Journal of Field Archaeology

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/yjfa20

Laser-Aided Profile Measurement and Cluster Analysis of Ceramic Shapes

Peter Demján, Peter Pavúk & Christopher H. Roosevelt

To cite this article: Peter Demján, Peter Pavúk & Christopher H. Roosevelt (2023) Laser-Aided Profile Measurement and Cluster Analysis of Ceramic Shapes, Journal of Field Archaeology, 48:1, 1-18, DOI: 10.1080/00934690.2022.2128549

To link to this article: https://doi.org/10.1080/00934690.2022.2128549

JOURNAL OF FIELD ARCHAEOLOGY
2023, VOL. 48, NO. 1, 1–18
https://doi.org/10.1080/00934690.2022.2128549

Laser-Aided Profile Measurement and Cluster Analysis of Ceramic Shapes

Peter Demján [1,2], Peter Pavúk [2], and Christopher H. Roosevelt [3]

[1]Institute of Archaeology of the Czech Academy of Sciences, Prague, Czech Republic; [2]Charles University, Prague, Czech Republic; [3]Koç University, Istanbul, Turkey

ABSTRACT
Ceramics are one of the commonest sources of archaeological information, yet their abundance often confounds documentation and analysis. This article presents a new method of documenting and analyzing ceramics that includes laser-aided profile measurement to capture ceramic shape and other information quickly and accurately, resulting in digital outputs suitable for both publication and morphometric analysis. Linked software and database solutions enable unsupervised machine learning to cluster shapes based on similarity, eventually assisting typological analysis. Following an overview of current practices in ceramic recording and both standard and computational shape classification analyses, the new approach is discussed in full as a documentary and analytical tool. A case study from the Middle and Late Bronze Age site of Kaymakçı in western Anatolia demonstrates the benefits of the recording method and helps show that a combination of automated and manual shape clustering techniques currently remains the best practice in ceramic shape classification.

KEYWORDS
digital recording; computational ceramic classification; unsupervised machine-learning; automated shape matching; Kaymakçı; western Anatolia

Introduction

Pottery is the single-most abundant category of archaeological finds recovered from most ancient sites that postdate its invention. The shape of the profile of an individual pot or ceramic vessel or a fragment of one—defined as a cross-section perpendicular to the plane of the rim—is the primary diagnostic attribute used in ceramic shape classification. Classification of ceramic shapes, along with their fabrics, surface treatments, and other characteristics, as well as the stratigraphic relations and radiocarbon dates acquired from their associated deposits, continues to be the most frequently consulted source of information for the relative chronologies and periodizations of pottery-using cultures across the world. With increasingly thorough excavation and careful recording techniques, however, the abundance of pottery presents a challenge, because the number of ceramic fragments from even a single site can reach tens or even hundreds of thousands. Not only are such large assemblages practically impossible to document using traditional hand-drawing techniques, but it is also becoming increasingly difficult to establish accurate and useful shape classifications, especially if ceramic shapes cannot be assigned to pre-existing classification schemes. Several previous studies have addressed these challenges by proposing new methods of digital and 3D documentation and automated classification, but because of bottlenecks resulting from the time required by 3D scanning, they were either applied only to relatively small assemblages or relied on data produced by digitizing hand-drawn profiles from existing publications (e.g., Gilboa et al. 2012; Grosman et al. 2014; Selden, Perttula, and O'Brien 2014).

The first step in any ceramic recording process is usually the assessment of wares, fabrics, surface treatment, decoration, and morphology for any given fragment. Focusing on the latter characteristic, this paper opens up new possibilities for the processing of large ceramic assemblages when it comes to recording and classifying shapes. While the other aspects of recording mentioned above naturally remain very relevant, even in the digital age (cf. Roosevelt et al. 2015), we present here a new technique of rapidly documenting the profiles, diameters, and simple decorations of ceramic fragments, along with photographs, that almost instantly produces digital illustrations. All data recorded in this newly proposed manner is curated in an integrated database that enables easy management, quick access, and display of individual ceramics, as well as groups of ceramics selected according to desired attributes. Furthermore, the approach enables a new automated method of matching ceramics stored in the database according to shape, which potentially enables us to establish computer-assisted shape classifications of large assemblages of ceramics by subdividing them into human-manageable clusters using unsupervised machine-learning techniques. This approach and the associated methods were developed owing to the practical need to process two large ceramic assemblages: one from the Neolithic settlement site at Svodín (southwestern Slovakia, central Europe), excavated in the 1970s and '80s (Němejcová-Pavúková 1995; Demján 2012), and one from the Middle–Late Bronze Age hilltop site at Kaymakçı (Manisa province, western Turkey), undergoing excavation since 2014 (Roosevelt et al. 2018).

Description of this new approach aims to demonstrate the following primary points. First, the shape of a ceramic vessel fragment can be captured precisely using laser-aided profile measurement. Second, the variability of shapes within an assemblage thus measured can be documented and

CONTACT Peter Demján peter.demjan@gmail.com Letenská 123/4, 118 01 Malá Strana, Praha, Czech Republic.
Supplemental data for this article can be accessed at https://doi.org/10.1080/00934690.2022.2128549

computationally analyzed for matching with examples from within or outside the same assemblage based on their profiles. And, finally, the results of these measurement and shape-matching analytical approaches can be combined with the help of database integration to order large assemblages into shape classes. An important element in all of this is the goal of "total" recording of all diagnostic fragments to avoid biases inherent in sampling or other selection procedures. A subsequent brief case study on ceramics from Kaymakçı allows for assessment of the successes and challenges of the new approach, as well as discussion of further possibilities of how this data can be integrated with the remaining dataset from Kaymakçı and beyond.

We describe our methods in three sections. The first deals with the acquisition of pottery profiles and production of digital drawings; the second briefly describes our approach to data storage; and, the third presents how data collected in this way can be used for computer-aided classification of ceramic shapes. Due to the high complexity and specific nature of the collected data and subsequent analytical techniques, it was decided early on that an integrated software solution was necessary and that an open-source solution would ensure replicability, reusability, and scientific scrutiny of results. Our own case study demonstrates the promise of the integrated system just as it points the way forward for next steps.

Laser-Aided Profile Measurement

Standard ceramic shape recording

The standard method of documenting ceramic shapes is still drawing profiles by hand. This is typically achieved either with pencil drawing and/or inking before scanning and further processing in vector- or image-editing software, or with a digitizing tablet, which omits the need for scanning and vectorizing steps. In both cases, the process requires the use of measuring tools such as calipers or profile gauges. Furthermore, the process requires either a professional draftsperson or personal training in drawing, as well as expert but general knowledge of the pottery shapes encountered, to be able to orient the fragments correctly and capture all necessary details. Expert knowledge is especially important for hand-drawing ceramic profiles based on fragments of vessels, where tactile experience of the material and drawing can reap distinct benefits (Morgan and Wright 2018).

For most shapes, the profile is drawn oriented as if the vessel was standing on a flat surface. The correct orientation is determined based on the rim or base of the fragment (if preserved) or on the circular marks left on vessel walls by the process of forming on a wheel. This also applies to any horizontally grooved decoration or sharp profiling, which enable a correct orientation even with body sherds. Hand-drawing can also illustrate surface textures and treatments, such as painted or plastic decoration, and final digital drawings are often combined with photographs. Documenting one vessel or a vessel fragment in this way typically takes between only a few minutes and tens of minutes, depending on the skill and experience of the draftsperson, as well as the size and complexity of the ceramic. The resulting illustration can be used in a printed or electronic catalog, which can then be visually compared with the catalogs of ceramic illustrations from other sites to assess similarities. For more in-depth metric analysis, or computational shape matching, the profile has to be extracted from the drawing by using manual or semi-automated vectorization techniques that usually involve some degree of programming (e.g., the OpenCV library; Bradski 2000). In this work, it is also important to consider the scale of the drawing so that the vectorized profile retains correct units.

As 3D recording technologies become more and more accessible, they have become established as regular parts of some archaeological documentation workflows (e.g., Seguchi and Dudzik 2019). With photogrammetric techniques, for instance, it is possible to produce a 3D model of an artifact with sub-millimeter precision using only a common digital camera (e.g., Luhmann et al. 2019; Göttlich et al. 2021). Other techniques include laser or structured-light scanning (e.g., Bitelli et al. 2020). The three-dimensional nature of most archaeological evidence naturally calls for such methods of documentation. Depending on the technology used and the size and complexity of the artifact, however, acquiring and subsequently processing the 3D model takes, again, at least tens of minutes (with most 3D-modeling methods requiring subsequent "cleaning"). Göttlich and colleagues (2021) report an average processing time of 37.5 minutes per ceramic fragment but also note that only 2.75 minutes of this time is needed for data capture (photography) itself. Like 2D illustrations, 3D models can then be used in electronic catalogs, can produce rendered representations of the artifact (requiring further processing), or can be directly employed in metric and/or spatial analysis. This approach also allows the production of high-precision drawings in semi-automated ways using specialized software (Karasik and Smilansky 2008). Creating a profile drawing from such models, however, takes additional skill and time. When high-resolution 3D scanning becomes more affordable and applicable in environments where light, temperature, and dust conditions cannot be controlled, we expect it to become standard in ceramic documentation, especially with published case studies demonstrating the production of between 15–20 (Karasik and Smilansky 2008), 40 (Roosevelt et al. 2015), and 50 (Poblome et al. 1997) ceramic illustrations per day (cf. Karasik et al. 2014, 210).

Computer aided drawing

Two key steps for producing a reconstruction drawing of a ceramic vessel based on a rim fragment are capturing the profile and calculating the diameter. Traditionally, the profile is drawn by sight, tracing the profile, or using a profile gauge, and the diameter is estimated by matching the arc of the preserved portion of the rim to a diameter chart, a prepared set of concentric circles of known diameters. The profile drawing can then be digitized and converted into vector format, allowing for computational analysis. A more direct approach is digitization of the physical fragment itself. Karasik and Smilansky (2008) devised a method to extract profiles and diameters from 3D scans of ceramic fragments, which can be acquired using either structured-light scanning or photogrammetry. Challenges related to maintaining appropriate lighting and climatic conditions for 3D scanners, as well as the relatively long model-capture and processing times required for each fragment, however, limit the applicability of this approach for use at excavations producing large quantities of new finds each season.

For these reasons, we decided to adopt a new method that had already been employed successfully for processing Neolithic pottery (Demján 2016) and that promised functionality in a broader range of lighting and climatic conditions, as well as faster processing with no loss of quality. The new approach uses static, industrial-grade laser line projectors and cameras to capture only the vessel profile and horizontal curvature directly. This device enables the operator to scan profiles even under strong ambient light and is portable and robust enough to sustain long-term operation in a dusty and hot environment. The instrument, called the Laser-Aided Profiler and developed by Peter Demján and Vladimir Držík (2018), consists of a frame holding one color and two monochrome cameras, two laser line projectors, one light source, a glass pane to support the captured fragment, and associated cabling and control circuitry (Figure 1). A USB 3.0 port ensures power supply and data connection to a personal computer on which dedicated software is used to control the process.

The maximum dimensions of a captured profile are 250 × 100 mm, but a stitching function allows for acquisition of profiles of theoretically unlimited length. This function can also be used to combine multiple captures of a profile in case of more complex shapes where the laser cannot illuminate the whole profile at once. Besides profiles, the device can also capture color photographs of fragments at a resolution of 2592 × 1944 pixels. Using the dedicated control software, the user first captures the profile of a fragment via the projection of a laser line that is captured by the monochrome cameras and rendered on screen in real time. In the next step, the user rotates the profile to orient it properly, which can be assisted by placing the fragment on its rim on the glass pane and using the partially captured outline as a guide. Next, the user calculates the diameter of the vessel at its rim, perpendicular to the profile, via the projection of a laser line parallel to, and thus representing, the rim curvature. From these measurements, the basic shape of the vessel is immediately reconstructed, and the user can add additional details typical to ceramic profile drawings in archaeology, such as inflection lines, handle sections, illustration of plastic decorations, etc., in addition to color photographs captured by the color camera. Optional textual and numeric descriptors relating to the fragment or its context can also be entered here (Figure 2). Following this method, a skilled user can create over a hundred digital drawings per day, exportable in commonly used formats (e.g., JPG, SVG, DXF, and PDF) and ready for further analysis and publication.

