2 and 3 (1.2 ± 0.3).

Principal component analysis with varying numbers of elements

To explore why Hughes (2009) found 11 distinctive samples while we had now found 20 using BDS, PCA was performed on both the original concentration data and on logarithmic data, using progressively fewer chemical elements (see bottom of Table 3 for the elements included in each PCA). The rationale was to progressively eliminate elements of similar chemistry, so that each periodic group was rep-

resented by increasingly fewer elements.

Although they are somewhat subjective, the resulting PC plots clearly show a progression that is element-choice dependent (Table 3), and hence may explain the apparent discrepancy with the earlier analysis by Hughes (2009).

Figure 2a (29 elements; PC1 vs PC2) sorts the post-medieval polychrome tin-glazed floor tiles (FT2) from the other samples, mainly on the basis of their high CaO content. In the original publication (Hughes, 2009), based on comparative analysis, it was concluded that these samples were products of the Low Countries and unlike any other ceramic material analyzed from Hill Hall. Moreover, in the bottom left hand corner of the plot lie many of the samples that we defined as Group 3, based on our BDS analysis. Without that *a priori* classification of these samples as Group 3 in the PCA, they would not appear as a clearly distinguishable cluster of samples, but rather as

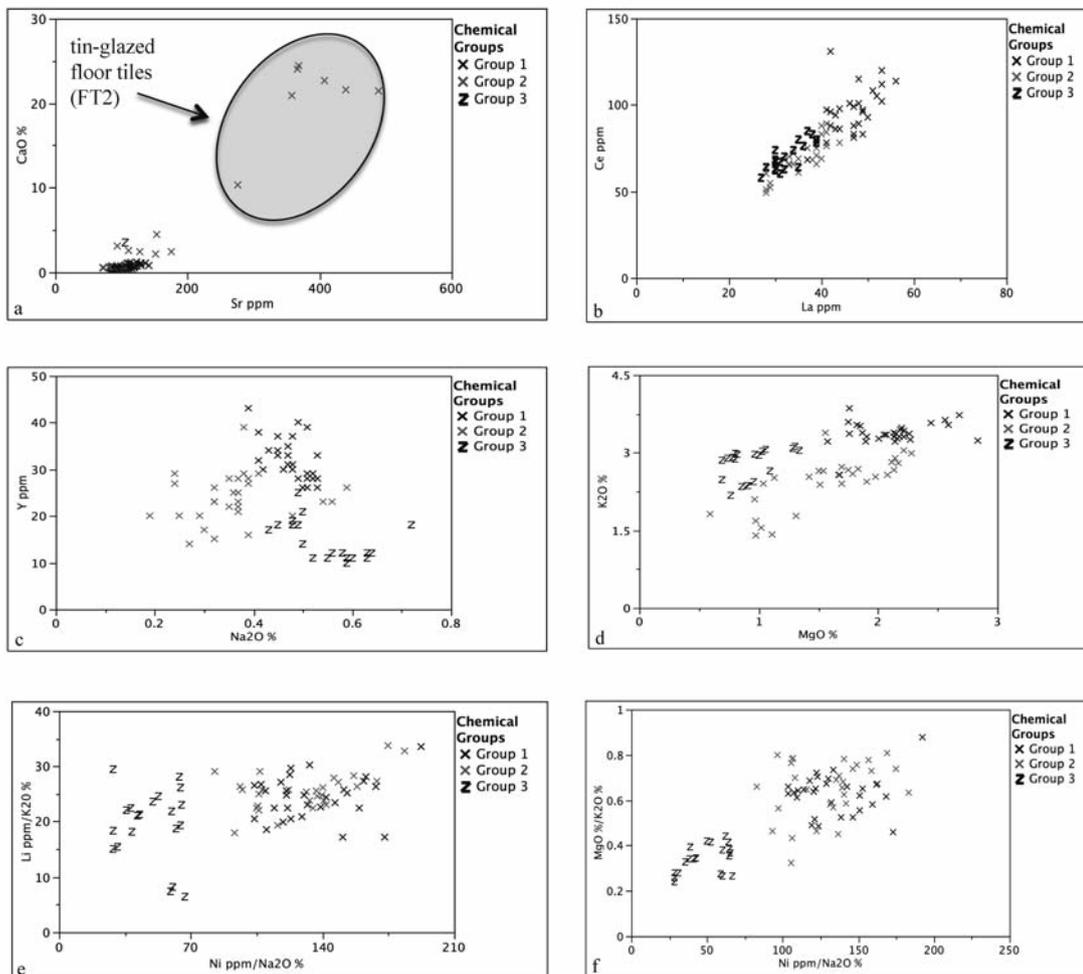


Figure 1. Scattergram of a) Sr ppm vs $CaO\%$; b) La ppm vs Ce ppm; c) $Na_2O\%$ vs Y ppm; d) $MgO\%$ vs $K_2O\%$; e) Ni ppm/ $Na_2O\%$ vs Li ppm/ $K_2O\%$; and f) Ni ppm/ $Na_2O\%$ vs $MgO\%/K_2O\%$.

