

## RESEARCH ARTICLE

# Obtaining new resolutions in carnivore tooth pit morphological analyses: A methodological update for digital taphonomy

Lloyd A. Courtenay<sup>1\*</sup>, Darío Herranz-Rodrigo<sup>2,3</sup>, Rosa Huguet<sup>4,5,6</sup>, Miguel Ángel Maté-González<sup>1,7,8</sup>, Diego González-Aguilera<sup>1,7</sup>, José Yravedra<sup>2,3</sup>

**1** Department of Cartographic and Terrain Engineering, Higher Polytechnic School of Ávila, University of Salamanca, Ávila, Spain, **2** Department of Prehistory, Complutense University, Madrid, Spain, **3** C. A. I. Archaeometry and Archaeological Analysis, Complutense University, Madrid, Spain, **4** Institut Català de Paleoeologia Humana i Evolució Social (IPHES), Tarragona, Spain, **5** Department d'Historia i Historià de l'Art, Universitat de Rovira i Virgili (URV), Tarragona, Spain, **6** Unit Associated to CSIC, Departamento de Paleobiología, Museo de Ciencias Naturales, Madrid, Spain, **7** Gran Duque de Alba Institution, Diputación Provincial de Ávila, Ávila, Spain, **8** Department of Topographic and Cartography Engineering, Higher Technical School of Engineers in Topography, Geodesy and Cartography, Technical University of Madrid, Madrid, Spain

\* [ladc1995@gmail.com](mailto:ladc1995@gmail.com)



## OPEN ACCESS

**Citation:** Courtenay LA, Herranz-Rodrigo D, Huguet R, Maté-González MÁ, González-Aguilera D, Yravedra J (2020) Obtaining new resolutions in carnivore tooth pit morphological analyses: A methodological update for digital taphonomy. *PLoS ONE* 15(10): e0240328. <https://doi.org/10.1371/journal.pone.0240328>

**Editor:** Efthymia Nikita, The Cyprus Institute, CYPRUS

**Received:** April 24, 2020

**Accepted:** September 24, 2020

**Published:** October 8, 2020

**Copyright:** © 2020 Courtenay et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data used are held in a public repository and directly cited in the text. Data as well as any code used can be found at [https://github.com/LACourtenay/GMM\\_Measurement\\_Accuracy\\_Tools](https://github.com/LACourtenay/GMM_Measurement_Accuracy_Tools) while an additional supplementary video has been provided via <https://vimeo.com/409256777>

**Funding:** L.A.C. is supported by the Spanish Ministry of Science, Innovation and Universities with a FPI Predoctoral Grant (Ref. PRE2019-

## Abstract

Modern day investigation in fields of archaeology and palaeontology can be greatly characterised by an exponential growth of integrated new technologies, nevertheless, while these advances are of great significance to multiple lines of research, their evaluation and update over time is equally as important. Here we present an application of inter and intra-observer analysis in taphonomy based geometric morphometrics, employing robust non-parametric statistical analyses for the study of experimental carnivore tooth pit morphologies. To fully understand the influence of measurement errors in the collection of this data, our statistical assessment was performed on fully superimposed, partially superimposed and raw landmark coordinates collected from 3D surface scanning. Experimental samples used to assess these errors includes wolf and dog tooth pits used in modern day ecological livestock predation analysis. Results obtained from this study highlight the importance of landmark type in the assessment of error, emphasising the value of semi-landmark models over the use of ambiguous Type III landmarks. In addition to this, data also reveals the importance of observer experience for the collection of data alongside an interesting increase in error when working with fully superimposed landmarks due to the “Pinocchio Effect”. Through this study we are able to redefine the geometric morphometric models used for tooth pit morphological analyses. This final hybrid Type II fixed landmark and semi-landmark model presents a significant reduction in human induced error, generating a more metrically reliable and replicable method that can be used for data pooling in future inter-institutional research. These results can be considered a fundamental step forward for carnivore inspired studies, having an impact on archaeological, palaeontological, modern-day ecological research as well as applications in other forensic sciences.

089411) associated to project RTI2018-099850-B-I00 and the University of Salamanca. R.H. is supported by MICINN/FEDER PGC2018-093925-B-C32, the AGUAR project number SGR 2017-1040; the Universitat Rovira i Virgili (2014, 2015 and 2016 PFR-URV-B2-147) and CERCA Programme/Generalitat de Catalunya. D.H.R. is supported by the Ministry of Science, Innovation and Universities, under the contract REF: PEJ2019-005420-A as part of the i+D+I Garantía Juvenil. Finally L.A.C., M.Á.M.G. and D.G.A. are supported by the General Foundation of the University of Salamanca under the Plan TCUE 2018-2020; Project "WOLF\_FOOTPRINT" (PC-TCUE18-20\_2013); lead researcher D.G.A. Funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## 1. Introduction

The identification and analysis of the Bone Surface Modifications (BSMs) produced by different taphonomic agents is a fundamental component in modern archaeological, palaeontological and, more recently fields of applied forensic sciences. Among the different types of BSMs, in prehistory tooth marks present valuable information that greatly contribute to the understanding of site formation processes, palaeolandscapes and their associated ecologies. Under this premise, many authors have paid particular attention to the different tooth marks that different animals may produce, whether this be for the consumption of meat or other practices of biological origin such as osteophagia or general foraging activities.

Among tooth mark studies, mammalian carnivores are a particularly important protagonist in archaeological research [1–5], considering their competition for resources with hominins throughout the Pleistocene. Interestingly herbivores [6–8], omnivores [9, 10], insects [11], and birds [12, 13], are also known modifiers of osteological remains, causing some confusion in the classification and interpretation of BSMs. Rodents such as the porcupine (*Hystrix sp.*) are an additional well documented bone accumulator [4, 14–16], proving to be problematic agents in the interpretation of faunal accumulations. Likewise, chondrichthyans [17, 18] as well as large [19, 20], and very large reptiles [21–23], are also of great importance especially in paleontological research. Last but not least, human tooth modifications to bone are also a compelling line of investigation with great value in archaeology [24, 25].

Through these studies, current understanding of carnivore BSMs enumerate 4 main types of alterations produced by teeth [1, 5]. These include rounded circular depressions caused by the direct imprint of the cuspid on bone (*pits*); elongated depressions with a circular base produced by the dragging of teeth across the cortical surface (*scores*); circular holes produced by the direct penetration of the tooth through the cortical walls (*puncture*); and the progressive deletion of large portions of bone through continuous chewing (*furrowing*).

To date a great deal of investigation has focused their efforts on the systematic study of pits and scores. This type of research has gone through numerous different phases, starting with a simple qualitative *in-visu* characterisation of these traces [1–5, 17, 18, 22, 23], to a more in depth integration of quantitative analyses [10, 21, 25–29]. In many cases, the use of metric variables such as the length and width of tooth marks have proven valuable variables in the interpretation of archaeological and paleontological sites [19, 30–34, *inter alia*]. Nevertheless, a common limitation noted by most is the ability to go beyond the size of the carnivore chewing and specifically identify the agent(s) present in a site.

New possibilities have recently been made available through Geometric Morphometric Methods (GMM). GMM are a powerful multivariate statistical toolset for morphological analyses [35, 36], originating primarily from fields of biology, evolutionary sciences, systematics, and physical anthropology [37–44]. Outside of anatomy, however, GMM has had other successful integrations into other aspects of archaeology, such as the study of lithic materials [45, 46, *inter alia*].

For geometric morphometric analysis, the identification and quantification of points of interest, known as *landmarks*, are required. These points can be defined as “a point of correspondence on an object that matches between and within populations” [47: pg. 3]. One advantage of landmarks is that they can either be recorded as 2D or 3D coordinates in space, providing a direct means of visualising variations between morphologies through the distortion or displacement of each landmark on each individual [48, 49]. The configuration of these points in space additionally present the advantage of preserving the full geometry of the specimens being studied [37]. Finally with the introduction of powerful 3D modelling equipment and methodological protocols, the capturing of morphological data is much more precise.

In BSM taphonomy, GMM were initially introduced as a means of quantifying cut mark morphologies for both 2D derived data [50] and entire 3D models [51]. These advances have obtained exceptional resolutions, presenting means of identifying the lithic tool and raw material used to produce each cut mark in archaeology [50–52, *inter alia*], while also arising in forensic research for sharp force trauma identification [53, 54, *inter alia*]. Likewise, initial efforts in the quantification of tooth score and pit morphologies reached up to 80% classification rates differentiating between carnivore agencies [55–58], with Machine Learning algorithms obtaining > 95% accuracy [58, 59].

