Hans-Joachim Mucha – Hans-Georg Bartel – Jens Dolata

Finding Roman Brickyards in Germania Superior by Model-Based Cluster Analysis of Archaeometric Data

Abstract: Chemical analysis of ancient ceramics and of other archaeologically important materials has been used frequently to support archaeological research. Often the dimensionality of the measurements has been high. Therefore, multivariate statistical techniques such as cluster analysis have to be applied. The aim of the present paper is to give a review of the research on bricks and tiles from Roman military brickyards in Germania Superior and to present the main results obtained by multivariate statistical analysis. In particular, new adaptive cluster analysis methods and modified model-based clustering are applied on archaeometric data (Mucha / Bartel / Dolata 2002; 2003a; 2005b; in press; Bartel / Dolata / Mucha 2000; 2003). The main result was the discovery of military brickyards that were not known when the project began about ten years ago. Recently, they have been discovered by the application of these multivariate statistical analysis models. Newly developed visualization methods support and facilitate the interpretation of both the data set and the results of grouping. This means archaeologists can easily identify a new finding of a Roman brick or tile by comparing its chemical fingerprint with those from the detected provenances.

Introduction

About 1000 Roman stamped bricks and tiles from the Upper Rhine area were the objects under investigation by methods of chemical and statistical analysis. Using both archaeological information and the results of mineralogy and chemistry allowed archaeologists to develop a complex model of brick- and tile-making in the Roman period. In south-west Germany a few large brickyards existed, of which the operating authority was the Roman army. The period of operation was from the middle of the first century AD until the end of the fourth century. Archaeologists are most interested in the location of brickyards and in the chronology of the production-marks, which are found on the building-material.

At present we have a long history of chemical and statistical analysis of Roman brick and tile making in Germania Superior (Dolata 1996; 1998; 1999; 2000; 2001; Dolata / Werr 1999; Werr 1998;

Fig. 1. (left) Staatliche Museen Berlin Preußischer Kulturbesitz (Antikensammlung) TC 5778a: later (brick) with stamp of legio XXII, findspot: Neuwied-Niederbieber, provenance: 'not yet known 1'. (right) Landesmuseum Mainz ZS 1581: tegula (tile) with stamp of legio XXII Primigenia Pia Fidelis, findspot: Mainz, provenance: Frankfurt-Nied.

Bartel / Dolata / Mucha 2000; 2000 / 2001; 2001; 2002; 2003; Dolata / Mucha / Bartel 2001; 2003a; 2003b; 2004; 2006; 2007; Mucha / Bartel / Dolata 2002; 2003a; 2003b; 2005a; 2005b; in press; Mucha et al. 2006; Swart et al. 2004).

The data of ancient coarse ceramics from Germania Superior used here is described by 19 variables: nine oxides (measured in mass-%: SiO_2, TiO_2, Al_2O_3, Fe_2O_3, MnO, MgO, CaO, Na_2O, K_2O) and ten trace elements (measured in ppm: V, Cr, Ni, Zn, Rb, Sr, Y, Zr, Nb, Ba). Besides the different scales of the variables, often problems with outliers and with long-tailed (skew) distributions of the variables in the archaeometric data were addressed, see recently (Baxter 2006).

Fig. 1 shows two objects from our data base that were provided with a stamp. Consequently they appear to be very valuable documents.

The chemical components of the bricks and tiles were measured by Gerwulf Schneider at the Free University Berlin using the X-ray fluorescence analysis (XRF, concerning this method see for instance Leute 1987). At the start a total of 613 Roman bricks and tiles were analyzed. However, there is an ongoing research process. Day-to-day new findings can be reported. For example, quite substantial new findings come from Boppard and Mainz. Current details on both the history and the ongoing research can be obtained from the web site http://www.ziegelforschung.de.