lying along one end of the spectrum of variation in PC1 scores. Figure 2b (10 elements, PC1 vs PC2) has twelve, or more, samples from Group 3 in the upper left quadrant that are clearly separated from the main group. It is only when one whittles down to less than ten elements (e.g. 4 elements, PC1 vs PC2; Figure 2f) that the 20 Group 3 samples are clearly separated from the rest.

first suggested the presence of a distinct small group with as many as 20 sample members. The consequences of such group membership variability can be of great importance to archaeologists, as they try to assess how often

certain raw materials sources were used, how common *foreign* materials were, or how widely used certain ceramic paste recipes were, for example. In our specific case study, the enlargement of the number of samples with

Centered log ratio-principal component analysis

To confirm the analyses of Hughes (2009), centered log ratio-principal component analysis (CLR-PCA) was employed (Aitchison, 1986; Aitchison *et al.*, 2002; Leese *et al.*, 1989; Pollard *et al.*, 2006). In CLR, each elemental concentration was first converted to the natural logarithm of the ratio of the specific elemental concentration to the geometric mean of all elemental concentrations in each sample. The new data set was then submitted to a PCA. In all the cases we considered enough principal components to achieve a cumulative variance of $\geq 90\%$. The summary results are presented in Table 3. Figure 2d shows that when all 29 elements are used in CLR-PCA, eleven Group 3 samples at the bottom right quadrant of the graph clearly separate from the rest of the samples, as do also the tin-glazed floor tiles (FT2). Figure 2e shows that 11 Group 3 samples in the lower right quadrant are clearly separated, and if one knew that the bulk of the other samples were tightly clustered on the centre-left, then more Group 3 samples are obvious in the upper right of the plot. As with PCA, it is only when the numbers of included elements are well below ten that the group separation is clear, as is displayed in Figure 2f, the 4 element CLR-PCA plot.

Discussion

In Hughes' original publication (2009) of the chemical composition of 79 brick and tile samples from Hill Hall, two distinct groups were defined: a large group with 68 samples and a small group with 11 samples. In this study, we found that membership to these groups varied based on the data exploration method used and on the number of chemical elements included in the analysis. Specifically, in our study, the membership of the small group varied from 11 samples, when CLR-PCA was used with 29 and 25 elements, to as many as 20 samples when PCA or CLR-PCA were used with 6 or 4 elements. Moreover, it was the careful examination of bivariate plots of chemical elements and inter-elemental ratios that

Table 2. Groupings found by this work and those found in Hughes (2009) (large and small).

Samples	This work - BDS			Hughes (2009) - PCA	
	G1	G2	G3	Large	Small
FT6	4	-	1	4	1
Med B1	-	6	-	6	-
Med B2	-	6	-	6	-
MB1-9	1	-	4	5	-
TB-1	-	-	9	3	6
RT4	6	-	-	6	-
FT10-13	6	-	-	6	-
RT3	3	-	-	3	-
Ongar church	-	6	-	6	-
Terracotta	5	1	-	6	-
Glazed terracotta	5	-	-	5	-
FT1	-	3	-	3	-
FT2	-	7	-	7	-
Med B3	-	-	1	1	-
Med B4	-	-	1	-	1
MB10	-	-	4	1	3
Total	30	29	20	68	11

BDS, bivariate data splitting; PCA, principal component analysis; G1, group 1; G2, group 2; G3, group 3; FT6, type of medieval floor tile; Med B1, Med B2, MB1-9, and TB-1, types of medieval brick; RT4, type of medieval roof tile; FT10-13, FT1 and FT2, types of floor tiles; RT3, type of roof tile; Med B3, Med B4, and MB10, types of medieval brick.

Table 3. Principal component analysis (original data or log10 data) and centered log ratio-principal component analysis (original data) sample assignments.

