since 1961

Baltica

BALTICA Volume 32 Number 2 December 2019: 139–155
https://doi.org/10.5200/baltica.2019.2.2

Case study of geochemical clustering as a tool for tracing sources of clays for archaeological and modern bricks

Ričardas Taraškevičius, Vaidotas Kazakauskas, Saulius Sarcevičius, Rimantė Zinkutė, Sergej Suzdalev

Taraškevičius, R., Kazakauskas, V., Sarcevičius, S., Zinkutė, R., Suzdalev, S., 2019. Case study of geochemical clustering as a tool for tracing sources of clays for archaeological and modern bricks. *Baltica, 32 (2),* 139–155. Vilnius. ISSN 0067-3064.

Manuscript submitted 28 February 2019 / Accepted 23 April 2019 / Published online 20 June 2019

Abstract. The study was conducted using 14 hierarchical clustering ways and combining them with 4 inter-related sets of elements, i.e. the contents of Al, Ba, Ca, Fe, Ga, K, Mg, Mn, Na, Nb, P, Rb, Si, Sr, Th and Ti determined by EDXRF in 44 splits of 10 archaeological and 4 modern bricks, as well as in 38 clay samples from two quarries. Empirical scoring of tree dendrograms of archaeological samples helped to identify Complete Linkage, Weighted Pair Group Average and Ward's methods as the most suitable for sourcing. Successful identification of geochemical clustering methods for fingerprinting sources of bricks is determined by the intrinsic features of the geochemical composition of stonework or quarries: their similarity, determined by the geochemical peculiarities of clay indicators, such as Al, Rb, Ga, K, Th, Fe, Ti and Nb, and differences, expressed by the elements that are more abundant in sands and silts (Si, Na), carbonates (Ca, Sr, Mg), organic matter (P) and other lithological-mineralogical tracers. It has been found that the mean values of the geochemical composition of the allied objects have much more useful fingerprinting properties. It is strongly recommended for source fingerprinting to select not only typical lithological-mineralogical samples, but also homogeneous sampling sets excluding possible outliers. It has been shown that each raw clay material has its own specific geochemical features. This is an essential useful feature for source fingerprinting using clustering of the objects of interest.

Keywords: *EDXRF analysis; inter-element correlation; hierarchical cluster analysis; cluster dendrogram; fingerprinting; scores; archaeometry*

✉ *Ričardas Taraškevičius (ricardas.taraskevicius@gmail.com), Nature Research Centre, Institute of Geology and Geography, Akademijos St. 2, LT-08412, Vilnius, Lithuania; Vaidotas Kazakauskas (vaidotas.kazakauskas@gamtc.lt), Nature Research Centre, Institute of Geology and Geography, Akademijos St. 2, LT-08412, Vilnius, Lithuania; Saulius Sarcevičius (sausarii@gmail.com), Lithuanian Institute of History, Urban Research Department, Kražių St. 5, LT–01108 Vilnius, Lithuania; Rimantė Zinkutė (rimante.zinkute@gamtc.lt), Nature Research Centre, Institute of Geology and Geography, Akademijos St. 2, LT-08412, Vilnius, Lithuania; Sergej Suzdalev (sergej.suzdalev@apc.ku.lt), Klaipėda University, Marine Research Institute, Universiteto Ave. 17, LT-92294 Klaipėda, Lithuania*

INTRODUCTION

Tracing sources of air, soil and sediments pollution is a topical task for the world community. Fingerprinting (tracing sources) is also used in many other spheres, e. g. in forensic science, archaeology, etc. In archaeology, when searching for sources of products (ceramics, metals, flint, glaze, pigments of artwork, etc.), researchers inevitably encounter the need to have different geological, especially geochemical (including mineralogical and petrographic) information.

In the wide review by Owens *et al.* (2016) about fingerprinting and tracing sources of soils and sediments, geoarchaeological studies are mentioned as one of the fields of application. The author also describes achievements in sourcing clays for pottery presuming that mineralogical, petrological and geochemical signatures of raw materials are preserved in ceramics and therefore the task of sourcing is quite feasible. In their research, Gutsuz *et al.* (2017*)* as well as other researchers (Weigand *et al.* 1977; Pollard, Heron 1996; Glascock *et al.* 2004; Hein *et al.* 2004, etc.) rely on the well-established simple assumption, which is known as the provenance postulate (the investigated ceramic product is chemically identical to clay as a source material) and is verified in numerous archaeometric studies. However, the task is not easy, because ceramic artefacts are a combination of fine clay matrix and temper, the problem lies in the unmixing of target samples (Owens *et al.* 2016). Montana *et al.* (2011) also state that direct linking of clays to ceramic products is possible only when raw material is used without any modifications. Owens *et al.* (2016) pay attention to the fact that different analytical methods are used for analysing ceramics, e.g. tempers may be analysed by petrological methods (e.g., Gonzales *et al*. 2015), clay – by geochemical methods. The fingerprinting applied by Owen *et al.* (2013) revealed differences even between the raw material of modern and ancient pottery.

Algorithms and peculiarities of most hierarchical clustering methods were described long ago (Sokal, Sneath 1963; Ward 1963). A common equation for these algorithms with different coefficients was provided by Lance and Williams (1967) and supplemented by Wishart (1969). Advantages and disadvantages of hierarchical clustering methods were described by Kopp (1978 a, b, c). Some researchers tested the performance of different cluster analysis methods using simulated data (Kuiper, Fisher 1975; Bayne *et al.* 1980; Jain *et al.* 1986; Saracli *et al.* 2013), while others real data, i.e. phytoplankton (Carteron *et al.* 2012), gene expression (Jaskowiak *et al.* 2014). Cluster dendrograms are also widely used in archaeology, unfortunately, without indicating clustering methods (Glascock *et al.* 2004; Waksman 2017) or without explanation why a certain method is used (Buxeda i Garrigos *et al.* 2001; Gutsuz *et al.* 2017), the same applies to our own publications (Taraškevičius *et al.* 2013; Sarcevičius, Taraškevičius 2015). In data analysis description, Hall (2017) also does not explain the choice of a suitable clustering method, which can predetermine conclusions of the investigation, because there is a wide variety of methods, as well as opinions about them. Of course, mathematicians have extensively developed the theory of cluster analysis, e.g. have described its 3 standard strategies ("the nearest neighbour", "the furthest neighbour" and "centroid") and their 3 properties, have defined the so-called "chaining effect" and emphasized the importance of the distance measure selection (Lance, Williams 1967), have explained admissibility properties (Fisher, Van Ness 1971; Van Ness 1973), importance of the monotonicity requirement, criteria for evaluating dendrograms (Sokal, Rohlf 1962) and clustering methods (Rand 1971), cluster validity (Dubes, Jain 1979).

Therefore, the aim of our research is to test 14 clustering ways by including the contents of various combinations of chemical elements determined by EDXRF in selected archaeological bricks (with the known origin of the respective stonework) as well as in modern bricks and clay used as a raw material for their production. After scoring the dendrograms obtained, the next aim is to describe the most attractive variants of geochemical clustering as a tool for fingerprinting.

MATERIALS, METHODS, ANALYSIS

Study design

The study design was determined by the sequence of steps for the implementation of the following tasks:

1) to test some options of hierarchical cluster analysis methods taking into account the following 3 main factors: sets of inter-related variables, distance measure and the method of amalgamation using the geochemical data of archaeological brick splits;

2) to attribute empirical scores to the aforementioned cluster analysis dendrograms and analyse them based on the information about the origin of archaeological bricks aiming to choose 3 most useful clustering ways and geochemical clustering tools for the recognition of this origin presuming that the same ways and tools may be also used for fingerprinting the clays used for brick production;

3) to test the suitability of selected ways and tools for geochemical source fingerprinting of clay as a raw material for brick production.

Material, sampling, study area

In order to identify the most proper ways and tools for the "fingerprinting*"* task (Owens *et al.* 2016), 30 random samples (splits) from 11 archaeological bricks were collected (Table 1, Fig. 1 a, b). Archaeological investigations of the bricks had been carried out earlier and the bricks had been described by various researchers (Kitkauskas 2009; Levandauskas 2012). Their primary geochemical characterisation was given by Taraškevičius *et al.* (2013). For test-ing of selected clustering options to reveal sources of clay as a raw material for bricks, we collected 14 random split samples from 4 randomly selected monolithic and hollow modern bricks produced from raw clay from Pašaminė and Rėvai quarries (Table 1).

To perform the geochemical fingerprinting task, 38 samples of raw clay material were collected, 31 sample of which was from the Pašaminė quarry and 7 samples were from the Rėvai clay quarry (Fig. 1a).

The Pašaminė section is located in an operating clay quarry, 7 km northeast of Švenčionėliai town. It is represented by a 400 m long, 6.5–7.5 m high inclined exploitation wall. The clay pit has been exploited until these days. Glaciolacustrine clayey sediments from the studied Pašaminė section (Fig. 1a) reflect sedimentation in a local proglacial lake during the East Lithuanian phase of the Baltija stadial (Uchman *et al.* 2008, 2009) or Grūda stadial (Guobytė 1999) of the Nemunas (Weichselian) Glaciation. Sediments occur in a 2 km wide and 4 km long area surrounded by marginal till deposits. Maximal thickness of the glaciolacustrine clayey sediments in the western part of the area reaches 23 m, while in the other parts only 5–6 m. Our investigated section of glaciolacustrine sediments (7.4 m thick) is mostly composed of homogeneous (massive) dark brown or brown clay (about 87% of the section) with laminated intervals of grey silt or silty sand (about 13% of the section). For the detailed analysis of the vertical distribution of chemical elements, 22 samples from the Pašaminė main section (Figs. 2, 6) were collected. To analyse the lateral and vertical distribution of elements along the clay quarry wall, 9 additional samples were taken from 3 short sections (a, c, d) and 3 samples (b short section) were used from the main section. These sections were located at every 100 m (Fig. 2).