Regardless of this success, however, it is important to consider the value of the landmark models being used for geometric morphometric data collection. To name one primary issue, landmarks can be further divided into three different types; easily locatable landmarks (Type I) and landmarks whose location are more ambiguous (Types II and III). Type I landmarks in biology are usually anatomical features of a precisely defined nature (e.g. tendon insertions), while Type II landmarks are defined by local properties (e.g. maximum curvature) and Type III landmarks are constructed points across the structure (e.g. a centroid) [35, 47: pg. 3–4]. From this concept, each type of landmark possesses a certain degree of error and subjectivity that must be addressed [60, 61]. Possible issues can be extrapolated to include; the experience of the researcher placing the landmark and analyst inconsistencies throughout measuring sessions [62–69]; the natural variability and preservation rate of the landmark itself [62, 70–73]; the landmark's definition and how easy it is to pinpoint [60–62, 67, 70]; and finally the precision of the measuring device being used [67–69].

Similarly, variability according to inter- and intra-observer error additionally has a significant effect on the comparison of information produced in different studies [67]. The ability to share data is a fundamental component of modern day research [74], especially in areas where accessibility to certain types of data is costly. This is a considerable issue when considering data derived from carnivore feeding, where in many cases large sample sizes are hard to obtain or access to wild animals may not be easy for researchers who reside on different continents [75]. Moreover, special care must be taken when working with animals, considering the conditions required to ensure correct welfare practices [76]. Under this premise, the ability to share data must be considered fundamental in order for research to advance.

While some initial approximations have been made to assess the reliability of geometric morphometric models in carnivore BSM analyses [55: pg. 97, 57: S2 Appendix], an in-depth study is yet to be carried out. For example, the current model for tooth pit analyses is known to produce highly complex dataset of high dimensionality. One particular issue with these datasets also includes the low proportion of cumulative variability represented by each variable [59]. Moreover, tooth pit models are based primarily on Type III landmarks, some of which are questionably much harder to locate than others. To ensure the reliability of this methodological approach, and where possible improve its precision, a detailed assessment is necessary.

The present study thus demonstrates how a detailed revision and update of Aramendi et al. [55]'s original model can facilitate future analyses of carnivore tooth pits with greater statistical reliability and accuracy. This in turn can be considered an important obstacle to supersede in order to make carnivore tooth mark data readily available to other researchers. Such advances are of great value to the study of carnivore BSMs; applicable to archaeological, palaeontological and modern day ecological research.

## 2. Methods

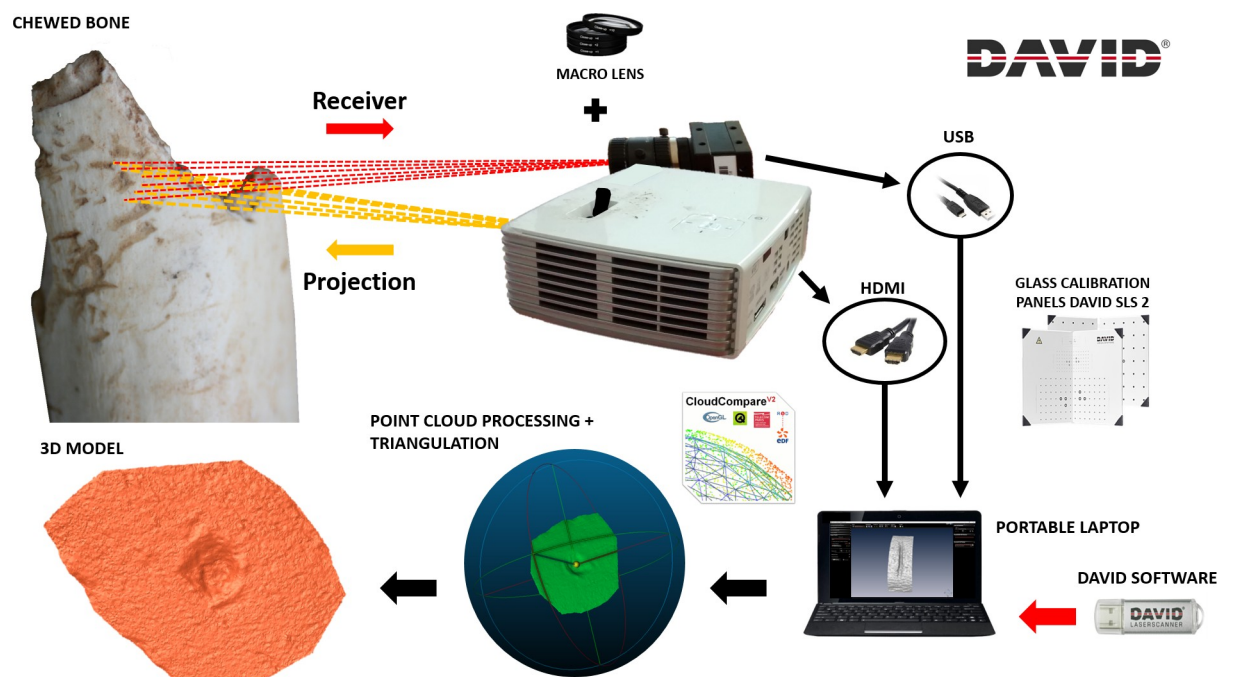
### 2.1. Tooth mark samples and digitisation

A total of 60 carnivore tooth pits were studied and compared for the purpose of this study. All marks were collected on the long bone diaphyses of large sized animals, including tibiae and

radii. Diaphyses were chosen considering their greater density than epiphyses and are thus more likely to survive during carnivore feeding. Samples were produced in a controlled setting and include tooth marks produced by *Canis lupus signatus* wolves (pits = 30) and an Irish Setter gundog (pits = 30). Tooth mark samples from both carnivores were obtained in a controlled setting during the feeding of these animals. For the case of wolves, these samples were obtained from semi-captive animals (>1000m<sup>2</sup>) from the Cabárceno natural park (Cantabria, Spain, <http://www.parquedecabarceno.com>) following standard feeding protocol established by the park. Wolves were left with the bones for 1 week before the samples were recollected by park employees. In the case of gundog tooth samples, bones were provided using product obtained from a local butcher. In this case, bones were left with the dogs for less than a day in an equally controlled environment. All samples were donated for use in the study, and no permits were required for their collection. Once collected, if necessary bones were cleaned in boiling water prior to analysis.

Each of these samples were originally studied by Yravedra et al. [58], identifying significant differences in tooth pit morphology for the identification of livestock predation in Europe. These bone surface modifications can thus be used as a control sample considering how morphological variations are already known in detail.

Digital reconstructions (Fig 1) of each mark were performed using the DAVID SLS-2 Structured Light Surface Scanner located at the TIDOP Research Group of the Higher Polytechnic School of Ávila (University of Salamanca, Spain). Equipment was calibrated using a 15 mm marker board while both the projector and the camera were equipped with additional macro-lenses (x2 to x10) for optimal resolution at microscopic scales [77]. Digitisations using this equipment are carried out in ca. 1 minute producing a point cloud density of up to 1.2 million points.



**Fig 1.** A graphical description of the digital reconstruction protocol used in the present study; producing 3D models of carnivore tooth marks on bone according to Maté-González et al. [77].

<https://doi.org/10.1371/journal.pone.0240328.g001>

## 2.2. Geometric morphometric data collection

For the purpose of this study, landmark data collection was performed by three separate individuals (A1-3) with varying degrees of experience. The first analyst (A1) is an experienced taphonomist, familiar with the different types of bone surface modifications produced by carnivores, yet has no prior experience working with geometric morphometric data. This analyst was chosen as a true analogy of a researcher in taphonomy who may be interested in applying GMM to their own analyses. The second analyst (A2), on the other hand, is a specialist in remote sensing and the handling of three-dimensional data, having worked in the past with geometric morphometric data. Finally, the third analyst (A3) is both a specialist in taphonomy and has the greatest extent of experience working with geometric morphometric data.

The main landmark configuration used for this study was that proposed by Aramendi et al. [55], consisting in a mixture of 17 fixed Type II and Type III 3D landmark points on the exterior and interior surfaces of each pit (Fig 2A). Each of the analysts performed collection of landmark data in separate locations, all using the free Landmark Editor software v. 3.0.0.6 [78]. Before data collection began, each of the participants were provided with the original paper [55] describing the landmark configuration to be used, alongside some general instructions on how to use the software. Throughout the process, none of the participants were allowed contact during the procedure so as to avoid influencing their placement of points.

