Journal of Archaeological Science: Reports 37 (2021) 102891

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

A computer tool to identify best matches for pottery fragments

Josef Wilczek [a,b,*], Fabrice Monna [b], Nicolas Navarro [c,d], Carmela Chateau-Smith [e]

[a] Department of Archaeology, University of Hradec Králové, 50003 Hradec Králové, Czech Republic
[b] ARTEHIS, UMR CNRS 6298, Université Bourgogne Franche-Comté, 21000 Dijon, France
[c] Biogéosciences, UMR CNRS EPHE 6282, Université Bourgogne Franche-Comté, 21000 Dijon, France
[d] EPHE, PSL University, 75014 Paris, France
[e] CPTC, EA 4178, Université Bourgogne Franche-Comté, 21000 Dijon, France

ARTICLE INFO

Keywords:
Archaeology
Best match
Automation
Pottery
ICP (Iterative Closest Point)
DCT (Discrete Cosine Transform)
RDP (Ramer-Douglas-Peucker)
RTC (Radius Tangent and Curvature)

ABSTRACT

Archaeologists working with pottery spend a considerable amount of time on a fundamental task – providing precise descriptions of pottery fragments. This study presents a survey of existing computational solutions to identify the best matches for a given fragment, based on its shape. Four methods (ICP, DCT, RDP, and RTC) are compared, using a pottery dataset from Graufesenque (southern France), dated to the Roman Period. The first three methods produced successful and very similar results for rim fragments (within the five best candidates for 95% of the dataset). The ICP algorithm produced the best overall results for rim fragments, and can also be used for non-rim fragments. A practical computer application, including all the above methods, was developed in R programming language, with an easy-to-use graphical interface, and is now made freely available to the archaeological community for future studies, and further development.

1. Introduction

Pottery is one of the most abundant materials present in archaeological excavations. It provides information about chronology, and the evolution of technique and style. It may also provide evidence of social organisation. Unlike precious artefacts belonging only to the elite, ceramics have been used by all social strata. The socio-economic dynamics of ancient populations can be reconstructed, based on pottery features: clay, fabrication technique, shape, decoration, spatial distribution, discovery context, etc. (e.g. Buko, 1990; Orton, 1980; Orton et al., 1993; Rice, 1987). However, pottery is fragile, and can be damaged or even destroyed by post-depositional processes, thus reducing the information available for archaeological inquiries.

Among archaeological investigations, one strategy is to use a set of descriptors (e.g. clay, colour, decoration, etc.) with pre-defined classes for the precise characterization of pottery fragments. Here, the main focus is on shape descriptors, which are generally related to period, origin, function, and/or aesthetics (e.g. Bahn and Renfrew, 2015).

The specialist often tries to find the best match for a given fragment from within a well-established classification system, which can be composed of morphological classes (e.g. Macháček, 2001; Ness, 2015; Vaginay and Guichard, 1988; Venclová, 1998, 2001). This attribution process can be time-consuming, particularly when processing thousands of fragments. To overcome these problems, many quantitative methods have been proposed for the automatic retrieval of pottery fragments.

Almost all archaeological pottery vessels can be considered to be rotationally symmetrical. These 3D objects are thus usually represented by a 2D profile, corresponding to any cross-section between the vessel and the plane passing through its rotational axis. Quantitative methods are therefore often based on the calculation of similarities between 2D profiles, usually represented by polar or cartesian coordinates (Liu et al., 2005; Maiza and Gaildrat, 2005), or expressed as a function of radius (e.g. Jičín and Vašíček, 1971; Karasik and Smilansky, 2011; Mom, 2005; Smith et al., 2014), or its derivatives, such as tangent (e.g. Gilboa et al., 2004; Karasik and Smilansky, 2011; Leese and Main, 1983; Main, 1987, 1986; Saragusti et al., 2005; Smith et al., 2014), or curvature (Gilboa et al., 2004; Hristov and Agre, 2013; Karasik et al., 2005; Saragusti et al., 2005). Profiles may sometimes be represented by b-spline coefficients associated with segments of the profile curve (e.g. Adler et al., 2002; Hlaváčková-Schindler et al., 2001; Kampel and Sablatnig, 2007, 2003, 2002, 1999; Laflin, 1986; Schurmans et al., 2001), by polylines (Lucena et al., 2016), or by dominant feature points that are extracted using the medialness measurement (Piccolli et al., 2015). Similarities between two pottery fragments can be calculated from Euclidean distances

* Corresponding author at: Department of Archaeology, University of Hradec Králové, 50003 Hradec Králové, Czech Republic.
E-mail address: josef.wilczek@uhk.cz (J. Wilczek).

https://doi.org/10.1016/j.jasrep.2021.102891
Received 18 June 2020; Received in revised form 10 February 2021; Accepted 18 February 2021

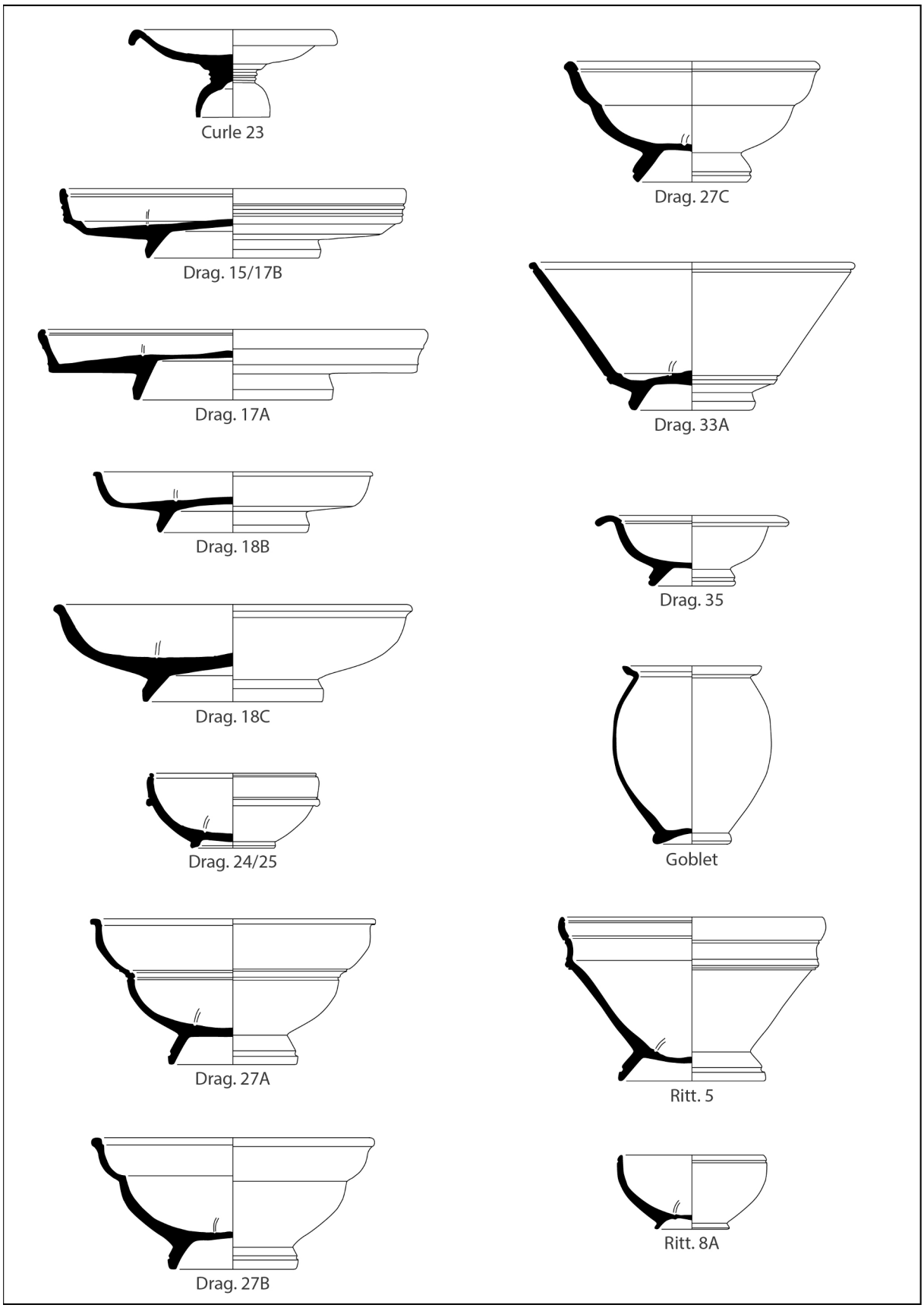

Fig. 1. Morphological sub-groups used in this study. Abbreviations: Drag. – Dragendorff, Ritt. – Ritterling. Scale 1/2.

