



HAL
open science

A computer tool to identify best matches for pottery fragments

Josef Wilczek, Fabrice Monna, Nicolas Navarro, Carmela Chateau-Smith

► **To cite this version:**

Josef Wilczek, Fabrice Monna, Nicolas Navarro, Carmela Chateau-Smith. A computer tool to identify best matches for pottery fragments. *Journal of Archaeological Science: Reports*, 2021, 37, pp.102891. 10.1016/j.jasrep.2021.102891 . hal-03182145

HAL Id: hal-03182145

<https://hal.science/hal-03182145v1>

Submitted on 12 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 A computer tool to identify best matches for pottery
2 fragments

3

4 Josef Wilczek^{1,2*} ; Fabrice Monna² ; Nicolas Navarro^{3,4} ; Carmela Chateau-Smith⁵

5

6 1 : Department of Archaeology, University of Hradec Králové, 50003 Hradec Králové, Czech Republic

7 2 : ARTEHIS, UMR CNRS 6298, Université Bourgogne Franche-Comté, 21000 Dijon, France

8 3 : Biogéosciences, UMR CNRS EPHE 6282, Université Bourgogne Franche-Comté, 21000 Dijon, France

9 4 : EPHE, PSL University, 75014 Paris, France

10 5 : CPTC, Université Bourgogne Franche-Comté, 21000 Dijon, France

11

12

13

14

15

16

17

18

19 * : Corresponding author, josef.wilczek@uhk.cz, tel: +420 606 880 806

20 **Abstract:**

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

32

33

34 **Keywords:** archaeology, best match, automation, pottery, ICP (iterative closest point), DCT
35 (discrete cosine transform), RDP (Ramer-Douglas-Peucker), RTC (radius, tangent, and
36 curvature)

37 1. Introduction

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

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

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

57 Almost all archaeological pottery vessels can be considered to be rotationally symmetrical.
58 These 3D objects are thus usually represented by a 2D profile, corresponding to any cross-
59 section between the vessel and the plane passing through its rotational axis. Quantitative
60 methods are therefore often based on the calculation of similarities between 2D profiles,
61 usually represented by polar or cartesian coordinates (Liu et al., 2005; Maiza and Gaildrat,
62 2005), or expressed as a function of radius (e.g. Jičín and Vašíček, 1971; Karasik and
63 Smilansky, 2011; Mom, 2005; Smith et al., 2014), or its derivatives, such as tangent (e.g.
64 Gilboa et al., 2004; Karasik and Smilansky, 2011; Leese and Main, 1983; Main, 1987, 1986;
65 Saragusti et al., 2005; Smith et al., 2014), or curvature (Gilboa et al., 2004; Hristov and Agre,
66 2013; Karasik et al., 2005; Saragusti et al., 2005). Profiles may sometimes be represented by
67 b-spline coefficients associated with segments of the profile curve (e.g. Adler et al., 2002;
68 Hlaváčková-Schindler et al., 2001; Kappel and Sablatnig, 2007, 2003, 2002, 1999; Laflin,
69 1986; Schurmans et al., 2001), by polylines (Lucena et al., 2016), or by dominant feature
70 points that are extracted using the medialness measurement (Piccolli et al., 2015).
71 Similarities between two pottery fragments can be calculated from Euclidean distances
72 between their profiles (Maiza and Gaildrat, 2005), or from the similarity of the coefficients
73 expressing their shape (Gilboa et al., 2004). The best match for a given fragment (and hence
74 its morphological class) is obtained by minimising differences with potential candidates in a
75 referential database.

76 The above-mentioned approaches have already been tested on real-world artefacts, with a
77 high success rate for classification (Karasik and Smilansky, 2011), and for best-match
78 retrieval (Lucena et al., 2016). It should, however, be noted that several major issues have
79 not yet been fully addressed. For example, if the position of the outline in relation to the
80 rotational axis is not appropriately constrained, the shape of the pottery fragment or vessel
81 may be drastically deformed, leading to misidentification. Methods requiring successive

82 derivatives generally suffer from numerical instability (Karasik and Smilansky, 2008).
83 Although the idea underlying morphological correspondence between profiles appears
84 simple, achieving perfect correspondence between individual points on profiles is
85 challenging, even more so when working on fragments. Many approaches have been
86 evaluated solely on the outer surface of the profile, or only on rim fragments. Although the
87 mathematical basis of all these approaches has already been explored (e.g. Hristov and Agre,
88 2013; Liu et al., 2005; Mom, 2005; Mom and Paijmans, 2007; Smith et al., 2014), no practical
89 solution is currently available to the broader archaeological community for routine use. It is
90 important for such proposals to be tested on a wider variety of artefacts, as no method is
91 likely to be universally applicable.

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

103

104 **2. Material and methods**

105 **2.1. Corpus**

106 The collection of ceramic vessels found at Graufesenque (southern France), dating from the
107 Roman period (first to mid-second century AD), is published as an illustrated paper catalogue
108 (Genin et al., 2008). The fact that vessels in this catalogue are already labelled by a widely
109 adopted morphological classification scheme (e.g. Brulet et al., 2012; Passelac and Vernhet,
110 1993; Py et al., 1993; Schucany et al., 1999; Tyers, 1996), makes it suitable for method
111 comparison. The test corpus identified within this vast collection contains complete vessels
112 only, and includes all morphological classes (and sub-groups) represented by ten or more
113 vessels in the catalogue. The test corpus is thus composed of 319 vessels, including plates,
114 bowls, cups, and goblets (Fig. 1; Supplementary Materials SM1), already divided into 14 sub-
115 groups, belonging to 10 morphological classes (Table 1).

116

117 **2.2. Profile acquisition and data preparation**

118 The drawings of the 319 vessels thus selected were scanned at 600 dpi from the paper
119 catalogue. Profile outlines were extracted using a modified set of functions described in
120 Claude (2008), and positioned according to their rotational axis. The outlines were
121 automatically divided into outer and inner segments, using the rim as a reference point (Fig.
122 2:A).

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

127 For rim fragments, all outlines were size-normalised using baseline registration (Bookstein,
 128 1991), sending the rim point and the point of intersection between rim plane and the
 129 rotational axis to the (-1,0) and (0,0) coordinates (Fig. 2:Bi). Rim fragments were then
 130 obtained by virtually cutting the outer and inner profile parts at a distance of 50 percent of
 131 the rim radius from the rim point, expressed by 100 equally spaced points (50 for the inner
 132 and 50 for the outer segment; Fig. 2:Bii). Virtual rim fragments therefore measured 11-35%
 133 of the profile length of the original complete vessel.

