Determination of cultural areas based on medieval pottery using an original divisive hierarchical clustering method with geographical constraint (MapClust)

Lise Bellanger, Arthur Coulon, Philippe Husi

Determination of cultural areas **based on** medieval pottery using an **original divisive** hierarchical clustering method with geographical constraint (MapClust)

Lise Bellanger[1], Arthur Coulon[2] and Philippe Husi[2]

Abstract

Spatial representation based on the distribution of artefact, especially pottery, is widely used in archaeology. The questions raised are often related for the construction of economic or cultural areas, based on fabrics, style or types of vessels. The middle Loire Valley with the study of an important corpus of medieval pottery constitutes an essential framework for understanding this mechanisms over long time. The corpus of data collected and studied for over twenty years includes archaeological sites with reliable chrono-stratigraphic sets whose pottery assemblages are sufficiently large to respond to the problem posed.
The permanent increase of the corpus of data, especially since the last publication on this subject in 2013 for which the approach was purely empirical, requires the development of statistical tools adapted here to the spatial clustering analysis of the data. This article deals with a divisive hierarchical clustering method with geographical constraints: MapClust. It is based on spatial indicators called spatial patches. This statistical tool allows the archaeologist to visualize and analyse geographical patterns easily. The comparison of the present results with those of 2013 is a way to see the relevance of the method.

Keywords: archaeology, pottery, cultural areas, middle Loire Valley, divisive hierarchical clustering, geographical constraints

[1] Université de Nantes Laboratoire de Mathématiques Jean Leray – UMR CNRS 6629, 2 rue de la Houssinière BP 92208, 44322 Nantes Cedex 03, France

[2] CNRS/Université de Tours, UMR 7324 CITERES, Laboratoire Archéologie et Territoires, 40 rue James Watt, ActiCampus 1, 37200 Tours, France

1 Introduction

The objective of an exhaustive study of pottery is to answer socio-economic and cultural questions from one of the major archaeological sources for the study of domestic life in pre-industrial periods. Spatialization of archaeological material like pottery is widely used in archaeology. The questions raised are often related to the distribution of products from the workshops, to the supply of consumer centres, to the spatial distribution of particular types of vessels for the construction of economic or cultural areas. These illustrative maps often remain very descriptive in the presentation of distribution by intensity or size of points, arrows linking sites, juxtaposed areas; drawn according to the knowledge and feeling of the archaeologist. This mode of representation has been frequently used since the 1970s, a period when data were still few in number and from a limited number of sites (Orton, 1969; Hodger, 1981). The development of preventive (rescue) archaeology since the 1990s and the mass of accumulated data has largely changed previous historical interpretations (Husi, 2010). This calls for the development of reflection on new analytical tools at the interface between statistics and geomatics. Progress in particular in cartographic tools has made it possible to improve the quality of maps without abandoning an often largely descriptive approach to spatialized data (Vince, 1993; Delage, 1998; Husi, 2003, Legoux, 2010; Husi, 2013; Chabert 2019; Allan, 2020; Henigfeld; 2021 to be published). The aim here is not to question the results and interpretations of such studies, but to attempt to overcome some of the limits of the spatial approach by adopting a statistical approach based on data analysis. At the same time, data analysis methods (clustering methods; factor analysis) have been widely used and for a long time in archaeometry, especially - but not only - in ceramology for the chemical and petrographic characterisation of production (Madsen, 1988; Tyers and Orton, 1991; Baxter, 2006; Vince, 2011; Holmqvist et al., 2018; Dervin and Bocquet, 2020; Dilmann et al. 2020).

Cluster analysis allows the construction of groups or clusters of homogeneous observations in a data corpus. A wide range of numerical approaches exists to detect clusters (Kaufman and Rousseeuw, 1990; Everitt et al., 2001; Aggarwa et al., 2014). Cluster techniques have been employed in numerous disciplines such as archaeology for example, to investigate the relationship between various types of artefact: see for example Everitt et al. (2001) for a brief description or Baxter (2006). The two most widely used clustering algorithms are partitional and hierarchical clustering. In this paper, we concentrate on hierarchical clustering whose approach consists in developing a binary tree-based data structure called the dendrogram. Hierarchical clustering techniques may be subdivided into agglomerative methods that aggregate iteratively the individuals into clusters and divisive methods that split up each cluster into smaller ones. Agglomerative and divisive clustering are complementary. In the literature, divisive methods are largely ignored. The main reason is the computational aspect: divisive clustering is more expensive in terms of computing time than agglomerative. Nevertheless, it does not need to generate the complete divisive hierarchy to analyse a top-level partition. The scarcity of methods using divisive algorithms led us to focus on them. A number of splitting procedures were designed in the past, the oldest one being by Williams and Lambert (1959) is said to be monothetic in the sense that object sets are split according to the values of only one variable. Another approach using all variables simultaneously is said to be polythetic, it has the advantage to only depend on inter object distances or dissimilarities. The method of Macnaughton-Smith et al. (1964) combine the advantages to be divisive, polythetic and computationally manageable. It is implemented in DIANA (Divisive ANAlysis) clustering algorithm available on R free software environment.

In this work, we propose a new constrained Divisive Hierarchical Clustering (DHC) method named MapClust that takes into account geographical constraints and we apply it on an archaeological pottery dataset from middle Loire Valley (France). To study the contribution of the constraint, we also present the results obtained with DIANA. The MapClust clustering method was developed in a project on medieval and modern pottery from the middle Loire Valley (Husi dir., 2003, 2013). This project, which began more than twenty years ago, is one of the first research projects at the national level to have tackled the question of the study of pottery over a long period of time (6th to 17th c.), in a collective and systemic way, particularly in the structuring and analysis of data. This research has made possible the development of the european information network on medieval and modern ceramics named ICERAMM. The website associated with this network is built as an online database referencing typological tools and information on pottery assemblages for France and Belgium (http://iceramm.univ-tours.fr/). This work, which led to the creation of an important corpus of pottery, has opened up a new field of investigation in statistical data analysis, a collaboration between archaeologists from the Laboratoire Archéologie et Territoires (UMR CITERES, CNRS/Université de Tours) and statisticians from the Laboratoire de Mathématiques Jean Leray (UMR 6629, CNRS/Université de Nantes). This long-standing collaboration has contributed to the development of archaeo-statistical methods adapted to voluminous artefact data. The statistical tools developed cover both the estimation of dating and the clustering of archaeological contexts for the construction of periodization or spatial studies. This research has resulted in several publications in international journals (Bellanger et al., 2006, 2008, 2015; 2021; Bellanger and Husi, 2012).

As the pottery corpus continues to grow with the addition of new sites, a purely descriptive approach to the spatialization of the results is insufficient for a rigorous analysis of the pottery data. This is why we developed MapClust, a clustering method to deal with data that are not homogeneous in their spatial distribution and often patchy, for which traditional spatial statistics tools like kriging are inadequate. MapClust makes it possible to answer the problems of clustering and spatialization of archaeological contexts associating (i) pottery data that compose them, (ii) geographical distance between sites. This approach is essential for the construction of cultural areas, in this case pottery areas, in accordance with the distribution of data and the greater or lesser geographical proximity of the sites. The constructed clusters correspond to geographical areas formed by archaeological sites (i) close in distance and (ii) similar in terms of technical, typological and stylistic criteria of the pottery data.

The paper is outlined as follows. Section 2 describes the archaeological background. Section 3 presents the data corpus and the Mapclust clustering method. The results are shown in Section 4. Finally, we discuss the benefits of the novel clustering approach in Section 5.

2. Archaeological background: first results on cultural areas of the Middle Loire Basin based on pottery

The first two parts of the research programme, completed in 2013, laid the foundations for a reflection on the transformation of cultural areas based on pottery over a long period of time and in an area with territorial coherence such as the Middle Loire basin (Husi dir., 2003, 2013). The aim was already to better identify the permanent features and changes. Is the Middle Loire area composed of one or more economic or cultural entities? What are the pertinent spatial analysis scales to understanding changes in the area? How does this or these territorial networks change over time? What role does the Loire or its affluents play in the structuring of this area? If there is a middle Loire entity, what is its geographical limit and therefore its area of influence? These are all questions that pottery alone can only answer imperfectly, but which have the advantage of being an omnipresent source in archaeological sites, unlike other types of artefacts. Identifying economic and cultural spaces therefore requires a precise and quantified analysis of the data according to the technical, stylistical and functional traditions of pottery, including innovation, imitation and competition between potters. One way is to compare technical traditions, such glaze, painting, engobe, the presence of printing decorations or the choice of firing and therefore the colour of the wares.

We present here the results acquired in 2013 which have guided the methodological developments of MapClust. We have therefore chosen to focus the presentation on the period between the 8th and 10th c. For this generally poorly documented period, pottery represent one of the best sources to understand the socio-economic and cultural mechanisms. The pottery corpus selected comes from excavations carried out in the main towns of the middle Loire Valley and some more minor sites. The sites are selected according to the typological quality of the pottery assemblages and the chronological quality of the stratigraphic set (domestic occupation, dumps...). Each stratigraphic set is included in an archaeological site located in a place (city, town...). The results of 2013 show a map at two scales (Fig. 1): (a) local economical areas around major consumption centres rarely exceeding 50 km around a production site or a major consumption centre ; (b) two larger cultural spaces (red and green on the map) with different manufacturing traditions. These two spaces are already distinguished by their pottery typology but also especially (i) in the southwest by a majority of white-beige ware with painted bands and glazed ware; (ii) in the northeast by a large presence of engobed ochre-red ware and the absence of glazed ware (Husi dir., 2013: 220-252). Although there may be small local evolutions between the 8th and 10th centuries, these technical and decorative traditions are widely represented in North-West Europe during these three centuries. This observation justifies the choice of this time step to understand how the Loire area fits into a wider European cultural area (Husi, 2010).

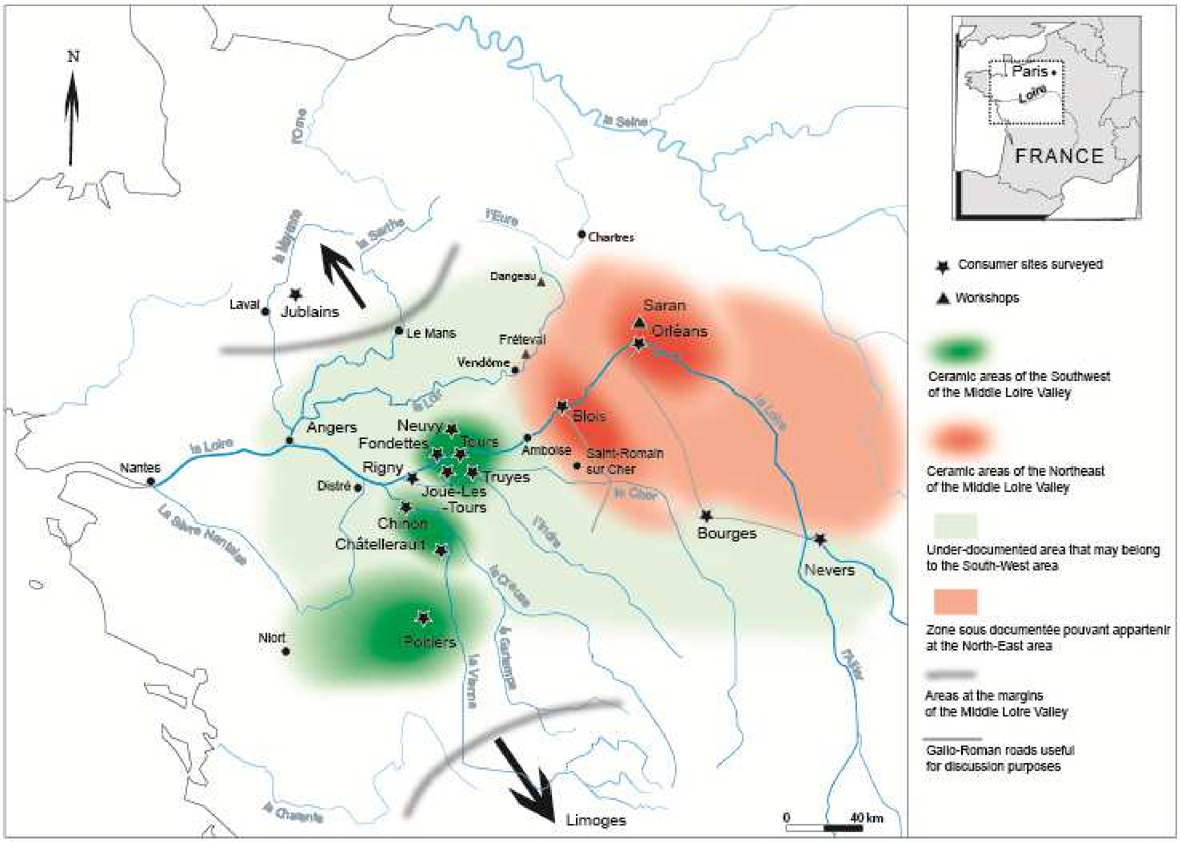

Fig. 1: Map showing the distribution of areas based on pottery traditions on the scale of the Middle Loire valley and local economic areas. (8th – 10th c.) (Husi dir. 2013 : 243).