Table 1
Test corpus used in the study (from Genin et al., 2008). For morphological class and sub-group codes: C – Curle, G – Goblet, D – Dragendorff, R – Ritterling. See Fig. 1 for corresponding images.

Class	C23	D15/17		D18		D24/25	D27			D33	D35	G	R5	R8
Sub-group	C23	D15/17B	D17A	D18B	D18C	D24/25	D27A	D27B	D27C	D33A	D35	G	R5	R8A
Total	10	21	11	52	11	61	10	49	11	15	22	15	17	14

between their profiles (Maiza and Gaildrat, 2005), or from the similarity of the coefficients expressing their shape (Gilboa et al., 2004). The best match for a given fragment (and hence its morphological class) is obtained by minimising differences with potential candidates in a referential database.

The above-mentioned approaches have already been tested on real-world artefacts, with a high success rate for classification (Karasik and Smilansky, 2011), and for best-match retrieval (Lucena et al., 2016). It should, however, be noted that several major issues have not yet been fully addressed. For example, if the position of the outline in relation to the rotational axis is not appropriately constrained, the shape of the pottery fragment or vessel may be drastically deformed, leading to misidentification. Methods requiring successive derivatives generally suffer from numerical instability (Karasik and Smilansky, 2008). Although the idea underlying morphological correspondence between profiles appears simple, achieving perfect correspondence between individual points on profiles is challenging, even more so when working on fragments. Many approaches have been evaluated solely on the outer surface of the profile, or only on rim fragments. Although the mathematical basis of all these approaches has already been explored (e.g. Hristov and Agre, 2013; Liu et al., 2005; Mom, 2005; Mom and Paijmans, 2007; Smith et al., 2014), no practical solution is currently available to the broader archaeological community for routine use. It is important for such proposals to be tested on a wider variety of artefacts, as no method is likely to be universally applicable.

This paper implements supervised classification, i.e. the attribution of a given fragment to one of the predefined classes. The aim of this study is to compare three existing approaches for the mathematical matching of pottery fragments based on morphology, together with a new method based on Discrete Cosine Transform (DCT). The goal is not to reconstruct complete vessels, but to identify the best matches to the fragment within the referential dataset, by shape similarity, thus indicating which shape label or labels would best suit the fragment. Attribution accuracy is evaluated on an already labelled real-world dataset. All the approaches tested are made available as a set of functions encoded in R programming language (R Core Team, 2021). An easy-to-use graphical interface was also developed, using Shiny GUI (Chang et al., 2019), and is made freely available to the archaeological community, to simplify data retrieval.

2. Material and methods

2.1. Corpus

The collection of ceramic vessels found at Graufesenque (southern France), dating from the Roman period (first to mid-second century AD), is published as an illustrated paper catalogue (Genin et al., 2008). The fact that vessels in this catalogue are already labelled by a widely

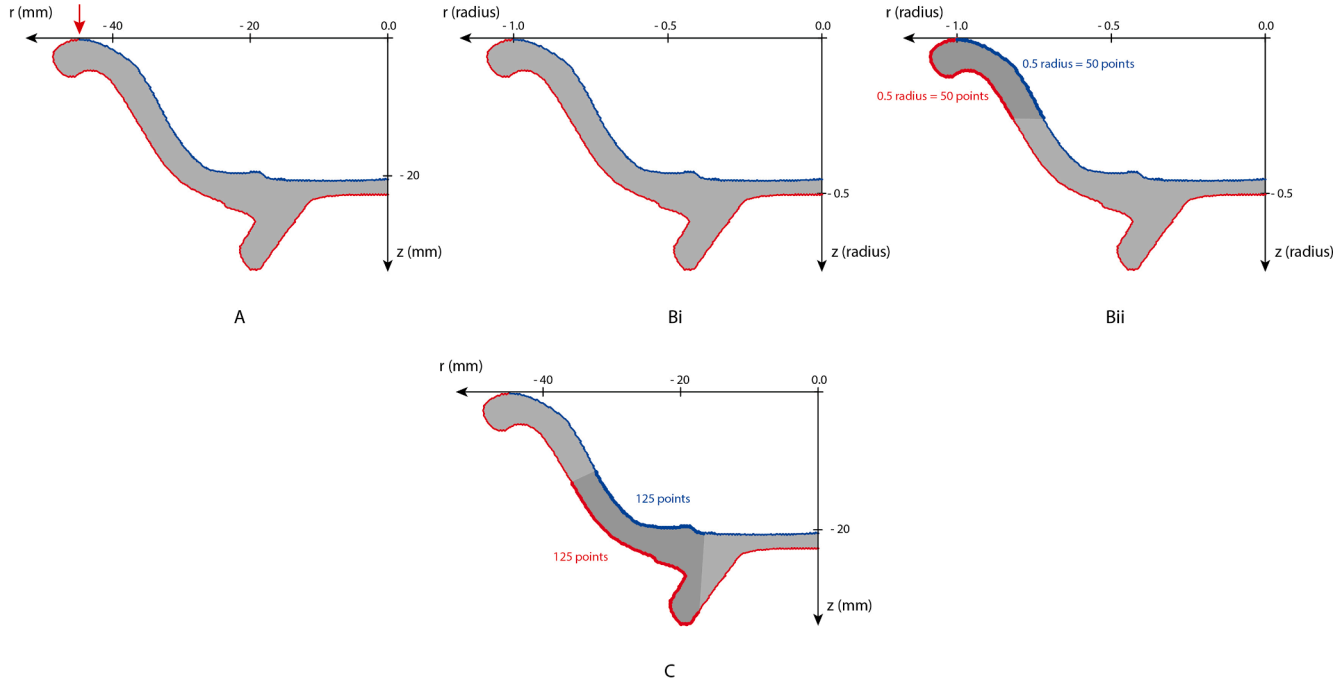

Fig. 2. Profile acquisition and data preparation. A) Scaling the outline to the original size of the vessel, positioning it according to the axis of symmetry (i.e. z axis), and r axis, and segmenting it to the outer (red) and inner (blue) segments using the rim (red arrow) as a reference point. Bi) Size normalisation of the profile by baseline registration (here the rim point and the point of intersection between rim plane and the axis of the symmetry are set to (−1, 0) and (0, 0) respectively; note that the coordinate system is now expressed in rim radius units). Bii) Virtual extraction of the rim fragment: the outer and inner part of the profile is virtually broken at the distance of 0.5 rim radius along the profile from the rim point, and resampled by 100 equally spaced points, 50 for the outer and 50 for the inner part. C) Virtual extraction of the non-rim fragment: the outer and inner part of the profile are extracted approximately in the middle of the vessel, and resampled by 250 points, with 125 points equally spaced along the outer and inner parts.