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

137

138 **2.3. Best match searching algorithms**

139 *2.3.1. Iterative Closest Point (ICP)*

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

$$148 \min_{a,b,\varphi,S} \left(\sqrt{\frac{1}{M} \sum_{j=1}^M d(P_j, T_i)^2} + \sqrt{\frac{1}{N} \sum_{k=1}^M d(P_k, T_o)^2} \right)$$

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

155 Note that scaling (S), and translation along the z axis (b) are the only rigid transformations
 156 that do not alter the original shape of the vessel, unlike a and φ , which are strongly
 157 constrained by the position of the rotational axis. Ancient pottery vessels are often
 158 considerably fragmented, and were not always perfectly regular, thus making it difficult to
 159 determine the precise rotational axis to be used for profile extraction. In real-life situations,
 160 the researcher may therefore decide to relax the constraints of the rotational axis position a

161 little, by allowing a and φ to vary to some extent, although some shape modification will
162 occur.

163 2.3.2. Discrete Cosine Transform (DCT)

164 Outlines can also be treated by Discrete Cosine Transform (DCT), used in signal treatment for
165 compression and de-noising. This Fourier-based method transforms open outline
166 coordinates (Fig. 4:A) into a set of harmonics (Fig. 4:B), each represented by a pair of
167 coefficients that may be used as shape variables (e.g. Dommergues et al., 2007; Forel et al.,
168 2009). The higher the number of coefficients, the more precise the reconstruction of the
169 outline (Fig. 4:B). Shape information can generally be preserved with a low number of
170 harmonics (Hurth et al., 2003; Schmittbuhl et al., 2003). The similarity between two profiles
171 is then expressed as the RMSD between their corresponding coefficients.

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

176 2.3.3. Radius, tangent, and curvature (RTC)

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

186 2.3.4. Ramer-Douglas-Peucker (RDP) polyline

187 Pottery profile coordinates can also be approximated by a polyline with a fixed number of
188 segments, obtained with the Ramer-Douglas-Peucker (RDP) algorithm (Douglas and Peucker,
189 1973; Lucena et al., 2016). Given a profile curve with endpoints a and b , this algorithm seeks
190 the most distant point on the curve from the segment ab . Once this point, c , is located, the
191 segment ab is then replaced by two new segments, ac and cb (Fig. 6:A). The entire
192 procedure is then repeated until the desired number of segments (L) is obtained on the
193 profile outline (Fig. 6:B-D). The points generated in this way then serve as shape variables.
194 Similarity between two profiles is expressed as the RMSD between their corresponding
195 points.

196

197 2.4. Evaluating fragment retrieval

198 All four algorithms are *a priori* suitable for rim fragment retrieval, but only the ICP algorithm
199 can be used to identify the best matches for non-rim fragments. The leave-one-out
200 procedure was used to evaluate the algorithms: one vessel was selected from the test
201 corpus of 319 complete vessels. A fragment extracted from this vessel was matched to the
202 remaining 318 complete vessels in order to identify the best matches. This procedure was

203 performed for all 319 vessels. Rim fragment retrieval was evaluated with the strategy used
204 by Lucena et al. (2016) and Martínez-Carrillo et al. (2010), where fragment attribution is
205 considered correct when the original class label is present among the k best matches. The
206 traditional values for classification performance evaluation (with k equal to 1, 3, and 5) were
207 used here for rim fragments. Stricter conditions were applied for non-rim fragments, where
208 fragment attribution was considered correct only if all k best matches corresponded to the
209 original class label (with k values ranging from 1 to 10).

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

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

219

220 **2.5. Programming**

221 The code for archaeological pottery identification was written in R language, version 4.0.3 (R
222 Core Team, 2021), with the aid of libraries 'sp' (Pebesma et al., 2020), 'MASS' (Ripley et al.,
223 2016; Venables and Ripley, 2002), and 'kmlShape' (Genolini and Guichard, 2016). The
224 graphical user interface was programmed with the 'shiny' package (Chang et al., 2019),
225 combined with RStudio (RStudio Team, 2019). All software and packages used here are
226 freely available. The application, with manual and sample data, is provided as
227 Supplementary Materials SM3, and is also accessible via the public Git repository
228 (<https://github.com/jwilczek-dotcom/RACORD>).

229

230 **3. Results and discussion**

231 **3.1. Rim fragment retrieval**

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

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

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

249

250 **3.2. Non-rim fragment retrieval**

251 Non-rim fragments were judged to be attributed correctly only if all k best candidates
252 belonged to the same sub-group. This choice was made to show the practical use of ICP for
253 matching, and to stress that a fair proportion of non-rim fragments, but not all, can be
254 attributed with very high accuracy. Table 4 shows the percentage of classifiable fragments,
255 and the accuracy of attribution, as k increases. With five best candidates ($k = 5$),
256 approximately 25% of fragments can be judged classifiable, with very high accuracy (93.7%
257 here).

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

271

272 **4. Concluding remarks**

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

284 All the algorithms presented here were implemented in a simple computer application, with
285 the open-source R software (Fig. 9). This program provides an extension to (semi-)automatic
286 systems dedicated to pottery fragment orientation and profile extraction, based on 3D
287 models (e.g. Karasik and Smilansky, 2008; Mara and Sablatnig, 2006; Wilczek et al., 2018).

288 The user has full control over the criteria used for best-match retrieval (e.g., the range of
289 searching values in the optimisation step). Criteria should be set in relation to the
290 presupposed quality of the rotational axis position and orientation, the size of the objects
291 stored in the dataset, and the number of points sampled along profile outlines. By arbitrarily
292 fixing several criteria (e.g. the percentage of unique classes within k -best candidates),
293 problematic fragments (e.g. typical of many classes) can be identified, and the entire process
294 can be fully automated. The proposed class labels can then be verified, both visually by
295 inspection of superimposed profiles, and quantitatively from the RMSD values. The quality of
296 the reference database and existing typological schemes will obviously affect the quality of
297 the output, whatever the algorithm used. The archaeologist always retains full control over
298 the final and most important options with regard to fragment attribution (i.e. whether
299 attribution is possible, and if so, into which class).

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

310 **TABLES**

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

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

314

Algorithm	Rim fragments	Non-rim fragments
Iterative Closest Point	$a \in (-0.1, +0.1)$ $\varphi \in (-3^\circ, +3^\circ)$ $S \in (-2\%, +2\%)$ a values are expressed in rim radius units	$a \in (-1\text{ cm}, +1\text{ cm})$ $\varphi \in (-3^\circ, +3^\circ)$ $S \in (-2\%, +2\%)$ $b \in (0\text{ cm}, T_h\text{ cm})$ 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	-

