DE GRUYTER Open Archaeology 2022; 8: 1183–1217

Research Article

Martin Hinz*, Caroline Heitz

Unsupervised Classification of Neolithic Pottery From the Northern Alpine Space Using t-SNE and HDBSCAN

https://doi.org/10.1515/opar-2022-0274
received June 9, 2022; accepted November 10, 2022

Abstract: Terms of "Neolithic cultures" are still used to describe spatial and temporal differences in pottery styles across central Europe. These terms date back to research periods when absolute dating methods were lacking and typological classification was used to establish chronologies. Those terms are charged with problematic, biasing notions of social configurations: cultural homogeneity, spatial boundedness, and immobility. In this article, we present an alternative approach to pottery classification by using ceramics from dendrochronologically and C14-dated sites of the 40th–38th c. BC located in the northern Alpine Foreland. The newly developed methodology uses a computational unsupervised classification based on profile shape and additional nominal characteristics using t-Distributed Stochastic Neighbour Embedding and Hierarchical Density-Based Spatial Clustering of Applications with Noise for cluster analyses. Its role in our project was to provide a quantitative, algorithm-based approach to classify large datasets of pottery while simultaneously account for a large number of variables. This enabled us to find similarity structures that would escape human cognitive capacities on which typological classification is based on. It formed one pilar of a mixed method research approach combining qualitative and quantitative methods of pottery classification. Our results show that the premises of cultural homogeneity are untenable but can be methodologically overcome by using the proposed classification approaches.

Keywords: unsupervised classification, morphometrics, ceramic analysis, HDBSCAN, t-SNE

1 Introduction

This article presents a quantitative approach to pottery classification based on unsupervised computer-assisted methods. We applied this methodology using pottery data from wetland sites located in the Northern Alpine Foreland, dated to the period between 3975 and 3750 BC and chronologically corresponding dryland sites from adjacent regions. This analysis aimed to compare different pottery assemblages within this period to reveal similarities and differences between ceramics across the various settlements and the regional pottery styles. The novelty of our quantitative pottery classification method is twofold: first, it transcends classical typological schemata but complements new alternative qualitative pottery

---

Special Issue on Digital Methods and Typology, edited by Gianpiero Di Maida, Christian Horn & Stefanie Schaefer-Di Maida.

*** Corresponding author: Martin Hinz,** Department for Prehistoric Archaeology, and Oeschger Centre for Climate Change Research (OCCR), Institute of Archaeological Sciences, University of Bern, Bern, Switzerland,
e-mail: martin.hinz@unibe.ch
Caroline Heitz: Institute of Pre- and Protohistoric Archaeology, CRC 1266 'Scales of Transformations – Human-Environmental Interaction in Prehistoric and Archaic Societies', Kiel University, Kiel, Germany

classification approaches; second, being part of a new bottom-up mixed method pottery classification approach that we propose (Hafner, Heitz, & Stapfer, 2016; Heitz, 2018), it uncovers material cultural entanglements that are related to phenomena of spatial mobility and would otherwise have been obscured by the classical pottery typologies that draw on conventional – sometimes even ethnic interpreted – problematic notions of 'Neolithic cultures'.

Furthermore, using a computer-based algorithmic approach for this type of analysis allows for intersubjective reproducibility of the pottery classification and its results. Since this approach can process large amounts of data simultaneously, both synchronous and diachronic comparative analyses can be conducted. Because the latter exceeds the human brain's capacity, new structures of similarities and relationships can be discovered that are not visible to the archaeologist's eye. The empirical results of the analyses were subsequently integrated into an overall analysis and have proven to be an essential tool for questioning the concepts of 'Neolithic cultures' anchored in the cultural-historical research paradigm.

The ceramic finds preserved in the numerous dendrochronologically dated Neolithic wetland sites of the Northern Alpine Foreland are a unique yet under-researched basis for addressing various socio-archaeological questions and fundamental issues of archaeological classification. In the scope of the SNSF project No 100011_156205 'Mobilities, Entanglements and Transformations in Neolithic Societies of the Swiss Plateau (3900–3500 BC)[1] (Hafner et al., 2016), short MET project, we elaborated a new methodology to gain a deeper understanding of spatial mobility and social transformations of Neolithic settlement communities based on the material entanglements of pottery (Heitz, 2017, 2018; Stapfer, 2019; Stapfer, Heitz, Hinz, & Hafner, 2019). The methodology is grounded by process philosophical ontological considerations of the New Materialism and New Realism, which conceptualize things – such as ceramic vessels – relationally. Pierre Bourdieu's praxeology serves as epistemology and his habitus theorem as well as his theory of practice as a social theoretical approach, which were combined with further concepts of, e.g., mobility and appropriation (Heitz, 2018, pp. 89–139). On this theoretical foundation, a mixed methods research (MMR) design was elaborated, in which different qualitative and quantitative archaeological and archaeometrical methods were combined to classify pottery regarding their material, technological, and stylistic features (Hafner et al., 2016). This includes newly elaborated qualitative and quantitative classification methods of stylistic-morphological aspects of pottery. The detailed description and argumentation of this approach as well as a full analysis, in which the qualitative and quantitative methods of classification were used, can be taken from the doctoral thesis of Heitz (2018) that is in the process of being published (Heitz, 2022). In this article, the quantitative classification approach, which was developed in collaboration of M. Hinz and C. Heitz, will be presented in full detail: a computational unsupervised classification based on profile shape and additional nominal characteristics using t-Distributed Stochastic Neighbour Embedding (t-SNE) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) for cluster analyses. The approach is currently used in the ongoing SNSF-Postdoc.Mobility-project No. 194326 'Time and Temporality in Archaeology. Approaching Rhythms and Reasons for Societal (Trans) formations in Prehistoric Central Europe (TimeArch).[2]

Two general demands drove the development of the computer-based unsupervised classification approach: first, the development of a variable-based classification method for pottery that should be able to utilize the abundance of metric and numeric features that can be compiled from pottery from numerous sites across different microregions. Second, this classification method, which is less influenced by the subjectivity of the archaeologist, should complement the qualitative visual classification methods based on the typological information of the pottery. The specific research questions that guided these objectives were as follows:

- How do the designs from different pottery styles group into clusters?
- Are there stylistically mixed clusters?
- Are there clusters encompassing vessels of different settlements?

---

[1] https://p3.snf.ch/Project-156205 (accessed 22 April 2022).

[2] https://p3.snf.ch/Project-194326 (accessed 22 April 2022).

- Is the computer-assisted classification beneficial in gaining information about entanglements between the production practices of the potters?
- What could the results reveal about cultural homogeneity and past communities' mobilities?

However, these questions and objectives unfold their full potential not in isolation, but only in the context of the MET project's overarching research topics and questions.

1.1 Overall Research Topic and Objectives of the MET Project

Throughout the research history, the understanding of the Neolithic in the Northern Alpine Foreland – as in the entire Central European Neolithic research – was characterized by concepts of "archaeological cultures" defined based on pottery. The epistemological consequences associated with this will be outlined very briefly below – since a more in-depth discussion is not possible here for reasons of space and, beyond that, has already taken place elsewhere (Furholt, 2008; Hafner & Suter, 2003; Heitz, 2017, pp. 258–262; 2018, pp. 15–19). As common in the cultural-historical research paradigm, pottery was first and foremost used to establish typo-chronologies before scientific dating techniques like dendrochronological and radiocarbon dating were available. Based on certain vessel types or the quantitatively predominant main characteristics of pottery, newly excavated sites could be assigned to the respective cultures (Hafner & Suter, 2003, pp. 2–4; Hafner et al., 2016). Hence, the Neolithic was structures in spatial-temporal "cultural blocks", most obvious in typo-chronological schemes and maps of cultures (Figure 1).

These typo-chronological units were subsequently equated with ethnic groups at that time, as was the case everywhere in Europe. Over time this led to a problematic research practice in which Neolithic cultures were defined exclusively by using pottery and conceptualized as homogeneous, spatially clearly definable units. In consequence, the farming communities assigned to these to these Neolithic cultures were imagined as sedentary, spatially barely mobile groups (Heitz, 2018, pp. 20–52; there further literature). The premises were that one settlement equates with one pottery style and one ethnic group and that Neolithic societies were living rather autarkic, subsistent-based, and predominantly sedentary lives where "cultural contacts" happen rather sporadically, mainly in onetime events of migration (Heitz, 2017, p. 259; Heitz & Stapfer, 2021).

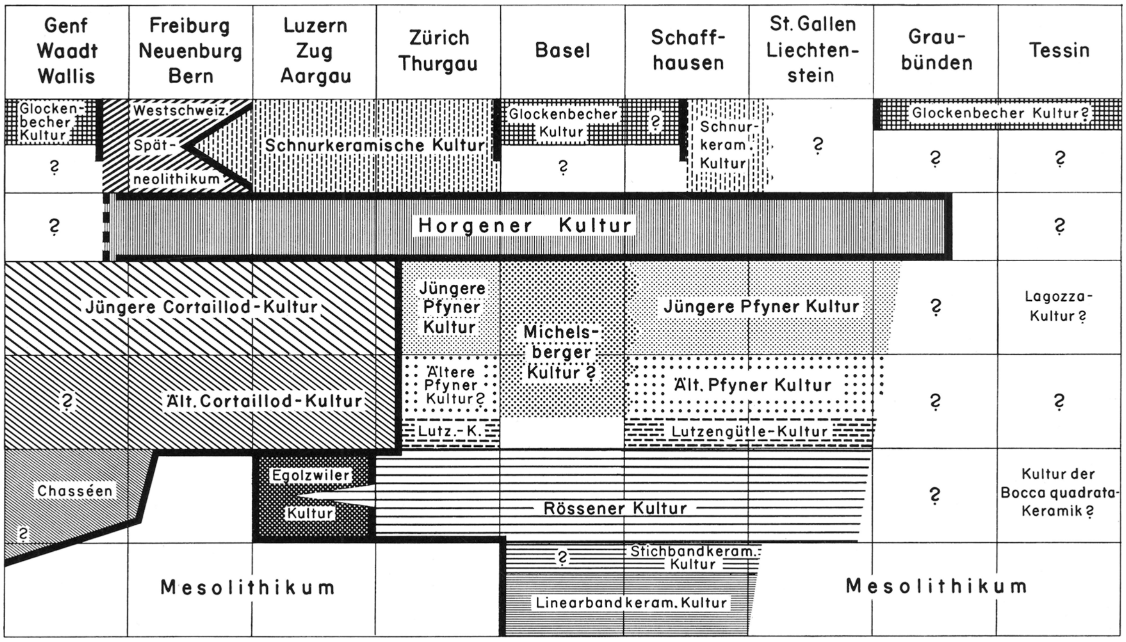


Figure 1: Vogt's (1964) Chronology Scheme of Neolithic Cultures in Switzerland.

One way to test and question this premise about ‘Neolithic cultures’ is to examine spatial mobility and accompanying cultural relations in depth by drawing on ceramic finds and analysing spatial and temporal similarities and differences. One way to test and challenge this premise about ‘Neolithic cultures’ is to examine spatial mobility and the associated cultural relationships in depth using ceramic finds by highlighting similarities and differences in pottery practices across space and time. For understanding spatial mobility based on pottery, it is decisive that regional stylistic differences can be separated from chronological ones. However, this circumstance is not given if pottery itself is used to establish typo-chronologies in different regions as only means of dating. In such cases, no degrees of freedom are left for further social-archaeological analyses. Morphological differences and similarities can no longer be used as chronology-independent source of cultural and social information. In this respect, working with pottery assemblages deriving from prehistoric wetland sites in the circum-alpine region offers a research potential of enormous value (Hafner, 2014). Due to the anaerobic preservation conditions of the waterlogged remains of settlements, thousands of organic artefacts are preserved. Among them are the wooden remains of houses that – in the optimal cases – allow us to date the settlement histories to the exact year by means of dendrochronology. Hence, combined with stratigraphic information and radiocarbon dates, an absolute chronological grid can be established for Neolithic wetland sites in the Northern Alpine Foreland.

The wide application of dendrochronological dating in Switzerland and adjacent countries makes chronologies independent from pottery typologies today. With the discovery of culturally “foreign pots” in dendrochronologically dated settlements, e.g., Arbon-Bleiche 3 (3384–3370 BCE), at Lake Constance, the premises-loaded notion of “Neolithic cultures” was challenged (De Capitani, Deschler-Erb, Leuzinger, Marti-Grädel, & Schibler, 2002, pp. 209–26, 220–221, 223; Freudiger-Bonzon, 2005; in general see also Burri-Wyser, 2007; Matuschik, 2010, 2011; Matuschik & Müller, 2016; Schlichtherle, 1998). During our project, an initial review of already published catalogues of pottery deriving from wetland sites in the Northern Alpine Foreland dated to the first half of the 4th millennium BCE quickly revealed that stylistic plurality in the pottery deriving from these sites was rather the norm than the exception (Heitz & Stapfer, 2016, pp. 150–151; 2021).

Pottery assemblages that derive from these settlement phases offer a rare possibility to separate temporal from spatial cultural differences. Furthermore, thanks to the favourable preservation conditions, an abundance of pottery per site can be analysed whose low fragmentation allows to reconstruct the full profile shapes. Accordingly, the methodology elaborated here based on such high-quality data promises to be ground-breaking and applicable to other projects.

To gain a deeper understanding of this phenomenon related to spatial mobility and social configurations, we developed two classification methods to approach stylistic and morphological similarities and dissimilarities between the pottery of different sites:

1. a qualitative impressionistic classification of the vessels’ designs that allows to understand social practices of pottery production from the actors’ perspective and
2. a quantitative computational unsupervised classification using t-SNE and HDBSCAN for cluster analyses based on the pottery vessels’ profile shape and further variables that allows to explore those social practises connected to the pottery consumption practice across several settlements.

Using both combined, the problematic models of Neolithic societies following the cultural-historical paradigm can be deconstructed by inquiring directions and rhythms of spatial mobility based on material entanglements in pottery that are indicative for social configurations.

1.2 Computer-Assisted and Visual Classification

In archaeology, classification is one of the recurring basic problems. The purpose of classification in general is to enable statements about a specific set of objects and thus to go beyond the description of individual items (Nakoinz & Hinz, 2015, p. 226). Classification is the process of gathering similar entities into groups called classes. The entities of a group should be more similar ("high within-group homogeneity") than two entities from different groups ("low between-group homogeneity") (Rice, 2015, pp. 220–221). Classification is based on a selection of features (variables), which are usually examined for their different characteristics in all entities to be classified. In principle, the entities to be classified, such as ceramic vessels, have an endless number of possible features that can be investigated (Rice, 2015, pp. 221, 224). Ceramic vessels, or rather ceramic fragments, can be classified in a variety of different ways and with a variety of different objectives. The assessment of which of these are relevant and thus to be selected is one of the most subjective moments in the entire investigation process, which cannot be avoided also by computer-assisted methods (for an overview of different classification methods, see Heitz, 2018, pp. 185–194).

The process of classification can be done impressionistically intuitively by eye/brain – what we refer to as qualitative classification – or by means of a quantitative computer-aided procedures, referred to as quantitative classification (Heitz, 2018, pp. 199–203). A classification based on a visual approach in the framework of subjective decisions does have various advantages (Heitz, 2018, pp. 190, 200–201): first, it certainly comes closest to the methodology in which the contemporaries of the production of the vessels would have classified them. Practically, it is usually quicker to carry out; moreover, humans can concentrate each time on different dimensions of similarity but combine these again in an overall view and distinguish situationally important from unimportant criteria. Furthermore, visual classification built on the extraordinary capacity of the human brain to find pattern but is also dependent on the researcher's experience and knowledge. It is therefore subjective and the ability to verify the results of such classifications relies on intersubjectivity. Implementing this for computer-aided procedures is not trivial (Read, 2016, p. 27).

Quantitative computer-assisted classification is based on a statistical model (variables and their characteristics, distance measure, algorithm, etc.) created by humans (Rice, 2015, pp. 231–232). Similarities between the entities are calculated with the help of a mathematical distance measure and based on this, classes (clusters) are formed with the help of an algorithm. Two basic directions can be used: the so-called "supervised" and "unsupervised" approaches. In the first case ("machine learning"), the computer is taught in a learning phase to classify according to a predefined classification system. In the second case, the aim is to find unknown groups in a dataset based on certain characteristics ("automatic classification").

When such quantitative, inductive bottom-up approaches of computer-assisted classification emerged in archaeology at the end of the 1950s, the classification results were initially considered essentially objective and the classes discovered and types derived from them were considered real, i.e., relevant for past societies (Hutson & Markens, 2002, pp. 9–14). However, it soon became apparent that subjective moments also exist in such classifications, where decisions must be made by humans: (1) in the selection of the variables to be considered at all, (2) in the weighting of the same, (3) in the choice of the algorithm, (4) in the assessment of the calculated classes and possible outliers, etc. (Hutson & Markens, 2002, pp. 9–14). The same selection of ceramic vessels can thus be classified very differently using different statistical models.