The Rėvai section is located in the already closed clay quarry, 9 km southeast of Alytus town. The varved clay of the studied Rėvai section (Fig. 6) reflects sedimentation in the local glaciolacustrine basin during the South Lithuanian phase of the Baltija stadial of the Nemunas (Weichselian) Glaciation. Our investigated section (2.7 m thick) of glaciolacustrine sediments is composed of brown or dark brown clay layers of varying thickness (about 40% of the section) and greyish brown silt or silty clay (about 60% of the section) with local convolute deformations in

Table 1 Description of split samples of brick artefacts in Vilnius and of modern bricks made of clay from Pašaminė and Rėvai quarries

[a] Common origin (IDs) and (coordinates of objects)	Specific features of the object	[b] Brick IDs	[c] Splits IDs	[d] n	Age
Vilnius Lower Castle (M) (54°41′08″N; 25°17′20″E)	[e] Stonework M16	M16	M16-a, M16-b, M16-c	3	2nd half of 13th – 1st half of 14th century
	[e] Stonework M22	M22	M22-a, M22-b, M22-c	3	
	[e] Stonework M25	M25	M25-a, M25-b, M25-c	3	~1323
Defence wall (D) (54°40′35″N; 25°17′31″E)	Fragment of stonework	DW	DW-a, DW-b, DW-c	3	1503–1522
Church of St. Anna & Holy Barbara (A) (54°41′15″N; 25°17′25″E)	[f] LNM VNLS, AV 40:25	AB25	AB25-a, AB25-b	2	1551–1572
	[f] LNM VNLS, AV 40:36	AB36	AB36-a, AB36-b	2	
	[f] LNM VNLS, AV 40:42	AB42	AB42-a, AB42-b	2	
	[f] LNM VNLS, AV 40:102	AB102	AB102-a, AB102-b, AB102-c	3	
Burning furnace (F) (54°41′16″N; 25°16′35″E)	Brick from the furnace construction	BBF	BBF-a, BBF-b, BBF-c	3	2nd half of 15th – 1st half of 16th century
Bishop's Palace (B) (54°41′08″N; 25°17′13″E)	[g] Secondary use	28'	28'-a, 28'-b, 28'-c	3	[h] 1st half of the 17th century [7]
		35'	35'-a, 35'-b, 35'-c	3	
Bricks (P'B) made of clay from the Pašaminė quarry (P'Q) (55°12′26″N; 26°05′29″E)	Monolithic brick	P'B-M	P'B-M-a, P'B-M-b, P'B-M-c, P'B-M-d, P'B-M-e, P'B-M-f	6	First decade of 21th century
	Hollow brick	P'B-H	P'B-H-a, P'B-H-b, P'B-H-c, P'B-H-d	4	
Bricks (R'B) made of clay from the Rėvai quarry (R'Q) (54°19′21″N; 23°57′55″E)	Monolithic brick	R'B-M	R'B-M-a, R'B-M-b	2	
	Hollow brick	R'B-H	R'B-H-a, R'B-H-b	2	

[a] Location of objects is shown in Fig. 1, IDs (M, A and B) are used in Table 5; [b] Identification of bricks in Fig. 3 and cluster dendrograms; [c] Identification of brick splits in cluster dendrograms; [d] Number of splits; [e] By agreement, the stonework of Vilnius Lower Castle is marked with the letter "M" (Kitkauskas 2009); [f] The numbers of bricks from the Church of St. Anna and Holy Barbara provided by the Department of the Middle and Newest Ages of the Lithuanian National Museum (LNM VNLS) are as follows: AV 40:25, 40:36, 40:42, 40:102; [g] The Bishop's Palace was built using bricks of the previously demolished buildings. The bricks were stamped with the numbers "28" and "35"; [h] Hypothetical date.

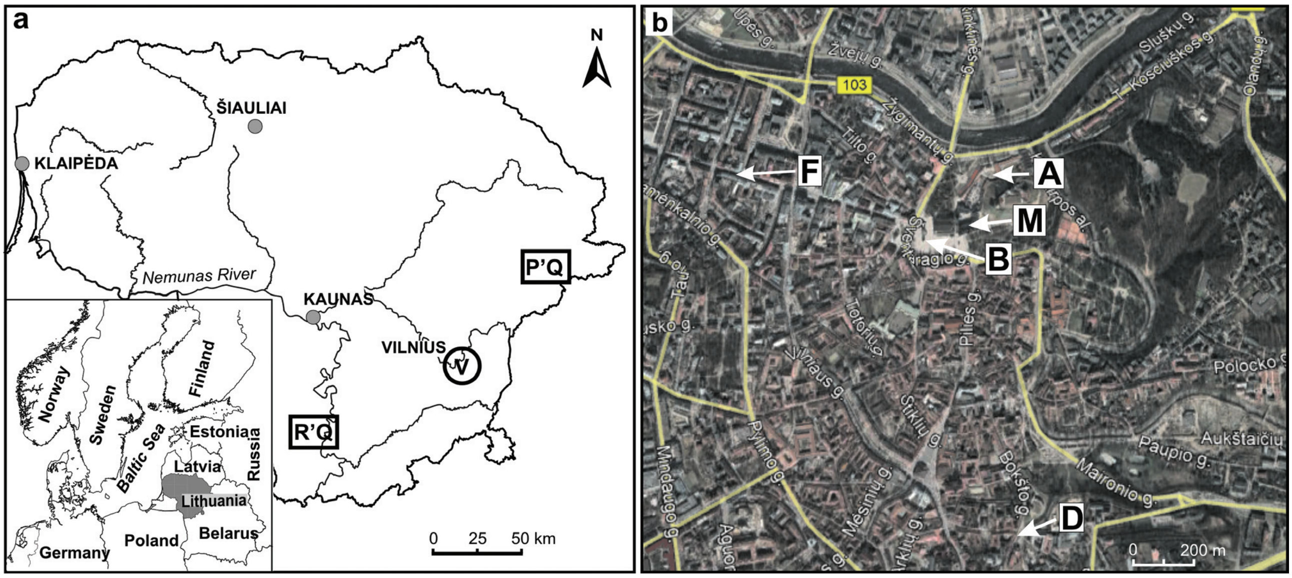

Fig. 1 Location map. (a) Locations of the investigated glaciolacustrine clay sections in Pašaminė (P'Q), Rėvai (R'Q) and brick (V) samples. (b) Locations of the investigated archaeological bricks: A – Vilnius Lower Castle, B – Bishop's Palace, M – Church of St. Anna & Holy Barbara, F – Burning furnace, D – Defence wall

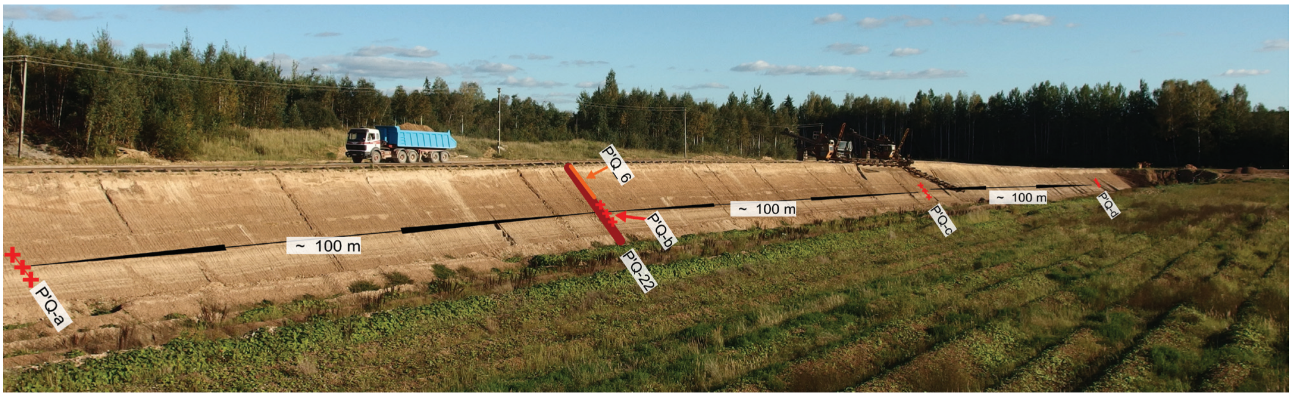

Fig. 2 The Pašaminė quarry and location of the investigated sections: P'Q-22 – a thoroughly investigated section with 22 samples collected for the geochemical analysis; P'Q_6 – the upper part of the section; P'Q-a, P'Q-b, P'Q-c, P'Q-d – short sections with 3 samples collected from each of them for the investigation of the vertical and horizontal distribution of chemical elements

the upper part of the section. To analyse the vertical distribution of chemical elements, 7 samples were collected from the Rėvai section.

Chemical analysis

Samples of brick splits and raw clay from the quarries were air-dried and milled using a MM400 mixer mill in zirconium oxide grinding jars. The milled material was divided into two parts and paired sub-samples were prepared. From each sub-sample, the mixture of 0.90 g of Licowax binder and 4.00 g of milled material was prepared and a pressed pellet with a 32 mm diameter was made. The pellets were pressed for ~3 min using 15 KN.

All sub-samples were analysed using the EDXRF equipment Spectro Xepos (Kleve, Germany, using the Turboquant for the pressed pellet calibration procedure elaborated by the manufacturers) to determine the contents of Al, As, Ba, Br, Ca, Cl, Co, Cr, Cu, Fe, Ga, Hf, Y, K, Mg, Mn, Na, Nb, Ni, P, Pb, Rb, S, Si, Sr, Th, Ti, Tl, U, V, W, Zn and Zr. The median values of the relative standard deviation of paired sub-samples (RSD) were considered as an important factor for the choice of a suitable element group. Only 25 elements (from the 33 aforementioned) characterised by the RSD<7.5% values (Environment Agency: www.mcerts.net), i.e. Rb, Sr, Fe, Al, Si, Mn, Ca, Zn, K, Mg, Y, Zr, Ti, P, Ni, S, Ba, Th, Pb, Nb, Ga, Cl, Na, Cu and Cr (ranked by increasing RSD values), were considered suitable. However, Zr and Y were eliminated from the list due to their presence in zirconium oxide grinding jars. Additionally, several other elements representing trace metals with the possible anthropogenic contribution (Taraškevičius *et al.* 2007; Taraškevičius *et al.* 2016) were discarded from the final analysis. The final list included 16 chemical elements: Al, Ba, Ca, Fe, Ga, K, Mg, Mn, Na, Nb, P, Rb, Si, Sr, Th and Ti. Materials of the International Soil-Analytical Exchange program (Wageningen University, period 2010–2017) and reference standard SRM 679 (Brick Clay) were prepared *pari passu* and used for the recalibration of the real total contents of the selected elements.

Statistical data treatment

Agglomerative hierarchical clustering results depend on 3 factors: variables included in the analysis, distance measure between the samples and the amalgamation method. Empirical experiments with clustering of archaeological brick splits from the ancient buildings, whose origin is known, were performed taking into account 3 aforementioned factors in order to reveal the most efficient clustering ways and tools for the fingerprinting task.

The experiments were carried out with the help of the STATISTICA 9 software. This software was also used to calculate values of Spearman's rank correlation coefficients and their significance and to test the non-parametric hypotheses using Mann-Whitney U-test.

While performing the first task, 3 more closely inter-related subgroups of variables (elements) were revealed by Ward's clustering method using 1-r distance, where r is Pearson's correlation coefficient. Then each of these variable-subgroups, as well as the whole group of 16 elements was used for the clustering of samples using 2 distance measures (Euclidean or City-block) and 7 methods of amalgamation, which are listed together with their abbreviations in Table 2. The City-block distance was chosen, because it gives results similar to Euclidean distance, but since differences in the formula are not squared, it is less affected by single outliers (see STATISTICA 9 software help).