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

318

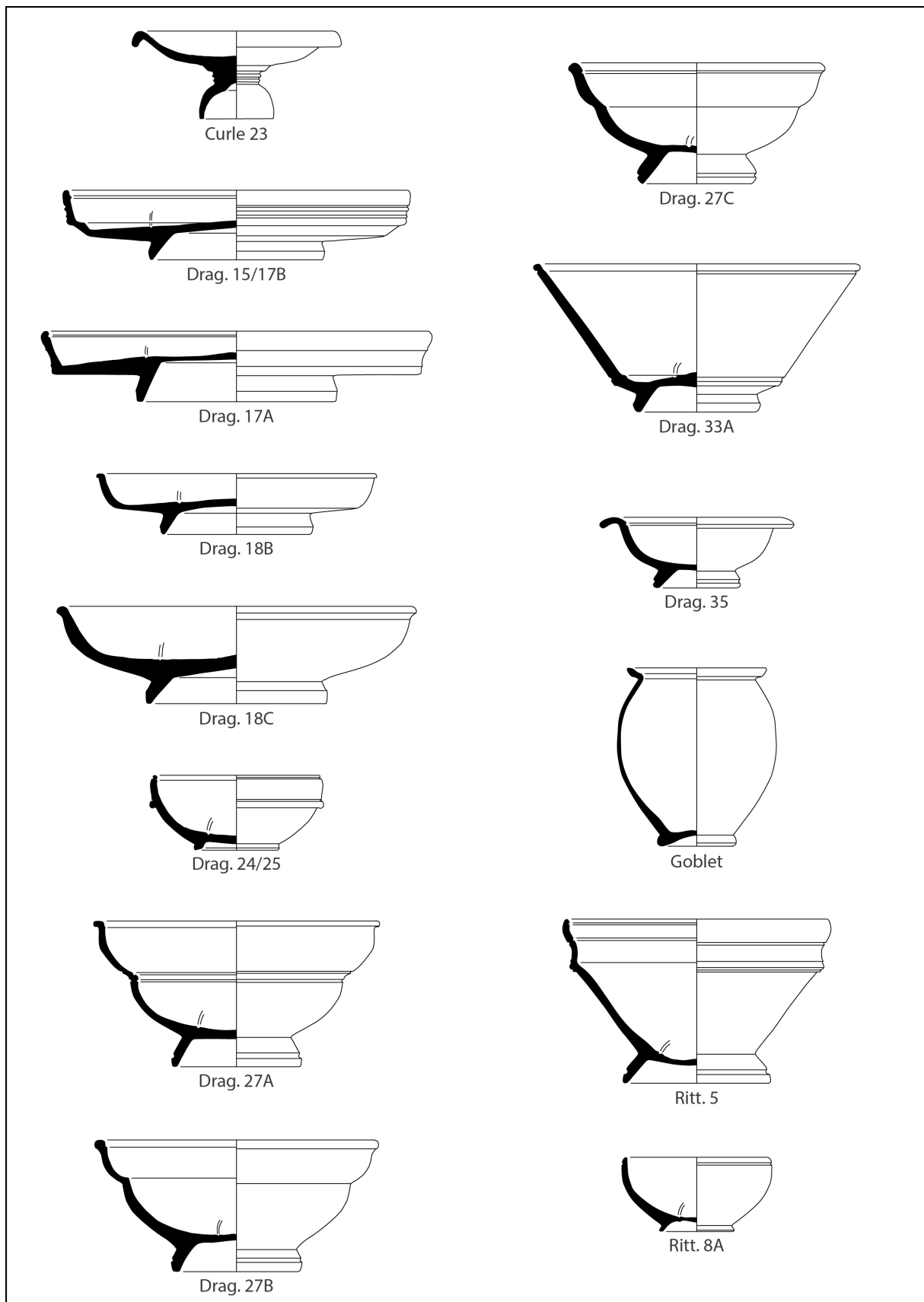
Algorithm	k		
	1	3	5
Radius, tangent, and curvature	56.4	74.3	81.2
Ramer-Douglas-Peucker	74.0	91.2	94.4
Discrete Cosine Transform	74.3	90.3	95.6
Iterative Closest Point	81.8	93.4	95.9

319 **Table 3.** Percentage of well-labelled rim fragments, when the correct class is among the k
 320 best candidates for the four algorithms tested. The best result for each k is highlighted in
 321 bold. See Supplementary materials SM4 for visualisation of confusion matrices obtained with
 322 $k = 1$.

<i>k</i>	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

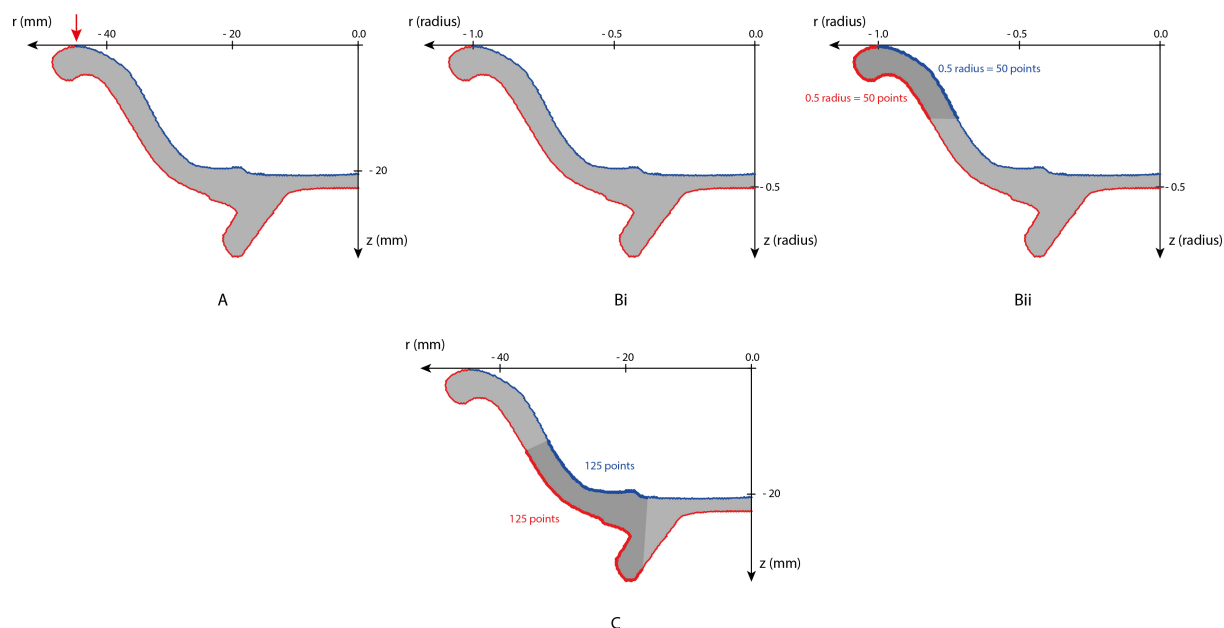
323 **Table 4.** Attribution for non-rim fragments, based on the *k* best candidates. Fragments were
324 judged classifiable only if all *k* best candidates belonged to the same sub-group. Accuracy
325 indicates the percentage of fragments judged classifiable that were attributed to the correct
326 sub-group.