Nevertheless, computer-based classifications also have decisive advantages: With the appropriate programming skills and computing power, an large number of variables can be processed for an equally large number of entities or datasets, which makes these methods particularly interesting for comparative studies of larger space/time windows. Furthermore, the methodological procedure can be documented step by step, so that a cluster solution becomes comprehensible and reproducible for third parties if the code, the software used, and the corresponding datasets are disclosed. Although these methods are not objective, they achieve a higher degree of intersubjectivity than intuitive classifications (see Heitz, 2018, pp. 178–179).

Because the pottery of the first half of the 4th millennium BCE is barely decorated, the vessels' shapes were an important feature for morphological and stylistic differentiation in the case study presented here. Therefore, the morphometrics of the half-profiles of pottery vessels – as represented by pottery drawings in general – were of central importance (Heitz, 2018, pp. 202–208). A classification of vessels based on profile images or profile curves has already been suggested several times in archaeology (Jammet-Reynal, 2006; Lăzărescu & Mom, 2016; Mom, 2007; Saev, 2015; van der Maaten, Lange, & Boon, 2010; Karasik &

Smilansky, 2011). The computer-assisted classification method proposed in this article provides solutions to two sets of problems that occur in archaeological pottery classification and haven't been solved yet:

1. It combines the profile images with other variables and makes them evaluable for cluster analysis. So far, no other procedure in archaeology makes this possible. Thus, a completely new explorative clustering procedure is tested here, which enables automatic and thus comparatively unsupervised classification of the vessels, including their metrics and nominal variables.
2. It considers that morphological standardization is limited in handmade Neolithic pottery and that the metrics of individual vessels of the same design can vary considerably. Style – understood here as material style associated with particular pottery production practices (see Section 2) – is not a rigid category but a fluent phenomenon. Since styles emerge in a correspondence of the material, the individual and the social, they include standardization but also variations. In consequence – and when taking mobility into account too – it is to be expected that the standardization of pottery found in Neolithic settlements is limited and tends to include marked deviation and outliers. This calls on the one hand for a targeted selection of variables that consider the intentional and unintentional/random emergence of pottery features. On the other hand, statistical techniques are needed that deal with the high-dimensional structure and the noisiness of such data.

Both issues were addressed in the classification method propose here. The concrete procedure of our method and the results achieved are directly reproducible through the disclosure of the data and R scripts. We also hope that the statistical model itself will be useful for other research projects. Therefore, in the following, after explaining our theoretical approach, we will briefly present the ceramic objects analysed in the classification and then explain the methods used for the computer-assisted classification. We will then outline and discuss the results of our analysis.

2 Theoretical Approach

The MET project sought to avoid the problems associated with the concept of archaeological cultures. Concurrently, a better access to pottery practices, as well as mobility, cultural appropriation, and transformations by means of observable material entanglements in ceramics, was central (Hafner et al., 2016). Therefore, we drew on Pierre Bourdieu's Reflexive Anthropology (Bourdieu, 2013) as a basic research attitude and transferred it into what we call Reflexive Archaeology (Heitz, 2018, pp. 109–116). The Reflexive Anthropology, from ontology to epistemology, social theory, and methodology, forms a coherent and consistent basis, which is particularly suitable to be appropriated in archaeological research.

By drawing on the mathematician, physicist, and philosopher Blaise Pascal (1623–1662), but also on the phenomenologists Martin Heidegger (1889–1976) and Maurice Merleau-Ponty (1908–1961), Pierre Bourdieu proposes a non-cartesian ontology of human beings. He understands body and mind relational, thus not as two separate entities, but as mutually constitutive (Bourdieu & Wacquant, 2013, p. 161; Bourdieu & Chartier, 1989, p. 54). Accordingly, human beings are part of the world and can relate to the world in different ways, as objects and as subjects. In consequence, there are two ways of knowing the world, subjectivism and objectivism (Heitz, 2018, p. 110). Based on his ontology, these two epistemological positions are not contradictory but complementary, as he conceptualizes them as two different modes of knowledge production that should be combined within research. The duality of this epistemology is what Pierre Bourdieu calls praxeology (Bourdieu, 2014, pp. 246–247, 255–256). The concept of objectivism needs further explanation. Objectivism is understood by Pierre Bourdieu as scientific form of knowledge production that accesses and describes the world in a standardized quantitative way with the aim of exploring and explaining structures and processes through systematic analyses. A comparable epistemological approach underlies most of the natural sciences, such as physics, which measures its observations to develop models in mathematical language based on specific experiments. In social sciences, it is the use of statistics that Pierre Bourdieu identifies as objectified mode of knowledge production (Wacquant, 2013, pp. 25–26). It is important to emphasize that Pierre Bourdieu considers both, objectivism, and subjectivism, as one-sided and biased.

Reflexive anthropology and praxeology provide an ontological and epistemological foundation that overcomes the problematic dualism between objectivism and subjectivism by understanding them as different yet limited and incomplete modes of knowledge gain (Bourdieu, 2014; Bourdieu & Wacquant, 2013). Furthermore, praxeology offers a methodological framework in which qualitative and quantitative methods can be operationalized within an ontological, epistemological, and social theoretical framework. The two methodological directions are integrated into the methodological approach as mutually complementary means of gaining knowledge and are combined with each other (cf. below).

Praxeology is used as a central epistemological tool to overcome the problems of research-dependent biases of subjectivist and objectivist approaches. Pierre Bourdieu proposed the conduction of three basic analytical steps – or rather adopting three different modes of research with changing epistemological perspectives (see Bourdieu, 2013, pp. 287, 294):

1. *Reflexivity*: The critical reflection of the history of research and the paradigms in that field as well as of our scholastic biases, ontological commitment, and its epistemological approaches and thus our relational entanglement within the research process and our constitutive role as research in knowledge production.
2. *Subjectivism*: The adoption of a subjective mode of research that allows us to understand social practices the past from the actor's perspective – for example, the production of ceramic vessels and socially shared pottery designs and styles using qualitative visual classification methods.
3. *Objectivism*: The adoption of an objectified mode of research that allows us to explore latent overarching structures in pottery practices that might not be obvious for the human eye – for example, the quantitative analysis of similarities and differences of pottery vessels and styles using computational approaches (cluster analyses) for pottery classification.

Methodologically these research steps are combined with in MMR design (Heitz, 2018, pp. 106–111; Creswell & Plano Clark, 2010; Fries, 2009; Guest, 2013; Sommer Harrits, 2011; Teddlie & Tashakkori, 2009). In consequence, different classification solutions of the same pottery vessels are not understood as conflicting but complementary since each perspective has its own specific biases.

In this article, we focus on the analytical level of morphological features of vessels and the differentiation of pottery styles only, setting aside other levels of analyses like technical aspects of pottery production as well as clays and temper. Two concepts need further explanation here: our understanding of pottery vessels as "things" and our understanding of "style" (Heitz, 2018, pp. 117–129).

Regarding "things", we chose a relational conceptualization that draws on Pierre Bourdieu's habitus theorem, Etienne Wenger's theories of communities of practice, and Tim Ingold's understanding of the making of things (Bourdieu, 2007, 2009; Ingold, 2013; Wenger, 1998). Such an approach – grounded in practice theory and theories of the new materialism – emphasizes that a pottery vessel receives its material form in the process of making. This process is relational in two respects (Heitz, 2018, pp. 117–118, 126–129): the properties of the materials correspond with the skills of the potter in a mutually constitutive human-material relation (Ingold, 2013) as well as the maker and his community of practice are mutually related (Wenger, 1998) by socially shared pottery production practices through the habitus (Bourdieu, 2007). The habitus – internalized collective dispositions that guide the actions – can be understood as predictive routines that grow out of our experiences and that inform our actions in often unconscious ways (Bourdieu, 2009, p. 169; Wacquant, 2013, pp. 39–40). Hence, pottery making is understood as a relational process between the makers and their social group of belonging as well as the materials, tools, and the surroundings that become materialized in the habitus of the respective pottery vessel (Heitz, 2017, pp. 263–75; Heitz, 2018, pp. 126–29, 142).

As Pierre Bourdieu and Loïc J. D. Wacquant have argued, social reality and the habitus exist in a dual way, not only in people's minds, bodies, actions, and practices (embodiment) but also in the things they make and use, making them material representations of their actions and practices (Bourdieu, 2007, pp. 282–283, 287–288; 2014, p. 113; Bourdieu & Wacquant, 2013, p. 161). However, following Tim Ingold, crafting a pot is not about forcing a “cultural” idea like a preconceived form on an inert “natural” substance. It is rather “bringing forth of potentials immanent in a world of becoming” (Ingold, 2013, p. 31). The different designs that once existed in the mind of the potters (Ingold, 2013, p. 66) in their “anticipatory reach of imaginative foresight” (Ingold, 2013, p. 72) seem to be recognizable. Accordingly, the habitus of a vessels is not only related to the potter’s habitus but also dependent on its own the material qualities (Heitz, 2018, p. 120).

For our understanding of “style”, we use a conceptualization that was proposed by Dietler and Ingrid (1998) (Heitz, 2017, pp. 263–275; 2018, pp. 119–125). There, a “material style” is understood as “the result of characteristic ranges of responses to interconnected technical, formal, and decorative choices made at all stages of a *chaîne opératoire* of production” (Dietler & Herbich, 1998, p. 246). Such “characteristic ways of doing things” – or a distinctive “style of action” or “techniques” – are generative for “material styles” (Dietler & Ingrid, 1998, pp. 246–248). This might include a pot’s shape, its decorations and colours, the structure of the surface, its lustre, etc. (Heitz & Stapfer, 2021, p. 110). In conclusion, pottery vessels made in the same material style share characteristics of visually discernible features. And it is the morphological features of their designs and styles that we address by the qualitative but also the quantitative methods of classification proposed in this paper. Neither of the existing pottery typologies takes this into account because they were set up for chronological purposes. Linked to the concepts of cultures, they are not only biased but also incompatible and thus not suitable for a multi-site, cross-regional analysis.

The objectivism-based quantitative classification approach – but also the qualitative one – draws on the theoretical presupposition that social practices lead to structural regularities in the material world – and are structured by them in return. The reason is that recurring features of designs like applied lugs on the rim rips or knobs at the shoulder, attached handles, etc., call for intentionally made decisions. The iteration of particular designs thus indicates a certain standardization in pottery making. To repeat a design to a certain degree of conformity means to be largely in control of the potter’s clay. This can be achieved by following the appropriate movements and gestures all along a proven to be successful *chaîne opératoire*. A potter’s way of working towards a certain design thus becomes habitual with growing iteration. It results in a particular “characteristic way of doing things,” a particular “style of action” (Dietler & Ingrid, 1998, pp. 236, 246; Heitz, 2017, pp. 263–275; 2018, pp. 119–125).

Regarding the subjectivism-based qualitative classification approach, which was an integral part of the research strategy adopted in the MET project, one must emphasize that, of course as archaeologist, we cannot observe past actors and have no access their emic perspectives on pottery. But tools like the *chaîne opératoire* (Lemonnier, 1976) enable us to approach and subjectively comprehend past actions regarding pottery production and consumption practices (Hahn, 2005, p. 46). With regard to Pierre Bourdieu’s praxeology, a qualitative subjective pottery classification that draws on the process of making allowed us to investigate some of the Neolithic potters’ (habitual) actions (Grenfell, 2014, p. 26). By examining the individualized, intentional aspects of pottery, the level that reflects the conscious decisions of the potters could be approached. This was relevant to identify the process of making as such and thus capture shared pottery production practices as well as the potter’s individual “handwritings.” Since we have outlined the corresponding qualitative approach to our classification system already elsewhere (Heitz, 2017, pp. 261–282; 2018, pp. 199–201, 215–217), we will not elaborate on this here further. The ends of this procedure were ultimately twofold: on the one hand, it served as independent classification on its own right, with focusing in much more details on the individual cases and resulting in the identification of intentional decisions in the process of ceramic production; on the other hand, it was also used to underpin and justify the selection of relevant variables and datasets (assemblages of pottery/sites) for the quantitative classification. This will be explained further in the following (see also Section 3.2.3).

Materialized regularities can be transferred into variables and statistically analysed (Bourdieu, 2009, p. 162; 2014, p. 75). Bourdieu refers to this step of the “field analysis” as mapping out objective structures (Grenfell, 2014, p. 26). He himself has mainly used multiple correspondence analysis as a statistical technique because it reveals structural relations in data (Bourdieu & Wacquant, 2013, pp. 125–126; Grenfell & Lebaron, 2014, p. 3; Lebaron & Bonnet, 2014, p. 123). Such an approach can be adopted to archaeological pottery by choosing primarily metric but also nominal variables and let the computer to perform an unsupervised classification. This provides intersubjectively replicable and reproducible results in which single pottery vessel are clustered into groups that refer to material and thus objective structures in pottery production and consumption practices. To operationalise this analytical step, the relevant features of pottery must be selected as variables first. Therefore, it is vital to gain a deeper understand of the process of making a pottery vessel and the different parameters and factors that determine its stylistic outcome. This was achieved by the qualitative classification. Furthermore, the latter showed that the morphological standardization of pottery designs was limited in our case study. In the case of this hand-built ceramics, which were made using naturally occurring raw materials and then fired in pits or in open firing (Heitz, 2018, p. 333), the observable morphological variety within the ceramic vessels made in a respective design must be considered. Beside the individual “handwritings” or “micro styles” of pottery mentioned above, the qualitative differences of the materials as well as the different course of events such as ceramic firing, which cannot be completely controlled, do have their influences. This circumstance was considered by using a density-based clustering algorithm for the quantitative classification – a novelty in archaeology.

3 Materials and Methods

The selection of pottery units that we used for the unsupervised classification included two steps: the selection of pottery assemblages from precisely datable feature units within archaeological sites and the selection of vessels – pottery drawings respectively – of all existing designs, whose semi-profiles were completely preserved or could be reconstructed. The latter was key for the applied classification method that is based on numeric values of the vessels’ body shapes as well as nominal variables.

3.1 Selected Sites and Pottery Vessels

We applied our new methodological approach to assemblages of pottery dating to the period of 4000–3750 BC, accuracy depending on the dating method.[3] Taking an actor’s perspective, we can expect this span of 250 years to cover up to 8–10 generations. This is a long time relative to a potter’s life, the making of pottery in communities of practice and thus shared forms of habitus. Hence, within this time span, we must take transformations in pottery production practices into account. Therefore, we established a spatial–temporal model that separates phenomena of regional stylistic differences from local transformations (Heitz, 2018, pp. 145–153, Figure 5.7). The model founded on stratified pottery assemblages that can be reliably attributed to absolutely dated structures. Furthermore, sites encompassing as many as possible complete semi-profiles of pottery vessels were given priority, since our aim was to compare the vessels’ shapes and designs based on their silhouettes.

The different kinds of archaeological structures from which the chosen pottery assemblages derive have a considerable impact on the nature of the data that we generated out of them. Except from one case, the burial pits of Didenheim-Lerchenberg (Alsace, F) (Schweitzer, 1987), the chosen pottery originates from different kinds of settlement structures: cultural layers of lakeshore, bog, river, and cave sites as well as pits and trenches from hilltop sites and earthworks. These different find contexts affect the number and preservation of pottery vessels and the accuracy of dating: while wetland sites might be dated to the exact years by means of dendrochronology and thus show the different length of time in which the pottery accumulated, trenches tend to have several cultural layers and thus fillings while pits might represent short-lived closed finds. The latter two, however, can only be dated by ^{14}C into a probability range of several

---

[3] In the case of the sites Sutz-Lattrigen (BE, Switzerland) and Schenkon (LU, Switzerland) also pottery assemblages from younger (~100 years) sites needed to be chosen, because older ones were lacking. The typologically dependent morphological difference, however, is expected to be neglectable.

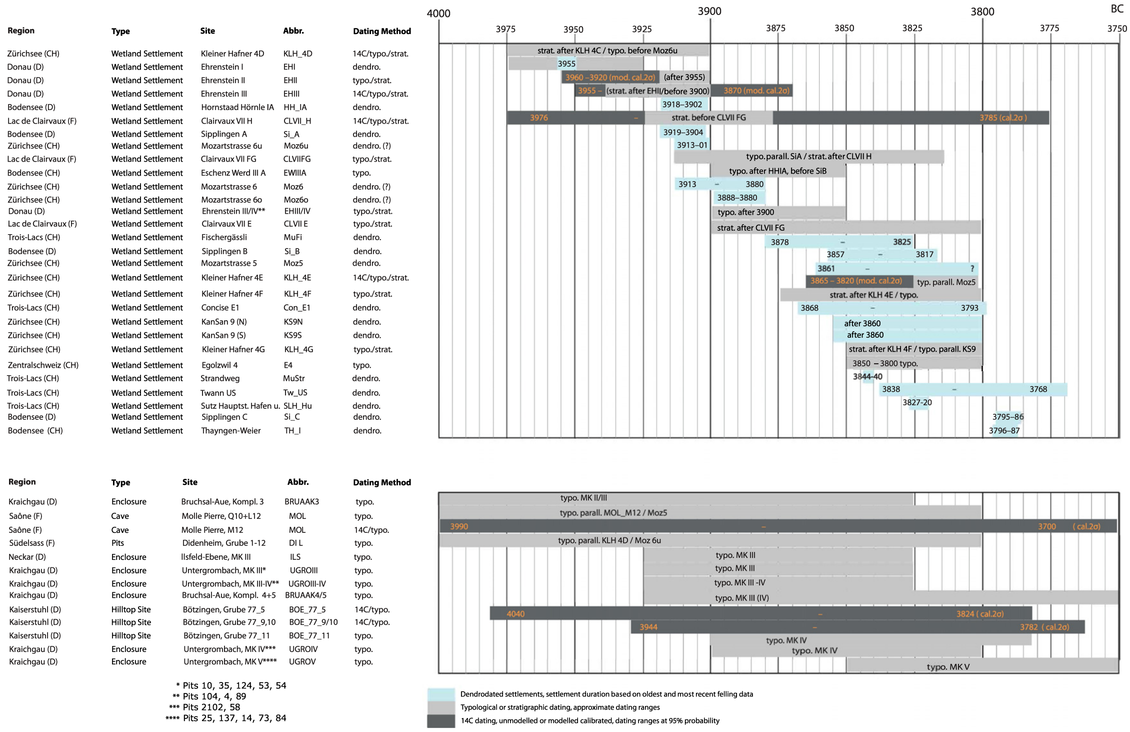


Figure 2: Temporal model of considered settlements and assemblages of pottery (figure: after Heitz, 2018, Figure 5.7). MK = Michelsberg.