In 2013, the corpus studied for this period was 14 places, 47 sites, 214 stratigraphic sets and about 13,000 vessels (Minimum Vessels Count); it is now 51 places, 87 sites, 333 sets or about 25,000 vessels. The continuous augmentation of the corpus, especially the number of sites and sets between 2013 and 2020, accentuates the spatial dispersion of the data at the origin of the development of MapClust. This tool is a decision aid for the archaeologist in the construction of cultural areas when the corpus becomes too large to interpret the data without recourse to statistics.

3. Materials and methods

3.1 Data collection

In this paper, which main aim is to illustrate the interest of the MapClust method, we have chosen only two of the numerous technical and decorative traditions contained in the 2020 data corpus. We will present here the analyses of the spatial distribution of white-beige or ochre-red pottery between the 8th and 10th c. mobilizing for this question 22 places out of the 51 recorded for the Middle Loire basin (Fig. 2). This chronological choice is guided by the changes that took place during this period, when light-coloured productions made in oxidising firing replace those of dark black or grey colour in reducing firing.

We have chosen the technique of Minimum Vessels Count (MINVC) to quantify pottery by fabric (Orton and Tyers, 1992). It is the quantifying technique nearest to historical reality without, for example, problems of fragmentation rates such as the number of sherds.

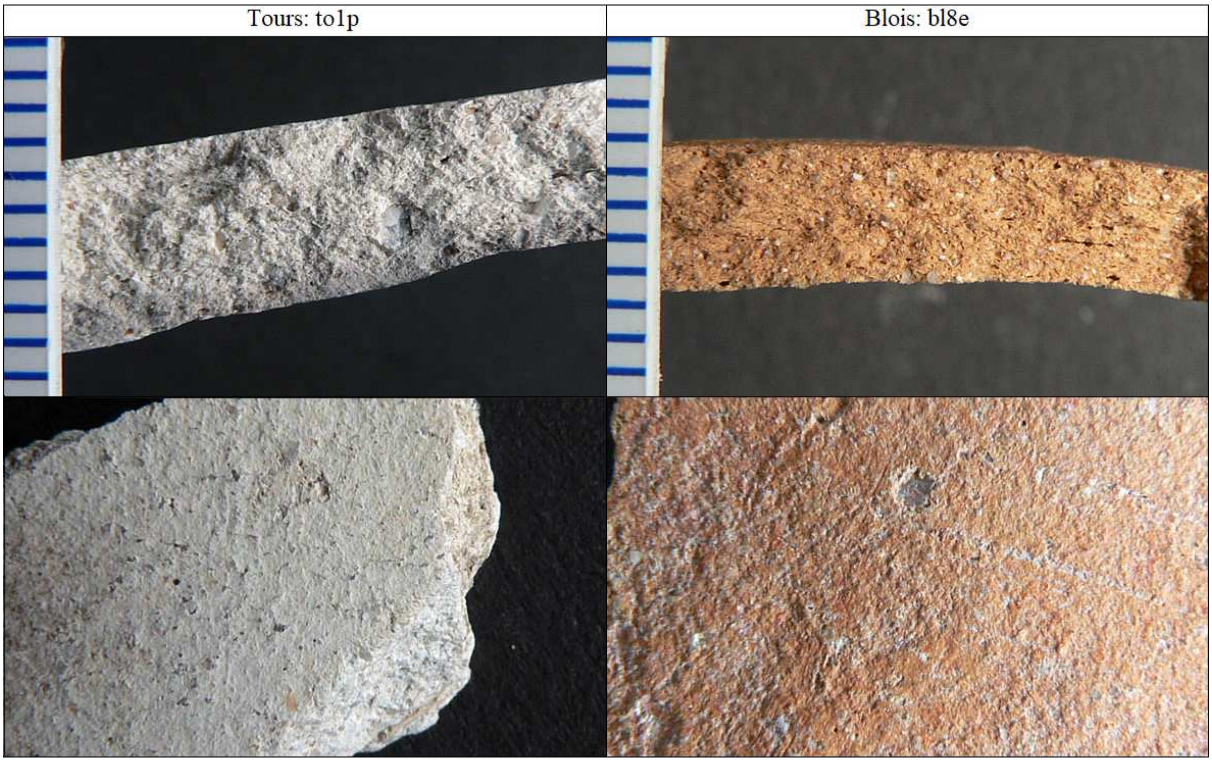

Fig. 2. Example of fabrics : Tours white-beige (fabric to1p); Blois ochre-red (fabric bl8e)

In this research work, we want to understand the transformation of cultural areas on the scale of the middle Loire Valley based on the evolution of pottery craft traditions. The manufacturing traditions corresponding to the pottery categories are defined as productions with identical or similar technical and stylistic characteristics. They come either from the same workshop, or from different workshops competing for the manufacture of fashionable pottery. In other words, the question is to identify spatially and over time the evolution of consumption habits, competition between products and potters' know-how.

3.2 Data analysis: a divisive hierarchical clustering approach with geographical constraints

In this work, we focus on Divisive Hierarchical Clustering method (Kaufman and Rousseeuw, 1990; Everitt et al., 2001; Aggarwa et al., 2014) to answer the problems of clustering and spatialization of archaeological contexts associating (i) pottery data that compose them, (ii) geographical distance between sites. The scarcity of methods using divisive algorithms led us to focus on them. Morerover, Macnaughton-Smith et al. (1964) argue that “divisive methods are safer than agglomerative ones, because “wrong” decisions in the early stages of an agglomerative analysis cannot be corrected later on, and one is mostly interested in the large clusters” (see Kaufman and Rousseeuw, 1990 page 273).

We propose a new Divisive Hierarchical Clustering method named MapClust that takes into account geographical constraints and compare it to the classical polythetic method DIANA (Divisive ANAlysis) (Macnaughton-Smith et al., 1964).

3.2.1 The proposed divisive hierarchical clustering method with geographical constraints

MapClust is based on spatial indicators intended to capture spatial patterns of a set of individuals (Woillez et al., 2009): the Centre of Gravity and the number of spatial patches.

Centre of Gravity and number of spatial patches

The centre of gravity CG is defined as the mean geographic location of the studied individuals. Noting $s_i = (x_i, y_i), i = 1, \dots, N$ the points in two-dimensional space and $z(s_i)$ the value of the regionalized variable measured in s_i, CG is defined by:

$$CG = \frac{\sum_{i=1}^{N} s_i z(s_i)}{\sum_{i=1}^{N} z(s_i)} \qquad (Eq. 1)$$

To make sense, the regionalized variable of interest z in Eq. 1 must be positive such as for example a frequency, a count or a probability density value. In this work, z will be associated to the Minimum Vessel Counts of the studied stratigraphic set (see appendix B).

As the spatial distribution of the data may be heterogeneous and present local aggregations or patches (e.g. archaeological pottery data), Woillez et al. (2007) proposed an algorithm to identify them "by attributing each sample to the nearest patch, with respect to a maximal threshold distance to its CG" (see Eq. 1). A point s_i is assigned to a patch according to its value $z(s_i)$ and its distance to other existing patches. The position of a patch is then determined by its CG. The algorithm starts with the largest value $\max_i z(s_i)$ and then considers each observation i in descending order of $z(s_i)$ value. The highest value initiates the first patch (Fig. 3.1). Then, the considered observation is assigned to the nearest patch, provided that its distance to the patch's CG is smaller than a fixed threshold distance (noted $dlim$: limit distance) (Fig. 3.2). Otherwise, the observation forms a new patch (Fig. 3.3).

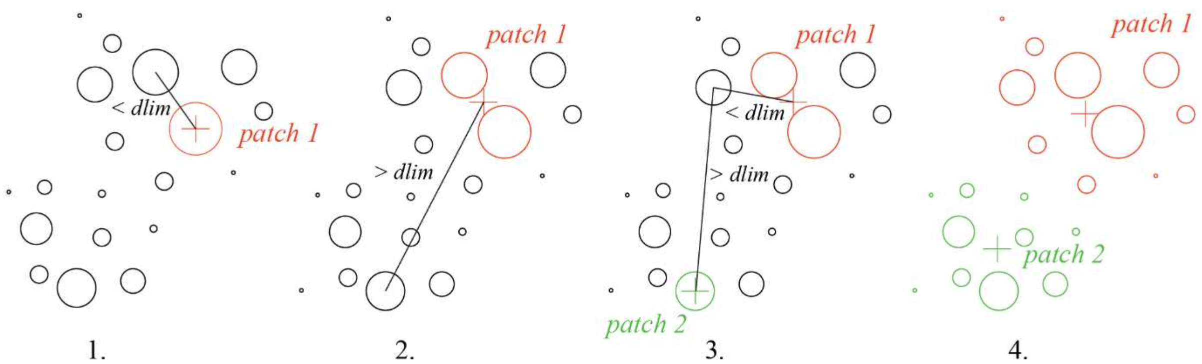

Fig. 3. Main steps of spatial patches detection algorithm for a fixed threshold distance dlim. Simple case of 2 patches.

The number of detected spatial patches is of course influenced by the choice of the $dlim$ threshold and the location of the highest z values: the smaller dlim is, the more patches will be built. Based on this observation, we have designed a divisive clustering procedure with geographical constraints by varying $dlim$ in the algorithm written to identify patches (Woillez et al., 2007).

MapClust principle

In order to better understand the nesting of geographical spaces, the previous algorithm used to determine the number of spatial patches is repeated in a divisive hierarchical approach by decreasing the maximum acceptable threshold distance $dlim$ (between observation points and the CG of a patch) until the first split in two. This approach leads to a divisive clustering approach (Everitt, 2001) based at each step on a k-means algorithm with $k = 2$: each node (parent cluster) of the hierarchy is splitting to obtain a bipartition of two clusters (child node) until only one observation per cluster is obtained (see Appendix A for algorithm details). The result can be visualized as a dendrogram, which shows the hierarchical divisions performed at each step of the process. For MapClust clustering procedure, the lengths of the steams (heights) of the dendrogram represent the *dlim* values at which clusters are split up (Fig. 4).

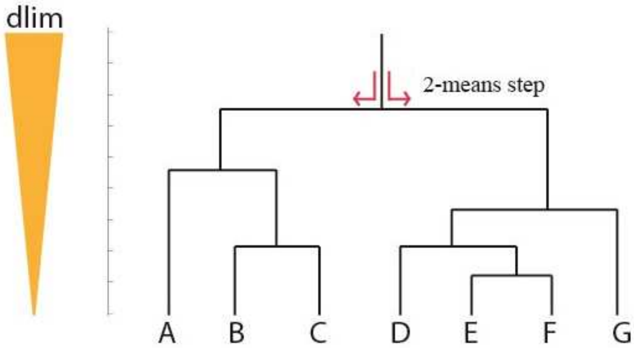

Fig. 4. Dendrogram for MapClust clustering method.

As we are not interested by the complete hierarchy but only in one selected partition, obtained from it, we must select one of the solutions in the nested sequences of clustering that comprise the hierarchy. This is equivalent to cut the dendrogram at a particular height to determine the number of clusters to be retained (sometimes called best cut). Large changes in division levels are taken to indicate the best cut. In this work, two methods based on the optimization of a criteria have been used to determination of the optimal number of clusters in the data: elbow method (with the total Within cluster Sums of Squares (WSS) criterion to minimize over a range of possible values for the number of clusters, see appendix C for details) and silhouette method (with the average silhouette width criterion to maximize over a range of possible values for the number of clusters, see appendix C for details). In addition, the archaeologist's expertise is called upon to select from the number of possible clusters the one that provides the most archaeologically interpretable clusters.

