Unsupervised model-based clustering for typological classification of Middle Bronze Age flanged axes

J. Wilczek a,b,⁎, F. Monna a, M. Gabillot a, N. Navarro c, L. Rusch a, C. Chateau d

a ArTeHiS, UMR CNRS 6298, Université Bourgogne Franche-Comté, 21000 Dijon, France
b Ústav archeologie a museologie, Masarykova univerzita, 602 00 Brno, Czech Republic
c Biogéosciences, UMR CNRS 6262, Ecole Pratique des Hautes Etudes, Université Bourgogne Franche-Comté, 21000 Dijon, France
d UFR SVTE, Université Bourgogne Franche-Comté, 21000 Dijon, France

⁎ Corresponding author at: ArTeHiS, UMR CNRS 6298, Université Bourgogne Franche-Comté, 21000 Dijon, France.
E-mail address: wilczek@mail.muni.cz (J. Wilczek).

1. Introduction

Many types of flanged axes produced in Western Europe during the Middle Bronze Age (1650–1350 BC) have been recognised by archaeologists (e.g. Abels, 1972; Briard and Verron, 1976; Gomez de Soto, 1980; Kläntschi, 1995; Kibbert, 1980; David-Elbiali, 2000; Gabillot, 2003; Michler, 2013). Most types have numerous variants, so that fine typological classification on the sole basis of their shape is generally problematic. The situation is even more complex because typologies generally combine several criteria, such as edge height, the possible presence of ornaments, and the total size of the object, but do not always take all of them into account. These descriptive criteria are not always given the same weight in type definition. Briard and Verron (1976) merged axe types into two generic groups: namely “Atlantic” (Atlantique) and “Eastern” (Orientale). Each of these generic groups, however, is highly polymorphic, so that it is often very difficult for the operator to classify individual axes with absolute confidence and certainty. In order to overcome such problems, a new shape classification is proposed, using morphometric analysis (Elliptic Fourier Analysis) followed by unsupervised model-based clustering and discriminant analysis, both based on Gaussian mixture modelling. Together, these methods produce a clearer pattern, which is independently validated by the spatial distribution of the findings, and multinomial scan statistics. This approach is fast, reproducible, and operator-independent, allowing artefacts of unknown membership to be classified rapidly. The method is designed to be amenable by the introduction of new artefacts, in the light of future discoveries. This method can be adapted to suit many other archaeological artefacts, providing information about the material, social and cultural relations of ancient populations.

© 2015 Published by Elsevier Ltd.
may have served as a means of exchange. Their potential for use as genuine axes cannot be excluded (Nicolardot and Verger, 1998; Delrieu et al., 2015).

The present study aims at systematising the typological classification of these flanged axes. Our approach is based exclusively on object shapes and their treatment by objective statistical techniques, reproducible in time and space by any operator. Since the 1960s, many morphometric methods have been developed. They are based on linear and angular measurements of objects (e.g. Roe, 1968; Hodson, 1971; Barker, 1975; for Bronze Age axes see Lull, 1983), sometimes simplified by deduced categorical variables (e.g. Hodson et al., 1966; Sackett, 1966; Vaginay and Guichard, 1988), and they have proved their worth in archaeological classification. More recently, morphometrics applied to archaeology has evolved into more complex methods including more information (e.g. Brande and Saragusti, 1996; Gilboa et al., 2004; Lycett, 2009; Karasik and Smilansky, 2008, 2011). These methods are known to allow a better description of the entire shape and a separation of shape and size. They provide a continuous morphospace allowing more complex statistical analyses, including the reconstruction of the mean shape and shape diversity within the group of interest (Adams et al., 2004; Navarro, 2003; Zelditch et al., 2004; Slice, 2005; Wilczek et al., 2014). Two recent studies undertaken on Bronze Age palstaves (Forel et al., 2009; Monna et al., 2013) have already demonstrated that combining geometric morphometrics with spatial analyses can be very effective for the better understanding of artefact production and use.

Our first goal was to apply these techniques to closed contours obtained from a corpus of 247 axes (all available as drawings, either in published literature or in personal collections), in an area circumscribed by the French Atlantic coast, the Rhine valley and Switzerland. A new classification approach, based on shape similarities, unsupervised clustering with Gaussian mixture modelling, and discriminant methods, was then developed. The performance of this model was spatially checked using multinomial scan statistics and compared to classifications currently used in the study area. Finally, 21 axe-ingots were introduced into the typological model, for attribution to one of the newly established groups.