Model-based Cluster Analysis in a Nutshell

The aim of clustering (grouping, unsupervised classification) based on chemical components was both to confirm supposed sites of brickyards and to find places of those ones that are not yet identified. Clustering was accompanied by multivariate graphical presentations of the objects and the clusters found (see the references above and especially (Bartel / Dolata / Mucha 2002) and (Dolata / Mucha / Bartel 2003b). Moreover, clustering was followed by validation of (a) the number of clusters, (b) the stability (reproducibility) of each cluster, and (c) the reliability of the class membership of each object to its cluster (see for example Mucha / Bartel / Dolata in press; Dolata / Mucha / Bartel 2007).

Generally, the final goal of cluster analysis is to find meaningful and stable clusters that can be reproduced to a high degree. When done well, clustering techniques can support scientists of various research areas in their search for hypothesis. Details on cluster analysis can be found in numerous papers (Everitt 1980; Späth 1985; Mucha 1992; Bartel 1996; Papageorgiou et al. 2001). In this instance we present some basic formulae based on distance measures that were also used in the following practical problem of the identification of new findings from the clusters found.

Let a sample of I independent observations (objects) be given in the space R^J and denote by $X = (x_{ij})$ the corresponding data matrix consisting of I rows and J columns (variables), where the element x_{ij} provides a value for the jth variable describing

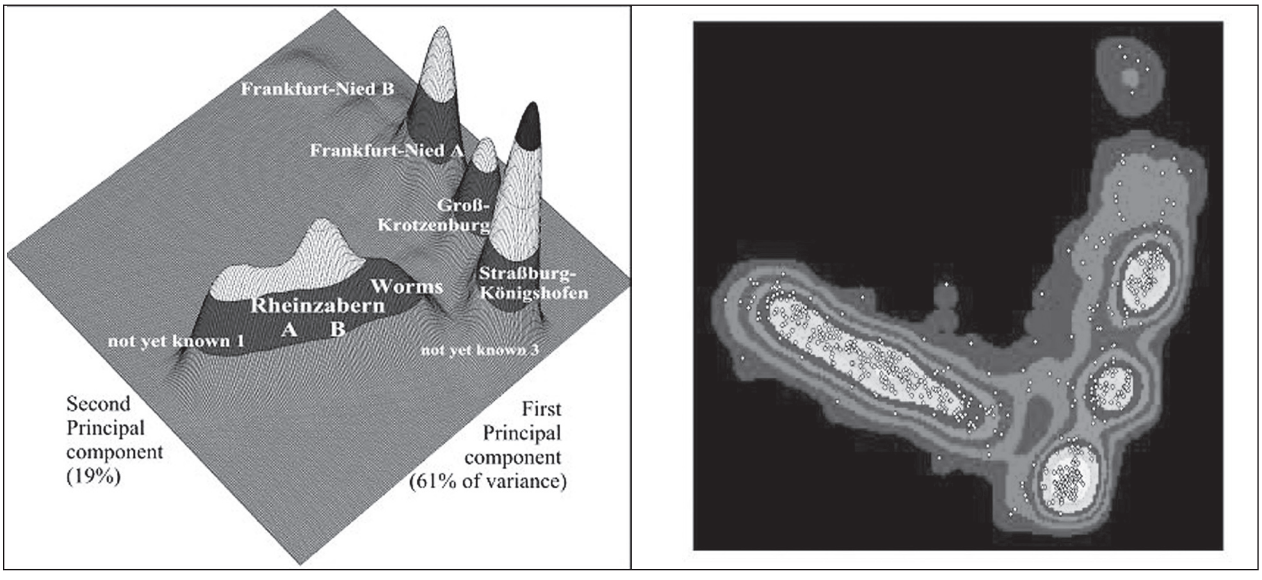

Fig. 2. (left) Bivariate density based on the first two principal components. (right) Several cuts of the density and the underlying objects projected as points bringing together the PCA-plot of points and the continuous density plot.

Fig. 3. Localization of the Roman theater in Mainz.

the ith object. Here the objects are archaeological findings. Further more, let $C = \{x_1, \ldots, x_i, \ldots, x_I\}$ denote the finite set of the I objects. Alternatively written shortly as $C = \{1, \ldots, i, \ldots, I\}$.