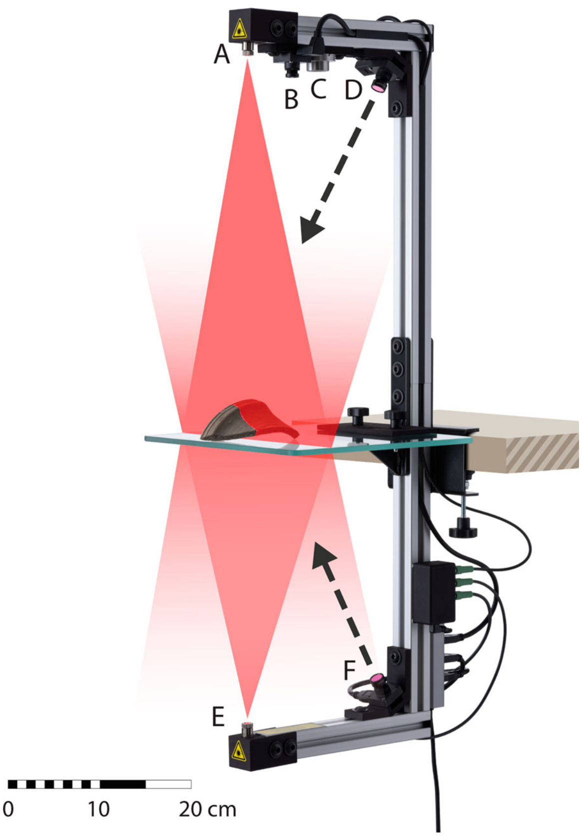

Figure 1. The Laser-Aided Profiler hardware with schematic representation of its function. The device consists of A, E) two laser line projectors; D, F) two monochrome cameras which capture the projected laser lines; and, B) one color camera with C) associated LED diode to photograph the finds and capture their outlines. Red areas represent projection planes of the laser lines. Dashed arrows represent the monochrome cameras' directions of view.

Database-Integrated Storage and Descriptive System

The control software of the Laser-Aided Profiler is linked to the custom graph database management system Deposit (Demján 2021a), where all elements of the drawing are stored separately and can be easily retrieved for analytical purposes or further editing. Using a graph database allows us to maintain high granularity of the data, in which all components of a drawing are recorded as individual, interlinked records (Figure 3).

Stored drawings can be accessed as a whole to generate catalogs in PDF format, but one can also retrieve individual components of a drawing, such as the profile or the arc capturing the vessel's curvature, all stored as coordinates in Well-Known Text (WKT) format at correct scales with sub-millimeter precision.

The Deposit software was developed together with the control application for the Laser-Aided Profiler to serve as a data storage and management platform at both the collection as well as analysis stage. The decision to create a custom solution was due to the specific requirements such software has to fulfill, which were not covered by any open-source and free data-management package.

At the data collection stage, it is important to maintain a high granularity of the data. Different components of the drawing have to be stored separately using one-to-many relations (a ceramic vessel has one main profile line and can contain multiple additional details, photographs, information on context, etc.). As the project progresses and additional archaeological information (stratigraphic, typological, spatial, etc.) becomes available, we will need to flexibly modify and extend the database schema. For such purposes, the graph database format is much more suitable than a relational format. Specifically, we use a directed graph that has a hierarchical structure that can be split into sub-trees which can again be merged into other trees, allowing for a streamlined merging of collected datasets to a main database or the extraction of subsets for purposes of data analysis or exchange.

A graph format is also beneficial at the analysis stage, which will be more closely described in subsequent sections. Automated shape matching produces dissimilarity matrices which can then be used for hierarchical clustering of the shapes. The resulting hierarchical relations between the shapes and their membership in clusters, as well as the

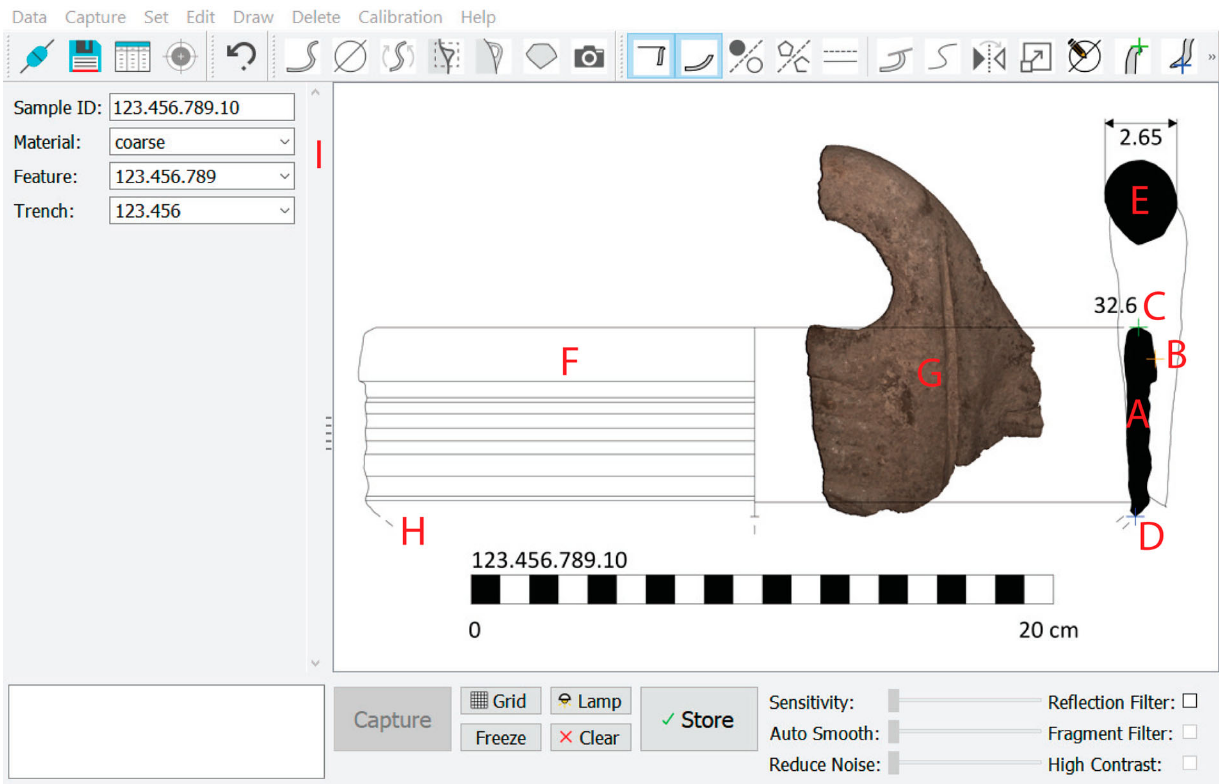

Figure 2. The Laser-Aided Profiler Control Application interface. Elements of the drawings are stored as separate interlinked records: A) profile; B) curvature arc and its point of capture; C) rim point; D) bottom point; E) additional details (e.g., handle section); F) inflection lines; G) photograph; H) profile break lines; and, I) optional additional descriptors.

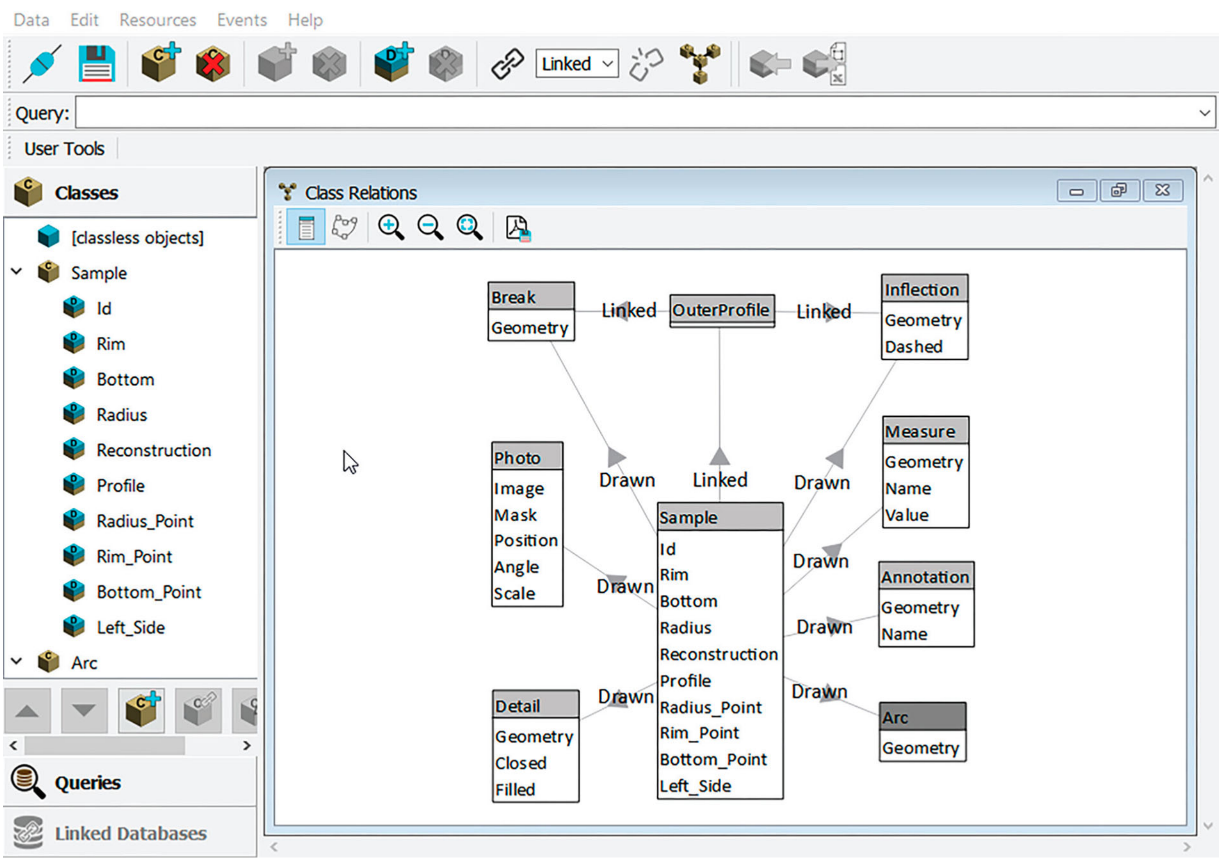

Figure 3. Deposit database management software interface showing a visual representation of the database schema generated by the Laser-Aided Profiler Control Application.

measures of dissimilarity between each pair of shapes, can also be stored as weighted relations in a graph database. This structure can be easily extended as new shapes are added. The tree structure produced by automated hierarchical clustering can further serve as a basic template for subsequent classification of the ceramic vessels either manually or using additional stratigraphic, chronological, or typological data. All standard graph operations can be performed directly on the dataset, without the need to convert its structure. While all this is technically possible to achieve using a relational database format, where links are represented using join tables, there would be major performance issues with relationship queries more than a couple of steps deep.

The reason we did not opt for an existing graph database solution was that we needed a free, open-source system capable of working both as a client-server where the storage is kept on a SQL server as well as on individual computers out of the box, without the need to set up a specific environment. Another technical requirement is the ability to store the data both in a graph (node-link) format and as a standard relational database for purposes of archiving and easy access by external applications. The data queries can also be exported in standard spreadsheet (CSV or XLS) formats.

Automated Shape Matching and Classification

Ceramic shape classification

Ceramic shape classification should never be an end in and of itself but must always be a means for gaining deeper insights. Various forms of seriation based on ceramic shapes were introduced to archaeology quite early, yet while attempts have been made to conceptualize and formalize these approaches along the way (summarized most recently in Albero Santacreu, Calvo Trias, and Rosselló 2017 and Bartolini 2017; cf. also Plog 1983, Dunnell 1986, and Orton and Hughes 2013, all of which make clear that the issues raised here have remained unresolved for years), many archaeologists practice them rather intuitively or simply follow their teachers, who in turn followed their teachers, as is more common in academic archaeological practice. Experience shows, however, that producing a useful, workable classification, which does not become one's master but rather provides a good service, is not as easy as it may seem and lies somewhere between art and science.

In what follows here, we discuss a specific type of classification that targets the chronological development of shapes, because it is vessel shapes and their accurate and efficient documentation that stand at the center of this work. An efficient classification should fulfill several criteria: it should classify the material at hand; it should allow for subdivisions, but not too many, or else it becomes too difficult to remember; it should be structured in such a way so as to enable a better grasp of the chronology of the materials it represents (not only their functional differences); and, it should be replicable by other experts, ideally at other sites. What makes a classification most useful and efficient, then, is its structure, and each type of material, in combination with its chronology and degree of complexity, potentially needs a different structural approach. To a certain degree, it is intuitive and requires a certain feeling or talent, and also flexibility, in lumping and splitting like and unlike. Whereas some categories of finds have naturally or functionally simple classification structures, ceramic shapes are certainly not one of them. For ceramics, possibly more so than for other fields of material production, human creativity factors strongly—not necessarily that of the modern humans classifying it, but that of the ancient potters who produced the materials that have become part of the archaeological record. Depending on the level of complexity and the specialization of everyday activities in a given society, we see different ranges of shapes and decorations. With more complex pottery-producing societies, we typically see large ranges of said qualities in the output of pottery, also possibly reflecting multiple distinctive potters or workshops. It is this human input and unpredictable degrees of creativity and/or changing fashions that make automated classification of shapes so difficult.