2. Material and methods

2.1. Corpus

The choice of the corpus was guided by several constraints: (i) the objects had to be intact and undamaged by use or corrosion, and (ii) their silhouette must not have been drastically reworked after they came out of the mould. Axe preservation was estimated visually from available items, or obtained from the literature (Bocquet, 1970; Abels, 1972; Gomez de Soto, 1980; Kibbert, 1980; Gabillot, 1997, 2003; Nicolardot and Verger, 1998; Mélin, 2012; Gabillot et al., 2014; Thevenot, unpublished). Although the above-mentioned constraints considerably reduced the number of individuals available (approximately 50–60% of available items were kept for further analysis), this selection process is expected to produce robust results. The final corpus consists of 247 reasonably contemporaneous flanged axes (126 from the Atlantic group, and 121 from the Eastern group), discovered in 132 sites, located in what is now France, Switzerland and Germany. Other more specific types, visually very different from the corpus of interest (Fig. 1b), or simply very scarce (e.g. type Strasbourg, Herbrechtingen, Luzern, Riquewihr), were not integrated into the present study. Finally, four generic groups of axes (concave and narrow blades for the Atlantic generic group and Salez and Neyruz types for the Eastern generic group) were retained in the study. The spatial distribution of these axes (Fig. 2) is marked by a clear gap between the two groups, possibly due to the relative absence of archaeological
exploration in that area, but which could also reflect an archaeological reality. This corpus was supplemented by 21 “axe-ingots” (e.g. Nicolardot and Verger, 1998), found in hoards located in the area roughly separating the zones of Atlantic and Eastern flanged axes.

2.2. Morphometrics

2.2.1. Data acquisition, drawings, and outlines

Acquisition follows the procedure described in Forel et al. (2009), Monna et al. (2013) and Wilczek et al. (2014). Briefly, all published images obtained from the available documentation (Supplementary materials S1, Fig. 3ab) were redrawn on tracing paper by a single operator (Fig. 3c). These drawings were then scanned at 300 dpi, and each silhouette was orientated vertically, using the first eigenvector computed on the x-, y-coordinates (Fig. 3d). After this step, each silhouette was subsampled, using 200 equally spaced points, starting from the point possessing the minimum y-coordinate value (Fig. 3e). Original drawings were hand-produced by the same operator, but may differ to some extent from the original object. However, given the number of artefacts processed, it is reasonable to think that any possible awkwardness in the drawings of some specimens will be insignificant for the final results.

2.2.2. Extraction of morphological data

The contour of the axe is taken to be a parametric equation defined as \( x(t) \) and \( y(t) \), where \( t \) is the displacement step along the outline, \( t \in [1, \ldots, 200] \). Decomposition is then performed by Elliptic Fourier Analysis (EFA), producing four coefficients by harmonics \( i \) (i.e. \( A_i, B_i, C_i \), and \( D_i \)). The greater the number of harmonics, the better the reconstruction of the original contour. These coefficients are commonly used as new variables to describe the shape (Kuhl and Giardina, 1982; Lestrel, 1989; Navarro et al., 2004). Normalisation of flanged axes was performed by the major axis of the first harmonic (Kuhl and Giardina, 1982; Rohlf and Archie, 1984; Furuta et al., 1995; Zhan and Wang, 2012), and coefficients were size-normalised, using the square root of the harmonic amplitudes. The first three coefficients of the first harmonic \( (A_1, B_1 \) and \( C_1 \) become constants and can be ignored in further calculations. Note that the fourth coefficient \( (D_1) \) of the first harmonic is retained, as it represents the minor axis of the first ellipse, and contains
Fig. 3. The data acquisition pipeline: a) example of original drawings accessible in the literature (here Kibbert, 1980), b) original drawing of a flanged axe, c) axe outline redrawn by the operator and scanned at a resolution of 300 dpi, d) vertical orientation of the axe, using the first eigenvector computed on the x-, y-coordinates, and e) sampling of 200 equally spaced points along the outline.

Fig. 4. The quality reconstruction of flanged axes, based on an increasing number of harmonics. The original silhouette corresponds to the black line while the shape reconstructed by a given number of harmonics is expressed as a grey polygon. Further calculations were performed using 11 harmonics (see grey boxes).
information about elongation (Iwata et al., 1998; Michaux et al., 2007; Helvaci et al., 2012).