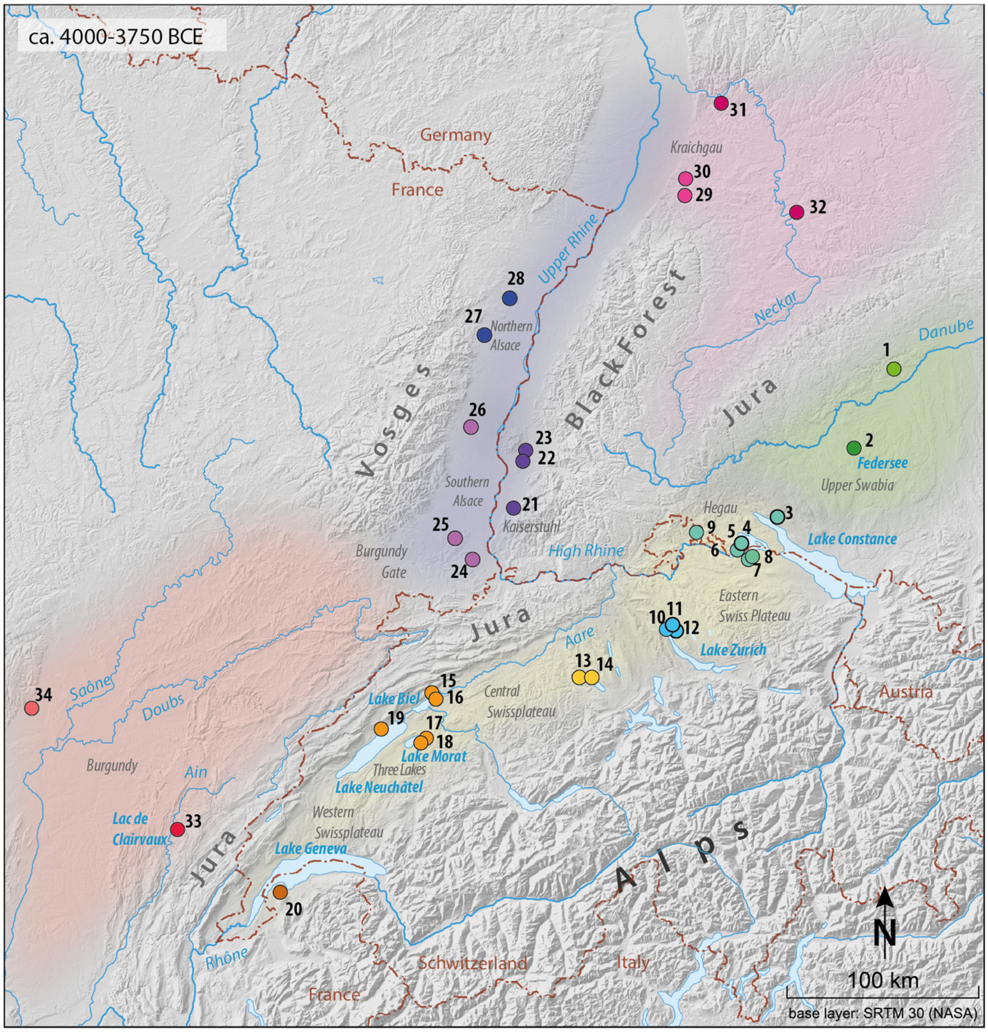
decades.[4] In cases where no absolute pottery-independent dating of the structures was available, the pottery assemblages could be matched with absolute dated ones based on typo-chronological arguments (Heitz, 2018, pp. 154–182). The resulting chronological framework that considers the differences between wetland and dryland site conditions, can be seen in Figure 2, which also indicates the different dating methods.

Within the core region of the MET project, the Northern Alpine Foreland, the settlement phases of dendro-dated wetland sites were assigned to an absolute chronology based on the oldest and youngest known felling dates per site. In the case of wetland sites with several stratigraphically separable settlements or settlement phases, cultural layers lacking scientific dating could be relatively dated within these stratigraphic sequences. Thus, it was possible to allocate their pottery assemblages to a certain absolute time span in a second step. In case of existing ^{14}C-dates they sometime could be modelled using the Bayesian “sequence-models.” Thanks to the high density of precisely dendro-dated wetland sites of the Swiss Plateau the typo-chronological transformation of pottery can be closely traced. This allowed us also to include only typologically datable pottery assemblages into the absolute time model too. Most of the sites in Southern Germany and Eastern France cannot be dated more precisely then to the overall time span of the 250 years. However, since typical pottery-designs from these regional styles, early NMB (Néolithique Moyen Bourguignon, Petrequin & Pétrequin, 2015), Michelsberg (primarily the periods MK III and IV, after Lüning, 1968; Seidel, 2017), and Munzingen (Jammet-Reynal, 2017) occur in the dendro-dates wetland sites, the relative dating of the chosen pottery assemblages could be verified.

While the time model is basically based on dendrochronologically dated settlement phases and stratigraphy, the grey and black bars in Figure 2 indicated approximate timespans and in the case of ^{14}C dates the probability intervals (95% likelihood) within these settlements existed and the pottery assemblages formed, respectively. The longer time spans thus are only a result of lesser accurate dating methods, not necessarily a result of long-duration settlement occupations.

---

[4] Only already published ^{14}C dates were used, and future large-scale ^{14}C-projects are something most desirable.



NORTHERN ALPINE FORELAND (EAST)

Danube	1. Ehrenstein-Blaustein (I, II, III, III/IV)
Federsee	2. Alleshausen-Hartöschle
Lake Constance	3. Sipplingen-Osthafen (A, B, C)
	4. Hornstaad-Hörnle (IA, II)
	5. Wangen-Hinterhorn (KS 1)
	6. Eschenz-Insel Werd III (A)
	7. Steckborn-Turgi
	8. Steckborn Schanz
	9. Thayngen-Weier (I)
Lake Zurich	10. Zürich-Kleiner Hafner (4C, 4D, 4E, 4 F, 4G)
	11. Zürich-Mozartstrasse (6u, 6o, 5)
	12. Zürich-KanSan (9N, 9S)

NORTHERN ALPINE FORELAND (WEST)

Central Swiss Lakes	13. Egolzwil 4 (H I-IV)
	14. Schenkon-Trichtermoos
Three Lakes	15. Twann-Bahnhof (US)
	16. Sutz-Lattrigen Hauptstation-Hafen (u)
	17. Muntelier-Fischergässli (4, 3, 2, 1)
	18. Muntelier-Strandweg
	19. Concise-Sous-Colachoz (E1)
Lake Geneva	20. Corsier-Port

UPPER RHINE PLAIN

Upper Rhine-Südbaden	21. Bötzingen-Häuslinsberg
	22. Munzingen-Tuniberg
	23. Eichstetten-Gutensberg
Upper Rhine-Southern Alsace	24. Magstatt-Le Bas
	25. Didenheim-Lerchenberg
	26. Houssen-Gravières
Upper Rhine-Northern Alsace	27. Rosheim-Sablière Maetz
	28. Wolfisheim-Pipeline

NECKAR- AND TAUBER-GÄUPLATTEN

Kraichgau	29. Untergrombach-Michelsberg
	30. Bruchsal-Aue
Neckar	31. Heidelberg-Handschuhsheim
	32. Ilsfeld-Ebene

RIVER VALLEYS EASTERN FRANCE

Lac de Clairvaux	33. Clairvaux-Les-Lacs
Saône	34. Mavilly-Mandelot, Grotte de la Molle Pierre

Figure 3: Selection of sites dating approx. 4000–3750 BCE and their geographical location in the northern Alpine Foreland and adjacent regions, highlighted in black: sites with pottery assemblages used in this paper for the unsupervised classification approach (figure: C. Heitz., map base according to Jammet-Reynal, 2017).

For the quantitative analysis, 1,046 vessels were used, which are distributed over 14 microregions (Figure 3) and 44 strata or diagnostic feature units of sites from the time between 4000 and 3750 BC. The dataset with a basic description of the metadata can be found here: doi: 10.5281/zenodo.7258694.

The vessels were selected according to the principle of weighted random sampling. The aim was to record the range of forms and decoration of the pottery of the respective site feature unit based on the already published catalogues of pottery drawings. In the case of the wetland sites, it was also possible to sift through the original material in the archaeological collections. For each site and ceramic assemblage, all vessels with reconstructed semi-profiles as presented in the respective catalogues were selected. This ensured that a maximum of morphological features could be recorded. In the case of very large ceramic assemblages, a representative selection was made (Heitz, 2018; Stapfer, 2019). The drawings were scanned and kept as separate “png” files per individual ceramic vessel (see Section 3.2.3).

3.2 Methodology

3.2.1 Considerations for Method Selection

The mentioned peculiarities regarding the character of the data as well as the ambition of the project required some considerations about the most appropriate classification method. A feature-based classification system is best suited for studies of entanglement and transformations across ceramic styles, which must therefore be carried out syn- and diachronically as well as across settlements and regions. The aim was to combine differently scaled variables with each other to consider not only the characteristics of the vessels (nominal variables) but also their absolute dimensions (metric variables). Only in combination can the designs intended in the production of the vessels (imagined ideal types of vessels) be captured, which respond to the demands of the consumers.

In the production of hand-built vessels, we must expect a low degree of standardization. Or to put it another way, the material concrete form of the ceramic vessels never corresponds exactly to the ideal type of an imagined design, so that we must assume considerable variability between vessels of the same design (see Section 2). From a statistical point of view, we must assume that the data for such ceramic vessels or their combinations of features will be rather noisy. This means that the expected classes contain a large data variance.

3.2.2 Methods of Data Analysis: PCA and Iterative Clustering, t-SNE, and HDBSCAN

To address this initial situation, a statistical model was devised and implemented that involves a combination of several successive multivariate statistical procedures: first, all nominal variables are mapped to a presence/absence matrix as dummy variables as 0 or 1. In this way, each variable specification becomes a variable that can be weighted. The weighting is carried out by means of a separately preceding principal component analysis (PCA). This is followed by the ordination of the data by the t-SNE algorithm to achieve a reduction in the dimensions in the high-dimensional data space. This is used to find latent pattern, i.e., similarity structures, in the data. For the grouping of these into the final classes, a cluster analysis is performed, which uses an algorithm that is particularly suitable for noisy data: HDBSCAN. Our application of the clustering method adopts an iterative approach. This means that within the clusters found in the cluster solution, clustering is carried out again in a second step to find clusters again on a second, third, etc., hierarchical level. The algorithm is implemented in the statistical environment R.

The methods were chosen because of the presumed high noise in the data. To achieve a better structuring of the data, instead of the usual PCA, t-SNE was used as the ordinating method, for which the R package “Rtsne” was used. The mathematical principles are explained in Van Der Maaten and Hinton (2008) and Van Der Maaten (2014). Conducting cluster analyses based on previous ordination is a procedure that is frequently used, as explained in detail in Ding and He (2004), for example.

t-SNE is an algorithm for representing high-dimensional data in a low-dimensional space, which is often used in data mining and machine learning. Multivariate datasets can have similarity structures at different levels in the high-dimensional space. This is also the case in multivariate datasets of ceramic characteristics (Van Der Maaten & Hinton, 2008; Van Der Maaten, Lange, & Boon, 2009). To make them manageable, these similarity structures need to be reduced to fewer dimensions. The advantage of t-SNE as an ordinating method compared to PCA is that a non-linear dimensional reduction takes place and that local similarities are weighted higher than global dissimilarities. As a result, the original clustering in high-dimensional space – i.e., the multidimensional similarity structures – can be mapped in a space of two or three dimensions in a more reduced meaningful way. As much as possible of the statistical (dis)similarity between the original data is preserved in the process. The t-SNE algorithm is preceded by a PCA by default. This circumstance could be used here to control the weighting of the variables by replacing the PCA already provided in Rtsne with a weighted version from the R package "FactoMineR." The weighted ordination is then processed further using t-SNE.

To convert the similarity structures in the ceramic data into separate classes, a cluster analysis was carried out based on the t-SNE results, which constitutes the actual classification process. Instead of traditional methods (hierarchical cluster analysis or kmeans), a density-based clustering method without the requirement to specify a number of classes (non-partitioning) was chosen here, which utilizes the HDBSCAN algorithm (Campello, Moulavi, Zimek, & Sander, 2015; Campello, Moulavi, & Sander, 2013; Ester, Hans-Peter, Jörg, & Xiaowei, 1996). This algorithm, used and developed for data mining, is particularly suitable for noisy data. For explanations of the underlying mathematics and programming, see McInnes, Healy, and Astels (2017).

HDBSCAN has the following advantages:

1. In contrast to common methods, not only circular cluster structures can be found by the density-based method but also irregularly shaped similarity structures.
2. Not every data point has to be allocated to a cluster. Unique ceramic objects that deviate strongly from the more standardized production practice, as well as serially produced designs that appear only sporadically in the dataset, do not necessarily represent an ideal type and should therefore not constitute or be part of a cluster. Instead, they are dropped from the final cluster solution as noise. This results in less data dispersion and thus greater similarity within the clusters.

HDBSCAN is therefore better suited for mapping commonly shared groups of characteristics that exist across individuals in the sense of ideal-typical designs. The R package "hdbscan" was used to implement the cluster analysis.

Furthermore, an iterative procedure was implemented. This results in clusters on several hierarchical levels which are, however, independent of each other. Overall, this considers that different features can be relevant at different classificatory levels, which is close to human classification practice. In the individual clusters created, classification is again carried out within the now limited feature space to further subdivide them as far as possible. As the lower limit, five ceramic vessels were chosen as the minimum number per cluster. Thus, the systematic iterative clustering ends when there are only five vessels left within a cluster, without considering those that have been identified as noise. Another criterion for terminating further subdivisions is when the vessels within a (sub)cluster are so similar that no further subdivision can be calculated. In total, the vessels could be divided into up to three hierarchical levels according to the number of vessels of ever smaller similarity groups, whereby the same logic was applied as for the entire cluster solution.

A special feature of t-SNE lies in its stochastic nature: since t-SNE implements a random process, the results of the individual runs can be slightly different – even if they are fundamentally similar in each case. To solve this challenge, a consensus clustering approach with majority voting, similar to the meta-clustering algorithm (cf. Strehl & Ghosh, 2002), was used. For this purpose, several clustering runs are performed in sequence, and finally those objects are combined into clusters, which in most cases were in the same cluster. In our case, the final cluster solution was achieved after 101 automated successive passes. This made it possible to achieve a stabilization and thus also a reproducibility of the result. The number 101 is an arbitrary choice that we believe represents a good balance between stability of results and speed of calculation. Only the fact that the number is odd is a mandatory requirement: to achieve a majority decision in any case and avoid a stalemate.

3.2.3 Methods of Data Collection: Selection of Variables

As already indicated, a combination of metric and nominal variables and values derived from converted ceramic drawings of 400 measurement segments from the centre to the outside of the vessel were used for the ordination and classification of the vessels.

The variables were selected according to the following criteria (Heitz, 2018, pp. 207–208):

- Comprehensibility and reproducibility: The variables should be clearly identifiable from the ceramic drawings and their assessment should therefore as less subjective as possible.
- Representation with regard to morphological phenomena on the vessel: The selection of the characteristics to be coded as well as their weighting (cf. below) is based on the observable morphological aspects on the vessel. Thus, the variables or their characteristics can be bundled into 10 variable groups per phenotype: profile shape, break shape, vessel height, closedness, base shape, handles, eyelets, knobs, ledges, and decorations.
- Relevance with regard to intentional action in the production process of the vessels: As far as possible, variables were selected from which heuristic conclusions can be drawn about intentional action and thus clear decisions about action by the producers in the production process (design approach). The profile forms hold the greatest potential for unintentional action, as different, individual producer handwritings become apparent here.
- Distinction quality: Preference was given to all variables that promised predictable distinguishing characteristics about the different production practices, to obtain a dataset with as much structure as possible. For example, the rather rare decorations (“décor”) were recorded only on a coarse scale.