Following the definition of the starting point of cluster analysis, then let us formalize the simplest (elementary) solution to the clustering problem with a fixed number of clusters K: a partition $\{C_1, \ldots, C_K\}$ of C, where every pair of two subsets (clusters) C_k and C_l has an empty intersection, and where the union of all K clusters give the total set C.

In the simplest kind of model-based Gaussian cluster analysis the sum of within-clusters sum of squares criterion has to be minimized. This criterion can be formulated as

$$V_K = \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i \in C_k} \sum_{\substack{l \in C_k \\ l > i}} d_{il}$$

by using the squared Euclidean distance

$$d_{il} = d(\mathbf{x}_i, \mathbf{x}_l) = (\mathbf{x}_i - \mathbf{x}_l)^T (\mathbf{x}_i - \mathbf{x}_l) = \left\| \mathbf{x}_i - \mathbf{x}_l \right\|^2$$

between two objects i and l. Here n_k is the number of objects in cluster C_k.

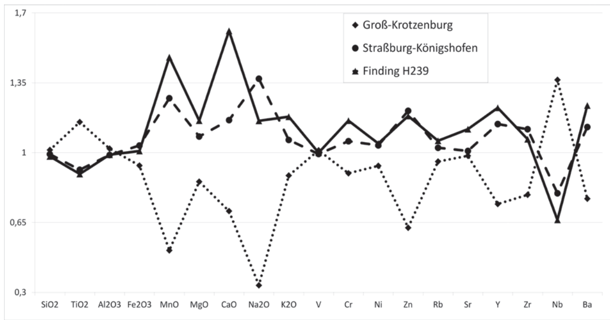

Fig. 5. Graphical presentation of the identification of brick H239 to the provenance Straßburg-Königshofen using chemical profiles.

In the case of adaptive cluster analysis used here the criterion above is modified by considering adaptive weights of variables in the definition of the distance measure (Mucha/Bartel/Dolata 2005a). The squared weighted Euclidean distance

$$d_{il}^{\mathrm{w}} = \sum_{j=1}^{J} \bar{s}_j^{-2} (x_{ij} - x_{lj})^2$$

is such an adaptive distance, where $\bar{s}_j$ is the pooled standard deviation of the variable j (Mucha/Bartel/Dolata 2002).

Roman Bricks and Tiles Classified

Successful applications of simple model-based Gaussian clustering of Roman bricks and tiles have already been reported (Mucha / Bartel / Dolata 2002). In this case new adaptive distances were applied for finding provenances of production of military brickyards. As a result the following locations of military brickyards are identified: Frankfurt-Nied, Groß-Krotzenburg, Rheinzabern, Straßburg-Königshofen, Worms (initially 'not yet known 2') and two with respect to their provenience not yet

Fig. 4. (left) A brick from the Lothary brickyard from Mainz (19th century). (right) A tile with a stamp of the legio XII that was recently found in Boppard (Rhine).

known ones. In *Fig. 2* the 'mountains' (clusters) of the bivariate density estimation and several cuts of this density are shown. This estimation is based on the first two components of the principal component analysis (PCA) of 613 objects. The PCA is a projection technique that, hopefully, makes visible the essentials of the data in two dimensions (for details see GREENACRE 1984; JACKSON 1991; MUCHA 1992). The quality of projection into two principal components is high (80% of variance). The bivariate density looks like a main data body with two arms. Some of the clusters are compact ones; others are not clearly isolated from one another. The right arm consists of three compact clusters. In opposition, the ridge at the left hand side shows neither clear peaks nor clear separation into clusters.

Usually the locations of findings are geographically different from the locations of military brickyards. They all have in common that they are located nearby rivers. The transport was done by cargo ships.