In developing a chronologically sensitive shape classification, one can start with more general classes, such as plates, cups, bowls, pots, jars, jugs, and so on. These can be subdivided into more specific classes—for instance, globular cups, ogival cups, and carinated cups—based on the curvature of the body and/or depending on other diagnostic features, such as the shape of the lip, the type and number of handles, or decoration, and variants thereof. All such classes and sub-classes then need to be checked against stratigraphy (whether vertical or horizontal) to identify the chronological significance of shape changes. Even if some general shapes can remain in use over centuries, most shapes do change over time in some identifiable way. Within a broader evaluation, one should also consider correlations between shape and fabric, ware, and decorative motifs, combinations of which can also show development over time. This is where flexibility in lumping vs. splitting, grouping vs. dividing is crucial, so that too many shape sub-classes are not created. Ideally, a well-defined shape in a given fabric or ware should be constrained to one or two chronological phases so that it remains a valuable chronological tool, especially for synchronization, whether between trenches or between sites. Such a shape is then termed "chronologically sensitive," and a full archaeological assemblage might have a range of such shapes. Finding the appropriate middle road between grouping and dividing, too general and too specific, then, is often a sensitive iterative process of trial and error.

Computational classification

Methods for more formalized classifications of ceramic vessels based on statistical analysis of quantified nominal attributes of shape and decoration were proposed as early as the 1950s (Spaulding 1953). With the wider availability of computers and digitization technology, it also became possible to capture and computationally analyze vessel shapes based on their measured geometry (Turpin and Neely 1977; Gero and Mazzullo 1984; Durham, Lewis, and Shennan 1995). While earlier techniques relied on completely preserved vessels, later approaches also allowed working with sufficiently preserved rim fragments (Gilboa et al. 2004; Karasik and Smilansky 2008, 2011; Smith et al. 2014). These techniques, which can be described generally as shape matching, depend on calculating the degree of (dis)similarity between individual vessels based on the shapes of their profiles. A dissimilarity matrix can be constructed using these values, which can in turn be used as the input for cluster analysis, thus assigning each vessel into a cluster of similar specimens. These clusters can then be interpreted as shape classes. The calculated degree of similarity can also be used to determine the assignment of a new find to an already established class or sub-class by comparing it with already assigned finds. Generally, this approach represents so-called unsupervised machine learning; that is, it does not require an already-established classification schema and a training dataset to assign the finds to classes based on their shape. While supervised machine learning approaches, concentrating mainly on decorated pottery, are showing promising results in recent years (Gualandi, Gattiglia, and Anichini 2021; Pawlowicz and Downum 2021), studies using unsupervised approaches remain scarce.

One reason we still see no widespread use of these techniques is their reliance on large datasets, ideally drawn from several sites to cover the full spectrum of shapes in particular regions and periods. These datasets must contain vessel profiles represented at correct scales as vector polygons, which can be further used for mathematical shape analysis. To create such datasets, robust, easy-to-deploy technologies are needed for acquiring vessel profiles and producing vectorized drawings in large numbers. One of the aims of the methods presented here—coupling laser-aided profile measurement with a specialized database management system—is to enable archaeologists to achieve such goals within convenient time spans.

Automated interactive shape clustering

Quantifying similarity

Automating the process of identifying ceramic fragments with similar profiles, the basic prerequisite of a shape classification, requires a method of quantifying the similarity (or dissimilarity) of two profiles based on their geometric representations. Such a method has to combine the differences between the sets of coordinates of each profile mathematically into a single expression in a way that best captures the character of each overall shape. Saragusti and colleagues (2005) proposed a function that compares a specific geometric property of the profile curve (e.g., radius, tangent, or curvature) along the arc-length of the curve—that is, the distance drawn along the surface of the profile from a zero point set at the rim (Figure 4A). This approach proved to be well suited for pottery shapes with relatively uniform thickness along the whole length (Karasik, Smilansky, and Beit-Arieh 2005). For more complex shapes, however, such as vessels with a strongly pronounced lip, comparison based on arc-length proved problematic. The problem arises when the thickness of one or both of the vessel profiles fluctuates, which results in the function selecting points for comparison that do not correspond to the same parts of each profile and hence should not be compared (Figure 4B). To solve this problem and enable comparison of profiles with variable thickness or pronounced lips, we need to use a function that is independent of profile thickness. Accordingly, we simplify each profile to a central axis that runs from the rim to the lowest preserved extent and use it as a reference line for subsequent comparisons.

To construct the geometric axis of a profile (Figure 5A), we first rasterize and reduce it to a topological skeleton (Figure 5B). The skeleton is subsequently vectorized into a set of non-branching curves using a shortest-path algorithm (Figure 5C). In the final step (Figure 5D), the short curves are pruned, and the remaining ones are joined to form the profile axis. The axis calculation is implemented in the profile_axis function in the fnc_matching module of the CeraMatch application (Demján 2021b; see description in the next section).

In order to take into account different aspects of the similarity between two profiles (e.g., Figure 6A), the application calculates four different metrics of dissimilarity: diameter, axis, Dice, and rim Dice. All dissimilarity functions are implemented in the fnc_matching module of the application (Demján 2021b). Diameter dissimilarity is the average ratio of the horizontal distances of points along the axes of the profiles, measured so that the zero point of both profiles on the x-axis is the axis of rotation (the horizontal midpoint of the vessel) and the zero point of both profiles on the y-axis is the rim. Only those points where both profiles overlap on the vertical axis are considered (Figure 6B; Equation 1). The equation can be expressed as follows:

$$D_{\mathrm{diameter}} = 1 - \frac{\sum_{y=0}^{ymax} ax_y}{\sum_{y=0}^{ymax} bx_y} \tag{1}$$

where *ymax* is the maximum y-coordinate (measured from the zero point) and ax_y is the smaller and bx_y the larger of the two x-coordinates of the profiles at a specific y-

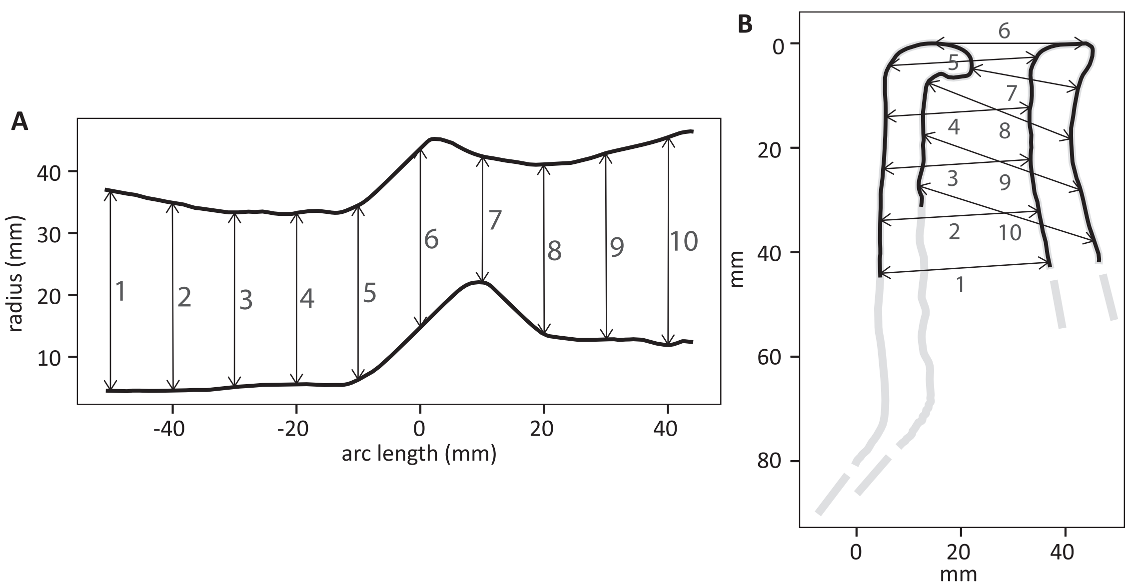

Figure 4. Radius function A) using arc length as a reference to compare two profile curves, with B) illustration of which parts of the profiles are compared.

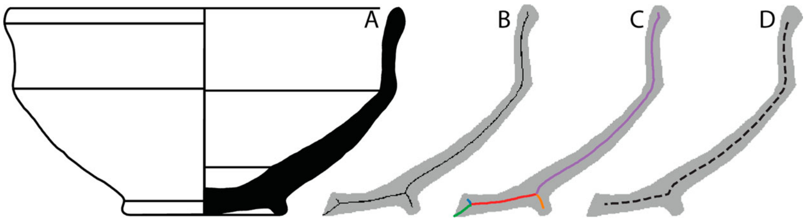

Figure 5. A) Extraction of the axis of a profile by B) topological skeletonization, C) shortest-path analysis, and D) pruning.

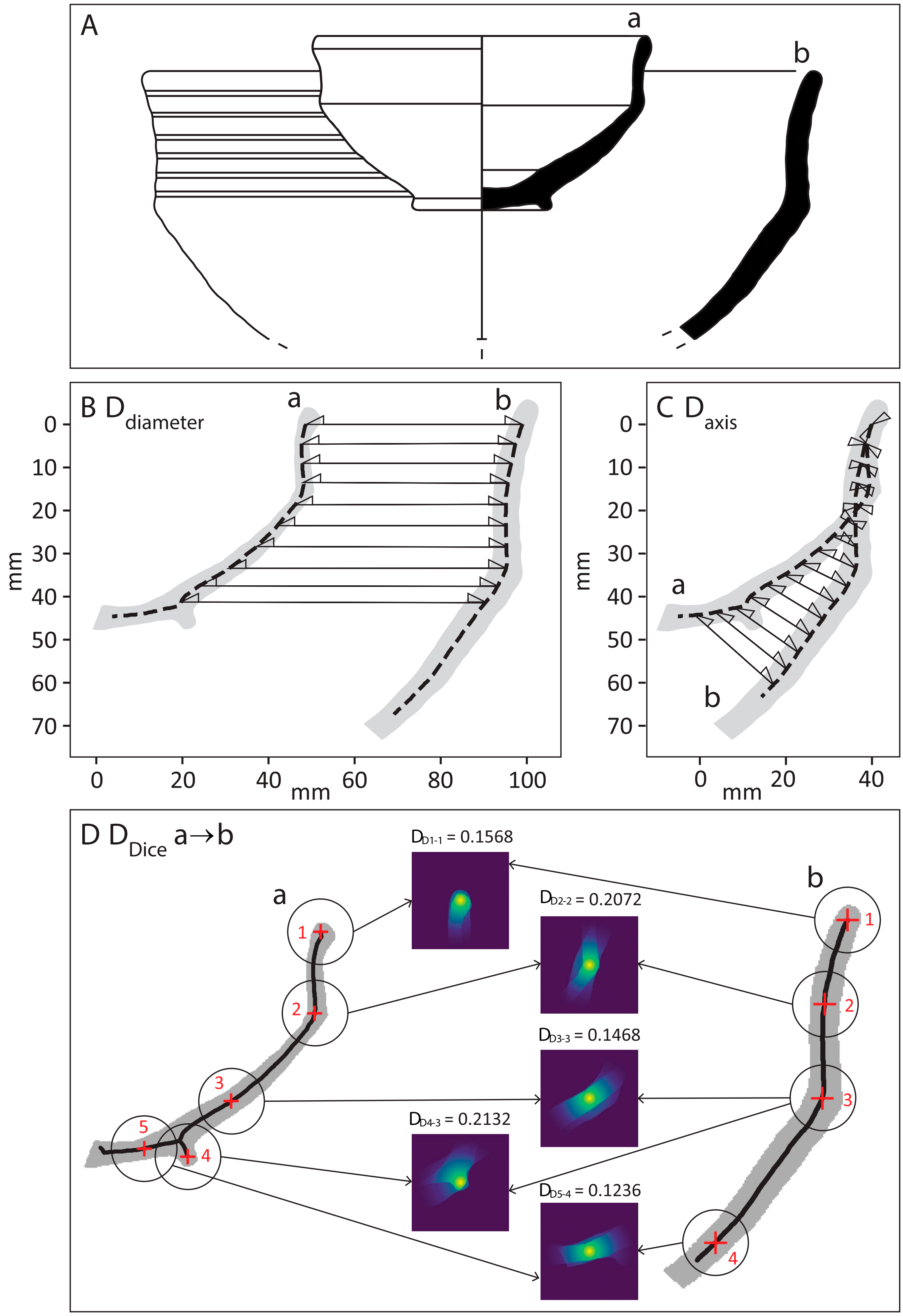

Figure 6. A) Different metrics of similarity between profiles of two vessels: B) diameter dissimilarity; C) axis dissimilarity; and, D) Dice dissimilarity, with key points illustrated in red.

coordinate. Diameter dissimilarity is calculated in the diameter_dist function of the application.