A total of 584 variable attributes were recorded across the 10 variable groups. These were thus considered proxies for latent complex variables, which reflect the design intentions of the producers (designs) and thus also fundamental demands for use on the part of the consumers. By recording the profile shapes based on 400 points per vessel, corresponding to 400 variables, a method-immanent weighting arose. The profile form as a group of variables was very dominant compared to the remaining 184 variables, so that the latter hardly had any effect in the cluster solution. To control the weighting, the PCA included in t-SNE was inactivated and a weighted version was implemented in which the variables were weighted beforehand. Each of the 10 groups of variables was given the same weight of 1, making them equally important.

This weighting was implemented for clustering on the first hierarchy level. For the clustering on the following hierarchy levels, the controlled weighting was omitted, whereby the method-related high weighting of the profile forms affects the formation of the clusters. We tested three variants (no controlled weighting, controlled weighting throughout, and the aforementioned option) and then empirically decided on this version. This has the advantage that the vessels are first divided into coarse groups based on as many variables as possible, within which these can then be classified into finer subgroups aimed at the similarities and differences of the profile forms. We found this to be a very welcome characteristic of the resulting classification, which was decisive for the selection of this variant.

The external validation of the cluster solutions, i.e., the check whether meaningful clusters were calculated, was carried out visually for the nominal variables based on the ceramic drawings of the vessels in the clusters and based on univariate statistics, such as the mean values and standard deviations of the profile shapes or the vessel height distribution of the ceramic individuals within a cluster. Overall, the weighting of the variables is one of the most sensitive, subjective, and adjustable steps in the whole process of autonomous classification and will certainly offer further possibilities in the future, depending on the issues at stake.

Another possibility to adapt the algorithm in the future is to consider or omit absolute measurement distances such as the total height of the objects. Currently, this is included and has a significant influence on the result. However, if this is not added, and since the 'measurement distances' are collected as a relative measure on a scaled drawing, the shaping of objects of different dimensions can also be synchronized with each other in principle.

3.2.4 Implemented Workflow

These mentioned considerations led to a specific selection of data and methods. The largest amount of the data represents 400 individual values which describe the shape of each vessel. For this purpose, the drawings of the vessels were aligned perpendicularly (as good as possible) and the vessel shape was filled in black. This was done by hand using standard graphics processing software (GIMP). The resulting representation of the vessel profiles was halved (centreline of the vessel on the right of the resulting image) because the vessel body could be understood as a rotational body for analytical simplification. In this context, the handles and other three-dimensional decorative elements were also removed from the drawing and their presence was later reintegrated into the dataset by adding nominal variables. The resulting pure black and white images were then scaled to 400 px to ensure uniform data recording. This of course also removed any reference to the absolute vessel height or true-to-scale image. Therefore, information on absolute measures was included as separate parameters in the dataset. The resulting pictures were saved as PNG graphics and served as a basis for further analysis, which was carried out by means of R. An automation of this process was developed in the course of the analysis (https://github.com/ISAAKiel/shapAAR), but it was not yet used in the study presented here.

In R, the individual images were loaded using the command readPNG (package png) and converted into a matrix consisting of 0 and 1. In this matrix, the last occurrence of a white pixel was detected in each row from the left (i.e., from the outside of the vessel, represented by 1 in the matrix) and this was subtracted from the total width of the image. Thus, the distance of the profile line to the centreline of the vessel can be obtained for each of the 400 sampling lines. This information represents the complete information available about the vessel profile, as it (for the comparatively rather simply shaped vessels of our case study) can be reconstructed unambiguously from these values.

The database thus obtained regarding the vessel shape was enriched by further information, as can be seen in figure (Figure 4). Without going into detail here (see Heitz, 2018, pp. 195–197), nominal classifications were compiled on three-dimensional applications (position, shape, and series, if applicable), handles (position, shape, and series, if applicable), decorative elements, as well as aspects of the shaping of design elements that are difficult to grasp with the scanning of the overall shape (bottom shape, break). Furthermore, the absolute height of the vessel (bottom to rim) as well as two diameters (minimum and maximum diameter excluding the bottom, in the case of bowls minimum and maximum diameter including the bottom) given as landmarks were added as metric values to provide a closeness attribute.

The weighting is chosen in such a way that all variable groups from Figure 4 each have a total weight of 1, so that each variable within a group receives a weight of 1/n__variable__per__group. Thus, for example, the total height of the vessel has the same weighting as the total shape, as well as the bottom shape. This setting may seem arbitrary, but due to the clearly different quality of the variables in each group, on the one hand, and the most neutral possible parameterization, on the other hand (all definitely different qualities are weighted equally), this setting seemed to us to be the model with the lowest amount of assumptions. The PCA command (Package FactoMineR; Husson, Josse, Le, & Mazet, 2017) was used to achieve a weighted principal component analysis.

The results of the principal component analysis are used as input values for the t-SNE in the next step using the Rtsne command. One of the few and perhaps the only relevant parameter for t-SNE is perplexity, a parameter that balances how much the local neighbourhood is weighted against the global neighbourhood for ordination. A perplexity of 30 was selected as the default value, but this was reduced if necessary for small datasets (in the following subclustering process), so that it always corresponded to a maximum of 1/3

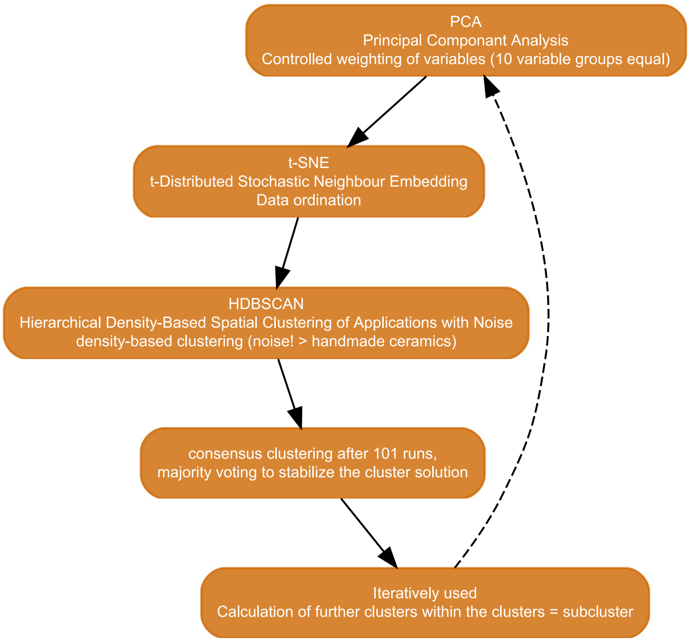


Figure 4: Full workflow of the computer-aided classification (figure: M. Hinz).

of the number of data – 1. This dynamization was necessary because a large perplexity value resulted in an error in such cases where there were few vessels within a cluster. At the same time, a PCA was skipped during t-SNE (pca = false) since it had already been performed. The parameter theta, which determines the compromise between speed and accuracy, was set to 0 to obtain an exact ordination. An output dimensionality of 2 dimensions was chosen.

In the third step, the ordination thus obtained was clustered using hdbscan. The parameter minPts, which acts as a minimum cluster size to detect and as a 'smoothing' factor of the density estimates, was set to 5 under the assumption that less than five vessels cannot reasonably represent an ideal type.

The fact that t-SNE is a stochastic algorithm results in each run producing a slightly different configuration of the ordination. Consequently, the result of the subsequent clustering becomes stochastic, so that in each run slightly different assignments of particularly ambiguous objects can occur. To stabilize the result, the procedure of majority-based consensus clustering was chosen, in which 101 ordination/classification runs were generated, and then, the objects, which occurred together in the majority of cases, were mapped into new, stabilized clusters.

This workflow (Figure 5) was first applied to the entire dataset. After this run, the same routine was applied recursively to the objects within these clusters until either a hierarchy depth of 3 was reached or the cluster in question contained less than five vessels (with the same reasoning as for the setting of the minPts parameter in HDBSCAN).

The resulting clusters were stored as folder structures in the file system, and the resulting data of the cluster-internal objects and the images of the vessels were saved here. In a further step, descriptive statistics and diagrams could be generated for each cluster.

It was also helpful for the evaluation that a "typical" vessel in the sense of a centroid could be determined for the respective cluster, on the one hand, as a virtual, ideal vessel and, on the other hand, by identifying the vessel, which could be regarded as the most typical real vessel. The centroid was calculated and then, the vessel was selected that was closest to the centroid. This vessel then served as a typical example for the hierarchy graph, which was generated by the program Graphviz.

Finally, all statistics and diagrams as well as the images of the vessels included in each cluster were converted into a PDF using an R Markdown script, and the PDF then provided the basis for the evaluation and interpretation of the results.

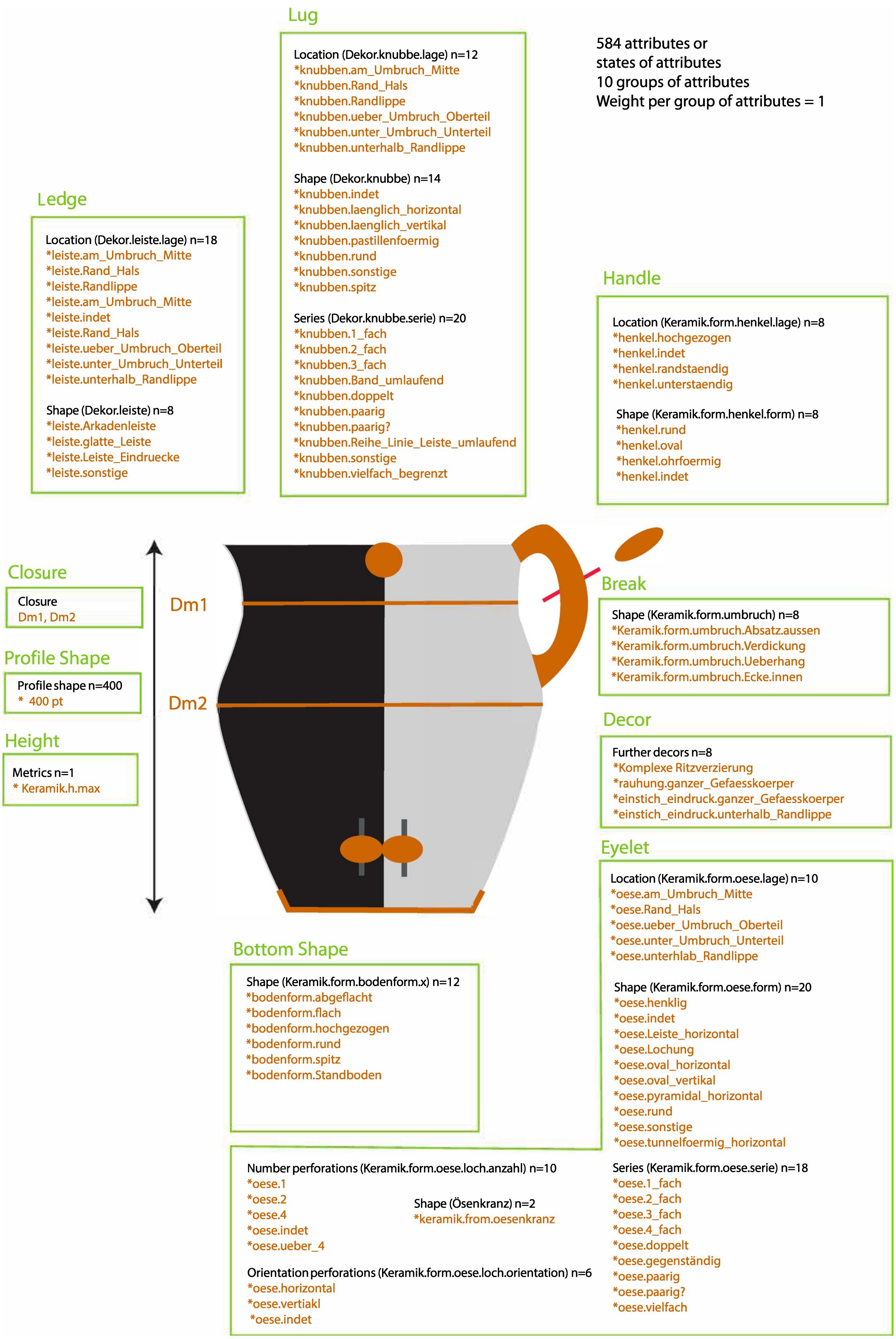


Figure 5: Attributes or states of attributes and their weighting (figure: C. Heitz, for an English translation of all variables, see doi: 10.5281/zenodo.7258694).

4 Results

4.1 Description of the Cluster Solution

The unsupervised classification procedure separated the pottery vessels into 90 clusters on three different hierarchical levels. On the first hierarchical level, the vessels were grouped into 27 clusters, while some vessels, the black dots (marked as cluster 0) in Figure 6, were excluded as noise. In the ordination by t-SNE as discussed above, the local similarity of the data is taken more into account than the global one. Consequently, only the relative position of clusters connected by a point cloud of noise objects is meaningful. This applies to the centrally mapped group of clusters, but not to the position of the isolated individual clusters. The latter, clusters 6, 19, 14, and 23, are not surrounded by noise and are clearly separated from the more central group of other clusters. The number of data points (ceramic vessels) per main cluster varies considerably, from $n = 6$ to $n = 209$. Especially the large main clusters can be subdivided into subclusters or sub-subclusters on the second and third hierarchical levels.

Following the implemented iterative procedure, similarity groups were again formed within all of the 27 main clusters – omitting the noise – and again data points were marked as noise. On the second hierarchical level, again excluding the noise, a total of 32 clusters were formed, whereby not every main cluster has subclusters. In the same way, similarity groups were also calculated on a third hierarchy level and a total of 29 clusters were formed anew, whereby again not every second level cluster could be further subdivided. Since the minimum number of vessels was limited to $n = 5$, the third hierarchy level could not be calculated

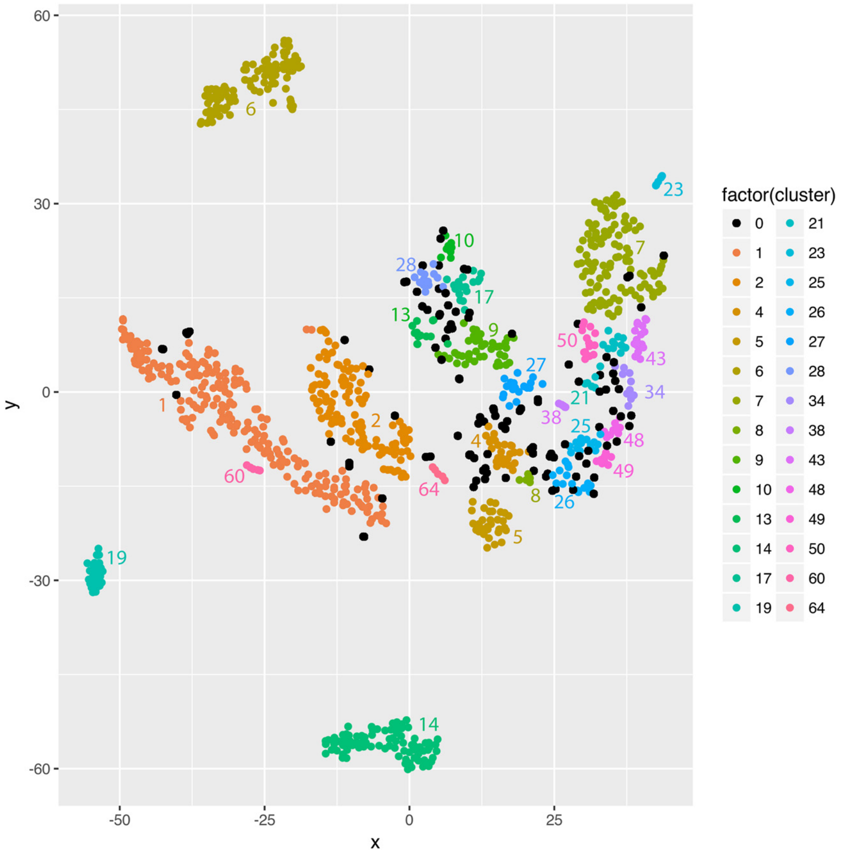


Figure 6: Representation of the 1st hierarchy level (main cluster) of the HDBSCAN cluster analysis on the two-dimensional t-SNE visualization; data points identified as noise are indicated in black (cluster 0) (figure: M. Hinz).

in any case. The entire representation of the cluster hierarchy is seen in Figure 7, whereby the individual groups are represented by the “most typical” vessel (objects nearest to the centroids).

Within the (sub)clusters of each hierarchical level, various uni- and bivariate statistics and graphical representations were automatized based on the variables recorded in the projects database. The following turned out to be particularly suitable for characterizing the (sub)clusters:

- box plots for data scattering of vessel heights;
- histograms and kernel densities for frequency distribution of vessel heights;
- plots of superimposed curves of the profile shapes (400 points);
- mean values and standard deviation respectively 1 and 2 σ-intervals for the profile shapes; and
- bar charts of the absolute frequencies of the bottom shapes.

In addition, the individual statistics and drawings of the vessels per (sub)cluster are compiled automatically, the centroid vessel is indicated, and any noise and further subclusters are shown on the t-SNE visualizations.