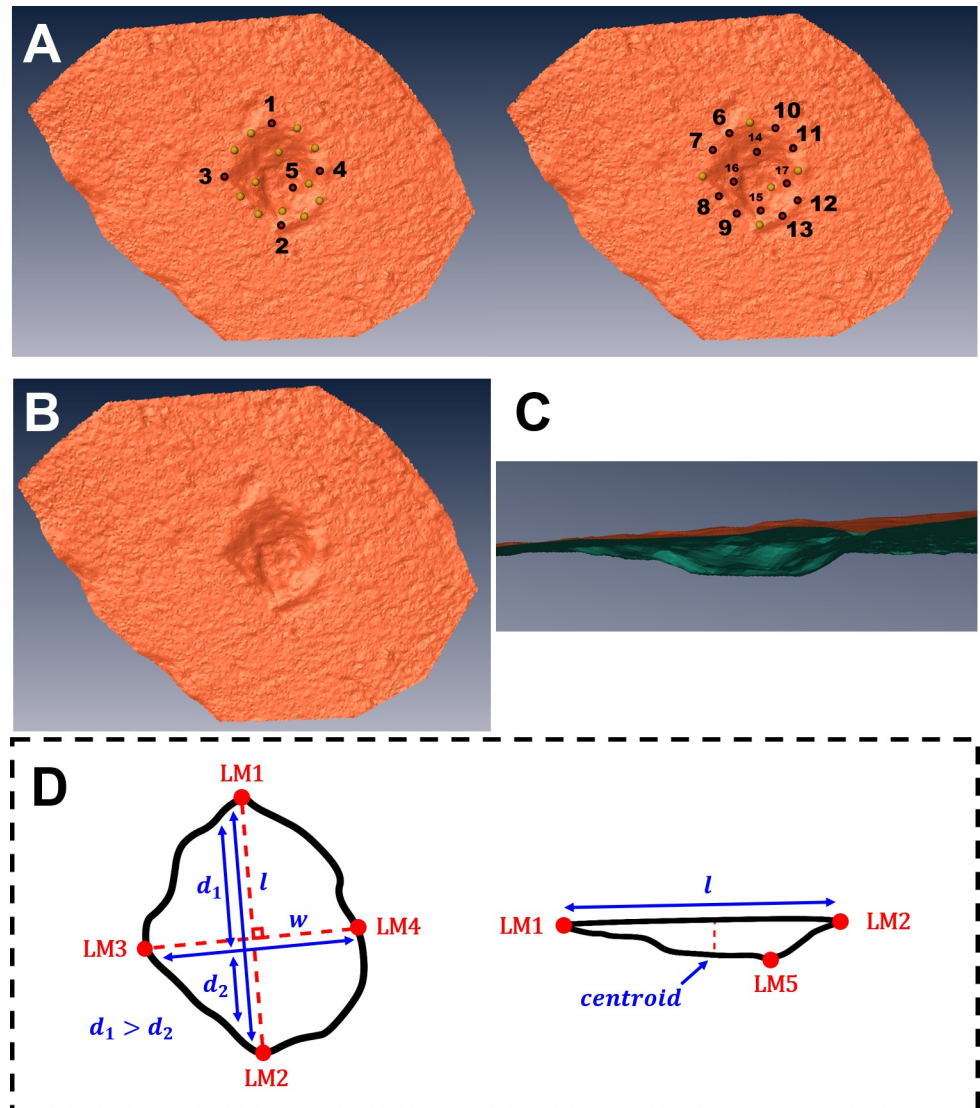
To ensure an identical orientation of the pit by each observer, thus facilitating any consequent analyses, the only additional instruction provided was on how to identify the position of landmark n° 1. Landmark n° 1 was thus defined as the point along the axis marking the maximum length furthest away from the perpendicular axis marking the maximum width.

Once 3D landmark data had been collected by each of the analysts, coordinates were formatted and imported into R v. 3.5.1. (<https://www.r-project.org/>) for further statistical processing.

## 2.3. Inter- and intra-observer analyses

As described by Cramon-Taubadel et al. [65], four main approaches exist for the comparison of inter and intra-observer errors. The first, and arguably most common approach, consists in projecting repeated measurements onto a common coordinate system [79]. This is most commonly performed through Generalised Procrustes Analyses (GPA) [41, 79, 80], consisting in a series of superimposition procedures, including scaling, rotation and translation. If after GPA the measured landmarks fall into a suitable range of variation, then the configuration can be considered reliable. The second method employs Euclidean distances between repeat measurements and the centroid of the configuration to assess the relative repeatability of each landmark [38]. The third approach consists in the repeated digitisation of a specimen that is held in a constant orientation, thereby directly assessing the error of a landmark coordinate through variation in landmark location [81]. The final approach consists in the partial superimposition of landmark configurations so as to reduce the impact of the “*Pinocchio effect*” [82], a phenomenon described by the distortion of measurements influenced by scaling procedures in GPA from the first and second approaches [65].

For the purpose of this study, a mixture of these approaches were employed on the analysis of observers studying different carnivore tooth pits. For the first part of these analyses, Euclidean distance calculations between points placed by each analyst were performed prior to any superimposition procedures. This was carried out on the raw coordinates obtained directly from the 3D models, providing a metrically accurate true approximation in millimetres to the margin of error of each landmark [81]. Additional calculations for inter-analyst variability in raw feature space was performed by calculating the Euclidean distance between the point placed by each analyst and the centroid between all three analysts [69].



**Fig 2. A detailed graphical description of the landmark models employed within this study.** (A) The original 17-Landmark model [51], placed on the pit presented in (B) and (C). (D) The detailed description of the precise locations of landmarks 1 to 5 in the revised version of this model.

<https://doi.org/10.1371/journal.pone.0240328.g002>

This was followed by the full superimposition of landmarks through GPA (*shape space*) and the calculation of mean reference configurations for each of the analysts separately, the carnivore samples, and for the entire sample as a whole. Principal Components Analyses (PCA) were then performed to examine the distribution of specimens in morphospace. If error among analysts is low, PCA should present a tight clustering of individuals regardless of the analyst recording the data. From the derived PC scores, Multivariate Analyses of Variance (MANOVA) tests were performed to statistically assess these variances. For MANOVA, either the “Hotelling-Lawley” or “Wilk’s Lambda” formulae were used for homogeneously and inhomogeneously distributed coordinates respectively. For homogeneity testing, both numerical tests and Quantile-Quantile (Q-Q) plots were used on landmark point clouds, PCA results and distance calculations to avoid issues induced by sample size [83].

From the fully imposed feature space, Procrustes distances were also calculated between a reference shape and the configuration of landmarks placed by each of the analysts. Here the mean configuration of all individuals was used as the reference shape. From these Procrustes distances, further statistical tests and metrics were used to evaluate the amount of error and variability produced. While most traditional studies of intra and inter-observer measurement errors represent within-group variance as an estimation of the within-observer mean squares or mean distances [e.g. 63, 66, 67], in this study robust statistical metrics were employed [83], using the Biweight Midvariance (BWMV) (Eqs 2–4), calculated via the median of errors and their consequent Median Absolute Deviation (MAD) (Eq 1).

$$MAD = m(|x_i - m_x|) \quad (1)$$

$$BWMV = \frac{n \sum_{i=1}^n a_i (x_i - m)^2 (1 - U_i^2)^4}{\left(\sum_{i=1}^n a_i (1 - U_i^2) (1 - 5U_i^2)\right)^2} \quad (2)$$

$$a_i = \begin{cases} 1, & \text{if } |U_i| < 1 \\ 0, & \text{if } |U_i| \geq 1 \end{cases} \quad (3)$$

$$U = \frac{x_i - m}{9MAD} \quad (4)$$

MAD is a calculation of the median ( $m$ ) of the absolute deviations from the data's median ( $m_x$ ) which is normally reported as a normalized value (NMAD). Normalisation of MAD is calculated through multiplying by the constant 1.4826. This is carried out to provide a scale factor of the true domain's standard deviation, using a 0.75 percentile as reference (Eq 5). Under this premise, MAD values are scaled by a factor of (Eq 5):

$$\frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \approx 1.4826 \quad (5)$$

Where  $\Phi^{-1}$  is the inverse cumulative distribution function of a standard Gaussian distribution [84].

NMAD has the distinct advantage of being a more robust measurement that is resilient to outliers in a dataset [83]. This is particularly useful for the analysis of datasets that do not follow a Gaussian distribution. The BWMV (reported as its square-root) is a further non-parametric measurement that additionally presents robustness of efficiency, proving a valuable substitute in non-parametric calculations [85]. NMAD and the square-root of the BWMV have proven to be important variables in accuracy control and system assessment analysis in remote sensing applications [83, 85–89]. Because of this, these metrics were employed as a substitute for standard deviation error calculations where Gaussian distributions were not detected, while the central tendency is reported as the median. Nevertheless, where deemed necessary by the normality of the dataset's distribution, further mean and standard deviation calculations were also reported.