Axis dissimilarity is the average Euclidean distance of each point of one profile axis to the point of the second profile axis equidistant along the axis from the zero coordinate. The zero coordinates of both axes are set to their uppermost point. The result is normalized by the maximum possible dissimilarity of two axes of the same given length —that is, the value one would get if the profiles were straight and set horizontally opposite each other (Figure 6C; Equation 2). The equation can be expressed as follows,

$$D_{axis} = \frac{\sum_{d=0}^{dmax} \sqrt{(ax_d - bx_d)^2 + (ay_d - by_d)^2}}{\sum_{d=0}^{dmax} d^2} \tag{2}$$

where d is the distance along the respective axis from the zero point, $dmax$ is the length of the shorter axis, (ax_d, ay_d) is the coordinate of the axis of the first profile, and $(bx_d, by_d)$is the coordinate of the second profile at distance d from the zero point. Axis dissimilarity is calculated in the axis_dist function of the application.

Dice dissimilarity is a metric based on calculating Dice's coefficient, where the similarity of two samples is expressed as double their area of overlap, divided by the total area of both samples. In this implementation, the application compares two equally long sections of two profiles and calculates the area of their overlap. The sections are selected along the profile axis at specific key points where its curvature changes. The key points are identified according to the following algorithm (see also function find_keypoints): 1) For each point of the profile, take a segment centered around this point with the length of twice the average profile thickness. Calculate profile curvature at this point as the area between the curve and the line connecting the first and last point of the segment. 2) Assign curvature to each point of the profile axis as the higher of the curvature values of the closest points of the inner and outer profile. 3) Find key points by proceeding from axis points with highest to lowest curvature, always leaving a gap equal to at least one average profile thickness between two key points.

During the matching, each profile section is rotated around the key point until the best match (lowest Dice dissimilarity) is achieved. This is done so that the function is independent of the overall orientation of the profiles. Each section from one profile is matched against nearby sections from the other profile, keeping only the best match. The results are weighted by distance from the key point to emphasize the center of the section and reduce mismatching resulting from cutting of the profiles at the section boundaries (Figure 6D). The results are further normalized by the weighted and summed areas of both sections. The root mean square (RMS) of all results for two profiles, calculated for the first versus the second profile and then vice versa, is the final Dice dissimilarity of those profiles (Equation 3). The equation can be expressed as follows,

$$D_{Dice} = \sqrt{\frac{\sum_{p=1}^{m} 1 - \frac{2I_p}{A_p + B_p} + \sum_{q=1}^{n} 1 - \frac{2I_q}{A_q + B_q}}{m + n}} \tag{3}$$

where p is one of m key points of the first profile, and q is one of n key points of the second profile. $I_p$and I_q are the weighted areas of overlap of profile sections at key points p and q. A_p and A_q are the weighted areas of the section of the first profile at the respective key points, and B_p and B_q are the same for the second profile. Dice dissimilarity is calculated in the dice_dist function of the application.

The application also calculates Dice dissimilarity separately for the rim. The process is similar to that used for the whole profile, except only the first key point (closest to the rim) is used. The purpose of this separate calculation is to emphasize the importance of rim shape in the final matching result.

Computational classification

As mentioned above, the definition of a ceramic shape class or sub-class depends on a multitude of criteria and requires a deep knowledge of the material and its historical context. The ability to find similar shapes in a large group is just one aspect of this process; with increasing assemblage sizes, however, it becomes an almost insurmountable obstacle. The automated clustering technique presented here serves to overcome this initial challenge by pre-sorting the material in a way that puts visually similar shapes close together in a hierarchical tree structure. To enable a ceramics expert to achieve a valid classification, we developed the CeraMatch application (Demján 2021b), which serves as an interface to the integrated custom Deposit database of digitized fragments and has three basic functions: calculate (dis)similarity between samples; perform automatic clustering; and, allow users to rearrange clusters freely until they represent a valid classification. Furthermore, CeraMatch quantifies the degree of similarity of each fragment to a certain class, allowing for a probabilistic approach to classification in which each vessel profile has quantifiable probabilities of belonging to particular classes or sub-classes.

The application uses a hierarchical cluster analysis (HCA) algorithm to assign each fragment into a cluster of similar specimens, a prerequisite for classification. The algorithm requires the input of a dissimilarity matrix to calculate distances between newly formed clusters. In the previous steps, the dissimilarity of a particular profile (e.g., Figure 7A) to others was calculated by four different metrics of dissimilarity (diameter, axis, overall Dice, and rim Dice; Figure 7B–D). These metrics can be recorded in the form of a $n \times n \times 4$ matrix, where n is the number of fragments in the dataset.

The standard HCA algorithm requires a method of calculating the distances between newly-formed clusters, for which a single dissimilarity metric is needed. The first step is to reduce the dissimilarity matrix to $n \times n$ dimensions. This is achieved by normalizing the four metrics of dissimilarity $D_{ij1..4}$ for each of a set of n fragments (Equations 4 and 5) and calculating the RMS to combine the metrics into one value (Equation 6; Figure 7E).

$$D' = D - \begin{pmatrix} \min\limits_{i,j \in [1,n]} D_{ij1} \\ \min\limits_{i,j \in [1,n]} D_{ij2} \\ \min\limits_{i,j \in [1,n]} D_{ij3} \\ \min\limits_{i,j \in [1,n]} D_{ij4} \end{pmatrix} \tag{4}$$

$$D'' = \frac{D'}{\begin{pmatrix} \overline{D_{ij1}} \\ \overline{D_{ij2}} \\ \overline{D_{ij3}} \\ \overline{D_{ij4}} \end{pmatrix}} \tag{5}$$

$$D_{RMS} = \sqrt{\frac{\sum_{k=1}^{4} D''^{2}_{ijk}}{4}} \tag{6}$$

The RMS combined dissimilarity is useful for finding the closest match for each fragment. To arrive at a reasonable clustering, however, only those shapes that are sufficiently similar according to all metrics should be grouped together. In this way, the application mimics the way archaeologists perform shape classification, where vessels with relatively large differences in diameter or profile orientation, for instance, can be considered similar and of the same type if they match according to other aspects, such as rim shape or overall profile curvature. The application takes this variability into account in its HCA algorithm by introducing a limit parameter when calculating distances between newly formed clusters. The limit parameter determines the maximum normalized dissimilarity D'' between fragments for them to be considered members of the same cluster. If the limit condition is met, the application uses the distance between cluster centroids to determine which two clusters are closest and thus which are the next to be joined in the hierarchy. The modified HCA algorithm uses the following formula to calculate the distance between centroids c_I and c_J of clusters I and J, implemented in the get_clusters function of the application (Equation 7):

$$\|c_I - c_J\| = \begin{cases} \sqrt{\dfrac{\sum_{\substack{i,j \in I \cup J \\ i<j}} D^2_{RMS\,ij} - (|I| + |J|)\left(\dfrac{\sum_{\substack{i,j \in I \\ i<j}} D^2_{RMS\,ij}}{|I|} + \dfrac{\sum_{\substack{i,j \in J \\ i<j}} D^2_{RMS\,ij}}{|J|}\right)}{|I||J|}}, & \max\limits_{i,j \in I \cup J} D''_{ij} \le limit \\ \infty, & \max\limits_{i,j \in I \cup J} D''_{ij} > limit \end{cases} \tag{7}$$

The number of clusters formed depends on the limit criterion, which can take values between 0 and 1. In our experience, it is best to start with a value of $limit = 0.68$ and to experiment by increasing or decreasing it in increments of 0.05. This approach proved more intuitive than the required manual selection of the number of clusters when using the standard HCA algorithm, because it is independent of the size and variability of the dataset.

The dissimilarities calculated between each pair of recorded fragments are stored in the Deposit database in the form of weighted relations. Because the calculation for a large dataset (1000 + fragments) can take several days, this method of storage allows for calculation of dissimilarities for newly added samples separately, without the need to re-calculate the whole dataset. Clustering results are also stored in the form of a tree graph, linking the samples in a hierarchical structure that can be used to visualize the clustering without the need for re-calculation (Figure 8). Full database integration also ensures the availability and portability of the complete dataset and easy replicability of the results of shape analysis.

After clustering, the samples (ceramic fragments) are represented as drawings arranged in a tree structure (dendrogram), in which branches can be freely moved, individual samples can be reassigned to different clusters, and new clusters can be formed (Figure 9). It is also possible to import a (partial or complete) cluster assignment in the form of a spreadsheet table, which is then used as the basis to construct a dendrogram and assign previously unprocessed samples to the clusters with the most similar shapes.

To sum up the shape classification process, it can follow two approaches. First, the application can build a classification "from scratch" by auto-clustering the shapes of better-preserved fragments. Second, it can assign new fragments to previously established classes based on the degree of similarity to individual shape clusters. These approaches can be alternated as the dataset grows and new classes or sub-classes are established.

Ceramics from Kaymakçı: A Brief Case Study

Kaymakçı is a fortified hilltop with associated extramural settlement situated in central western Anatolia, on the western flanks of the Marmara Lake Basin, north of the Gediz River. The site was discovered during a regional survey in 2001, was intensively documented via non-invasive methods by the Central Lydia Archaeological Survey (CLAS), and has been the focus of excavations under the Kaymakçı Archaeological Project (KAP) since 2014 (Roosevelt and Luke 2017). The primary stratigraphic and architectural phases identified so far reflect the most intensive habitation during the local Late Bronze Age (LBA) in the middle centuries of the 2nd millennium B.C., yet finds and absolute dates demonstrate activities stretching back into the Middle Bronze Age, at least to the beginning of that millennium. Seven excavation areas dispersed over a fortification system, an inner citadel and surrounding slopes, and a broad and densely occupied terrace reflect at least defensive, ritual, storage, domestic, and productive activities and abundant ceramic finds deriving from them (Luke and Roosevelt 2017; Roosevelt et al. 2018) (Figure 10).

The ceramic assemblage of Kaymakçı is principally embedded in local western Anatolian traditions. Because it is the first systematically excavated assemblage in the immediate region from this period, there is no local and pre-existing ceramic classification with which to compare it. Slightly further afield, the quite well-defined typo-chronological systems of contemporary Minoan and Mycenaean pottery on Crete and the Greek mainland across the Aegean Sea (Mountjoy 1998; Hallager 2010; Rutter 2010) provide some comparative value, with Troy at the northern tip of the western Anatolian coast (Pavúk 2014) being a somewhat better match. Closer to our region of interest are sites in the gulf of Izmir, in the central part of the western Anatolian coast (see below), but other coastal sites still await full publication. The best-known site inland is Beycesultan in the Upper Meander valley, often considered southwestern Anatolia (Lloyd and Mellaart 1965; Mellaart and Murray 1995; Dedeoğlu and Abay 2014). While again of some typologically

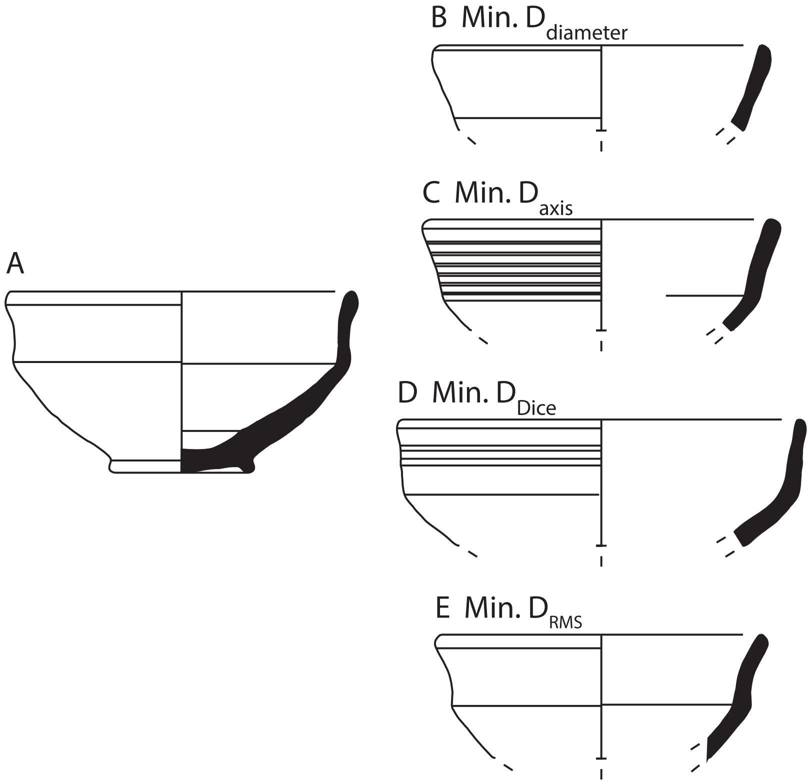

Figure 7. Example of how different metrics of dissimilarity identify different "most similar" vessels. At right are the shapes most similar to A) a vessel according to B) diameter, C) axis, D) Dice, and E) combined dissimilarity using root mean square.