As an example, the results regarding cluster 14, respectively, subclusters 14.1 and 14.2, as well as the clusters on the third hierarchical level, clusters 14.1.2, 14.1.6, and 14.1.8 are shown (Figures 8 and 9).

Overall, this results in a classification that primarily groups according to vessel size, compactness (height-width proportions) and closure (Dm 1 to Dm 2, see Figure 4). On the first hierarchical level, the

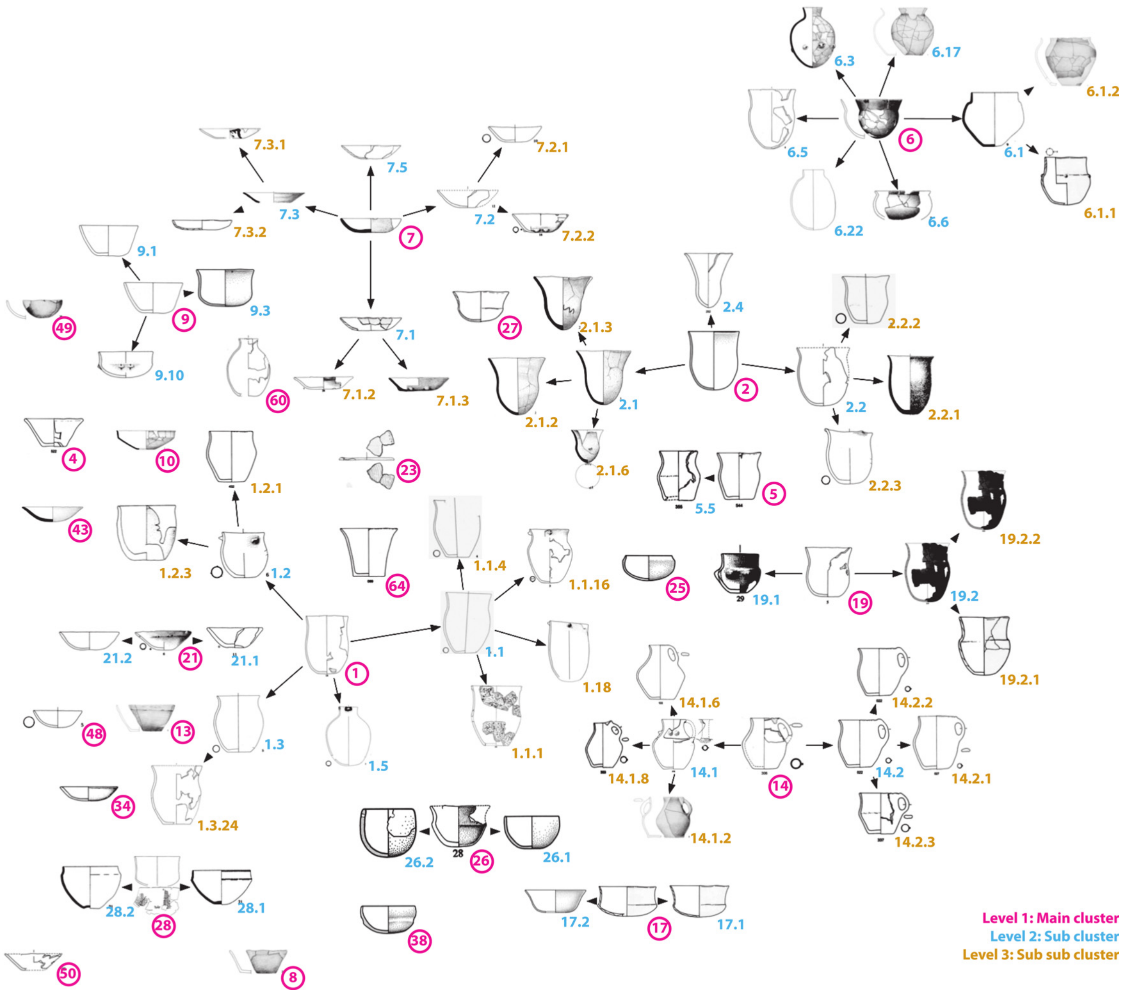


Figure 7: The clusters and their subcluster on three hierarchical levels, represented by the centroid vessels, displayed as hierarchical network (figure: M. Hinz and C. Heitz, Copyright of pottery drawings: doi: 10.5281/zenodo.7258694.).

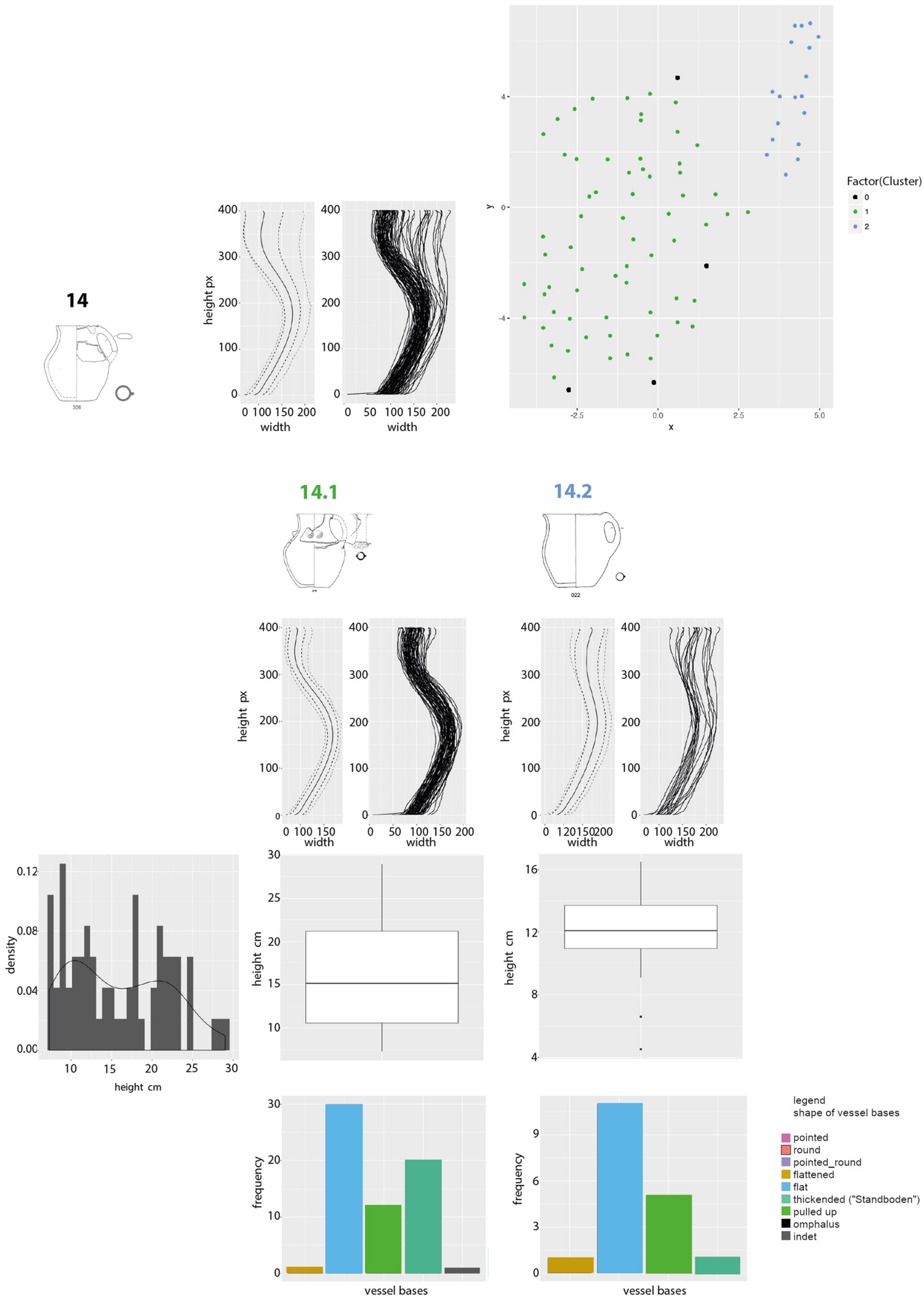


Figure 8: Cluster 14 and subclusters 14.1 and 14.2 (figure: M. Hinz and C. Heitz, for the picture credits of the pottery drawings, see doi: 10.5281/zenodo.7258694).

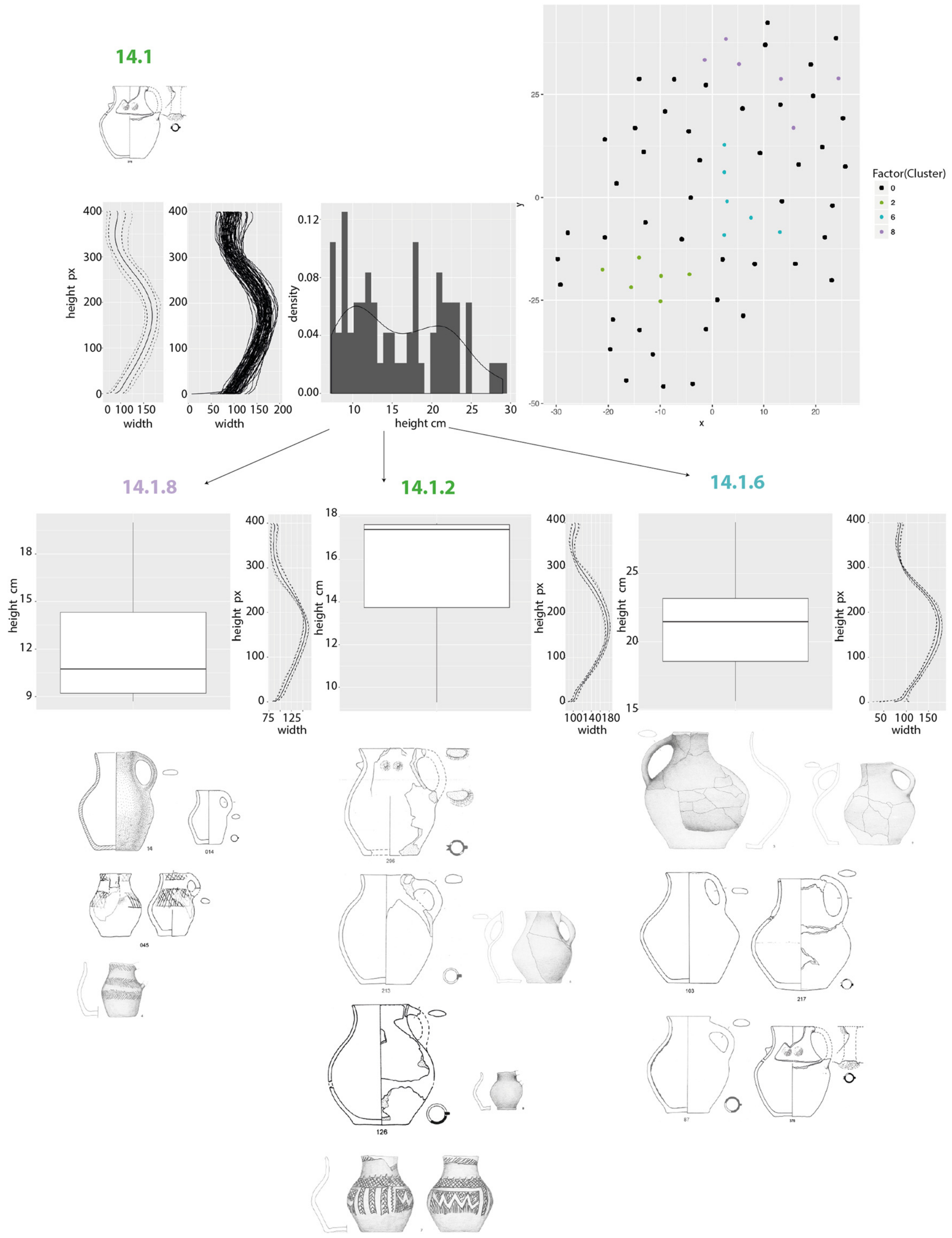


Figure 9: Subcluster 14.1, 14.1.8, 14.1.2, and 14.1.6 on the third hierarchical level (figure: M. Hinz and C. Heitz, for the picture credits of the pottery drawings, see doi: 10.5281/zenodo.7258694).

height tends to be more dominant and on the second, the profile shape. Characteristics, such as bottom and break shapes, handles, and ornamental elements, play a subordinate role but are nevertheless incorporated. In some (sub)clusters, the nominal variables occur in different degrees. This is less due to the chosen statistical model itself than to the dataset as such: The total variance in the data is relatively large. If the number of similar vessels was larger with respect to the variables, clusters with lower dispersion would result. The relatively small number of sufficiently preserved vessels, the large vessel spectrum of the various Late Neolithic ceramics, and the rather low standardization of ceramic production all contribute to this.

The existing or lacking seriality (Heitz, 2017; 2018, pp. 119–126, 300) of the vessels produced also plays a role. Cluster 14, vessels with handles, was produced in relatively large numbers in a relatively standardized manner, i.e., clearly in series. Not only does cluster 14 separate itself from the rest of the vessels without noise points, but subclusters 14.1 and 14.2 also have a finer subdivision into jar-shaped and cup-shaped handle vessels, i.e., taller and more closed, contrasting with wider and more open forms. Here, therefore, a very clear separation is achieved by computer-based classification.

A further example are subclusters 2.1, 2.2, and 2.4 of cluster 2, in which the Michelsberg bag and tulip beakers ('Tulpen- und Beutelbecher' according Lüning, 1968), which are also produced in a highly standardized way, are grouped and separated from other bulbous tulip-shaped vessels (Figure 10). Subcluster 2.1 ($n = 26$) groups bag and tulip Beakers together, although these are even more strongly subdivided into subclusters 2.1.2, 2.1.3, and 2.1.6 based on their profile shape. Subcluster 2.2 ($n = 72$) collects much more bulbous vessels compared to subcluster 2.1, but they are also strongly open and tend to be tulip-shaped. The group is relatively heterogeneous. Flat and raised bases also occur, as well as applications of all kinds. The distinction into further three sub-subclusters, 2.2.2, 2.2.3, and 2.1.6, allows an even finer subdivision according to the profile shape and thus the squatness and closedness of the vessels. Subcluster 2.4 ($n = 10$) classifies compared to 2.1 more slender vessels together, which are called slender tulip beakers ('schlanke Tulpenbecher') according to Lüning's typology (Figure 10).

Other clusters, such as cluster 6, unite morphologically very different vessels, which, however, still show comprehensible similarities: prominent rims and spherical vessel bodies. On the second hierarchical level, these are then subdivided according to their different profile forms, closeness, and compactness (Figures 11 and Figures 12). Especially in the case of the deep and shallow bowl-shaped vessels, it is noticeable that the dominance of the variable height on the first hierarchical level provides for separation. Since with such forms, the height seems less relevant from an archaeological point of view than, for example, the width or the closeness, there is rather a too fine subdivision here. In addition to sorting out noise points using HDBSCAN, the integrative procedure implemented is particularly helpful here, as clusters at different hierarchical levels can be used for further investigations.

The division of the vessels into clusters makes it obvious that the classification neither follows qualitatively defined typo-chronological units such as so-called "archaeological cultures" nor spatial units such as settlements or regions (Figure 13). The distribution of the settlements (diagnostic units) among the main clusters is very heterogeneous: none of the clusters consists solely of vessels of one settlement. This shows that the unsupervised classification detects similarity groups that basically connect the vessel spectra of the different settlements. A second look at the distribution of the traditional typo-chronological ceramic groups also shows that similarity structures are found with this method that cross, so to speak, the common "cultures" or potentially combine different ceramic practices.