After scoring clustered data, 3 clustering ways with the highest scores were chosen to be used for the recognition of modern bricks as well as for fingerprinting (sourcing) the clay used as a raw material for their production (in this case, with the known source). These data were tested in 2 ways. Firstly, all 43 samples (including both brick splits and clay) were grouped and after the analysis of the dendrograms had been performed, an attempt to recognise splits of each of the 14 bricks (R'B-M and R'B-H had 2 splits each, P'B-M had 6 splits and P'B-H had 4 splits) and to relate them to the respective clay quarries (7 samples from Rėvai (R'Q-7) and 22 samples from the main section of the Pašaminė quarry) was made.

Secondly, average values in each of the 11 subsets (3 of them are related to the Rėvai object and 8 to the Pašaminė object) were used for compiling dendrograms. The Rėvai object was represented by the average contents of elements in cross section (R'Q-7) and by splits of monolithic (R'B-M) and hollow (R'B-H) bricks produced of this clay. The Pašaminė object was represented in dendrograms by average values of elemental contents in the following 8 subsets: a) in 22 clay samples from the Pašaminė main section (P'Q-22, see Fig. 2); b) in 6 clay samples (P'Q_6) from the upper part (0–2.7 m) of the Pašaminė section (this interval corresponds to the thickness of the Rėvai quarry section); c) in 4 short sections of the Pašaminė quarry with a 100 m distance between them (P'Q-a, P'Q-b, P'Q-c, P'Q-d, each of them is characterised by 3 samples); d) in splits of monolithic (P'B-M) and hollow (P'B-H) bricks produced of Pašaminė clay. The results were generalised using empirical scoring analogous to the scoring of dendrograms of archaeological brick splits.

Empirical criteria and respective scores for the selection of the preferable distance and method

The first criterion for the inter-comparison of algorithm variants concerns bricks: splits of the same brick (their number is either 2 or 3) should form their own first-level cluster (they can be formed at a different linkage distance, for UC and WC methods at a different step of amalgamation). The scores given to the brick criterion (*BSc*) are as follows: *BSc* = 2 if all splits of the same brick belong to the first-level cluster, *BSc* = 1 if at least 2 of 3 splits of the same brick are in the first-level cluster, but the third one joins a higher-level cluster, to which splits of other bricks also belong, *BSc* = 0 if each split consecutively joins a higher-level cluster also containing other brick splits (the latter case is the so-called "chaining effect"). The total number of bricks was 11, so the maximum possible sum of *BSc* scores was 22.

The second criterion was intended for brick splits of the same-origin object. The respective scores (*OSc*) take into account only the clustering of splits from the 3 main objects (Fig. 1b): the numbered bricks of the Church of St. Anna & Holy Barbara (A), Bishop's Palace (B) and Vilnius Lower Castle (M). It can be formulated as follows: splits of bricks of each of these objects should form their own cluster, which should be as pure as possible. The calculation of *OSc* is as follows: *OSc* = 2 if all brick splits of the selected object belong to the same cluster without any admixture of splits from the other 2 objects, *OSc* = 1 in two cases: a) if all splits of the selected object's bricks form their own cluster, but there is some admixture of splits of another object's bricks in this cluster, b) if splits of at least 2 bricks of the selected object form their own cluster, but the remaining splits belong to another cluster where splits of another object prevail;

OSc = 0, if first-level clusters of brick splits of selected object do not form their own higher-level cluster, but one-by-one join a higher-level cluster where brick splits from another object prevail (the latter case is higher-level "chaining effect"). Since 3 objects were used for this criterion, the maximum possible sum of *OSc* scores for archaeological bricks was 6.

The third complex criterion (with *Csc* scores) took into account the clustering of 24 brick splits of the same 3 objects (A, B, M). The *Csc* values were the sum of positive and negative scores. Each brick split received 1 positive score, so the maximum possible sum of *Csc* values could be 24. Negative scores were given only to those brick splits which were not in "their own cluster" formed at the lowest linkage distance, i.e. to those which "ran away for their own cluster". The value of the negative score given to each such brick split was equal to the number of hierarchical steps until this split joined its "own cluster", but mixed with splits of other bricks.

RESULTS

The main geochemical parameters of the selected archaeological and modern bricks

According to its average content in both archaeological and modern bricks (307582 mg/kg), Si is confidently leading in their geochemical composition (Fig. 3). Its mean contribution to the total amount of all the studied 16 chemical elements is 58%. The arrangement of the other 9 chemical elements, listed by Rudnick and Gao (2003) as major elements of the upper continental crust, in descending sequence according to their average amount is as follows: Al>Ca>Fe>K>Mg>Na>Ti>P>Mn. The overall contribution of all major elements to the total amount reaches 99.75%. As for other 6 elements, the contribution of Ba to the total amount (0.085%) is 1.5 times higher than the overall contribution of the remaining 5 elements, i.e. Rb, Sr, Ga, Nb, Th (0.055%).

The average contents of Ba, Fe, Ga, Mn, P, Ti, Th, Nb and Mg in bricks are closer to the concentrations in the upper continental crust given by Rudnick and Gao (2003) than to those given by Wedepohl (1995), meanwhile the contents of Al, K and Ca in bricks are more similar to these data.

Clustering of archaeological brick splits

The clustering of brick splits by 7 methods, i.e. UA, WA, UC, WC, SL, CL and W, using Euclidean and City-block distances (Table 2) was obtained (Figures in Supplements, Figs. S1–S4).

In order to select variables for the compilation of tree diagrams demonstrating clustering of brick splits, the relationships between elemental contents were assessed by analysing their Spearman's rank correlation

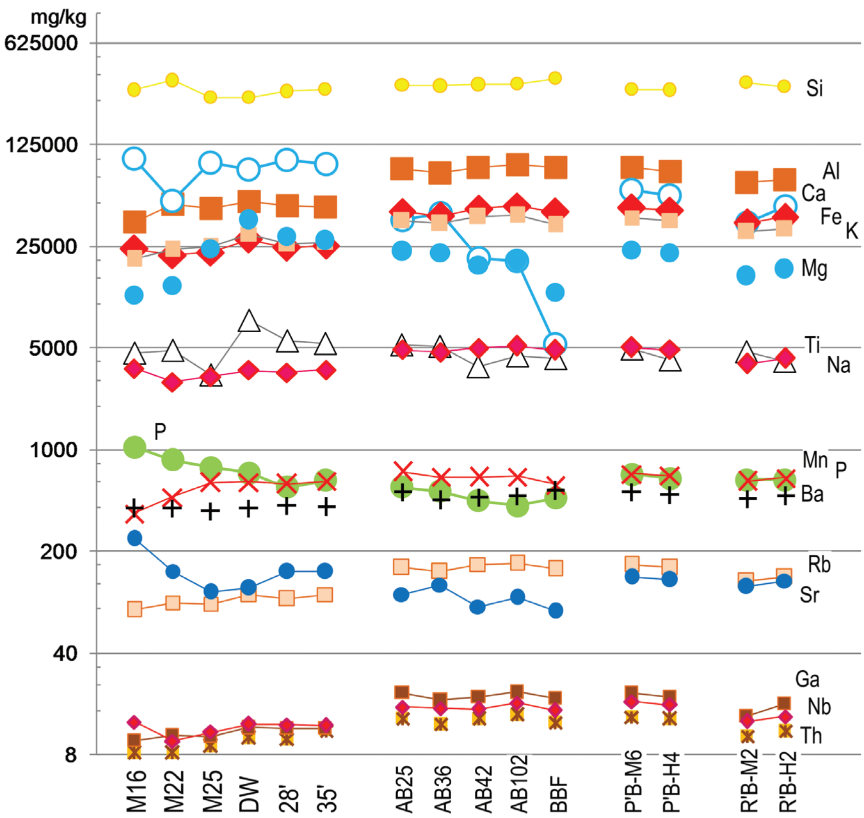

Fig. 3 Distribution of average contents of elements in archaeological and modern bricks. Abbreviations are given in Table 1

coefficients and the cluster dendrogram of their paragenetic associations (Fig. 4).

The first group of tree diagrams of brick splits (Figures in Supplements, Fig. S1) was compiled using 8 closely inter-correlated chemical elements, which are clay proxies with Al and Fe prevailing and Rb, Ga, K, Th, Ti and Nb closely related to them (Fig. 4c). The second group of tree diagrams (Figures in

Table 2 Terms used and their abbreviations

Terms (explanation)	Abbreviation
Unweighted Pair Group Average method	UA
Weighted Pair Group Average method	WA
Unweighted Pair Group Centroid method	UC
Weighted Pair Group Centroid (median) method	WC
Single Linkage method	SL
Complete Linkage method	CL
Ward's method	W
Euclidean distance	E
City-block distance	CB
Clustering way, i.e. a combination of the clustering method and distance measure	C-way, e.g. [a] CL&E
[a] Complete Linkage method with Euclidean distance	CL&E
Geochemical clustering, i.e. a combination of the clustering way and a set of chemical elements used in cluster analysis for fingerprinting tasks	GC-tools, e.g. [b] CL&E/16
[b] Complete Linkage method with Euclidean distance using 16 variables (chemical elements)	CL&E/16
[c] Clay sample from the Rėvai quarry with indicated depth, m	R'Q-2.69
[c] Clay sample from the Pašaminė quarry with indicated depth, m	P'Q-2.99
Average value, mg kg^{-1}	AV
Standard deviation, mg kg^{-1}	SD

[a] An example describing clustering method and distance.
[b] An example describing clustering method and distance and showing the number of chemical elements used in cluster analysis.
[c] An example describing the sample from the clay quarry.

a

	Si	Al	Ca	Fe	K	Mg	Na	Ti	P	Mn	Ba	Rb	Sr	Ga	Nb
Al	**0.62**														
Ca	***-0.79***	***-0.89***													
Fe	0.37	**0.82**	***-0.69***												
K	0.39	**0.90**	***-0.75***	**0.93**											
Mg	***-0.67***	-0.14	0.41	-0.01	0.10										
Na	-0.30	-0.13	0.28	-0.04	-0.03	**0.58**									
Ti	0.43	**0.70**	***-0.67***	**0.92**	**0.80**	-0.26	-0.19								
P	-0.48	***-0.82***	**0.67**	***-0.86***	***-0.89***	0.00	0.12	***-0.74***							
Mn	0.01	0.53	-0.35	**0.68**	**0.77**	0.37	0.06	0.53	***-0.62***						
Ba	**0.64**	**0.78**	***-0.79***	**0.73**	**0.75**	-0.28	-0.12	**0.72**	***-0.76***	0.47					
Rb	0.51	**0.89**	***-0.78***	**0.93**	**0.97**	0.02	-0.04	**0.82**	***-0.90***	**0.71**	**0.78**				
Sr	-0.38	***-0.80***	**0.77**	***-0.67***	***-0.77***	0.09	0.42	***-0.58***	**0.70**	-0.46	***-0.64***	***-0.73***			
Ga	0.44	**0.85**	***-0.71***	**0.92**	**0.94**	0.06	0.04	**0.81**	***-0.90***	**0.70**	**0.77**	**0.95**	***-0.68***		
Nb	0.37	**0.68**	***-0.63***	**0.87**	**0.81**	-0.21	-0.11	**0.92**	***-0.73***	**0.57**	**0.72**	**0.77**	-0.56	**0.79**	
Th	0.39	**0.82**	***-0.70***	**0.93**	**0.95**	0.08	-0.07	**0.84**	***-0.90***	**0.77**	**0.78**	**0.95**	***-0.71***	**0.94**	**0.80**

