Identifying accurate artefact morphological ranges using optimal linear estimation: Method validation, case studies, and code

Alastair Key, Metin I Eren, Michelle R Bebber, Briggs Buchanan, Alfredo Cortell-Nicolau, Carmen Martín-Ramos, Paloma de la Peña, Cameron A Petrie, Tomos Proffitt, John Robb, et al.

Journal of Archaeological Science 162 (2024) 105921

Contents lists available at ScienceDirect

Journal of Archaeological Science

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

Identifying accurate artefact morphological ranges using optimal linear estimation: Method validation, case studies, and code

Alastair Key [a,*], Metin I. Eren [b,c], Michelle R. Bebber [b], Briggs Buchanan [d], Alfredo Cortell-Nicolau [a], Carmen Martín-Ramos [a], Paloma de la Peña [a,e,f], Cameron A. Petrie [a], Tomos Proffitt [g], John Robb [a], Konstantina-Eleni Michelaki [h], Ivan Jarić [i,j]

[a] Department of Archaeology, University of Cambridge, Cambridge, CB2 3DZ, UK
[b] Department of Anthropology, Kent State University, Kent, OH, 44242, USA
[c] Department of Archaeology, Cleveland Museum of Natural History, Cleveland, OH, 44106, USA
[d] Department of Anthropology, University of Tulsa, Tulsa, OK, 74104, USA
[e] Departamento de Prehistoria y Arqueología, Universidad de Granada, Campus Universitario de Cartuja s/n, 18071, Granada, Spain
[f] Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa
[g] Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArEHB), Universidade do Algarve, Faro, Portugal
[h] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, 85281, USA
[i] Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Gif-sur-Yvette, 91190, France
[j] Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic

ARTICLE INFO

Keywords:
Morphometrics
Incomplete record
Fragmentary data
Archaeological discovery
Lithics
Ceramics
Metals

ABSTRACT

A fundamental goal of archaeologists is to infer the behaviour of past humans from the attributes of the artefacts they left behind. The archaeological record is, however, fragmented and often provides a partial record of the total artefacts produced by a given population. In turn, there is potential for population-level morphometric data, and therefore behavioural inferences, to be biased relative to the trends realised in the past. This includes morphological range data which are important for identifying similarities and differences between artefact groups, and for contextualising artefacts relative to external variables such as human anatomy, ecology, climate and chronology. Here, we investigate whether optimal linear estimation (OLE) modelling can be used to accurately identify the upper and lower limits of artefact morphological ranges (including those represented by sparse datasets). First, we test whether OLE reliably identifies morphological ranges using randomly sampled subsets of 'known and complete' replica artefact assemblages. Using morphometric data from lithic, ceramic, and metal archaeological case studies, we then identify how much further the upper and lower form limits of these artefact types would have been in the past, relative to the ranges evidenced by excavated (i.e., known partial) records. Validation tests demonstrate OLE to be capable of providing broadly accurate estimates for the true morphological range of artefact assemblages. Estimate accuracy increases relative to the percentage of the total assemblage used and the method is shown to function well using as few as five records (k) from an assemblage. The case studies reveal how OLE can overhaul or reinforce our understanding of artefact morphological ranges. In some instances, it is clear that the archaeological record provides a highly accurate representation of artefact morphological ranges and the overlap between artefact groups. For others, it is demonstrated that our understanding of the extreme artefact forms produced by past people is likely inaccurate

1. Introduction

A fundamental goal of archaeologists is to infer the behaviour of past populations using assemblages of artefacts (human, hominin or primate modified cultural objects). Sometimes individual artefact 'types' are used to create behavioural inferences, sometimes groups of different artefact types are combined. These assemblages are, however, rarely complete and often represent a fragmented dataset relative to the true range of artefactual (cultural) information present at the time of production (Dunnell and Dancey, 1983; Lucas, 2014; Wylie, 2017;

* Corresponding author.
E-mail address: ak2389@cam.ac.uk (A. Key).

https://doi.org/10.1016/j.jas.2023.105921
Received 12 June 2023; Received in revised form 14 December 2023; Accepted 15 December 2023

Pétursdóttir and Olsen, 2018; Huggett, 2020). Although there may be exceptions (e.g. intact artefact caches, sunken transport vessels), it is rare for archaeologists to be confident that an artefact assemblage represents the full range of morphological, raw material, or technological variation created by past individuals or populations. The earlier the archaeological period, the more fragmented and sparse artefact datasets typically are.

This problem is acutely felt in morphometric studies which use size and shape data to identify behavioural differences and/or similarities between groups of artefacts (and, in turn, human populations) (Lycett, 2009; Kovarovic et al., 2011; Okumura and Araujo, 2019). For example, how can an archaeologist be confident of limited shape-space overlap between two artefact assemblages when they are knowingly using partial datasets? Alternatively, can external variables be accurately linked to tool-form range changes if the extremes of these ranges are not accurately represented in excavated assemblages? The solution is often to use as large a dataset as possible. The more artefacts investigated then the greater the proportion of the total cultural variation represented (i.e., the 'true' variation present at the time of production), and in turn, the more secure any morphometric inferences are likely to be. This is the logic behind many archaeological morphometric studies.

While the behavioural inferences derived from such studies are important and will often be accurate, two unavoidable issues persist. First, no matter the size of the assemblage studied, or its proportional representation of the total excavated sample, it is usually impossible to be sure that past morphological variation did not extend beyond that represented in the collection(s) studied, or that central value tendencies were the same. Pooling multiple artefact assemblages together does not solve the issue, and arguably increases the possibility for unknown artefact forms – potentially as a function of the number of assemblages (populations) investigated – while simultaneously distorting individual assemblage form-distribution information. This is akin to problems that arise during summed distribution treatment of ^{14}C data (Ramsey, 2017). Indeed, trying to resolve the incompleteness of the archaeological record by increasing the volume of data analysed – or applying 'big data approaches' – "does not resolve these problems—if anything, they set them aside or risk covering them up" (Huggett, 2020: s13).

Secondly, archaeologists often only have access to very small or partial datasets, be it due to poor preservation rates, relatively few artefacts being produced in the past (and therefore being sparse within archaeological landscapes), or modern political, historical and economical factors limiting access to, or the discovery of, artefacts. This contributes to the aforementioned possibility of unknown artefact forms existing, but it also means that archaeologists frequently deal with highly fragmented datasets, complicating understanding of differences and similarities between groupings or categories of artefacts, and giving incomplete pictures of artefact distribution ranges. These problems are not limited to studies of modern humans (*Homo sapiens*) (Foley, 1981; Turq et al., 2013; Dibble et al., 2017), nor to investigations of hominin material culture (Gowlett, 2015; Visalberghi et al., 2015; Proffitt et al., 2022, 2023).

Archaeologists acknowledge these issues by stating that greater size and/or shape variation may have been present in the past relative to an analysed dataset. In the study of ecological or cultural diversity, these issues have long been known and methods have been developed to identify or address them (e.g., Chao, 1984; Chao and Jost, 2012; Chao et al., 2014; Colwell, 2013; Colwell and Elsensohn 2014; Buchanan et al., 2017; Eren and Buchanan, 2022 and papers therein; Eren 2012; Eren et al., 2012, 2016; Leonard and Jones, 1989 and papers therein). To our knowledge, however, there are no methods routinely applied in archaeological research that reconstruct this missing morphometric information; that is, the upper and lower tail ends (limitations, boundaries) of a morphological trait's range that is not reflected in an excavated artefact sample. Although kernel density estimate curves can suggest there to be morphometric range extensions beyond known limits (e.g., Bretzke and Conard, 2012; Ramsey, 2017; Birch and Martinón-Torres, 2019). This is despite the statistical summarisation of fragmented datasets being common in archaeological dating research (see: Crema et al., 2017; Ramsey, 2017; Timpson et al., 2020). If it is possible to model (reconstruct) the true range of morphometric variation for an artefact type using only the partial, fragmented excavated datasets available, then the reliability of any past behavioural inferences would increase. Put another way, if the aforementioned shape-space overlap between two artefact assemblages remains similarly limited even after the 'missing portions' of each assemblage's shape range has been modelled, then it is more likely that the artefactually-derived morphometric differences accurately reflect behavioural differences.

Here, we repurpose optimal linear estimation (OLE) modelling, a statistical technique used to reconstruct origination and cessation dates for artefact phenomena (Key et al., 2021a), for a new methodological use: reconstructing the full morphological range of human, extinct hominin and non-human primate artefacts. After validating the method using experimentally replicated artefact assemblages, we present lithic, ceramic and metal archaeological case studies to investigate whether OLE could provide a more accurate understanding of artefact morphological ranges.

2. Methods

OLE modelling was first developed by Roberts and Solow (2003) to provide more accurate estimates for the extinction timings of faunal species relative to their last known sightings or fossil occurrences (e.g., Solow, 2005; Clements et al., 2013; Pimiento and Clements, 2014; Wang and Marshall, 2016; Zhang et al., 2020). Many of the model's data-structure assumptions are not, however, unique to the temporal presence of species' populations. In recent years, this has led to its use in identifying origination and cessation timings for archaeological phenomena using the partially known (i.e., discovered) records available to archaeologists (Key et al., 2021a; Bebber and Key, 2022; Djakovic et al., 2022; Vidal-Cordasco et al., 2022, 2023).

Faunal and artefactual records both use OLE because they deal with fragmentary and often sparse datasets characterised by density distributions with approximately Weibull-form upper and lower tail ends. Population densities through time are not, however, the only phenomena to display such distributions, and the tail ends of artefact morphological ranges can similarly be interpreted to display approximate Weibull-form distributions. That is, in a sample of artefacts there are often extreme morphological outliers, but as one moves incrementally towards the central values of the tool-form range, greater artefact densities will be observed (Fig. 1). Weibull distributions incorporate a diverse range of probability shapes, including those formed by skewed data. As OLE has no scale limitations and displays few underlying assumptions (Rivadeneira et al., 2009; Boakes et al., 2015; Key et al., 2021a), it could feasibly be used as a method to reconstruct the full morphological ranges of archaeological artefacts.

2.1. Optimal linear estimation method

The formulaic expression of OLE is widely available through open-access sources (e.g., Pimiento and Clements, 2014; Roberts et al., 2021; Key et al., 2021b). Fundamentally, the method uses the spacing (intervals) and distribution of data points on a linear scale to make predictions about the upper or lower limits of an investigated phenomenon on that scale. It does this by fitting a Weibull distribution curve to these data and the end-point of this distribution marking the estimated start or end of the investigated phenomenon (Fig. 1). That is, OLE is based on the joint distribution of the k largest or smallest data points for a given metric having the same approximate 'Weibull form', regardless of the characteristics of the full record. In the case of scaled morphometric measurements, the model effectively asks, given the distribution of recorded data, what is the probability another artefact would not have existed beyond a given size or shape, from which we can

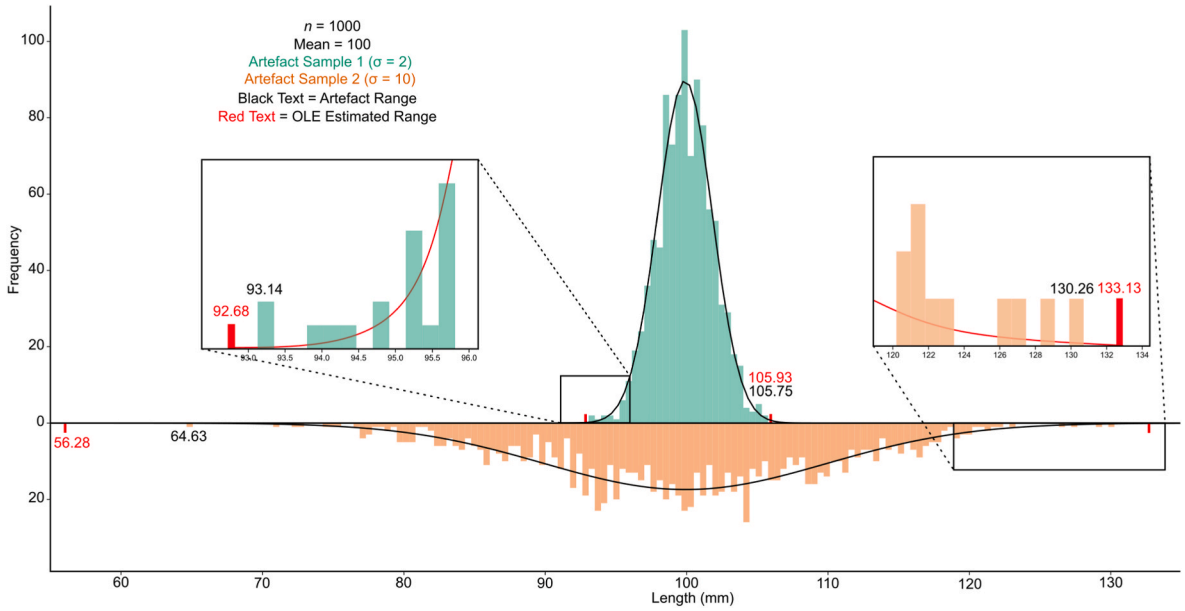

Fig. 1. Frequency distributions describing the length (mm) of artefacts in two hypothetical normally distributed artefact samples ($n = 1000$). Both samples display a mean length of 100 mm but differ in their standard distribution. When $\sigma = 2$, the distribution has relatively short tails (green). When $\sigma = 10$, the tails of the distribution are substantially longer (orange). OLE estimates for these hypothetical samples are depicted in red and demonstrate how the method can be used to predict the full morphological range of artefact samples despite the fragmentary nature of the archaeological record. The red lines in the insets illustrate how the morphological distribution of these samples present Weibull–form frequency curves. Note that these are large and relatively unfragmented datasets. Range estimations will likely be more substantial the more fragmented the artefact scenario. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

infer an upper or lower limit to a morphometric trait's range (for further information see, Solow, 2005; Rivadeneira et al., 2009; Clements et al., 2013; Boakes et al., 2015).

One advantage of the method is, therefore, that estimates are only impacted by the distribution of data at the upper or lower limits of the phenomenon's known (in this case artefactually recorded) range (Roberts and Solow, 2003; Solow, 2005). This is advantageous for cultural artefacts where diverse variables of influence interact in complex ways, potentially leading to bimodal or other distributions in central portions of an artefact's morphometric range. Moreover, the method requires relatively few data points, making it ideal for fragmented artefact assemblages with sparse datasets. This does not, however, mean the method cannot be used in data-rich scenarios. It is able to account for the amount of data available, and in the case of scenarios where morphometric distributions are densely populated, estimated end/start points would likely position close to known artefact forms. Note that while OLE can be applied to investigate diverse phenomena, it is archaeologists who define the artefact groupings used, and any specific phenomenon investigated, and in turn the research question, varies according to these definitions.

2.2. Model assumptions

Three important assumptions underpin OLE when applied in an archaeological morphometric scenario. First and foremost, the investigated metric's data – whether the upper or lower tail – needs to display an approximate Weibull-form distribution. This is a reasonable inference for most cultural artefacts governed by cultural evolutionary pressures, where specific object-forms are preferentially selected for, but variation is introduced to a sample by external factors including differing functional (incl. ecological) conditions, raw material factors, and drift mechanisms (Bebber, 2021; Bebber et al., 2019; Buchanan et al., 2014; Eren et al., 2015, 2022; Hamilton and Buchanan 2009; Lycett, 2011; Lycett and von Cramon-Taubadel 2015; Mesoudi 2011), or indeed, the pooling of artefacts from multiple populations via taphonomic or recovery/excavation processes (Bailey, 2008; Malinsky-Buller et al., 2011; Pardo-Gordó et al., 2018). Scales of variation, and in turn Weibull distributions, will be highly context dependent, particularly when considering industrially produced vs. hand-made artefacts. Further to this, OLE assumes that the artefact's morphological range likely spans beyond the current artefactually evidenced boundaries. As already mentioned, this is a safe assumption for most excavated artefact forms; particularly those that are not produced through industrial machining processes. Indeed, the 'hand crafted' nature of pre-industrial artefacts lends itself to the creation of morphological outliers, although this can of course occur during industrial fabrication too.