Substantial New Findings in Mainz and Boppard

During the excavation of a Roman theater in Mainz *(Fig. 3)* many bricks and tiles were found. The chemical compositions of 70 objects were measured. The question arose: Was there a Roman military brickyard in Mainz? In order to answer this question a comparison with bricks from Worms, Rheinzabern, Frankfurt-Nied, and from the modern brickyard of Christian Lothary was done. The latter was located in Mainz. Thirty bricks from this brickyard *(Fig. 4)* were analyzed.

The multivariate statistical comparison based on the chemical composition delivered the mathematically confirmed result that without any doubt the objects from the theater were produced in Worms (DOLATA / MUCHA / BARTEL 2006; MUCHA ET AL. 2006).

During an important excavation in Boppard, 43 additional bricks and tiles were discovered *(Fig. 4)*. It could be shown that these objects can be assigned to Worms with high probability (MUCHA / BARTEL / DOLATA 2005a). Additionally the archaeological knowledge about the military brickyard in Worms and its importance in the Rhine area could be improved.

Provenance Identification of New Findings

As demonstrated above, the assignment (identification) of provenance to new findings is an important task in archaeometry. Beside identification of large sets of objects, often individual objects have to be assigned to known brickyards or in general to identified locations of production of any kind of artifacts. An identification of such objects can be based on the same distance measures that are used in cluster analysis. We recommend the K-nearest-neighbor technique (BARTEL / DOLATA / MUCHA 2004).

Fig. 5 gives an additional graphical insight when comparing chemical profiles (fingerprints). For simplicity reason here the comparison is restricted to three profiles only: the pooled profile of members of the clusters Straßburg-Königshofen and Groß-Krotzenburg, and the chemical fingerprint of the finding H239 with unknown location of production. The numerical distance values are: d (H239, Straßburg-Königshofen) = 0.37 << d (H239, Groß-Krotzenburg) = 4.13. Therefore, H239 is assigned to Straßburg-Königshofen by the nearest-neighbor rule. By visual inspection of *Fig. 5* the profile of H239 looks much more similar to Straßburg-Königshofen.

Summary

During the past dozen years, a complex model of history and relations of Roman brick and tile production in south-west Germany has been developed by archaeologists. Clustering techniques are able to support the archaeologists in their search for hypothesis. The cluster analysis referred to here comes with validated results and outstanding multivariate graphics of objects and clusters. The graphics are visual methods for obtaining a better understanding of the statistical results obtained by the archaeologists.

References

BARTEL 1996

H. BARTEL, Mathematische Methoden in der Chemie (Heidelberg 1996).

BARTEL / DOLATA / MUCHA 2000

H. BARTEL / J. DOLATA / H. MUCHA, Klassifikation gestempelter römischer Ziegel aus Obergermanien. Archäometrie und Denkmalpflege, 2000, 86–88.

BARTEL / DOLATA / MUCHA 2000/2001

H. BARTEL / J. DOLATA / H. MUCHA, Klassifikation von 613 Proben als Referenzen für die Herstellungsprovenienzen römischer Baukeramik im nördlichen Obergermanien. Mainzer Archäometrische Zeitschrift 7/8, 2000/2001, 275–300.

BARTEL / DOLATA / MUCHA 2001

H. BARTEL / J. DOLATA / H. MUCHA, Parametrische und nichtparametrische Identifikationsmethoden, dargestellt am Beispiel römischer Baukeramik aus Obergermanien. Archäometrie und Denkmalpflege, 2001, 104–106.

BARTEL / DOLATA / MUCHA 2002

H. BARTEL / J. DOLATA / H. MUCHA, Automatische Klassifikation in der Archäometrie. Berliner und Mainzer Arbeiten zu oberrheinischen Ziegeleien in römischer Zeit. Berliner Beiträge zur Archäometrie 19, 2002, 31–62.

BARTEL / DOLATA / MUCHA 2003

H. BARTEL / J. DOLATA / H. MUCHA, Über eine Modifikation eines graphentheoretisch basierten partitionierenden Verfahrens der Clusteranalyse. Match 48, 2003, 209–223.