The minimum number of harmonics needed to properly reconstruct shape was investigated by the harmonic power (Lestrel, 1997). First, the power carried by the harmonics i was defined as $P_i$:

$$P_i = \frac{A_i^2 + B_i^2 + C_i^2 + D_i^2}{2}. \tag{1}$$

The reconstruction quality obtained for a given number of harmonics, say $n$, was estimated by calculating the relative cumulative power of the first $n$ harmonics, expressed as a percentage of the total power, i.e. the sum power carried by all harmonics (Reuãd et al., 1999; Helvaci et al., 2012). The relative cumulative power stabilises at a value close to 100%, on average, from 7 harmonics (Supplementary materials S2), but 11 harmonics (corresponding to 41 Fourier coefficients) were finally retained, to ensure a finer reconstruction of artefact shapes in all cases (Fig. 4).

2.2.3. Statistical treatment of morphological data

Fourier coefficients were treated by Principal Component Analysis, PCA (Jolliffe, 2002), a procedure which maximises the variance carried by the first axis. This analysis produces a projection of the artefacts into a low-dimensional, visually friendly, Euclidian space, or morphospace, in morphometric terms. However, one major drawback with PCA is that the representation is strongly affected by the most outlying artefacts, i.e. those contributing most to the variance. This implies that any possible structure within the set of the most common artefacts, that is to say those projecting close to the centre of the PCA space, will be blurred. To circumvent this issue, self-organizing maps (SOMs), a class of neural-network algorithms (Kohonen, 2001), were also computed. These techniques are now increasingly used for data visualisation, clustering and classification of large datasets (Yin, 2008). They allow a representation of a multidimensional dataset by nonlinear projection of artefacts in a lower dimension space, usually represented by discrete locations in a regular 2D lattice. Despite the loss of linearity in the output space, the topological relationships between objects (the order of the distances) are preserved (Liu and Weisberg, 2005). In our case, the use of SOMs can provide a first, exploratory step for further clustering of large datasets (Yin, 2008). Artefacts are assigned to the most similar prototype (also called codebook) vectors, which represent a set of locations summarizing the original data. Interestingly, the density of codebook vectors increases with the density of artefacts, making possible (unlike PCA) a high level of detail in the structure of the data where the artefacts are the most numerous. Practically, the codebook vectors are initiated at random. They are then progressively displaced by an iterative process, following an algorithm reasonably similar to the one applied for the widely used k-means clustering (more details about the procedure can be found in Wehrens and Buydens, 2007; Wehrens, 2011; Kung, 2014). It should be noticed however that different tunings (number of codebook vectors, type of topology used, etc.) will produce different outputs. Several tries are recommended. If a relevant structure exists within the data, the maps produced will tend to be similar.

The final determination of groups was then performed using the increasingly popular Gaussian mixture models (McLachlan and Peel, 2000), where membership association of individuals is probabilistic (Wehrens, 2011). The clusters are assumed to follow multivariate normal distributions here. The point is to find the density of each cluster, their covariance matrix and mean, as well as the conditional probability of membership. This is achieved using a two-step iterative Expectation–Maximisation (EM) algorithm (Dempster et al., 1977), where the conditional probabilities are estimated by the Expectation step and the features of the cluster by the Maximisation step. The optimal number of clusters for classification is determined by likelihood. However, the likelihood is expected to increase with the number of clusters, so that the final decision should be taken by examining a measure penalized by the number of parameters sought. The choice of the number of clusters was made using the Bayesian Information Criterion, or BIC (Schwarz, 1978; Baudry et al., 2010; Wehrens, 2011), and more particularly by close examination of the function:

$$\Delta \text{BIC} = 2(\log L_i - \log L_j) + (p_j - p_i) \log n \tag{2}$$

where $j$ is the number of groups considered and $i$ is this number of groups minus 1, $L$ corresponds to the likelihood, $p$ to the number of free parameters in the model, and $n$ to the sample size. In other words, $\Delta \text{BIC}$ expresses the gain in information when an additional group is considered. The posterior membership attribution, computed using an increasing number of clusters, was also examined. If the model is relevant and artefacts well classified in the clusters thus created, the posterior probabilities should most of the time be close to 1. Discriminant Analysis based on Gaussian finite mixture modelling was then applied to establish group membership for unknown artefacts (here, axe-ingots).