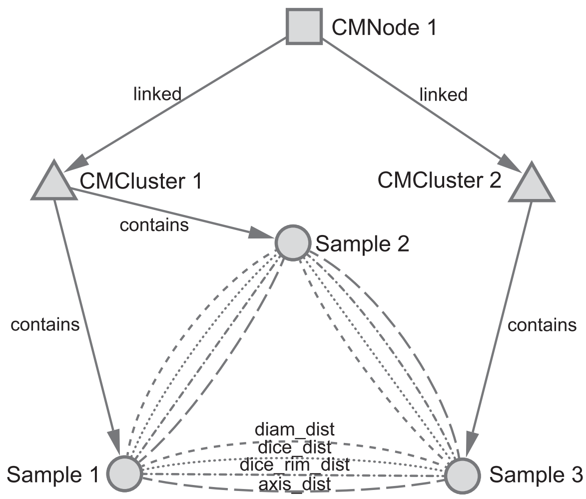

Figure 8. Deposit database schema of storing dissimilarity metrics between samples and clustering results. The graph nodes represent Deposit objects (equivalent to records in a relational database); the edges represent relationships between objects. The objects are members of different classes (equivalent to tables in a relational database). The CMCluster class (triangles) represents clusters generated in the CeraMatch (CM) application, the CMNode class (square) represents nodes in the clustering hierarchy tree, and the Sample class (circles) represents individual ceramic samples. The weighted relations diam_dist (dashed line), dice_dist (dotted line), dice_rim_dist (dot-dash) and axis_dist (long-dash) represent different metrics of dissimilarity.

comparative value, the Beycesultan assemblage is really that of an inland western Anatolian ceramic province. Further east, Late Bronze Age pottery assemblages are dominated by north-central Anatolian or Hittite traditions (Glatz 2009; Schoop 2011; Mielke 2017), which are radically different in terms of both production and shape classes, as well as functional use.

The closest comparative proxies for ceramics from Kaymakçı, then, remain those of the coastal zone around the bay of Izmir and, to a lesser degree, Troy. Ceramic traditions at these sites are broadly known, but their locations on the Aegean coast result in substantial influence from across the Aegean Sea, as well as actual imports (Ersoy 1988; Mangaloğlu-Votruba 2015; Erkanal-Öktü 2018; Aykurt 2020). However, Aegean influenced pottery and actual imports are rather rare further inland, yet this may partly result from the lack of excavations and the fact that the majority of inland western Anatolian ceramics is known only from surveys (summarized in Pavúk 2015; see also Roosevelt and Luke 2017; Pavúk and Horejs 2018). In the absence of stratigraphic sequences, full reliance on survey ceramics even gave some the impression of the near absence of developed LBA activities in the area until recently. The importance of Kaymakçı in this respect is clear. The 2nd millennium B.C. ceramic wares and fabrics of Kaymakçı and the Marmara Lake basin had already been defined according to surface survey ceramics prior to excavation (Luke et al. 2015), yet the full variety of shapes still had to be defined. For this, a more robust dataset of excavated samples was needed.

Ceramic documentation

From its inception, the Kaymakçı Archaeological Project aspired to rapid and holistic recording of excavated pottery, simultaneous with the process of excavation (Roosevelt et al. 2015). Following washing, sorting, counting, and weighing according to ware groups, rims, handles, and bases are separated for more detailed documentation. At first, a system of 3D laser scanning of selected diagnostic pieces was initiated, focusing on rims because of their diagnostic potential, along with extraction of 2D profile drawings via Karasik and Smilansky's (2008) software. This resulted in highly accurate profile drawings, yet it quickly became clear that

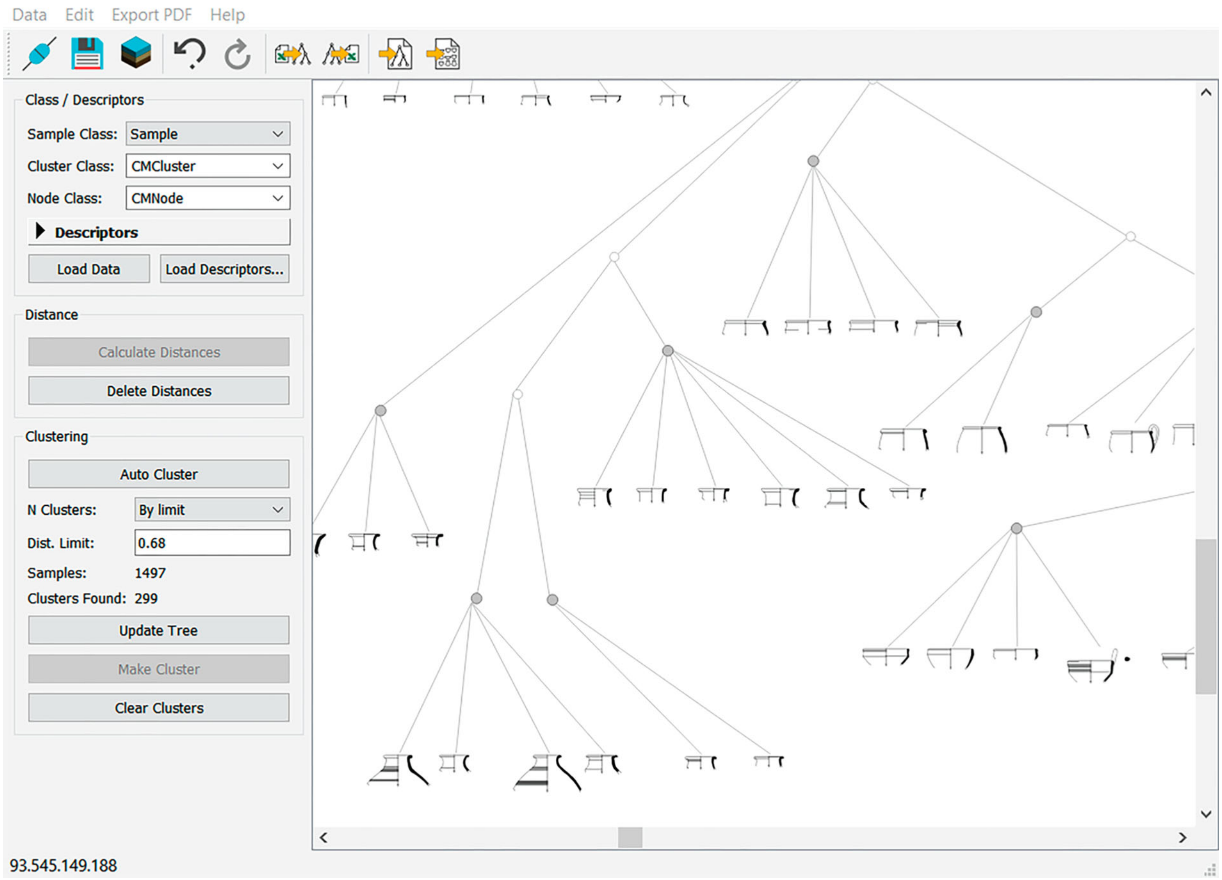

Figure 9. The CeraMatch application used to calculate, visualize, and modify the clustering of ceramic shapes.

the speed of our workflow was insufficient for handling the volume of material recovered from a site like Kaymakçı, with tens of thousands of diagnostic fragments excavated each year. While around 40 ceramics could be recorded via 3D laser scanning a day (Roosevelt et al. 2015, 336), subsequent software processing into profile drawings added more time to the process, resulting in our inability to maintain the desired pace of recording.

The workflow improved in 2017 with the introduction of laser-aided profile measurement, as described above. That year, a total of 1188 fragments were recorded. After familiarization with recording routines, 70 drawings on average were produced per day, with a maximum of 150 per day using one device operated by one person. Increasing familiarization and fine-tuning of the instrument and software resulted in 837 drawings in the 2018 season (over half the number of days, with an average of 76 drawings per day), 1693 drawings in 2019 (with an average of 81 and a maximum of 254 drawings per day), and 775 drawings in 2021 (with an average of 48 and a maximum of 112 drawings per day). In total, 4493 profile drawings have now been generated from the assemblage of Middle and Late Bronze Age pottery from Kaymakçı.

Since currently available digitization techniques do not allow for capturing the profile of complete vessels with closed shapes—because the scanner camera cannot "see" inside the vessel—these are not part of our dataset. Since complete vessels represent only a very small fraction of all ceramic samples (25 out of the 61,961 recorded to date), we expect that excluding them does not significantly change the results of our analysis. Nevertheless, we are working on a method to estimate even hidden interior profiles and thus be able to include complete vessels in the future.

Towards a shape classification

Automated clustering

As the body of digital data produced via laser-aided profile measurement grew, it became possible to consider automated clustering to aid shape classification, as described above. In the first attempt, the full dataset of several thousand fragments was used. The automatically generated clusters CeraMatch created from this full dataset were at first unsatisfactory: clusters that should have represented potential classes and sub-classes did not withstand the visual test of experienced ceramicists. Accordingly, a more refined selection of samples drew only from the more complete rim profiles, from the rim down to the carination or the widest diameter, still including several hundred pieces. The results of this clustering attempt were better, yet still mixed, with both successful (Figure 11) and less successful matches (Figure 12), and with the majority falling somewhere between these extremes.

Automated clusters representing successful shape matches generally include shapes with relatively simple body profiles, distinctive axes, and rim forms, or a combination of these characteristics. Simple shapes such as hemispherical lipless bowls, for instance, clustered very well (Figure 11A). This was also the case with plates and simple carinated cups. Some small and medium bowls clustered very well (Figure 11B), but less well in most other cases, seemingly dependent on the complexity of the body profile and rim form (see below). Semi-closed jars also clustered very well (Figure 11C), probably a result of a distinctive profile axis combined with a relatively simple rim form (see Figure 11C). The same applies to jars with everted rims (Figure 11D).

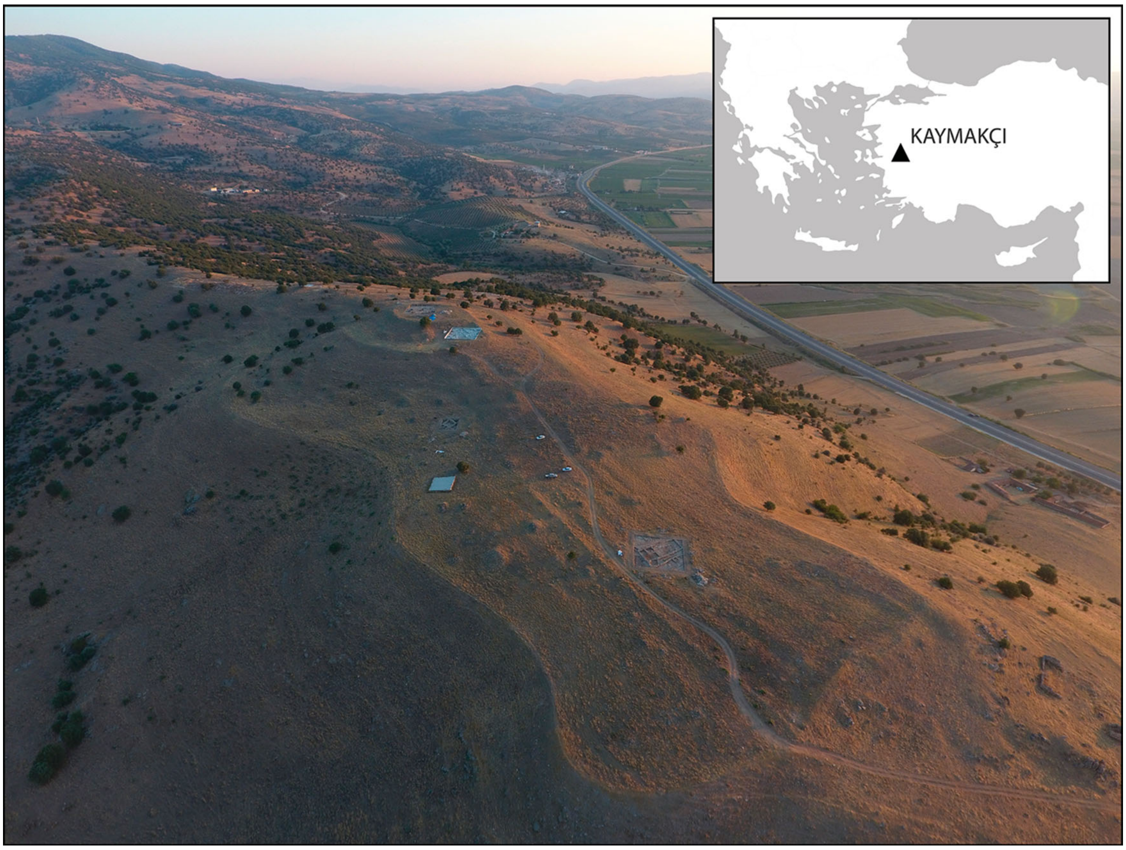

Figure 10. Oblique view of the Kaymakçı citadel towards the north. Note the three cars in the middle of the image for reference scale.