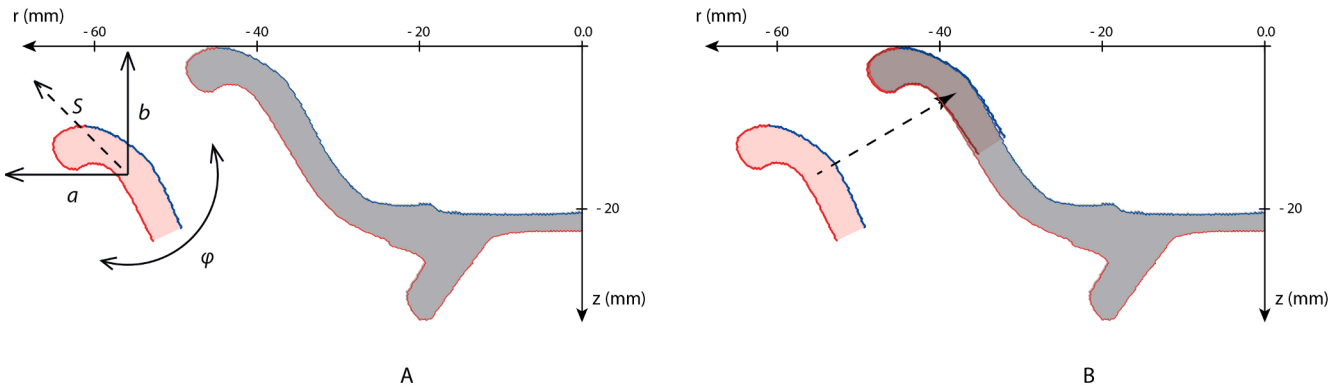

Fig. 3. Alignment process between source fragment (red polygon) and complete target vessel (grey polygon). A) The fragment is translated along the r and z axes (*a*, *b*), rotated (φ), and scaled (*S*) to minimise the distance with the complete vessel. B) Final alignment of the source on the target.

adopted morphological classification scheme (e.g. Brulet et al., 2012; Passelac and Vernhet, 1993; Py et al., 1993; Schucany et al., 1999; Tyers, 1996) makes it suitable for method comparison. The test corpus identified within this vast collection contains complete vessels only, and includes all morphological classes (and sub-groups) represented by ten or more vessels in the catalogue. The test corpus is thus composed of 319 vessels, including plates, bowls, cups, and goblets (Fig. 1; Supplementary Materials SM1), already divided into 14 sub-groups, belonging to 10 morphological classes (Table 1).

2.2. Profile acquisition and data preparation

The drawings of the 319 vessels thus selected were scanned at 600 dpi from the paper catalogue. Profile outlines were extracted using a modified set of functions described in Claude (2008), and positioned according to their rotational axis. The outlines were automatically divided into outer and inner segments, using the rim as a reference point (Fig. 2A).

To evaluate the potential of the four methods under comparison, two test subsets of synthetic data (rim fragments and non-rim fragments) were then produced from these outlines, by virtually fragmenting the profiles of the original complete vessels. Rim fragment analysis was based only on shape, and not on size.

For rim fragments, all outlines were size-normalised using baseline registration (Bookstein, 1991), sending the rim point and the point of intersection between rim plane and the rotational axis to the (−1, 0) and (0, 0) coordinates (Fig. 2Bi). Rim fragments were then obtained by virtually cutting the outer and inner profile parts at a distance of 50 percent of the rim radius from the rim point, expressed by 100 equally spaced points (50 for the inner and 50 for the outer segment; Fig. 2Bii). Virtual rim fragments therefore measured 11–35% of the profile length of the original complete vessel.

Non-rim fragments, representing 40–45% of the profile length of the original vessel, were extracted approximately from the middle of the profile (Fig. 2C). Each non-rim fragment was expressed by 250 points (125 each for outer and inner segments).

2.3. Best match searching algorithms

2.3.1. Iterative Closest Point (ICP)

The Iterative Closest Point algorithm (ICP; Liu et al., 2005; Maiza and Gaildrat, 2005) is widely applied in graphics and computer vision for 3D model alignment (e.g. Besl and McKay, 1992; Fitzgibbon, 2003; Turk and Levoy, 1994). To identify the best match, a complete profile (target), is positioned in the coordinate system, and the fragment to be aligned (source), is iteratively translated, scaled, and rotated (Fig. 3A), to minimise (using a Simulated Annealing algorithm; Bélisle, 1992; Cortez, 2014; Hajek, 1988) the sum of root-mean-square deviations (RMSDs) of the distances between the points of the source and those of the target (Fig. 3B):

$$\min_{a,b,\varphi,S}\left(\sqrt{\frac{1}{M}\sum_{j=1}^{M} d(P_j,T_i)^2}+\sqrt{\frac{1}{N}\sum_{k=1}^{M} d(P_k,T_o)^2}\right)$$

For the source profile, *a* and *b* correspond to translation along the r and z axes, φ to rotation, and *S* to scaling; *M* and *N* are the total number of points on the inner and outer segments of the source profile. The expression $d(P_j,T_i)$ is the distance between the *j*-th point, *P*, on the source segment and its closest point on the inner target segment, T_i. The expression $d(P_k,T_o)$ is the distance between the *k*-th point, *P*, on the source segment and its closest point on the outer target segment, T_o.

Note that scaling (*S*), and translation along the z axis (*b*) are the only rigid transformations that do not alter the original shape of the vessel, unlike *a* and φ, which are strongly constrained by the position of the rotational axis. Ancient pottery vessels are often considerably fragmented, and were not always perfectly regular, thus making it difficult

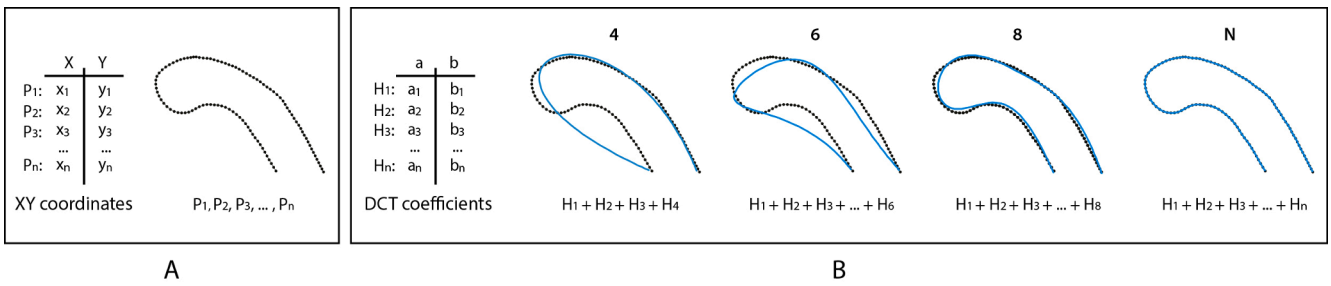

Fig. 4. Discrete Cosine Transform. A) The original cartesian coordinates of the outline. B) The original cartesian coordinates of the outline are decomposed into a set of harmonics. A given number of harmonics can be used to reconstruct the approximation of the original contour (here 4, 6, 8, and a full set of harmonics were used for reconstruction).