3.2.2 MapClust clustering process for non positive or multivariate regionalized variable

If the variable of regionalized interest z does not take positive values or is not univariate, CG could not be defined as defined in (Eq. 1). However, these situations are not rare in practice. Therefore, to adapt and extend our method to these cases, we propose an approach inspired by the density-based clustering method proposed by Hartigan (1981). The N points $z(s_i)$ are treated as a sample from a population with probability density function f which takes real values between 0 and 1. $\hat{f}$, a probability density estimate of f is obtained by the kernel method (Parzen, 1962), a non-parametric way to estimate f. Then z is replaced by $\hat{f}(z)$ in Equation (Eq. 1) such that CG can be written:

$$CG = \frac{\sum_{i=1}^{N} s_i \hat{f}(z(s_i))}{\sum_{i=1}^{N} \hat{f}(z(s_i))} \qquad (Eq. 2)$$

The use of probability density estimation $\hat{f}$ to compute CG in (Eq. 2) makes it possible to implement MapClust clustering algorithm (Appendix A) in non-positive or multidimensional variable of regionalized interest cases. It is also possible to (i) reduce the dimensionality of the data set using factorial analysis methods (Bellanger and Tomassone, 2014; Pagès, 2014) like Principal Component Analysis (PCA) if variables are quantitative or Correspondence Analysis (CA) for non negative ratio scale data, (ii) apply the MapClust method using (Eq. 2) where $\hat{f}$ is the estimated density of the chosen principal components. Factorial analysis allows to convert sets of higher-multidimensional feature into lower dimensional feature sets that preserve as much information as possible. The selection of the number of components depends on the data and must be chosen carefully. But it has no rigorous solution. This choice must allow the essential data to be extracted while eliminating less significant information during the clustering step.

Finally, to study the contribution of the constraint, we also compare our results with MapClust to those obtained with DIANA clustering algorithm (Macnaughton-Smith et al., 1964). All statistical analyses are performed using the free software R (R Core Team, 2016).

4. Results

The aim is to compare with the use of MapClust the present results about culture areas with those of the 2013 map (Fig. 1), only partially here with the example of the white-beige and ochre-red pottery, all the criteria become part of a future larger publication on the subject in the Loire Valley (Husi dir., 2021). The difference with 2013 comes from a much larger site and pottery corpus and from a statistical and not simply descriptive approach of the data, central question of this paper.

4.1. Data description

The general corpus of 2020 is 51 places, 87 sites, 333 sets or about 25,000 vessel (§ 2).The reduced corpus of data mobilised to compare the distribution of white-beige and ochre-red pottery from the 8th to the 10th c. is structured by stratigraphic sets and pottery categories (see §3.1). It is defined in 22 places, 44 sites and 116 stratigraphic sets. The pottery data are quantified in Minimum Vessel Count (MINVC) by categories and represent a total corpus of

5668 vessels (Table 1). Among the 116 sets between 8th to the 10th centuries, there are 99 sets with white-beige production and 103 sets with ochre-red production. For the multidimensional case 101 sets contain at least one of the productions in quantity satisfying the condition of using a correspondence analysis (sum of all productions MINVC greater than 5 individuals).

Table 1. Description of white-beige and ochre-red categories of pottery.

			Vessels (MINVC)			
	Sites	Stratigraphic sets	Total	Median **(per set)**	Mean **(per set)**	Standard deviation **(per set)**
White-beige	30	99	1713	8	17.30	29.19
Ochre-red	35	103	3955	11	38.39	67.80

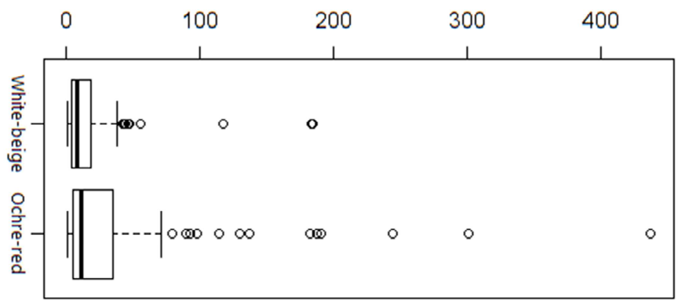

Fig. 5. MINVC for white-beige (top) and ochre-red (bottom) in the stratigraphic sets (99 for white-beige, 103 for ochre-red).

For both white-beige and ochre-red, half of the stratigraphic sets have less than 10 vessels (Fig. 5, Table 1). The studied dataset also contains some stratigraphic sets with large quantities of ochre-red pottery (10 sets with more than 100 vessels) and of white-beige pottery (3 sets with more than 100 vessels). These sets explain the difference between the median and the mean and the relatively high value of standard deviation (Table 1). It can be noted that for a relatively similar number of sets (of which 86 are common) the number of ochre-red pottery is twice as important as that of the white-beige production. The median is close between the two types of production but not the mean value. The number of sets with large quantities of ochre-red pottery is greater than those with large quantities of white-beige production. As the total quantity of pottery is very different from one set to another, the raw MINVC data for white-beige and ochre-red are unusable without transformation. Therefore in all the following, for the unidimensional cases, we use variable of interest derived from MINVC (Appendix B).

4.2. From the data to the determination of homogeneous areas using unidimensional information

The colours of the vessels depend on the preferences of the consumers but also on the choice and the technical constraints of the potters. These constraints depend on the characteristics of the clays, in our case on the importance of iron oxide for the red colour, but

also on the firing atmosphere. As the same clays are present all over this part of the middle Loire Valley, their extraction is done locally according to the colour chosen by the potters. Consequently, the spatial distribution based on the colour of the recipients is rather cultural, linked to a stylistic choice, than dictated by potential technical constraints.
The objective is to first look at the ochre-red and white-beige data separately and secondly performing a multidimensional analysis by integrating the two data sets.

4.2.1 Ochre-red productions between the 8th and the end of the 10th century

The aim here is to analyse the distribution of ochre-red pottery (see Fig. 1) by stratigraphic set taking account of the geographical distance between places. Each place being composed of several sets sharing the same coordinates, we have chosen to present the results of MapClust clustering on places, the best scale for the definition of cultural areas.

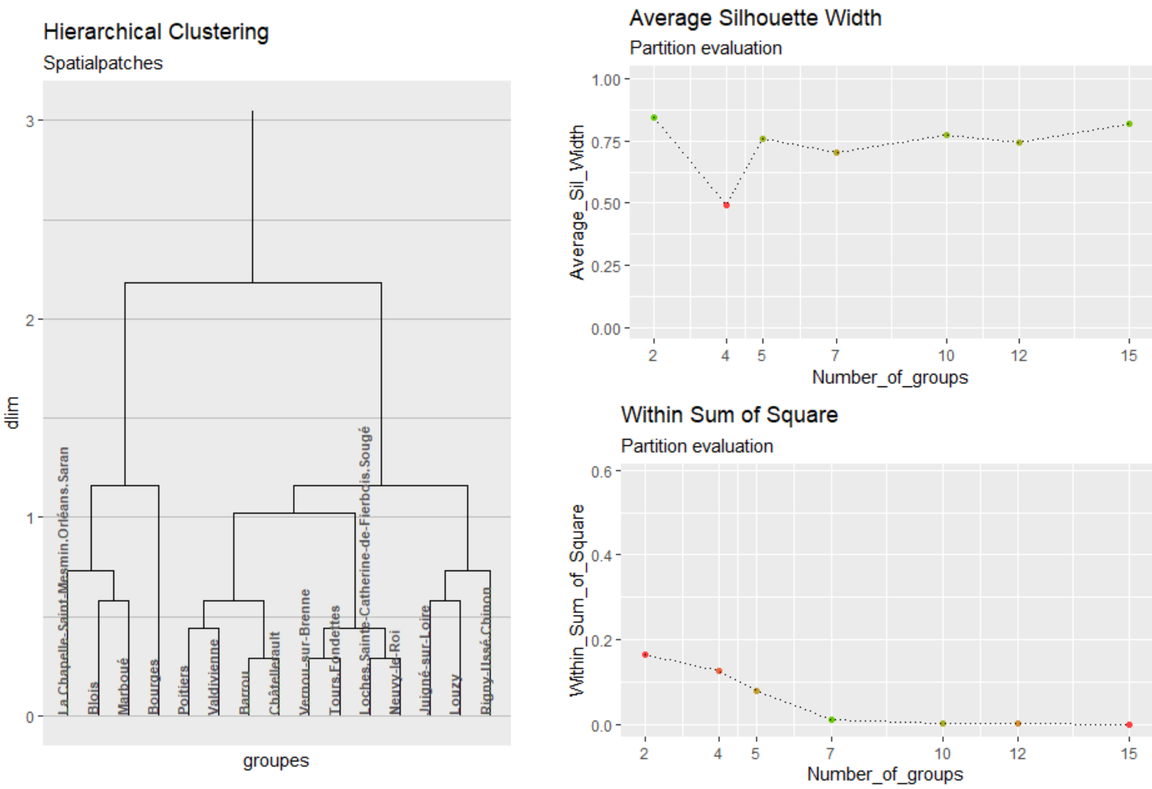

Fig. 6. Ochre-red - Dendrogram obtained with MapClust (left).
Average silhouette width (top right) and WSS (bottom right).

Referring to the average silhouette criterion, calculate on weighted Euclidean distances (Appendix C), the MapClust partitions are correct except for the one corresponding to a partition into 4 clusters (Fig. 6). The WSS indicates that the optimal number of clusters can be chosen between 5 and 7. The choice was made to retain the partition with 5 clusters, a finer partition with more clusters having historically little meaning.

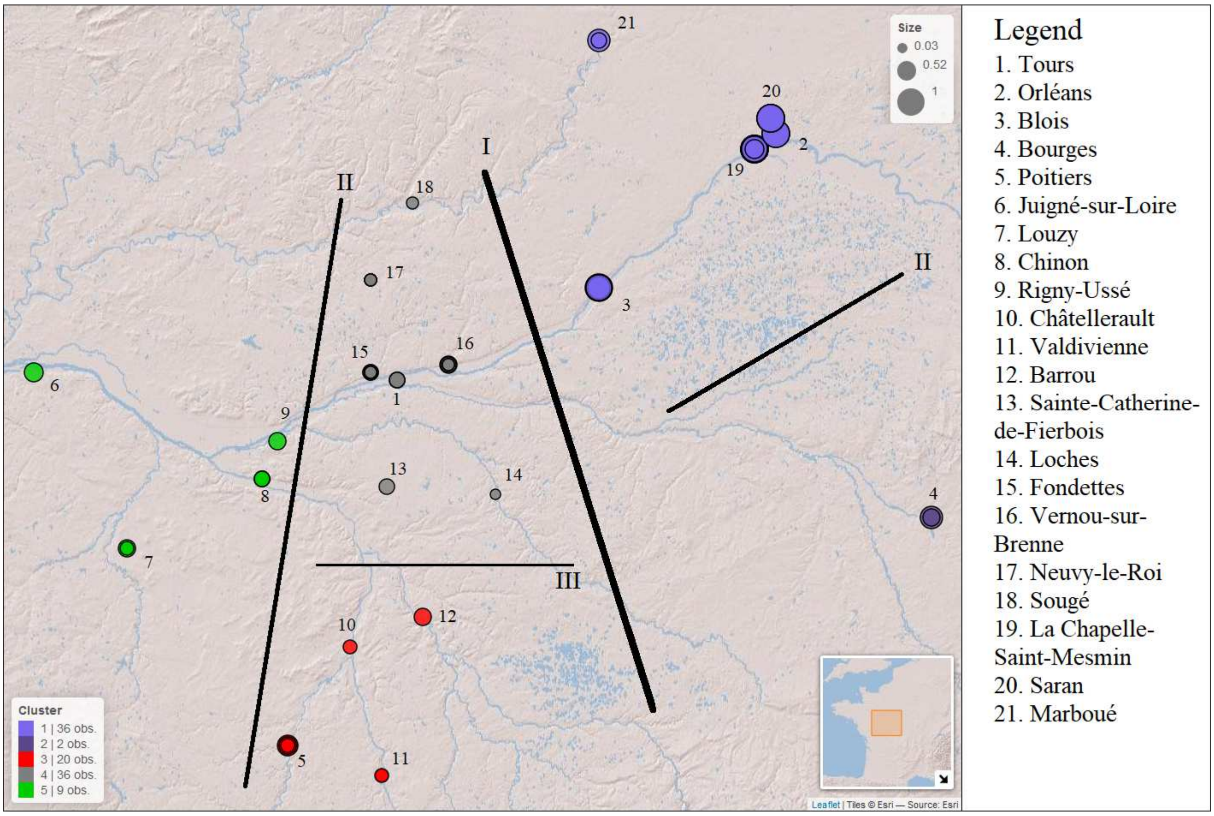

Fig. 7. Ochre-red - 5 clusters partition (spaces) resulting from the MapClust algorithm on 103 sets. The lines give the hierarchy of partitions (I to III, because level 2 occurs in both sub cluster in the same time). The size of the circle increases with the frequency of the productions.