	Ceramic materials							Total	BestPC plot
	FT6	Med B3	Med B4	MB1-9	MB10	TB1	Med B2		
PCA									
29 elements*	1	0	1	2	4	8	0	≤ 16	PC1-PC2
25 elements ^o	1	0	1	2	4	8	0	16	PC1-PC2
17 elements [†]	1	0	1	1	4	5	0	12	PC1-PC2
10 elements [§]	1	0	1	4	4	9	0	19	PC1-PC2
6 elements [^]	1	1	1	4	4	9	0	20	PC1-PC2
4 elements [§]	1	1	1	4	4	9	0	20	PC1-PC2
CLR-PCA									
29 elements*	1	0	0	1	4	5	0	11	PC1-PC2
25 elements ^o	1	0	0	1	4	5	0	11	PC1-PC2
17 elements [†]	1	0	0	1	4	8	0	14	PC1-PC2
10 elements [§]	1	0	0	1	4	5	0	11	PC1-PC2
10 elements [§]	1	0	1	1	4	8	0	15	PC1-PC3
6 elements [^]	1	1	1	4	4	9	0	20	PC1-PC2
4 elements [§]	1	1	1	4	4	9	0	20	PC1-PC2

FT6, type of medieval floor tile; Med B3, Med B4, MB1-9, MB10, TB1, and Med B2, types of medieval brick; PC, principal component; PCA, principal component analysis; PC1, principal component one; PC2, principal component two; CLR-PCA, centered log ratio-principal component analysis; PC3, principal component three. *Al, Fe, Mg, Na, K, Ti, Co, Cr, Cu, Li, Ni, Sc, V, Y, Zn, Zr, La, Ce, Nd, Sm, Eu, Dy, Yb, Ca, Mn, P, Ba, Pb, Sr; ^oAl, Fe, Mg, Na, K, Ti, Co, Cr, Cu, Li, Ni, Sc, V, Y, Zn, Zr, La, Ce, Nd, Sm, Eu, Dy, Yb, Mn, P; [†]Al, Fe, Na, K, Ti, Mg, Co, Cr, Li, Ni, Sc, V, Y, Zn, La, Ce, Nd; [§]Al, Mg, Na, K, Ti, Ni, Li, Cu, Y, La, Ce; [^]Mg, Na, K, Ni, Li, Y; [§]Mg, Na, K, Ni.