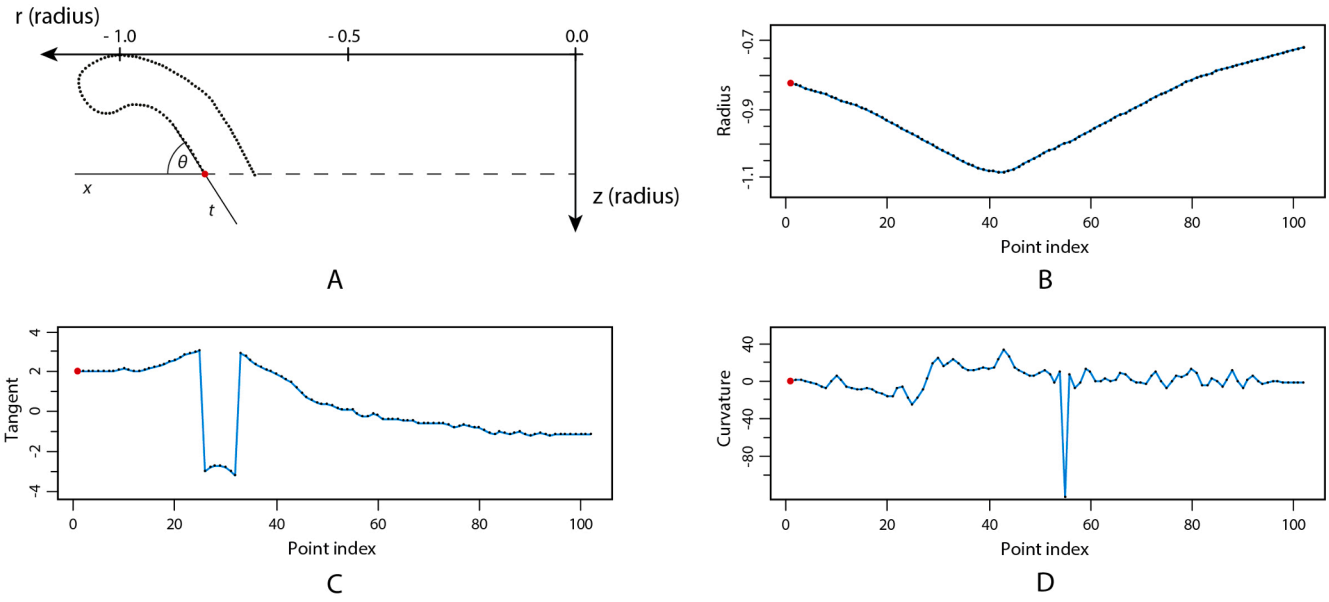

Fig. 5. Radius, Tangent, and Curvature. The first point on the outline and its corresponding representations are highlighted in red. A) Position of the fragment according to rotational axis and tangent angle calculation. The tangent angle θ of the point is calculated as an angle between the tangent line passing through the point, *t*, and the line parallel to the rim, *x*. B) The radius function represents the distance between each point on the profile and the rotational axis. C) Tangent function represents the angle between the tangent of each point on the profile and the line parallel to the rim (see A for details). D) Curvature function represents the rate of change of the tangent function.

to determine the precise rotational axis to be used for profile extraction. In real-life situations, the researcher may therefore decide to relax the constraints of the rotational axis position a little, by allowing α and φ to vary to some extent, although some shape modification will occur.

2.3.2. *Discrete Cosine Transform (DCT)*

Outlines can also be treated by Discrete Cosine Transform (DCT), used in signal treatment for compression and de-noising. This Fourier-based method transforms open outline coordinates (Fig. 4A) into a set of harmonics (Fig. 4B), each represented by a pair of coefficients that may be used as shape variables (e.g. Dommergues et al., 2007; Forel et al., 2009). The higher the number of coefficients, the more precise the reconstruction of the outline (Fig. 4B). Shape information can generally be preserved with a low number of harmonics (Hurth et al., 2003; Schmittbuhl et al., 2003). The similarity between two profiles is then expressed as the RMSD between their corresponding coefficients.

The calculation of DCT is strongly constrained by the position of the rotational axis. Thus the researcher does not control variation of position, size, and orientation of the source fragment, as applied in ICP. Note that these facts also concern the two remaining algorithms: radius, tangent, and curvature, and Ramer-Douglas-Peucker polyline.

2.3.3. *Radius, Tangent, and Curvature (RTC)*

The profile curve (Fig. 5A) can be represented by three mathematical functions: radius, tangent, and curvature. The radius function of the profile outline represents the distance between the rotational axis and each point on the profile (Fig. 5B). The tangent function represents the angle between the tangent of each point on the profile and the rotational axis (Fig. 5C). The curvature function represents the rate of change of this angle (Fig. 5D). The similarity between two profiles can then be obtained by calculating the sum of the RMSDs of these three functions, weighted to stress the relative importance of different parts of the profiles if required (e.g. Adan-Bayewitz et al., 2009; Hristov and Agre, 2013; Karasik and Smilansky, 2011; Smith et al., 2014).

2.3.4. *Ramer-Douglas-Peucker (RDP) polyline*

Pottery profile coordinates can also be approximated by a polyline with a fixed number of segments, obtained with the Ramer-Douglas-Peucker (RDP) algorithm (Douglas and Peucker, 1973; Lucena et al., 2016). Given a profile curve with endpoints *a* and *b*, this algorithm seeks the most distant point on the curve from the segment *ab*. Once this point, *c*, is located, the segment *ab* is then replaced by two new segments, *ac* and *cb* (Fig. 6A). The entire procedure is then repeated until the desired number of segments (*L*) is obtained on the profile outline (Fig. 6B–D). The points generated in this way then serve as shape variables. Similarity between

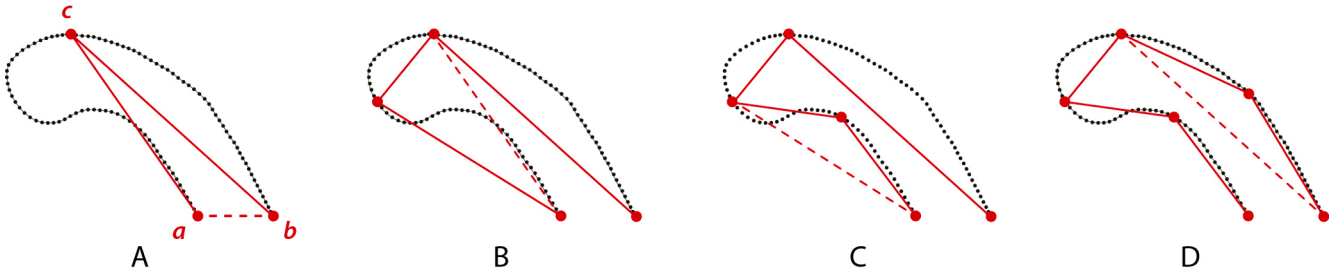

Fig. 6. Ramer-Douglas-Peucker polyline algorithm. A) Given a profile curve (black dots) with endpoints *a* and *b* (red dots), this algorithm seeks the most distant point on the curve from the segment *ab* (dashed red line). Once this point, *c*, is located (red point), the segment *ab* is replaced by two new segments, *ac* and *bc* (red lines). B–D) The procedure is repeated until the desired number of segments is obtained (here 5 segments).

Table 2
Parameters and ranges of acceptable values used to evaluate algorithms. See Supplementary materials SM2 for visualisation of the reconstruction quality for 20 DCT harmonics and 20 RDP polyline segments.

Algorithm	Rim fragments	Non-rim fragments
Iterative Closest Point	$a \in (-0.1, +0.1)$ $\varphi \in (-3°, +3°)$ $S \in (-2\%, +2\%)$ a values are expressed in rim radius units	$a \in (-1\ cm, +1\ cm)$ $\varphi \in (-3°, +3°)$ $S \in (-2\%, +2\%)$ $b \in (0\ cm, T_h\ cm)$ a and b values are expressed in real units T_h corresponds to the height of the target profile
Discrete Cosine Transform	20 harmonics	-
Radius, tangent, and curvature	Equal weights for radius, tangent, and curvature	-
Ramer-Douglas-Peucker	20 segments	-

two profiles is expressed as the RMSD between their corresponding points.

2.4. Evaluating fragment retrieval

All four algorithms are *a priori* suitable for rim fragment retrieval, but only the ICP algorithm can be used to identify the best matches for non-rim fragments. The leave-one-out procedure was used to evaluate the algorithms: one vessel was selected from the test corpus of 319 complete vessels. A fragment extracted from this vessel was matched to the remaining 318 complete vessels in order to identify the best matches. This procedure was performed for all 319 vessels. Rim fragment retrieval was evaluated with the strategy used by Lucena et al. (2016) and Martínez-Carrillo et al. (2010), where fragment attribution is considered correct when the original class label is present among the k best matches. The traditional values for classification performance evaluation (with k equal to 1, 3, and 5) were used here for rim fragments. Stricter conditions were applied for non-rim fragments, where fragment attribution was considered correct only if all k best matches corresponded to the original class label (with k values ranging from 1 to 10).