Likewise, while most traditional analyses of observer measurement errors employ the use of Type II Analysis of Variance (ANOVA) calculations or Tukey's pairwise post-hoc comparisons [63, 64, 66–68, 90, 91, *inter alia*], these tests are statistically and mathematically conditioned by the underlying assumption of normality in the data's distribution [92, 93]. In cases where distances and errors do not fit a Gaussian distribution, parametric ANOVA tests are unlikely to truly capture the true nature of variances in sample distributions [94]. Because of this, when

non-Gaussian distributions were detected, the present study employed the use of an adaptation of ANOVA known as a non-parametric Kruskal-Wallis robust statistical test to assess differences in distribution medians [95; pg. 585, 587 & 598]. This test employs a rank-based H statistic, similar to the more traditional  $\chi^2$  or F statistic, that is able to “detect the kinds of difference of real interest” while making “only general assumptions. . . about the kind of distributions from which the observations come” [95; pg. 585].

Further testing considered the use of *Repeatability Measures* (RM); a calculation of the proportion of variance due to true variation among individuals [63, 66]. RMs can be defined as calculations involving the sum of squared distances ( $d^2$ ) through (Eq 6):

$$RM = \frac{\left( \frac{1}{n-1} \sum_{j=1}^J d_j^2 \right)}{\left( \frac{1}{n-1} \sum_{i=1}^I \sum_{j=1}^J d_{ij}^2 \right)} \quad (6)$$

where  $n$  is the total number of individuals in all of the samples,  $I$  is the total number of samples and  $J$  is the number of individuals in each of the samples.  $d$  values used to calculate this metric were the Procrustes distance. The RM metric is reported as a value ranging from 0 to 1; the latter indicating variance in measurements being attributable to the analyst under study ( $A_j$ ), rather than the true variation within the dataset [63].

From Procrustes distance calculations, further analyses included the generation of Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees. UPGMA results were analysed according to tree topologies, testing to see whether tooth pits clustered together according to the animal producing the tooth mark or the observer recording landmark data. Neighbour Joining (NJ) tree algorithms were also contemplated, nevertheless in this case NJ proved harder to read and was therefore discarded

Following this, two unsupervised Machine Learning algorithms were trained on fully superimposed coordinate data to test and see whether landmarks could be clearly defined through pattern recognition techniques. These analyses were performed in Python v.3.7 (<http://www.python.org>), using the SciKit-Learn package. Python applications were additionally implemented into the rest of the R workflow using the reticulate package and the RStudio development environment (<https://rstudio.com/>).

The first algorithm used was a Density-Based Spatial Clustering Algorithm with Noise (DBSCAN). DBSCAN employs a series of non-parametric mathematical and logical theories that differentiate between points that can be classed as reachable from its clustering core, and those that are considered as outliers or *noise points* ([96]:  $\{p \in D \mid \forall i: p \notin C_i \text{ where point } p \text{ in dataset } D \text{ is not associable to any of the clusters } C\}$ ). DBSCAN requires the definition of 2 hyperparameters; the  $\epsilon$  value defining the neighbourhood of a point and the minimum number of points (MinPts) that are required to form a cluster (i.e. the density of points). MinPts was set to 30, considering each of the analysts were required to digitise 30 individuals per sample for this study.  $\epsilon$  values were established in accordance with both the corresponding MinPts value, as well as the nature of the dataset's distribution. This value can be objectively defined mathematically using  $k$ -distance graphs according to the nearest neighbour [96], while the “elbow-joint” method can be used to establish the optimal  $\epsilon$  value for a dataset through calculating the point of maximum curvature in the plotted  $k$ -distance graph. The final  $\epsilon$  value for this study was thus calculated at 0.04 using the “kneed” algorithm in Python [97].

The second pattern recognition algorithm used was the non-parametric Mean-Shift algorithm (MS). MS, much like DBSCAN, works through calculating areas of high point density



with the use of a kernel function, also allowing for the mathematical detection of outliers [98–100]. MS only requires the definition a single hyperparameter, known as the bandwidth, which is used to define the size scale of the algorithm's internal kernel function. In this study, the bandwidth was established using Scikit-Learn's bandwidth estimation functions, employing a calculated quantile value of 0.05 to fine tune this estimation.

The objectives of using pattern recognition algorithms consisted in defining the degree of separation between landmarks. Landmarks that can be clearly defined by all three analysts should theoretically appear clearly differentiable by both algorithms with as little noise as possible. The results produced by these algorithms were evaluated by considering the number of identified clusters, the number of noise points, and the number of points that were correctly or incorrectly grouped with other points within each cluster.

The final part of this analysis performed calculations on partially superimposed landmarks, carried out by removing the scaling procedure from GPA (*form space*). This ensures that each of the configurations are centred and superimposed without creating statistical noise from the scaling process and thus factoring in the possible distortion that could be created by the “*Pinocchio effect*” in GMM [65, 82]. On partially superimposed data, PCA and Procrustes distance calculations were performed. From these distances, UPGMA trees were additionally calculated.

## 2.4. Semilandmark models

Once the accuracy and reliability of each landmark had been assessed, trials were performed removing landmarks most prone to inter and intra-analyst error and replacing them with semilandmarks (S1 Appendix) [35, 73, 101]. Semilandmarks are computed equidistant points that slide along the surface of the 3D model [73]. This overcomes many issues in accuracy when dealing with features that can only be represented by Type III fixed landmarks. Considering these landmarks are most prone to error and often incorrectly placed [61], substituting these points with semilandmarks presents a more efficient means of objectively quantifying complex morphological features such as curves and irregular surfaces [68].

For the process of placement of these semilandmarks, patches were drawn over each pit in the Landmark Editor v.3.0.0.6 software and different numbers of semilandmarks were computed. Trials were performed to find the optimal number of semilandmarks for tooth pit analysis, including 5x5, 6x6, 7x7, 8x8, 9x9 and 10x10 configurations.

The final part of this study consisted in both a full and partial GPA of the final landmark data to assess the variances in shape and form space captured by the newly updated landmark configurations. PCA were then performed to represent these differences graphically and to reduce the dimensionality of the dataset for further multivariate statistical processing. Comparison of PCA results reported the dimensionality of the consequent feature space ( $\mathbb{R}^n$ ), the proportion of variance represented by the first 2 PC scores as well as the cumulative proportion of variance represented by the first 10 PC scores. These were followed by MANOVA calculations to assess the degree of separation between wolf and dog samples according to the landmark configuration used. Finally, a Canonical Variance Analysis (CVA) was performed, alongside calculations of Mahalanobis and Procrustes distances and permuted ( $n = 999$ )  $p$ -values to evaluate the degree of separation between the samples.

## 3. Results

### 3.1. Inter- and intra-observer errors

Throughout the study, clear patterns emerged indicating those landmarks that were easier to locate and those that were more difficult. In all three trials, A1-3 noted difficulties on locating

Landmarks (LM) 14 through to 17, while LM1-5 were considered the most straightforward. A1-3 also noted that LM6-13 were easy to locate once LM1-4 had been clearly established. Nevertheless, *in-visu* inspection of each of the measurements taken on the 3D models documented that some analysts were more meticulous upon placing these points than others. Interestingly, A1 and A2 noted that wolf tooth marks were harder to process, observing that the 3D models obtained from this carnivore presented weaker topographical variations and were thus harder to analyse. Dog tooth marks on the other hand were noted as much deeper, easier to recognise and therefore digitisation procedures were easier overall. While the variations between these marks may be due to a number of variables dependent on the samples used, these go beyond the scope of the present study and are pending further future research.

**3.1.1. Raw measurement calculations.** Raw errors obtained directly from the 3D models indicate an error range of 1.95 mm across the entire data set (Table 1: Min error documented = 0.095 mm, Max = 2.045 mm). Mean and Median errors for all individuals were reported at 0.343 mm and 0.303 mm respectively, with a NMAD of 0.119 mm and square root of the BMWV at 0.121 mm. In all cases, no significant differences were detected when comparing the animal being processed by each analyst (Fig 3:  $\text{Chi}^2 = 0.027$ ,  $p = 0.8682$ ). Nevertheless, in accordance with the personal observations by A1 and A2, errors calculated on wolf tooth mark samples still present a slightly more varied distribution as opposed to dogs (Fig 3: Wolf Skewness = 4.28, Dog Skewness = 1.61). This indicates that wolves may be harder to process.

When separating samples according to the observer (Fig 3), a tendency can be observed for larger errors to be recorded when comparisons include the least experienced individuals, nevertheless, kruskal-wallis testing indicates no significant differences ( $\text{Chi}^2 = 0.98$ ,  $p = 0.61$ ).