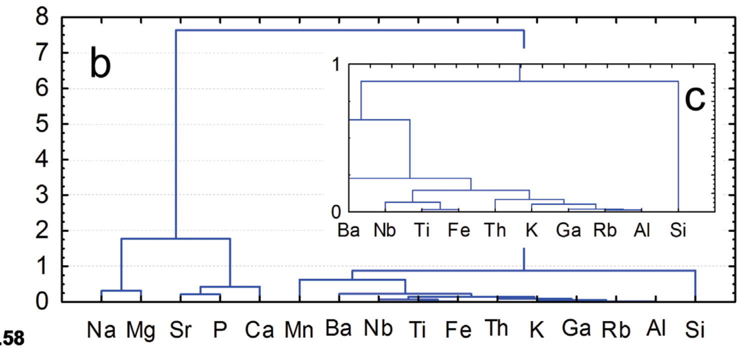

Fig. 4 Spearman's rank correlation coefficients between element contents in the splits of archaeological bricks collected from Vilnius (a) and clusters of the associated chemical elements (b). Links between clay indicators (Nb, Ti, Fe, Th, K, Ga, Rb, Al) are given in more detail (c). If the coefficient of correlation (r_{50}) is in bold, it is significant at $p < 0.001$

Supplements, Fig. S2) was obtained by adding Ba, Mn and Si to the aforementioned clay indicators and using 11 chemical elements (Mn, Ba, Nb, Ti, Fe, Th, K, Ga, Rb, Al and Si) from the right branch of the tree dendrogram (Fig. 4b) in cluster analysis. For the compilation of the third group (Figures in Supplements, Fig. S3), the number of variables was reduced to 9, i.e. only Si was added to clay indicators. This decision was taken, because the contribution of Si to the total amount of the study elements in bricks was the highest and its correlation with Al (r = 0.62) was strong (significant at $p < 0.001$) (Fig. 4a). To form the fourth group (Figures in Supplements, Fig. S4), all 16 chemical elements were used.

The summary of scores (*BSc)* according to the criterion for the recognition of the same brick splits is given in Table 3. The highest sum of *Bsc* (80) was obtained by CL&CB clustering way. The dendrograms compiled using the other three clustering way options (CL&E, WA&CB, W&CB) follow CL&CB and have only a slightly lower score (79). Tree diagrams based on the selected groups of 8, 11, 9 and 16 elements, i.e. geochemical clustering options CL&CB/8, CL&CB/11, CL&CB/9 and CL&CB/16 are given in Fig. 5.

Attention should be paid to the fact that when alone, this criterion is insufficient, e.g. splits of AB102 brick and AB42 brick (Fig. 5d) are more strongly related to brick splits of the Brick Burning Furnace (BBF) than to splits of the other two bricks of the same object, i.e. Church of St. Anna & Holy Barbara. Therefore other scores (*OSc)* according to the criterion for the recognition of the same origin object brick splits were also necessary for inter-comparison of the variants. The summary of *OSc* is given in Table 4.

There are several drawbacks to the 2 aforemen-

Table 3 Summary of scores (*Bsc*) according to the criterion for the recognition of the same archaeological brick splits

[a] C-way (method and distance)	*BSc*				Sum of *BSc (C-way)*
	[b]/8	/11	/9	/16	
UA&E	16	19	19	21	75
UA&CB	16	19	20	22	77
WA&E	17	20	19	21	77
WA&CB	17	20	20	22	79
UC&E	15	17	19	20	71
UC&CB	15	18	20	20	73
WC&E	12	14	19	20	65
WC&CB	15	19	20	20	74
SL&E	16	18	19	20	73
SL&CB	16	18	20	20	74
CL&E	18	20	19	22	79
CL&CB	17	20	21	22	80
W&E	16	20	21	21	78
W&CB	17	20	20	22	79
Sum of *BSc(n)*	223	262	276	293	

[a] Abbreviations are given in Table 2.
[b] The number (n) of chemical elements used for geochemical clustering (given after a slash).

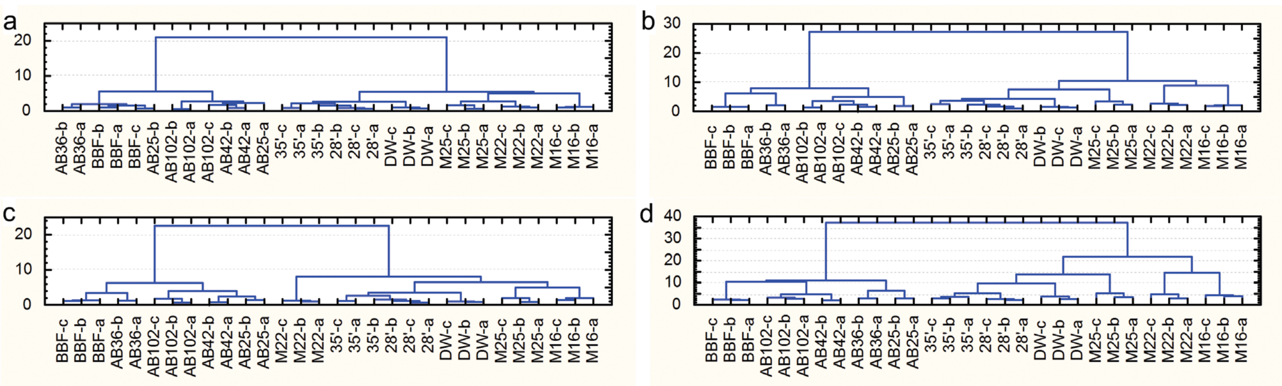

Fig. 5 Grouping of archaeological brick splits by geochemical clustering options CL&CB/8 (a), CL&CB/11 (b), CL&CB/8 (c) and CL&CB/16 (d)

Table 4 Summary of scores (*Osc*) according to the criterion for the recognition of brick splits of the same origin archaeological object

[a] C-way (method and distance)	*Osc*																[g] Sum of *Osc* (*C-way*)
	[b]/8				/11				/9				/16				
	[c]A	[d]B	[e]M	[f]S	A	B	M	S	A	B	M	S	A	B	M	S	
UA&E	0	1	1	2	2	1	1	5	2	2	0	4	2	2	1	5	16
UA&CB	0	1	1	2	2	1	1	5	0	2	0	2	2	2	1	5	13
WA&E	0	2	2	4	2	2	1	5	2	2	0	4	2	2	1	5	18
WA&CB	0	2	2	4	2	2	0	4	0	2	0	2	2	2	1	5	15
UC&E	0	2	0	2	2	1	1	5	2	2	0	4	2	2	0	4	14
UC&CB	0	2	1	3	2	2	0	4	0	2	0	2	2	2	0	4	13
WC&E	0	2	0	2	2	2	1	5	2	2	0	4	2	2	0	4	15
WC&CB	0	2	1	3	2	2	0	4	0	2	0	2	2	2	0	4	13
SL&E	0	1	0	1	2	1	1	5	0	2	0	2	2	2	0	4	11
SL&CB	0	1	0	1	2	1	0	3	0	2	0	2	2	2	0	4	10
CL&E	0	2	2	4	2	2	1	5	0	2	1	3	0	2	1	3	15
CL&CB	0	2	2	4	0	2	1	3	0	2	1	3	0	2	1	3	13
W&E	0	1	1	2	2	2	1	5	2	2	0	4	2	2	1	5	16
W&CB	0	2	2	4	2	2	1	5	0	2	2	4	2	2	2	6	19
[h] Sum of *Osc(n)*				38				59				42				61	

[a, b] Explanation is given in Tables 1, 2, 3.
[c, d, e] Sum of scores for the objects A, B, M.
[f] Sum of scores over all the 3 objects at selected n.
[g] Sum of scores over all the 3 objects and all n at selected C-way.
[h] Sum of scores over all the 14 C-ways at selected n.

Table 5 Summary of scores (*Csc)* according to the complex criterion for the recognition of brick splits of the same origin archaeological object with rated values for the "chaining effect"

[a] C-way (method and distance)	*Csc*																[g] Sum of *Csc* (C-way)
	[b]/8				/11				/9				/16				
	[c]A	[d]B	[e]M	[f]S	A	B	M	S	A	B	M	S	A	B	M	S	
UA&E	9	2	5	16	9	2	5	16	7	5	7	19	9	5	8	22	73
UA&CB	8	4	5	17	9	4	5	18	6	5	7	18	9	6	8	23	76
WA&E	3	5	8	16	9	4	8	21	7	5	6	18	9	5	8	22	77
WA&CB	3	4	9	16	8	5	7	20	7	5	7	19	9	6	8	23	78
UC&E	0	5	2	7	8	2	5	15	8	5	7	20	9	4	7	20	62
UC&CB	3	5	3	11	8	2	6	16	7	5	7	19	9	4	7	20	66
WC&E	0	5	1	6	7	4	7	18	8	5	7	20	9	4	7	20	64
WC&CB	4	5	3	12	8	4	7	19	7	5	7	19	9	4	7	20	70
SL&E	3	4	3	10	9	2	4	15	8	5	6	19	9	4	7	20	64
SL&CB	3	3	4	10	8	3	4	15	7	5	6	18	9	4	7	20	63
CL&E	2	5	9	16	9	4	8	21	7	5	8	20	7	6	7	20	77
CL&CB	0	5	9	14	5	5	8	18	8	5	8	21	7	6	8	21	74
W&E	-1	3	4	6	9	4	7	20	9	5	6	20	9	5	8	22	68
W&CB	0	5	9	14	8	5	8	21	7	5	9	21	9	6	9	24	80
[h] Sum of *Csc(n)*				171				253				271				297	

[a, b, c, d, e, f, g, h] Explanations are provided in Tables 1, 2, 3, 4.

tioned criteria, i.e. the brick-criterion (*Bsc*) and the object-criterion (*Osc*): 1) since there are much more bricks (22) than the main objects (3), the maximum possible *Bsc* (22) greatly exceeds the maximum possible *Osc* (6), which means that much more attention is given to bricks, than to objects; (2) scoring of "chains" in both criteria is too low (*Bsc* = 0 and *Osc* = 0). Therefore, complex criterion scores *(Csc)* were calculated for each of the 3 objects (Table 5). The sum of these values characterises the performance of the clustering way when the number of chemical elements is 8, 11, 9 and 16. The total of these sums *Csc (C-way)* gives the general characterisation of the clustering way. An important peculiarity of this criterion is that it gives a higher score when the "chaining effect" is observed.