327

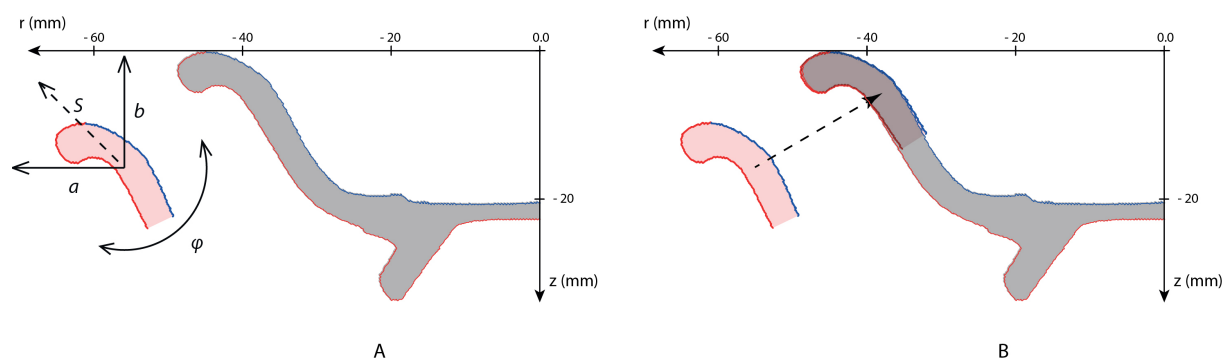


329
330
331

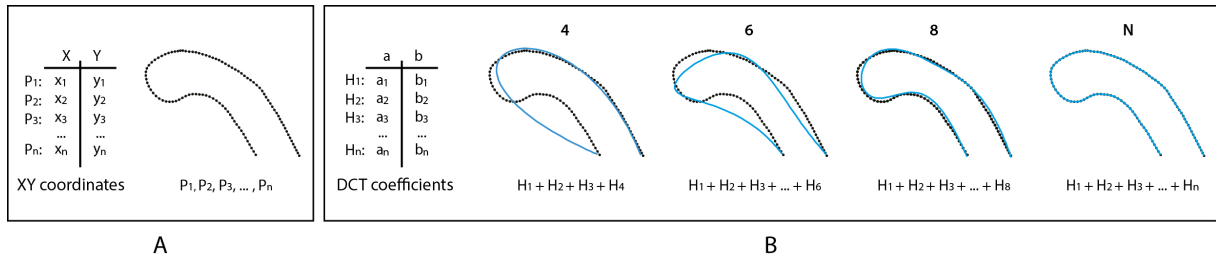
Figure 1. Morphological sub-groups used in this study. Abbreviations: Drag. – Dragendorff, Ritt. – Ritterling. Scale 1/2.



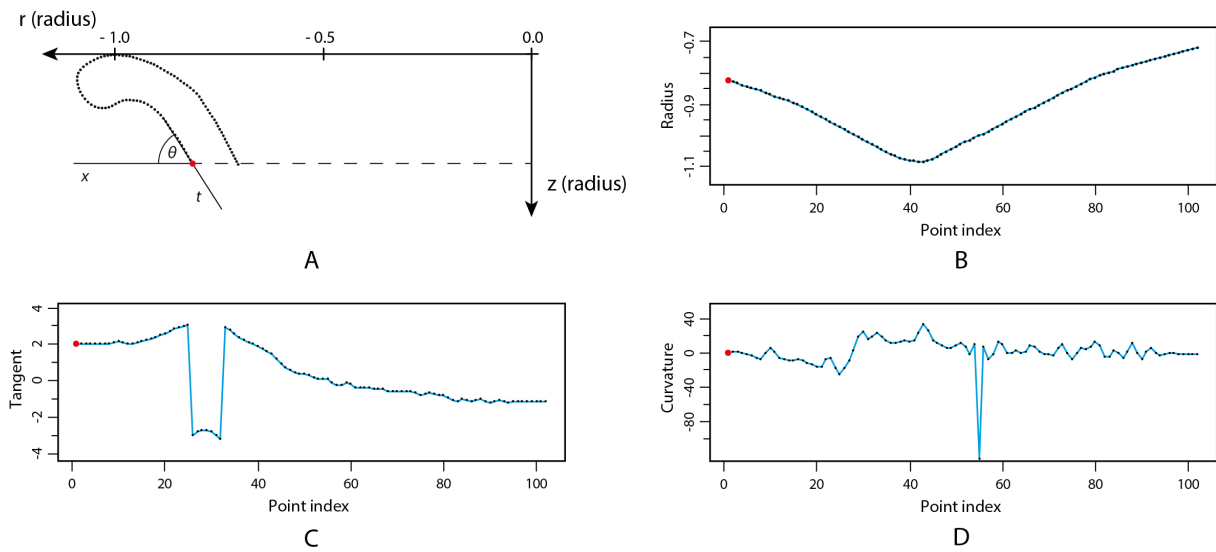
332
 333 **Figure 2.** Profile acquisition and data preparation. A) Scaling the outline to the original size of
 334 the vessel, positioning it according to the axis of symmetry (i.e. z axis), and r axis, and
 335 segmenting it to the outer (red) and inner (blue) segments using the rim (red arrow) as a
 336 reference point. Bi) Size normalisation of the profile by baseline registration (here the rim
 337 point and the point of intersection between rim plane and the axis of the symmetry are set
 338 to (-1,0) and (0,0) respectively; note that the coordinate system is now expressed in rim
 339 radius units). Bii) Virtual extraction of the rim fragment: the outer and inner part of the
 340 profile is virtually broken at the distance of 0.5 rim radius along the curvilinear abscissa from
 341 the rim point, and resampled by 100 equally spaced points, 50 for the outer and 50 for the
 342 inner part. C) Virtual extraction of the non-rim fragment: the outer and inner part of the
 343 profile are extracted approximately in the middle of the vessel, and resampled by 250
 344 points, with 125 points equally spaced along the outer and inner parts.



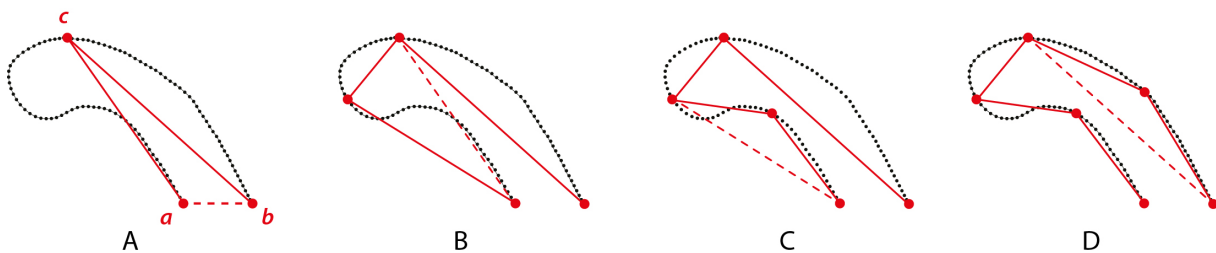
345
 346 **Figure 3.** Alignment process between source fragment (red polygon) and complete target
 347 vessel (grey polygon). A) The fragment is translated along the r and z axes (a , b), rotated (φ),
 348 and scaled (S) to minimise the distance with the complete vessel. B) Final alignment of the
 349 source on the target.