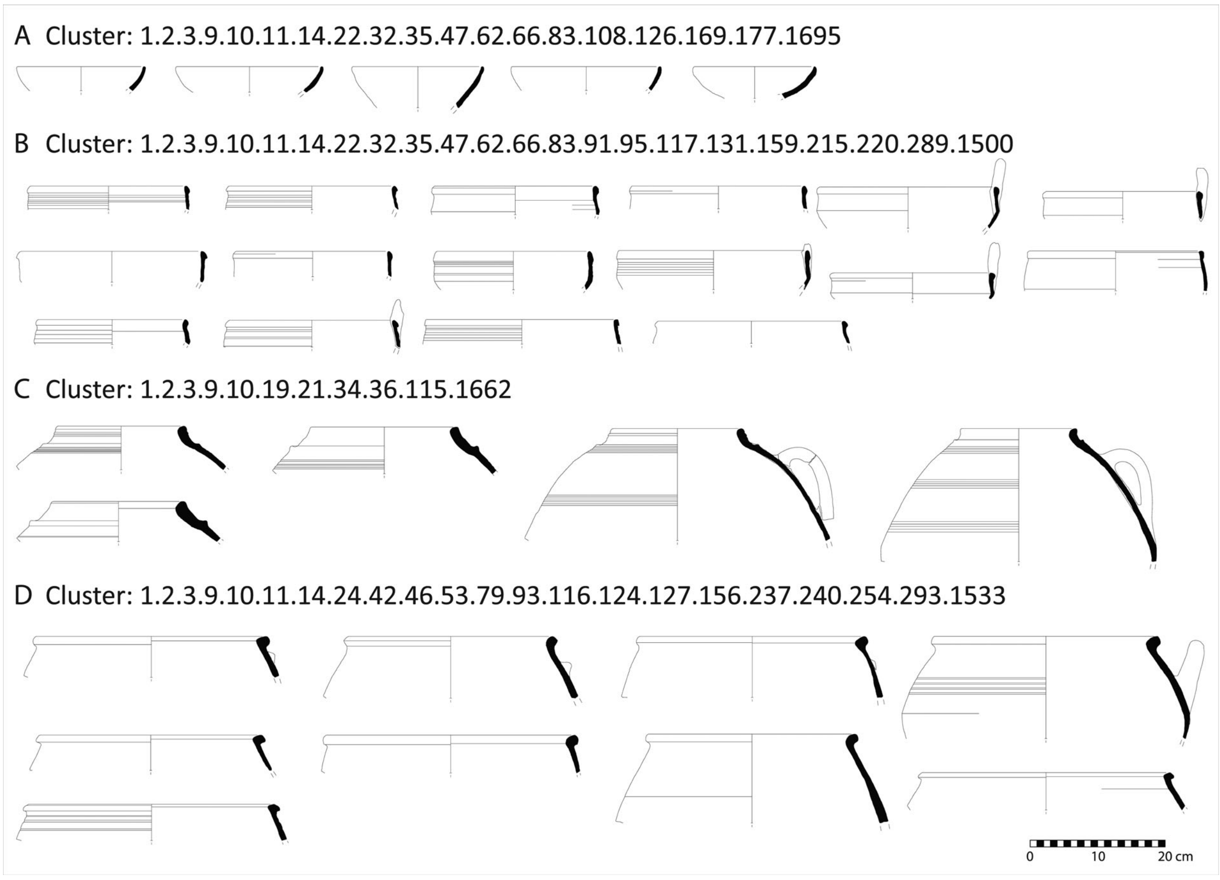

Figure 11. Examples of successful automated clustering: A) hemispherical lipless bowls; B) small to medium bowls; C) semi-closed jars; and, D) jars with everted rim. The cluster numbers reflect their hierarchical order and correspond to the labels in Supplemental Materials 1 and 2.

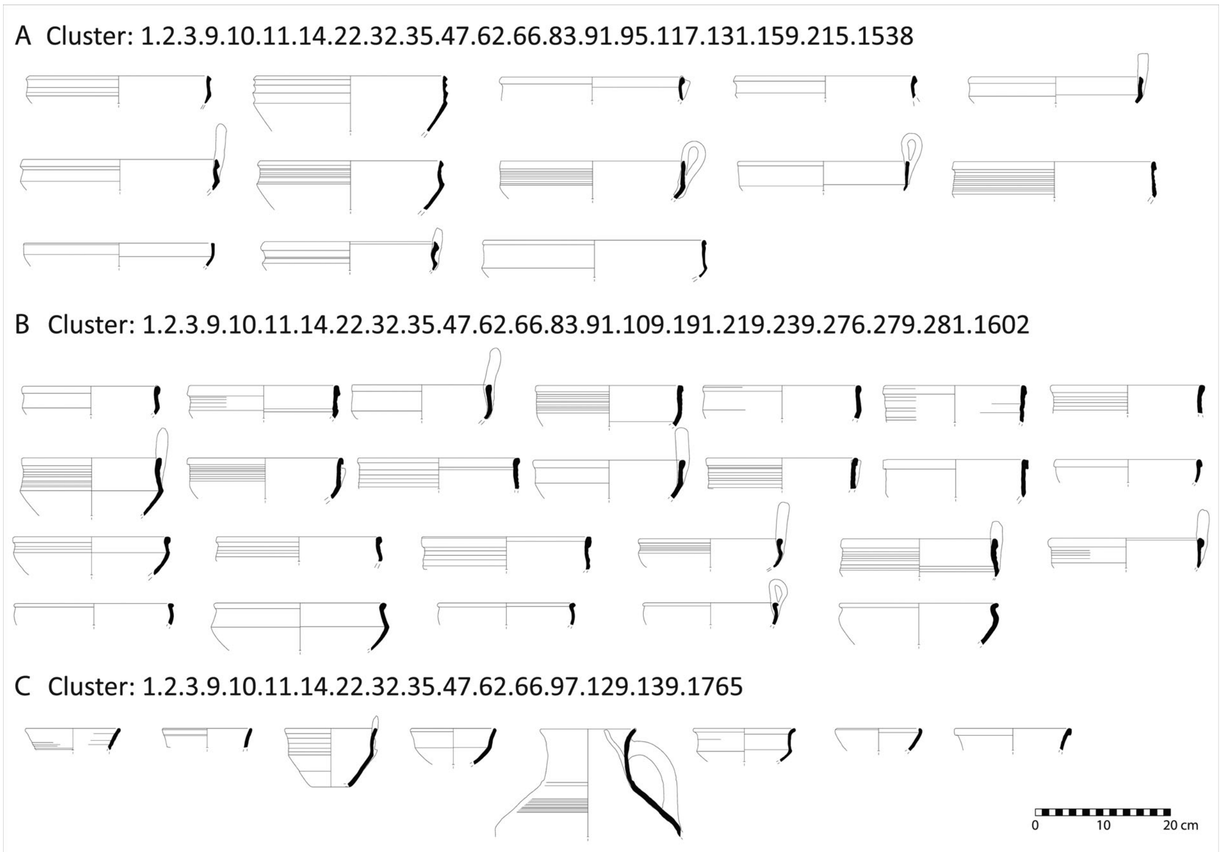

Figure 12. Examples of less successful automated clustering: A–B) carinated bowls and C) a mixed cluster. The cluster numbers reflect their hierarchical order and correspond to the labels in Supplemental Materials 1 and 2.

Automated clusters representing less successful shape matches almost invariably include carinated forms, whether plain or ridged, and the least fully preserved profiles. Three examples illustrate the problems (see Figure 12). In each of these automated clusters, CeraMatch included a variety of body profiles, axes, and rim forms that is greater than acceptable in standard shape classification. In clusters of carinated bowls (see e.g., Figure 12A–B), it is only the overall size that seems to unite the samples—including a combination of diameter and wall thickness—while the axis and especially the rim form are often quite divergent. In other clusters (see e.g., Figure 12C), jugs and both carinated bowls and simple bowls are mixed, suggesting that CeraMatch settings weighted the upper profile axis more significantly than other characteristics. While our approach includes a separate metric for rim shape (the rim Dice dissimilarity), future inclusion of other non-shape metrics, such as fabric, ware, and surface treatment or decoration, would likely improve clustering results in cases where the shape alone is not a sufficient discriminant.

Manual clustering and comparison with automated clustering

Until automated clustering can produce satisfactory shape classes completely independently, a combination of manual and automated clustering may be the most productive way forward. A manually produced classification can also serve as a useful check against the challenges of automated clustering described above. For these reasons, a preliminary classification was produced manually for Kaymakçı according to the standard practices outlined above.

Using the several hundred samples of more complete profiles selected for automated clustering, every piece was manually assigned to a shape class or sub-class. Unsurprisingly, the Kaymakçı assemblage comprised the broad categories of shapes known well from nearby Aegean coastal sites. These include plates, globular cups, carinated cups, lipless rounded bowls, carinated bowls with rolled rim (the so-called bead-rim bowls), carinated bowls with ribbed shoulder (ridged bowls), and carinated bowls with concave shoulder (A60-bowls at Troy), among many other shape classes. Unexpected, based on comparanda, was the high degree of variability within these classes, such that the dividing lines between them were often rather fluid. Kaymakçı potters appear to have combined the individual elements that define ceramic classes rather freely and were also very good at scaling. They produced almost identical shapes—with similar profiles, axes, and rim forms, for example—in a variety of sizes, ranging from larger cups through small and medium-sized bowls to large open jars. Nevertheless, broad shape categories were defined by this manual process.

As a kind of reverse control, we then compared manual clusters against CeraMatch's automated clusters (Supplemental Materials 1, 2; Demján et al. 2021). In the two examples illustrated here (Figure 13), the degree of agreement between manual clusters and automated clusters is captured visually by illustrating the samples of a manual cluster first, followed by the associated automated clusters into which the same samples fall. Samples in automated clusters that belong to the selected manual cluster are shown in black, while those found in other manual clusters are shown in gray.

In both examples, CeraMatch assigns the components of the manual clusters that were deemed similar to the

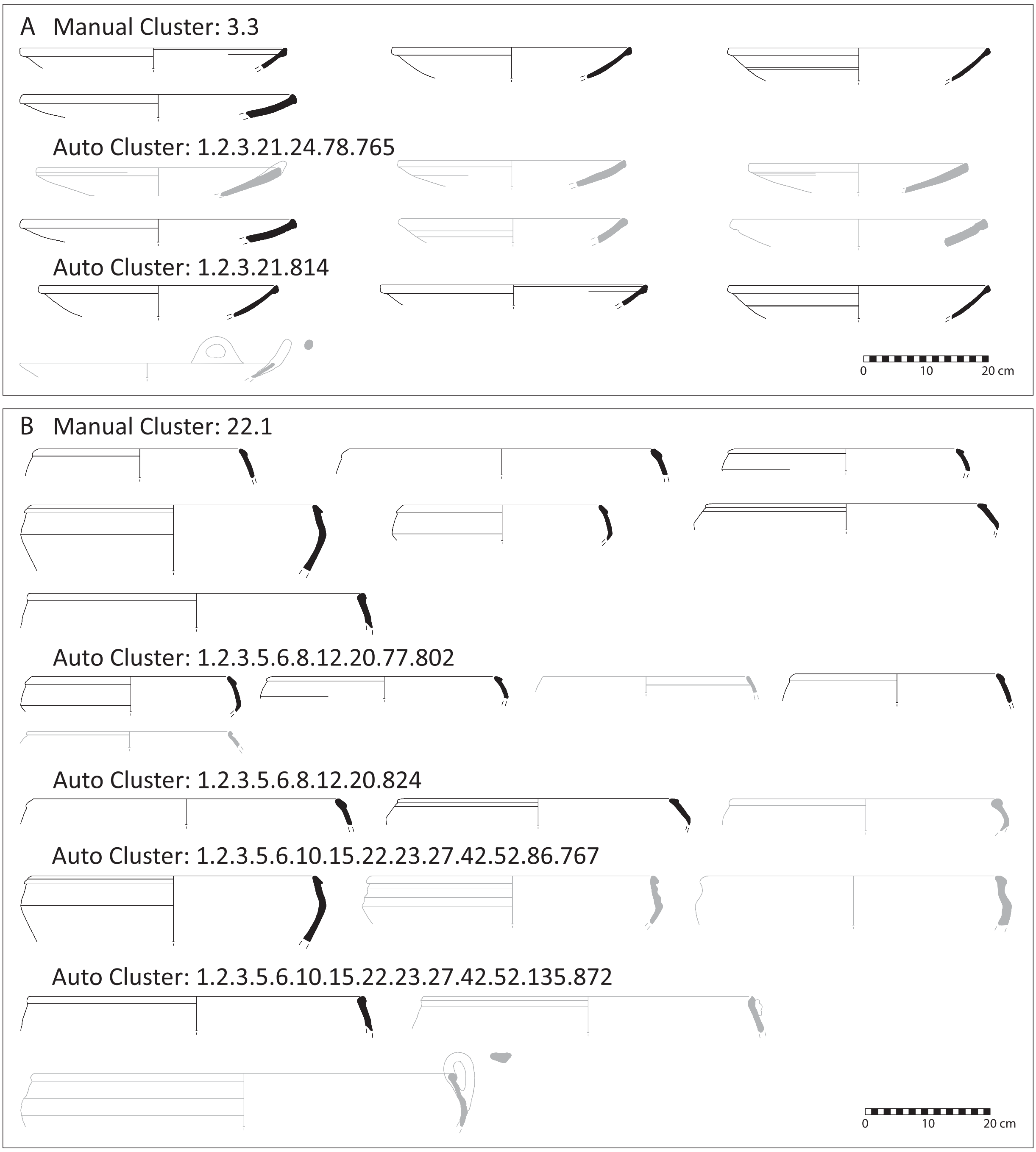

Figure 13. Comparison of two manual clusters (A and B) with associated automated clusters. Vessels in automated clusters not present in selected manual clusters are displayed in gray. The cluster numbers reflect their hierarchical order and correspond to the labels in Supplemental Materials 1 and 2.

human eye to multiple separate automated clusters (two clusters in the first example and four in the second). This shows that CeraMatch weighted certain parameters (e.g., diameter, thickness, or axis) more heavily than did the manual clustering, resulting in automated clusters that are more divided/split than grouped/lumped. In some cases, this may be a useful indication that manual clusters should be further divided, but such determinations must also be based on the functional and chronological sensitivity of the particular shape. For the full comparison of manual vs. automated clusters, see Supplemental Materials 1 and 2.