2.2.4. Geographical treatment of morphological data

Axe distributions were mapped by applying a Gaussian kernel function to the location of artefact finds. Map smoothness depends greatly on bandwidth, a parameter of the kernel function: the higher the bandwidth, the smoother the kernel surface (Baxter et al., 1997). Optimal values may be obtained by cross-validation, as recommended by Wand and Jones (1994), but this algorithm tends to create roughness in the estimate. To reduce the risk of overinterpretation, a bandwidth of 65 km was applied, following the procedure of Stevens et al. (2009). Multinomial scan statistics (Jung et al., 2007) were used to complement the geographical mapping of different groups of axes. The aim of this method is to identify non-random spatial patterns (i.e. clusters) in the geographical space. The null hypothesis assumes that the probability of belonging to a given type is the same everywhere. The alternative hypothesis states that, at least for one group, membership probability is not uniform in all parts of the area. Basically, a zone of interest is progressively scanned using a scanning window of increasing diameter, and the number of items in each category is counted, inside and outside the scanning window (Jung et al., 2007). Significant geographical clusters, where type distributions are different from the rest of the map, are identified using likelihood ratio tests. A Monte Carlo procedure allows p-values to be calculated (Jung et al., 2010).

In practice, acquisition and statistical treatments were performed using the set of functions found in Claude (2008), together with the MASS (Venables and Ripley, 2002), mclust (Fraley and Raftery, 2002, 2007; Fraley et al., 2012), monomocs (Bonhomme et al., 2014), kohonen (Wehrens and Buydens, 2007), and ks (Duong, 2007) packages, all written for the free R software (R Core Team, 2014). Multinomial scan statistics and geographical mapping were performed using SatScan v9.3.1 (http://www.satscan.org/; Kulldorff et al., 1998) and Quantum GIS 2.6.1. (http://www.qgis.org; QGIS Development Team, 2015), both of which are freely available software programs.

3. Results and discussion

3.1. Evaluating current typology (Atlantic vs Eastern generic groups)

Axes belonging to the Atlantic and Eastern groups were projected onto a PCA-based morphospace, in which the first two components account for almost 86% of the total variance (Fig. 5). The first principal component (PC1) is characterised by an overall enlargement of flanged axes. The second, PC2, depicts the ratio between cutting edge length and body width. Atlantic and Eastern axes occupy approximately the same area in the PCA-based morphospace, so that neither generic group can be clearly distinguished by this analysis. It should, however, be recalled that PCA is not specifically designed to test the presence of possible
groups, which is why a multivariate analysis of variance (MANOVA) was computed on the Fourier coefficients. The difference between the two generic groups, Atlantic and Eastern, is highly significant \( p < 10^{-15} \). A significant percentage of specimens are a posteriori well classified, when linear discriminant analyses (LDA) are applied, supplemented by either leave-one-out (85%) or two-fold cross-validation (88%). Provided that both generic groups, previously determined by archaeologists, are to some extent validated in terms of statistics, it could be tempting to cease our investigations here. Finer structuration of axe shapes was nonetheless explored, using self-organizing maps, SOMs (Fig. 6), as this procedure may allow local features to be better recognised than by PCA (Wehrens, 2011). The SOM is broadly coherent with the results of the MANOVA and LDA, as there is no major overlap between Atlantic and Eastern axes (Fig. 6a). However, the Atlantic axes are projected on two opposite regions of the SOM, separated by the Eastern group, which does not fit the original interpretation, with only two groups.

3.2. Exploring a new typological system

The evidence from the SOM clearly suggests the possible existence of finer structuration within the two existing generic groups, at least for the Atlantic corpus. Identifying this internal structure deserves more attention, and this hypothesis was therefore explored with Gaussian mixture models. A VEI model (varying volumes, equal shapes, and identity orientation of clusters) was selected by applying a \( \Delta \text{BIC} \) procedure. It appears that, after 6 clusters, there is no notable improvement in terms of information (Supplementary materials S3b), so that further modelling was performed using this value. For each flanged axe, the probability of belonging to each of the six clusters formed was also