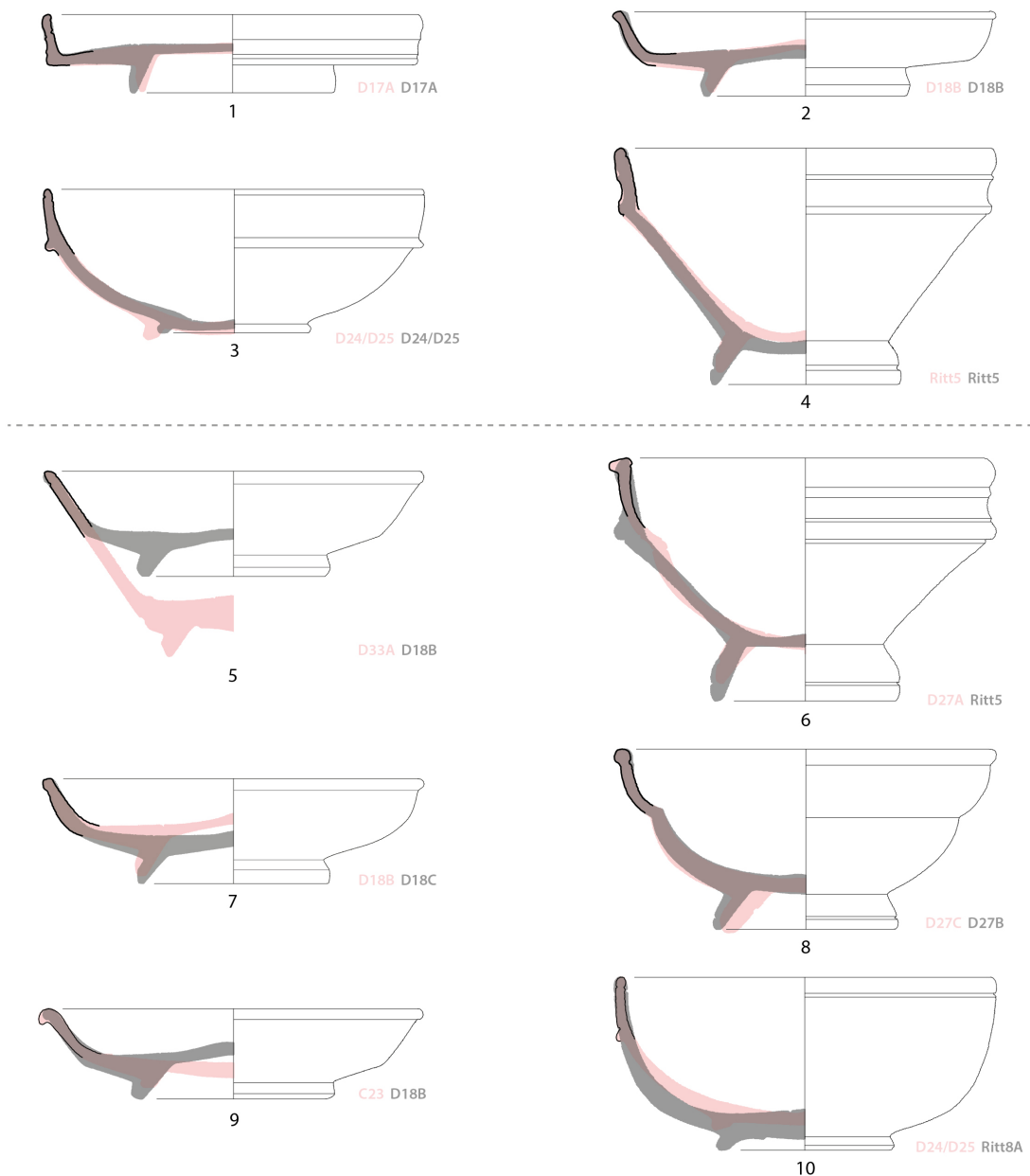
350
 351 **Figure 4.** Discrete cosine transform. A) The original cartesian coordinates of the outline. B)
 352 The original cartesian coordinates of the outline are decomposed into a set of harmonics. A
 353 given number of harmonics can be used to reconstruct the approximation of the original
 354 contour (here 4, 6, 8, and a full set of harmonics were used for reconstruction).



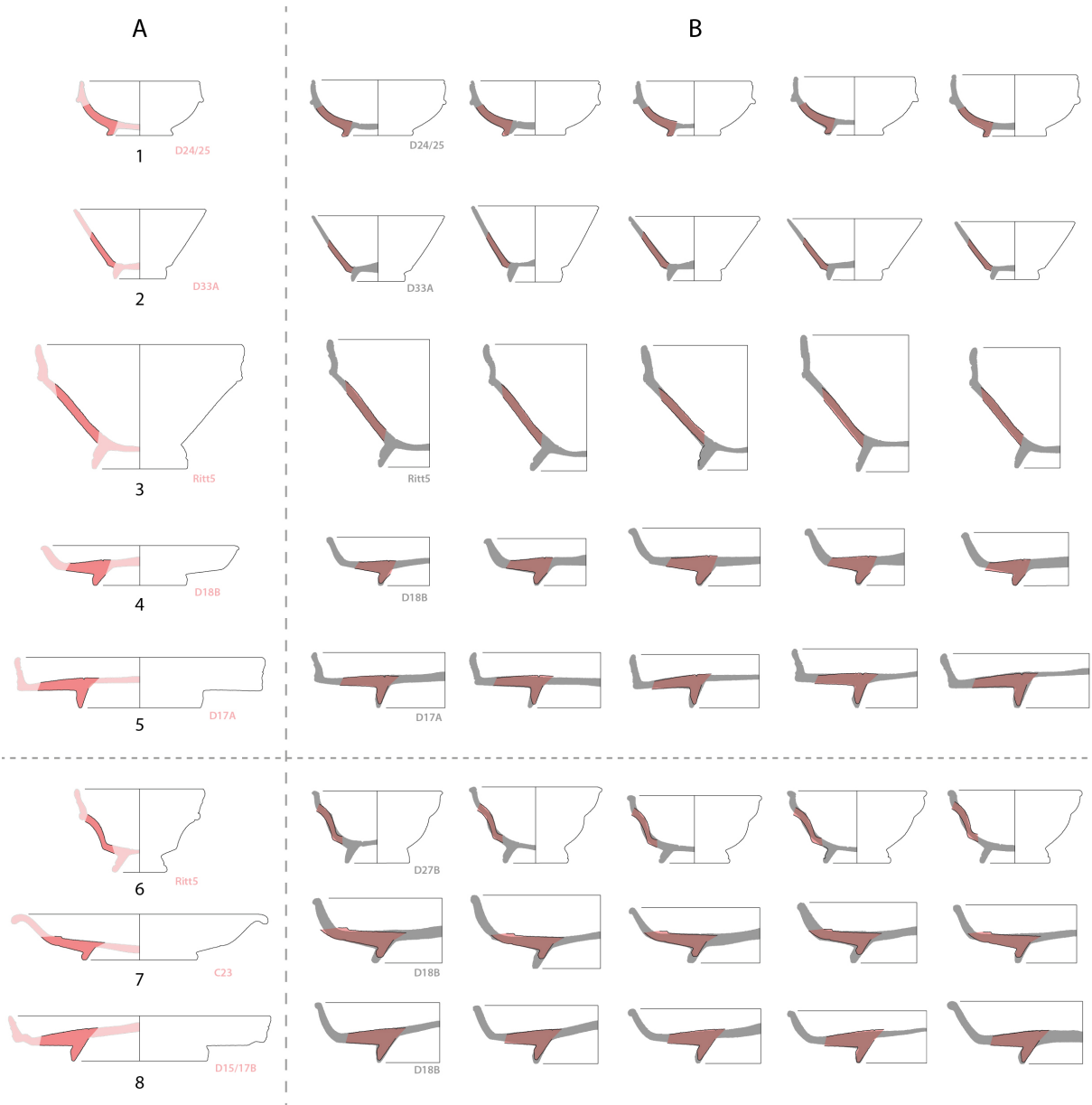
355
 356 **Figure 5.** Radius, tangent, and curvature. The first point on the outline, p , and its
 357 corresponding representations are highlighted in red. A) Position of the fragment according
 358 to rotational axis and tangent angle calculation. The tangent angle θ of the point (here p) is
 359 calculated as an angle between the tangent line passing through the point, t , and the line
 360 parallel to the rim of vessel, x . B) The radius function represents the distance between each
 361 point on the profile and the rotational axis. B) Tangent function represents the angle
 362 between the tangent of each point on the profile and the line parallel to the rim (see A for
 363 details). C) Curvature function represents the rate of change of the tangent function.