OLE also assumes all investigated data to be independent. This is a straightforward assumption for most morphometric data, where each artefact represents a distinct creative process. It is, again, industrially produced artefacts where this assumption may become complicated, as multiple artefacts could be produced through single events/processes (for example, consider candle making, where wicks may be dipped by hand, but many wicks may be lowered in a single action). There is also the assumption that no portions of a given morphological range are more or less likely to be present in artefact assemblages. That is, all artefacts on a given morphological range are assumed to have been subject to the same rates of discovery. This is more problematic for some artefact types compared to others. For example, artefacts with forms at or below sieve mesh sizes (e.g., beads, lithic microdebitage) will often display biases in favour of the discovery of larger artefact forms. In these instances, OLE could still be suitable, so long as all data entered into the model were subject to the same biases. For example, if the lower size boundary of a glass bead type would be modelled, and all lengths entered into the model were under the 10 mm mesh used on-site, then all data could have been subject to the same discovery rates.

2.3. Method validation and code

To better understand the ability of OLE to estimate the full range of a morphological trait, it was necessary to test the model within a context

where all artefactual units produced by a population are known – a feat that is impossible in almost all archaeological contexts. To circumnavigate this issue, we use modern assemblages of replica artefacts as this allows random subsamples from a much larger, known distribution of artefacts to be created. These subsamples are then used in an analogous way to the fragmentary portions of artefact records discovered by archaeologists. Importantly, irrespective of the artefact subsample 'discovered', the upper and lower limits of the full distribution of any given morphological trait will be known. It is also possible to confirm that the complete replica assemblage (sample) conforms to the assumptions of the OLE model (see above). Two replica stone tool assemblages produced by the authors are used here, but the results can be applied within any type of material context. Importantly, the replica assemblages were produced before the present study was conceived, and thus serve as a blind test.

2.3.1. Acheulean handaxe shape and size

Handaxes are one of two artefact types that characterise Acheulean assemblages, a period of Early Stone Age and Lower Palaeolithic stone tool production from c.1.8 to 0.2 million years ago (de la Torre, 2016; Wynn and Gowlett, 2018; Key et al., 2021b). 500 replica Acheulean handaxes were produced by Key and Lycett (2017) on British flint over the course of 18 months for use in a series of experimental functional studies (Fig. 2). The tools were designed to have forms that went "beyond those typically found in the archaeological record in order to push the ranges of variability" (Key and Lycett, 2017: 517). Subsequent investigations have demonstrated these replica tools to display greater length, thickness, breadth, elongation, relative thinness, and 3D shape variation than comparable individual handaxe artefact assemblages (Key, 2019; Key and Gowlett, 2022). Importantly, the replica assemblage displays similar data distributions (shapes) to those observed in handaxe artefact assemblages, albeit with greater skewness and kurtosis in some metrics (Supplementary Table 3 [Key and Gowlett, 2022]). Descriptive size and 3D shape (expressed as principal components where PC1 and PC2 are driven by length and thickness records, respectively) data for the full replica handaxe assemblage can be viewed in Table 1, while the distributions of key morphometric variables can be seen in Supplementary Figs. 1 and 2. The methods used to collect these morphometric data are described in Key and Lycett (2017).

2.3.2. Early Archaic point size

To understand how OLE may function when faced with a much smaller 'complete' assemblage of artefacts, we also investigated morphometric data from 45 replica Early Archaic points (Fig. 2). The Early Archaic period in North America occurred during the Early Holocene. These points, consistent with the Early Archaic side-notched "Thebes cluster" style (Justice 1987), have not been used in any other published studies, and were originally knapped by M.I.E. in 2021 with the intention of conducting a ballistics experiment involving their breakage. However, this experiment never came to fruition, and the specimens have since been curated in the Kent State University Experimental Archaeology Laboratory. Morphological data for the replica Early Archaic point assemblage can be viewed in Table 1, while the distributions of key morphometric variables can be seen in Supplementary Figs. 1 and 2. The metric data were recorded in centimetres from plan-view and profile-view digital images.

2.3.3. Method validation techniques and code

No morphometric studies have tested the robustness of OLE estimates under different data conditions, but previous temporal studies recommend entering c.10 data points into the model (Roberts and Solow, 2003; Solow, 2005; Rivadeneira et al., 2009). We follow this recommendation here, and use the 10 (k) largest or smallest measurements of a given trait, depending on whether the model is applied in the forward or reverse direction. In light of the sometimes considerable artefact samples available to archaeologists, we also assessed the use of k values of 5, 20 and 30 for all validation tests.

The replica handaxe assemblage was randomly subsampled at rates of 5%, 10%, 20%, and 50% relative to the total number of records in the assemblage. These randomly created subsamples were then entered into the OLE model, with only the k (5, 10, 20, or 30) largest or smallest values being used. As the archaic point sample was substantially smaller, it was not possible to run any analyses where $k = 30$. It was also necessary to increase the random sampling percentages for the archaic points such that they started at 20% and increased by 10% until they reached 50%, as sampling rates of 5% and 10% were not sufficient to create a sample large enough to use in the model. One thousand iterations of each process were performed for all subsampling percentages and k, and results were presented as median values across all iterations. OLE produces two sets of values relevant to understanding assemblage morphological ranges. The first is the morphological limit (T_E), on both sides of the known data range, inferred to be the most likely range endpoint according to the Weibull distribution curve fitted to the data. The second are the associated 95% confidence intervals (T_{CI}), beyond which artefact forms have a less than 5% chance of occurring.

Suitability of the OLE method to identify morphological range boundaries is first assessed by the proximity of the median T_E value across 10,000 iterations relative to the known largest or smallest values for a given morphological attribute. In addition, we apply the approach of Rivadeneira et al. (2009) which investigates how often confidence intervals include or span over the true end of a reconstructed range (i.e, in the case of simulated, complete ranges of morphological attributes, their true maximum and minimum values) across the 1000 iterations. A perfect coverage would be one that is equal to 1-alpha (with alpha of 0.05 used here). Coverage below 0.95 (0–0.94) indicates that OLE tends to underestimate the given morphological range limit, and that it may be more prone to Type I error. Coverage values above 0.95 (0.96–1.00) indicate that the OLE method is overly conservative and tends to overestimate the modelled morphological limit (Type II error). Importantly, values on opposing sides of perfect coverage (i.e., above and below 0.95) cannot be directly compared, as the range from 0.95 to 1.00 is more limited (condensed) than 0 to 0.95. All handaxe results created under the $k = 30$ scenario were excluded when the overall (combined) results were compared between the two replica assemblages, as these data were not available for the archaic points.

OLE does not work when records with equal values are included, and so all duplicates were automatically removed before the validation analyses (this was only relevant to mass, length, breadth and thickness) (also see duplicate discussion below). As OLE depends on the use of data from a distribution's 'long tail', and duplicates will converge on centrality, few data points relevant to the modelling were actually removed. It is also worth noting that the sampling probability was equally distributed across the dataset's range, and the model provides estimates for a single range endpoint. When sampling percentages are high, this is not likely an issue as the highest or lowest records entered into the model will have come from the relevant Weibull-like tail. However, when sampling percentages are low and k is high, there is the possibility that records from the distribution tail in opposition to the one being modelled (for example, a record from the lower tail if the upper one is being modelled) could have been used to create the OLE estimates. This would have decreased the accuracy of these estimates, but as OLE gives greater weight in the model to the highest or lowest records (respectively, for modelling higher or lower values), any impact is likely to be marginal. Further, any impact will make OLE appear less accurate in these validation tests relative to its application in an archaeological morphometric context. All analyses were undertaken in 'R' version 4.3.0. The script used to run all validation analyses is available in the supplementary information.

2.4. Artefact case studies

Ten case studies are investigated here to understand how OLE

Fig. 2. An example of each artefact type investigated here, including the two replica tool types. Replica handaxe (A), replica Archaic point (B), copper socketed tang point (C), Boxgrove handaxe (D), Old Park flake (E), Neolithic Stentinello pot sherds (F), Middle Historic bowl (G), Sibudu bipolar cores (H), Clovis point (I), Neolithic microliths (J), capuchin flake (K), and Olduvai cleaver (L). Each scale is in centimetres.

influences our understanding of the archaeological record under diverse assemblage conditions (Fig. 2). Case studies cover tools produced or used by Holocene humans (*Homo sapiens)*, Early-to-Late Pleistocene hominins (*H. erectus*, *H. heidelbergensis s.l., Homo sapiens*) and bearded capuchins (*Sapajus libidinosus*). The hominin case studies include Acheulean stone flakes from Old Park (UK, c.600 thousand years ago [ka]), Acheulean handaxes from Boxgrove (UK, c.500 ka), Acheulean cleavers from Olduvai Gorge Beds III-IV (Tanzania, c.1300-600 ka), Howiesons Poort bipolar cores from Sibudu Cave (South Africa, 63.8 ± 2.5 ka), Paleoindian Points from across North America

Table 1
Descriptive morphometric data for the two replica stone tool assemblages used in the validation analyses. Length, breadth, and thickness values are in millimetres (mm), while mass is in grams (g). Note that the mean values for PC1 and PC2 are not presented as by definition they approach zero with only very minor variation.

		Mean	SD	CV	Minimum	Maximum
Handaxe Size and Shape (n = 500)	Length	135.9	38.4	28.2	38.8	296.3
	Breadth	91.9	26.3	28.7	24.7	200.3
	Thickness	40.7	17.4	42.6	7.4	106.3
	Mass	576.7	559.0	96.9	8	4485
	Elongation	0.686	0.120	17.5	0.308	1.100
	Refinement*	0.439	0.131	29.7	0.188	1.091
	PC1	–	1.687	–	−5.041	6.862
	PC2	–	0.977	–	−2.591	4.550
Archaic Point Size (n = 45)	Length	57.7	10.1	17.6	37.0	83.8
	Shoulder Breadth	33.4	5.8	17.2	21.2	48.4
	Neck Breadth	19.7	4.4	22.4	11.2	33.1
	Basal Breadth	34.1	4.9	14.3	25.4	47.6
	Thickness	9.62	1.7	17.4	6.5	12.9

*Refinement is used here to describe the relative thickness of a tool (thickness divided by breadth).

Pakistan (600–1200 CE). We present diverse case studies to illustrate how OLE can be used to investigate artefact morphometrics across archaeological periods and material types. In addition, we investigate wild bearded capuchin (*Sapajus libidinosus*) flakes from Serra da Capivara National Park, Brazil, to demonstrate OLE's potential for better understanding non-human primate material culture and tool use. We were keen to include case studies with diverse sample sizes, to explore how estimates may be influenced by the fragmentedness of a dataset. In all cases, morphometric data were collected prior to the present study being conceived. Duplicates were removed for most, but not all, datasets (see below). The script used to run all case study tests is available in the supplementary information.

2.4.1. Lithic (hominin)

2.4.1.1. Lower Palaeolithic flakes. Old Park, also known by the moniker 'Fordwich',[1] is an MIS 15 to MIS 13 Palaeolithic site in southern Britain containing Lower Palaeolithic awls, scrapers, handaxes and flakes recovered from fluvial contexts in the Stour Valley, Canterbury, Kent (Key et al., 2022). Of interest here are the hundreds of flint flakes recovered in excavations since 2020. Although the total flake sample is now significantly higher, we limit the present case study to the whole flakes excavated from MIS 15 sediments in the 2020 season ($n = 173$).

Table 2
OLE estimates for all variables considered in the replica handaxe validation tests, alongside the 'true' upper and lower range limits represented in the complete assemblage. Results of all *k* values and sampling percentage conditions are presented. 'na' represents instances where results could not be returned as there were not enough records in the sample relative to those required to be entered into the model.

	Percentage of Total Assemblage Sampled (%)											
	Minimum Range Limit						Maximum Range Limit					
	k	5	10	20	50	True Value	*k*	5	10	20	50	True Value
Mass (g)	**5**	−0.2	−2.5	−9.9	−6.7	8	**5**	3153.3	4254.1	4693.1	5553.7	4485
	10	11.4	8.3	−3.8	−6.2		**10**	2821.9	3834.4	4521.5	5279.4	
	20	48.4	12.5	2.1	−5.9		**20**	2618.6	3589.6	4365.8	5265.9	
	30	na	23.1	3.9	−3.3		**30**	na	3436.8	4246.9	4905.1	
Length (mm)	**5**	52.8	40.0	27.8	27.5	38.8	**5**	267.3	279.7	301.9	328.3	296.3
	10	58.9	45.7	32.8	28.1		**10**	258.4	270.8	290.9	324.1	
	20	63.8	48.3	35.8	28.6		**20**	249.7	268.5	287.8	315.5	
	30	na	49.9	38.9	29.7		**30**	na	264.8	286.5	309.2	
Breadth (mm)		34.9	26.9	18.3	16.7	24.7	**5**	177.9	202.5	215.3	214.7	200.3
	10	38.7	30.5	21.0	16.9		**10**	170.9	192.9	210.0	213.9	
	20	41.9	33.6	22.9	17.6		**20**	165.2	189.8	205.6	214.0	
	30	na	34.6	24.9	18.1		**30**	na	187.9	203.3	212.9	
Thickness (mm)	**5**	9.2	7.8	6.5	4.3	7.4	**5**	99.9	105.1	106.5	118.0	106.3
	10	10.3	8.9	7.4	5.2		**10**	96.5	102.1	104.2	114.1	
	20	11.8	9.5	7.7	6.4		**20**	94.5	99.3	103.3	113.3	
	30	na	10.3	7.8	5.5		**30**	na	98.2	102.1	113.7	
Elongation	**5**	0.387	0.358	0.279	0.245	0.308	**5**	1.017	1.047	1.089	1.156	1.100
	10	0.401	0.377	0.322	0.249		**10**	0.993	1.03	1.059	1.184	
	20	0.416	0.385	0.346	0.254		**20**	0.975	1.021	1.052	1.167	
	30	na	0.393	0.357	0.263		**30**	na	1.015	1.048	1.152	
Refinement	**5**	0.194	0.189	0.175	0.173	0.188	**5**	0.817	0.924	1.201	1.259	1.091
	10	0.203	0.198	0.179	0.174		**10**	0.783	0.877	1.028	1.261	
	20	0.208	0.202	0.181	0.176		**20**	0.761	0.865	0.971	1.233	
	30	na	0.204	0.185	0.177		**30**	na	0.851	0.947	1.209	
PC1	**5**	−4.239	−4.629	−5.127	−5.928	−5.041	**5**	5.173	5.661	5.729	6.655	6.862
	10	−4.041	−4.320	−4.987	−5.747		**10**	4.863	5.440	5.599	8.455	
	20	−3.789	−4.215	−4.860	−5.537		**20**	4.573	5.232	5.561	8.144	
	30	na	−4.130	−4.767	−5.632		**30**	na	5.062	5.483	5.843	
PC2	**5**	−2.543	−2.651	−2.668	−2.885	−2.591	**5**	2.968	3.559	4.215	6.085	4.549
	10	−2.392	−2.584	−2.606	−2.783		**10**	2.693	3.291	3.899	5.782	
	20	−2.267	−2.551	−2.592	−2.709		**20**	2.507	3.146	3.734	5.544	
	30	na	−2.519	−2.580	−2.760		**30**	na	3.069	3.711	4.669	

(~13,500–10,000 calBP), geometric microliths from the Iberian Late Mesolithic (*ca* 8500-7500 calBP), Old Copper Culture socket tanged points from midwestern North America (6000-3000 BP), Neolithic Impressed Ware and Stentinello pots from Penitenzeria, Italy (7500 - 7000 BP), and Middle Historic ceramic bowls from Bannu Basin,

[1] The Old Park archaeological site was previously named after the village of Fordwich (e.g., Wymer, 1968), but there are several distinct Palaeolithic sites in the area and the locality is more clearly identified by the Old Park Site of Special Scientific Interest (SSSI) where it is located. In many places, the Palaeolithic bearing sediments are closer to Canterbury than Fordwich.

Table 3
OLE estimates for all variables considered in the replica archaic point validation tests, alongside the 'true' upper and lower range limits represented in the complete assemblage. Results of all *k* values and sampling percentage conditions are presented. 'na' represents instances where results could not be returned as there were not enough records in the sample relative to those required to be entered into the model.