For the origin-based clustering of brick splits of the main objects (A, B, M), the W&CB clustering way appeared to be the most suitable according to values of the sum of *Osc(C-way)* (Table 4), while SL&E and SL&CB clustering ways were found to be the least suitable. The evaluation of dendrograms according to complex criterion scores *Csc* shows that 2 highest values of the sum of *Csc*(*C-way*) are obtained by W&CB and WA&CB clustering ways (Table 5).

Geochemical description of Pašaminė and Rėvai clay quarries

The geological sections of Pašaminė and Rėvai glaciolacustrine clayey sediments with the distribution of chemical elements contents are presented in Fig. 6. According to lithology and character of lamination, the Pašaminė quarry clay section can be subdivided into 2 intervals (Fig. 6). The lower interval (7.1–4.5 m) is composed of varved sediments represented by thick layers of greyish brown or dark brown clay and thin layers of grey silt. The upper (4.5–0.55 m) interval consists of light yellow silt, massive clay with local convolute deformations and 2 horizons of loaded sand lenses. The Rėvai quarry clay section (Fig. 6) is characterised by various thickness layers of brown or dark brown clay and light brown silt with local convolute deformations in the upper part of the section.

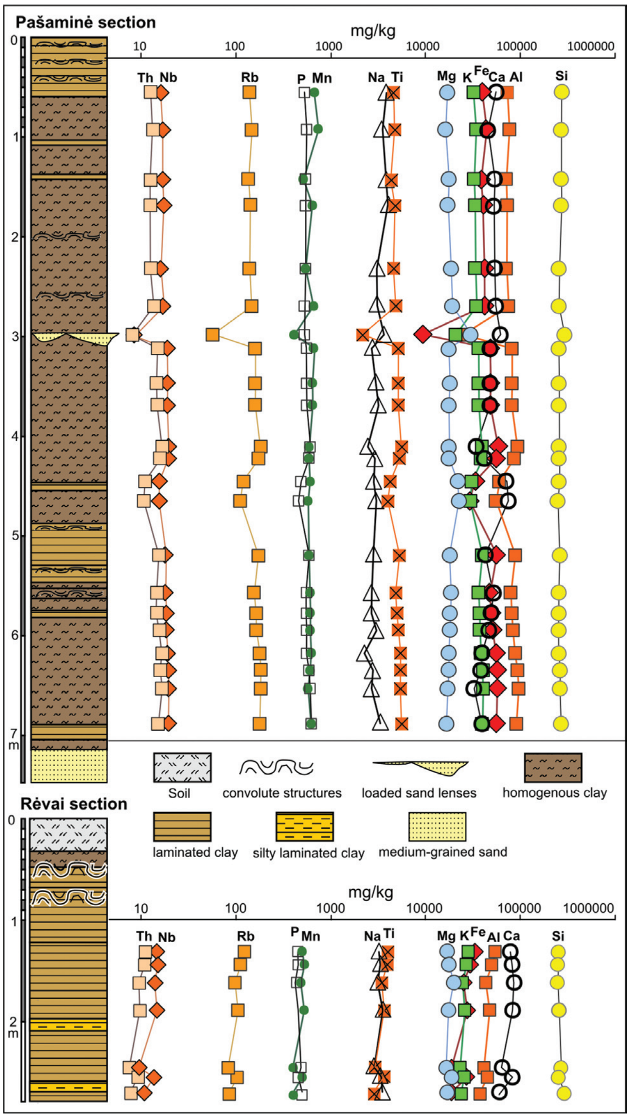

Fig. 6 Clayey sediment sections of Pašaminė and Rėvai quarries and the distribution of chemical elements contents

The distribution of chemical elements in the Pašaminė main section (P'Q-22) is rather stable, except for the sandy layer (at 2.99 m depth), which has high average content of Si and Mg, but much lower (than average) contents of Al, Ca, Fe, K, Ti, Mn, Rb, Nb and Th. A quite similar, but less expressed pattern can be seen in the layer of massive, but more silty clay at a depth of 4.4–4.6 m. The lateral inter-comparison of 4 short vertical sections (P'Q-a, P'Q-b, P'Q-c, P'Q-d, each being characterised by 3 samples, Fig. 2) did not reveal any significant ($p < 0.01$, Mann-Whitney U-test) differences in the contents of each selected element.

The distribution of chemical elements in the Rėvai section (R'Q-7) is rather stable with fluctuations observed in the lower part of the section (at a depth of 2.3–2.7 m, see Fig. 2).

The comparison of average elemental contents in Pašaminė clayey sediments with reference data (Table 6) shows that only contents of Al, Fe, Ti, Rb, Th are similar (within the percentage interval [80,120] of the reference values), as contents of Si, Ca, K, Mg, Nb are higher (>120%), while those of Na, P, Mn, Ba, Sr, and Ga are lower (<80%).

DISCUSSION

Statistical criteria and empirical scores

The discussion about choosing the proper clustering method started long ago and is still continuing without providing an unambiguous answer or a common agreement (Jain *et al.* 1986; D'haeseleer 2005), since the choice depends on the shape of the cluster, data spreading from its centre and the criteria used for the evaluation of dendrograms, e.g. cophenetic correlation coefficient, 2-norm-criterion (Carteron *et al.* 2012), the extent of retrieval or misclassification when data clusters are already known (Jain *et al.* 1986), etc. When generalizing our empirical scoring of dendrograms, the main attention was given to methods and only partial to distance measures. Our results of scoring clustering methods are in accordance with the early findings of statisticians as can be seen from 5 points listed below.

1. Despite the fact that the scores of complex criterion (*Csc*) were attributed treating the "chaining effect" as a rather positive feature, all our 3 empiri-

Table 6 The main geochemical statistical parameters of sediments from Pašaminė and Rėvai quarries and respective bricks, mg kg^{-1}

	a10th	b90th	cP'Q	a10th	b90th	dR'Q	eRV	fP'B-M	gP'B-H	hR'B-M	iR'B-H
	Pašaminė quarry			Rėvai quarry				Pašaminė bricks		Rėvai bricks	
Si	253918	271481	262618	254411	283276	265423	73000	299961	295131	332739	311087
Al	61710	93477	78032	40142	52399	45969	80000	87357	81197	68810	71465
Ca	38391	60953	49638	62543	84155	76704	22100	60185	55996	36332	46326
Fe	34067	56471	45755	18837	31157	25776	47200	45602	44019	35977	39465
K	31495	40154	35738	23702	27643	25603	26600	39308	37599	31386	32869
Mg	17065	21915	18805	16777	19328	17861	15000	23386	22375	15574	17412
Na	2654	3811	3065	2995	3492	3246	9600	4843	4110	4655	4002
Ti	4215	5393	4801	2866	3941	3479	4600	5014	4835	3874	4227
P	528	614	563	429	484	451	700	676	630	619	620
Mn	551	670	612	396	519	471	850	680	661	602	640
Ba	428	499	464	405	479	446	580	515	490	462	478
Rb	124	184	154	83	114	99	140	160	155	123	131
Sr	112	123	117	126	167	153	300	132	127	114	125
Ga	15.5	24.8	20.2	10.8	15.0	12.9	30	21.0	19.6	14.6	17.8
Nb	15.7	19.7	17.7	10.0	14.3	12.7	11	18.5	17.6	13.5	14.6
Th	11.0	16.4	14.0	7.4	10.6	9.05	12	14.5	14.1	10.8	11.7

[a, b] 10th and 90th percentile values in sediments of Pašaminė and Rėvai quarries, respectively.
[c, d] Average values in sediments of Pašaminė and Rėvai quarries, respectively.
[e] Reference values for continental clay reported by Turekian and Wedepohl (1961).
[f] Average values in monolithic bricks produced using clay from the Pašaminė quarry.
[g] Average value in hollow bricks produced using clay from the Pašaminė quarry.
[h] Average value in monolithic bricks produced using clay from the Rėvai quarry.
[i] Average value in hollow bricks produced using clay from the Rėvai quarry.

cal scores obtained for the CL-method exceed those given to the SL-method, which is most often characterised by the "chaining effect" in dendrograms when the structure of data is "smeared". This fact indicates that we prefer compact classes obtained by the CL-method (Kopp 1978 b) to the groups which appear as "branching and intertwined lines" often obtained using the SL-method (Kopp 1978 a). The latter method is suitable for "long-shaped classes" (Kopp 1978 c).

2. The ranges of *Bsc* and *Csc* values obtained by both "group-average strategy" methods (UA, WA) exceed the respective ranges obtained by both "centroid strategy" methods (UC, WC). This corresponds to the criticism of "centroid strategy" (Lance, Williams 1967; Kopp 1978 c; Fisher, Van Ness 1971), because the monotonicity (mathematical term) requirement is not always met by an algorithm, i.e. inversions appear.

3. Our scores for UA and WA methods are higher than for the SL-method, because these methods use all information about clusters (Čekanavičius, Murauskas 2004) and, thus, according to Kopp (1978 c), the "average linkage method" (UA and WA) "profits from stability" of SL and "the homogeneous property" of CL.

4. We gave preference for the WA-method over the UA-method because, when using weighting, both small and large groups are considered to the same extent, which is in line with Kopp's (1978 c) recommendations. This option is useful, because although the number of the same-brick splits (either archaeological or modern) is quite similar, higher-level clusters (either stonework or clay from quarries) include a very different number of samples.

5. Our selection of Ward's method as one of the 3 suitable ones is in line with Kopp's (1978 c) statement that this method is widely used in practice, e.g. it was applied by Gutsuz *et al.* (2017) to distinguish 4 clusters of clay samples from Amuq (Turkey) according to geochemical data. The two drawbacks to this method mentioned by Čekanavičius and Murauskas (2004) are either a merit (many small clusters in dendrograms) or can be easily overcome (sensitivity to outliers) as has been done by Gutsuz *et al.* (2017) for the elements which have large scattering or the contents of which are close to the detection limit. In our opinion, not only the detection limit but also the magnitude of RSD is important for the selection of elements. We recommend to estimate RSD values according to the results of duplicate samples and to select only the elements with low RSD (in our case, the limit value was 7.5%). This was the basis for the selection of 16 chemical elements in our research and showed successful fingerprinting results when scoring was based on all 16 analytes.

Performance of the 3 most suitable methods selected in our research is sometimes evaluated quite differently, e.g. according to the cophenetic correlation coefficient criterion using E-distance, Carteron *et al.* (2012) assigned the highest rating (1) to the performance of the UC-method, followed by that of the UA-method (2), meanwhile performance of the CL-method (6) as well as that of the W-method (7) was rated the lowest. Similar findings were obtained using the same criterion and CB-distance by Saracli *et al.* (2013). This can be explained by differences in criteria and the data used.