364
 365 **Figure 6.** Ramer-Douglas-Peucker polyline algorithm. A) Given a profile curve (black dots)
 366 with endpoints a and b (red dots), this algorithm seeks the most distant point on the curve,
 367 from the segment ab (dashed red line). Once this point, c , is located (red point), the segment
 368 ab is replaced by two new segments, ac and bc (red lines). 2-4) The procedure is repeated
 369 until the desired number of segments is obtained (here 5 segments).

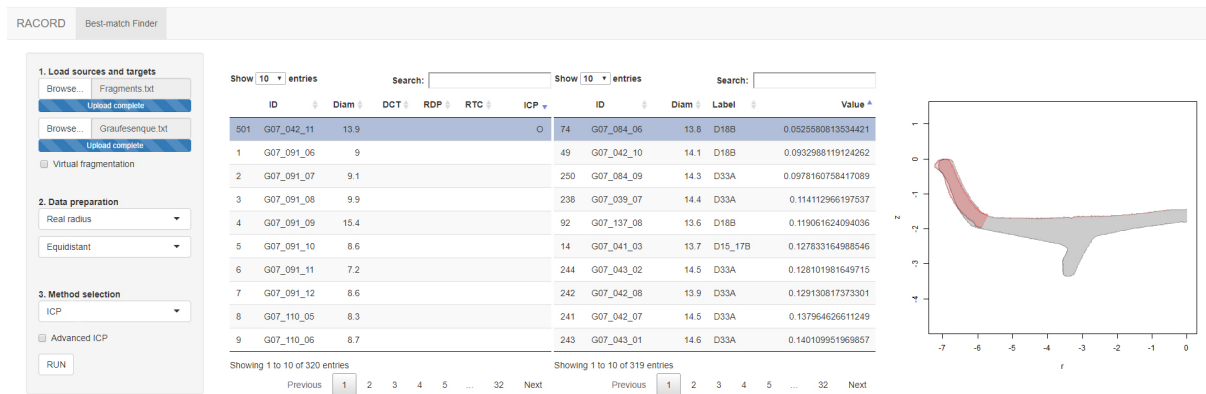


370
 371 **Figure 7.** Evaluation of rim fragment retrieval. 1-4) Example of well-labelled fragments. 5-10)
 372 Example of incorrectly labelled fragments. Source fragments correspond to black outlines.
 373 Original complete vessels of source fragments are represented by red polygons, with correct
 374 sub-groups in red. Complete target vessels are represented by grey polygons, with correct
 375 sub-groups in grey. Rims of all fragments and vessels are scaled to the size of the radius.



376
 377
 378
 379
 380
 381
 382
 383

Figure 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.



384
 385 **Figure 9.** Screenshot of the application. The source rim fragment to be labelled
 386 (G07_042_11; red polygon) is shown superimposed onto the most similar target complete
 387 vessel (G07_084_06; grey polygon) found in the test corpus, containing 319 individuals. The
 388 target vessel sub-group (here D18B) can be used to label the source fragment.

389 **SUPPLEMENTARY MATERIALS**

390 **SM1.** Information about the test corpus. A description of rim diameters, sub-groups, groups,
391 and bibliographic references for a set of 319 vessels from the site of Graufesenque (southern
392 France) dated to the Roman Period.

393 **SM2.** The reconstruction quality of pottery fragment outlines, based on an increasing
394 number of DCT harmonics (A), and RDP segments (B). The original outline corresponds to the
395 black points; the shapes reconstructed by a given number of harmonics and segments are
396 expressed as a blue curve and a red curve. Evaluations were performed using 20 harmonics
397 and 20 segments (see black box).

398
399 **SM3.** The suite of functions for best-match identification implemented in the RACORD
400 application, with instruction manual, and examples.

401 **SM4.** Four confusion matrices of rim fragment labelling when the 1st best candidate is the
402 correct sub-group (for $k = 1$). Each table represents the confusion matrix between the
403 original sub-groups and sub-groups obtained with corresponding method. Correct
404 correspondences are highlighted in bold. Dashed rectangles delimit sub-groups of vessels
405 belonging to the same group. See also Table 3.