	Percentage of Total Assemblage Sampled (%)											
	Minimum Range Limit						Maximum Range Limit					
	k	20	30	40	50	True Value	*k*	20	30	40	50	True Value
Length (cm)	**5**	3.392	3.35	3.264	3.187	3.699	**5**	7.860	7.825	7.833	9.946	8.384
	10	na	3.515	3.454	3.317		**10**	na	7.669	7.690	8.792	
	20	na	na	na	3.427		**20**	na	na	na	7.742	
Shoulder Breadth (cm)	**5**	2.066	2.045	1.995	1.709	2.121	**5**	4.740	4.899	5.034	5.525	4.835
	10	na	2.134	2.074	1.844		**10**	na	4.776	4.891	5.296	
	20	na	na	na	1.923		**20**	na	na	na	5.168	
Neck Breadth (cm)	**5**	1.120	1.128	1.091	0.875	1.117	**5**	2.949	3.022	3.072	3.366	3.310
	10	na	1.183	1.182	1.106		**10**	na	2.922	2.945	3.047	
	20	na	na	na	1.208		**20**	na	na	na	2.897	
Basal Breadth (cm)	**5**	2.335	2.334	2.407	2.441	2.539	**5**	4.488	4.698	4.921	5.544	4.762
	10	na	2.378	2.402	2.428		**10**	na	4.527	4.621	4.747	
	20	na	na	na	2.417		**20**	na	na	na	4.544	
Thickness (cm)	**5**	0.575	0.595	0.612	0.626	0.654	**5**	1.336	1.344	1.364	1.382	1.298
	10	na	0.597	0.607	0.617		**10**	na	1.313	1.323	1.355	
	20	na	na	na	0.619		**20**	na	na	na	1.333	

We do this to allow future contextualisation of the OLE estimates relative to the total sample of flakes excavated across a greater number of field seasons. Flake length (mm), width (mm), thickness (mm) and mass (g) are investigated and recorded using the methods in Key et al. (2022).

2.4.1.2. Acheulean handaxes. Boxgrove is a well known ca. 500 ka Acheulean occurrence from southern Britain that has been subject to diverse morphometric investigations (e.g., Roberts and Parfitt, 1999; Lycett and von Cramon Taubadel, 2015; Hosfield et al., 2018; García-Medrano et al., 2019). Roughly 500 flint handaxes were excavated from low energy fluvial contexts, with the tools subsequently displaying exceptional preservation (Roberts and Parfitt, 1999). Due to the lack of disturbance, the site is argued to display a series of highly temporally constrained episodes of behaviour by *H. heidelbergensis s.l.* (Roberts and Parfitt, 1999; Pope et al., 2020). Here we investigate a ~50% subset of the total handaxe assemblage from Boxgrove ($n = 254$ [$n = 214$ for the shape data]), and investigate their length (mm), width (mm), thickness (mm), mass (g), elongation index, refinement index and 2D shape. The methods used to collect these data can be viewed in Key and Lycett (2017).

2.4.1.3. Acheulean cleavers. Cleavers are, together with handaxes, one of the most representative morpho-technological tool types of the Acheulean techno-complex. They are defined as large cutting tools (LCTs), generally made on flake blanks, whose characteristic transversal distal edge (the cleaver bit or bisal) results from the intersection of two unretouched planes (Tixier, 1957). The cleaver sample analysed here (n = 134) is part of a larger LCT assemblage from Olduvai Gorge Beds III and IV (ca.1300-600 ka) excavated by Mary Leakey and colleagues between the 1969 and 1971 (n = 562) (Leakey and Roe, 1994). Olduvai Gorge (northern Tanzania) is a UNESCO World Heritage Site and one of the most prominent paleoanthropological and archaeological localities in the world (Leakey and Leakey, 1964; Leakey, 1971). Nevertheless, despite a well-documented geostratigraphic sequence that ranges from the Lower Pleistocene up to the Holocene (Hay, 1994; Deino et al., 2021; Njau et al., 2021), at Olduvai most investigations have focused, until recently (Arroyo and de la Torre, 2020; Pante et al., 2020; Martín-Ramos, 2022), in the Oldowan/Acheulean transition (de la Torre and Mora, 2005; de la Torre et al., 2011; Diez-Martín et al., 2015; Proffitt and Martín-Ramos, 2019). Here we investigate cleaver length (mm), breadth (mm), thickness (mm), edge length (mm) and mass (g). The methods used to collect these data have been previously described in Martín-Ramos (2022).

2.4.1.4. Howiesons Poort bipolar cores. Sibudu Cave in South Africa is one of the most well known and heavily studied Howiesons Poort sites (Wadley and Jacobs, 2006; Wadley, 2008; Jacobs and Roberts, 2008; de la Peña 2015a). A sample of small, quartz, bipolar cores are present at the site (in the Howiesons Poort and post-Howiesons Poort layers), with most excavated from the 'Grey Sand Layer' which dates to 63.8 ka (Jacobs and Roberts, 2008; de la Peña and Wadley, 2014). Bipolar cores represent an effective exploitation method that produces expedient, small flakes (de la Peña, 2015b). These flakes may be used in-hand or hafted. The technique is notable for its ability to produce flakes from small nodules or pebbles that may not be flaked through freehand or other techniques (Pargeter and de la Peña, 2017). The small size of the Sibudu cores (usually <15 mm in length), which largely stems from the small size of the local raw materials, makes them an interesting case study for the OLE method. Here we investigate the quartz bipolar cores from the Grey Sand layer at Sibudu cave, as described by de la Peña and Wadley (2014) and de la Peña (2015b). We use length (mm), breadth (mm), thickness (mm) and mass (g) data recorded from each core following the method outlined by de la Peña and Wadley (2014).

2.4.1.5. Paleoindian Points. To further explore the explanatory power of OLE within morphometric investigations, we wanted to see how the technique would perform when faced with multiple defined artefact typologies that overlap in their morphometric attributes. We chose early Paleoindian lithic points from North America (~13,500 - 10,000 calBP) as they display a series of defined lanceolate forms that overlap morphologically (e.g. Buchanan et al., 2018; Eren et al., 2022). Here, we focus on Clovis (incl. Eastern), Folsom and Midland points only. These four point types often provide a focal point for Palaeoindian morphometric discussions, in part due to their strong association with the peopling of North America during the Late Pleistocene (Eren and Buchanan 2016; Jennings and Smallwood 2019; Meltzer 2021). More detailed information on the spatial and temporal overlap of these point types can be found in Buchanan et al. (2019, 2022), Collard et al. (2010), and Jennings (2012, 2016). Data are derived from Buchanan and Hamilton's (2021) study of scaling laws in Paleoindian points, and the methods used to collect the length (mm), breadth (mm), thickness (mm) and mass (g) data, which we investigate here, are outlined in this study.

2.4.1.6. Mesolithic and Neolithic microliths. In a similar vein to the Paleoindian points, we also investigated Iberian Mesolithic geometric microliths as they display a series of defined typologies with strong distinctive morphometric features. Although geometric microliths can be found in European assemblages since *ca* 20.000 BP (Straus, 2002),

they acquire a specific limelight during the Holocene, and become one of the key cultural markers of the European Mesolithic and the transition to farming (Marchand and Perrin, 2017). European microlithic geometrism is a highly complex phenomenon, both spatially and chronologically. Here, we will focus only on its last Mesolithic phase (e.g., Castelnovian or Blade and Trapeze Complex). More specifically, we analysed a sample of the geometric microliths of Cueva de la Cocina (Cortell-Nicolau et al., 2023), one of the key sites of the Iberian Mesolithic. With 2177 geometric microliths, the site contains one of the largest assemblages of this type of arrowhead in Europe and, counting occupations from the Late Mesolithic until the Bronze Age, including the Early Neolithic, it is key for understanding the adoption of farming in the Western Mediterranean. Here, we only use geometric microliths from stratigraphic layers that can be confidently assigned to either phase A (ca. second half of the IX millennium BP) or phase B (ca. first half of the VIII millennium BP) in the regional Mesolithic sequence (see García-Puchol et al., 2023). We only use microliths that preserve at least 95% of their original size, according to the procedure described in Cortell-Nicolau (2019). The process of capturing the thickness of the geometric microliths was not automatised and, therefore, we did not include microliths lacking these measures. Finally, two outliers have been removed from the assemblage of phase B after the performance of exploratory principal component analyses (PCAs). All in all, this leaves us with a total of 344 geometrics for phase A and 390 for phase B.

2.4.1.6.1. Accounting for duplicate data. In addition to the OLE procedure run for all other case studies, we decided to run a series of additional analyses for the microlith sample that were also able to include duplicate records. These additional analyses added minor randomised variation to each record such that no two records would be the same. To do this, we resampled each value from a normal distribution, where we used the value itself as the mean, and a very small standard deviation (0.001). We repeated the OLE procedure with randomised values for 1000 iterations, and in our results we show the mean values obtained after the iterations. This ensures and corrects for potential not aplicable (na) values on the resulting output. This was repeated for the variables length (mm), breadth (mm), thickness (mm) and area (mm^2), but not for principal components PC1 and PC2, since these do not present duplicate records.

2.4.2. Lithic (non-human primates)

2.4.2.1. Flakes made by bearded capuchins. In recent years it has been identified that a number of stone tool-using primate species unintentionally and repeatedly detach sharp edged flakes from both stone hammers and anvils during various percussive tasks (Proffitt et al., 2016; Falótico et al., 2018; Luncz et al., 2022; Proffitt et al., 2023). Wild long-tailed macaques (*Macaca fascicularis)* in Phang Nga Bay, Thailand, are known to produce stone flakes as a by-product of cracking oil palm nuts on large embedded limestone anvils with hammerstones (Proffitt et al., 2023). Additionally, bearded capuchin monkeys from Serra da Capivara National Park in Brazil frequently produce conchoidal flakes as an unintentional by-product of stone on stone percussion (SoS) (Proffitt et al., 2016). This behaviour consists of striking quartzite cobbles embedded in conglomerate with a quartzite hammerstone in order to pulverise the surface of the embedded cobble to produce silica dust, which is then ingested (Mannu and Ottoni, 2009; Falótico and Ottoni, 2016). An analysis of the resulting lithic assemblage noted that flakes detached from the hammerstones during this process share many of the same technological and metric attributes to those produced by Plio-Pleistocene hominins (Proffitt et al., 2016). Due to the shared mechanisms between capuchin stone on stone (SoS) percussion and hominin flake production, where one stone is struck against another, a sample of capuchin SoS flakes (n = 31) were included in this study. Flake length (mm), breadth (mm), thickness (mm) and mass (g) were used in this study, following methods described by Proffitt et al. (2016).

2.4.3. Metal

2.4.3.1. Copper socketed tang points. The North American Old Copper Culture (OCC; ca. 6000-3000 BP) provides a unique perspective from which to analyse the initial use of copper metal as a tool making media. The OCC stands out in global material culture evolution in that here, forged copper tools, made from local copper sources, were manufactured by mobile hunter-gatherer groups for several millennia, and then abruptly ceased ca. 3000 BP. Despite early debate as to the origin of these ancient copper toolmakers (Bebber and Chao 2022; Martin 1999), it is now well established that Indigenous peoples, over an extensive area, produced a wide variety of copper tools including projectile points, knives, axes, awls, and fishing gear. These native copper tools represent some of the earliest examples in the world of metal being used as tool media with recent research projecting the origin of copper tool production in North America to ca. 9000 BP (Bebber and Key 2022). Native copper, which occurs in abundance throughout the western Great Lakes region, is a relatively soft metal that can be worked in its raw form with no need of smelting or casting. Research shows that members of the OCC used combined production methods of hot hammering, cold hammering, and annealing to manufacture their copper implements (LaRonge 2001; Vernon 1990). Given the malleability of the raw material, copper projectile points were manufactured in a wide variety of forms; the most iconic OCC style being the socketed tang point (Fig. 2). Here, we investigate the potential morphological range of socketed tang points using length (mm), breadth (mm), thickness (mm), and mass (g).

2.4.4. Ceramic

2.4.4.1. Middle historic ceramic bowls. Fired ceramic vessels are one of the most prevalent forms of archaeological artefacts, and throughout their history of use, there have been a range of important innovations in their production. For instance, the use of rotation of various speeds, including the use of tournettes and kick wheels have made it possible to mass-produce pottery that is relatively standardised (Blackman et al., 1993; Courty and Roux 1995; Roux 2019; Ceccarelli et al., 2020). During the Middle Historic period in ancient Pakistan (c.600–1200 CE) there is clear evidence for the mass production of simple bowls using rotation (Petrie, 2020). These vessels appear in somewhat standardised dimensions and show clear signs of mass production, including string cut bases, lack of attention to surface finish, and impromptu repairs to throwing mistakes (Petrie 2020). Here, we use a sample of these ceramic bowls from the Bannu Basin. We investigate the base diameter, rim diameter, rim thickness and height of the bowls, which were measured using diameter boards with 1 cm increments, callipers and a metal rule (see: Petrie, 2020). As the bowls are often fragmentary, sample sizes vary between each variable.

2.4.4.2. Neolithic Impressed Ware and Stentinello pots. A small dataset of Neolithic pottery was used to provide ceramic data of a different kind with varying data distributions. These vessels were excavated at the Neolithic site of Penitenzeria, Calabria, Southern Italy; they date to between 5500 and 5000 BCE (Robb, 2007). They represent small-scale production by a small, closely networked group of village pottery-makers. The vessels were hand-formed, principally by coiling. They were decorated by smoothing and by impressing and stamping patterned motifs in the unfired clay in the “Impressed Ware” and “Stentinello” styles, and they were fired in open pit fires. Most of the vessels are small hemispherical bowls; a few are medium or large serving or storage vessels. Although these potters were highly skilled, they were not specialists in the conventional economic sense of the term, and an individual potter may have made only a few vessels for their own use each year. In contrast to the Bannu Basin assemblage, this dataset affords data from an assemblage which was not mass-produced by specialists using mechanical aids. It is also a small assemblage of highly

fragmented vessels, and thus can show how OLE can improve our sense of an assemblage's particular characteristics when they can be difficult to estimate simply using the statistical distribution of the data. We investigate rim thickness (mm) and rim diameter (mm) following Robb (2007).

3. Results

3.1. Validation tests

The results of each OLE estimate (T_E) produced through the randomised subsampling procedures are presented in Tables 2 and 3.

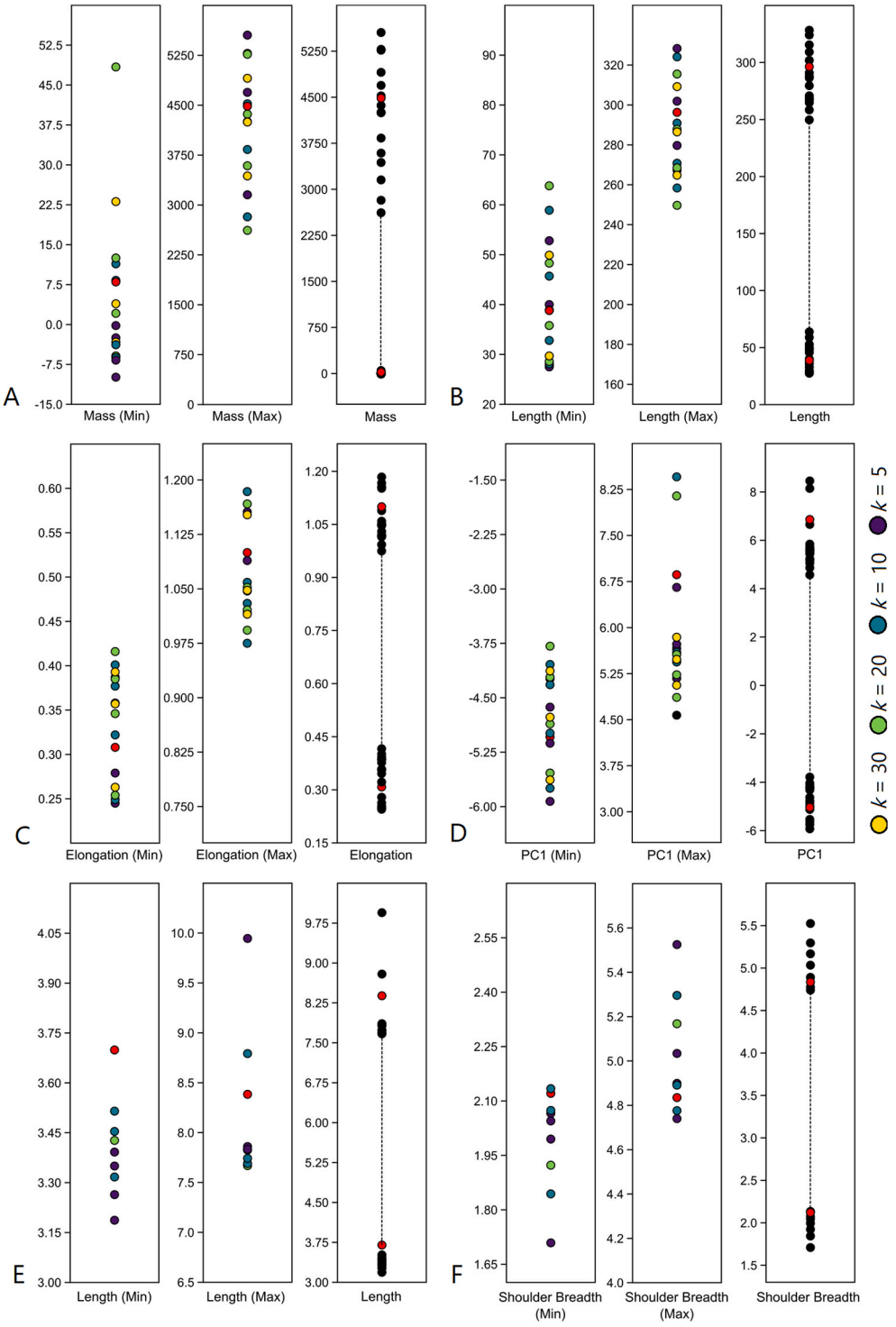

Fig. 3. OLE estimates (T_E) produced during the validation tests. Minimum, maximum and full range data are presented for all *k* values and sampling percentages. Figures A to D illustrate test data from the replica handaxe assemblage, while Figures E and F present data from the Early Archaic point tests. Red dots represent the 'true' upper and lower limits of each variable as observed in the complete replica tool assemblages. Purple, teal, green, and yellow dots represent OLE estimates produced using *k* values of 5, 10, 20, and 30, respectively. All data are in Tables 2 and 3 Mass is recorded in grams while length and breadth values are in millimetres. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Presented alongside these data are the 'true' upper and lower limits of each morphological attribute in the complete replica handaxe and archaic point assemblages. It is clear that varying *k* values and sub-sampling percentages impact the accuracy of the OLE estimates. Further, it is evident that for each variable there are median modelled results providing accurate estimates of the 'true' upper or lower range limits observed in the replica assemblages. When the results of all *k* values and sampling percentages are combined for a given variable, estimates can be considered broadly accurate across multiple model conditions (Tables 2 and 3; Fig. 3). Variation is present, but the 'true' value is almost always at the centre of the estimates (Fig. 3), and estimate deviation is low relative to the complete range of the variable considered. The only clear exception is maximum handaxe mass (Fig. 3A), where upper range estimates vary between 5554 and 2629 g.