Sourcing clay for bricks

Taking into account the highest sums of scores for different criteria (Tables 3–5), the following 3 methods of amalgamation (W, CL, WA) using either Euclidean or City-block distances were selected as possible clustering ways for fingerprinting sources of clay for modern bricks. The geochemical clustering of modern brick samples and clay from 2 known quarries was analogous to that used for splits of archaeological bricks, i.e. the number of chemical elements used in cluster analysis was 8, 9, 11 and 16 (all selected elements).

The analysis of the structure of geochemical clustering options presented in 24 dendrograms (Figures in Supplements, Figs. S5–S6) showed that in its branches, none of the tree diagrams demonstrates "pure clusters", which include only "related" brick or clay samples or at least shows their amalgamation as a "chain" (Table 7). Several (8) most acceptable geochemical clustering options can be arranged as follows (see Figures in Supplements, Figs. S5, S6): WA&E/8 (Fig. S5e), WA&E/9 (Fig. S5k) > CL&E/9 (Fig. S5i), CL&CB/9 (Fig. S5j) > W&E/9 (Fig. S5g) > CL&E/16 (Fig. S6i), CL&CB/16 (Fig. S6j)> WA&CB/16 (Fig. S6l), in the last 3 options, the amalgamation is as a "chain". However, 3 samples from the Pašaminė quarry (3PQ in Table 7) always go together with 4 bricks produced from Rėvai clay (4RB) and 7 samples from Rėvai quarry (7RQ in Table 7).

These 3 samples are anomalous among all 22 samples from the Pašaminė section. Sample P'Q-2.99 is the most sandy, it is distinguished by the highest content of Si and by the lowest contents of clay indicators Nb, Ti, Fe, Th, K, Ga, Rb, Al, besides, also of Ba, Mn and Sr. The peculiarity of the other 2 samples, i.e. P'Q-4.46 and P'Q-4.63, is that the contents of Ca and Mg exceed the 90th percentile levels, while the contents of Nb, Ti, Fe, Th, K, Ga, Rb, Al, Si, P, Sr, Ba are below the 10th percentile (Fig. 6, Table 6). Therefore, it is quite clear that in all dendrograms, these 3 samples with quite different geochemical composition moved from the group of the Pašaminė object to the dendrogram branch with Rėvai object samples. If attention is not paid to these 3 samples, the most proper geochemical clustering options for amalgamation of the related object samples are obtained when Si is added to 8 clay proxies (4 cases from 8, see Table 7). It is quite possible that this fact is not accidental, because Si is not merely one of the elements in the chemical composition of clayey sediments, but also the element

Table 7 Generalized schemes of the dendrograms obtained using geochemical clustering options for the grouping of brick and clay samples related to Rėvai and Pašaminė objects

Method	Euclidean distance (E)	City-block distance (CB)
	Nb, Ti, Fe, Th, K, Ga, Rb, Al, [a] /8	
W	[b]{7RQ, 2RB, 2RB, 6PQ}-{16PQ, 4PB, 6PB}	{7RQ, 2RB, 2RB, 6PQ}-{16PQ, 4PB, 6PB}
CL	{7RQ, 2RB, 3PQ}-{7PQ-[[2RB, 3PQ]-[9PQ, 4PB, 6PB]}	{2RQ, 1PQ}-{[5RQ, 4RB, 8PQ]-[13PQ,4PB,6PB]}
WA	{**7RQ**, **4RB,** 3PQ}-{19PQ, 4PB, 6PB}	{7RQ, 2RB, 3PQ}-{[6PQ, 2RB]-[13PQ, 4PB, 6PB]}
	Nb, Ti, Fe, Th, K, Ga, Rb, Al and Si, /9	
W	{**4RB**, **7RQ,** 3PQ}-{19PQ, 10PB}	{7RQ, 4RB, 9PQ}-{13PQ, 10PB}
CL	{**4RB**,**7RQ**, 3PQ}-{19PQ, 10PB}	{**4RB**,**7RQ**, 3PQ}-{19PQ, 10PB}
WA	{**4RB**, **7RQ**, 3PQ}-{19PQ, 10PB}	{2RQ, 1PQ}-{[5RQ, 4RB, 2PQ]-[19PQ, 10PB]}
	Nb, Ti, Fe, Th, K, Ga, Rb, Al, Si, Ba and Mn, /11	
W	{7RQ, 4RB, 5PQ}-{17PQ, 10PB}	{7RQ, 4RB, 10PQ}-{12PQ, 10PB}
CL	{7RQ, 3PQ}-{[4RB, 12PQ]-[7PQ, 10PB]}	{7RQ, 3PQ}-{[4RB, 7PQ]-[12PQ, 10PB]}
WA	{7RQ, 3PQ}-{7PQ-[[4RB, 2PQ]-[10PQ, 10PB]]}	{2RQ, 1PQ}-{[5RQ, 8PQ, 4RB]-[13PQ, 10PB]}
	Nb, Ti, Fe, Th, K, Ga, Rb, Al, Si, Ba, Mn, Sr, Ca, Mg, P and Na, /16	
W	{7RQ, 1PQ}-{13PQ-[[4RB,8PQ]-10PB]}	{7RQ, 3PQ}-{14PQ-[4RB, 5PQ, 10PB]}
CL	{**7RQ**,3PQ}-{**4RB-[**19PQ, 10PB]}	{**7RQ**, 3PQ}-{**4RB**-[13PQ, 10PB]}
WA	{7RQ, 1PQ}-{7PQ-[14PQ-[4RB, 10PB]]}	{**2RQ,** 1PQ}-{[**5RQ,** 2PQ]-[**4RB**-[19PQ, 10PB]]}

[a] Number of variables (elements).

[b] The codes used for the Pašaminė object (are PB brick produced from Pašaminė clay) and PQ (clay from Pašaminė quarry). The codes used for the Rėvai object are RB (a brick produced from Rėvai clay) and RQ (clay from Rėvai quarry). The number before the code indicates how many samples were amalgamated. Curley brackets indicate the main two (highest-level) branches of the dendrograms, while box brackets show some additional information about clustering within these branches, but only in those cases when Rėvai object samples are not in the same main branch. The part of the dendrogram scheme containing samples of the Rėvai object is emboldened if the respective option of geochemical clustering provides the most acceptable variant.

playing an increasing role when clay is tempered (by adding sand) during the mixture preparation process for brick production (Kizinievič 2006). Cohen *et al.* (2019) stated that it can be difficult to compare the chemical composition of raw clays and archaeological ceramics using the technique of instrumental neutron activation analysis. It is understandable, because this analytical method is not suitable for Si content detection. This case is important from methodological point of view as it demonstrates the importance of the proper selection of variables. It seems to confirm the ideas of Owens *et al.* (2016) about using different tracer properties and their combinations.

Totally different results were obtained when the average elemental contents characterising 3 subsets (R'B-H, R'B-M, R'Q-7) related to the Rėvai object and 8 subsets related to that of Pašaminė (P'B-H, P'B-M, P'Q-22, P'Q_6, P'Q-a, P'Q-b, P'Q-c, P'Q-d) were used (Figures in Supplements, Figs. S7, S8).

The methodology of scoring is analogous to the one used for compiling Table 5, and the respective summary is provided in Table 8. It should be emphasised that unlike in the first case (when 43 samples were amalgamated), the clustering results of these 11 subsets were more successful, i. e. the related subsets appeared to be in the same branch of dendrograms or at least showed their amalgamation as a "chain". When the number of geochemical variables was the lowest (8), none of the clustering ways provided successful results. The reason for that is the average value in a P'Q_6 subset, which moves to the Rėvai object. This subset is distinguished by relatively lower average contents of clay proxies (markers), i.e. Nb, Ti, Fe, Th, K, Ga, Rb and Al than other subsets characterising Pašaminė quarry clayey sediments. It can be explained by more expressed silt and clay lamination with sand lenses in the upper part of the Pašaminė section than that in the lower part. It suggests the importance of selecting a sufficient number of samples for proper characterisation of the study object. This would enable elimination of anomalous (accidental) samples before cluster analysis, because they can distort dendrograms. If samples are taken accidentally or non-professionally, i.e. without taking into account the lithological characteristics of the study object, the non-typical (anomalous) samples can move to the cluster with samples of quite different composition, and, in such a case, wrong conclusions can be drawn. The clustering results of 43 single samples obviously demonstrate such a possibility. A greater sample size containing at least 8 specimens from each object (Molenaar *et al.* 2018) should be the main prerequisite for proper sourcing (fingerprinting).

Geochemical signatures of present-day bricks and clay in two quarries

U-test has revealed (Table 9) that clay from the Rėvai quarry has significantly higher contents of elements related to carbonates, while that from the Pašaminė quarry higher contents of elements related to clay. These differences are seen in the relative average values (Fig. 7). Bricks produced from clay of 2 quarries also differ significantly.

Of course, there are geochemical differences between bricks and the clay used for their production due to the technological processes applied. These differences may hinder proper clustering. But geochemical signatures of Pašaminė clay and bricks made from it are rather similar (Fig. 7), meanwhile Rėvai clay and bricks made from it differ more, especially according to Ca content.

When only clay proxies are used (Table 8), the performance of geochemical clustering is not very good, i.e. *Csc* (8) = 30, but it greatly improves, when more elements, first of all Si, are added. When all 16 chemical elements are used, the performance of geochemical clustering becomes slightly poorer, i.e. *Csc* (16) = 57. This fact confirms the idea presented by Owens *et al.* (2016) about the choice of suitable fingerprints.

Table 8 Summary of scores (*Csc*) according to the complex criterion for recognizing Rėvai and Pašaminė objects

[a] C-way (method and distance)	*Csc*												Sum of *Csc* (C-way)
	[b] /8			/9			/11			/16			
	[c] R	[d] P	[e] S	R	P	S	R	P	S	R	P	S	
W&E	2	3	5	3	8	11	3	8	11	1	7	8	35
W&CB	2	3	5	2	3	5	3	8	11	3	8	11	32
CL&E	2	3	5	3	8	11	3	8	11	3	8	11	38
CL&CB	2	3	5	3	8	11	3	8	11	2	3	5	32
WA&E	2	3	5	2	8	10	3	8	11	3	8	11	37
W&CB	2	3	5	2	8	10	2	3	5	3	8	11	31
Csc(n)			30			58			60			57	

[a, b] Abbreviations are given in Tables 2, 3.
[c] Sum of scores for Rėvai object subsets (clay and bricks).
[d] Sum of scores for Pašaminė object subsets (clay and bricks).
[e] Sum of scores for both objects.

Table 9. Significant ($p < 0.01$) differences between elemental contents in raw clay and brick objects

Objects compared	U-test results and respective ratios of average values
P'Q vs R'Q	[a] P'Q (10): Fe(1.8)> Al(1.7)> Rb, Ga(1.6)> Th(1.5)> K, Nb, Ti (1.4)> Mn(1.3)> P(1.2) R'Q (2): Ca(1.6)> Sr(1.3)
P'B vs R'B	P'B (13): Ca, Mg (1.4) > Nb, Th, Ga (1.3) > Rb, Ti, Al, K, Fe (1.2) > Sr, Mn, Ba (1.1)
P'Q vs P'B	P'Q (8): Na (1.5), Mg, Ca, P (1.2), Si, Sr, Mn, Ba (1.1)
R'Q vs R'B	R'Q (8): Al, Fe (1.5), P (1.4), Na, Mn, Rb (1.3), K, Si (1.2) R'B (1): Ca(1.9)

[a] One of the objects (for abbreviations see Table 1) compared by U-test, the number of elements, the contents of which are significantly higher than in the other object, is given in parentheses. The elements are arranged in descending sequence according to the ratios of their respective average values.