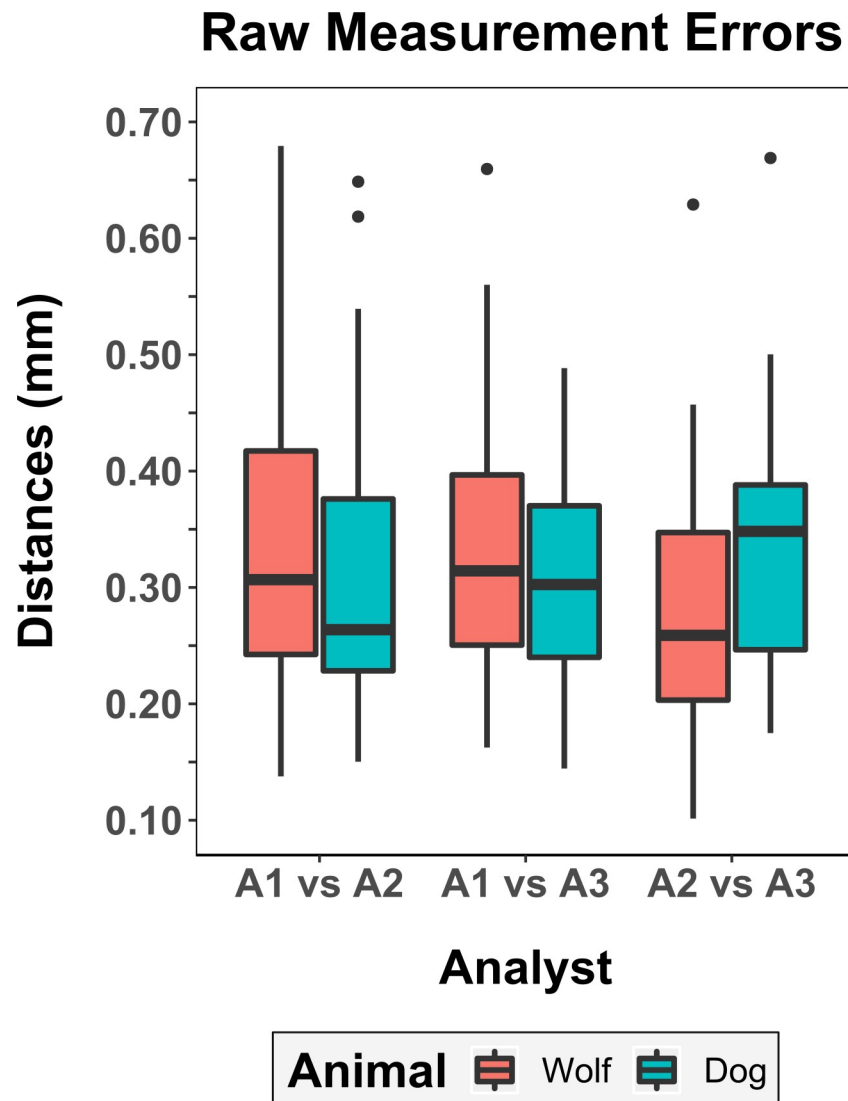
Further experimentation with the number of landmarks included reveal that as LM14-17 are removed, error calculations drop to less than half (Table 1: Median = 0.146 mm, NMAD = 0.077 mm, BMWV = 0.070 mm). Likewise, when only considering LM1-5, errors documented on wolf and dog samples are considered the most equivalent ( $\text{Chi}^2 = 0.0018$ ,  $p = 0.9658$ ). Moreover, when considering each of the LMs separately (Table 2), LM14 presents the highest recorded error among analysts (Median = 0.352 mm, NMAD = 0.292 mm, BMWV = 0.175 mm) while LM5 the lowest (Median = 0.175 mm, NMAD = 0.122 mm, BMWV = 0.075 mm).

When analysing inter-analyst variability according to each individual landmark (Fig 4, S1-S4 Tables of S1 File), errors begin to appear indicating a significant negative correlation between the experience of the analyst and the error produced (Kendall's  $\tau = -0.03$ ,  $p = 0.017$ ), proving the error to be significantly lower in analysts of more experience working with GMM than those in the process of learning. Nevertheless, differences in inter-analyst error are only

**Table 1. Overall documented metric errors according to the landmarks included in the study and the animal being analysed.** BMWV values are reported as the square root of the BMWV. All measurements are recorded in mm. For frame of reference; the studied pits range from 1.176 to 4.448mm in total length and 0.8382 to 3.5128mm in total width.

		Min	Sample Mean	Standard Deviation	Median	NMAD	BWMV	Max
LM 1–17	Absolute	<b>0.095</b>	0.343	0.227	<b>0.303</b>	0.119	0.121	<b>2.045</b>
	Wolf	<b>0.101</b>	0.363	0.291	<b>0.305</b>	0.128	0.130	<b>2.045</b>
	Dog	<b>0.095</b>	0.322	0.134	<b>0.303</b>	0.109	0.113	<b>0.966</b>
LM 1–13	Absolute	<b>0.048</b>	0.172	0.084	<b>0.146</b>	0.077	0.070	<b>0.464</b>
	Wolf	<b>0.053</b>	0.172	0.090	<b>0.143</b>	0.075	0.084	<b>0.448</b>
	Dog	<b>0.048</b>	0.171	0.078	<b>0.158</b>	0.067	0.070	<b>0.464</b>
LM 1–5	Absolute	<b>0.002</b>	0.172	0.107	<b>0.139</b>	0.083	0.092	<b>0.586</b>
	Wolf	<b>0.002</b>	0.177	0.120	<b>0.139</b>	0.081	0.097	<b>0.586</b>
	Dog	<b>0.009</b>	0.166	0.094	<b>0.140</b>	0.085	0.087	<b>0.509</b>

<https://doi.org/10.1371/journal.pone.0240328.t001>



**Fig 3. Boxplot representing the distribution of metric errors according to the analyst placing landmarks and the animal being studied.**

<https://doi.org/10.1371/journal.pone.0240328.g003>

revealed to be significant when considering wolves separately (S3, S4 Tables of [S1 File](#);  $\text{Chi}^2 = 11.6$ ,  $p = 0.003$ ), while landmark placement appears much more consistent among analysts when analysing dogs (S3, S4 Tables of [S1 File](#);  $\text{Chi}^2 = 0.36$ ,  $p = 0.8$ ).

**3.1.2. Fully superimposed calculations.** Fully superimposed coordinates present highly non-Gaussian spatial distributions on all accounts (Shapiro  $w > 0.78$ ,  $p < 2.2e-16$ ), while distribution of landmark centroids reveal interesting differences among the distribution of landmarks according to the analyst (Figs 5 and 6). In most cases, the greatest differences can be detected for LM14-17. Moreover, LM14-17 can clearly be seen to present the highest differences among analysts by observing point cloud distributions ([Fig 5](#)) as well as differences in centroid location ([Fig 6](#)). Likewise, LM1 to LM5 appear to be the most consistent.

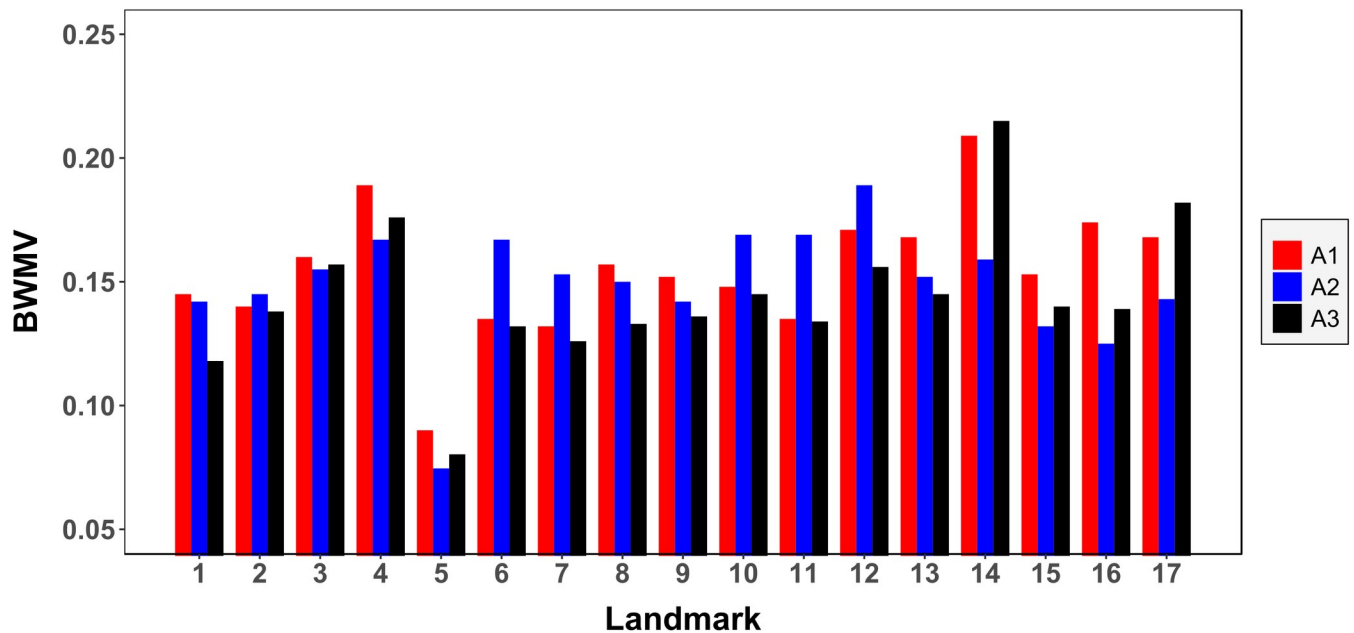
PCA results on superimposed coordinate data ([Fig 7](#) and S1 Fig of [S1 File](#)) reveal an inhomogeneous (Shapiro  $w = 0.99$ ,  $p = 0.0002$ ) wide dispersal of points with no clustering occurring according to the pit being studied (S1 Fig of [S1 File](#)). Clear differences among the analysts