The acceptable ranges for parameter values used to evaluate algorithms are given in Table 2. The optimal values required for very good approximation of the original fragment profile were 20 DCT harmonics and 20 RDP polyline segments (see Supplementary Materials SM2).

The time required for ICP calculation depends on the maximum number of iterations used in the Simulated Annealing optimisation algorithm. Rim fragments were optimised with 1000 iterations. To speed up the calculation time for non-rim fragments, the initial raw position of the source fragments on target complete vessels was estimated with the maximum number of iterations set to 500. The ten best candidates were then selected, and the optimisation procedure was repeated with the maximum number of iterations increased to 5000.

2.5. Programming

The code for archaeological pottery identification was written in R language, version 4.0.3 (R Core Team, 2021), with the aid of libraries 'sp' (Pebesma et al., 2020), 'MASS' (Ripley et al., 2016; Venables and Ripley, 2002), and 'kmlShape' (Genolini and Guichard, 2016). The graphical user interface was programmed with the 'shiny' package (Chang et al., 2019), combined with RStudio (RStudio Team, 2019). All software and packages used here are freely available. The application, with manual and sample data, is provided as Supplementary Materials SM3, and is also accessible via the public Git repository (https://github.com/jwilczek-dotcom/RACORD).

3. Results and discussion

3.1. Rim fragment retrieval

Rim retrieval results can be seen in Table 3 (see also Supplementary materials SM4 for confusion matrices obtained with $k = 1$). The RTC representations, although used with success to create a classification of rim fragments dated to the Iron Age from the region of Levante (Karasik and Smilansky, 2011), did not correctly classify ca. 20% of the test dataset, while the other three methods all produced better results, reaching 95.9% for ICP with $k = 5$. Interestingly, these successful retrieval rates, based only on rim fragments, were very close to those obtained from analyses of whole ceramic profiles (Lucena et al., 2016).

Examples of visual outputs of rim fragment retrieval can be seen in Fig. 7. This figure shows that, when a direct correspondence for the fragment is present in the corpus, the correct morphological class is identified, and the superimposition of the source fragment and target complete vessel is almost perfect (Fig. 7:1–4). Retrieval was not considered successful for rim fragments *(i)* with no direct correspondence in the corpus (Fig. 7:5–6), or *(ii)* which were attributed to a subgroup of the same morphological class (Fig. 7:7–8), or *(iii)* which could

Table 3
Percentage of well-labelled rim fragments, when the correct class is among the k best candidates for the four algorithms tested. The best result for each k is highlighted in bold. See Supplementary materials SM4 for visualisation of confusion matrices obtained with $k = 1$.

	k		
Algorithm	**1**	**3**	**5**
Radius, Tangent, and Curvature	56.4	74.3	81.2
Ramer-Douglas-Peucker	74.0	91.2	94.4
Discrete Cosine Transform	74.3	90.3	95.6
Iterative Closest Point	**81.8**	**93.4**	**95.9**

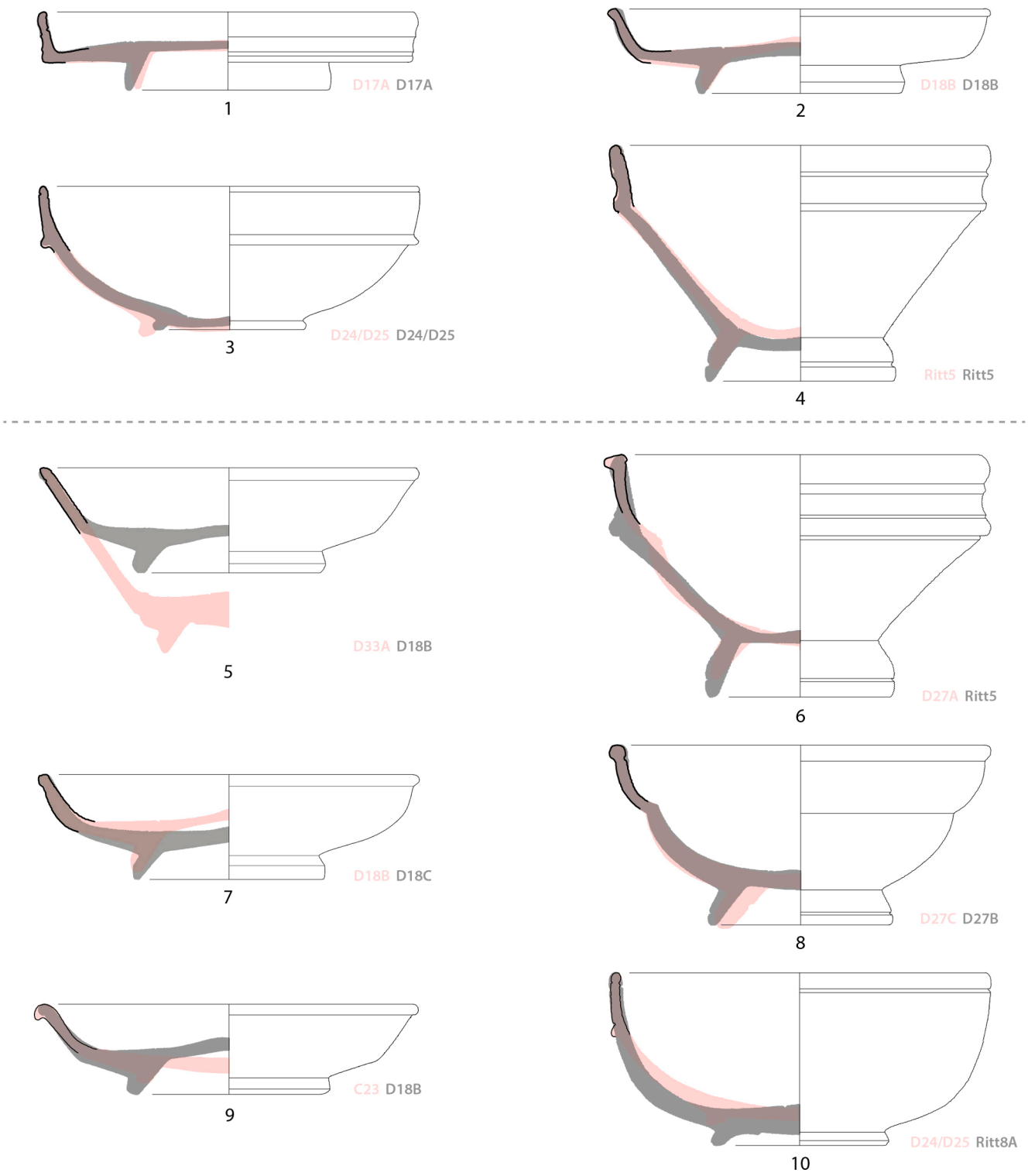

Fig. 7. Evaluation of rim fragment retrieval. 1–4) Example of well-labelled fragments. 5–10) Example of incorrectly labelled fragments. Source fragments correspond to black outlines. Original complete vessels of source fragments are represented by red polygons, with correct sub-groups in red. Complete target vessels are represented by grey polygons, with correct sub-groups in grey. Rims of all fragments and vessels are scaled to the size of the radius.

have been attributed to several different classes (Fig. 7:9–10). However, the first problem would easily be identified by the archaeologist, who could therefore decide not to attribute that fragment to a specific class. The remaining two problems clearly illustrate the underlying limits of morphological attribution in archaeology.