BARTEL / DOLATA / MUCHA IN PRESS

H. Bartel / J. Dolata / H. Mucha, Über Identifikationsmethoden, dargestellt am Beispiel römischer Baukeramik aus Obergermanien. Berliner Beiträge zur Archäometrie 21, in press.

BAXTER 2006

M. BAXTER, A Review of Supervised and Unsupervised Pattern Recognition in Archaeometry. Archaeometry 48, 2006, 671–694.

DOLATA 1996

J. DOLATA, Hin zu einer archäologischen Nutzanwendung geochemischer Analytik römischer Baukeramik. Mainzer Archäometrische Zeitschrift 3, 1996, 105–125.

DOLATA 1998

J. DOLATA, Archäologische und archäometrische Untersuchung an römischer Baukeramik und Ziegelstempeln. Archäometrie und Denkmalpflege, 1998, 93–95.

DOLATA 1999

J. DOLATA, Ingenieurtechnische Untersuchung an antiken Ziegelsteinen aus Mainz: Interdisziplinäre Erforschung römischer Baukeramik und Ziegelstempel. Ziegel Zeitschrift 6, 1999, 421–423.

DOLATA 2000

J. DOLATA, Römische Ziegelstempel aus Mainz und dem nördlichen Obergermanien. PhD. Thesis (Frankfurt am Main 2000).

DOLATA 2001

J. DOLATA, Inventarisation und Forschung zu archäologischem. Römische Ziegelstempel aus Mainz – Erreichtes und Perspektiven. Denkmalpflege in Rheinland-Pfalz 52–56, 2001, 488–494.

DOLATA / MUCHA / BARTEL 2001

J. DOLATA / H. MUCHA / H. BARTEL, Untersuchungsperspektiven zum kombinierten archäologischen und materialanalytischen Nachweis römischer Ziegelherstellung in Mainz und Worms. Archäometrie und Denkmalpflege, 2001, 107–109.

DOLATA / MUCHA / BARTEL 2003a

J. DOLATA / H. MUCHA / H. BARTEL, Statistische Untersuchung zur Aufklärung der Binnenstruktur römischer Ziegelproduktion von Frankfurt-Nied. Archäometrie und Denkmalpflege, 2003, 40–42.

DOLATA / MUCHA / BARTEL 2003b

J. DOLATA / H. MUCHA / H. BARTEL, Archäologische und mathematisch-statistische Neuordnung der Orte römischer Baukeramikherstellung im nördlichen Obergermanien. Xantener Berichte 13, 2003, 381–409.

DOLATA / MUCHA / BARTEL 2004

J. DOLATA / H. MUCHA / H. BARTEL, Eine Anwendung der hierarchischen Clusteranalyse. 'Unbekannt 1' als Provenienz einer bekannten obergermanischen Heeresziegelei? Archäometrie und Denkmalpflege, 2004, 28–30.

DOLATA / MUCHA / BARTEL 2006

J. DOLATA / H. MUCHA / H. BARTEL, Provenienz von Ziegeln aus dem römischen Theater in Mainz. Archäologische Bewertung von clusteranalytischen Resultaten. Archäometrie und Denkmalpflege, 2006, 150–152.

DOLATA / MUCHA / BARTEL 2007

J. DOLATA / H. MUCHA / H. BARTEL, Uncovering the Internal Structure of the Roman Brick and Tile Making in Frankfurt-Nied by Cluster Validation. In: R. DECKER / H.-J. LENZ (EDS.), Advances in Data Analysis (Berlin 2007) 663–670.

DOLATA / WERR 1999

J. DOLATA / U. WERR, Wie gleich ist derselbe? – Homogenität eines römischen Ziegels und Aussagegrenzen geochemischer Analytik aufgrund von Meßtechnik und Materialvarianz. Mainzer Archäometrische Zeitschrift 5/6, 1999, 129–147.