The ochre-red colour productions between the end of the 8th c. and the 10th c. are divided into 5 clusters representing five areas, (Fig. 7). The off-centre geographical position of the Marboué and Louzy sites in relation to the core of the middle Loire Valley is related to their poor ranking according to the silhouette index. The first partition in two clusters separates a North-Eastern area of the middle Loire Valley, which includes the majority of ochre-red production, from a south-eastern area where they are generally less represented.

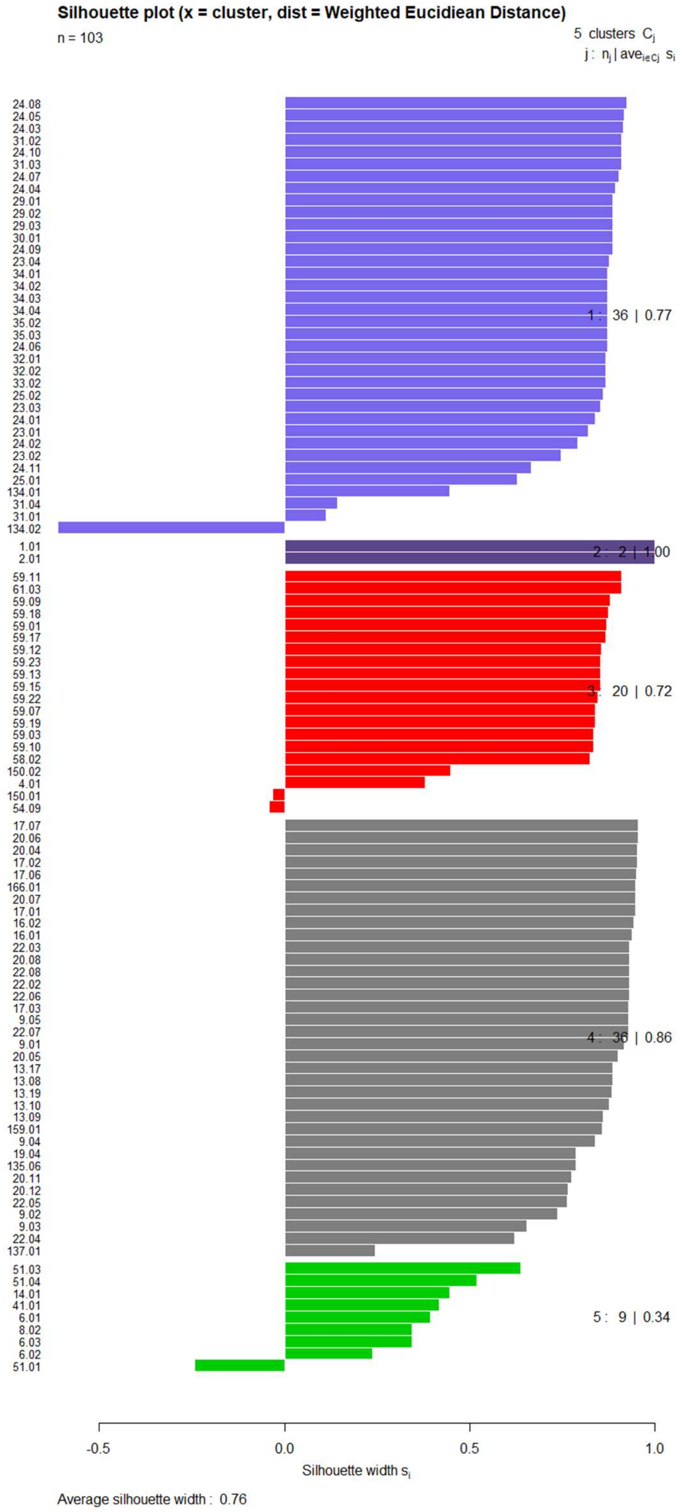

Fig. 8. Ochre-red - Silhouette index by MapClust cluster.

Two sites, from Marboué (134.02) and Louzy (51.01) are misclassified (Appendix C, silhouette index) (Fig. 8). The ochre-red productions are less significant at Marboué than at the other sites in the cluster 2, and the opposite is true for the Louzy site in the cluster 5. However, because of their geographical position, they are included in these clusters despite the fact that their pottery production is not similar. These sets are located in places at the margins of our space and are influenced by other production areas.

4.2.2 White-beige productions from the 8th to the end of the 10th century

As for ochre-red pottery, the objective is to analyse the distribution of white-beige pottery based on their quantified in Minimum Vessel Count by stratigraphic set and taking account of geographical distance between places. Each place is composed of several sets that share the same coordinates.

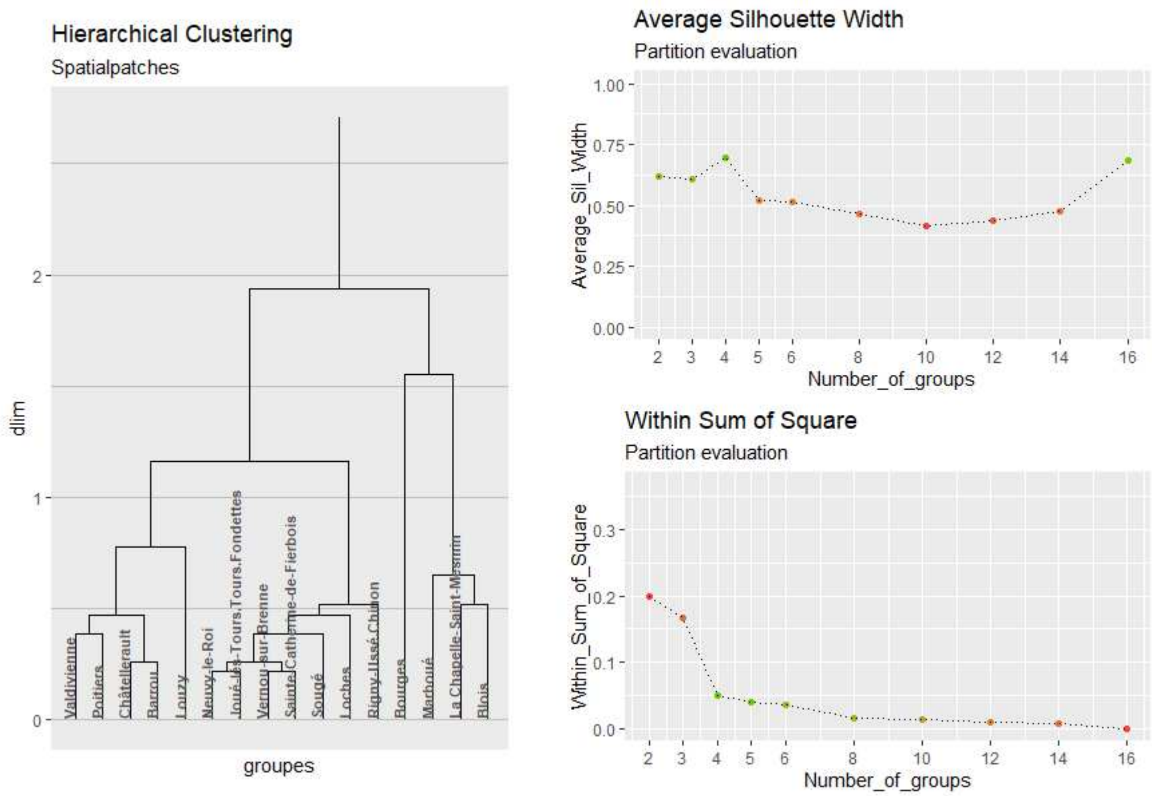

Fig. 9. White-beige - Dendrogram obtained with MapClust (Left).
Average silhouette width (top right) and WSS (bottom right).

Referring to the average silhouette and WSS criteria, the best value corresponds to a partition into 4 clusters (Fig. 9).

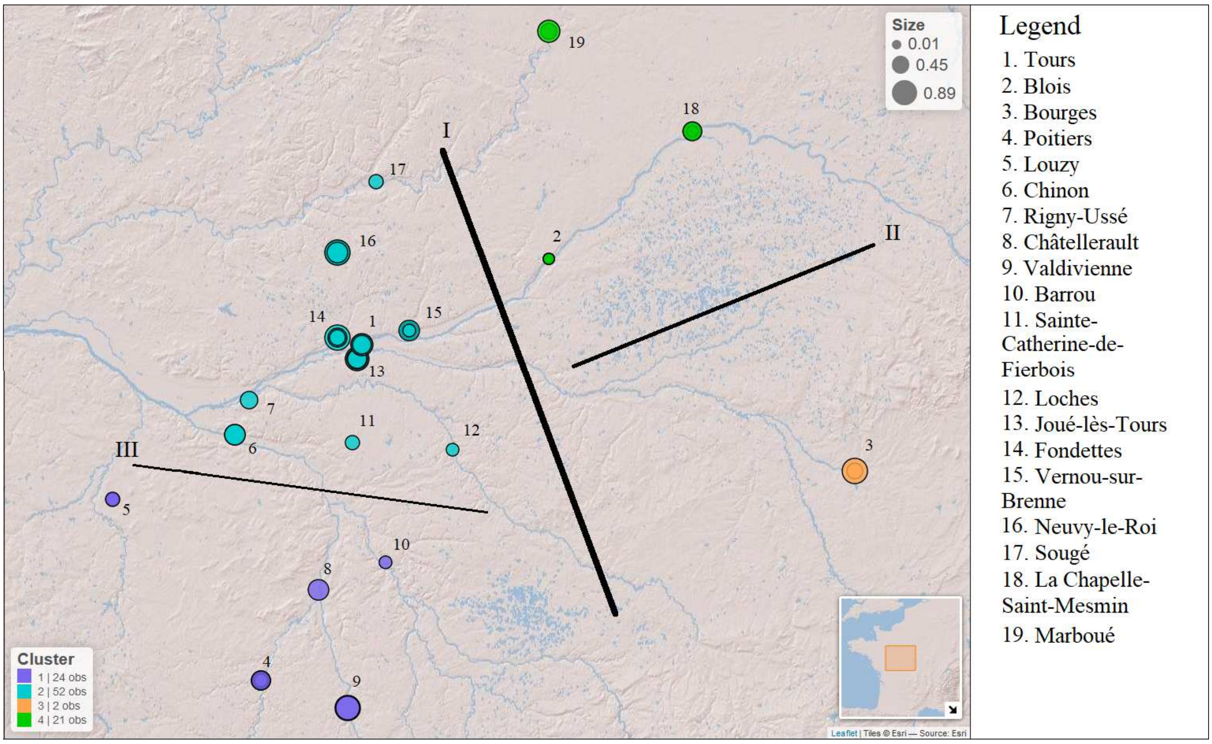

Fig. 10. White-beige - 4 clusters partition (spaces) resulting from the MapClust algorithm on 99 sets. The lines give the hierarchy of partitions (I to III). The size of the circle increases with the frequency of the productions.