3.2. *Non-rim fragment retrieval*

Non-rim fragments were judged to be attributed correctly only if all k best candidates belonged to the same sub-group. This choice was made to show the practical use of ICP for matching, and to stress that a fair proportion of non-rim fragments, but not all, can be attributed with very high accuracy. Table 4 shows the percentage of classifiable fragments,

Table 4
Attribution for non-rim fragments, based on the *k* best candidates. Fragments were judged classifiable only if all *k* best candidates belonged to the same sub-group. Accuracy indicates the percentage of fragments judged classifiable that were attributed to the correct sub-group.

k	1	2	3	4	5	6	7	8	9	10
Classifiable (%)	100.0	57.1	37.3	28.8	24.8	20.7	16.3	14.7	10.3	8.8
Accuracy (%)	63.3	77.5	83.2	89.1	93.7	93.9	96.2	95.7	97.0	96.4

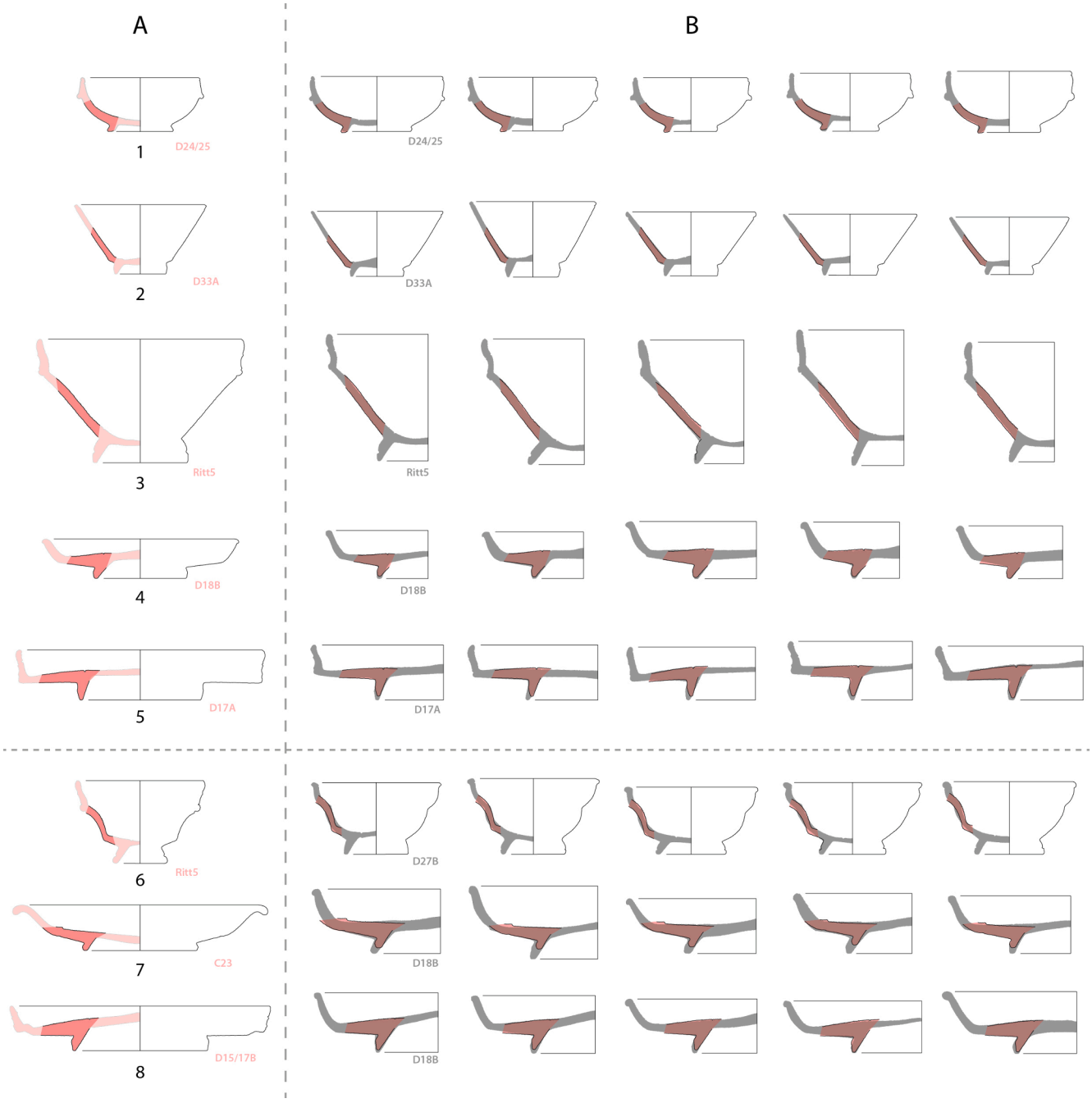

Fig. 8. Evaluation of non-rim fragment retrieval, using ICP, for $k = 5$. 1–5) Example of well-labelled fragments. 6–8) Example of incorrectly labelled fragments. A) Source fragments (darker red polygon bounded by black outlines) and their position on the original complete vessels (lighter red polygons). Source fragment sub-group code is shown in red. B) Source fragments (darker red polygons bounded by black outlines) are superimposed on the five most similar target complete vessels (grey polygons). Sub-group code for the five best complete vessels is shown in grey. Scale 1/3.

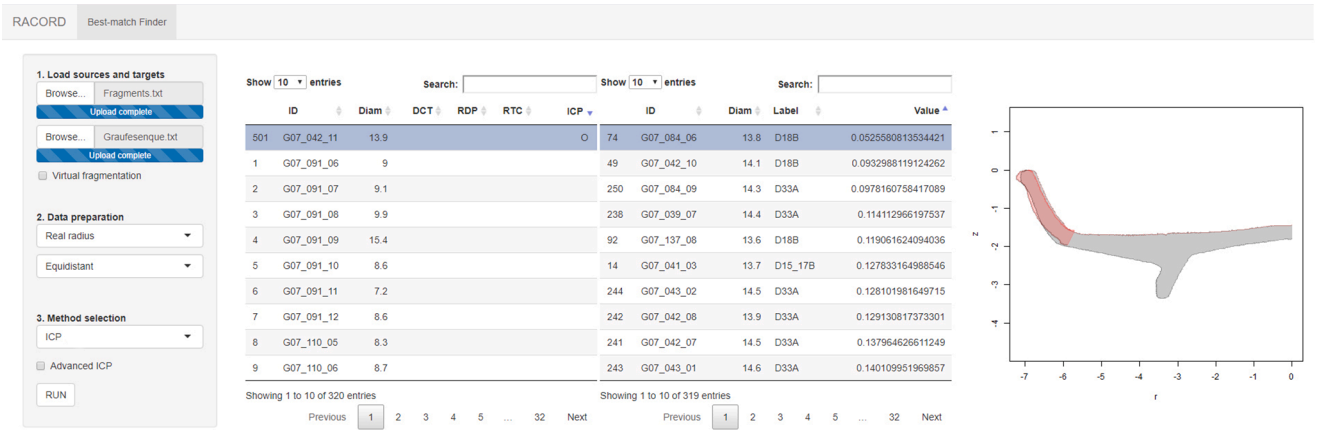

Fig. 9. Screenshot of the application. The source rim fragment to be labelled (G07_042_11; red polygon) is shown superimposed onto the most similar target complete vessel (G07_084_06; grey polygon) found in the test corpus, containing 319 individuals. The target vessel sub-group (here D18B) can be used to label the source fragment.

and the accuracy of attribution, as k increases. With five best candidates ($k = 5$), approximately 25% of fragments can be judged classifiable, with very high accuracy (93.7% here).

Examples of several correctly and incorrectly labelled fragments, for $k = 5$, are shown in Fig. 8. Very good results can be seen in Fig. 8:1–5: the superimposition of source fragments with target complete vessels is almost perfect, and all candidates belong to the same sub-group as the fragment to be labelled. Fragments without a direct correspondence in the dataset, i.e. which are not well aligned on the best candidates, can easily be identified and filtered out by the archaeologist (Fig. 8:6–7), thus minimising errors in attribution. However, in some cases, the source fragment may seem to be well labelled when it is not - it is perfectly aligned with vessels from a different sub-group than the original (Fig. 8:8). The reason is that such fragments do not possess features characteristic of the original class (here the specific shape of the rim) and/or that some parts of vessel profiles are the same for several sub-groups, and so they cannot be discriminated solely on geometric information. As in the case of rim retrieval, such objects show clearly the limits of automatic attribution.