406

407 **References**

- 408 Adan-Bayewitz, D., Karasik, A., Smilansky, U., Asaro, F., Giauque, R.D., Lavidor, R., 2009.
409 Differentiation of ceramic chemical element composition and vessel morphology at a
410 pottery production center in Roman Galilee. *Journal of Archaeological Science* 36,
411 2517–2530. <https://doi.org/10.1016/j.jas.2009.07.004>
- 412 Adler, K., Kampel, M., Kastler, R., Penz, M., Sablatnig, R., Schindler, K., 2002. Computer Aided
413 Classification of Ceramics: Achievements and Problems.
- 414 Bahn, P., Renfrew, C. 2015. *Archaeology. Theories, Methods and Practice* (6th edition all in
415 colour). Thames & Hudson, London.
- 416 Bélisle, C.J.P., 1992. Convergence Theorems for a Class of Simulated Annealing Algorithms on
417 Rd. *Journal of Applied Probability* 29, 885–895.
- 418 Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on*
419 *Pattern Analysis and Machine Intelligence* 14, 239–256.
420 <https://doi.org/10.1109/34.121791>
- 421 Bookstein, F.L., 1991. *Morphometric tools for landmark data: Geometry and Biology*.
422 Cambridge University Press, Cambridge.
- 423 Buko, A., 1990. *Ceramika wczesnopolska. Wprowadzenie do badań*. Polska Akademia Nauk,
424 Wrocław - Warszawa - Kraków - Gdańsk - Łódź.
- 425 Chang, W., Cheng, J., Allaire, J.J., Yihui, X., McPherson, J., 2019. shiny: Web Application
426 Framework for R.
- 427 Claude, J., 2008. *Morphometrics with R*. Springer.
- 428 Cortez, P., 2014. *Modern Optimization with R*. Springer International Publishing.
- 429 Di Angelo, L., Di Stefano, P., Guardiani, E., Morabito, A.E., 2019. A 3D information framework
430 for automated archaeological pottery archival. 2019 IMEKO TC-4 International
431 Conference on Metrology for Archaeology and Cultural Heritage Florence, Italy,
432 December 4-6, 2019, pp. 178–183.
- 433 Dommergues, C.H., Dommergues, J.-L., Verrecchia, E.P., 2007. The Discrete Cosine
434 Transform, a Fourier-related Method for Morphometric Analysis of Open Contours.
435 *Mathematical Geology* 39, 749–763.
- 436 Douglas, D.H., Peucker, T., 1973. Algorithms for the reduction of the number of points
437 required to represent a digitized line or its caricature. *Cartographer* 10, 112–122.
- 438 Fitzgibbon, A., 2003. Robust registration of 2D and 3D point sets. *Image and Vision*
439 *Computing* 21, 1145–1153.
- 440 Forel, B., Gabillot, M., Monna, F., Forel, S., Dommergues, C.H., Gerber, S., Petit, C., Mordant,
441 C., Chateau, C., 2009. Morphometry of Middle Bronze Age palstaves by Discrete
442 Cosine Transform. *Journal of Archaeological Science* 36, 721–729.
- 443 Genin, M., Dejoie, C., De Parseval, P., Relaix, S., Schaad, D., Schenck-David, J.-L., Sciau, P.,
444 2008. *La Graufesenque (Millau, Aveyron) - Volume II - Sigillées lisses et autres*
445 *productions*, Edition de la Fédération Aquitania.
- 446 Genolini, C., Guichard, E., 2016. K-Means for Longitudinal Data using Shape-Respecting
447 Distance. <https://CRAN.R-project.org/package=kmlShape>
- 448 Gilboa, A., Karasik, A., Sharon, I., Smilansky, U., 2004. Towards computerized typology and
449 classification of ceramics. *Journal of Archaeological Science* 31, 681–694.
- 450 Gualandi, M.L., Scopigno, R., Wolf, L., Richards, J., Buxeda i Garrigos, J., Heinzelmann, M.,
451 Hervas, M.A., Vila, L., Zallocco, M., 2016. ArchAIDE Archaeological Automatic
452 Interpretation and Documentation of cERamics, in: Catalano, C.E., De Luca, L. (Eds.),
453 EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2016), 4 p.

454 Hajek, B., 1988. Cooling schedules for optimal annealing. *Mathematics of operations*
455 *research* 13, 311–329.

456 Hlaváčková-Schindler, K., Kampel, M., Sablatnig, R., 2001. Fitting of a Closed Planar Curve
457 Representing a Profile of an Archaeological Fragment. *Proceedings of the 2001*
458 *conference on Virtual reality, archeology, and cultural heritage*, pp. 263–270.

459 Hristov, V., Agre, G., 2013. A Software System for Classification of Archaeological Artefacts
460 Represented by 2D Plans. *Cybernetics and Information Technologies* 13, 82–96.
461 <https://doi.org/10.2478/cait-2013-0017>

462 Hurth, E., Montuire, S., Schmittbuhl, M., Le Minor, J.-M., Schaaf, A., Viriot, L., Chaline, J.,
463 2003. Examination of the tooth morphospace of three *Mimomys* lineages
464 (*Arvicolinae*, *Rodentia*) by elliptical Fourier methods. *Coloquios de Paleontología* 1,
465 325–334.

466 Jičín, R., Vašíček, Z., 1971. K možnostem srovnávání tvarů na základě podobnosti, in: Bouzek,
467 J., Buchvaldek, M. (Eds.), *Nové Archeologické Metody*. Universita Karlova, Praha, pp.
468 131–139.

469 Kampel, M., Sablatnig, R., 2007. Rule based system for archaeological pottery classification.
470 *Pattern Recognition Letters* 28, 740–747.
471 <https://doi.org/10.1016/j.patrec.2006.08.011>

472 Kampel, M., Sablatnig, R., 2003. Automated Archivation System for Pottery.

473 Kampel, M., Sablatnig, R., 2002. Automated segmentation of archaeological profiles for
474 classification. *Pattern Recognition, 2002. Proceedings. 16th International Conference*
475 *on, Volume: 1*.

476 Kampel, M., Sablatnig, R., 1999. On Estimating the Position of Fragments on Rotational
477 Symmetric Pottery. *2nd International Conference on 3D Digital Imaging and Modeling*
478 *(3DIM '99), 4-8 October 1999, Ottawa, Canada; 01/1999*.

479 Karasik, A., Smilansky, U., 2011. Computerized morphological classification of ceramics.
480 *Journal of Archaeological Science* 38, 2644–2657.

481 Karasik, A., Smilansky, U., 2008. 3D scanning technology as a standard archaeological tool for
482 pottery analysis: practice and theory. *Journal of Archaeological Science* 35, 1148–
483 1168. <https://doi.org/10.1016/j.jas.2007.08.008>

484 Karasik, A., Smilansky, U., Beit-Arieh, I., 2005. New Typological Analyses of Early Bronze Age
485 Holemouth Jars from Tel Arad and Southern Sinai. *Journal of the Institute of*
486 *Archaeology of Tel Aviv University* 32, 20–31.

487 Laflin, S., 1986. Use of a Sinclair Spectrum for Shape Analysis, in: Laflin, S. (Ed.), *Conference*
488 *Proceedings Presented at the Computer Applications in Archaeology 1986*. Centre for
489 *Computing and Computer Science, University of Birmingham*, pp. 83–90.

490 Leese, M.N., Main, P.L., 1983. An Approach to the Assessment of Artifact Dimension as
491 Descriptors of Shape, in: Haigh, J.G.B. (Ed.), *Computer Applications in Archaeology*.
492 *University of Bradford, Bradford*, pp. 171–180.

493 Liu, D., Razdan, A., Simon, A., Bae, M., 2005. An XML-based information model for
494 archaeological pottery. *Journal of Zhejiang University - Science A: Applied Physics &*
495 *Engineering* 6A, 447–453.