4.2 Interpretation of the Cluster Solution

In view of the Late Neolithic, morphologically highly variable pottery, however, the statistical model chosen succeeds in creating a surprisingly stringent, archaeologically meaningful, unsupervised classification: at different hierarchical levels, the vessels are roughly separated into similarity groups regarding morphologically fundamental properties. These were composed of the absolute size, the degree of closeness and compactness, and the profile curve. Variables such as break (shoulder rsp. belly) or bottom shapes and attachments are of secondary importance. This fails in separating types or designs in the way one would do impressionistically. However, computer-aided classification provides something else which the human eye and memory can hardly achieve – especially not for more than 1,000 vessels: a separation into different morphological similarities, including size, on which the ceramic dataset is based. Thus, the cluster solution

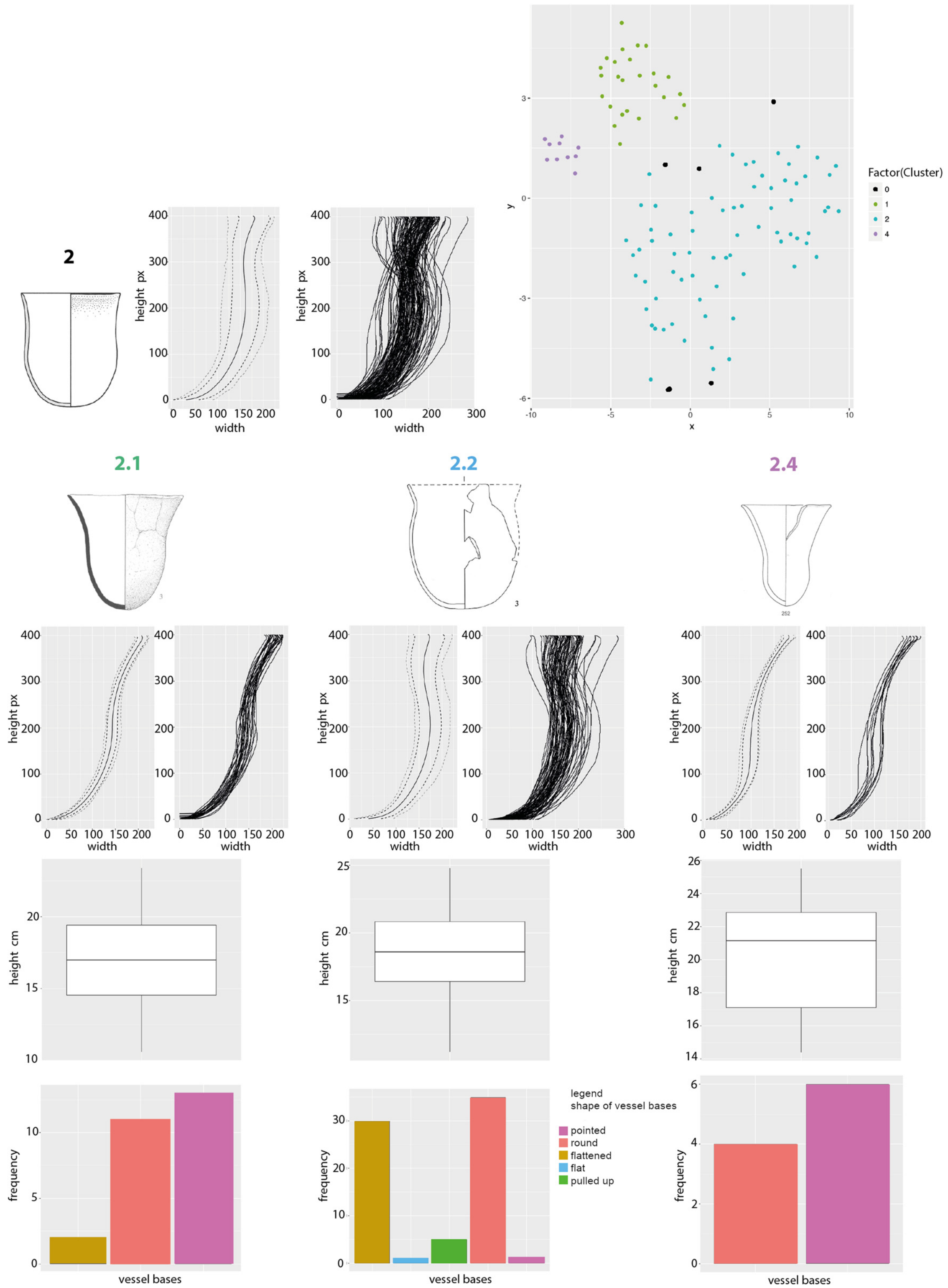


Figure 10: Cluster 2 and subclusters 2.1, 2.2, and 2.3 (figure: M. Hinz and C. Heitz, for the picture credits of the pottery drawings, see doi: 10.5281/zenodo.7258694).

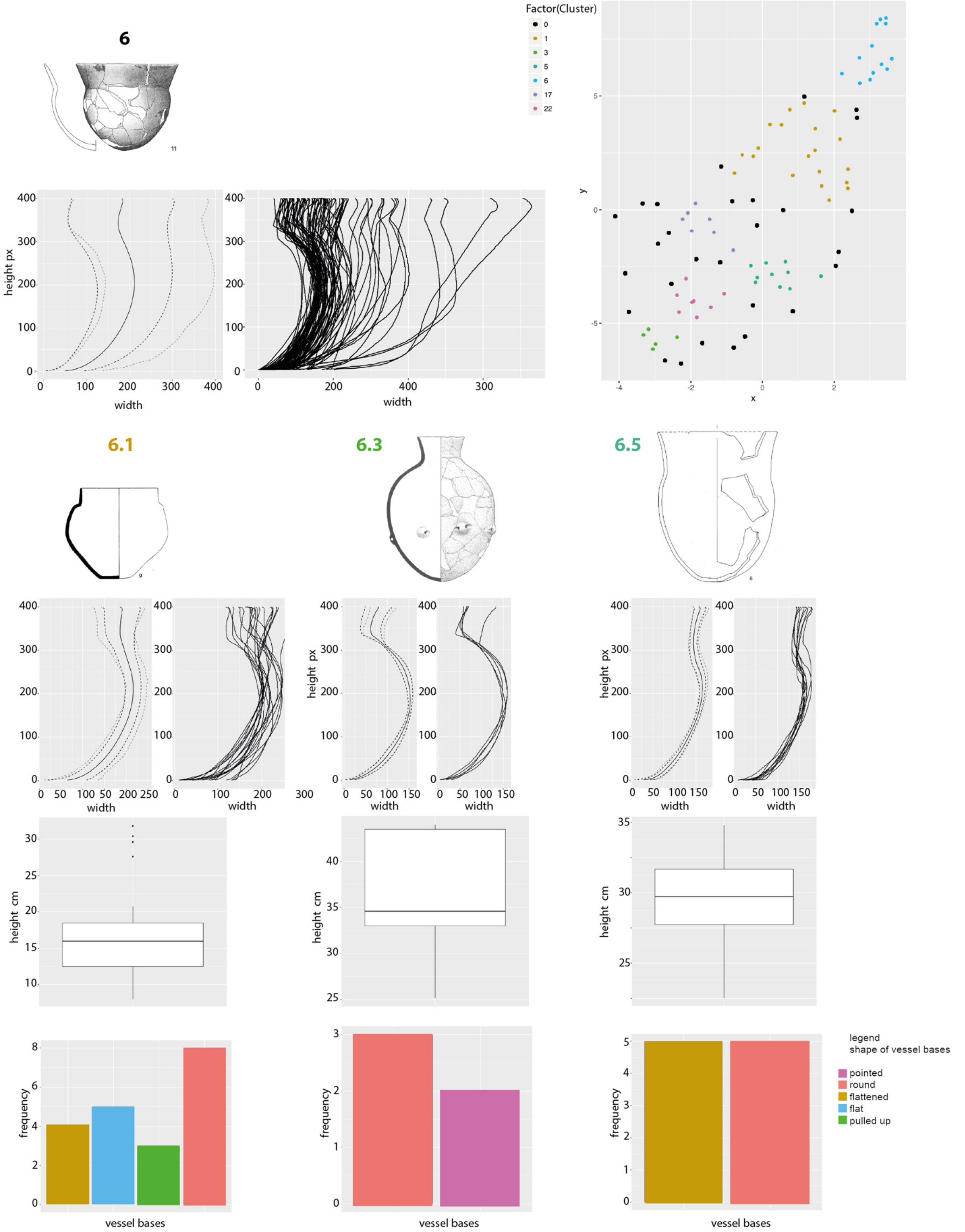


Figure 11: Cluster 6 and subclusters 6.1, 6.3, and 6.5 (figure: M. Hinz and C. Heitz, for the picture credits of the pottery drawings, see doi: 10.5281/zenodo.7258694).

shows primarily a grouping according to "basic forms" in different degrees of refinement. Interestingly, these now lead the perspective from production to the consumption level, since basic forms are associated with "functions" or rather "latent potentials of vessel use" (Heitz, 2018, pp. 289, 300; 2022, pp. 390–392): The basic morphology of the vessels, the basic form, has an influence on phenomena of affordance: on what

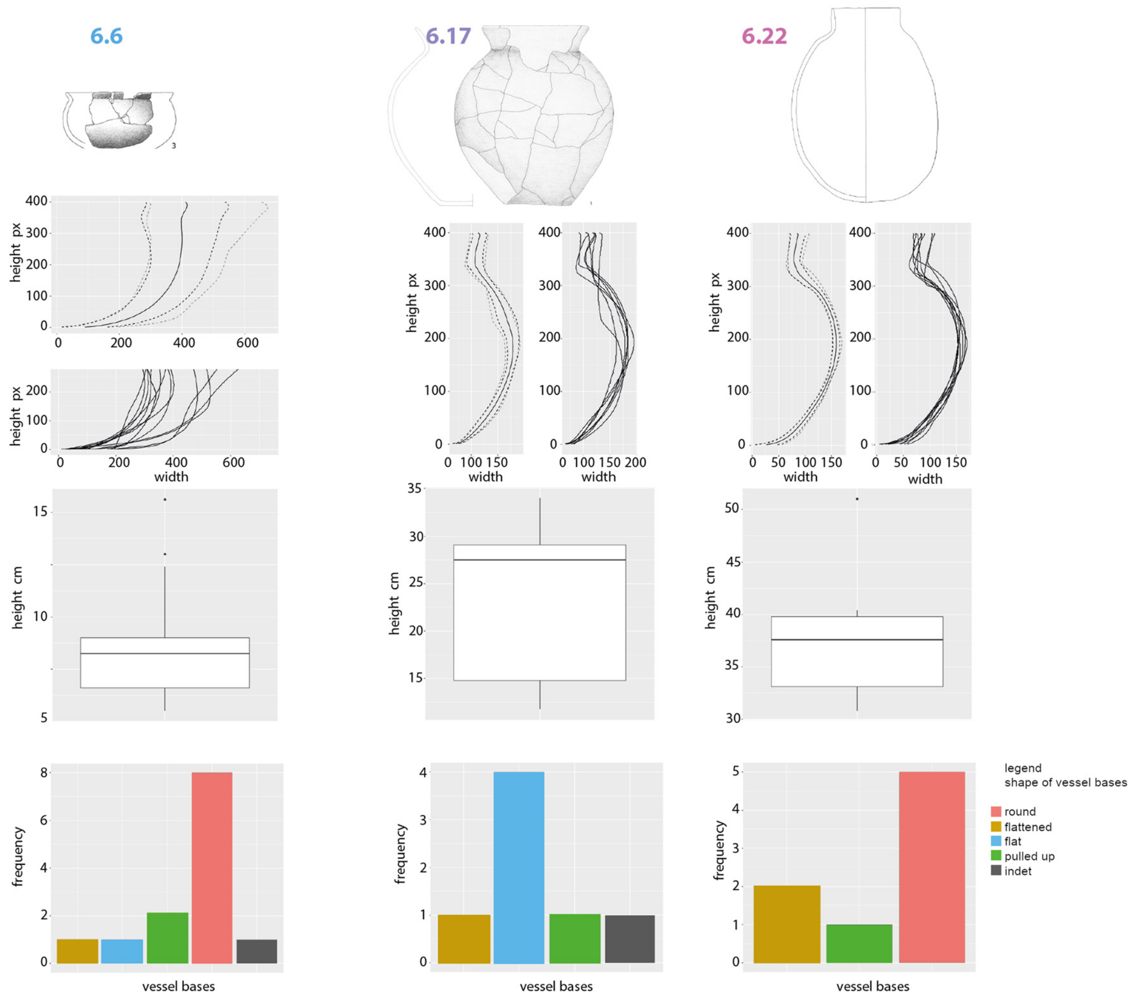


Figure 12: Subclusters 6.6, 6.17, and 6.22 (figure: M. Hinz and C. Heitz, for the picture credits of the pottery drawings, see doi: 10.5281/zenodo.7258694).

the vessels are considered suitable and useful for and the different actions in which they can be integrated, in short: the practices of vessel use.

In this respect, an article by Wotzka (1997) is instructive, in which he examines the connections between metric ceramic features and the orally stated functions of vessels from ethnological investigations using multivariate statistics (main component and discriminant analyses). According to Gallay, Huysecom, and Honegger (1991), he distinguishes three different functional concepts – and thus emic categorization types:

1. Primary function: the intended use in the production of vessels by the producers.
2. Secondary function: This is to be understood as what the consumers indicate in the survey as the intended use of the vessels, essentially these are ethnotaxonomies.
3. Tertiary function: actually, observable numerous functions, i.e., the uses of the vessels in practice.

On the one hand, this shows that there is a fundamental relationship between the size, proportions and closeness of vessels and their functionality – or rather, their potential uses. In the sense of affordance, however, a whole spectrum of actual uses per basic form is to be expected. In short (Heitz, 2022, pp. 390–392): the basic forms of the vessels refer to groups of possible, similar utilizations. Therefore, we define basic, latent groups of use potentials by means of unsupervised classification. The clusters can be described as latent groups because they do not necessarily catch the eye from a typological point of view when looking at the vessel drawings. Furthermore, the groups of use potentials are also latent because the

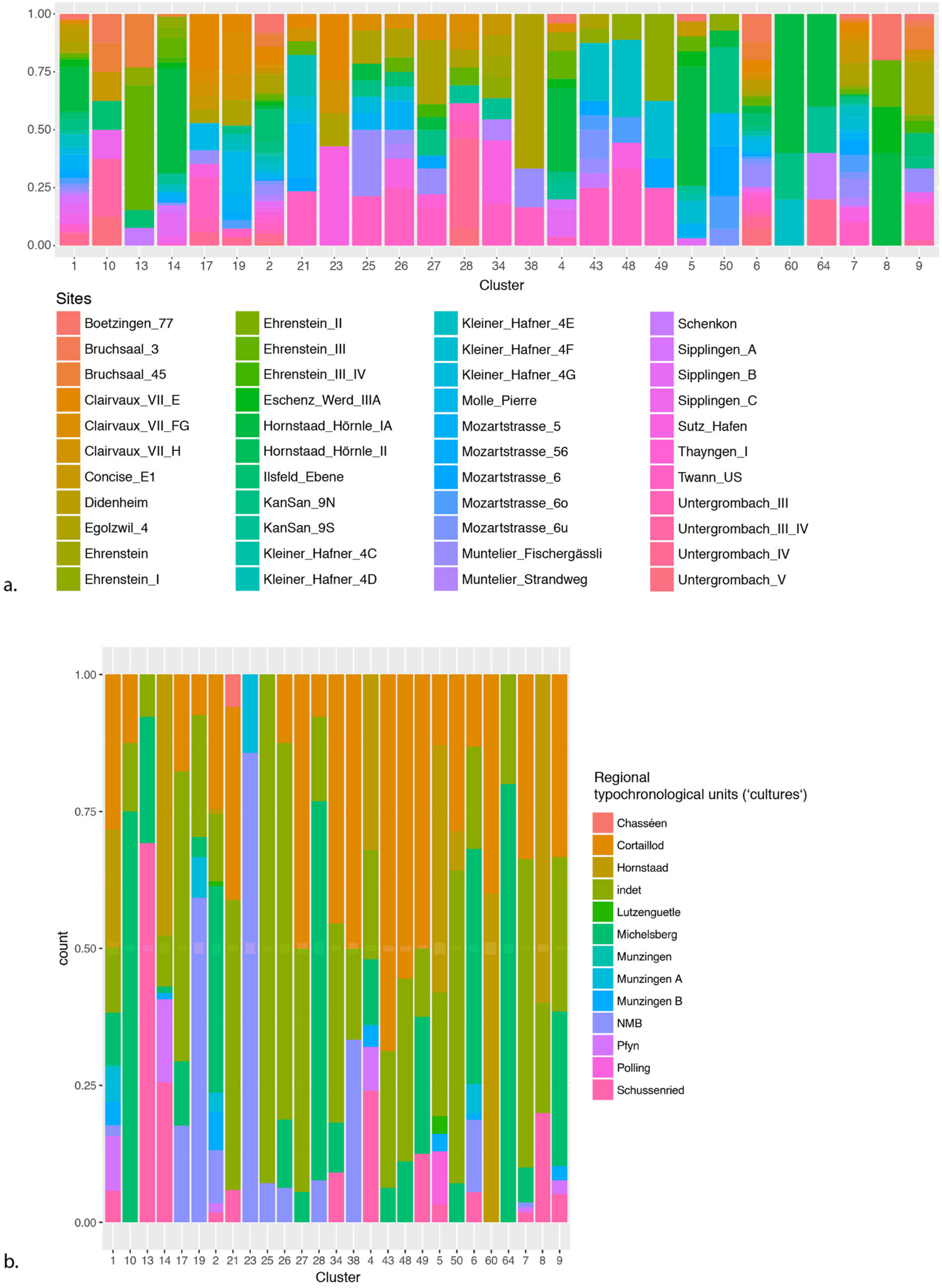


Figure 13: Result of the unsupervised classification regarding temporal and spatial attribution of the potter, (a) proportions (relative frequency) of sites (pottery assemblages of feature units) per cluster of the first hierarchy level and (b) of the typochronological ceramic groups per cluster on the first hierarchy level (figure: M. Hinz).

vessels elude any fixed functional attribution in the sense of affordance. The relationship between form and function is thus not stable and rigid but involves a certain unsteadiness that results from the dynamics of the mutual, situationally varying human-object relationships. The benefit of the computer-aided procedure in general lies furthermore in the reproducibility of its results, but also in the fact that it reduces subjectivity – to at least partly – for the classification result. This opens new perspectives on the finds, which complement those of conventional classification.