Supplementary Tables 2 and 3, which report how often OLE confidence intervals (T_{CI}) include or span over the end of a reconstructed range (across 1000 iterations), illustrate how estimates vary depending on the *k* values and sampling percentage used. Fig. 4 demonstrates that for both tool types, estimates for the minimum range limit of variables display higher accuracy than models estimating maximum range limits. That is, minimum estimates appear less likely to over- or underestimate a given range limit. Fig. 5 breaks these values down according to each tool type, before being further separated according to the *k* values or sampling percentage used to provide data for each model. For the $n = 500$ replica handaxe assemblage, it is clear that for both minimum and maximum range estimates, model accuracy increases in-line with the percentage of data available from the complete assemblage (Fig. 5). Indeed, the median value under the 50% sampling condition is close to 0.95 (i.e., 1-α), while for the 5% sampling condition it is approximately 0.62, indicating that sample percentages below 50% of the total sample underestimate range limits. A sampling percentage of 20% still returned median values above 0.82, indicating broadly acceptable estimates. For the $n = 45$ replica archaic point assemblage, all sampling percentages returned median values ≥ 0.8, and there were no clear trends whereby greater sampling percentages equated to greater estimate accuracy.

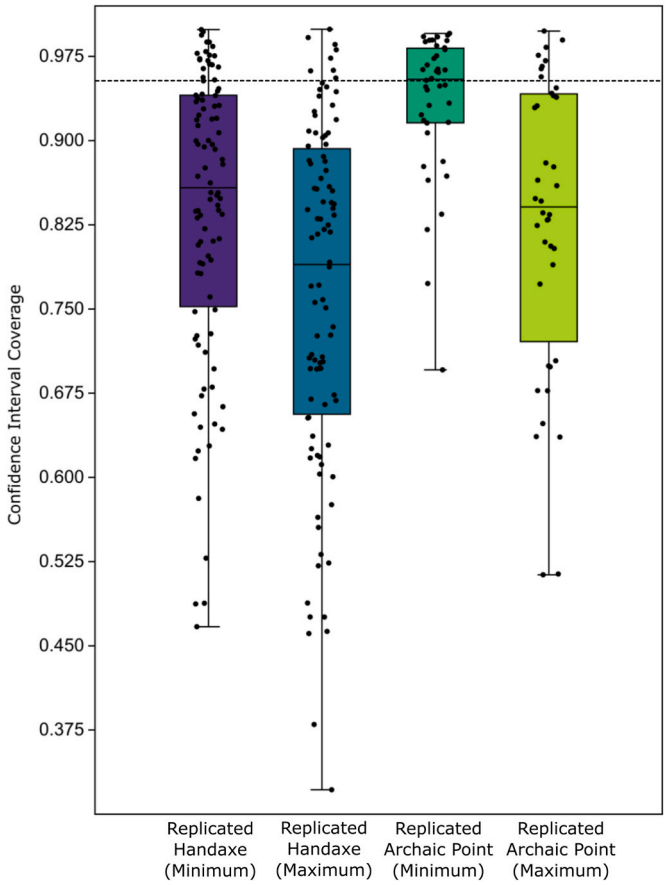

Fig. 4. Confidence interval coverage (how often confidence intervals include or span over the end of a reconstructed range) where perfect coverage is equal to 0.95 (dotted line), coverage below 0.95 indicates OLE is underestimating the morphological range limit, while values above 0.95 are overly conservative and OLE is overestimating range limits. Data from all analyses are combined and displayed according to tool type and minimum or maximum range limit (excluding the k = 30 handaxe results). Lines represent median values and boxes represent 25–75 percent quartiles.

The impact of varying *k* values is consistent across both replica assemblages and both the minimum and maximum range limit scenarios (Fig. 5). For all, it is clear that $k = 5$ is less likely to return confidence intervals that underestimate range limits, and for the minimum range of the archaic points it more often than not overestimates the value (Supplementary Tables 2 and 3; Fig. 5D). For *k* values of 10 through to 30, estimate accuracy declines in a sequential order, with OLE more often underestimating morphological range limits. Ranked median CI coverage using *k* values of 5 and 10 are presented in Table 4. Generally, *k* values of 5 more consistently provided coverage closer to 0.95, but the top ranked condition tended to overestimate true range endpoints and should not be considered the best performing. On a relative basis, such divergence from perfect coverage (e.g., 0.962 coverage) within a narrower range between perfect and maximum coverage (i.e., 0.95–1.00) is equivalent to a considerable underestimation. Consequently, the maximum archaic point estimate using $k = 10$ is likely the best performing scenario (i.e., closest to the perfect coverage [0.95]). Thus, a *k* value of 10 can still be considered to provide adequate estimates, particularly when applied to upper range limits, which here display relatively long distribution tails (Supplementary Tables 2 and 3; Fig. 5A and C). In sum, OLE can be considered to provide broadly accurate estimates for the upper and/or lower limits of artefact morphological ranges using fragmented datasets, but the most accurate results are returned when *k* is around 5 to 10 and record samples are as complete as possible.

3.2. Artefact case studies

In all instances artefact ranges were extended by the OLE estimates (Tables 5–7). In some instances additions were marginal, for others ranges were extended considerably. Extensions were largely on a relative basis following the density and distribution of artefact samples prior to the artefactually-evidenced range boundaries. Often, these densities reflected the sample size of the investigated assemblage. Smaller artefact samples, such as the copper socketed tang points, had relatively low densities at the tail end of some attributes – and therefore comparatively large morphological gaps between 'archaeologically observed artefact records' – meaning that estimated range extensions were sometimes substantial. For example, OLE increased the $n = 93$ copper socketed tang point breadth range by 62.9% (Table 5). Larger artefact samples generally resulted in more limited range extensions (Fig. 6B). For example, out of the four Paleoindian point types, the Clovis assemblage was by far the largest (n = 810), and experienced the smallest range expansion percentage across all investigated attributes (all were ≤12.2%; Table 6; Fig. 7). The Neolithic pot assemblages provide another example where range expansion percentage aligns well with sample size (Table 5). When results from multiple artefact types are pooled, and sample size and range extension are correlated, these relationships become apparent (Fig. 6).

These broad assemblage-size trends did not hold true in all instances (Fig. 6A). The substantial Boxgrove assemblage ($n = 254$), for example, estimated there to be considerably more elongated and less refined handaxes made by this *H. heidelbergensis s.l.* population, with range extensions of 36.2% and 38.3% being returned, respectively. Indeed, OLE extended the lower range limit of handaxe elongation by 0.158 (Table 5). Contrariwise, handaxe refinement was principally extended

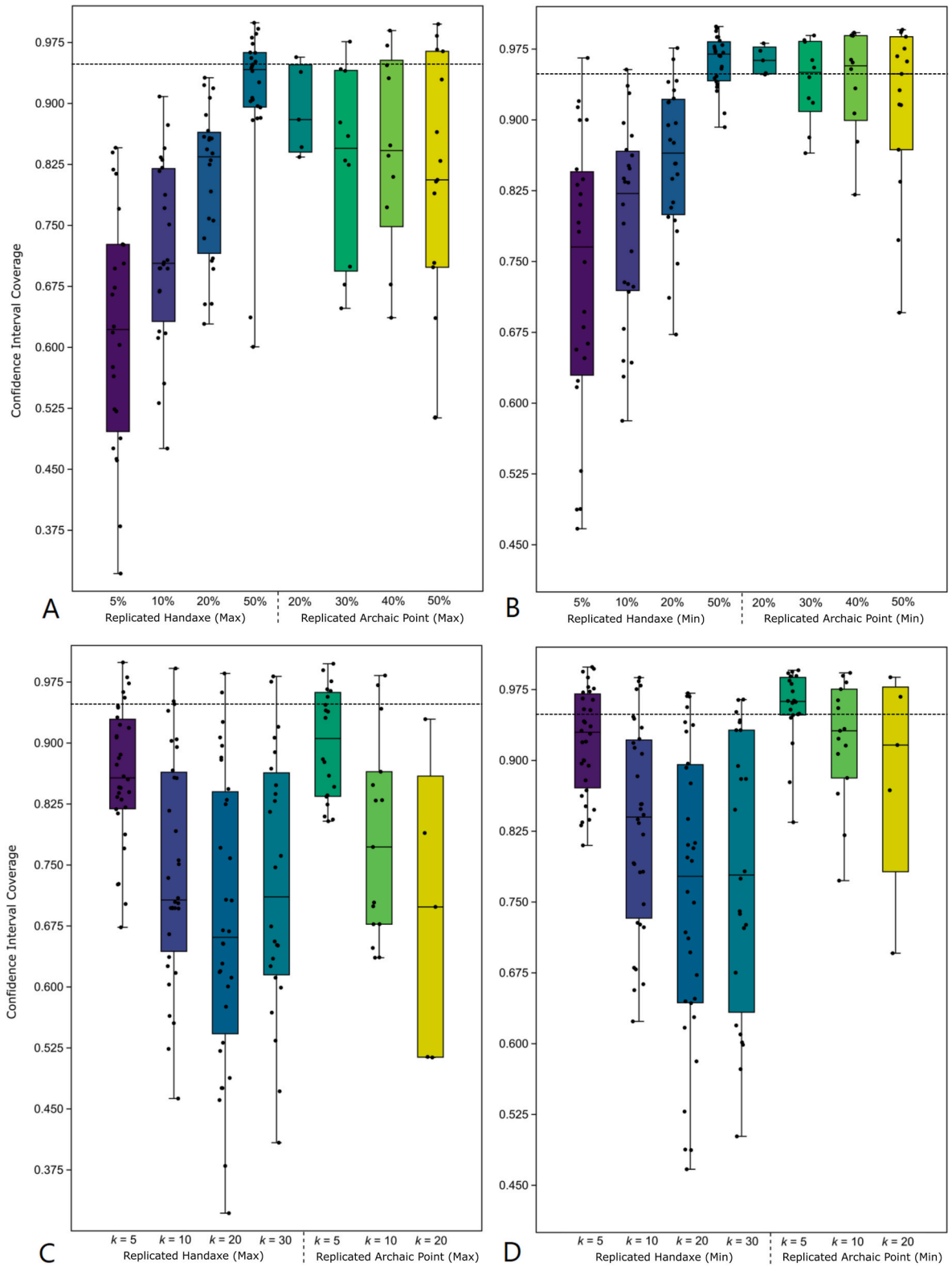

Fig. 5. Confidence interval coverage where perfect coverage is equal to 0.95 (dotted line), coverage below 0.95 indicates OLE is underestimating the morphological range limit, while values above 0.95 suggest OLE is overestimating range limits. Data from all analyses are separated according to their respective *k* values and sampling percentages. The $k = 30$ handaxe results are excluded from the sampling percentage comparisons. Lines represent median values and boxes represent 25–75 percent quartiles.

Table 4
Ranked median confidence interval coverage for all $k = 5$ and 10 value scenarios for the replica handaxe and archaic point validation tests.

Rank	Median CI Coverage	K value	Estimate Direction	Range Estimated
1	0.962	5	Overestimate	Archaic Point, Maximum
2	0.931	10	Underestimate	Archaic Point, Maximum
3	0.930	5	Underestimate	Handaxe, Maximum
4	0.906	5	Underestimate	Archaic Point, Minimum
5	0.857	5	Underestimate	Handaxe, Minimum
6	0.840	10	Underestimate	Handaxe, Maximum
7	0.772	10	Underestimate	Archaic Point, Minimum
8	0.707	10	Underestimate	Handaxe, Minimum

by OLE at its upper limit (0.227; Table 5). The rim diameter of the Middle Historic bowls, which display the largest sample in any case study ($n = 1466$), is another example of OLE creating a considerable range extension (40.7%) in spite of the substantial sample size (Table 5). Alternatively, the mass-range of the Midland Paleoindian type only increased by 5.1%, despite the assemblage only consisting of 64 artefacts (Table 6).

Occasionally, range extensions varied substantially within the same case study. Capuchin flake mass ($n = 31$), for example, was extended by 56%, while flake length, breadth and thickness extensions were more limited – albeit not low –- at 22.3, 21.2 and 26%, respectively (Table 5). The additional resampling procedures undertaken for the geometric microliths reveal the removal of duplicate records to have little-to-no impact on OLE morphometric estimates, and in turn, range extensions. For all variables, bar the minimum length of geometric microliths in phase A, results are identical across the 'duplicate removed' and 'duplicate randomised' model conditions (Table 7). Beyond the geometric microlith case study, duplicate data only restricted the use of OLE in the ceramic bowl and pot tests, where one example each of rim thickness, rim diameter and base diameter could not be returned (Table 5).

4. Discussion

Archaeologists are acutely aware that morphometric-based behavioural inferences drawn from fragmentary artefact samples have potential to be inaccurate. While many behavioural conclusions will be correctly determined using such datasets, these assemblages are nonetheless incomplete representations of past human behaviour. Previously, it has been difficult to accurately identify which artefact forms remain undiscovered and unknown. Here we investigated whether optimal linear estimation (OLE) modelling can be used to model the extreme range limitations of diverse morphological attributes in lithic, ceramic and metal artefacts. Validation tests using replica stone tool assemblages

Table 5
Range data for each morphological variable investigated for each artefact case study characterised by a single artefact type (i.e., excluding the Paleoindian point and Mesolithic and Neolithic microliths). Alongside these data are the OLE estimated limits for each attribute (T_E) and how much the original range data has been extended by the OLE estimates, expressed as a percentage. Length, breadth and thickness values are recorded in millimetres (mm). Mass is recorded in grams (g). The ceramics display varying n for each metric as the artefacts are often fragmentary shards, meaning that not all information is available for every artefact. 'Na' indicates that estimates were not able to be returned using the OLE method, likely because there were multiple records of the same value entered into the model. Note that the mean values for PC1 and PC2 are not presented as by definition they approach zero with only very minor variation.