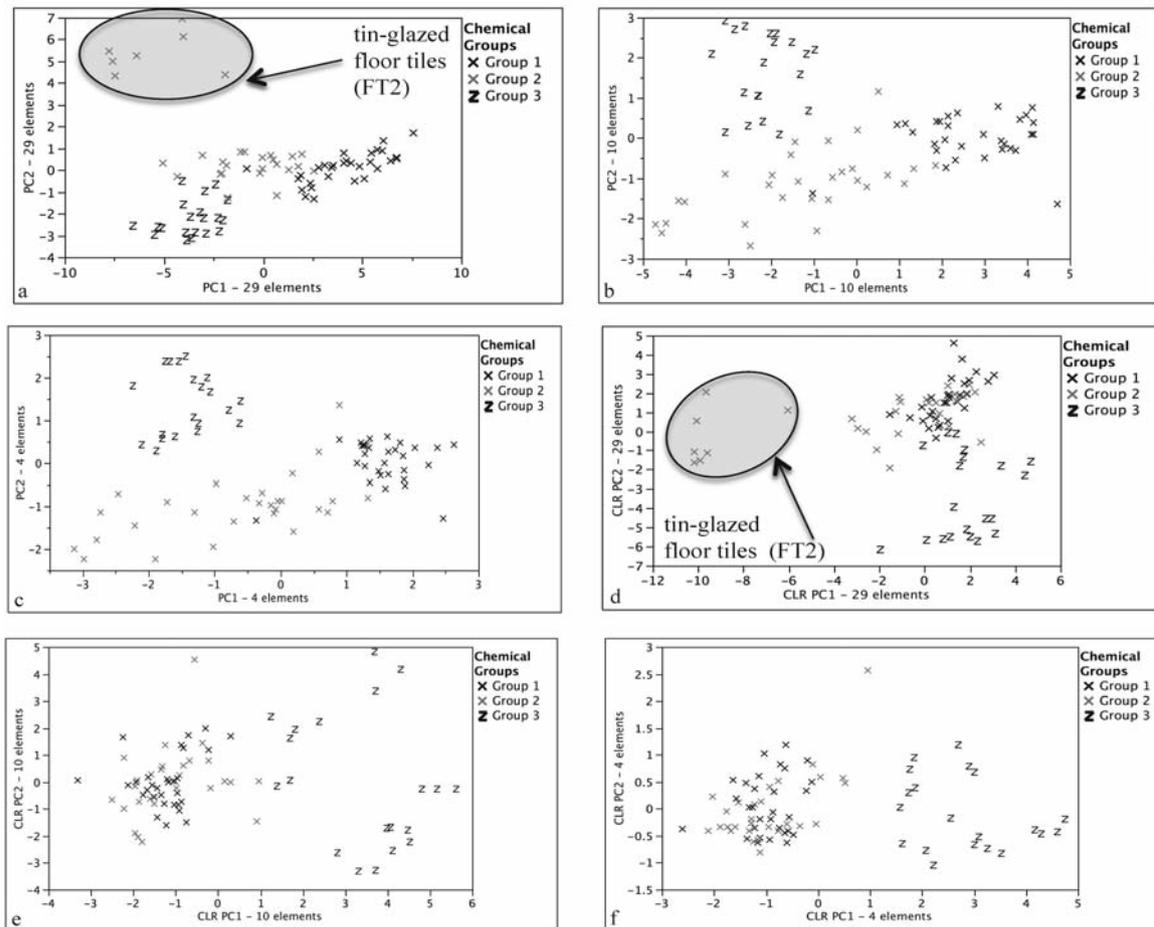


Figure 2. a) PC1 vs PC2 using 29 elements; b) PC1 vs PC2 using 10 elements; c) PC1 vs PC2 using 4 elements; d) CLR-PC1 vs CLR-PC2 using 29 elements; e) CLR-PC1 vs CLR-PC2 using 10 elements; and f) CLR-PC1 vs CLR-PC2 using 4 elements. 29 elements=Al, Fe, Mg, Na, K, Ti, Co, Cr, Cu, Li, Ni, Sc, V, Y, Zn, Zr, La, Ce, Nd, Sm, Eu, Dy, Yb, Ca, Mn, P, Ba, Pb, Sr; 10 elements=Al, Mg, Na, K, Ti, Ni, Li, Y, La, Ce; 4 elements=Mg, Na, K, Ni.

Group 3 chemistry (*i.e.* 20 rather than 11 samples) had a most interesting effect: now many medieval bricks (all of TB1 and most of MB1-9) and all of the post-medieval bricks (MB10, Med B3 and Med B4) are displaying the same Group 3 chemistry. One medieval floor tile (FT6) sample continues to be part of Group 3, as initially suggested by Hughes (2009).

Conclusions

Overall, in this study, the order of analysis implemented to render clarity in the chemical sorting of the samples analyzed was: inter-element concentration ratio plots; PCA; CLR-PCA; and bivariate element concentration plots. The finding that the method of data exploration and the number of elements included in analysis can influence chemical group membership and, thus, have an impact on archaeological

interpretations suggests that we must apply as much attention to assessing and sorting analytical data as we do to creating data sets.

References

- Aitchison J, 1986. The statistical analysis of compositional data. Chapman and Hall Ltd., London, UK.
- Aitchison J, Barceló-Vidal C, Pawlowsky-Glahn V, 2002. Some comments on compositional data analysis in archaeometry, in particular the fallacies in Tangri and Wright's dismissal of logratio analysis. *Archaeometry* 44:296-304.
- Baxter MJ, Freestone IC, 2006. Log-ratio compositional analysis in archaeometry. *Archaeometry* 48:511-31.
- Baxter MJ, Jackson CM, 2001. Variable selection in artifact compositional studies. *Archaeometry* 43:253-68.
- Hughes MI, 2009. ICP-AES analysis of ceramic building material. In: P. Drury, R. Simpson (eds.) *Hill Hall: a singular house devised by a Tudor intellectual*. The Society of Antiquaries of London, London, UK, pp 366-79.
- Leese MN, Hughes M, Stopford J, 1989. The chemical composition of tiles from Bordesley: a case study in data treatment. In: S. Rahtz, J. Richards (eds.) *Computer applications and quantitative methods in archaeology*. British Archaeological Reports, Oxford, UK, pp 241-9.
- Michelaki K, Hancock RGV, 2011. Chemistry versus data dispersion: Is there a better way to assess and interpret archaeological data? *Archaeometry* 53:1259-79.
- Pollard A, Blockley SPE, Lane CS, 2006. Some numerical considerations in the geochemical analysis of distal microtephra. *Appl Geochem* 21:1692-714.