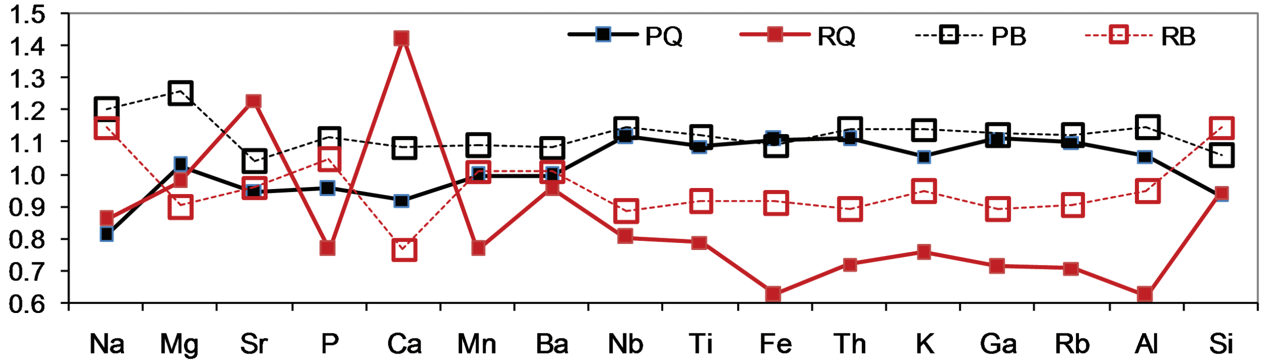

Fig. 7 The relative average values of chemical elements in Pašaminė and Rėvai objects: clay and bricks (P'Q, R'Q and P'B, R'B, respectively). Each average value was divided by their median value

Comparative analysis of the geochemical composition of clay deposits from Lithuania and other countries

The comparative examination of arithmetic averages of clay deposits from recent studies of various countries and geological formations within the 10%–90% percentile range of chemical elements obtained from chemical analyses of clay in Pašaminė and Rėvai quarries was carried out (Table 10).

The average values (AV) of chemical elements in clay deposits of various age and geological formations, obtained from recent studies, vary within a very wide range: 35% of AV (169) are lower than 10th percentile, 33% of AV (163) are within the range of 10th–90th percentiles, and 32% of AV are higher than 90th percentile of our investigated clay sections. Clay deposits from various countries show: (a) similar contents (more than 67% of AV are within the range of 10th–90th percentiles) for Al (78% of AV), Fe (78%), Ti (67%); (b) lower values (more than 58% of AV are lower than 10th percentile) for Si (73%), K (97%), Ba (89%), Rb (58%) and (c) higher values (more than 54% of AV are higher than 90th percentile) for Ca (53% of AV), Mg (67%), P (96%), Mn (57%) and Sr (54%). The comparative analysis of the chemical composition of clay deposits from different countries and that of the clay quarries under our investigation reveals some similarities and differences in the chemical composition (Tables 6, 10). The above-mentioned differences result from the interaction of complex factors, such as weathering, topography, sedimentation environment, and anthropogenic influences (e.g., land use). All of these factors generally lead to stratigraphic and spatial variations in sedimentary succession and consequently differ in their chemical compositions (Gutsuz *et al.* 2017; Wronkiewicz, Condie 1987; Condie *et al.* 1992; Garver, Scott 1995). In Lithuania, the chemical composition of glaciolacustrine clay was influenced by different age Pleistocene tills, whose chemical composition varies with location of till beds. Clay from the investigated Rėvai quarry is more calcareous than that from the Pašaminė quarry (Table 6), although both clay deposits were formed during the same Late Nemunas (Late Weichselian) Glaciation retreat from Lithuania. The investigated sections were not homogeneous: laminated intervals of clay and silt layers in the lower part of the section and sand lenses at a depth of 3 m were observed in the vertical section of the Pašaminė quarry. Therefore, it is very important that the samples collected for geochemical investigations be representative and those that are not characteristic of the clay bed (in our case, the sample from the sand lenses at a depth of 2.99 m, see Fig. 6) be eliminated. Our research suggests that it is necessary to collect a sufficient number of clay samples for geochemical analysis and focus on analysing average values of

Table 10 Scores of elemental average levels of some clay deposits in comparison with 10^{th}–90^{th} data of the present study

	[a] 10^{th}–90^{th}	[b] 1 [j] n9	[c] 2 n7	[d] 3 n15	[e] 4 n2	[f] 5 n1	[g] 6 n1	[h] 7 n1	Sum of scores (of all clay deposits)		
		Scoring of given elemental average values compared to the extreme values of 10^{th}–90^{th} percentile ranges of our data							[k] -	[l] ~	[m] +
Si	253918–283276	-2 ~ 7	-6 + 1	-13 ~ 2	[n] nd	-1	-1	-1	24	9	0
Al	40142–93477	~5 + 4	~7	-1 ~ 14	+2	+1	~1	~1	1	28	7
Ca	38391–84155	-2~2 + 5	-1 + 6	-2 ~ 7	-2	+1	~1	+1	7	10	1
Fe	18837–56471	~7 + 2	~6 + 1	~13 + 2	+2	+1	~1	~1	0	28	8
K	23702–40154	-9	-7	-15	-2	-1	~1	-1	35	1	0
Mg	16777–21915	-6 ~ 3	+7	-2 + 13	nd	-1	-1	-1	11	3	20
Na	2654–3811	-2~1 + 6	-5 + 2	-3 + 12	+2	+1	-1	+1	11	1	24
Ti	2866–5393	~4 + 5	-1 ~ 6	-1 ~ 12 + 2	+2	~1	-1	~1	3	24	9
P	429–614	+9	+7	+15	nd	+1	nd	-1	1	0	23
Mn	396–670	-1 ~ 3 + 5	~1 + 6	-2 ~ 7 + 6	+2	+1	nd	~1	3	12	20
Ba	405–499	-9	-6 ~ 1	-15	+2	+1	nd	-1	31	1	3
Rb	83–184	-5 + 4	-4	-5 ~ 10	-2	-1	nd	-1	18	10	3
Sr	112–167	+9	+7	-5 ~ 10	~1+1	+1	nd	+1	5	11	19
Ga	10.8–24.8	nd	-3 ~ 4	nd	nd	~1	nd	~1	3	6	0
Nb	10.0–19.7	nd	-3 ~ 4	nd	nd	-1	nd	~1	4	5	0
Th	7.4–16.4	nd	-7	-5 ~ 10	~2	~1	nd	~1	12	14	0

[a] Extreme values of 10^{th} and 90^{th} percentile ranges in sediments of Pašaminė or Rėvai quarries.
[b,c,d,e, f, g.h] The references to average values of clay deposits, given by: [b] 1 – Montana *et al*. (2011) for Sicily, [c] 2 – Gutsuz *et al.* (2017) for Turkey, [d] 3 – Hein *et al.* (2004) for Crete, [e] 4 – Cohen *et al.* (2019) for Mexico, [f] 5 – Owen *et al.* (2013) for Ghana, [g] 6 – Sedmale *et al.* (2017) for Latvia, [h] 7 – Christidis *et al.* (2014) for Aegina island, Greece.
[j] The number of described clay deposits.
[k] The number of clay deposits with average values lower than 10^{th} percentile.
[l] The number of clay deposits with average values within the range of 10^{th} and 90^{th} percentiles.
[m] The number of clay deposits with average values higher than 90^{th} percentile.
[n] No data.

their chemical composition, thus eliminating random samples that are not typical of that clay bed.

CONCLUSIONS

The study on suitability of 14 clustering ways (7 methods, each tested with 2 distances) combined with selected sets of geochemically related chemical elements by attributing empirical scores to dendrograms shows optimistic possibilities for the cluster analysis application in recognizing archaeological brick splits and in fingerprinting (tracing) sources of clay as a raw material for production of modern bricks.

Our study revealed that Complete Linkage, Weighted Pair Group Average and Ward's clustering methods using a City-block distance measure are the most efficient in recognizing, i.e. "piecing together", splits of the same origin archaeological brick. The highest percentage of correctly recognized samples was obtained using all the 16 chemical elements studied, i.e. Al, Ba, Ca, Fe, Ga, K, Mg, Mn, Na, Nb, P, Rb, Si, Sr, Th and Ti. The aforementioned 3 methods using both City-block and Euclidean distances have probably the best fingerprinting properties for the recognition of objects, i.e. stonework, as the scoring of dendrograms has shown.

The study of modern bricks and their raw clay sources (from quarries) has revealed that values of average element contents in objects, which are compared according to geochemical composition, have much more useful fingerprinting properties than contents of elements in single samples. Therefore, for the purpose of raw clay source fingerprinting, it is insistently recommended not only to collect such samples that represent the lithological-mineralogical composition of each object well (it is an essential prerequisite), but also to increase the number of collected samples so that it would be possible to analyse the statistical homogeneity of the respective dataset and to remove possible outliers.

The two main dialectically related peculiarities of the geochemical composition of the objects to be linked (either different ancient buildings or bricks and clay from quarries as their raw material) predetermine the successful selection of geochemical clustering tools for fingerprinting: their similarity, which is influenced by such clay proxies as Al, Rb, Ga, K, Th, Fe, Ti and Nb, and their differences, which depend on e.g. the elements more abundant in sands and silts (Si, Na), carbonates (Ca, Sr, Mg), organic matter (P) and other lithological-mineralogical tracers. These peculiarities are the key to solving the problem. It is insufficient to use only clay proxies. Datasets with all major elements (with obligatory inclusion of Si) and the number of trace elements as large as possible should be accumulated as geochemical reference data both for archaeological artefacts and clay deposits as their possible source.

The elements with high values of the analytical relative standard deviation (higher than 7 or 10%) should be highlighted. Selection of appropriate equipment and sample preparation method is and will be a key success factor in fingerprinting.

The performed study of Rėvai and Pašaminė quarries has shown that even rather similar geological age clay deposits have their own specific geochemical features. Moreover, their comparison with clay deposits of different geological formations from other countries revealed that all clay deposits show some differences in the geochemical composition. This essential property of differences in chemical composition of clay deposits is useful for fingerprinting (sourcing) clay for bricks by means of clustering objects. In the future, it would be useful to analyse not only chemical element contents, but also their ratios (such as Si/Al, Si/K, Ca/Mg, Ca/Si, Mg/Si, Ca/Sr, Al/K, K/Rb, K/Th, Fe/Ti, Ti/Nb, etc.) by selecting mutually correlating or antagonistic pairs of indicators.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for essential remarks and useful recommendations.