Artefact Case Study	Variable	Mean	Minimum	OLE Minimum	Maximum	OLE Maximum	OLE Range Extension %
Old Park Flakes ($n = 173$)	Length	22.9	9.4	**8.3**	55.6	**67.2**	**27.5**
	Breadth	20.1	8	**7.6**	53.1	**68.3**	**34.6**
	Thickness	5.7	1.3	**0.8**	18.6	**20.7**	**15.0**
	Mass	3.9	0.2	**0.062**	40.2	**54.6**	**36.3**
Boxgrove Handaxes ($n = 254$)	Length	117.8	44.5	**38.4**	196.4	**217.1**	**17.6**
	Breadth	76.2	34.3	**26.9**	114.6	**116.9**	**12.1**
	Thickness	28.9	13.1	**10.4**	53.1	**57.8**	**18.5**
	Mass	286.2	23	**11.1**	959.7	**1059.0**	**11.9**
	Elongation	0.654	0.405	**0.247**	0.883	**0.898**	**36.2**
	Refinement	0.384	0.223	**0.203**	0.862	**1.087**	**38.3**
	PC1	–	−0.590	**−0.621**	1.027	**1.760**	**47.2**
	PC2	–	−0.468	**−0.536**	0.726	**0.948**	**24.3**
Olduvai Bed IV Cleavers (n = 134)	Length	142.9	93	**84.0**	205	**216.3**	**18.1**
	Breadth	87.6	57	**49.5**	119	**126.8**	**24.7**
	Thickness	42.2	22	**19.5**	60	**60.9**	**8.9**
	Edge Length	259.7	20	**−10.6**	481	**501.2**	**11.0**
	Mass	588.3	252	**224.6**	1269	**1557.8**	**31.1**
Sibudu Bipolar Cores ($n = 131$)	Length	15.8	9.6	**9.4**	24.8	**25.1**	**3.3**
	Breadth	10.6	5.4	**5.1**	17.1	**17.3**	**4.3**
	Thickness	6.2	2.5	**2.2**	10.9	**11.3**	**8.3**
	Mass	1.236	0.11	**0.046**	4.6	**5.2**	**17.2**
Capuchin Flakes ($n = 31$)	Length	41.3	18.8	**15.9**	71.0	**79.7**	**22.3**
	Breadth	28.9	11.6	**8.6**	52.4	**58.0**	**21.2**
	Thickness	13.5	5.1	**4.9**	33.5	**40.7**	**26.0**
	Mass	21.1	1.3	**1.1**	108.6	**168.5**	**56.0**
Copper Socketed Tang Points ($n = 93$)	Length	126.6	59.2	**43.7**	234.7	**258.9**	**22.6**
	Breadth	24.5	11.2	**7.9**	54.9	**79.1**	**62.9**
	Thickness	9.6	4.3	**3.1**	16.1	**18.1**	**27.1**
	Mass	49.9	10	**7.7**	148	**151.9**	**4.5**
Middle Historic Ceramic Bowl	Rim Diameter ($n = 1466$)	15.6	6.0	**3.5**	24.0	**28.8**	**40.7**
	Rim Thickness ($n = 1463$)	2.4	1.1	**na**	4.7	**4.8**	**na**
	Base Diameter ($n = 77$)	5.1	4.0	**na**	6.0	**na**	**na**
	Height ($n = 59$)	59.9	50.0	**47.4**	67.0	**68.5**	**24.3**
Neolithic Impressed Ware Pots	Rim Diameter ($n = 38$)	11.7	5.0	**1.3**	18.0	**19.8**	**42.3**
	Rim Thickness ($n = 54$)	6.9	4.1	**3.3**	10.1	**10.6**	**21.7**
Neolithic Stentinello Pots	Rim Diameter ($n = 114$)	14.6	5.0	**na**	34.0	**36.8**	**na**
	Rim Thickness ($n = 146$)	6.3	3.0	**2.2**	11.7	**12.1**	**13.8**

Table 6
Mean and range data for the Paleoindian point samples, alongside the minimum and maximum values (T_E) for these four artefact types modelled using OLE. The extent of range extension through the OLE method, expressed as a percentage, is also displayed. Sample sizes varied between point types (Clovis [$n = 810$, except mass where $n = 101$], Eastern Clovis [$n = 228$, except mass where $n = 63$], Folsom [$n = 179$, except mass where $n = 125$], Midland [$n = 64$]). Length, breadth and thickness were recorded in mm following Buchanan and Hamilton (2021). Mass was recorded in grams using digital scales.

		Mean	Minimum	OLE Minimum	Maximum	OLE Maximum	OLE Range Extension %
Length (mm)	**Clovis**	67.3	21.9	**20.5**	230.5	**243.4**	**6.9**
	E. Clovis	54.2	27.5	**26.6**	151.0	**211.0**	**49.3**
	Folsom	40.5	16.7	**11.7**	92.7	**105.8**	**23.8**
	Midland	42.0	19.1	**14.5**	68.0	**70.5**	**14.8**
Breadth (mm)	**Clovis**	27.1	13.0	**12.6**	64.4	**66.3**	**4.5**
	E. Clovis	24.5	11.9	**7.5**	41.0	**46.7**	**34.8**
	Folsom	20.4	10.9	**8.8**	36.3	**40.1**	**23.3**
	Midland	19.5	12.5	**10.4**	30.5	**35.1**	**37.1**
Thickness (mm)	**Clovis**	7.2	3.0	**2.8**	13.7	**14.9**	**12.2**
	E. Clovis	6.8	3.0	**1.8**	14	**19.4**	**60.0**
	Folsom	4.2	2.3	**2.1**	11.1	**15.0**	**46.2**
	Midland	4.3	3.0	**2.9**	6.0	**6.5**	**19.6**
Mass (g)	**Clovis**	38.8	1.9	**1.4**	196.2	**201.8**	**3.1**
	E. Clovis	10.2	2.6	**1.6**	23.9	**27.6**	**21.9**
	Folsom	4.3	0.5	**−0.4**	32.5	**57.9**	**82.3**
	Midland	4.1	1.2	**1.0**	8.6	**8.8**	**5.1**

Table 7
Range data for each morphological variable investigated in the geometric microlith case study. Alongside these data are OLE estimated limits for each attribute (T_E) and data range extensions following the OLE estimates expressed as a percentage. Length, breadth, thickness and area values are recorded in millimetres (mm). 'Na' indicates that estimates were not able to be returned because there were duplicate records. Note that the mean values for PC1 and PC2 are not presented as by definition they approach zero with only very minor variation.

Case Study	Variable	Duplicates Removed or Randomised	Mean	Min.	OLE Min.	Max.	OLE Max.	OLE Range Extension %
Geometric microliths phase A ($n = 344$)	Length	Removed	17.6	9.8	**8.9**	31.8	**35.8**	**22.3**
		Randomised	17.6	9.8	**9.0**	31.8	**35.8**	**21.8**
	Breadth	Removed	9.2	5.3	**5.0**	14.8	**15.5**	**10.5**
		Randomised	9.2	5.3	**5.0**	14.8	**15.5**	**10.5**
	Thickness	Removed	2.3	1.0	**0.8**	4.8	**6.1**	**39.5**
		Randomised	2.3	1.0	**0.8**	4.8	**6.1**	**39.5**
	Area	Removed	9.1	3.9	**3.6**	18.7	**21.3**	**19.6**
		Randomised	9.1	3.9	**3.6**	18.7	**21.3**	**19.6**
	PC1	Removed	–	−7.3	**−7.4**	13.1	**18.3**	**26.0**
		Randomised	–	−7.3	**−7.4**	13.1	**18.3**	**26.0**
	PC2	Removed	–	−3.2	**−3.3**	5.7	**6.5**	**10.1**
		Randomised	–	−3.2	**−3.3**	5.7	**6.5**	**10.1**
Geometric microliths phase B ($n = 390$)	Length	Removed	17.2	8.2	**6.6**	28.3	**31.7**	**24.9**
		Randomised	17.2	8.2	**6.6**	28.3	**31.7**	**24.9**
	Breadth	Removed	8.2	4.9	**4.4**	12.8	**na**	**na**
		Randomised	8.2	4.9	**4.4**	12.8	**12.9**	**7.6**
	Thickness	Removed	2.1	1.1	**na**	4.4	**5.2**	**na**
		Randomised	2.1	1.1	**1.0**	4.4	**5.2**	**27.3**
	Area	Removed	7.4	3.3	**na**	14.3	**16.1**	**na**
		Randomised	7.4	3.3	**3.1**	14.3	**16.1**	**18.2**
	PC1	Removed	–	−13.9	**−16.3**	7.5	**8.0**	**13.6**
		Randomised	–	−13.9	**−16.3**	7.5	**8.0**	**13.6**
	PC2	Removed	–	−3.7	**−3.9**	10.4	**3.12**	**148.9**
		Randomised	–	−3.7	**−3.9**	10.4	**3.12**	**148.9**

subjected to diverse sampling scenarios demonstrate OLE to provide broadly accurate estimates for the unknown (undiscovered) morphological extremes of partial artefact assemblages. Diverse artefact case studies illustrate how OLE has potential to both substantially increase the morphological range of defined artefact assemblages and/or classifications, or provide little extension relative to ranges already known.

As with any modelling process, the present results should only be treated as theoretically robust predictions (when the model assumptions are met). In addition, the use of OLE does not mean that these extreme morphologies *were* created in the past, they instead represent the best available *estimate* for a range limit given the artefacts available and artefact-groupings (typologies, raw materials, production stages, or other logical clusters) created by archaeologists. With this in mind, our data demonstrates that OLE has potential to play an important role in understanding past artefact morphological variation.

4.1. How accurately did OLE perform?

Range estimation varied according to the data sampling conditions used and number of records entered into the model. Across all conditions it was evident that the more complete the sample, the more accurate the estimated ranges were. This was most clear in the replica handaxe scenario. This is not surprising, but it is noteworthy that sampling conditions of 10–20% (of a total assemblage) still provided median confidence interval [CI] coverages of 0.822 and above (Fig. 5A and B). Arguably, the 5% handaxe sample also returned broadly accurate estimates with a median CI coverage of 0.765. While archaeologists will very rarely know the complete (original) size of an artefact assemblage, it is clear that OLE can be used in conditions with highly fragmentary datasets. Archaeologists should, however, endeavour to use as complete (large) a dataset as possible to maximise the accuracy of results.

The Early Archaic points (EAP) were surprising insofar as all

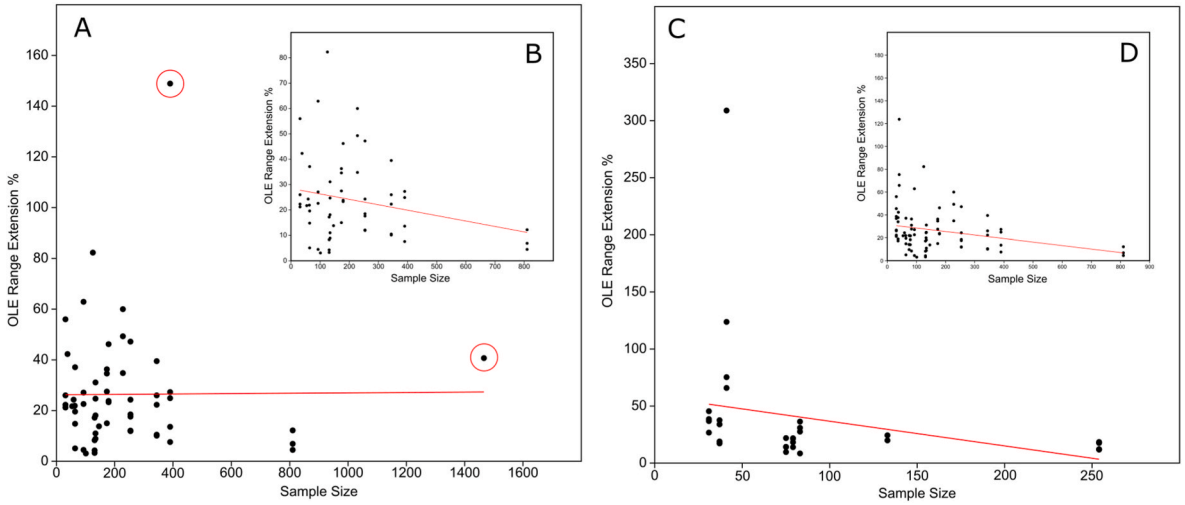

Fig. 6. Correlation of sample size against the range extensions, expressed as a percentage, created by OLE. Only length, breadth, thickness and mass data are used, and thus, each artefact type or site has the potential of contributing four data points to each plot. Figure A presents data from all results present in Tables 5–7 (minus the duplicate randomised results in Table 7). Highlighted in red are two outliers, which when removed, result in a negative relationship between range extension percentage and sample size (B). Additional OLE tests were conducted on length, breadth, thickness and mass data from eight Acheulean handaxe assemblages previously analysed by Key (2015). Figure C demonstrates this negative relationship to also be present in these data, while Figure D presents the pooled data from the main case studies and these new stone tool data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

sampling percentage conditions (20%, 30%, 40%, and 50%) provided broadly comparable CI coverages. Reasoning for this difference relative to the handaxes may lie in the distribution of each sample's data. The handaxe variables (particularly their upper range) typically displayed greater kurtosis, with records being taken from this long tail, while the EAP data displayed steeper distribution curves and relatively reduced kurtosis, with the small 'complete' assemblage size meaning records were almost always taken from the same distribution shape irrespective of the sampling condition (Supplementary Table 3). In turn, there was potentially less possibility for the modelled Weibull curve to vary depending on the EAP sampling condition. When combined with Rivadeneira et al.'s (2009) finding that gradual faunal extinctions (shallow distributions, long tails [low kurtosis]) more often produce OLE CI coverage below 0.95, while sudden extinctions (steep distributions, short tails [high kurtosis]) erred on greater coverages values, it suggests that even when EAP sampling percentages were reduced, the modelled distribution curve would have leaned towards higher coverage relative to the handaxe sample, and there may not have been much potential for this to vary.

Similarly, the impact of varying distribution shapes helps to explain the differing overall CI coverage between the combined (all k values and sampling percentages) handaxe and EAP results, where the latter displays higher values (Fig. 4). It potentially also explains the greater CI coverage demonstrated by the minimum range estimates of both handaxes and EAPs compared to maximum estimates (Fig. 4). Indeed, minimum values for variables in both tool types displayed steeper distributions (Supplementary Figs. 1 and 2). OLE estimates are, therefore, likely more accurate if samples display relatively low standard deviations, low kurtosis, and consequently comparatively short distribution tails. The adequate performance of the small EAP sample was to a degree surprising as we had anticipated that records from the tail opposing the one being modelled, or the distribution's peak, could have occasionally (as part of the random sampling) contributed to estimates; particularly under the higher k value scenarios. While it is reassuring that there were no major implications for the method's use, future use of the method should take this into account, and potentially lower the k value used when sample sizes are particularly small.

The removal of duplicate data is revealed to have little impact on OLE estimates (Table 7). Potentially in some samples the removal of duplicates may have led to the creation of lower kurtosis, and in turn potentially resulted in less accurate inferences, following the above argument. However, as Table 7 demonstrates, any impact is likely very marginal, but would increase on a relative basis in line with the number of duplicates removed. If duplicates were present at the extreme ranges of artefact samples, then this could have impacted results, but these instances are likely rare (see: Section 2.1.4). If a dataset is known to display multiple identical records at its morphological extremes, then the resampling procedure in Section 2.4.6.1 may be preferential to use. Alternatively, more accurate morphometric data (i.e., to a higher decimal place) should be collected to avoid duplication.

Previous temporal-based research has recommended the use of ~10 records (k) when producing OLE estimates (Solow, 2005; Rivadeneira et al., 2009). We tested the impact of this methodological choice when modelling metric data, running scenarios using 5, 10, 20 and 30 records. Our results are consistent with these earlier studies, with 5–10 records performing well (Fig. 5C and D). In the future, archaeologists using OLE to identify morphometric ranges may wish to run models under both conditions, or use a compromise of $k = 8$. Disentangling the best k value for each assemblage and variable is, however, complicated and varies dependent on the data investigated (Supplementary Tables 1 and 2). Irrespective, these results support the use of OLE in highly fragmented data scenarios where relatively few artefacts exist.