One further attempt was made to guide CeraMatch in the identification of which parameters were most significant. Morphologically distinctive and well-preserved key shapes that served as the most representative samples of manual clusters were taken as the potential seeds of automated clusters, similar in concept to supervised classification. Despite several iterations and modified settings, this experiment yielded only very general clusters. The most likely reason for this generalization is that while CeraMatch is good at finding the most similar shape for each fragment based on a singular distillation of metrics (the RMS combined value described above), it is not yet as good at selecting the most significant variables in each case and the requisite degree of similarity (between profile axes, diameters, rims, etc.) that make two fragments with relatively low overall morphometric similarity still belong to the same class.

General evaluation

Ceramic documentation via laser-aided profile measurement is the most successful method we have encountered for the Kaymakçı ceramics so far, enabling the project's goal of rapid documentation. The method achieves a satisfying speed of recording, as well as versatile data management, manipulability, and export, with no loss in the level of accuracy required for morphometric shape matching and classification. While this method of ceramic documentation demonstrated clear success, automated clustering of the dataset in CeraMatch for shape classification encountered more challenges, especially when compared against the preliminary manual classification based on the same dataset.

As described above, CeraMatch's automated clustering routine grouped certain shapes more successfully than others, with visually different degrees of within-cluster homogeneity, and almost always produced a larger number of clusters than manual clustering from the same selection of samples. CeraMatch almost always took a finer approach than manual clustering, generally tending towards splitting rather than lumping samples. The input parameters and tolerances clearly gave more weight to wall thickness than did manual clustering, and this resulted in a newly clarified understanding that wall thickness is indeed a good proxy for overall vessel size. CeraMatch was also sensitive to general profile axis, or orientation, the presence or absence of a sudden change in profile axis (e.g., carination), and diameter, as predictable from the four metrics of shape dissimilarity it evaluates.

Compared to manual clusters, automated clusters were almost too sensitive to some parameters, however, such as diameter measurements, which are given larger ranges in the manual classification. Conversely, automated clusters seem to have glossed over some finer details, such as grooved or ridged decoration and rim form. The experience of expert ceramicists suggests that some of these finer within-cluster differences, which may seem insignificant to the algorithmic or inexperienced eye, are in fact quite crucial in identifying chronologically sensitive shape classes. It is hoped that statistical evaluation will help untangle how significant these differences are following the completion of stratigraphic modeling.

Discussion

Tying together various points raised above, both general and specific, brief discussion here emphasizes the benefits of this new method of documentation and shape clustering and looks forward to future prospects and possibilities.

First, we return to the brief definition of an efficient shape classification offered above. There is likely no way for an automated computational shape classification to achieve the full sophistication of a manual approach, with all the necessary splitting and merging it entails, and especially if it excludes the stratigraphic and fabric/ware information that is almost always incorporated into final, more holistic ceramic typologies. Nonetheless, an automated or semi-automated computational shape classification is clearly of use when handling a large dataset, produces interpretable results, and thus provides a complementary approach. A well-integrated combination of computational and manual approaches is likely the best way forward. One should keep in mind the reason this research was undertaken: the need to record large masses of pottery quickly, accurately, and in a way that avoids bias and other problems deriving from sampled datasets as much as possible. While manual classification will always serve as a verification for computational classifications, it is only the latter that can begin to provide evaluation of whole-site assemblages.

Looking to the future, the obvious next step for this research is to incorporate into the analysis other components of the rich datasets recovered from most archaeological sites and, for us, from the Kaymakçı Archaeological Project. This work can be seen as the first step in an ongoing program that gradually includes more and more components of the KAP dataset, beginning with those already at hand. One next step would be to evaluate both the computational and manual shape classifications against stratigraphic records and/or archaeometric dating evidence. This could begin from those excavation areas with the fullest stratigraphic representation and extend by cross-checking with others. Another obvious next step would be to integrate information about fabric, ware, and surface treatment/decoration. Along with stratigraphic relations and dating evidence, the aim here would be to create a holistic typo-chronological model. Because the graph-based database system employed here enables recording of complex relations between all types of finds and contexts, eventual incorporation of these datasets is possible as well, from small finds to botanical and faunal evidence and other contextual data. It is through this type of modeling that we hope the additional benefits of computational classifications and clustering will become even more apparent; because they are able to capture patterns in large-scale, cross-trench analyses, they will help reveal connections and patterns that remain invisible to the human eye.

Conclusions

Improved methods of documentation and analysis are nearly constant desires in archaeology, and all the more so with respect to ceramics, which compose the bulk of finds from many archaeological sites. The technique of capturing ceramic shape information via laser-aided profile measurement described in this article increases the daily output of a draftsperson without compromising precision or information value when it comes to ceramic profiles and diameters, which are the essential features for shape classification. It is also fully capable of metrically recording additional vessel features, such as bases, handles, and plastic attachments, as well as photographically documenting color and other decorative characteristics. While axially asymmetric shapes and complex decorations will still benefit from manual drawing or 3D scanning, the bulk of most ceramic assemblages can be completely processed using this technique, at the same time eliminating the need for manual diameter estimation and drawing, as well as subsequent digitization and extraction of metric data for morphometric analysis.

Furthermore, the methods of recording and analysis introduced in this article invoke innovative shape-matching algorithms and automated clustering routines within a single database (Deposit) and software environment (CeraMatch) that enables a smooth workflow from initial recording to the classification of ceramic shapes. The advantage of using an unsupervised machine-learning approach based on traditional mathematical and statistical techniques (such as Euclidean distance metrics and hierarchical cluster analysis) is that it does not depend on large training datasets, as is the case with supervised machine-learning methods. The large, well-described, and digitized ceramic datasets necessary to train neural networks are not yet available. While the approach is unable to capture some subtle differences or their combinations in certain shapes, which can be crucial for identifying chronologically sensitive shape classes or subclasses, it does provide a good starting point for datasets that are too large to order manually. For the time being, it seems that combining automated and manual clustering techniques is the most profitable avenue of shape classification, preserving the part-science, part-art nature of the endeavor.

The ability to express ceramic class attribution in a probabilistic way should lead to improved accuracy in statistical modeling when combined with other archaeological information, for example, stratigraphy or radiocarbon dating. Probabilistic assignment of fragments to classes and, subsequently, classes to stratigraphic units also has the potential to improve the results of seriation, which traditionally relies on the quantified presence/absence of shape classes in stratigraphic features such as graves or settlement pits.

In attempting to achieve such classification purely computationally, the challenge of replicating human perception is clear. As explained above, a classification that serves as a useful tool and not just a descriptive categorization is a complex and iterative process, requiring steps forward and backward across fluid class boundaries and consideration of the reasons underlying ancient potters' choices. Because such ancient decision-making processes remain somewhat unknowable, nuanced human input is needed. Algorithmic, fully automated routines are thus far unable to compete with the polythetic expertise of the expert eye: different sets of characteristics are important for different shape classes and with changing importance from class to class. Finding a unified set of characteristics that applies to all shape classes is simply beyond the possible. Improvement in this area could come from introducing multi-step automated clustering into the workflow—clustering first more generally according to one set of criteria and parameter weights and subsequently, within resultant clusters, according to other sets of the criteria and parameter weights—yet proof of this concept, among other future developments such as inclusion of photographic metrics or stratigraphic information, awaits future testing.

Acknowledgements

This research was funded by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World," Reg. No. CZ.02.1.01/0.0/0.0/16_019/0000734; by Koç University; by the Merops Foundation; fellowships from the Research Center for Anatolian Civilizations (ANAMED); and, by private donors—to all of whom we remain grateful. For permissions and support, we thank the General Directorate of Cultural Heritage and Museums, Ministry of Culture and Tourism, Republic of Turkey, and its annual representatives, as well as the Manisa Museum of Archaeology and Ethnography. We would like to acknowledge and thank all participants in the work of the Kaymakçı Archaeological Project, and explicitly Ján Bobik, for working with the Laser-Aided Profiler, Tunç Kaner, for managing the workflow of ceramic documentation at Kaymakçı, and Christina Luke, for encouragement and support of bringing this research to publication.

Disclosure Statement

In accordance with Taylor & Francis policies and researcher ethical obligations, we report that Peter Demján has a license agreement with the DRŽÍK s.r.o. company that manufactures and sells Laser Aided Profiler devices, involving also software license fees for the sale of these devices. The results presented here are in no way considered promotion for or encouragement of that company's business operations, and, conversely, DRŽÍK s.r.o. business operations had no impact on the scientific and technical merits nor the research outcomes presented here.

Notes on Contributors

Peter Demján (Ph.D. 2016, Comenius University in Bratislava) is a researcher in the Department of Information Sources and Landscape Archaeology at the Institute of Archaeology of the Czech Academy of Sciences, Prague. His areas of research include computational modeling of prehistoric settlement landscapes, site chronologies, and ceramic shape analysis. He is a co-creator of the Laser Aided Profiler system.

Peter Pavúk (Ph.D. 2006, Eberhard-Karls-University, Tübingen) is the Director of the Institute of Classical Archaeology at the Charles University in Prague and a corresponding member of the German Archaeological Institute. His research focuses on the Aegean and Anatolian Bronze Age, with an overlap to the Balkans and central Europe, specializing in ceramic finds, chronology, and intercultural relations.

Christopher H. Roosevelt (Ph.D. 2003, Cornell University) is a professor of Archaeology and History of Art and Director of the Research Center for Anatolian Civilizations at Koç University, director of the Kaymakçı Archaeological Project, and co-director of Gygaia Projects. His research interests include archaeological and spatial technologies and the archaeology of Bronze and Iron Age Anatolia and the eastern Mediterranean.

ORCID

Peter Demján http://orcid.org/0000-0002-1589-4727
Peter Pavúk http://orcid.org/0000-0002-8739-9434
Christopher H. Roosevelt http://orcid.org/0000-0002-4302-4788

References

Albero Santacreu, D., M. Calvo Trias, and J. G. Rosselló. 2017. "Formal Analysis and Typological Classification in the Study of Ancient Pottery." In *The Oxford Handbook of Archaeological Ceramic Analysis*, edited by A. M. W. Hunt, 181–199. Oxford: Oxford University Press.

Aykurt, A. 2020. *Liman Tepe I: Orta Tunç Çağı Seramiği ve Küçük Buluntuları/Middle Bronze Age Ceramics and Small Finds. Bilgin Kültür Sanat Yayınları 145*. Ankara: Bilgin Kültür Sanat Yayınları.

Bartolini, E. 2017. "Typology and Classification." In *The Oxford Handbook of Archaeological Ceramic Analysis*, edited by A. M. W. Hunt, 651–670. Oxford: Oxford University Press.

Bitelli, G., F. Rinaudo, D. Gonzalez-Aguilera, and P. Grussenmeyer. 2020. *Data Acquisition and Processing in Cultural Heritage*. Multidisciplinary Digital Publishing Institute: Basel.

Bradski, G. 2000. "The OpenCV Library." *Dr. Dobb's Journal of Software Tools* 120: 122–125.

Dedeoğlu, F., and E. Abay. 2014. "Beycesultan Höyük Excavation Project: New Archaeological Evidence from Late Bronze Age Layers." *Arkeoloji Dergisi* 17: 1–39.

Demján, P. 2012. "Grave Typology and Chronology of a Lengyel Culture Settlement: Formalized Methods in Archaeological Data Processing." In *Theoretical and Methodological Considerations in Central European Neolithic Archaeology: Proceedings of the "Theory and Method in Archaeology of the Neolithic (7th - 3rd Millenium BC)" Conference Held in Mikulov, Czech Republic, 26th - 28th of October 2010*, edited by J. Kolář, and F. Trampota, 77–93. Oxford: Archaeopress.