![Fig. 5. Principal Component Analysis. Projection of the 247 items in a PC2 vs PC1 morphospace. The first two principal components together account for almost 86% of the total variance. Grey shapes, reconstructed by the inverse Fourier transform, represent 64 virtual axes.](image-url)

![Fig. 6. Projection of 247 flanged axes on a self-organizing map. Artefacts belong to a) the Atlantic generic group (blue dots) or the Eastern generic group (red triangles); and b) one of the six groups obtained by unsupervised model-based clustering. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
computed. If the clusters are properly defined, the probabilistic attribution of items to groups should not be ambiguous. As a result, the maximum membership probability should be close to 1. Out of a total of 247 items, 224 have a maximum probability above 0.99, with 240 achieving a probability of 0.90 (Supplementary materials S3a). Such results suggest that the six groups constructed do not strongly overlap in terms of shape. This result is confirmed by a model-based DA, with leave-one-out cross-validation, in which none of the specimens was erroneously classified. Interestingly, with six groups, the SOM also depicts clearly separated items (Fig. 6b). To reinforce cluster validation, two complementary analyses were performed for each group: (i) artefacts were plotted on the same graph, together with the mean shape, resulting in clearly visible differences between the six groups (Fig. 7), and (ii) the location where each item was found was mapped, to investigate possible geographical clusters (Fig. 8). The multinomial scan statistics show non-random spatial distribution. Four circular zones (clusters) were detected, in which the distribution of items (in terms of representation of groups) was significantly different from the rest of the map \((p < 0.001)\). Zone A corresponds to a greater relative abundance of the flanged axes belonging to G1 and G2, zone B has a greater number of axes from G3, zone C contains more axes from G3 and G4, while D, the final zone, encompasses most of the items from G5 and G6 (Fig. 8, Table 1).

The statistical analysis therefore confirms the existence of two traditions of flanged axes, corresponding to the Atlantic and Eastern generic groups mentioned above. Nevertheless, within this classification, a more complex internal organisation was discovered: the Atlantic axes can be subdivided into four groups (G1, G2, G5, and G6), while the Eastern axes form two groups (G3 and G4). This result explains why the initial linear discriminant analysis was able to differentiate between Atlantic and Eastern axes so efficiently.

The first two groups (G1 and G2) are composed of similar axes, characterised by a thin, rectangular body, and a rather linear cutting edge; both forms, but more particularly G2, are very stable in terms of shape (Fig. 7). They are distributed almost exclusively in the western part of France, close to the Atlantic coast (Fig. 8, cf zone A). Given the shape homogeneity of G2, the statistical analysis tends to consider these axes as belonging to a separate group, rather than attributing them to G1. Most of these axes could result from a homogeneous unit of production, explaining why they exhibit such low shape diversity. The presence of G2 might also result from the structure of the corpus taken into account, since almost all artefacts belonging to group G2 (36 out of 44) were discovered at a single location. It is worth noting that zone A corresponds to cultural entities already recognised in central western France (i.e. the Duffaits and Vindo-Médocain groups; Gomez de Soto, 1995). The vast majority of the axes belonging to G1 and G2 were previously described as narrow-blade flanged axes (Table 2).