4. Concluding remarks

The four algorithms presented here propose a set of potential best matches for a given fragment, ordered by shape similarity. For three of these algorithms, the results obtained were very good, both qualitatively and quantitatively. The ICP algorithm may be recommended for all types of fragments, because it gives the best labelling outputs, and can easily be applied to all parts of the profile (rims, bodies, and bases), and to both the outer and inner parts of the profile. The basic geometric transformations on which this algorithm is based are easy to grasp, allowing the user to make pertinent choices (e.g. relaxing constraints on the rotational axis). Shape attribution with ICP is rapid, as it takes less than one second to align a given fragment. This algorithm could also be adapted to 3D data to handle fragments that are not perfectly symmetrical around the rotational axis (e.g. deformed fragments, cubic shapes, fragments with plastic features, etc.).

All the algorithms presented here were implemented in a simple computer application, with the open-source R software (Fig. 9). This program provides an extension to (semi-)automatic systems dedicated to pottery fragment orientation and profile extraction, based on 3D models (e.g. Karasik and Smilansky, 2008; Mara and Sablatnig, 2006; Wilczek et al., 2018). The user has full control over the criteria used for best-match retrieval (e.g., the range of searching values in the optimisation step). Criteria should be set in relation to the presupposed quality of the rotational axis position and orientation, the size of the objects stored in the dataset, and the number of points sampled along profile outlines. By arbitrarily fixing several criteria (e.g. the percentage of unique classes within k-best candidates), problematic fragments (e.g. typical of many classes) can be identified, and the entire process can be fully automated. The proposed class labels can then be verified, both visually by inspection of superimposed profiles, and quantitatively from the RMSD values. The quality of the reference database and existing typological schemes will obviously affect the quality of the output, whatever the algorithm used. The archaeologist always retains full control over the final and most important options with regard to fragment attribution (i.e. whether attribution is possible, and if so, into which class).

This freely available tool can be maintained, developed, and tested by the archaeological community. It can be adapted to suit the requirements of various classification strategies. The best-match retrieval procedures can also be integrated into other existing computer-aided systems for the documentation, retrieval, and classification of archaeological pottery fragments (e.g. PIQD (Smith et al., 2014); ArchAIDE (Gualandi et al., 2016); GRAVITATE (Phillips et al., 2016); and more recent works (Di Angelo et al., 2019)). The tool can also calculate the minimum distance (i.e. similarity) between two aligned profiles. Calculated over a set of homologous profiles, these distances can be used for unsupervised classification, or to calculate shape variability, in order to explore artefact variability over space and time.

Acknowledgements

This article was supported by the project “Postdoctoral job positions at UHK” and by the FF UHK Editorial Board.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jasrep.2021.102891.

eferences

Adan-Bayewitz, D., Karasik, A., Smilansky, U., Asaro, F., Giauque, R.D., Lavidor, R., 2009. Differentiation of ceramic chemical element composition and vessel morphology at a pottery production center in Roman Galilee. J. Archaeol. Sci. 36 (11), 2517–2530. https://doi.org/10.1016/j.jas.2009.07.004.

Adler, K., Kampel, M., Kastler, R., Penz, M., Sablatnig, R., Schindler, K., 2002. Computer Aided Classification of Ceramics: Achievements and Problems.

Bahn, P., Renfrew, C. 2015. Archaeology. Theories, Methods and Practice (6th edition all in colour). Thames & Hudson, London.