Demján, P. 2016. "Svodín. Štruktúra a chronológia sídliska lengyelskej kultúry." Ph.D. diss., Comenius University, Bratislava.

Demján, P. 2021a. "Deposit (1.3.2)" [Computer software]. https://github.com/demjanp/deposit.

Demján, P. 2021b. "CeraMatch (1.0)" [Computer software]. https://github.com/demjanp/CeraMatch.

Demján, P., and V. Držík. 2018. "Laser Aided Profiler." https://laseraidedprofiler.com.

Demján, P., P. Pavúk, T. Kaner, J. Bobik, and C. H. Roosevelt. 2021. "Sample Data - Middle- to Late Bronze Age Pottery from Kaymakçı." *OSF*, https://doi.org/10.17605/OSF.IO/UX8VD.

Dunnell, R. C. 1986. "Methodological Issues in Americanist Artifact Classification." *Advances in Archaeological Method and Theory* 9: 149–207.

Durham, P., P. H. Lewis, and S. Shennan. 1995. "Artefact Matching and Retrieval Using the Generalised Hough Transform." In *Computer Applications and Quantitative Methods in Archaeology 1993*, edited by J. Wilcock, and K. Lockyear, 25–30. Oxford: British Archaeological Reports.

Erkanal-Öktü, A. 2018. *Panaztepe I: Die Friedhöfe von Panaztepe. Türk Tarih Kurumu Yayınları 51 a/1*. Ankara: Türk Tarih Kurumu.

Ersoy, Y. E. 1988. "Finds from Menemen/Panaztepe in the Manisa Museum." *The Annual of the British School at Athens* 83: 55–82.

Gero, J., and J. Mazzullo. 1984. "Analysis of Artifact Shape Using Fourier Series in Closed Form." *Journal of Field Archaeology* 11: 315–322. https://doi.org/10.1179/009346984791535467.

Gilboa, A., A. Karasik, I. Sharon, and U. Smilansky. 2004. "Towards Computerized Typology and Classification of Ceramics." *Journal of Archaeological Science* 31: 681–694. https://doi.org/10.1016/j.jas.2003.10.013.

Gilboa, A., A. Tal, I. Shimshoni, and M. Kolomenkin. 2012. "Computer-Based, Automatic Recording and Illustration of Complex Archaeological Artifacts." *Journal of Archaeological Science* 40: 1329–1339. https://doi.org/10.1016/j.jas.2012.09.018.

Glatz, C. 2009. "Empire as Network: Spheres of Material Interaction in Late Bronze Age Anatolia." *Journal of Anthropological Archaeology* 28: 127–141. https://doi.org/10.1016/j.jaa.2008.10.003.

Göttlich, F., A. Schmitt, A. Kilian, H. Gries, and K. Badreshany. 2021. "A New Method for the Large-Scale Documentation of Pottery Sherds Through Simultaneous Multiple 3D Model Capture Using Structure from Motion: Phoenician Carinated-Shoulder Amphorae from Tell el-Burak (Lebanon) as a Case Study." *Open Archaeology* 7 (Issue 1): 256–272. https://doi.org/10.1515/opar-2020-0133.

Grosman, L., A. Karasik, O. Harush, and U. Smilansky. 2014. "Archaeology in Three Dimensions: Computer-Based Methods in Archaeological Research." *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 2: 48–64. https://doi.org/10.5325/jeasmedarcherstu.2.1.0048.

Gualandi, M. L., G. Gattiglia, and F. Anichini. 2021. "An Open System for Collection and Automatic Recognition of Pottery Through Neural Network Algorithms." *Heritage* 4: 140–159. https://doi.org/10.3390/heritage4010008.

Hallager, B. 2010. "Minoan Pottery." In *The Oxford Handbook of the Bronze Age Aegean (ca. 3000-1000 BC)*, edited by Eric H. Cline, 405–414. Oxford: Oxford University Press.

Karasik, A., Z. Greenhut, J. Uziel, N. Szanton, L. Grosman, I. Zandbank, and U. Smilansky. 2014. "Documentation and Analyses on the National Scale at the Israel Antiquities Authority: The Story of One (Broken) Sherd." *Near Eastern Archaeology* 77: 209–213. https://doi.org/10.5615/neareastarch.77.3.0209.

Karasik, A., and U. Smilansky. 2008. "3D Scanning Technology as a Standard Archaeological Tool for Pottery Analysis: Practice and Theory." *Journal of Archaeological Science* 35: 1148–1168. https://doi.org/10.1016/j.jas.2007.08.008.

Karasik, A., and U. Smilansky. 2011. "Computerized Morphological Classification of Ceramics." *Journal of Archaeological Science* 38: 2644–2657. https://doi.org/10.1016/j.jas.2011.05.023.

Karasik, A., U. Smilansky, and I. Beit-Arieh. 2005. "New Typological Analyses of Early Bronze Age Holemouth Jars from Tel Arad and Southern Sinai." *Tel Aviv* 32 (1): 20–31. https://doi.org/10.1179/tav.2005.2005.1.20.

Lloyd, S., and J. Mellaart. 1965. *Beycesultan II. Middle Bronze Age Architecture and Pottery. Occasional Publications of the British Institute of Archaeology at Ankara 8*. London: British Institute of Archaeology at Ankara.

Luhmann, T., S. Robson, S. Kyle, and J. Boehm. 2019. *Close-Range Photogrammetry and 3D Imaging*. Berlin: De Gruyter.

Luke, C., and C. H. Roosevelt. 2017. "Cup-marks and Citadels: Evidence for Libation in the Second-Millennium BCE Marmara Lake Basin, Western Anatolia." *Bulletin of the American Schools of Oriental Research* 378: 1–23. https://doi.org/10.5615/bullamerschoorie.378.0001.

Luke, C., C. H. Roosevelt, P. J. Cobb, and Ç Çilingiroğlu. 2015. "Composing Communities: Chalcolithic Through Iron Age Survey Ceramics in the Marmara Lake Basin, Western Turkey." *Journal of Field Archaeology* 40 (4): 428–449. https://doi.org/10.1179/2042458215Y.0000000009.

Mangaloğlu-Votruba, S. 2015. "Liman Tepe During the Late Bronze Age." In *Nostoi: Indigenous Culture, Migration and Integration in the Aegean Islands and Western Anatolia During the Late Bronze and Early Iron Ages*, edited by N. C. Stampolidis, Ç Maner, and K. Kopanias, 647–668. Istanbul: Koç University Press.

Mellaart, J., and A. Murray. 1995. *Beycesultan III. 2. Late Bronze Age and Phrygian Pottery and Middle and Late Bronze Age Small Objects. Occasional Publications of the British Institute of Archaeology at Ankara 12*. London: British Institute of Archaeology at Ankara.

Mielke, D. P. 2017. "From »Anatolian« to »Hittite«. The Development of Pottery in Central Anatolia in the 2nd Millennium BC." In *Innovation Versus Beharrung: Was Macht den Unterschied des Hethitischen Reichs im Anatolien des 2. Jahrtausends v. Chr.? Internationaler Workshop zu Ehren von Jürgen Seeher, Istanbul, 23-24. Mai 2014*, edited by A. Schachner, 121–144. *Byzas* 23. Istanbul: Ege Yayınları.

Morgan, C., and H. Wright. 2018. "Pencils and Pixels: Drawing and Digital Media in Archaeological Field Recording." *Journal of Field Archaeology* 43 (2): 136–151. https://doi.org/10.1080/00934690.2018.1428488.

Mountjoy, P. A. 1998. "The East Aegean-West Anatolian Interface in the Late Bronze Age: Mycenaeans and the Kingdom of Ahhiyawa." *Anatolian Studies* 48: 33–67. https://doi.org/10.2307/3643047.

Němejcová-Pavúková, V. 1995. *Svodín: 1. Zwei Kreisgrabenanlagen der Lengyel-Kultur. Studia Archaeologica et Mediaevalia 2*. Bratislava: Comenius University.

Orton, C., and M. Hughes. 2013. *Pottery in Archaeology*. Cambridge: Cambridge University Press.

Pavúk, P. 2014. *Troia VI Früh und Mitte. Keramik, Stratigraphie, Chronologie. Studia Troica Monographien 3*. Bonn: Verlag Dr. Rudolf Habelt GmbH.

Pavúk, P. 2015. "Between the Aegeans and the Hittites: Western Anatolia in the 2nd Millennium BC." In *Nostoi: Indigenous Culture, Migration and Integration in the Aegean Islands and Western Anatolia During the Late Bronze and Early Iron Ages*, edited by N. C. Stampolidis, Ç Maner, and K. Kopanias, 81–113. Istanbul: Koç University Press.

Pavúk, P., and B. Horejs. 2018. "Ceramics, Surveys, and Connectivity in Western Anatolia: The Middle and Late Bronze Age Bakırçay/Kaikos Valley Restudied." *Egypt and Levant* 28: 457–485. https://doi.org/10.1553/AEundL28s457.

Pawlowicz, L. M. and Downum, C. E., 2021. "Applications of Deep LEarning to DEcorated CEramic TYpology and CLassification: A Case STudy Using Tusayan White Ware from Northeast Arizona." *Journal of Archaeological Science 130*. https://doi.org/10.1016/j.jas.2021.105375

Plog, S. 1983. "Analysis of Style in Artifacts." *Annual Review of Anthropology* 12: 125–142.

Poblome, J., J. van den Brandt, B. Michiels, G. Evesever, R. Degeest, and M. Walkens. 1997. "Manual Drawing Versus Automated Recording of Ceramics." In *Sagalassos IV. Acta Archaeologica Lovaniensia Monographiae 9*, edited by M. Walkens, 533–538. Leuven: Leuven University Press.

Roosevelt, C. H., P. Cobb, E. Moss, B. R. Olson, and S. Ünlüsoy. 2015. "Excavation is Destruction Digitization: Advances in Archaeological Practice." *Journal of Field Archaeology* 40 (3): 325–346. https://doi.org/10.1179/2042458215Y.0000000004.

Roosevelt, C. H., and C. Luke. 2017. "The Story of a Forgotten Kingdom? Survey Archaeology and the Historical Geography of Central Western Anatolia in the Second Millennium BC." *European Journal of Archaeology* 20 (1): 120–147. https://doi.org/10.1017/eaa.2016.2.

Roosevelt, C. H., C. Luke, S. Ünlüsoy, C. Çakırlar, J. M. Marston, C. R. O'Grady, P. Pavúk, M. Pieniążek, J. Mokrišová, C. B. Scott, N. Shin, and F. G. Slim. 2018. "Exploring Space, Economy, and Interregional Interaction at a Second-Millennium B.C.E. Citadel in Central Western Anatolia: 2014–2017. Research at Kaymakçı." *American Journal of Archaeology* 122 (4): 645–688. https://doi.org/10.3764/aja.122.4.0645.

Rutter, J. B. 2010. "Mycenaean Pottery." In *The Oxford Handbook of the Bronze Age Aegean (ca. 3000–1000 BC)*, edited by Eric H. Cline, 415–429. Oxford: Oxford University Press.

Saragusti, I., A. Karasik, I. Sharon, and U. Smilansky. 2005. "Quantitative Analysis of Shape Attributes Based on Contours and Section Profiles in Artifact Analysis." *Journal of Archaeological Science* 32: 841–853. https://doi.org/10.1016/j.jas.2005.01.002.

Schoop, U.-D. 2011. "Hittite Pottery: A Summary." In *Insights Into Hittite History and Archaeology*, edited by H. Genz and D. P. Mielke, 241–274. *Colloquia Antiqua* 2. Leuven - Paris - Walpole, MA: Peeters.

Seguchi, N., and B. Dudzik. 2019. *3D Data Acquisition for Bioarchaeology, Forensic Anthropology, and Archaeology*. San Diego: Elsevier Science.

Selden, R. Z., Jr, T. K. Perttula, and M. J. O'Brien. 2014. "Advances in Documentation, Digital Curation, Virtual Exhibition, and a Test of 3D Geometric Morphometrics." *Advances in Archaeological Practice* 2 (2): 64–79.

Smith, N. G., A. Karasik, T. Narayanan, E. S. Olson, U. Smilansky, and T. E. Levy. 2014. "The Pottery Informatics Query Database: A New Method for Mathematic and Quantitative Analyses of Large Regional Ceramic Datasets." *Journal of Archaeological Method and Theory* 21: 212–250. https://doi.org/10.1007/s10816-012-9148-1

Spaulding, A. C. 1953. "Statistical Techniques for the Discovery of Artifact Types." *American Antiquity* 18: 305–313.

Turpin, S. A., and J. A. Neely. 1977. "An Automated Computer Technique for Vessel Form Analysis." *Plains Anthropologist* 22: 313–320. https://doi.org/10.1080/2052546.1977.11908868.