The white-beige productions between the end of the 7th century and the 10th century are divided into four clusters representing four local economic areas (Fig. 10). The geographical position, off-centre of the Marboué site, revealing an economic area with little connection to the Loire, is certainly the cause of a poor allocation to clusters 4. The first partition in two clusters, which is also interesting, separates two areas, North-East and South-West, according to this manufacturing tradition: (i) the first with a marginal representation of these products in Blois and Orléans, (ii) the second which concentrates the majority of them between Haut-Poitou and Touraine. A set of the site of Marboué is badly clustered (Appendix D: 134.02). While white-beige products are well represented in this set, it belongs to the cluster 4, which is characterised by a low presence of such products. These set is located in a place at the margins of our space and is influenced by other production areas.

4.3. Multidimensional analysis of the productions of white-beige and ochre-red colour between the 8th and 10th century

We consider here the same question of defining pottery areas based on the colour of the productions, but analysing simultaneously the two productions, white-beige and ochre-red. Correspondence Analysis (CA) (Benzécri, 1973; Greenacre, 2007) was performed on both variables to describe the structure of the dataset. We use $\hat{f}$, the probability density estimation of the unique principal component of the CA as the variable of interest to compute CG in (Eq. 2) and obtain a partition with MapClust clustering method.

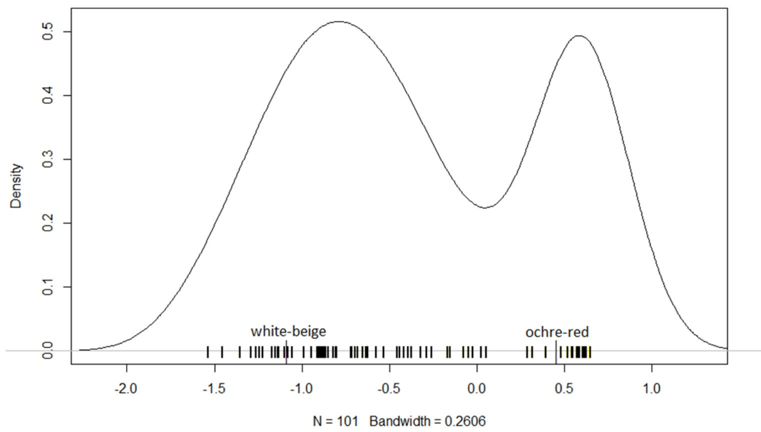

Fig. 11. Ochre-red & white-beige - Kernel density estimate of the only CA principal component. Contributions of the two variables and the sets are plotted on the abscissa.

Only those sets and pottery categories with more than 5 individuals have been selected for the CA. The ochre-red (negative values) and white-beige (positive values) colour productions are opposed on the first axis of the CA, thus characterising two different profiles (Fig. 11).

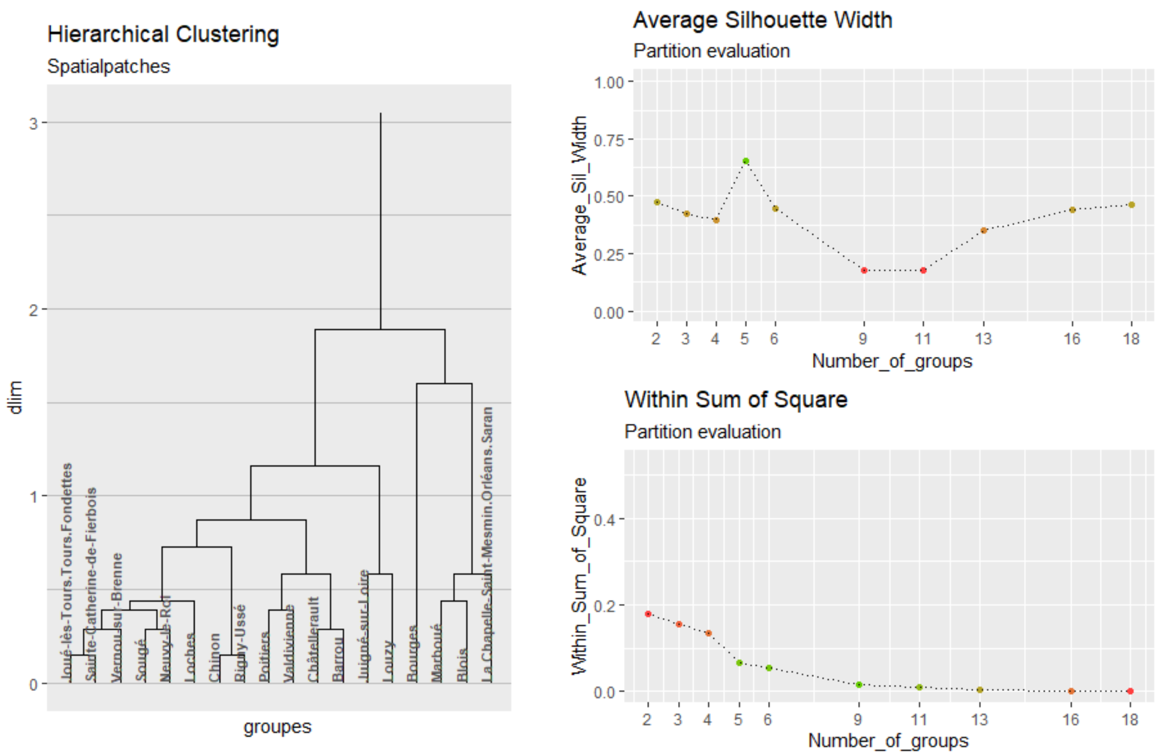

Fig. 12. Ochre-red & white-beige - Dendrogram obtained MapClust (left). Average silhouette widths (top right) and WSS (bottom right).

Referring to the average silhouette widths criterion, the best value correspond to the partition into 5 clusters (Fig. 12). The WSS criterion confirms the choice of a 5-clusters partition.

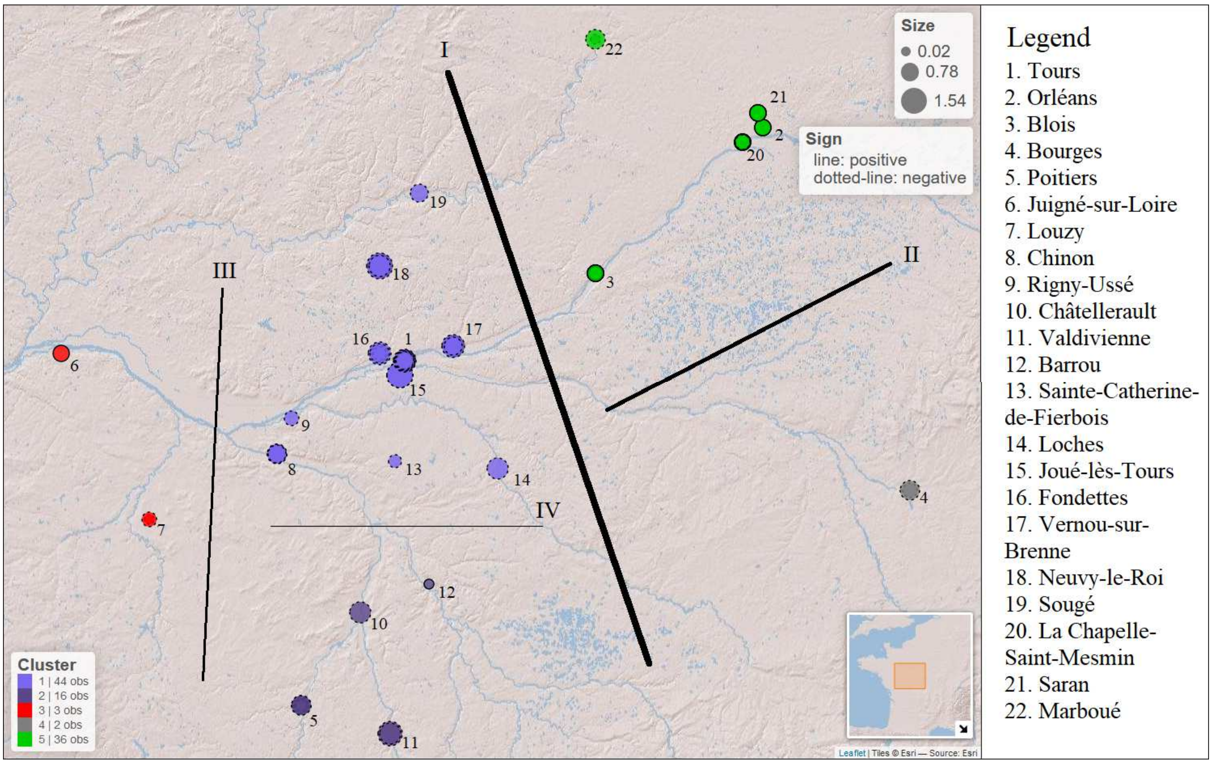

Fig. 13. Ochre-red & white-beige - 5 clusters partition (spaces) resulting from the MapClust algorithm on 101 sets. The lines give the hierarchy of partitions (I to IV). The size of the circle increases with the frequency of the productions.
Dotted circles: places with sets containing a majority of white-beige productions; solid lines for ochre-red productions.

Such a partition highlights five areas based - when they are sufficiently informed - on the main consumption centres of the middle Loire Valley, of which the best attested are Touraine, Haut-Poitou and Blésois-Orléanais. Two areas in the far west and Berry, on the steps of the middle Loire Valley, are still insufficiently informed to be associated with one or other of the traditions (Fig. 13). The three sets, of Juigné, Chinon and Vernou-sur-Brenne are misclassified (Appendix E: respectively sets 41.01; 6.01; 22.06). White-beige production is less important at these sites of Chinon and Vernou-sur-Brenne than at the other sites of cluster 1, where this production dominates, and ochre-red production is in the majority at the Juigné site, in contrast to the other elements of cluster 3 which clearly lacks data. The first partition (I) of the dendrogram very quickly separates a South-Western space of white-beige tradition (dashed circles) and a North-Eastern space of ochre-red tradition (solid circles), the South-Western/North-Eastern.

The multidimensional case has a slightly smaller average silhouette width (0.65, Appendix D) but is still good. We find the same overall clusters than previously. The multivariate analysis, together with the univariate analyses on each colour separately, allow to highlight the robustness of the results found on all these analyses (e.g. the first separation between North-East and South-West). Some partitions differences may be observed due to (i) particularity in the studied datasets, (ii) small amounts of MINVC for some stratigraphic sets

4.4 Comparison of MapClust and divisive hierarchical clustering DIANA results

In order to analyse the impact of the geographical constraint, we compared MapClust to the polythetic Divisive hierarchical clustering method DIANA. To compare the agreement of the partitions and the number of clusters fixed between MapClust and DIANA, we use: (i) the confusion matrix (e.g. contingency table) that summarizes the overlap between the two possible partitions, (ii) the Adjusted Rand Index (ARI) (Rand 1971) that measures the similarity of objects present in the clusters: the closer the value is to 1, the better the accuracy between the partitions is.

4.4.1 Ochre-red productions

The DIANA partition in 5 clusters shows a first separation between two geographical spaces of the study area (South-Western/North-Eastern) (Fig. 14).

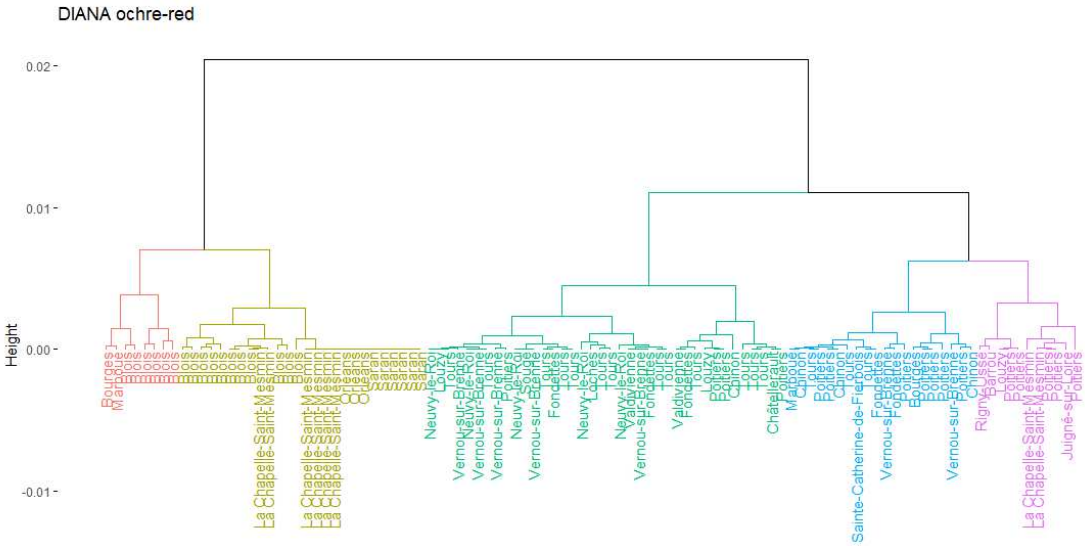

Fig. 14. Ochre-red - Dendrogram obtained with DIANA (5-clusters partition).