The axes from groups G5 and G6 exhibit the largest cutting edges and are the most heterogeneous in terms of shape (Fig. 7). They are concentrated in north-western France (more particularly in the Seine valley — zone D), and to some extent in Switzerland and southern Germany. Spatially, almost no items from G5 or G6 were discovered in zone A, while no artefacts from G1 or G2 were found in zone D.
Fig. 8. Kernel density maps for the six groups of flanged axes obtained by model-based clustering. Numbers correspond to the number of axes found in one location, while dashed circles represent the clusters (A–D) provided by scan statistics (purely spatial scan statistics under a multinomial probability model, circle scanning and Monte Carlo randomisation with 999 permutations). Red numbers in squares represent the number of axe-ingots attributed by probability to one of the six clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
occupying this geographical area therefore used a well-de-
site of Salez (Switzerland) are members of G4. Bronze Age communities
of the 25 analysed from the eponymous
types. Moreover, 19 artefacts out of the 25 analysed from the eponymous
types. Axes in G4 present good shape stability, and their distribution is
tween cultural entities, but could also be viewed as noise in the data.
France (Fig. 8). These axes (G3) may indicate long-distance contact be-
the north-eastern part for G4 (zone C). Some items attributed to G3
the entire territory for G3 (zone B and C), and almost exclusively in
found in an area approximately corresponding to Switzerland: over
is especially true of the G4 axes (Fig. 7). Most of these artefacts were
the outer shape of the flanged axes suggests a rel-
ating axes (Table 2).
The axes in groups G3 and G4 are the most abundant. Their summits,
flanks and cutting edges are more rounded than those in G1 and G2; this
is especially true of the G4 axes (Fig. 7). Most of these artefacts were
found in an area approximately corresponding to Switzerland: over
the entire territory for G3 (zone B and C), and almost exclusively in
the north-eastern part for G4 (zone C). Some items attributed to G3
can also be found in what is now Germany and in the western part of
France (Fig. 8). These axes (G3) may indicate long-distance contact be-
tween cultural entities, but could also be viewed as noise in the data.
Table 2 shows that axes from G3 correspond to Neyruz and Salez
types. Axes in G4 present good shape stability, and their distribution is
limited to zone C. All artefacts from G4 belong exclusively to the Salez
type. Moreover, 19 artefacts out of the 25 analysed from the eponymous
site of Salez (Switzerland) are members of G4. Bronze Age communities
occupying this geographical area therefore used a well-defined type of
axe (G4), which is statistically different from other types, at least in
shape. As G4 axes have not been discovered elsewhere (except for
some rare occurrences), it is reasonable to assume that they were not
massively exported.
Using the discriminant analysis based on Gaussian finite mixture
modelling, each of the 21 ingots can be associated with one of the six
previously defined groups. Most of the axe-ingots were attributed to
G3, but two artefacts were classed as G1, while the remainder were
associated to the most heterogeneous group, G5. This result is not
surprising, as most of the axe-ingots were found close to zone B, where the flanged G3 axes were the most abundant (Fig. 8).
The low shape disparity within groups of flanged axes suggests a rel-
atively well-established production pattern. Metalworkers produced
axes following set technological rules of fabrication, adapted to the so-
cial and cultural standards of Bronze Age communities (Monna et al.,
2013; Gabillot et al., in press). It is tempting to think that the zones
with the greatest number of axes represent the areas of indigenous pro-
duction. However, at least for the corpus studied here, it cannot formally
be decided whether artefacts were produced (i) in one particular place
and then transported to different locations, or (ii) in different places,
by different metalworkers, all following the same shape and style (Orton,
1980). Except for a few examples (e.g. in France, Gabrières, Ambert
and Barge-Mahieu, 1991; Saint-Veran, Barge, 2005), workshops and
mines providing raw materials have not yet been identified in the ar-
chaeological record, so that production and distribution processes are
still debated. In any case, if a strong spatial structure can still be
recognised today, it is because exchanges at the scale of the study area
were likely to be limited, as otherwise the distribution of artefacts
would have been more homogeneous.

4. Conclusion

The current classification for Atlantic and Eastern flanged axes, un-
systematically mixing several criteria (overall shape, supposed area of
distribution, presence of decoration, etc.) was statistically validated on
the outer shape of the flanged axes alone. However, the internal struc-
ture of these two groups can be much better identified by using a com-
bination of modern computational methods. Unlike the traditional
approach, the proposed classification is statistically established from
unsupervised, model-based clustering analyses of morphometric data,
and then validated independently by close examination of the spatial
patterns. There is no need to know a priori the proper classification
model, nor the number of groups. Artefacts of unknown membership
can be classified, as was the case here for axe-ingots. This approach is
quick, reproducible, operator-independent, easy to implement (as the
tools used are freely available), and straightforward to adapt for almost
any type of object where classification on the basis of shape is sought.

Unsupervised classification methods nevertheless have several limi-
tations. Clustering basically aims to attribute all artefacts to an optimal
number of groups. The issue is that what is considered as optimal today
does not necessarily reflect the reality of the field in the past. Model-
based clustering approaches need to have a sufficient number of items
in order to identify individual groups correctly. In archaeology, where
well-preserved artefacts may be scarce, some groups may have existed,
but their representatives may not have survived in sufficient quantities.
Model-based clustering will identify these items as extremes, but with-
out defining a proper group for them, whereas a skilled archaeologist
might be able to individualise them, and to treat them separately. None-
theless, identifying them as extremes may allow a shift in focus for those
specific artefacts.