REFERENCES

Bayne, C.K., Beauchamp, J.J., Begovich, C.L., Kane, V.E. 1980. Monte Carlo comparisons for selected clustering procedures. *Pattern Recognition 12*, 51–62.

Buxeda i Garrigos, J., Kilikoglou, V., Day, P.M. 2001. Chemical and mineralogical alteration of ceramics from a Late Bronze Age kiln at Kommos, Crete: the effect on the formation of a reference group. Archaeometry 43 (3), 349–371.

Carteron, A., Jeanmougin, M., Leprieur, F., Spatharis, S. 2012. *Ecological Informatics 9*, 64–68.

Christidis, G.E., Shriner, C.M., Murray, H.H. 2014. An integrated methodological approach for source-clay determination of ancient ceramics: the case of Aegina Island, Greece. *Clay and Clay Minerals 62*, 447–469.

Cohen, A.S., Pierce, D.E., Fisher, C.T. 2019. Geochemical analysis and spatial trends of ceramics and clay from Angamuco, Michoacán. *Journal of Archaeological Science: Reports 23,* 216–230.

Condie, K.C., Noll, P.D., Conway, C.M. 1992. Geochemical and detrital mode evidence for two sources of Early proterozoic sedimentary rocks from the Tonto Basin Supergroup, central Arizona. *Sedimentary Geology 77,* 51–76.

Čekanavičius, V., Murauskas, G. 2004. *Statistika ir jos taikymai. II.* TEV, Vilnius, 272 pp. [In Lithuanian].

D'haeseleer, P. 2005. How does gene expression clustering work? *Nature Biotechnology 23 (12),* 1499–1501.

Dubes, R., Jain, A.K. 1979. Validity studies in clustering methodology. *Pattern Recognition 11*, 235–254.

Fisher, L., Van Ness, J. W. 1971. Admissible clustering procedures, *Biometrika 58*, 91–104.

Garver, J.I., Scott, T.J. 1995. Trace elements in shale as indicators of crustal provenance and terrane accretion in the southern Canadian Cordillera. *Geological Society of America Bulletin 107,* 440–453.

Glascock, M.D., Neff, H., Vaughn, K.J. 2004. Instrumental neutron activation analysis and multivariate statistics for pottery provenance. *Hyperfine Interactions 154,* 95–105.

Gonzales, D.A., Arakawa, F., Koenig, A. 2015. Petrographic and geochemical constraints on the provenance of sanidine-bearing temper in ceramic potsherds, Four Corners Region, Southwest USA. *Geoarchaeology 30,* 59–73.

Guobytė, R. 1999. Quaternary Geological Map of Lithuania. Geological Survey of Lithuania, Vilnius.

Gutsuz, P., Kibaroğlu, M., Sunal, G., Hacıosmanoğlu, S. 2017. Geochemical characterization of clay deposits in the Amuq Valley (Southern Turkey) and the implications for archaeometric study of ancient ceramics. *Applied Clay Science 141*, 316–333.

Hall, M.E. 2017. X-ray fluorescence-energy dispersive (ED-XRF) and wavelength dispersive (WD-XRF) spectrometry. In: A. M. W. Hunt (ed.). *The Oxford handbook of archaeological ceramic analysis.* Oxford University Press, Oxford, 342–362.

Hein, A., Mommsen, H., Zender, G. 2004. Pliocene clays from Aegina (Greece): Reference material for chemical provenance studies on Bronze Age pottery from the Island. *Geoarchaeology: An International Journal 19,* 553–564.

Jain, N.C., Indrayan, A., Goel, L.R. 1986. Monte Carlo comparison of six hierarchical clustering methods on random data. *Pattern Recognition 19 (1)*, 95–99.

Jaskowiak, P.A., Campello, R.J.G.B., Costa, I.G. 2014. On the selection of appropriate distances for gene expression data clustering. *BMC Bioinformatics 15 (Suppl 2)*, 1–17.

Kitkauskas, N. 2009. *Lietuvos Didžiosios Kunigaikštystės valdovų rūmai.* „Kultūros“ leidykla, Vilnius, 319 pp. [In Lithuanian].

Kizinievič, O. 2006. *Pramoninių atliekų panaudojimo statybinėje keramikoje tyrimai.* Doctoral thesis. Vilniaus Gedimino technikos universitetas, Vilnius, 99 pp. [In Lithuanian].

Kopp, B. 1978a. Hierarchical classification. I. Single linkage method. *Biometrical Journal 20 (5)*, 495–501.

Kopp, B. 1978b. Hierarchical classification. II. Complete linkage method. *Biometrical Journal 20 (6*), 597–602.

Kopp, B. 1978c. Hierarchical classification. III. Average linkage, median, centroid, WARD, flexible strategy. *Biometrical Journal 20*, 703–711.

Kuiper, K.K., Fisher, L. 1975. A Monte Carlo comparison of six clustering procedures. *Biometrics 31*, 777–783.

Lance, G.N. and Williams, W.T. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computers Journal 9*, 373–380.

Levandauskas, V. 2012. *Lietuvos mūro istorija.* Vytauto Didžiojo universiteto leidykla, Kaunas, 456 pp.

Molenaar, J., Cofino, W.P., Torfs, P.J.J.F. 2018. Efficient and robust analysis of interlaboratory studies. *Chemometrics and Intelligent Laboratory Systems 175,* 64–73.

Montana, G. 2017. Ceramic raw materials. In: A.M.W. Hunt (ed.). *The Oxford handbook of archaeological ceramic analysis*. Oxford University Press, Oxford, 87–100.

Montana, G., Cau, Ontiveros, M.Á., Polito, A.M., Azzaro, E. 2011. Characterisation of clayey raw materials for ceramic manufacture in ancient Sicily. *Applied Clay Science 53,* 476–488.

Owen, J.V., Casey, J.L., Greenough, J.D., Godfrey-Smith, D. 2013. Mineralogical and geochemical constraints on the sediment sources of Late Stone Age pottery from the Birimi Site, Northern Ghana. *Geoarchaeology 28,* 394–411.

Owens, P.N., Blake, W.H., Gaspar, L., Gateuille, D., Koiter, A.J., Lobb, D.A., Petticrew, E.L., Reiffarth, D.G., Smith, H.G., Woodward, J.C. 2016. Fingerprinting and tracing the sources of soils and sediments: Earth and ocean science, geoarchaeological, forensic, and human health applications. *Earth-Science Reviews 162*, 1–23.

Pollard, A.M., Heron, C. 1996. *Archaeological Chemistry.* RSC, Cambridge. 376 pp.

Rand, W.W. 1971. Objective criteria for evaluation of clustering methods. *Journal of American Statistical Association 66,* 846–850.

Rudnick, R.L., Gao S. 2003. Composition of the continental crust. In: R.L. Rudnick (ed.). *The Crust, vol. 3.* Elsevier, pp. 1–64.

Saracli, S., Doğan, N., Doğan, U. 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications 203*, 1–8.

Sarcevičius, S., Taraškevičius, R. 2015. Vilniaus Šv. Mikalojaus bažnyčios plytos – nuo jų ženklinimų iki geocheminių sąsajų [Bricks of St. Nicholas church: link between marking and geochemistry]. *Archaeologia Lithuana 16,* 49–66 [In Lithuanian].

Sedmale, G., Randers, M., Rundans, M., Seglins, V. 2017. Application of differently treated illite and illite clay samples for the development of ceramics. *Applied Clay Science 146,* 397–403.

Sokal, R.R., Rohlf, F. 1962. The comparison of dendrograms by objective methods. *Taxon 11 (2)*, 33–40.

Sokal, R.R., Sneath, P.H.A. 1963. *Principles of Numerical Taxonomy.* W.H. Freeman, San Francisco, 359 pp.

Taraškevičius, R., Bliujienė, A., Karmaza, B., Merkevičius, A., Nemickienė, R., Rackevičius, G., Sarcevičius, S., Stakėnienė, R., Strazdas, D., Širvydaitė, S., Vaitkevičius, G., Zinkutė, R. 2013. Geochemical methods of investigation in archaeology – possibilities of application. In: A. Merkevičius (ed.). *Metodai Lietuvos archeologijoje. Mokslas ir technologijos praeičiai pažinti.* Vilniaus Universiteto leidykla, Vilnius, 249–304 [In Lithuanian].

Taraškevičius, R., Zinkutė, R., Nabažaitė, R. 2016. Potentially harmful elements in breakages of glazed tiles from three sites in Lithuania. In: *Abstract Book of the 18th International Conference on Heavy Metals in the Environment, 12 to 15 September 2016, Ghent, Belgium.* 667–668.

Taraškevičius, R., Zinkutė, R., Radzevičius, A., Strazdas, D. 2007. Chemical composition of tile glaze as potential source of soil contamination. In: *Environment, Technology, Resources: Proceedings of the 8th International Scientific and Practical Conference, June 20–22, 2007, Rezekne, Latvia*. *Vol. 1.* Rezekne, 175–181.

Turekian, K.K., Wedepohl, K.H. 1961. Distribution of the elements in some major units of the Earth's crust. *Geological Society of America Bulletin 72*, 175–192.

Uchman, A., Kazakauskas, V., Gaigalas, A. 2009. Trace fossils from Late Pleistocene lacustrine varve sediments in eastern Lithuania. *Palaeogeography, Palaeoclimatology, Palaeoecology 272*, 199–211.

Uchman, A., Gaigalas, A., Kazakauskas, V. 2008. Trace fossils from the Upper Pleistocene glaciolacustrine laminated sediments of Lithuania. *Geologija 50*, 212–226.

Van Ness, J.W. 1973. Admissible clustering procedures. *Biometrika 60*, 422–424.

Waksman, Y. 2017. Provenance studies. Productions and compositional groups. In: Hunt, A.M.W. (ed.). *The Oxford handbook of archaeological ceramic analysis*. Oxford University Press, Oxford, 148–161.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *Journal of Americal Statistical Association 58 (301)*, 236–244.

Wedepohl, K.H. 1995. The composition of the continental crust. *Geochimica et Cosmochimica Acta 59,* 1217–1232.

Weigand, P.C., Harbottle, G., Sayre, E.V. 1977. Turquoise sources and source analysis: Mesoamerica and the Southwestern U.S.A. In: Earle, T.K., Ericson, J.E. (eds.), *Exchange Systems in Prehistory.* Academic Press, New York, pp. 15–34.

Wishart, D. 1969. An algorithm for hierarchical classification. *Biometrics 25,* 165–170.

Wronkiewicz, D.J., Condie, K.C. 1987. Geochemistry of Archean shales from the Witwaterstrand Supergroup, South Africa: source area weathering and provenance. *Geochimica et Cosmochimica Acta 51,* 2401–2416.

Supported Online Material
Figures S1– S8