4.2. How did OLE impact the archaeological case studies?

It is worth emphasising that OLE is used here to model morphological extremes. The artefact forms identified in this study would not have been common in the past; quite the contrary, the frequency with which they were produced would have been incredibly low. Yet, identifying the morphological extremes of artefacts is important. It allows us to better understand how similar or distinct groups or classifications of artefact types and forms are. It potentially reveals the absolute limits of what is anatomically and cognitively possible by humans, extinct hominins and non-human primates. It demonstrates how selective pressures, drift and other cultural evolutionary mechanisms can create extreme deviation away from an original or supposedly 'preferred' form. It also allows us to better contextualise how external variables such as environmental markers and dietary choices relate to aspects of past

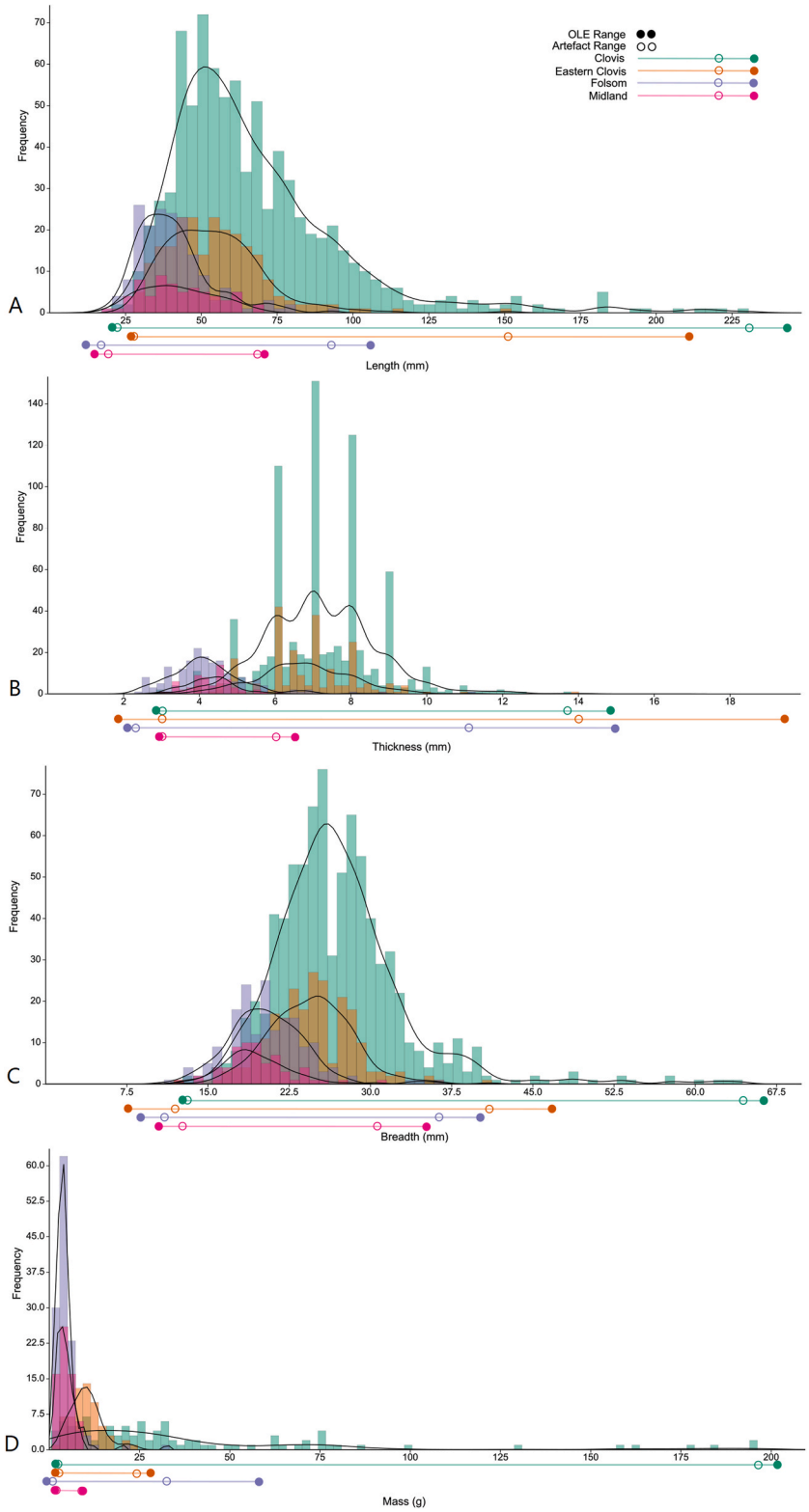

Fig. 7. Data distribution for length (A), thickness (B), breadth (C) and mass (D) in each of the four Paleoindian point types. Gaussian kernel density distributions for each are also presented. Morphometric ranges present in the artefact data are represented by hollow circles, while the OLE estimated ranges are depicted by solid circles.

material culture. Detailed discussion of each case study is not possible here. Instead, we highlight key examples to emphasise the value of OLE while simultaneously demonstrating how estimates vary according to the data records used.

Within Palaeolithic archaeology, flake technologies are often described as expedient tools with low-to-minimal consideration of their form beyond the necessity of producing a sharp, acute cutting edge (e.g., Borel et al., 2013; Gurtov and Eren, 2014; Key and Lycett, 2017; Vaquero and Romahnoli, 2018; Kuhn, 2021; Stemp et al., 2021). Certainly, this was how AK previously described the flakes from Old

Park, which vary highly in their form (Key et al., 2022). OLE increased this variability by 15%–36%, depending on the morphometric attribute investigated. For example, flake mass – at least for the technological strategies discovered so far (as there could be others [e.g., Sharon, 2010]) – should no longer be considered to vary from 0.2 to 40.2 g, but instead likely ranges from ~0.06 to 55 g. This does not change our understanding of how or why these flakes were produced. Instead, it suggests the smallest and largest flake artefacts produced by this hominin population(s) have not yet been excavated, and it supports the assertion that the smallest flakes produced during knapping processes are rarely recovered by archaeologists.

When the Old Park range extensions are compared to Boxgrove, some may be surprised to see similar values. Indeed, Boxgrove handaxes are often characterised by their "conservatism in outline shape" (Hosfield et al., 2018: 12) and "limited form variation relative to many other Acheulean sites" (Key, 2019: 563). These prior inferences of morphological homogeneity can now be argued to exclude the extreme tool-forms that probably remain buried at low frequencies elsewhere at the site (Roberts and Parfitt, 1999; Pope et al., 2020). Potentially, extreme forms were also removed from the locality to be used elsewhere, but the artefacts investigated here are not drawn from the wider landscape, and inferences must reflect the sample in hand. Although some Boxgrove range extensions were limited (e.g., mass at 11.9%), OLE estimated there to be considerable undiscovered shape diversity. Indeed,

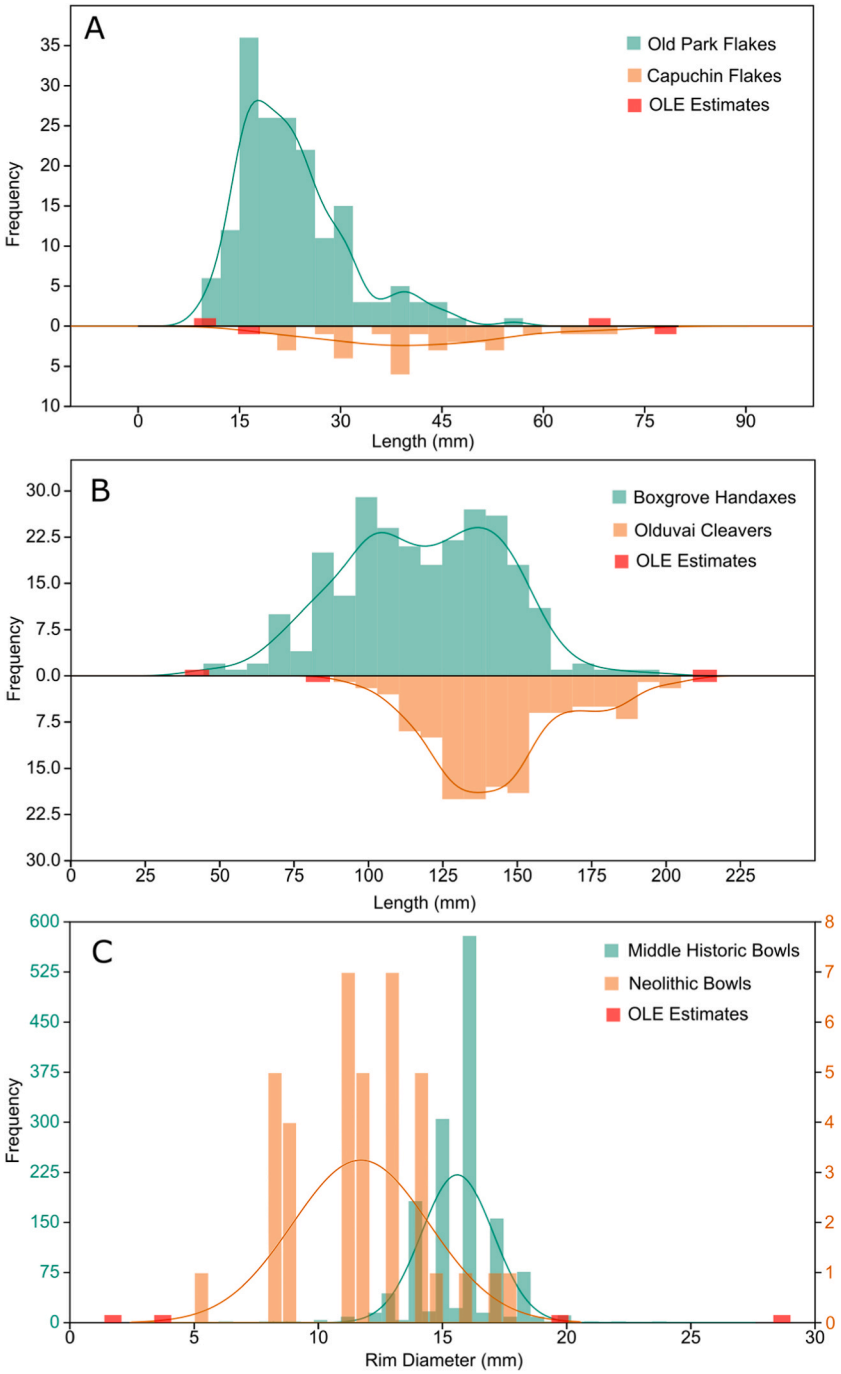

Fig. 8. Data distributions for the length in the Old Park flakes (A), capuchin flakes (A), Boxgrove handaxes (B), Olduvai cleavers (B), middle historic bowls (C) and Neolithic pots (C). Highlighted in red are the OLE range estimates for each distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

PC1 and PC2 are predicted to increase by 47% and 24%, respectively. This does not mean that Boxgrove does not display strong ovate central tendencies (Roberts and Parfitt, 1999; Lycett and von Cramon-Taubadel, 2015; Hosfield et al., 2018; García-Medrano et al., 2019), but rather, there are likely far more diverse forms being produced by these *H. heidelbergensis s.l.* individuals than previously recognised and tool-form ranges may be similar to other Acheulean sites (Hosfield et al., 2018). As noted above, some of this additional shape variation may be linked to increased elongation and reduced refinement levels, but it is not immediately clear how PC range extensions relate to the individual variables contributing to the shape data. Future discussion of shape standardisation in the Boxgrove handaxes should take this into consideration.

The two ceramic case studies were produced through distinct processes, with the Middle Historic bowls being wheel thrown en masse (Petrie, 2020) and the Neolithic pots being hand-built in more limited numbers (Robb, 2007). One would expect the former to err towards greater (positive) kurtosis and the latter to display more evenly distributed forms that lean towards lower (potentially negative) kurtosis. In turn, and as discussed above, the Middle Historic bowls could theoretically have displayed lower OLE range extensions (Rivadeneira et al., 2009). It is therefore interesting that, at least in terms of rim diameter – the one directly comparable metric – both ceramic types had their range similarly extended (40.7% and 42.3% for the Middle Historic and Neolithic Impressed Ware, respectively). The similarity may at first seem surprising; the Middle Historic bowl's rim diameter standard deviation was half that of the Neolithic pots (1.4 and 2.8 mm, respectively; Fig. 8). However, this does not take into account the rare instances when substantial additional morphological variation was introduced during the throwing process – be it intentionally or not – and these irregular forms were kept and fired. Given the size of the Middle Historic sample, a portion of this 'rare' additional variation has been captured and provides the data used to model the Weibull curve. In turn, a shallow curve with a long tail has been modelled relative to the distribution displayed by more central values in the sample (Fig. 8). The difference between the two ceramic samples is then, that on a relative basis, the likelihood of these more extreme forms being produced is substantially reduced for the Middle Historic bowls, and this variation may not have been captured in the sample had it been a similar size to the Neolithic pots (i.e., $n = 54$). What the OLE modelling instead reflects is the shared ability of both processes to introduce rim diameter variability via cultural evolutionary mechanisms (Mesoudi, 2011; Lycett, 2015), even if this is at varying frequencies.

The Middle Historic ceramic case study also highlights how the collection of low-resolution (i.e., low accuracy) morphometric data can create high numbers of duplicate data at morphological extremes, which in turns leads to the OLE estimates being uninformative. Indeed, data were only collected to one decimal place, meaning that at times there were multiple records with the same value entered into the model and the model could not be run (Table 5). While the resampling procedure outlined in Section 2.4.7.1 would result in data that could be used by the model, a more accurate method would have been to collect higher resolution morphometric data in the first instance (e.g., to three or more decimal places).

The Sibudu bipolar core's more limited range extensions (3.3–17.2 %) result from there being multiple tightly constrained records at the upper and lower limits of each variable. This suggests the artefact sample to provide a near-accurate representation of the core variability produced by these Middle Stone Age humans. It also indicates any pressure to avoid the upper and lower core form limits may be low; which may again be derived from the expedient nature of flake technologies (including bipolar [Gurtov and Eren, 2014; Gurtov et al., 2015]). At the same time, tightly constrained records before a morphological limit also suggest the presence of strong pressures beyond these limits. Potentially, this is because larger cores made from other raw materials local to Sibudu, such as hornfels or dolerite, can be knapped via different techniques into alternative tool forms, and there is no need to use bipolar reduction. In other words, it could be that quartz was particularly preferred for specific functional tasks which required small bipolar flake blanks (e.g. composite projectiles) (de la Peña et al., 2018). Further, de la Peña and Wadley (2014) proposed that in the Howiesons Poort of Sibudu quartz was initially knapped through free-hand percussion, but when this technique could no longer be applied because cores became too small, the knappers switched to anvil (bipolar) percussion. This, again, supports the inference of strong core-size thresholds for the bipolar technique, as demonstrated by the limited OLE range extensions. It also might be because the quartz nodules surrounding Sibudu were by default small (the source of this raw material is unknown for this site). It would be interesting if future work similarly investigated the range extensions experienced by bipolar flakes, as it has been proposed that bipolar cores produce more regular morphologies (Jeske and Lurie, 1993, Diez-Martín et al., 2009) than bipolar flakes, and therefore cores are easier to recognize in lithic analysis.

In sharp contrast is copper socketed tang point breadth, which displays the second largest range extension across all case studies (62.9%; Table 5). This breadth increase results almost entirely from a 24 mm range increase at the upper limit of these tools, which itself results from a single outlier more than 15 mm greater than the next widest artefact (Supplementary Fig. 3). Had this outlier not been present in the artefact sample then the sample's kurtosis would have been significantly lower, the OLE extension would have been substantially more limited, and we would have no understanding of this morphological attribute's long extended tail. This reinforces the need to use as large an artefact sample as possible when using OLE, and the need to visualise data distributions to accurately understand OLE estimates. The OLE estimated 79.1 mm upper range limit (Table 5) suggests the outlying artefact's form to be underpinned by pressures that could feasibly have created even broader tools. Given the utilitarian function of most copper points (Gibbon 1998; Penman 1977), the outlier artefact and modelled range limit may represent non-utilitarian artefacts, and the tight clustering of artefacts prior to 39 mm in breadth may represent a functional threshold.