EVERITT 1980

B. EVERITT, Cluster Analysis (New York 1980).

GREENACRE 1984

M. GREENACRE, Theory and Applications of Correspondence Analysis (New York 1984).

HARTIGAN 1975

J. HARTIGAN, Clustering Algorithms (New York 1975).

Jackson 1991
J. Jackson, A User's Guide to Principal Components (New York 1991).
Leute 1987
U. Leute, Archaeometry – An Introduction to Physical Methods in Archaeology and the History of Art (Weinheim 1987).
Mucha 1992
H. Mucha, Clusteranalyse mit Mikrocomputern (Berlin 1992).
Mucha 2007
H. Mucha, On Validation of Hierarchical Clustering. In: R. Decker / H.-J. Lenz (eds.), Advances in Data Analysis (Berlin 2007) 115–122.
Mucha / Bartel / Dolata 2002
H. Mucha / H. Bartel / J. Dolata, Exploring Roman Brick and Tile by Cluster Analysis with Validation of Results. In: W. Gaul / G. Ritter (eds.), Classification, Automation, and New Media (Berlin 2002) 471–478.
Mucha / Bartel / Dolata 2003a
H. Mucha / H. Bartel / J. Dolata, Core-Based Clustering Techniques. Methods, Software, and Applications. In: M. Schader / W. Gaul / M. Vichi (eds.), Between Data Science and Applied Data Analysis (Berlin 2003) 74–82.
Mucha / Bartel / Dolata 2003b
H. Mucha / H. Bartel / J. Dolata, Modellbasierte Clusteranalyse römischer Ziegel aus Worms und Rheinzabern. Archäologische Informationen 26/2, 2003, 471–480.
Mucha / Bartel / Dolata 2005a
H. Mucha / H. Bartel / J. Dolata, Model-Based Cluster Analysis of Roman Bricks and Tiles from Worms and Rheinzabern. In. C. Weihs / W. Gaul (eds.), Classification – The Ubiquitous Challenge (Berlin 2005) 317–324.
Mucha / Bartel / Dolata 2005b
H. Mucha / H. Bartel / J. Dolata, Techniques of Rearrangements in Binary Trees (Dendrograms) and Applications. Match 54, 2005, 561–582.
Mucha et al. 2006
H. Mucha / H. Bartel / J. Dolata / C. Swart / K. Graubner, Zur Ausweisung einer Ziegelreferenz »Mainz« durch Bewertung der Stabilität von Klassen- und Objektzuweisungen. Archäometrie und Denkmalpflege, 2006, 153–155.
Mucha / Bartel / Dolata in press
H. Mucha / H. Bartel / J. Dolata, Effects of Data Transformation on Cluster Analysis of Archaeological Data. In: C. Preisach / H. Burkhardt / L. Schmidt-Thieme / R. Decker (eds.): Data Analysis, Machine Learning, and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation, Berlin, 2007.
Papageorgiou / Baxter / Cau 2001
I. Papageorgiou / M. Baxter / M. Cau, Model-based Cluster Analysis of Artefact Compositional Data. Archaeometry 43, 2001, 571–588.
Späth 1985
H. Späth, Cluster Dissection and Analysis (Chichester 1985).
Swart et al. 2004
C. Swart / B. Paz / J. Dolata / G. Schneider / J. Simon / H. Bartel / H. Mucha, Analyse römischer Ziegel mit ICP-MS/-OES. Methodenvergleich zwischen RFA und ICP. Archäometrie und Denkmalpflege, 2004, 25–27.
Werr 1998
U. Werr, Grenzen der Aussagekraft chemischer Analytik für römische Baukeramik. Archäometrie und Denkmalpflege, 1998, 96–98.

Hans-Joachim Mucha

Weierstraß Institute for Applied Analysis and Stochastics (WIAS)
Mohrenstr. 39
10117 Berlin, Germany
mucha@wias-berlin.de

Hans-Georg Bartel

Institute for Chemistry
Humboldt University Berlin
Unter den Linden 6
10099 Berlin, Germany
hg.bartel@yahoo.de

Jens Dolata

Head Office for Cultural Heritage Rhineland-Palatinate (GDKE)
Große Langgasse 29
55116 Mainz, Germany
dolata@ziegelforschung.de