4.3 Regional Similarities and Differences as Indication for Spatial Mobility

As mentioned above, the unsupervised classification provides us with a hierarchical cluster solution, which allows to work with differentiations on several scales. From an archaeological point of view, not all clusters or subclusters are equally useful. As explained above, cluster 6 in itself is too heterogeneous, however, its subclusters, 6.1, 6.3, and 6.5 are more homogeneous and thus useful. Thus, it is necessary to validate the quantitative classification result critically to select the informative (sub)clusters for further archaeological analyses. Accordingly, to examine the distribution of latent groups of use potentials in settlements and time periods, one can select the most informative clusters or subclusters of the second or third hierarchical level, depending on the research question. The latter can be grouped visually into cluster groups, which has the advantage of yielding larger numbers of units per spatial or temporal unit for the comparison of functional spectra, as it was done for the MET project (Figure 14; Heitz, 2018, pp. 303–305). The results in Figures 15 and 16, which are described below, show that the analyses based on the combined cluster groups lead to the

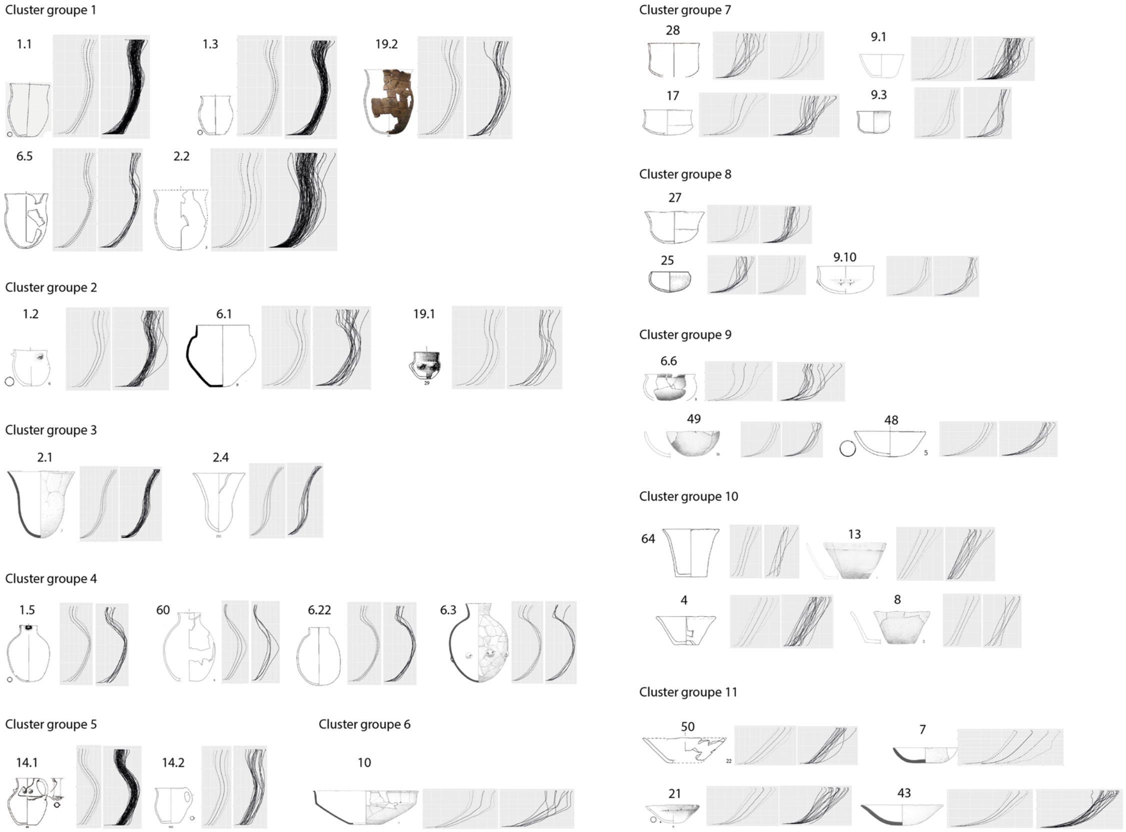


Figure 14: Choosen clusters and subclusters from different hierarchical level for further research question-dependent analyses within the MET project (after: C. Heitz, for the picture credits of the pottery drawings see doi: 10.5281/zenodo.7258694).

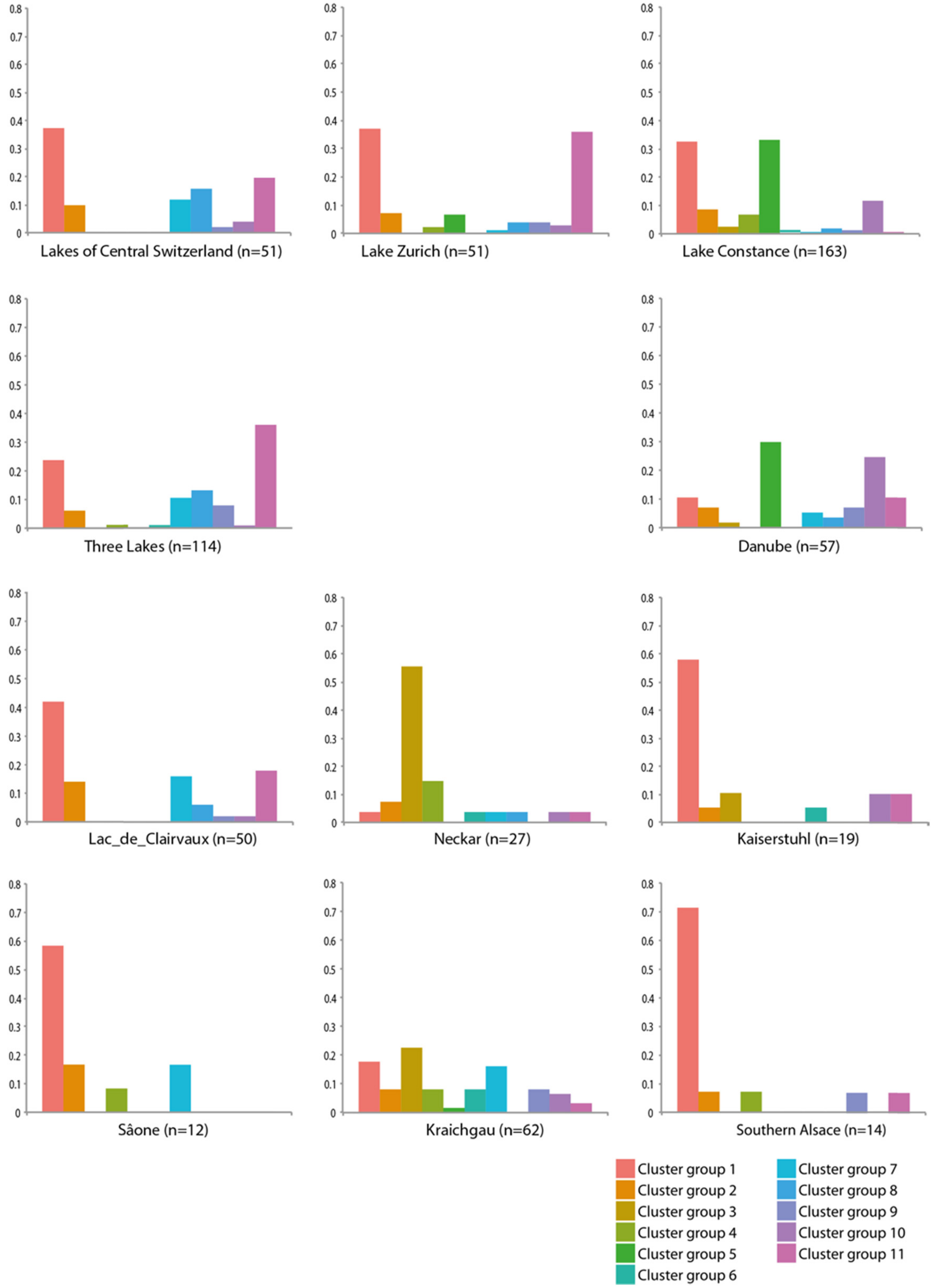


Figure 15: Relative frequencies of represented cluster groups per region (figure: C. Heitz).

Figure 16: Relative frequencies of represented clusters per region (figure: C. Heitz).

same results as the calculations based on individual clusters and subclusters. The only difference lies in the greater or lesser variability considered in each case.

Since the MET project aimed at a better understanding of spatial mobility between settlement communities based on material interconnections in pottery practices, the first question was to what extent similarities and differences exist between the small regions (Heitz, 2018, pp. 305–309). For this purpose, the spectra were compared based on the absolute or relative frequency (shares in %) of the formed cluster groups and the cultures as such per region (cf. Figure 3). When examining the spectra, we compared their basic patterns with each other, whereby the values of the frequencies were understood as tendencies depending on the number of cases.

If the spectra of cluster groups are grouped based on the relative frequencies, it becomes apparent that there are definitely patterns (Figure 15, see also Figure 3 for the spatial location of the regional groups). The effects of the different numbers of cases become particularly clear with this approach, whereby the spectra of the mineral soil sites are likely to be fundamentally thinned out. A quantitative investigation of the latter with respect to the relative frequencies of individual cluster groups in the spectrum is thus limited.

It is immediately apparent that each small region shows a slightly different spectrum in terms of the presence and frequency of cluster groups. Nevertheless, there are also clear similarities between individual small regions. In part, their spatial proximity to each other plays a role, but not exclusively.

Overall, the following lines of connection between the regions are revealed on a general level by the similarities and differences of the basic form spectra (Heitz, 2018, p. 310): the settlement communities living at the Lakes in Central Switzerland and in the Three Lakes region in western Switzerland used a more similar spectrum of basic vessel shapes as those Lac de Clairvaux or the Sâone region, respectively. This also applies to the settlement communities in the Lake Zurich area, although here there is a commonality with the pottery practice of the settlement communities at Lake Constance and in the Danube area, especially through the use of vessels with handles (jars, mugs). In the latter settlement areas, a similar spectrum of basic vessel shapes was used, whereby a connection to the areas around the Black Forest, the Kaiserstuhl, Kraichgau, and the Neckar area can be determined through individual groups of potential use such as cluster group 3. It is also clearly expressed that two spheres of different pottery practices and habitus groups overlapped in the area between Lake Constance and Lake Zurich during this period. This result corresponds to the insight from a cultural-historical perspective that there are fundamental differences between eastern Switzerland (Hornstaad/Pfyn) and central or western Switzerland (Cortaillod). The result is most satisfying, because it shows two things: with the statistical approach of unsupervised classification of “whole” pottery vessels elaborated here, quite reasonable, and straightforwardly interpretable results can be obtained in relation to settlement areas. On the general level of latent groups of use potentials mentioned so far, these do not contradict the basic findings of the conventional typological approach but confirm or rather refine and complement them.

Particularly meaningful are the spectra of clusters of small regions in which wetland settlements are present, which provide an optimal basis for comparison due to similar preservation conditions and numbers of cases. These include – listed according to their geographical position from west to east – the small regions Lac de Clairvaux, Three Lakes, Central Switzerland, Lake Zurich, Lake Constance, and Danube.

If the spectra of the clusters as such are plotted regarding their relative frequencies (Figure 16), the picture becomes even more clear (Heitz, 2018, pp. 311–312): within the fundamentally large similarities at the cluster group level when comparing adjacent regions, a closer look at the clusters within them reveals slight differences, which, however, speak overall for interconnections between the settlement communities in the small regions of Saône, Lac de Clairvaux, Trois-Lacs, and Lake Zurich. This shows that there was possibly the same need for certain basic functions of the vessels and that they were used in similar quantities – but possibly in slightly different functional variants or designs.

The heterogeneity or homogeneity in the clusters is interesting as far as the nominal variables are concerned (Heitz, 2018, pp. 319–321): shapes of the breaks, bottom shapes, decorations, and handles. However, this subjectively perceived, and to some extent statistically demonstrated, heterogeneity is much more likely to capture what actually takes place at the consumption level of ceramics from the actor’s perspective: the ever-changing affordances that result from the human-thing relationship between ceramic vessels and their users. Vessels can be used in a variety of ways. Their use is rarely reduced to one function, but rather materially similar vessels can be used similarly – or very differently. However, the fact that vessels were produced in certain similar basic shapes at all shows that the latter respond to a need for them on the part of the consumers.

It is significant that especially the clusters that have more than one distribution centre or are rather equally frequent in several regions show a greater inner variability. This can be exemplified by the vessels of cluster 1.1. Basically, these are tall, slender pot- and cup-shaped vessels with a fundamentally similar functional potential. In some of them, encrusted food remains are present, as we observed for example in vessels from Lake Zurich and Lake Constance. Such vessels have thus been used for cooking but are also suitable for storing food or other things. Thus, they occur in all small regions. The detailed profile or rim shapes are unlikely to have played a role in their use. This is probably more an expression of stylistic or practical aspects of production (e.g., ability). More decisive for the habitual use of these are handles, bottom shapes, or slip. These could refer to regionally different social practices of consumption. In their basic functional potential, however, the vessels are very similar.

Such regional differences can be found in almost all clusters. They point to the fact that vessels of the same potential functionality were produced in different designs in the individual small regions, with temporal differences also emerging within them. Thus, it should be emphasized once again that autoclassification provides a complementary perspective to the impressionistic design or style classification of ceramics.

5 Conclusion and Discussion

The use of computer-assisted classification in this study opened entirely new perspectives on regional and temporal dynamics in ceramic production and, above all, on the consumer perspective. We intentionally used an ordination method that considers the nature of stylistic transformation processes. For these, we must assume that changes take place at different speeds regarding different characteristics. At the same time, we must assume that different features occurred not only once but also several times in the shaping of ceramic objects. A method like t-SNE, which weights the local similarities more strongly than the global dissimilarities, can describe the complex processes of the emergence, mutual influence, and disappearance of certain ceramic styles better than it would be possible with simpler, common methods such as principal component analysis. In addition, the choice of the cluster algorithm, which included a distinction between the actual types and statistical noise, helped to reflect the nature of handmade ceramics better than such methods, which require each individual vessel to be assigned to a class.

Furthermore, by including nominal and metric variables, a variety of similarity structures and their entanglement could be considered. Such a result would not have been possible if we had limited ourselves to only one class of variables. Only in this way was it possible to incorporate archaeological expert knowledge about the significance of individual features in addition to the complex evaluation of profile shapes. As subjective as this expert knowledge may be, it contains an important level of information that would not be accessible to a purely metric evaluation. Thus, it was possible to consider aspects that could not be adequately described by metric variables (such as different forms of handles or bottoms).

Using a method of analysis that transcends classical typological schemata but intersects with the alternative qualitative classification by design elaborated in the MET project (Hafner et al., 2016; Heitz, 2017, 2018), it was also possible to uncover connections that would otherwise have been obscured by the predefined ‘conventional cultural boundaries’. Archaeological typological units are formed based on localities as regional-temporal constructs. In their description, they focus on similarities but also differences to group several sites into the same unit and distinguish them from other units. However, these systems of typological classification were made to establish relative chronologies long before scientific dating or computational approaches to classification were available. Hence, to keep the overview of centuries of ‘culture history’ and to be able to date the sites, they draw on the reduction of complexity in favour for handy culturally homogeneous spatial and temporal blocks in chronological schemes or blobs on culture maps. And as scientists, we have grown up and socialized with these classification systems that led to specific habits of looking at and understanding pottery. It is difficult for us to step back from this and to focus on other aspects that might indicate spatial mobility or stylistic appropriation as well as the potentials utilizations of pottery, in short: on pottery production, distribution, and consumption practices.

From this point of view, the spatial structures we have explored through quantitative classification are a level of affordance. They can be made visible through computer-assisted quantitative analysis. This sphere of understanding would hardly have been tangible without an overarching procedure that is decoupled from human decision-making processes in the first step. Such a completely new view can be contrasted excellently with classification schemes that have evolved over time and are certainly highly adapted for the respective original purposes.

Ultimately, however, it is the combination of qualitative and quantitative approaches that makes a new quality of archaeological investigation possible. Especially for the application of a model-based computer-assisted method, it is imperative to deal intensively with the theoretical foundations of data collection and evaluation. This makes it possible to reflect on the results, while at the same time operating in front of a blank slate and questioning traditional patterns of interpretation. In the synthesis of the quantitative and qualitative results, it still had to be considered that different statement qualities are possible with both approaches, the new quantitative and the new qualitative classification method, that we both had elaborated for the MET project while omitting classic culture-historical typologies. However, they complement each other excellently (Heitz, 2018, pp. 326–354).

The answering of questions in the humanities by means of computer-assisted methods falls into the area that is often referred to as digital humanities. As in the evaluation of text sources by means of the visualization of word frequencies or similarities and relationships in the statements, so in the case of ceramic analysis, too, the computer-assisted method is primarily a way of uncovering links and recognizing patterns beyond the capacity of the human brain. Here, as there, however, the determination of these patterns and their evaluation depends in turn on the skills and knowledge of the scientists investigating them. Just as the typical fingerprint of an author in the choice of words cannot be recognized and interpreted by the computer itself, the detection and analysis of similarity structures in pottery vessels as well as an understanding of their reasons and conditions are inevitably linked to the (inter)subjectivity of the research question and the investigating individuals or group of researchers. However, one of the great advantages of these methods is to externalize the pattern search from the mind of the investigator and thus make it obvious and discussable to the research community. In this sense, we hope that the method we have presented will meet with diverse criticism and but also with subsequent uses.