Finally, the Paleoindian point and geometric microlith case studies provide examples where morphological attributes could be used to differentiate between groupings of similar artefacts. In some cases, the OLE results had little to no impact on the artefactually observed differences between groups. For example, the lower breadth limit of all four point types remained broadly similar after OLE modelling, and while differences exist in their maximal breadth, OLE extended each by a similar amount, and thus maintained these differences (Fig. 7C). Likewise, the minimum and maximum lengths of microliths in phase A remained slightly larger than those in phase B, despite OLE revising their range limits. For other attributes, the OLE estimates have altered inferred differences or similarities between groupings. One of the most notable changes is observed in the length of Clovis and Eastern Clovis points. Prior to the present paper there was a 75 mm difference between the largest evidenced Eastern Clovis point (151 mm) and the largest Clovis point (231 mm). The OLE estimates reduced this difference substantially, with the modelled Clovis point maximum length being 243 mm, while the estimated maximum size of Eastern Clovis points is now 211 mm, 60 mm longer than previously thought. This substantially reduces the strength of any claimed size difference between these point types in terms of overall potential size range, although not necessarily in terms of central tendency or modal frequency. There are no examples of morphological differences between the two geometric microlith samples altering substantially as a result of the OLE modelling. This supports the assertion that artefactual differences between phase A and B are a largely accurate reflection of past behaviour, even if the scale of these differences is marginally altered by the OLE estimates.

5. Conclusion

Archaeologists rarely find complete artefact assemblages. That is,

sets of archaeological artefacts often do not represent the total cultural – and therefore behavioural – variation of past populations. In artefact morphometric studies, this inevitably leads to the conclusion that some artefact forms created in the past remain unknown. Usually, these forms would be at the extreme ends of morphometric ranges as they would have been infrequently produced in the past. Here, we investigated whether optimal linear estimation – a frequentist statistical technique able to estimate the upper and lower range boundaries of Weibull-distributed linear phenomena (Roberts and Solow, 2003) – can accurately identify the true range of artefact morphological attributes using the partial records in known (i.e., excavated) assemblages.

Validation tests using replica artefact assemblages demonstrate OLE to vary in accuracy depending on the number of records entered into the model, and the percentage of the original artefact sample available for investigation. Although more experimental testing and method validation are needed and encouraged (Eren et al., 2016a,b), the method is suggested to provide broadly accurate estimates for the morphological range of artefacts. It was revealed that as large a sample size as possible should be used, but estimates remain broadly accurate with samples as low as 5–10% of the original 'complete' assemblage. The optimal number of records (k) entered into the model is approximately 5–10 (with a compromise of $k = 8$ appearing suitable). Morphometric data from ten archaeological case studies, covering lithic, ceramic and metal artefacts produced by humans, extinct hominins and non-human primates, were investigated to see how OLE can further our understanding of past human behaviour. Results varied highly between each case study, but it is clear that OLE has potential to alter, or in some cases overhaul, our understanding of artefact morphological variation in the past. On occasion, only small morphometric range extensions were returned, but in other instances ranges were increased by upwards of 40%. At times, previously recognised differences or similarities between groups of artefacts were severely reduced or reversed.

CRediT authorship contribution statement

Alastair Key: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing - original draft, Writing - review & editing. **Metin I. Eren:** Conceptualization, Resources, Writing - original draft, Writing - review & editing. **Michelle R. Bebber:** Resources, Writing - original draft, Writing - review & editing. **Briggs Buchanan:** Resources, Writing - original draft, Writing - review & editing. **Alfredo Cortell-Nicolau:** Formal analysis, Resources, Writing - original draft, Writing - review & editing. **Carmen Martín-Ramos:** Resources, Writing - original draft, Writing - review & editing. **Paloma de la Peña:** Resources, Writing - original draft, Writing - review & editing. **Cameron A. Petrie:** Resources, Writing - original draft, Writing - review & editing. **Tomos Proffitt:** Resources, Writing - original draft, Writing - review & editing. **John Robb:** Resources, Writing - original draft, Writing - review & editing. **Konstantina-Eleni Michelaki:** Resources. **Ivan Jarić:** Formal analysis, Methodology, Resources, Writing - original draft, Writing - review & editing.

Declaration of competing interest

None

Acknowledgements

AK is grateful to Stephen Lycett for support during the production of the replica handaxe assemblage. Collection of capuchin flake data was made possible by a British Academy Fellowship and Leakey Foundation Grant to TP. Excavations to collect the Old Park flake data were funded by a Leakey Foundation Grant to AK. Data collection for the Olduvai cleavers was conducted as part of CMR's doctoral research project, funded by the London Natural Environment Research Council Doctoral Training Partnership (London NERC DTP training grant NE/L002485/1). TP was supported by grant CEECINST/00052/2021 funded by the Portuguese Foundation for Science and Technology. PDLP has a Ramón y Cajal Research contract (RYC2020-029506-I) at the Universidad de Granada (Spain) funded by European social fund and the Agencia Estatal de Investigación (Spain). ACN was supported by H2020-MSCA–IF–2020 No. 101020631 during the development of this research. This project came about through discussions between AK and ME during a Royal Society funded research trip (RGS\R1\211318).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jas.2023.105921.

References

Arroyo, A., de la Torre, I., 2020. Pitted stones in the Acheulean from Olduvai Gorge Beds III and IV (Tanzania): a use-wear and 3D approach. J. Hum. Evol. 145, 102837.
Bailey, G., 2008. Time perspectivism: origins and consequences. In: Holdaway, S., Wandsnider, L.A. (Eds.), Time in Archaeology: Time Perspectivism Revisited. Utah University Press, Utah, pp. 13–30.
Bebber, M.R., 2021. The role of functional efficiency in the decline of North America's Copper Culture (8000-3000 BP): an experimental, ecological, and evolutionary approach. J. Archaeol. Method Theor 28, 1224–1260.
Bebber, M.R., Norris, J.D., Flood, K., Fisch, M., Meindl, R.S., Eren, M.I., 2019. Controlled experiments support the role of function in the evolution of the North American copper tool repertoire. J. Archaeol. Sci.: Reports 26, 101917.
Bebber, M.R., Key, A.J.M., 2022. Optimal linear estimation (OLE) modeling supports early Holocene (9000–8000 RCYBP) copper tool production in North America. Am. Antiq. 87 (2), 267–283.
Bebber, M.R., Chao, A., 2022. The diversity of North America's "old copper" projectile points. In: Eren, M.I., Buchanan, B. (Eds.), Defining and Measuring Diversity in Archaeology: Another Step toward an Evolutionary Synthesis of Culture. Berghahn Books, New York, pp. 43–63.
Birch, T., Martinón-Torres, M., 2019. Shape as a measure of weapon standardisation: from metric to geometric morphometric analysis of the Iron Age 'Havor' lance from Southern Scandinavia. J. Archaeol. Sci. 101, 34–51.
Blackman, M.J., Stein, G.J., Vandiver, P.B., 1993. The standardization hypothesis and ceramic mass: technological, compositional, and metric indexes of craft specialization at Tell Leilan, Syria. Am. Antiq. 58 (1), 60–80.
Boakes, E.H., Rout, T.M., Collen, B., 2015. Inferring species extinction: the use of sighting records. Methods Ecol. Evol. 6 (6), 678–687.
Borel, A., Gaillard, C., Moncel, M.-H., Sala, R., Pouydebat, E., Simanjuntak, T., Semah, F., 2013. How to interpret informal flakes assemblages? Integrating morphological description, usewear and morphometric analysis gave better understanding of the behaviors of anatomically modern humans from Song Terus (Indonesia). J. Anthropol. Archaeol. 32 (4), 630–646.
Bretzke, K., Conard, N.J., 2012. Evaluating morphological variability in lithic assemblages using 3D models of stone artifacts. J. Archaeol. Sci. 39 (12), 3741–3749.
Buchanan, B., O'Brien, M.J., Collard, M., 2014. Continent-wide or region-specific? A geometric morphometrics-based assessment of variation in Clovis point shape. Archaeological and Anthropological Sciences 6, 145–162.
Buchanan, B., Chao, A., Chiu, C.H., Colwell, R.K., O'Brien, M.J., Werner, A., Eren, M.I., 2017. Environment-induced changes in selective constraints on social learning during the peopling of the Americas. Sci. Rep. 7, 44431.
Buchanan, B., Hamilton, M.J., 2021. Scaling laws of Paleoindian projectile point design. J. Archaeol. Method Theor 28, 580–602.
Buchanan, B., Andrews, B., O'Brien, M.J., Eren, M.I., 2018. An assessment of stone weapon tip standardization during the Clovis–Folsom Transition in the Western United States. Am. Antiq. 83, 721–734.
Buchanan, B., Andrews, B., Kilby, J.D., Eren, M.I., 2019. Settling into the country: comparison of Clovis and Folsom lithic networks in western North America shows increasing redundancy of toolstone use. J. Anthropol. Archaeol. 53, 32–42.
Buchanan, B., Kilby, J.D., LaBelle, J.M., Surovell, T.A., Holland-Lulewicz, J., Hamilton, M.J., 2022. Bayesian modeling of the Clovis and Folsom radiocarbon records indicates a 200-year multigenerational transition. Am. Antiq. 87, 567–580.
Ceccarelli, L., Moletti, C., Bellotto, M., Dotelli, G., Stoddart, S., 2020. Compositional characterization of Etruscan earthen architecture and ceramic production. Archaeometry 62 (6), 1130–1144.
Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Stat. 11, 265–270.
Chao, A., Jost, L., 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93, 2533–2547.
Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monogr. 84, 45–67.
Clements, C.F., Worsfold, N.T., Warren, P.H., Collen, B., Clark, N., Blackburn, T.M., Petchey, O.L., 2013. Experimentally testing the accuracy of an extinction estimator: Solow's optimal linear estimation model. J. Anim. Ecol. 82 (2), 345–354.
Collard, M., Buchanan, B., Hamilton, M.J., O'Brien, M.J., 2010. Spatiotemporal dynamics of the Clovis–Folsom transition. J. Archaeol. Sci. 37, 2513–2519.

Colwell, R.K., 2013. EstimateS: statistical estimation of species richness and shared species from samples. Version 9 and earlier. User's Guide and application. Published at: http://purl.oclc.org/estimates. (Accessed 2 June 2023).
Colwell, R.K., Elsensohn, J.E., 2014. EstimateS turns 20: statistical estimation of species richness and shared species from samples, with non-parametric extrapolation. Ecography 37, 609–613.
Cortell-Nicolau, A., 2019. Geomeasure: GIS and scripting for measuring morphometric variability. Lithic Technol. 44 (3), 153–165.
Cortell-Nicolau, A., García-Puchol, O., Juan-Cabanilles, J., 2023. The geometric microliths of cueva de la cocina and their significance in the mesolithic of Eastern Iberia: a morphometric study. Quat. Int. 677–678, 51–64.
Courty, M.A., Roux, V., 1995. Identification of wheel throwing on the basis of ceramic surface features and microfabrics. J. Archaeol. Sci. 22 (1), 17–50.
Crema, E.R., Bevan, A., Shennan, S., 2017. Spatio-temporal approaches to archaeological radiocarbon dates. J. Archaeol. Sci. 87, 1–9.
de la Peña, P., 2015a. Refining our understanding of Howiesons Poort lithic technology: the evidence from Grey rocky layer in Sibudu cave (KwaZulu-Natal, South Africa). PLoS One 10 (12), e0143451.
de la Peña, P., 2015b. The interpretation of bipolar knapping in African Stone Age studies. Curr. Anthropol. 56 (6), 911–923.
de la Peña, P., Wadley, L., 2014. Quartz knapping strategies in the Howiesons poort at Sibudu (KwaZulu-Natal, South Africa). PLoS One 9 (7), e101534.
de la Peña, P., Taipale, N., Wadley, L., Rots, V., 2018. A techno-functional perspective on quartz micro-notches in Sibudu's Howiesons Poort indicates the use of barbs in hunting technology. J. Archaeol. Sci. 93, 166–195.
de la Torre, I., 2016. The origins of the Acheulean: past and present perspectives on a major transition in human evolution. Phil. Trans. R. Sco. B. 371 (1698), 20150245.
de la Torre, I., Mora, R., 2005. Technological Strategies in the Lower Pleistocene at Olduvai Beds I & II. Etudes et Recherches archéologiques de l'Université de Liège (ERAUL).
de la Torre, I., McHenry, L., Njau, J., Pante, M., 2011. The origins of the Acheulean at Olduvai Gorge (Tanzania): a new paleoanthropological project in East Africa. Archaeol. Int. 15, 89–98.
Deino, A.L., Heil, C., King, J., McHenry, L.J., Stainstreet, I.G., Stollhofen, H., Njau, J.K., Mwankunda, J., Schick, K.D., Toth, N., 2021. Chronostratigraphy and age modeling of Pleistocene drill cores from the Olduvai basin, Tanzania (Olduvai Gorge coring project). Palaeogeogr. Palaeoclimatol. Palaeoecol. 571, 109990.
Dibble, H.L., Holdaway, S.J., Lin, S.C., Braun, D.R., Douglass, M.J., Iovita, R., McPherron, S.P., Olszewski, D.I., Sandgathe, D., 2017. Major fallacies surrounding stone artifacts and assemblages. J. Archaeol. Method Theor 24, 813–851.
Diez-Martín, F., Sánchez, P., Domínguez-Rodrigo, M., Mabulla, A., Barba, R., 2009. Were Olduvai hominins making butchering tools or battering tools? Analysis of a recently excavated lithic assemblage from BK (Bed II, Olduvai Gorge, Tanzania). J. Anthropol. Archaeol. 28 (3), 274–289.
Diez-Martín, F., Sánchez Yustos, P., Uribelarrea, D., Baquedano, E., Mark, D.F., Mabulla, A., Fraile, C., Duque, J., Díaz, I., Pérez-González, A., Yravedra, J., Egeland, C.P., Organista, E., Domínguez-Rodrigo, M., 2015. The origin of the Acheulean: the 1.7 million-year-old site of FLK west, Olduvai Gorge (Tanzania). Sci. Rep. 5, 17839.
Djakovic, I., Key, A., Soressi, M., 2022. Optimal linear estimation models predict 1400–2900 years of overlap between Homo sapiens and Neandertals prior to their disappearance from France and northern Spain. Sci. Rep. 12, 15000.
Dunnell, R.C., Dancey, W.S., 1983. The siteless survey: a regional scale data collection strategy. Adv. Archaeol. Method Theor. 6, 267–287.
Eren, M.I., 2012. Were unifacial tools regularly hafted by Clovis foragers in the North American Lower Great Lakes region? An empirical test of edge class richness and attribute frequency among distal, proximal, and lateral tool-sections. Journal of Ohio Archaeology 2, 1–15.
Eren, M.I., Chao, A., Hwang, W.H., Colwell, R.K., 2012. Estimating the richness of a population when the maximum number of classes is fixed: a nonparametric solution to an archaeological problem. PLoS One 7, e34179.
Eren, M.I., Buchanan, B., O'Brien, M.J., 2015. Social learning and technological evolution during the Clovis colonization of the New World. J. Hum. Evol. 80, 159–170.
Eren, M.I., Buchanan, B., 2016. Clovis Technology. eLS, pp. 1–9.
Eren, M.I., Lycett, S.J., Patten, R.J., Buchanan, B., Pargeter, J., O'Brien, M.J., 2016a. Test, model, and method validation: the role of experimental stone artifact replication in hypothesis-driven archaeology. Ethnoarchaeology 8, 103–136.
Eren, M.I., Chao, A., Chiu, C.H., Colwell, R.K., Buchanan, B., Boulanger, M.T., Darwent, J., O'Brien, M.J., 2016b. Statistical analysis of paradigmatic class richness supports greater Paleoindian projectile-point diversity in the Southeast. Am. Antiq. 81, 174–192.
Eren, M.I., Buchanan, B. (Eds.), 2022. Defining and Measuring Diversity in Archaeology: Another Step toward an Evolutionary Synthesis of Culture. Berghahn Books, New York.
Eren, M.I., Bebber, M.R., Knell, E.J., Story, B., Buchanan, B., 2022. Plains Paleoindian projectile point penetration potential. J. Anthropol. Res. 78, 84–112.
Falótico, T., Ottoni, E.B., 2016. The manifold use of pounding stone tools by wild capuchin monkeys of Serra da Capivara National Park, Brazil. Behaviour 153 (4), 421–442.
Falótico, T., Coutinho, P.H.M., Bueno, C.Q., Rufo, H.P., Ottoni, E.B., 2018. Stone tool use by wild capuchin monkeys (*Sapajus libidinosus*) at Serra das Confusões National Park, Brazil. Primates 59, 385–394.
Foley, R., 1981. Off-site archaeology: an alternative approach for the short-sited. In: Hodder, I., Isaac, G., Hammond, N. (Eds.), Pattern of the Past. Studies in Honour of David Clarke. Cambridge University Press, Cambridge, pp. 157–184.
García-Medrano, P., Ollé, A., Ashton, N., Roberts, M.B., 2019. The mental template in handaxe manufacture: new insights into Acheulean lithic technological behavior at Boxgrove, Sussex, UK. J. Archaeol. Method Theor 26, 396–422.
García-Puchol, O., McClure, S.B., Juan-Cabanilles, J., Cortell-Nicolau, A., Diez-Castillo, A., Pascual-Benito, J. Ll, Pardo-Gordó, S., Gallello, G., Ramacciotti, M., Molina-Balaguer, Ll, López-Montalvo, E., Bernabeu-Aubán, J., Basile, M., Real-Margalef, C., Sanchís-Serra, A., Pérez-Fernández, A., Orozco-Köhler, T., Carrión-Marco, Y., Pérez-Jordá, G., Barrera-Cruz, M., Escribá-Ruiz, P., Jiménez-Puerto, J., 2023. A multi-stage Bayesian modelling for building the chronocultural sequence of the late mesolithic at cueva de la Cocina (Valencia, Eastern Iberia). Quat. Int. 677–678, 18–35.
Gibbon, G., 1998. Old copper in Minnesota: a review. Plains Anthropol. 43 (163), 27–50.
Gowlett, J.A.J., 2015. Variability in an early hominin percussive tradition: the Acheulean versus cultural variation in modern chimpanzee artefacts. Philosophical Transactions of the Royal Society B 370, 1682, 20140358.
Gurtov, A.N., Eren, M.I., 2014. Lower Paleolithic bipolar reduction and hominin selection of quartz at Olduvai Gorge, Tanzania: What's the connection? Quat. Int. 322–323, 285–291.
Gurtov, A.N., Buchanan, B., Eren, M.I., 2015. "Dissecting" quartzite and basalt bipolar flake shape: a morphometric comparison of experimental replications from Olduvai Gorge, Tanzania. Lithic Technol. 40 (4), 332–341.
Hamilton, M.J., Buchanan, B., 2009. The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: quantifying Clovis evolutionary dynamics. J. Anthropol. Archaeol. 28, 55–69.
Hay, R.L., 1994. Geology and dating of Beds III, IV and the Masek Beds. In: Leakey, M.D., Roe, D.A. (Eds.), Olduvai Gorge Vol. 5: Excavations in Beds III, IV, and the Maselk Beds (1968-1971). Cambridge University Press, Cambridge, pp. 8–14.
Hosfield, R., Cole, J., McNabb, J., 2018. Less of a bird's song than a hard rock ensemble. Evol. Anthropol. 27 (1), 9–20.
Huggett, J., 2020. Is big digit data different? Towards a new archaeological paradigm. J. Field Archaeol. 45 (S1), S8–S17.
Jacobs, Z., Roberts, R.G., 2008. Testing times: old and new chronologies for the Howieson's Poort and Still Bay industries in environmental context. South Afr. Archaeol. Soc. Goodwin Ser. 10, 9–34.
Jennings, T.A., 2012. Clovis, Folsom, and Midland components at the Debra L. Friedkin site, Texas: context, chronology, and assemblages. J. Archaeol. Sci. 39, 3239–3247.
Jennings, T.A., 2016. The impact of stone supply stress on the innovation of a cultural variant: the relationship of Folsom and Midland. PaleoAmerica 2, 116–123.
Jennings, T.A., Smallwood, A.M., 2019. The Clovis record. SAA Archaeol. Rec. 19, 45–50.
Jeske, R.J., Lurie, R., 1993. The archaeological visibility of bipolar technology: an example from the Koster site. Midcont. J. Archaeol. 131–160.
Justice, N.D., 1987. Stone Age Spear and Arrow Points of the Midcontinental and Eastern United States: A Modern Survey and Reference. Indiana University Press, Bloomington.
Key, A.J.M., 2015. Form and Function in the Lower Palaeolithic. Unpublished PhD Thesis, University of Kent.
Key, A.J.M., 2019. Handaxe shape variation in a relative context. Comptes Rendus Palevol 18 (5), 555–567.
Key, A.J.M., Lycett, S.J., 2017. Influence of handaxe size and shape on cutting efficiency: a large scale experiment and morphometric analysis. J. Archaeol. Method Theor 24, 514–541.
Key, A., Roberts, D., Jarić, I., 2021a. Reconstructing the full temporal range of archaeological phenomena from sparse data. J. Archaeol. Sci. 135, 105479.
Key, A., Jarić, I., Roberts, D.L., 2021b. Modelling the end of the Acheulean at global and continental levels suggests widespread persistence into the Middle Palaeolithic. Humanities and Social Sciences Communications 8, 55.
Key, A., Gowlett, J.A.J., 2022. Intercomparison of form, size and allometry in a million-year-old and modern replicated handaxe set. Lithic Technol. https://doi.org/10.1080/01977261.2022.2125670.
Key, A., Lauer, T., Skinner, M.M., Pope, M., Bridgland, D.R., Nobel, L., Proffitt, T., 2022. On the earliest Acheulean in Britain: first dates and in-situ artefacts from the MIS 15 site of Fordwich (Kent, UK). R. Soc. Open Sci. 9 (6), 211904.
Kovarovic, K., Aiello, L.C., Cardini, A., Lockwood, C.A., 2011. Discriminant function analyses in archaeology: are classification rates too good to be true? J. Archaeol. Sci. 38 (11), 3006–3018.
Kuhn, S.L., 2021. The Evolution of Paleolithic Technologies. Routledge, London.
LaRonge, M., 2001. An experimental analysis of Great Lakes Archaic copper smithing. North Am. Archaeol. 22 (4), 371–385.
Leakey, L.S.B., Leakey, M.D., 1964. Recent discoveries of fossil hominids in Tanganyika: at Olduvai and near lake Natron. Nature 4927, 5–7.
Leakey, M.D., 1971. Olduvai Gorge Volume 3: Excavations in Beds I and II, 1960-1963. Cambridge University Press, Cambridge.
Leakey, M.D., Roe, D.A., 1994. Olduvai Gorge Volume 5: Excavations in Beds III, IV and the Masek Beds, 1968-1971. Cambridge University Press, Cambridge.
Leonard, R.D., Jones, G.T., 1989. Quantifying Diversity in Archaeology. Cambridge University Press, Cambridge.
Lucas, G., 2014. Evidence of what? On the possibilities of archaeological interpretation. In: Chapman, R., Wylie, A. (Eds.), Material Evidence. Learning from Archaeological Practice. Routledge, London, pp. 311–323.
Luncz, L.V., Arroyo, A., Falotico, T., Quinn, P., Proffitt, T., 2022. A primate model for the origin of flake technology. J. Hum. Evol. 171, 103250.
Lycett, S.J., 2009. Quantifying transitions: morphometric approaches to Palaeolithic variability and technological change. In: Camps, M., Chauhan, P. (Eds.), Sourcebook of Palaeolithic Transitions. Springer, Cham, pp. 79–92.