This result can be explained by a high production of ochre-red pottery in the North-Eastern part. The cities of Blois and Bourges are separated in two clusters (the clusters 1 and 2 and respectively the clusters 1 and 4) and it is globally difficult to identify geographical area with DIANA method (Table 2).

Table 2: Ochre-red - Confusion matrix between DIANA clusters and MapClust clusters.

Clusters		**DIANA**				
		1	**2**	**3**	**4**	**5**
MapClust	**1**	7	26	0	1	2
	2	1	0	0	1	0
	3	0	0	7	7	6
	4	0	0	28	8	1
	5	0	0	3	3	2

An ARI value of 0.43 indicates a mediocre agreement between the two partitions.

4.4.2 White-beige productions

DIANA results on white-beige dataset do not show a geographical separation of the study area (Fig. 15) as for the ochre-red dataset.

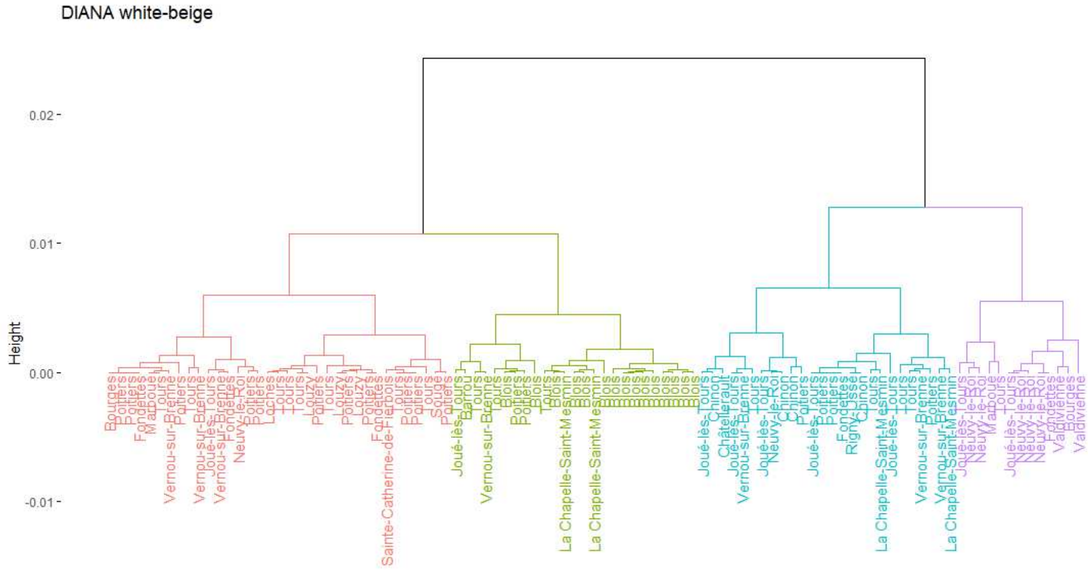

Fig. 15. White-beige - Dendrogram obtained DIANA (4-clusters partition).

The clusters are spread over large areas. It seems that DIANA method does not work properly for this type of data (Table 3).

Table 3: White-beige - Confusion matrix between DIANA clusters and MapClust.

Clusters		**DIANA**			
		1	**2**	**3**	**4**
MapClust	**1**	14	3	5	2
	2	19	5	19	9
	3	1	0	0	1
	4	1	17	2	1

It yields ARI value as 0.14 that indicates a very bad agreement between the two partitions.

The ochre-red dataset show a result with the first separation on the dendrogram. Despite this, the DIANA algorithm do not help for data with strong geographical context. This and the very few results with the white-beige dataset show the necessary need of geographical constraint in the context of univariate analysis.

4.4.3 White-beige and ochre-red productions

Starting from the contingency table of the two productions, we perform a correspondence analysis and use the coordinates of the unique principal component to generate the distance matrix. We then applied the DIANA clustering method (Fig. 16).

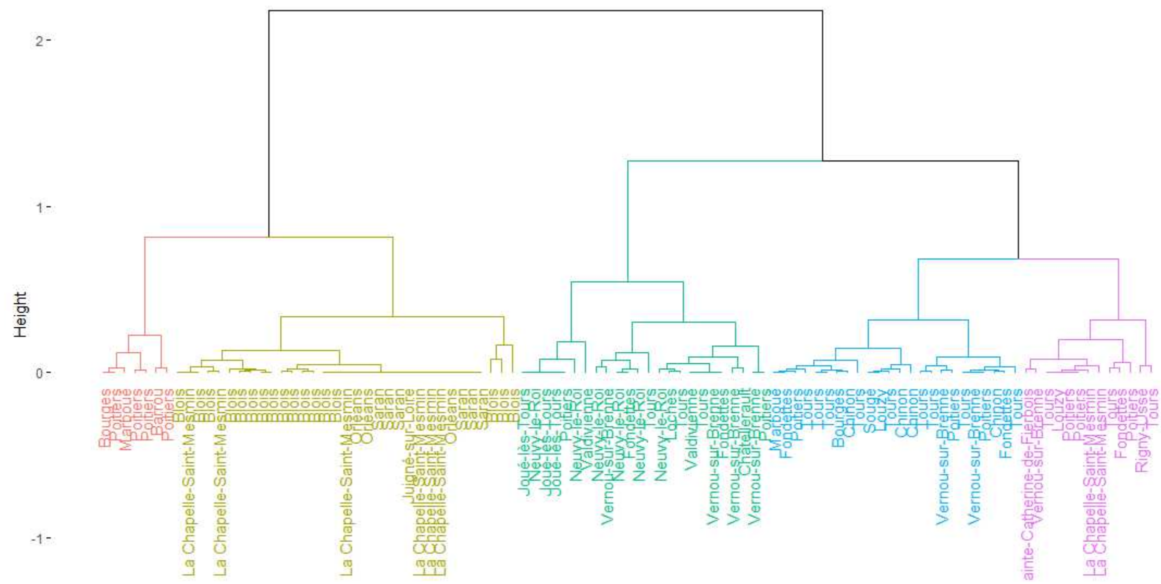

Fig. 16. Ochre-red & white-beige - Dendrogram obtained DIANA (5-clusters partition).

The results are very similar to those we obtained with DIANA with the ochre-red dataset. The agreement between DIANA and Mapclust partition is very poor (Table 4; ARI 0.45). DIANA gives also the first separation between two geographical areas (South-West/North-East).

Table 4: Ochre-red & white-beige - Confusion matrix between DIANA clusters and MapClust.

Clusters		**DIANA**				
		1	**2**	**3**	**4**	**5**
MapClust	**1**	0	0	19	18	7
	2	5	0	5	3	3
	3	0	1	0	1	1
	4	1	0	0	1	0
	5	1	32	0	1	2

An ARI value of 0.45 indicates a very bad agreement between the two partitions

The contribution of geographical constraints is essential for certain types of data in order to help determine clusters. The case of archaeology, characterised by the distribution of data in semi points, requires a review of how to integrate these constraints.

5. Discussion and conclusion

The spatial analysis of artefacts, particularly pottery, is generally realized only according to the intensity of the data, with a blurred boundary from one space to another, constructed without any real statistical processing. Moreover, in our case, archaeological data are measured at a limited number of locations. As a consequence, it is not possible to use geostatistical methods. Indeed, we choose to use simple spatial indicators without additional statistical assumptions to build our proposed clustering method. MapClust brings a new perspective to the archaeological problem of determination of cultural areas using pottery since it takes into account the data to build the clusters but also the distance between sites. The application of MapClust clustering method on pottery data leads to results near similar to those already observed on an empirical study in 2013, confirming the relevance of a statistical approach applied to a dataset too large for a simple descriptive analysis. Overall, there is great stability in the local economic areas located around the main consumption centres in the Loire area. Whether MapClust is used separately on white-beige or ochre-red dataset or with both of them, we can observe: (i) three main areas, Haut-Poitou, Touraine and Blésois-Orléanais, (ii) one or two secondary areas more on the periphery to the extreme west with Anjou and to the east with Berry. There is a strong tradition of common ochre-red manufacturing for the Blésois and Orléanais, always in the same cluster at a low level of partition (IV) of the dendrogram (Fig. 6, 9 and 12). In the same vein, it is necessary to wait for the third partition (III) to differentiate Touraine from Haut-Poitou, this statistical proximity reflecting a strong common anchoring in the white-beige tradition. Whatever the analysis, the first division with MapClust results in a similar partition of Loire space in two pottery area opposing South-West and Nord-East. By looking more globally at the structure of the partition, the first MapClust partition split this Loire region into two distinct cultural areas. The white-beige productions corresponding mainly to the south-west of the middle Loire Valley (Touraine and Haut-Poitou). The ochre-red productions are concentrated in the North-Eastern part of the middle Loire Valley with Orléanais and Blésois. More generally, the use of MapClust with the other criteria chosen for the definition of cultural areas, but not presented here, still reveals this same South-West and Nord-East partition (Husi dir. 2021 forthcoming).

From a statistical point of view, MapClust corresponds to the original divisive hierarchical clustering method with the addition of geographical constraints, which is new! Our choice of a hierarchical divisive clustering approach has many advantages over other approaches: (i) it is more efficient because it is often not necessary to generate the complete hierarchy, (ii) it is also more accurate than the agglomerative version because it takes into consideration the global distribution of data when making top-level partitioning decisions. Although clustering is regularly used in archaeology, in this work we present a new one, MapClust that is an original divisive hierarchical clustering method with geographical constraints providing a good way to answer the question of how to construct cultural areas from a large amount of data. The interest also comes from the fact that MapClust can be used for pottery but also for many other artefact. This methodological research is currently being pursued in a project of the French National Research Agency (ANR) entitled ModAThom for "Explanatory Model of Urban Construction of Angkor Thom". This project includes the study of an important pottery corpus, but also other artefact resulting from many excavations made in the capital of the Khmer empire in Cambodia (https://anr.fr/Projet-ANR-17-CE27-0008). In addition, MapClust has easy-to-use tools to choose the number of clusters and to characterize geographically the created clusters. It does not require a lot of practice of the R language to run, thanks in particular to a web application Shiny more intuitive for users. The choice was made to develop both approaches, classical R package and web application in order to enlarge the target audience. This statistical method is integrated in a larger interdisciplinary project dedicated to the processing of data in archaeology, whose tools will be implemented an update of the R package: **S**tatistical **PA**ttern **R**ecognition and da**T**ing using **A**rchaeological **A**rtefacts assemblage**S** (SPARTAAS, Coulon et al., 2021) available on the CRAN. In this study we present a new method, MapClust, designed to cluster archaeological data considering geographical constraints. It allows to keep a geographical consistency in the cluster analysis. Moreover, in the future, this method could find applications in many other fields.

Authors' contributions statement P.H. and L.B. designed the research, P.H. collected the data, P.H and A.C. checked the data quality, A.C. developed the R code, L.B. and A.C. performed the data analyses, and P.H., L.B. and A.C. wrote the paper. All authors reviewed the manuscript.

Declaration of competing interest the authors have no competing interests.