496 Lucena, M., Fuertes, J.M., Martinez-Carrillo, A.L., Ruiz, A., Carrascosa, F., 2016. Efficient
497 classification of Iberian ceramics using simplified curves. *Journal of Cultural Heritage*
498 19, 538–543. <https://doi.org/10.1016/j.culher.2015.10.007>

499

- 500 Macháček, J., 2001. Studie k velkomoravské keramice. Metody, analýzy a syntézy, modely.
501 Ústav archeologie a muzeologie. Filozofická fakulta Masarykovy univerzity v Brně,
502 Brno, 296 p.
- 503 Main, P.L., 1987. Accessing Outline Shape Information Efficiently within a Large Database II:
504 Database Compaction Techniques, in: Ruggles, C.L.N., Rahtz, S.P.Q. (Eds.), CAA87.
505 Computer and Quantitative Methods in Archaeology 1987 (BAR International Series
506 393). Oxford, pp. 242–251.
- 507 Main, P.L., 1986. Accessing Outline Shape Information Efficiently within a Large Database, in:
508 Lafin, S. (Ed.), Conference Proceedings Presented at the Computer Applications in
509 Archaeology 1986, pp. 73–82.
- 510 Maiza, Ch., Gaildrat, V., 2005. Automatic Classification of Archaeological Potsherds, in:
511 Dimitri, P. (Ed.), The 8th International Conference on Computer Graphics and
512 Artificial Intelligence, 3IA'2005, Limoges, France, 11/05/2005-12/05/2005. pp. 135–
513 147.
- 514 Mara, H., Sablatnig, R., 2006. Orientation of Fragments of Rotationally Symmetrical 3D-
515 Shapes for Archaeological Documentation. Third International Symposium on 3D
516 Data Processing, Visualization, and Transmission, pp. 1064–1071.
517 <https://doi.org/10.1109/3DPVT.2006.105>
- 518 Martínez-Carrillo, A.L., Lucena, M., Fuertes, J.M., Ruiz, A., 2010. Morphometric Analysis
519 Applied to the Archaeological Pottery of the Valley of Guadalquivir, in: Elwa, A.M.T.
520 (Ed.), Morphometrics for Nonmorphometricians. Springer, pp. 307–323.
- 521 Mom, V., 2005. SECANTO—The Section Analysis Tool, in: Figueiredo, A., Leite Velho, G. (Eds.),
522 The World Is in Your Eyes. Proceedings of the 33rd Computer Applications and
523 Quantitative Methods in Archaeology Conference. Tomar, March 2005.
- 524 Mom, V., Paijmans, J.J., 2007. SECANTO: A Retrieval System and Classification Tool for Simple
525 Artifacts, in: Layers of Perception: Proceedings of the 35th Computer Applications
526 and Quantitative Methods in Archaeology Conference. Berlin, Germany, April 2–6,
527 2007. pp. 1–5.
- 528 Ness, K. L., 2015. Classification Systems with a Plot: Vessel Forms and Ceramic Typologies in
529 the Spanish Atlantic. *International Journal of Historical Archaeology* 19/2, 309-333.
530 <https://doi.org/10.1007/s10761-015-0290-9>
- 531 Orton, C., 1980. *Mathematics in Archaeology*. Cambridge University Press, Cambridge.
- 532 Orton, C., Tyers, P., Vince, A., 1993. *Pottery in Archaeology*. Cambridge University Press,
533 Cambridge.
- 534 Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M.,
535 MacQueen, D., Lemon, J., O'Brien, J., O'Rourke, J., 2020. sp: Classes and Methods for
536 Spatial Data. <https://cran.r-project.org/web/packages/sp/index.html>
- 537 Phillips, S., Walland, P., Modafferi, S., Spagnuolo, M., Catalano, C.E., Oldman, D., Tal, A.,
538 Shimshoni, I., Hermon, S., 2016. GRAVITATE: Geometric and semantic matching for
539 cultural heritage artefacts, in: Proceedings of the Eurographics Workshop on
540 Graphics and Cultural Heritage. pp. 199–202.
- 541 Piccolli, Ch., Aparajeya, P., Papadopulos, G.Th., Bintliff, J., Leymarie, F.F., Bes, Ph., van der
542 Enden, M., Poblome, J., Daras, P., 2015. Towards the automatic classification of
543 pottery sherds: two complementary approaches. CAA 2013. Across Space and Time.
544 Papers from the 41st Conference on Computer Applications and Quantitative
545 Methods in Archaeology. Amsterdam University Press, pp. 463–474.
- 546 R Core Team, 2021. R: A Language and Environment for Statistical Computing.

547 Rice, P.M., 1987. Pottery analysis. A Sourcebook. The University of Chicago Press, Chicago,
548 London.

549 Ripley, A., Beables, B., Bates, D.M., Hornik, K., Gebhardt, A., Fith, D., 2016. Package “MASS”.
550 <https://cran.r-project.org/web/packages/MASS/index.html>.

551 RStudio Team, 2019. RStudio: Integrated Development Environment for R.

552 Saragusti, I., Karasik, A., Sharon, I., Smilansky, U., 2005. Quantitative analysis of shape
553 attributes based on contours and section profiles in artifact analysis. *Journal of*
554 *Archaeological Science* 32, 841–853.

555 Schmittbuhl, M., Allenbach, B., Le Minor, J.-M., Schaaf, A., 2003. Elliptical Descriptors: Some
556 Simplified Morphometric Parameters for the Quantification of Complex Outlines.
557 *Mathematical Geology* 35, 853–871.

558 Schurmans, U., Razdan, A., Simon, A., Mccartney, P., Marzke, M., Van Alfen, D., Jones, J.,
559 Rowe, J., Farin, G., Collins, D., Zhu, M., Liu, D., Bae, M., 2001. Advances in Geometric
560 Modeling and Feature Extraction on Pots, Rocks and Bones for Representation and
561 Query via the Internet, in: Burenhult, G., Arvidsson, J. (Eds.), *Archaeological*
562 *Informatics: Pushing The Envelope*. CAA2001. Computer Applications and
563 Quantitative Methods in Archaeology. Archaeopress, Oxford, pp. 191–204.

564 Smith, N.G., Karasik, A., Narayanan, T., Olson, E.S., Smilansky, U., Levy, T.E., 2014. The
565 Pottery Informatics Query Database: A New Method for Mathematic and
566 Quantitative Analyses of Large Regional Ceramic Datasets. *Journal of Archaeological*
567 *Method and Theory* 21.

568 Turk, G., Levoy, M., 1994. Zipped Polygon Meshes from Range Images. In *Proceedings of*
569 *SIGGRAPH'94* 311–318.

570 Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer, New York.

571 Vaginay, M., Guichard, V., 1988. L'habitat gaulois de Feurs (Loire). Fouilles récentes (1978-
572 1981). *Documents d'archéologie française*, 14. Ed. de la Maison des Sciences de
573 l'Homme, Paris.

574 Venclová, N. 1998. Mšecké Žehrovice in Bohemia. *Archaeological Background to a Celtic*
575 *Hero (3rd – 2nd cent. B.C.)*. Kronos, Sceux.

576 Venclová, N. 2001. *Výroba a sídla v době laténské*. Projekt Loděnice. Archeologický ústav AV
577 ČR, Praha.

578 Wilczek, J., Monna, F., Jébrane, A., Labruère-Chazal, C., Navarro, N., Couette, S., Chateau
579 Smith, C., 2018. Computer-assisted Orientation and Drawing of Archaeological
580 Pottery. *Journal on Computing and Cultural Heritage* December 2018.