Acknowledgements: We are very grateful to the Archaeological services of the Cantons of Bern, Fribourg, Neuchâtel, Vaud, Lucerne, Solothurn, Zürich, Thurgau, Schaffhausen in Switzerland as well as the Landesamt für Denkmalpflege Baden-Württemberg im Regierungspräsidium Stuttgart, Dienststelle Hemmenhofen, in Germany for their collaboration. We are also grateful for two anonymous reviews and the helpful comments made there. Finally, we thank the MET project team and especially Prof. Dr. Albert Hafner of the Institute of Archaeological Sciences of the University of Bern for collaboration and support.

Funding information: The research for this article is part of the SNSF project "Mobilities, Entanglements and Transformations of Neolithic Wetland Sites on the Swiss Plateau (4000-3500 BC)" (Project No. 100011 156205) as well as the SNSF-Postdoc.Mobility-project "Time and Temporality in Archaeology. Approaching Rhythms and Reasons for Societal (Trans)formations in Prehistoric Central Europe (TimeArch)" (Project No. 194326) and was supported by the Swiss National Science Foundation.

Author contributions: Martin Hinz: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Visualization; Caroline Heitz: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing – Original Draft, Visualization.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: The R code used for the classification, together with information about the program and version used, can be accessed at https://github.com/MartinHinz/unsupervised.classification.swiss.pottery.2022 or archived at https://doi.org/10.5281/zenodo.6603137. The original dataset including metadata is accessible here: doi: 10.5281/zenodo.7258694.

References

Bourdieu, P. (2007). *Die feinen unterschiede: Kritik Der gesellschaftlichen urteilskraft. Suhrkamp Taschenbuch Wissenschaft*. Frankfurt am Main: Suhrkamp.

Bourdieu, P. (2009). *Entwurf einer theorie der praxis auf der ethnologischen grundlage der kabylischen gesellschaft. Suhrkamp Taschenbuch Wissenschaft*. Frankfurt am Main: Suhrkamp.

Bourdieu, P. (2013). Die Praxis der reflexiven Anthropologie: Einleitung zum Seminar an der ecole des hautes etudes en sciences sociales, Paris, Oktober 1987. In P. Bourdieu & J. D. W. Loic (Eds.), *Reflexive anthropologie*. Frankfurt am Main: Suhrkamp.

Bourdieu, P. (2014). *Sozialer Sinn: Kritik der theoretischen vernunft. Suhrkamp Taschenbuch Wissenschaft*. Frankfurt am Main: Suhrkamp.

Bourdieu, P., & Chartier, R. (1989). Gens à histoire, gens sans histoire: Dialogue Bourdieu/Chartier. *Politix*, *2*(6), 53–60

Bourdieu, P., & Wacquant L. J. D. (Eds.). (2013). *Reflexive anthropologie*. Frankfurt am Main: Suhrkamp.

Burri-Wyser, E. (2007). *La Céramique du Néolithique Moyen: Analyse spatiale et histoire des peuplements*. Cahiers d'archéologie Romande: De La Bibliothèque Historique Vaudoise. Lausanne: Cahiers D'archéologie Romande.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-Based clustering based on hierarchical density estimates. Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery in Databases, PAKDD 2013. *Lecture Notes in Computer Science 7819*. doi: 10.1007/978-3-642-37456-2_14.

Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data* (*TKDD*), *10*(1), 1–51.

De Capitani, A., Deschler-Erb, S., Leuzinger, U., Marti-Grädel, E., & Schibler, J. (2002). *Die Jungsteinzeitliche Seeufersiedlung Arbon, Bleiche 3: Funde*. Archäologie im Thurgau 11. Frauenfeld: Departement für Erziehung und Kultur des Kantons Thurgau, Amt für Archäologie.

Creswell, J. W., & Plano Clark, V. L. (2010). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE Publications Ltd.

Dietler, M., & Herbich, I. (1998). Habitus, techniques, styles: An integrated approach to the social understanding of material culture and boundaries. In M. T. Stark (Ed.), *The archaeology of social boundaries*. Washington, DC: Smithsonian Institution Press.

Ding, C., & He, X. (2004). K-Means clustering via principal component analysis. *Proceedings of the 21st International Conference on Machine Learning*.

Ester, M., Hans-Peter, K., Jörg, S., & Xiaowei, X. (1996). A Density-Based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining* (*KDD-96*).

Freudiger-Bonzon, J. (2005). *Archaeometrical study* (*Petrography, mineralogy and chemistry*) *of Neolithic ceramics from Arbon Bleiche 3* (*Canton of Thurgau, Switzerland*). Fribourg: Université, Dép. de Géosciences, Minéralogie et Pétrographie.

Fries, C. J. (2009). Bourdieu's reflexive sociology as a theoretical basis for mixed methods research: An application to complementary and alternative medicine. *Journal of Mixed Methods Research*, *3*(4), 326–348. doi: 10.1177/1558689809336660.

Furholt, M. (2008). Pottery, cultures, people? The European Baden material re-examined. *Antiquity*, *82*(317), 617–628. doi: 10.1017/S0003598X0009726X.

Gallay, A., Huysecom, E., & Honegger, M. (1991). *Enquêtes éthnoarchéologiques au Mali: Rapport des deux premières missions* (*1988–89, 1989–90*). *Document du Département d'Anthropologie et d'Ecologie de l'Université de Genève*. Genève: Département d'Anthropologie et d'Ecologie de l'Université de Genève.

Grenfell, M. (2014). Part I: Bourdieu and Data Analysis. In M. Grenfell & F. Lebaron (Eds.), *Bourdieu and data analysis: Methodological principles and practices*. Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang.

Grenfell, M., & Lebaron, F. (Eds.). (2014). *Bourdieu and data analysis: Methodological principals and practices*. Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang.

Guest, G. (2013). Describing mixed methods research: An alternative to typologies. *Journal of Mixed Methods Research*, *7*(2), 141–151. doi: 10.1177/1558689812461179.

Hafner, A. (2014). UNESCO World Heritage "Prehistoric Pile-Dwellings around the Alps": Chances and challenges for management and research of cultural heritage under water. In M. Roio (Ed.), *The Changing coastal and maritime culture. The 5th Baltic Sea Region Cultural Heritage Forum, Tallin 18.–20. September 2013* (pp. 36–40). Tallin: Cultural Heritage Forum Tallin. doi: 10.7892/boris.58972.

Hafner, A., Heitz, C., & Stapfer, R. (2016). *Mobilities, entanglements, transformations. Outline of a research project on Pottery practices in Neolithic Wetland sites of the Swiss Plateau*. Bern Working Papers on Prehistoric Archaeology/Berner Arbeitspapiere Zur Prähistorischen Archäologie 1. Bern: Institut für Archäologische Wissenschaften.

Hafner, A., & Suter, P. (2003). Das Neolithikum in der Schweiz. *Journal of Neolithic Archaeology*, *5*, 1–75. https://www.jungsteinsite.uni-kiel.de/pdf/2003_hafnersuter_text.pdf.

Hahn, H. P. (2005). Stil und Lebensstil als Konzeptualisierungen der Bedeutungen Materieller Kultur. In T. L. Kienlin (Ed.), *Die Dinge als Zeichen: Kulturelles Wissen und materielle Kultur*. Bonn: Universitätsforschungen Zur Prähistorischen Archäologie.

Heitz, C. (2017). Making things, being mobile. Pottery as intertwined histories of humans and materials. In C. Heitz & R. Stapfer (Eds.), *Mobility and Pottery production. Archaeological & anthropological perspectives* (pp. 257–291). Leiden: Sidestone Press.

Heitz, C. (2018). *'Keramik Jenseits von, Kulturen'. Ein Praxeologischer Zugang zu Mobilität, Verflechtungen und Transformationen im Nördlichen Alpenvorland (3950–3800 v.Chr.)'*. Bern: Universität Bern.

Heitz, C. (2022). *Keramik Jenseits von, Kulturen. Mobilität, Verflechtungen und Transformationen im Nördlichen Alpenvorland (3950–3800 v.Chr.)*. Leiden: Sidestone.

Heitz, C., & Sapfer, R. (2016). Fremde Keramik = Fremde Menschen? Mobilität and Beziehungsnetzwerke. In Archäologisches Landesmuseum Baden-Württemberg and Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart (Eds.), *4.000 Jahre Pfahlbauten: Begleitband Zur Grossen Landesausstellung Baden-Württemberg 2016*. Ostfildern: Jan Thorbecke.

Heitz, C., & Stapfer, R. (2021). Itineraries of Pottery. A mixed methods approach to mobility. In C. D. Gibson, C. J. Frieman, & K. Cleary (Eds.), *Making journeys. Archaeologies of mobility* (pp. 107–119). Oxford: Oxbow Books.

Heitz, C., Stapfer, R., & Martin, H. (2022). *The MET-Pottery-dataset*. doi: 10.5281/zenodo.7258694.

Husson, F., Josse, J., Le, S., & Mazet, J. (2017). *FactoMineR: Multivariate exploratory data analysis and data mining*. https://CRAN.R-project.org/package = FactoMine.

Hutson, S. R., & Markens, R. (2002). Rethinking Emic Pottery Classification. *Kroeber Anthropological Society Papers*, *88*, 8–27.

Ingold, T. (2013). *Making: Anthropology, archaeology, art and architecture*. London, New York: Routledge.

Jammet-Reynal, L. (2006). *La Céramique de Clairvaux VII (Jura, France) Dans Son Contexte Regional: Typologie, Étude Quantitative et Seriation*. These. Genf: Université de Genève.

Jammet-Reynal, L. (2017). The munzingen culture in the southern upper rhine plain (3950–3600). In C. Heitz, & R. Stapfer. *Mobility and Pottery production. Archaeological and anthropological perspectives*. Leiden: Sidestone.

Karasik, A., & Smilansky, U. (2011). Computerized morphological classification of ceramics. *Journal of Archaeological Science*, *38*(10), 2644–2657. doi: 10.1016/j.jas.2011.05.023.

van der Maaten, L., Lange, G., & Boon, P. (2010). Visualization and automatic typology construction of Pottery profiles. In B. Frischer, J. W. Crawford, & K. David (Eds.), *Making History Interactive: Computer Applications and Quantitative Methods in Archaeology (CAA): Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26, 2009*, Computer applications and quantitative methods in archaeology (Organization), Colonial Williamsburg Foundation, and University of Virginia, 1–12. BAR International Series 2079. Oxford: Archaeopress.

Lăzărescu, V. A., & Mom, V. (2016). Pottery studies of the 4th-Century necropolis at Bârlad-Valea Seacă, Romania. In S. Campana, R. Scopigno, G. Carpentiero, & M. Cirillo. (Eds.), *CAA2015. Keep The Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Archaeopress Archaeology*. Oxford: Archaeopress Publishing Ltd.

Lebaron, F., & Bonnet, P. (2014). Introduction Fo part III. In M. Grenfell & F. Lebaron (Eds.), *Broudieu and data analysis: Methodological principals and practices*. Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang.

Lemonnier, P. (1976). La description des chaînes opératoires: Contribution à l'analyse des systèmes techniques. *Techniques & Culture*, *1*, 100–151.

Lüning, J. (1968). Die Michelsberger Kultur. Ihre Funde in räumlicher and zeitlicher Gliederung. *Bericht Der Römisch-Germanischen Kommission*, *48*, 1–350.

Matuschik, I. (2010). Michelsberg am Bodensee. In C. Lichter (Ed.), *Jungsteinzeit im Umbruch. Die 'Michelsberger Kultur' und Mitteleuropa Vor 6000 Jahren. Katalog zur Ausstellung im Badischen Landesmuseum Schloss Karlsruhe 20.11.2010–15.5.2011*. Darmstadt: Primus Verlag.

Matuschik, I. (2011). *Die Keramikfunde von Hornstaad-Hörnle I-VI. Besiedlungsgeschichte der Fundstelle und Keramikentwicklung im beginnenden 4. Jahrtausend v. Chr. im Bodenseeraum. Forschungen Und Berichte Zur Vor- Und Frühgeschichte in Baden-Württemberg 122*. Siedlungsarchäologie im Alpenvorland 12. Stuttgart: Konrad Theiss Verlag.

Matuschik, I., & Müller, A. (2016). Ein "Hotspot" am Bodensee: 16 x Sipplingen-"Osthafen". In A. L. Baden-Württemberg, & Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart. *4.000 Jahre Pfahlbauten: Begleitband zur Grossen Landesausstellung Baden-Württemberg 2016*. Ostfildern: Jan Thorbecke.

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*.

Mom, V. (2007). SECANTO, the section analysis tool. In A. Figueiredo & G. Velho (Eds.), *The World is in Your Eyes. Proceedings of the XXXIII Computer Applications and Quantitative Methods in Archaeology Conference*. Tomar: CAAPortugal.

Nakoinz, O., & Hinz, M. (2015). Modelle in der Archäologie. In B. Thalheim, & I. Nissen (Eds.), *Wissenschaft und Kunst der Modellierung. Kieler Zugang zur Definition, Nutzung und Zukunft* (pp. 219–249). Philosophische Analyse 64. De Gruyter.

Petrequin, P., & Pétrequin, A. M. (2015). Une adaption au milieu amphibie: Successions atratigraphiques, techniques de fouille et architecture néolitique. In P. Pétrequin & A. M. Pétrequin (Eds.), *Clairvaux et Le 'Néolithique Moyen Bourguignon'*. Les Cahiers de La MSHE Ledoux Série Dynamiques Territoriales. Besançon; Gray: Presses Universitaires de Franche-Comté; Centre de Recherche Archéologique de la Vallée de l'Ain.

Read, D. W. (2016). *Artifact classification: A conceptual and methodological approach*. New York: Routledge.

Rice, P. M. (2015). *Pottery analysis: A sourcebook*. Chicago, IL: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/P/bo19986220.html.

Saev, I. (2015). *Gefäßkeramik – Funktion, Form, Habitus. Universitätsforschungen zur Prähistorischen Archäologie*. Bonn: Habelt, R.

Schlichtherle, H. (1998). Was sucht Michelsberg in den Ufersiedlungen des Bodensees? In J. Biel, H. Schlichtherle, M. Strobel, & A. Zeeb (Eds.), *Die Michelsberger Kultur und ihre Randgebiete – Probleme der Entstehung, Chronologie und des Siedlungswesens*. Materialhefte zur Archäologie in Baden-Württemberg. Stuttgart: Konrad Theiss Verlag.

Schweitzer, J. (1987). Le site Michelsberg de Didenheim. *Cahier de l'Association Pour La Promotion de La Recherche Archéologique En Alsace*, *3*, 50–87.

Seidel, U. (2017). From Typo-Chronology to Inter- and Intra Site Variety. The "Michelsberg" Pottery of South Germany (4300–3600 BC). In C. Heitz & R. Stapfer (Eds.), *Mobility and Pottery production. Archaeological and anthropological perspectives*. Leiden: Sidestone.

Sommer Harrits, G. (2011). More than method? A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, *5*(2), 150–166.

Stapfer, R. (2019). *Souvenir, Import, Reiseproviant oder Imitation? Kontaktnetze und Mobilität neolithischer Gesellschaften im West- and Zentralschweizerischen Mittelland zwischen 3900 und 3500 v. Chr.* (Unpublished doctoral thesis). Universität Bern, Bern.

Stapfer, R., Heitz, C., Hinz, M., & Hafner, A. (2019). Interdisciplinary Examinations Carried out on Heterogeneous Coarse Ceramics from Neolithic Lakeside Settlements in the Northern Alpine Foreland (3900–3500 BCE): Analysis Strategy and Preliminary Results from a Test Series Using PXRF. *Journal of Archaeological Science: Reports*, *25*(June), 217–238. doi: 10.1016/j.jasrep.2019.03.018.

Strehl, A., & Ghosh, J. (2002). Cluster ensembles: A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, *3*, 583–617.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences*. Thousand Oaks: Sage.

Van Der Maaten, L. (2014). Accelerating T-SNE using tree-based algorithms. *Journal of Machine Learning Research*, *15*(1), 3221–3245.

Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using T-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Van Der Maaten, L., Lange, G., & Boon, P. (2009). Visualization and automatic typology construction of Pottery profiles. In B. Frischer, J. W. Crawford, & D. Koller (Eds.), Computer Applications and Quantitative Methods in Archaeology (Organization), Colonial Williamsburg Foundation, Digital History Center, University of Virginia, and Virtual World Heritage Laboratory. *Making History Interactive: Computer Applications and Quantitative Methods in Archaeology* (*CAA*)*: Proceedings of the 37th International Conference*. Oxford: Archaeopress Publishing Ltd.

Wacquant, L. J. D. (2013). Auf dem Wege zu einer Sozialpraxeologie. Strukturen und Logik der Soziologie Pierre Bourdieus. In P. Bourdieu & L. J. D. Wacquant (Eds.), *Reflexive anthropologie*. Frankfurt am Main: Suhrkamp.

Wenger, E. (1998). *Communities of practice. Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Wotzka, H. P. (1997). Keramikformen und -Funktionen. Wider die Systematische Trivialisierung kulturspezifischer Zusammenhänge. *Archäologische Informationen*, *20*(2), 269–299.