Bélisle, C.J.P., 1992. convergence theorems for a class of simulated annealing algorithms on Rd. J. Appl. Probab. 29, 885–895.
Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14 (2), 239–256. https://doi.org/10.1109/34.121791.
Bookstein, F.L., 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge University Press, Cambridge.
Buko, A., 1990. Ceramika wczesnopolska. Wprowadzenie do badań. Polska Akademia Nauk, Wrocław – Warszawa – Kraków – Gdańsk – Łódź.
Chang, W., Cheng, J., Allaire, J.J., Yihui, X., McPherson, J., 2019. Shiny: Web Application Framework for R.
Claude, J., 2008. Morphometrics with R. Springer.
Cortez, P., 2014. Modern Optimization with R. Springer International Publishing.
Di Angelo, L., Di Stefano, P., Guardiani, E., Morabito, A.E., 2019. A 3D information framework for automated archaeological pottery archival. 2019 IMEKO TC-4 International Conference on Metrology for Archaeology and Cultural Heritage Florence, Italy, December 4-6, 2019, pp. 178–183.
Dommergues, C.H., Dommergues, J.-L., Verrecchia, E.P., 2007. The discrete cosine transform, a Fourier-related method for morphometric analysis of open contours. Math. Geol. 39, 749–763.
Douglas, D.H., Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographer 10, 112–122.
Fitzgibbon, A., 2003. Robust registration of 2D and 3D point sets. Image Vis. Comput. 21, 1145–1153.
Forel, B., Gabillot, M., Monna, F., Forel, S., Dommergues, C.H., Gerber, S., Petit, C., Mordant, C., Chateau, C., 2009. Morphometry of middle Bronze age palstaves by discrete cosine transform. J. Archaeol. Sci. 36, 721–729.
Genin, M., Dejoie, C., De Parseval, P., Relaix, S., Schaad, D., Schenck-David, J.-L., Sciau, P., 2008. La Graufesenque (Millau, Aveyron) - Volume II - Sigillées lisses et autres productions, Edition de la Fédération Aquitania.
Genolini, C., Guichard, E., 2016. K-Means for Longitudinal Data using Shape-Respecting Distance. https://CRAN.R-project.org/package=kmlShape.
Gilboa, A., Karasik, A., Sharon, I., Smilansky, U., 2004. Towards computerized typology and classification of ceramics. J. Archaeol. Sci. 31, 681–694.
Gualandi, M.L., Scopigno, R., Wolf, L., Richards, J., Buxeda i Garrigos, J., Heinzelmann, M., Hervas, M.A., Vila, L., Zallocco, M., 2016. ArchAIDE Archaeological Automatic Interpretation and Documentation of cEramics, in: Catalano, C.E., De Luca, L. (Eds.), EUROGRAPHICS Workshop on Graphics and Cultural Heritage, 4 p.
Hajek, B., 1988. Cooling schedules for optimal annealing. Math. Oper. Res. 13, 311–329.
Hlaváčková-Schindler, K., Kampel, M., Sablatnig, R., 2001. Fitting of a closed planar curve representing a profile of an archaeological fragment. Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, pp. 263–270.
Hristov, V., Agre, G., 2013. A software system for classification of archaeological artefacts represented by 2D plans. Cybern. Inf. Technol. 13, 82–96. https://doi.org/10.2478/cait-2013-0017.
Hurth, E., Montuire, S., Schmittbuhl, M., Le Minor, J.-M., Schaaf, A., Viriot, L., Chaline, J., 2003. Examination of the tooth morphospace of three Mimomys lineages (Arvicolinae, Rodentia) by elliptical Fourier methods. Coloquios de Paleontología 1, 325–334.
Jičín, R., Vašíček, Z., 1971. K možnostem srovnávání tvarů na základě podobnosti. In: Bouzek, J., Buchvaldek, M. (Eds.), Nové Archeologické Metody. Universita Karlova, Praha, pp. 131–139.
Kampel, M., Sablatnig, R., 2007. Rule based system for archaeological pottery classification. Pattern Recogn. Lett. 28, 740–747. https://doi.org/10.1016/j.patrec.2006.08.011.
Kampel, M., Sablatnig, R., 2003. Automated Archivation System for Pottery.
Kampel, M., Sablatnig, R., 2002. Automated segmentation of archaeological profiles for classification. Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 1.
Kampel, M., Sablatnig, R., 1999. On Estimating the Position of Fragments on Rotational Symmetric Pottery. 2nd International Conference on 3D Digital Imaging and Modeling (3DIM '99), 4-8 October 1999, Ottawa, Canada, 01/1999.
Karasik, A., Smilansky, U., 2011. Computerized morphological classification of ceramics. J. Archaeol. Sci. 38, 2644–2657.
Karasik, A., Smilansky, U., 2008. 3D scanning technology as a standard archaeological tool for pottery analysis: practice and theory. J. Archaeol. Sci. 35, 1148–1168. https://doi.org/10.1016/j.jas.2007.08.008.
Karasik, A., Smilansky, U., Beit-Arieh, I., 2005. New typological analyses of early Bronze Age Holemouth Jars from Tel Arad and Southern Sinai. J. Inst. Archaeol. Tel Aviv Univ. 32, 20–31.
Laflin, S., 1986. Use of a Sinclair Spectrum For Shape Analysis. In: Laflin, S. (Ed.), Conference Proceedings Presented at the Computer Applications in Archaeology 1986. University of Birmingham, Centre for Computing and Computer Science, pp. 83–90.
Leese, M.N., Main, P.L., 1983. An approach to the assessment of artifact dimension as descriptors of shape. In: Haigh, J.G.B. (Ed.), Computer Applications in Archaeology. University of Bradford, Bradford, pp. 171–180.
Liu, D., Razdan, A., Simon, A., Bae, M., 2005. An XML-based information model for archaeological pottery. J. Zhejiang Univ. Sci. A Appl. Phys. Eng. 6A, 447–453.
Lucena, M., Fuertes, J.M., Martinez-Carrillo, A.L., Ruiz, A., Carrascosa, F., 2016. Efficient classification of Iberian ceramics using simplified curves. J. Cult. Heritage 19, 538–543. https://doi.org/10.1016/j.culher.2015.10.007.
Macháček, J., 2001. Studie k velkomoravské keramice. Metody, analýzy a syntézy, modely. Ústav archeologie a muzeologie. Filozofická fakulta Masarykovy univerzity v Brně, Brno, 296 p.
Main, P.L., 1987. Accessing outline shape information efficiently within a large database II: database compaction techniques. In: Ruggles, C.L.N., Rahtz, S.P.Q. (Eds.), CAA87. Computer and Quantitative Methods in Archaeology 1987 (BAR International Series 393). Oxford, pp. 242–251.
Main, P.L., 1986. Accessing Outline Shape Information Efficiently within a Large Database, in: Lafin, S. (Ed.), Conference Proceedings Presented at the Computer Applications in Archaeology 1986, pp. 73–82.
Maiza, Ch., Gaildrat, V., 2005. Automatic Classification of Archaeological Potsherds, in: Dimitri, P. (Ed.), The 8th International Conference on Computer Graphics and Artificial Intelligence, 3IA'2005, Limoges, France, 11/05/2005-12/05/2005. pp. 135–147.
Mara, H., Sablatnig, R., 2006. Orientation of fragments of rotationally symmetrical 3d-shapes for archaeological documentation. Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 1064–1071. https://doi.org/10.1109/3DPVT.2006.105.
Martínez-Carrillo, A.L., Lucena, M., Fuertes, J.M., Ruiz, A., 2010. Morphometric analysis applied to the archaeological pottery of the valley of Guadalquivir. In: Elwa, A.M.T. (Ed.), Morphometrics for Nonmorphometricians. Springer, pp. 307–323.
Mom, V., 2005. SECANTO–the section analysis tool, in: Figueiredo, A., Leite Velho, G. (Eds.), The World Is in Your Eyes. Proceedings of the 33rd Computer Applications and Quantitative Methods in Archaeology Conference. Tomar, March 2005.
Mom, V., Paijmans, J.J., 2007. SECANTO: a retrieval system and classification tool for simple artifacts, in: Layers of Perception: Proceedings of the 35th Computer Applications and Quantitative Methods in Archaeology Conference. Berlin, Germany, April 2–6, 2007. pp. 1–5.
Ness, K.L., 2015. Classification systems with a plot: vessel forms and ceramic typologies in the Spanish Atlantic. Int. J. Historical Archaeol. 19 (2), 309–333. https://doi.org/10.1007/s10761-015-0290-9.
Orton, C., 1980. Mathematics in Archaeology. Cambridge University Press, Cambridge.
Orton, C., Tyers, P., Vince, A., 1993. Pottery in Archaeology. Cambridge University Press, Cambridge.
Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J., O'Brien, J., O'Rourke, J., 2020. sp: Classes and Methods for Spatial Data. https://cran.r-project.org/web/packages/sp/index.html.
Phillips, S., Walland, P., Modafferi, S., Spagnuolo, M., Catalano, C.E., Oldman, D., Tal, A., Shimshoni, I., Hermon, S., 2016. GRAVITATE: Geometric and semantic matching for cultural heritage artefacts, in: Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage. pp. 199–202.
Piccolli, Ch., Aparajeya, P., Papadopulos, G.Th., Bintliff, J., Leymarie, F.F., Bes, Ph., van der Enden, M., Poblome, J., Daras, P., 2015. Towards the automatic classification of pottery sherds: two complementary approaches. CAA 2013. Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology. Amsterdam University Press, pp. 463–474.
R Core Team, 2021. R: A Language and Environment for Statistical Computing.
Rice, P.M., 1987. Pottery Analysis. A Sourcebook. The University of Chicago Press, Chicago, London.
Ripley, A., Bevables, B., Bates, D.M., Hornik, K., Gebhardt, A., Fith, D., 2016. Package "MASS". https://cran.r-project.org/web/packages/MASS/index.html.
RStudio Team, 2019. RStudio: Integrated Development Environment for R.
Saragusti, I., Karasik, A., Sharon, I., Smilansky, U., 2005. Quantitative analysis of shape attributes based on contours and section profiles in artifact analysis. J. Archaeol. Sci. 32, 841–853.
Schmittbuhl, M., Allenbach, B., Le Minor, J.-M., Schaaf, A., 2003. Elliptical descriptors: some simplified morphometric parameters for the quantification of complex outlines. Math. Geol. 35, 853–871.
Schurmans, U., Razdan, A., Simon, A., Mccartney, P., Marzke, M., Van Alfen, D., Jones, J., Rowe, J., Farin, G., Collins, D., Zhu, M., Liu, D., Bae, M., 2001. Advances in geometric modeling and feature extraction on pots, rocks and bones for representation and query via the internet, in: Burenhult, G., Arvidsson, J. (Eds.), Archaeological Informatics: Pushing the Envelope. CAA2001. Computer Applications and Quantitative Methods in Archaeology. Archaeopress, Oxford, pp. 191–204.
Smith, N.G., Karasik, A., Narayanan, T., Olson, E.S., Smilansky, U., Levy, T.E., 2014. the pottery informatics query database: a new method for mathematic and quantitative analyses of large regional ceramic datasets. J. Archaeol. Method Theory 21.
Turk, G., Levoy, M., 1994. Zipped Polygon Meshes from Range Images, in: Proceedings of SIGGRAPH'94, pp. 311–318.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer, New York.
Vaginay, M., Guichard, V., 1988. L'habitat gaulois de Feurs (Loire). Fouilles récentes (1978-1981). Documents d'archéologie française, 14. Ed. de la Maison des Sciences de l'Homme, Paris.
Venclová, N., 1998. Mšecké Žehrovice in Bohemia. Archaeological Background to a Celtic Hero (3rd–2nd cent. B.C.). Kronos, Sceux.
Venclová, N., 2001. Výroba a sídla v době laténské. Projekt Loděnice. Archeologický ústav AV ČR, Praha.
Wilczek, J., Monna, F., Jébrane, A., Labruère-Chazal, C., Navarro, N., Couette, S., Chateau Smith, C., 2018. Computer-assisted orientation and drawing of archaeological pottery. J. Comput. Cult. Heritage December 2018.