Acknowledgements

This paper was supported by the project ANR ModAThom (https://anr.fr/Projet-ANR-17-CE27-0008)

Appendix A: MapClust algorithm

Algorithm 1: MapClust

1. Initialization

- Root group initialization $C = s_1, \dots, s_N$
- Hierarchy Initialization $H_0 = C$
- Initialization of the first $dlim$: maximum distance from the point cloud
- Initialization of the step between two $dlim$

2. Preparation of variable z

if several variables are available **then**
Dimensional reduction with correspondence analysis (use of the first axes of the results)
Or use an multivariate kernel density estimation (PARZEN, 1962).
end if
if the variable of interest z is not equivalent to a count **then**
The variable z is replaced by the kernel density estimates.
Transformation from z to $f(z)$
end if

3. Construction of the hierarchy

while < $dlim > 0$ **and** groups have more than 1 observation > **do**
Retrieving groups from the hierarchy at the previous level
Create H_{tmp}: the current hierarchy level

for < each group at the previous level > **do**

Sort the observations in decreasing order of the variable of interest
Create group $C_x^* = s_i$ where $s_i = argmax\ z = s_{(N)}$ $\quad s_i \in C_x$ $\quad$ s_i corresponds to the observation with the maximal value of z

for < each observation > **do**
Recalculate the Centre of Gravity of C_x^*, $CG = \frac{\sum_{i=1}^{N} s_i z(s_i)}{\sum_{i=1}^{N} z(s_i)}$
Calculate the distance $D_i = d(s_i, CG)$ where d is the weighted euclidien distance defined as

$$d(i,j) = \sqrt{\left((x_i - x_j)^2 + (y_i - y_j)^2\right) * (z_i - z_j)^2}$$

if $D_i < dlim$ **then**
$C_x^* = C_x^* \cup s_i$
$C_x = C_x \setminus C_x^*$
end if
end for
for < each observations s_i > **do** $\quad$ try every observations without recalculate the centre of gravity
Calculate the distance $D_i = d(s_i, CG)$ $\quad$ without recalculate CG or CG'
Calculate the distance $D_i' = d(s_i, CG')$ where CG' is the center of gravity of the group C_x
if $D_i > D_i'$ **then**
$C_x = C_x \cup s_i$ $\quad$ if the observation is closer to the centre of gravity of C_x then
$C_x^* = C_x^* \setminus C_x$ $\quad$ the observation returns to the group C_x
end if
end for

Add C_x^* et C_x to the current hierarchy level H_{tmp}
end for

Add H_{tmp} to the hierarchy H
Delete H_{tmp}
Reduce $dlim$: $new\ dlim = dlim - step$

end while

Appendix B: MapClust regionalized variable of interest derived from Minimum Vessel Count

Unidimensional variable of interest

For each unidimensional case, ochre-red and white-beige, we start with a contingency table with all pottery categories as columns and all the sets concerned by the case as rows. Only sets with a non-zero MINVC of the focus production are kept. MINVC information on pottery is therefore contained in a contingency table $\mathbf{N}$ of size 103×J for the ochre-red case (resp. 99×J for the white-beige case).

$$\mathbf{N} = \begin{bmatrix} n_{1;1} & \cdots & n_{1;OR} & \cdots & n_{1;J} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{103;1} & \cdots & n_{103;OR} & \cdots & n_{103;J} \end{bmatrix}$$

Thus $n_{i\,;j}$ $(i = 1, \ldots, n_{OR} = 103$ (resp. $n_{WB} = 99$) and $j = 1, \ldots, J)$ represents the MINVC of pottery category j found in set (e.g. context) i for the ochre-red case (resp. white-beige case).

For the unidimensional analysis of ochre-red (resp. white-beige) productions, the variable of interest z_{OR} (resp. z_{WB}) in (Eq. 1) is defined for each set, as the MINVC of ochre-red pottery (resp. white-beige) divided by the total MINVC found in the set i ($\frac{n_{i;OR}}{\sum_{j=1}^{J} n_{i;j}}$), also divided by the sum over all sets. z_{OR} (resp. z_{WB}) is consequently given by

$$z_{OR}(s_i) = \left. \frac{n_{i;OR}}{\sum_{j=1}^{J} n_{i;j}} \middle/ \sum_i \frac{n_{i;OR}}{\sum_{j=1}^{J} n_{i;j}} \right. \qquad \text{(resp.} \quad z_{WB}(s_i) = \left. \frac{n_{i;WB}}{\sum_{j=1}^{J} n_{i;j}} \middle/ \sum_i \frac{n_{i;WB}}{\sum_{j=1}^{J} n_{i;j}} \right)$$

MapClust algorithm (Appendix A) can then be applied on z_{OR} (resp. z_{WB}).

Multidimensional variable of interest

For the multidimensional analysis of the productions of white-beige and ochre-red colour, A correspondence analysis (CA) of the two columns matrix $\mathbf{N} = \begin{bmatrix} n_{1;OR} & n_{1;WB} \\ \vdots & \vdots \\ n_{101;OR} & n_{101;WB} \end{bmatrix}$ **(**only sets with more than 5 individuals in both pottery categories taken together are kept) makes it possible to reduce the number of variables by constructing a unique synthetic variable (e.g. principal component). Then, $\hat{f}$, the probability density estimation of the unique principal component of the CA is used as the variable of interest to compute CG in (Eq. 2) and obtain a partition with MapClust clustering method.

Appendix C: Number of cluster selection and quality evaluation of the partitions

Several criteria exist to determine the number of clusters to be selected. Two indices have been selected: the average of the silhouettes widths and the Within Sum of Square (WSS). The expertise of the archaeologist, derived from his knowledge of the nature of the context and the data, was also taken into account in determining the number of clusters.

The silhouette index (Rousseeuw, 1987) is a widely used index for assessing the fit of individual observations in the clustering, as well as the quality of clusters and the entire partition. It is an index calculated for each observation. Let i be an observation belonging to cluster A, denote by C a cluster not containing i, the silhouette index is defined as:

$$s_{sil}(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where $\mathrm{a(i)}$ is defined as the average dissimilarity between i and all other observations in A, and $b(i) = \min_{C \neq A} d(i, C)$ with $\mathrm{d(i, C)}$ is the average dissimilarity between i and all observations in C. $b(i)$ is the mean dissimilarity to the observations of the nearest cluster.

In our case, we use a weighted dissimilarity defined as:

$$d(i,j) = \sqrt{\left(\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2\right) * (z_i - z_j)^2}$$

where x is the x-coordinate, y the y-coordinate and z the variable of interest.

$s_{sil}(i)$ ranges between −1 and 1. A Value near 1 indicates that the observation i is appropriately clustered. If *s*(*i*) is close to 0, the correct cluster for the observation i is doubtful. *s*(*i*) near −1 indicates obvious misclustering, it would be more appropriate if i was clustered in its neighbouring cluster .

The average of the silhouette indices defines the overall average of the silhouette widths in the different clusters that make up the score. To determine the number of clusters, the dendrogram is cut at different levels that provide different partitions and then the global average of the silhouettes widths is calculated for each possible partition. The partition that maximizes this quality index is retained.

Total Within cluster Sums of Squares (WSS) measures the degree of homogeneity between belonging to the same cluster. The objective is to obtain clusters that are as homogeneous as possible and thus to make the variability between observations of the same cluster as low as possible. To determine the number of clusters to be selected, WSS is calculated according to the number of selected clusters and then the number of clusters is chosen so that adding another cluster does not improve (decrease) WSS much. In other words, after plotting the WSS vs. number of clusters curve, we observe the location of large reductions (elbow method) in WSS on the curve to determine a number of clusters.

Appendix D: Silhouette plot from the white-beige productions MapClust clustering

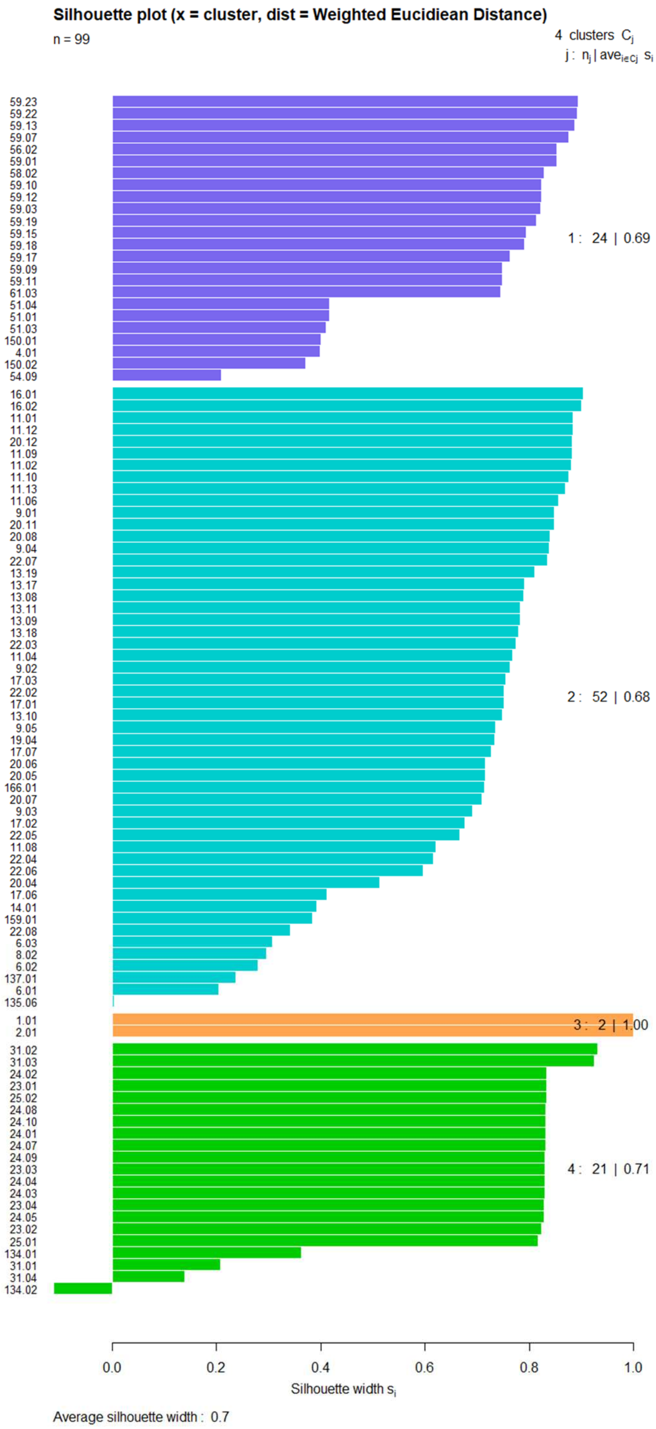

Fig. D.1. White-beige - Silhouette index by MapClust cluster.

Appendix E: Silhouette plot from the multidimensional MapClust clustering

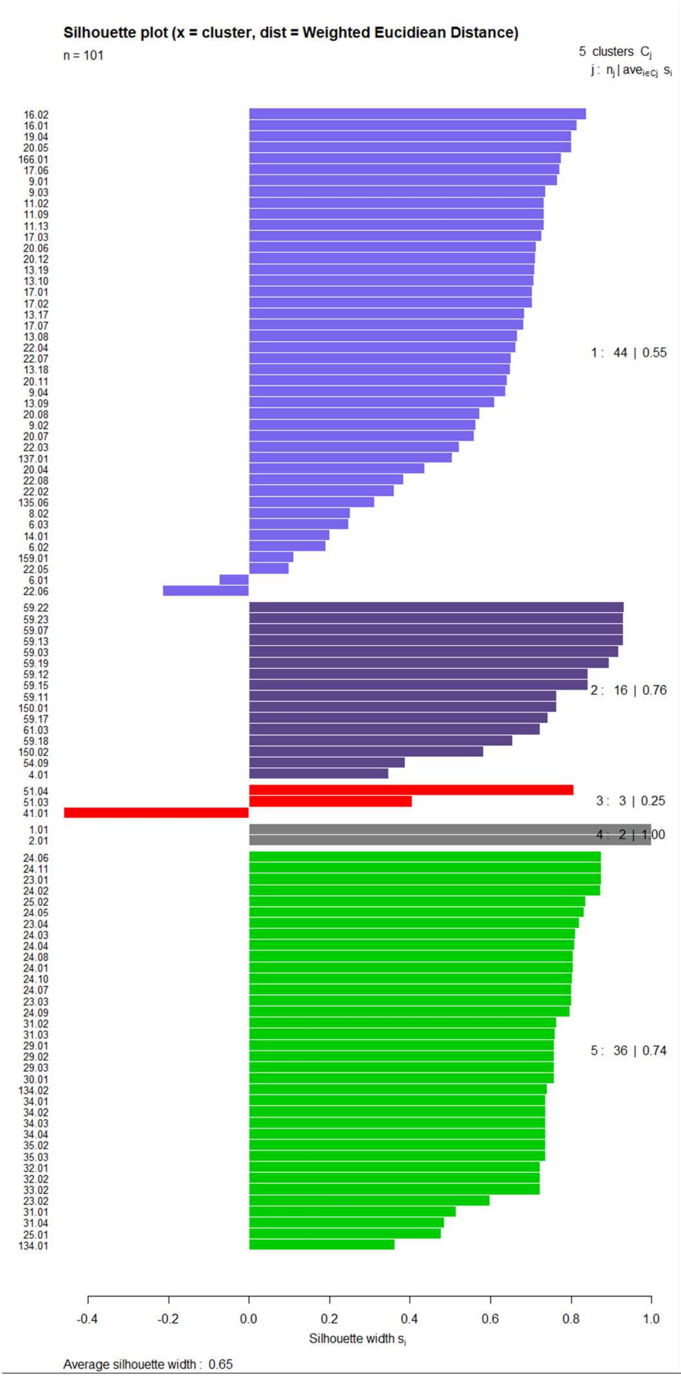

Fig. E.1. Ochre-red & white-beige - Silhouette index by MapClust cluster.