**Table 2. Metric errors according to the landmark being placed.** BWMV values are reported as the square root of the BWMV. All measurements are recorded in mm. For frame of reference; studied pits range from 1.176 to 4.448mm in total length and 0.8382 to 3.5128mm in total width.

Landmark	Median	NMAD	BWMV
LM1	0.282	0.191	0.108
LM2	0.279	0.216	0.121
LM3	0.324	0.211	0.125
LM4	0.342	0.229	0.137
LM5	0.175	0.122	0.075
LM6	0.279	0.162	0.103
LM7	0.287	0.176	0.111
LM8	0.315	0.216	0.123
LM9	0.271	0.214	0.116
LM10	0.321	0.233	0.121
LM11	0.288	0.221	0.132
LM12	0.339	0.209	0.135
LM13	0.297	0.172	0.114
LM14	0.352	0.292	0.175
LM15	0.26	0.189	0.122
LM16	0.278	0.203	0.123
LM17	0.315	0.202	0.13

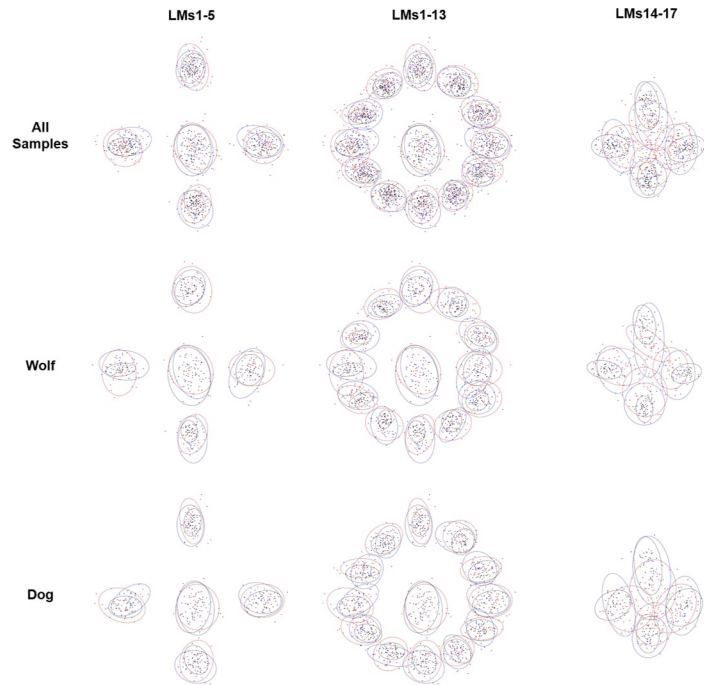
<https://doi.org/10.1371/journal.pone.0240328.t002>

can be observed with A1 and A2 occupying a vast area of the represented feature space while A3 presents the least amount of variation (Fig 7). MANOVA calculations obtained from this data indicate significant differences between observers ( $p = 0.001$ ). While significant differences are also detected between wolf and dog samples ( $p = 0.002$ ), the latter  $p$  values are slightly



**Fig 4. Inter-analyst raw measurement errors for each of the landmarks.** Measurements represent the square root of the BWMV of Euclidean distances (mm) from each of the points placed by the analyst to the landmark’s absolute centroid by all three analysts.

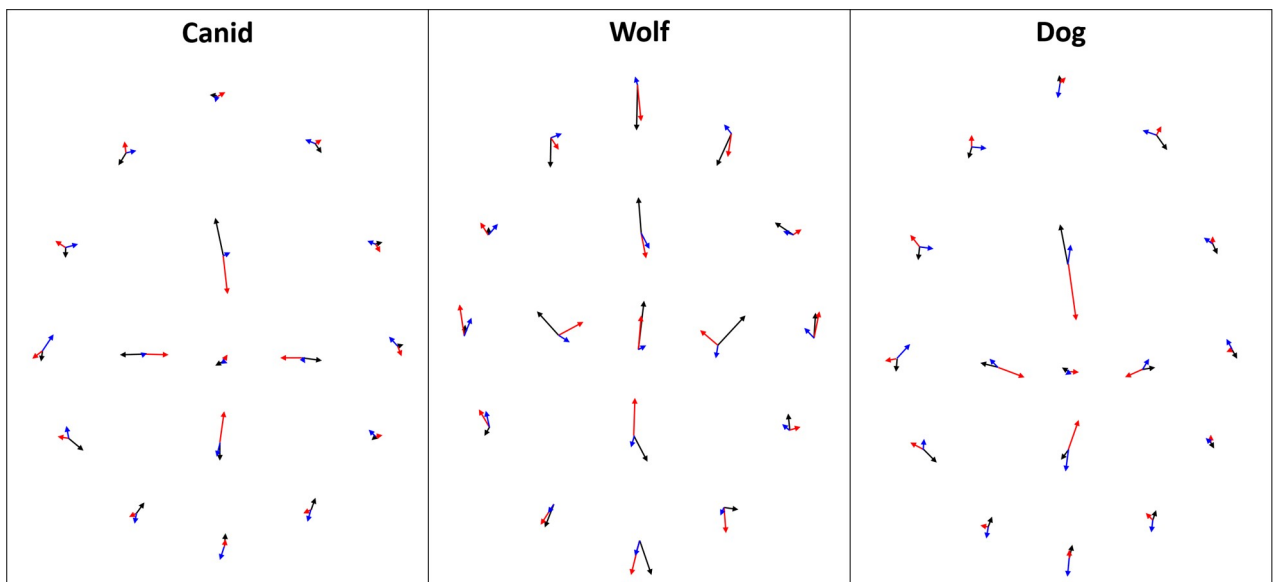
<https://doi.org/10.1371/journal.pone.0240328.g004>



**Fig 5. Fully superimposed landmarks in shape space and 95% Confidence ellipses calculated for spatial distributions (A1 = Red, A2 = Blue, A3 = Black).**