Lycett, S.J., 2011. Most beautiful and most wonderful": those endless stone tool forms. J. of Evolu. Psychol. 9 (2), 143–171.
Lycett, S.J., 2015. Cultural evolutionary approaches to artifact variation over time and space: basis, progress, and prospects. J. Archaeol. Sci. 56, 21–31.
Lycett, A.J., von Cramon-Taubadel, N., 2015. Toward a "quantitative genetic" approach to lithic variation. J. Archaeol. Method Theor 22, 646–675.
Malinsky-Buller, A., Hovers, E., Marder, O., 2011. Making time: 'Living floors', 'palimpsests' and site formation processes – a perspective from the open-air Lower Paleolithic site of Revadim Quarry, Israel. J. Anthropol. Archaeol. 30 (2), 89–101.
Mannu, M., Ottoni, E.B., 2009. The enhanced tool-kit of two groups of wild bearded capuchin monkeys in the Caatinga: tool making, associative use, and secondary tools. Am. J. Primatol. 71 (3), 242–251.
Marchand, G., Perrin, T., 2017. Why this revolution? Explaining the major technical shift in Southwestern Europe during the 7th millennium cal. BC. Quat. Int. 428 (Part b), 73–85.
Martin, S.R., 1999. Wonderful Power: the Story of Ancient Copper Working in the Lake Superior Basin. Wayne State University Press, Detroit.
Martín-Ramos, C., 2022. Acheulean Technology and Hominin Cognition. Analysis of Large Cutting Tools from Olduvai Gorge Beds III and IV, Tanzania. Unpublished PhD thesis. University College London.
Meltzer, D.J., 2021. First Peoples in a New World: Populating Ice Age America. Cambridge University Press, Cambridge.
Mesoudi, A., 2011. Cultural Evolution. University of Chicago Press, Chicago.
Njau, J.K., toth, N., Schick, K., Stanistreet, I.G., McHenry, L.J., Stollhofen, H., 2021. The Olduvai Gorge Coring Project: Drilling high resolution palaeoclimatic and palaeoenvironmental archives to constrain hominin evolution. Palaeogeogr. Palaeoclimatol. Palaeoecol. 561, 110059.
Okumura, M., Araujo, A.G.M., 2019. Archaeology, biology, and borrowing: a critical examination of geometric morphometrics in archaeology. J. Archaeol. Sci. 101, 149–158.
Pante, M.C., de la Torre, I., d'Errico, F., Njau, J., Blumenshine, R., 2020. Bone tools from Beds II–IV, Olduvai Gorge, Tanzania, and implications for the origins and evolution of bone technology. J. Hum. Evol. 148, 102885.
Pardo-Gordó, S., Garcia Puchol, O., Diez Castillo, A.A., McClure, S.B., Juan Cabanilles, J., Perez Ripoll, M., Molina Balaguer, L., Bernabeu Auban, J., Pascual Benito, J.L., Kennett, D.J., Cortell-Nicolau, A., Tsante, N., Basile, M., 2018. Taphonomic processes inconsistent with indigenous Mesolithic acculturation during the transition to the Neolithic in the Western Mediterranean. Quat. Int. 483, 136–147.
Pargeter, J., de la Peña, P., 2017. Milky quartz bipolar reduction and lithic miniaturization: experimental results and archaeological implications. J. Field Archaeol. 42 (6), 551–565.
Penman, J., 1977. The old copper culture: an analysis of old copper artifacts. The Wisconsin Archeologist (New Series) 58 (4), 3–23.
Petrie, C.A., 2020. Resistance at the Edge of Empires: the Archaeology and History of the Bannu Basin from 1000 BC to AD 1200. Oxbow Books, Oxford.
Pétursdóttir, P., Olsen, B., 2018. Theory adrift: the matter of archaeological theorizing. J. Soc. Archaeol. 18 (1), 97–117.
Pimiento, C., Clements, C.F., 2014. When did *Carcharocles megalodon* become extinct? A new analysis of the fossil record. PLoS One 9 (1), e111086.
Pope, M., Roberts, M., Parfitt, S., 2020. The Horse Butchery Site GTP17: A High-Resolution Record of Lower Palaeolithic Hominin Behaviour at Boxgrove. SpoilHeap Publications, UK.
Proffitt, T., Luncz, L.V., Falótico, T., Ottoni, E.B., de la torre, I., Haslam, M., 2016. Wild monkeys flake stone tools. Nature 539, 85–88.
Proffitt, T., Martín-Ramos, C., 2019. Oldowan/acheulean Succession at Olduvai Gorge. In: Encyclopedia of Life Sciences. John Wiley & Sons, Chichester, pp. 1–9.
Proffitt, T., Reeves, J.S., Pacome, S.S., Luncz, L.V., 2022. Identifying functional and regional differences in chimpanzee stone tool technology. R. Soc. Open Sci. 9 (9), 220826.
Proffitt, T., Reeves, J.S., Braun, D.R., Malaivijitnond, S., Luncz, L.V., 2023. Wild macaques challenge the origin of intentional tool production. Sci. Adv. 9 (10), eade8159.
Ramsey, C.B., 2017. Methods for summarizing radiocarbon datasets. Radiocarbon 59 (6), 1809–1833.
Rivadeneira, M.M., Hunt, G., Roy, K., 2009. The use of sighting records to infer species extinctions: an evaluation of different methods. Ecology 90 (5), 1291–1300.
Robb, J., 2007. The Early Mediterranean Village: Agency, Material Culture, and Social Change in Neolithic Italy. Cambridge University Press, Cambridge.
Roberts, M.B., Parfitt, S.A., 1999. Boxgrove. A Middle Pleistocene Hominid Site at Eartham Quarry. Boxgrove, West Sussex. English Heritage Archaeological Report 17.
Roberts, D.L., Solow, A.R., 2003. When did the dodo become extinct? Nature 426, 245.
Roberts, D.L., Rossman, J.S., Jarić, I., 2021. Dating first cases of COVID-19. PLoS Pathog. 17 (6), e1009620.
Roux, V., 2019. Ceramics and Society: A Technological Approach to Archaeological Assemblages. Springer, Cham.
Sharon, G., 2010. Large flake Acheulian. Quat. Int. 223–224, 226–233.
Solow, A.R., 2005. Inferring extinction from a sighting record. Mathematical Bioscience 195 (1), 47–55.
Stemp, W.J., Graham, E., Helmke, C., Awe, J.J., 2021. Expedient lithic technology in complex sedentary societies: use-wear, flake size, and edge angle on debitage from two ancient Maya sites. J. Anthropol. Archaeol. 61, 101243.
Straus, L.G., 2002. Selecting small: microlithic musings for the upper Paleolithic and mesolithic of western Europe. Archaeological Papers of the American Anthropological Association 12 (1), 69–81.
Timpson, A., Barberena, R., Thomas, M.G., Méndez, C., Manning, K., 2020. Directly modelling population dynamics in the South American Arid Diagonal using 14C dates. Phil. Trans. Biol. Sci. 376 (1816), 20190723.
Tixier, J., 1957. Le Hachereau dans l'Acheuleen Nord-Africain. Notes typologiques. In: Congrés Préhistorique de France, pp. 914–923. Poitiers.
Turq, A., Roebrokes, W., Bourguignon, L., Faivre, J.-P., 2013. The fragmented character of Middle Palaeolithic stone tool technology. J. Hum. Evol. 65 (5), 641–655.
Vaquero, M., Romahnoli, F., 2018. Searching for lazy people: the significance of expedient behavior in the interpretation of Paleolithci assemblages. J. Archaeol. Method Theor 25, 334–367.
Vernon, W.W., 1990. In: New Archaeometallurgical Perspectives on the Old Copper Industry of North America, vol. 4. Geological Society of America Centennial Special, pp. 499–512.
Vidal-Cordasco, M., Ocio, D., Hicker, T., Marín-Arroyo, A.B., 2022. Ecosystem productivity affected the spatiotemporal disappearance of Neanderthals in Iberia. Nature Ecology and Evolution 6, 1644–1657.
Vidal-Cordasco, M., Terlato, G., Ocio, D., Marin-Arroyo, A.B., 2023. Neanderthal coexistence with *Homo sapiens* in Europe was affected by herbivore carrying capacity. Sci. Adv. 9 (38), eadi4099.
Visalberghi, E., Sirianni, G., Fragaszy, D., Boesch, C., 2015. Percussive tool use by Taï Western chimpanzees and Fazenda Boa Vista bearded capuchin monkeys: a comparison. Phil. Trans. R. Soc. B 370 (1682), 20140351.
Wadley, L., 2008. The Howieson's poort industry of Sibudu cave. South Afr. Archaeol. Soc. Goodwin Ser. 10, 122–132.
Wadley, L., Jacobs, Z., 2006. Sibudu Cave: background to the excavations, stratigraphy and dating. South. Afr. Humanit. 18 (1), 1–26.
Wang, S.C., Marshall, C.R., 2016. Estimating times of extinction in the fossil record. Biol. Lett. 12 (4), 20150989.
Wylie, A., 2017. How archaeological evidence bites back: strategies for putting old data to work in new ways. Sci. Technol. Hum. Val. 42 (2), 203–225.
Wymer, J., 1968. Lower Palaeolithic Archaeology in Britain. Humanities Press Inc., New York.
Wynn, T., Gowlett, J., 2018. The handaxe reconsidered. Evol. Anthropol. 27 (1), 21–29.
Zhang, H., Jarić, I., Roberts, D.L., He, Y., Du, H., Wu, J., Wang, C., Wei, Q., 2020. Extinction of one of the world's largest freshwater fishes: Lessons for conserving the endangered Yangtze fauna. Sci. Total Environ. 710, 136242.