References

Aggarwal, C., Reddy, C., 2014. Data Clustering: Algorithms and Applications. Boca Raton: Chapman and Hall/CRC.

Allan, J., 2020. The importation of French Ceramics to South-West England, c AD 900-1600, In. Henigfeld Y., Husi P., Ravoire F. (eds.), 2020, L'objet au Moyen Âge et à l'époque moderne, Fabriquer, échanger, consommer et recycler, actes du XIe congrès international de la Société d'Archéologie médiévale, moderne et contemporaine (Bayeux, 28-30 mai 2015), CRAHAM, Presses Universitaires de Caen (PUC), Caen, p.456. : 181-196.

Baxter, M.J., 1994. *Exploratory multivariate analysis in archaeology*, Edinburgh University Press, Edinburgh.

Baxter, M.-J., 2006, A review of supervised and unsupervised pattern recognition in archaeometry, *Archaeometry* 48, 4, Oxford: 671- 6.4.

Bellanger, L., Tomassone, R., 2014. *Exploration de données et méthodes statistiques : Data analysis & Data mining avec R*. Collection Références Sciences, Editions Ellipses, Paris. 480 pages.

Bellanger, L., Husi, P., 2012. Statistical tool for dating and interpreting archaeological contexts using pottery, *Journal of Archaeological* Science 39, Issue 4, 777–790 https://doi.org/10.1016/j.jas.2011.06.031

Bellanger, L., Tomassone, R., Husi, P., 2008. A statistical approach for dating archaeological contexts, *Journal of Data Science* (JDS) 6, n° 2, revue en ligne, http://www.sinica.edu.tw/~jds/Content-v-6-2.html

Bellanger, L., Husi, P., Tomassone, R., 2006. Statistical aspects of pottery quantification for dating some archaeological contexts, *Archaeometry* 48, Oxford, 169-183.

Bellanger, L., Husi, P., Laghzali, Y., 2015. Spatial statistic analysis of dating using pottery: an aid to the characterization of cultural areas in West Central France*, in :* Traviglia A. ed., *Across Space and Time, Proceedings of the 41th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA-2013),* Perth (Australie), Amsterdam University Press : 276-282.

Bellanger, L., Coulon, A., Husi, P., 2021. PerioClust: A Simple Hierarchical Agglomerative Clustering Approach Including Constraints. In: Chadjipadelis T., Lausen B., Markos A., Lee T.R., Montanari A., Nugent R. (eds) *Data Analysis and Rationality in a Complex World*. IFCS 2019. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Cham. https://doi.org/10.1007/978-3-030-60104-1_1

Benzecri, J.-P., 1973. L'analyse des Données. Volume II. L'analyse des Correspondances. Paris: Dunod.

Chabert, S., 2019. La céramique arverne Durant l'Antiquité tardive et le haut Moyen Âge (fin IIIe – milieu VIIIe s.), approche chrono-typologique, économique et culturelle, 71eme Supplément à la revue Archéologique du Centre de la France, FERAC, Tours, p. 374.

Coulon, A., Bellanger, L., Husi, P., 2021. SPARTAAS: Statistical Pattern Recognition and daTing using Archaeological Artefacts assemblageS. R package version 1.0.0. https://CRAN.R-project.org/package=SPARTAAS.

Delage, R., 1998. Première approche de la diffusion des céramiques sigillées du Centre de la Gaule en occident romain, *SFECAG*, Actes du Congrès d'Istres : 271-314.

Dervin, S., Bocquet –Liénard, A., 2020. Les aires d'approvisionnement en céramique en Basse-Normandie de la fin du XIIe siècle à la fin du Moyen Âge, In. Henigfeld Y., Husi P., Ravoire F. (eds.), 2020, L'objet au Moyen Âge et à l'époque moderne, Fabriquer, échanger, consommer et recycler, actes du XIe congrès international de la Société d'Archéologie médiévale, moderne et contemporaine (Bayeux, 28-30 mai 2015), CRAHAM, Presses Universitaires de Caen (PUC), Caen, p.456. : 153-166.

Dillman, P., L'Héritier, M., Dissier, A., Leroy, S., Vega, E., 2020. Produire, échanger, utilizer les métaux ferreux au Moyen Âge : nouveaux éclairages interdisciplinaires In. Henigfeld Y., Husi P., Ravoire F. (eds.), 2020, L'objet au Moyen Âge et à l'époque moderne, Fabriquer, échanger, consommer et recycler, actes du XIe congrès international de la Société d'Archéologie médiévale, moderne et contemporaine (Bayeux, 28-30 mai 2015), CRAHAM, Presses Universitaires de Caen (PUC), Caen, p.456. : 197-210.

Everitt, B.S., Landau, S., Leese, M., 2001. Cluster Analysis. London UK: Arnold.

Greenacre, M.J., 2007. Correspondence Analysis in Practice. Chapman & Hall/CRC, BocaRaton, second edition.

Hartigan, J.A., 1981. Consistency of single linkage for high-density clusters. Journal of the American Statistical Association 76(374), 388–394.

Henigfeld, Y. dir., 2021, to be published. La céramique dans les pays de la Loire et en Bretagne de la fin du Xe au début du XVIIe siècle, Presses Universitaires de Rennes (PUC), collection Archéologie, Rennes.

Hodges, R., 1981. *The Hamwih pottery: the local and imported wares from 30 years' excavations at Middle Saxon Southampton and their European context*, cba research Report 37, London.

Holmqvist, E., Larsson, A.-M., Kriiska, A., Palonen, V., Pesonen, P., Mizohata, K., Kouki, P., Raïsänen, J., 2018. Tracing grog and pots to reveal Neolithic Corded Ware Culture contacts in the Baltic Sea region (SEM-EDS, PIXE), *Journal of Archaeological Science* 91, 77-91, https://doi.org/10.1016/j.jas.2017.12.009.

Hurst, J. G., 1969. Red-painted and glazed pottery in Western Europe from the eighth to the twelfth C., *Medieval Archaeology* XIII, 92-146.

Husi, P. (dir.), 2003. *La céramique médiévale et moderne du Centre-Ouest de la France (11^{e} – 17^{e} siècle), chrono-typologie de la céramique et approvisionnement de la vallée de la Loire moyenne,* 20^{e} supplément à la Revue Archéologique du Centre de la France, FERAC, Tours, 110 p. et un cédérom (Ouvrage imprimé épuisé) (Persée : Ouvrage en ligne : www.persee.fr/issue/sracf_1159-7151_2003_ant_20_1 ; Preuves en ligne : http://citeres.univ-tours.fr/cera2003).

Husi, P., 2010. Red-painted and glazed ware of the early medieval period in western France : new data for previous interpretations, an assessment for northwestern Europe*, in :* De Groote K., Tys D. Marnix P. (ed.), *Excavation Medieval Material Culture, Studies on archaeology and history presented to Frans Verhaeghe*, Relicta Monografieën, 4, Archeologie, Monumenten – en Landschapsonderzoek in Vlaanderren, Heritage Research in Flanders. : 81-91.

Husi, P. (dir.), 2013. *La céramique du haut Moyen Age dans le Centre-Ouest de la France : de la chrono-typologie aux aires culturelles*, 49^{e} Supplément à la Revue Archéologique du Centre de la France, FERAC-ARCHEA, Tours, 283 p. (volume papier, et en ligne Persée : https://www.persee.fr/issue/sracf_1159-7151_2013_arc_49_1; Preuves en ligne : http://citeres.univ-tours.fr/cera2013).

Husi, P. (dir.), 2021 to be published. *La céramique médiévale et moderne du bassin de la Loire moyenne (Centre-ouest de la France), transformation des aires culturelles dans le temps long (6^{e} – 19^{e} s.),* Supplément à la Revue Archéologique du Centre de la France, FERAC-ARCHEA, Tours, publication en ligne.

Kaufman, L., Rousseeuw, P. J., 1990. Finding Groups in Data - An Introduction to Cluster Analysis. A Wiley-Science Publication John Wiley & Sons.

Legoux, R., 2010. Diffusion en Île-de-France et en Picardie de la céramique ornée entre la fin du Ve et le VIIe siècle. In, Hincker V., Husi P. (dir.) 2006 - *La céramique du haut Moyen-Age dans le Nord-Ouest de l'Europe, Actes du colloque de Caen, mars 2004,* Éditions NEA, Condé-sur-Noireau, 429, 326-336.

Macnaughton-Smith, P., Williams, W.T., Dale, M.B., Mockett, L.G., 1964. Dissimilarity analysis: A new technique of hierarchical sub-division, *Nature,* **202**, 1034-1035.

Madsen, T., 1988. Multivariate statistics and archaeology, in Multivariate archaeology (ed. T. Madsen), Aarhus University Press, Aarhus, 7–27.

Orton, C. R., and Tyers, P. A., 1992. Counting broken objects: the statistics of ceramic assemblages, in New developments in archaeological science; a joint symposium of the Royal Society and the British Academy, February 1991 (ed. A. M. Pollard), 163–84, Proceedings of the British Academy, 77, published for The British Academy by Oxford University Press.

Pagès, J., 2014. Multiple Factor Analysis by Example Using R (1st ed.). CRC Press, 272 Pages. https://doi.org/10.1201/b17700

Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Stat. 33, 1065–1076.Rand, W. M., 1971, "Objective criteria for the evaluation of clustering methods". *Journal of the American Statistical Association*. American Statistical Association. **66** (336): 846–850. *arXiv*:*1704.01036*. *doi*:*10.2307/2284239*. *JSTOR 2284239*.

R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Avalaible at: https://www.R-project.org/

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Tyers, P. A., Orton, C. R., 1991. Statistical analysis of ceramic assemblages, in *Computer applications and quantitative methods in archaeology* (eds. K. Lockyear and S. Rahtz), 117–20, British Archaeological Reports, International Series 565, Oxford.

Vince, A., 1993. Form, functions and manufacturing techniques of late 9^{th} and 10^{th} century wheelthrown pottery in England and their Origins, In. PITON eds. La céramique du Ve au Xe siècle dans l'Europe du Nord-Ouest, actes du colloque d'Outreau (avril 1992) , n° hors-série de Nord-Ouest Archéologie, 151-164.

Vince, A., 2011. Characterising French whiteware imports in Britain and Northwest Europe. In Bocquet-Liénard A. et Fajal B., A propos de l'usage de la production et de la circulation des terres cuites dans l'Europe du Nord-Ouest autour des XIVe-XVIe siècles, Actes des tables rondes du CRAHM, publication du CRAHM, Caen.

Woillez, M., Poulard, J.C., Rivoirard, J., Petitgas, P., Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES J. Mar. Sci. 64, 537–550.

Woillez, M., Rivoirard, J., Petitgas, P., 2009. Notes on survey-based spatial indicators for monitoring fish populations. Aquat. Living Resour. 22, 155–164.