Here, the optimal number of groups was determined using ΔBIC: i.e.
$n = 6$ groups. It is worth mentioning that the operator must pay great
attention to this parameter, because, with an inappropriate number of
groups, classification may produce meaningless output. For example,
in our case, $n = 5$ could reasonably have been considered (Supplemen-
tary materials S3), but the output would not have been as finely differ-
tiated, because the axes classified as G5 and G6 would simply have
been merged together. Note that if the $n$ was lower, say 4 or 3, the ob-
erved spatial patterns might have not been identified. In general,
increasing $n$ above the optimal value is therefore not as unfavourable as
underestimating $n$. In any case, the determination of meaningful groups
in the dataset is a problem shared by all classification techniques
(e.g., Orton, 1980; Legendre and Legendre, 1998).
The methodology proposed here is unequivocally statistical: the
more data available, the better the output, meaning that this method

### Table 1
Concordance between model-based clustering and clustering performed by multinomial scan statistics. Bold font highlights the concordance between both methods.

<table>
<thead>
<tr>
<th>Model-based groups</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zones (Multinomial clusters)</td>
<td>A</td>
<td>15</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>29</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>38</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>18</td>
<td>33</td>
</tr>
<tr>
<td>Not associated</td>
<td>4</td>
<td>1</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>44</td>
<td>62</td>
<td>47</td>
<td>27</td>
<td>46</td>
</tr>
</tbody>
</table>

### Table 2
Correspondence between unsupervised typology and current archaeological classification. Bold font highlights the correspondence between both methods.

<table>
<thead>
<tr>
<th>Model-based groups</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaeological classification</td>
<td>Narrow-blade flanged axes</td>
<td>17</td>
<td>41</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Concave-blade flanged axes</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>32</td>
</tr>
<tr>
<td>Neyruz type</td>
<td>1</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Salez type</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>47</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>44</td>
<td>62</td>
<td>47</td>
<td>27</td>
<td>46</td>
</tr>
</tbody>
</table>
is not suitable for too small a corpus. Artefact assignment is probabilistic, so that it does not formally answer the question of strict attribution to a single type, except when the highest probability is considered. When the classification is based on this highest probability, subsequent analyses take for granted this classification and therefore no longer consider the level of uncertainty associated with it. It should be mentioned that the archaeologist experimenting typological attributions based on visual observation alone, faces the same restrictions, but unfortunately without any probabilistic control.

The groups defined by this approach cannot be thought of as definitive. Any new artefact added to the dataset will marginally modify the output. By its very nature, the proposed model for creating and validating the classification is one stage in a never-ending process. Any new artefact specimen or other observed descriptor for the artefact (e.g. decoration, chemical composition, or fabrication process) can test the stability of the proposed model and may thus contribute to its improvement.

It is noteworthy that the morphometric approach proposed here takes into account only the shape of the external outline of the axe. Although this feature was probably of some significance in ancient societies, as indicated by the spatial patterning found here, modern acquisition techniques allow researchers to develop better integrated descriptions of artefact shape variation, using 3D models, and not just focusing on the a priori more informative side-view of the artefact. Such features can be used and treated in similar ways.

The combination of this approach, applied to several types of objects, with other sources of information might contribute to a better understanding of the material, social and cultural relationships in ancient populations.

Supplementary data for this article can be found online at http://dx.doi.org/10.1016/j.jasrep.2015.06.030.

Acknowledgements

We are very grateful to M. Mélin and J.-P. Thevenot for allowing us access to their personal archives and two anonymous reviewers for their constructive comments.

References


Beaulieu, J., 1963. Bronze age connections across the North Sea, a study in prehistoric trade and industrial relations between the British Isles, the Netherlands, North Germany and Scandinavia, 1700–700 B.C. Palaeohistoria IX (Groningen, 286 pp.).


Gabillot, M., 1997. Le Bronze moyen en région Centre (Master thesis), Université de Bourgogne, Dijon (199 pp.).


Hodson, F.R., Sneath, P.H.A., Doran, J.E., 1966. Some experiments in the numerical analysis of axes, and industrial relations between the British Isles, the Netherlands, North Germany and Scandinavia, 1700–700 B.C. Palaeohistoria IX (Groningen, 286 pp.).


Thevenot, J.-P., unpublished study. Etude du dépôt du Bronze moyen du Bois de Mont Genèvre à Vic-de-Chaisenay (Côte-d'Or).