<https://doi.org/10.1371/journal.pone.0240328.g005>

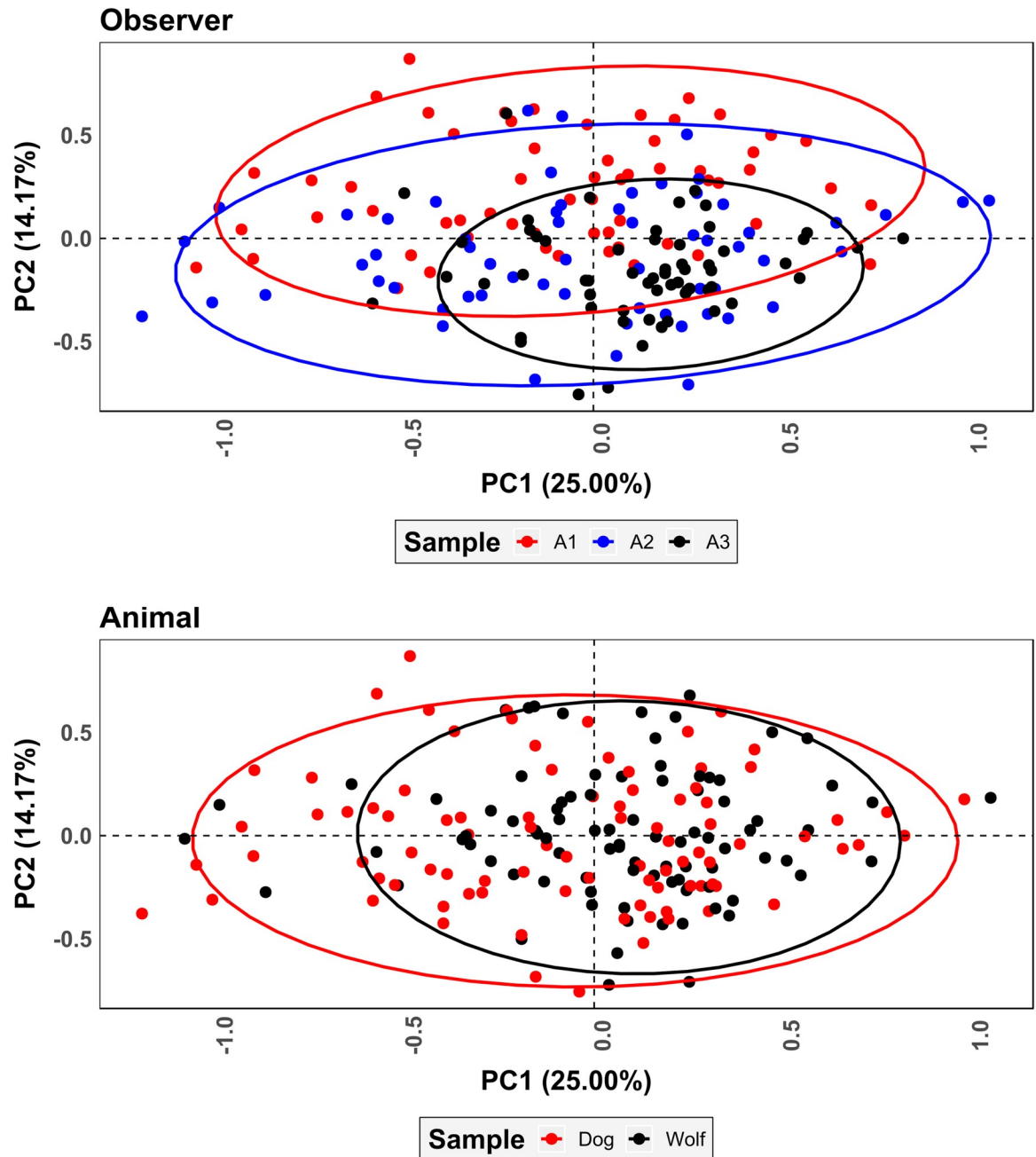
higher indicating the observer to have a stronger weight over variations detected in PCA than the carnivore.



**Analyst 1 – Analyst 2 – Analyst 3**

**Fig 6. Deviations of landmark centroids calculated from mean configurations in fully superimposed shape space.** The origin of each arrow marks the overall centroid for the corresponding landmark in the absolute mean configuration. The end of each arrow marks the position of the same landmark's centroid according to the analyst.

<https://doi.org/10.1371/journal.pone.0240328.g006>



**Fig 7. Scatter plot of principal components analysis results for fully superimposed landmark coordinates.** Ellipses represent 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0240328.g007>

Procrustes distance calculations (Table 3) reveal a mixture of significant and insignificant differences among the different analysts (Table 4), while comparisons with A1 and A3 show the greatest degrees of variation. These differences are equally evident when comparing carnivore samples separately. Overall distances highlight A3 to be the analyst closest to the absolute mean configuration presenting the least amount of intra-analyst variability (RM = 0.251). Nevertheless, this highlights how analysts with greater experience in GMM are significantly separable from those beginning (Table 3). RMs in general (A1 = 0.403, A2 = 0.357, A3 = 0.251)

**Table 3. Median, NMAD and square root of the BMWV values for Procrustes distances in fully superimposed shape space between each of the tooth pits processed by the analysts and the mean configuration.**

	Analyst	Median	NMAD	BWMV
Overall	A1	0.230	0.071	0.062
	A2	0.205	0.041	0.051
	A3	0.165	0.028	0.038
Wolf	A1	0.219	0.080	0.067
	A2	0.197	0.047	0.057
	A3	0.155	0.028	0.034
Dog	A1	0.233	0.061	0.054
	A2	0.208	0.044	0.052
	A3	0.171	0.024	0.037

<https://doi.org/10.1371/journal.pone.0240328.t003>

indicate that a great degree of variation detected within the samples is due to the analyst rather than the true variation produced by the carnivores. This is especially worrying considering how this variation increases according to the experience of the analyst. UPGMA tree topologies additionally reflect this, with absolutely no grouping according to the carnivore under study (S2 Fig of [S1 File](#)), and a slight agglomeration of A3 individuals grouped towards the middle of the tree's branches.

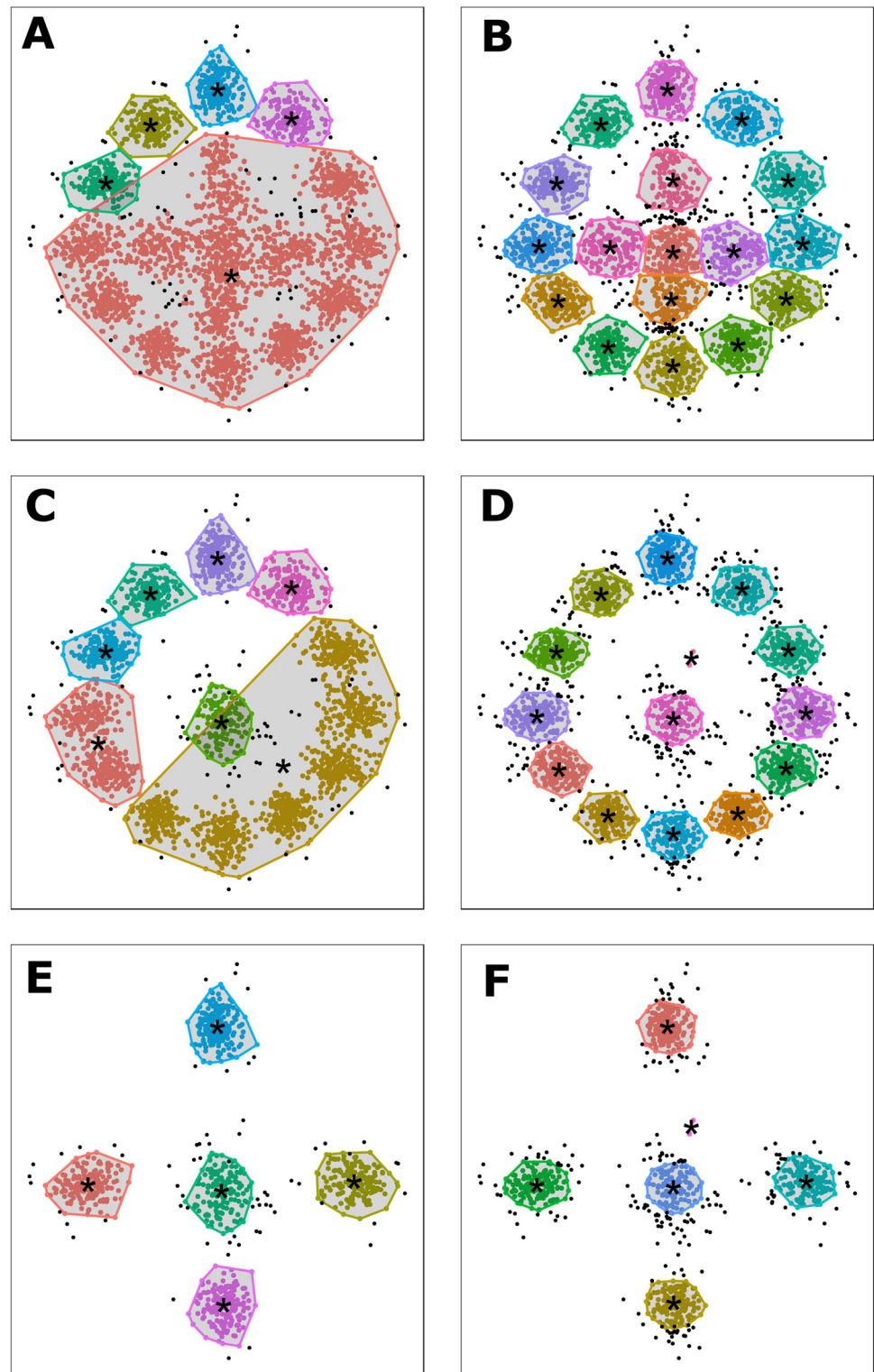
Pattern recognition results classify a high number of points as noise ([Fig 8](#) and S3 Fig & S5 Table of [S1 File](#); DBSCAN = 74, MS = 250), with the least amount of noise detected when only considering LM1-5 ([Fig 8E and 8F](#)). Likewise, the highest misclassification rates appear when LM14-17 are included in the models (S5 Table of [S1 File](#); DBSCAN = 2110, MS = 167). Visually, while MS is still able to clearly define the 17 landmarks ([Fig 8B](#)), misclassification and noise rates are considerably higher. DBSCAN on the other hand appears much more susceptible to confusion created by density, only clearly identifying points of interest when only LM1-5 are included ([Fig 8E](#)). In both cases, pattern recognition helps identify 5 clear points of interest that can be considered key landmarks ([Fig 8E and 8F](#)).

**3.1.3. Partially superimposed calculations.** Partially superimposed coordinates present highly non-Gaussian spatial distributions on all accounts (Shapiro  $w > 0.94$ ,  $p < 1.99e-10$ ). Procrustes distance calculations ([Table 5](#)) reveal insignificant differences among all three analysts ([Table 6](#)), while comparisons with A1 and A3 show the greatest degrees of variation. This is observed when comparing both carnivore samples together as well as separately ([Table 6](#)).

**Table 4. Kruskal-Wallis Chi<sup>2</sup> and p-values comparing Procrustes distances in fully superimposed shape space between each of the analysts and the overall mean configuration.** Calculations are provided for both carnivores as well as wolf and dog samples separately. Significant p-values under the standard alpha ( $\alpha$ ) threshold of 0.05 are marked in bold.

		Chi <sup>2</sup>	p-Value
Overall	A1 vs A2	1.56	0.21
	A1 vs A3	20.23	<b>6.9e-06</b>
	A2 vs A3	16.24	<b>5.6e-05</b>
Wolf	A1 vs A2	0.42	0.52
	A1 vs A3	9.00	<b>0.0027</b>
	A2 vs A3	7.97	<b>0.0047</b>
Dog	A1 vs A2	0.50	0.48
	A1 vs A3	10.29	<b>0.0013</b>
	A2 vs A3	7.89	<b>0.0050</b>

<https://doi.org/10.1371/journal.pone.0240328.t004>



**Fig 8.** Unsupervised DBSCAN (A, C, E) and MS (B, D, F) pattern recognition results trained on fully superimposed landmark coordinates. Detected clusters are marked by their corresponding colours and calculated convex hulls. Points classified as noise are marked in black. The centroid or point of interest for each cluster is marked by a single black \*. (A, B) LM1-17; (C, D) LM1-13; (E, F) LM1-5.

<https://doi.org/10.1371/journal.pone.0240328.g008>



**Table 5. Median, NMAD and square root of the BMWV values for Procrustes distances in partially superimposed shape space between each of the tooth pits processed by the analysts and the mean configuration.**

	Analyst	Median	NMAD	BWMV
Overall	A1	1.056	0.375	0.410
	A2	0.977	0.394	0.420
	A3	0.962	0.362	0.329
Wolf	A1	0.976	0.450	0.460
	A2	1.083	0.445	0.383
	A3	0.962	0.417	0.335
Dog	A1	1.040	0.318	0.375
	A2	0.959	0.351	0.447
	A3	0.946	0.382	0.383

<https://doi.org/10.1371/journal.pone.0240328.t005>

Considering each of the analyst's deviation from the mean reference configuration (Table 5), A3 is seen to present the smallest differences and the greatest consistency when placing landmarks (RM = 0.301). A2 (RM = 0.328) on the other hand is frequently seen separate while A1 presents the least amount of repeatability (RM = 0.383). Considering all three RM scores, a moderate degree of variation is detected within the sample due to the analyst rather than the true variation produced by the carnivore. UPGMA tree topologies additionally reflect this, with no grouping according to the carnivore under study (S4 Fig of S1 File).

PCA results on partially superimposed coordinate data (Fig 9 and S1 Fig of S1 File) yet again reveal an inhomogeneous (Shapiro  $w = 0.90$ ,  $p = 2.2e-16$ ) distribution of points. As opposed to fully superimposed data (Fig 7), however, clustering according to the pit being studied is stronger (S1 Fig of S1 File). At no point, however, is the grouping of points as strong as should be expected, with great variation within the represented feature space being product of the analyst. Once again A3 occupies a much more restricted proportion of feature space, however the difference between distributions is not as extreme. MANOVA calculations indicate equally significant differences between observers as in fully superimposed shape space ( $p = 0.001$ ), however differences between animal samples remain significant regardless of the observer ( $p = 0.001$ ).

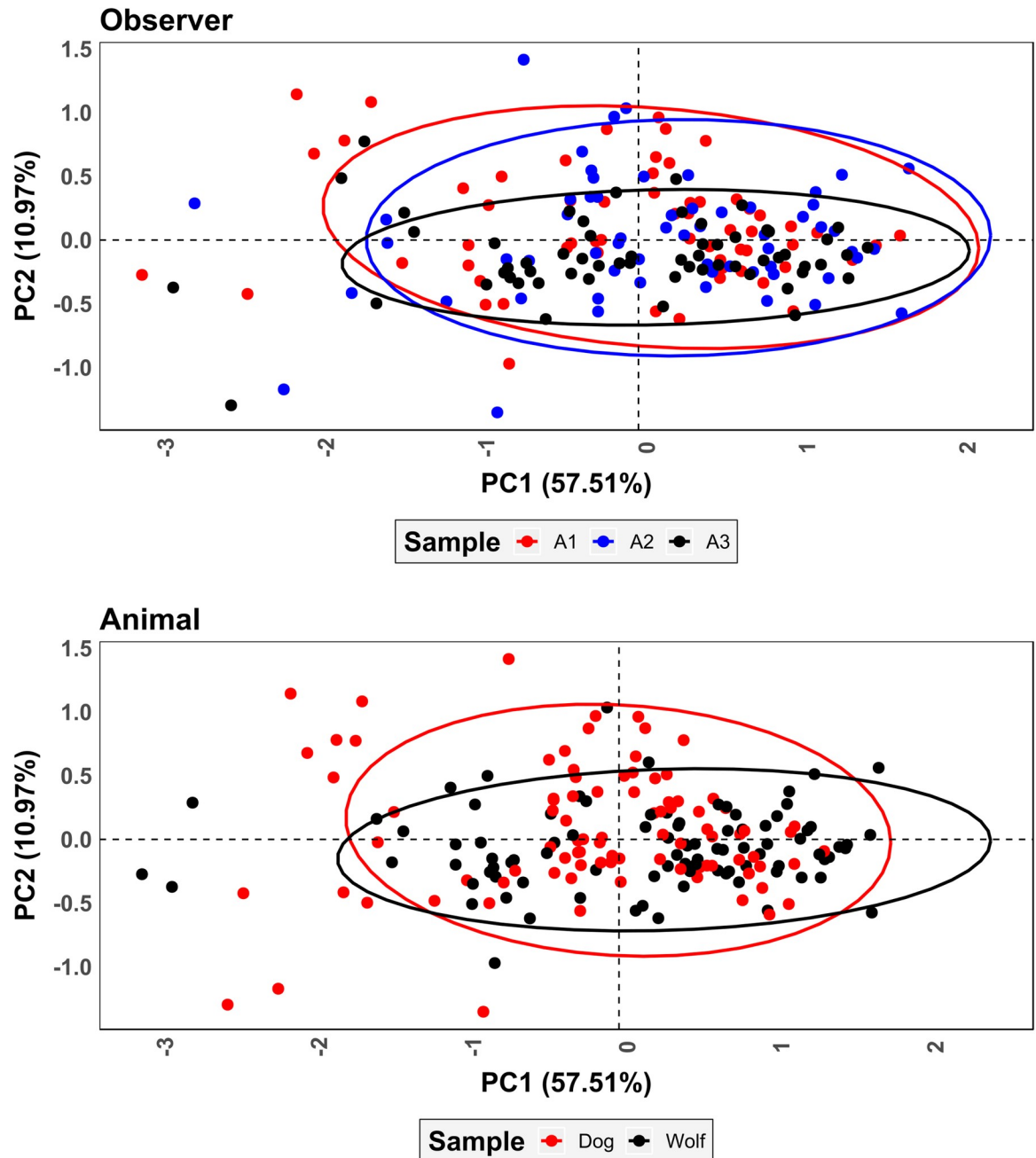
### 3.2. Overall evaluation of the 17-landmark model

Comparison of results obtained through raw coordinate data as well as fully and partially superimposed landmarks are sufficient in drawing two main conclusions about the 17

**Table 6. Median, NMAD and square root of the BMWV values for Procrustes distances in partially superimposed shape space between each of the tooth pits processed by the analysts and the mean configuration.**

	Analyst	Median	NMAD	BWMV
Overall	A1	1.056	0.375	0.410
	A2	0.977	0.394	0.420
	A3	0.962	0.362	0.329
Wolf	A1	0.976	0.450	0.460
	A2	1.083	0.445	0.383
	A3	0.962	0.417	0.335
Dog	A1	1.040	0.318	0.375
	A2	0.959	0.351	0.447
	A3	0.946	0.382	0.383

<https://doi.org/10.1371/journal.pone.0240328.t006>

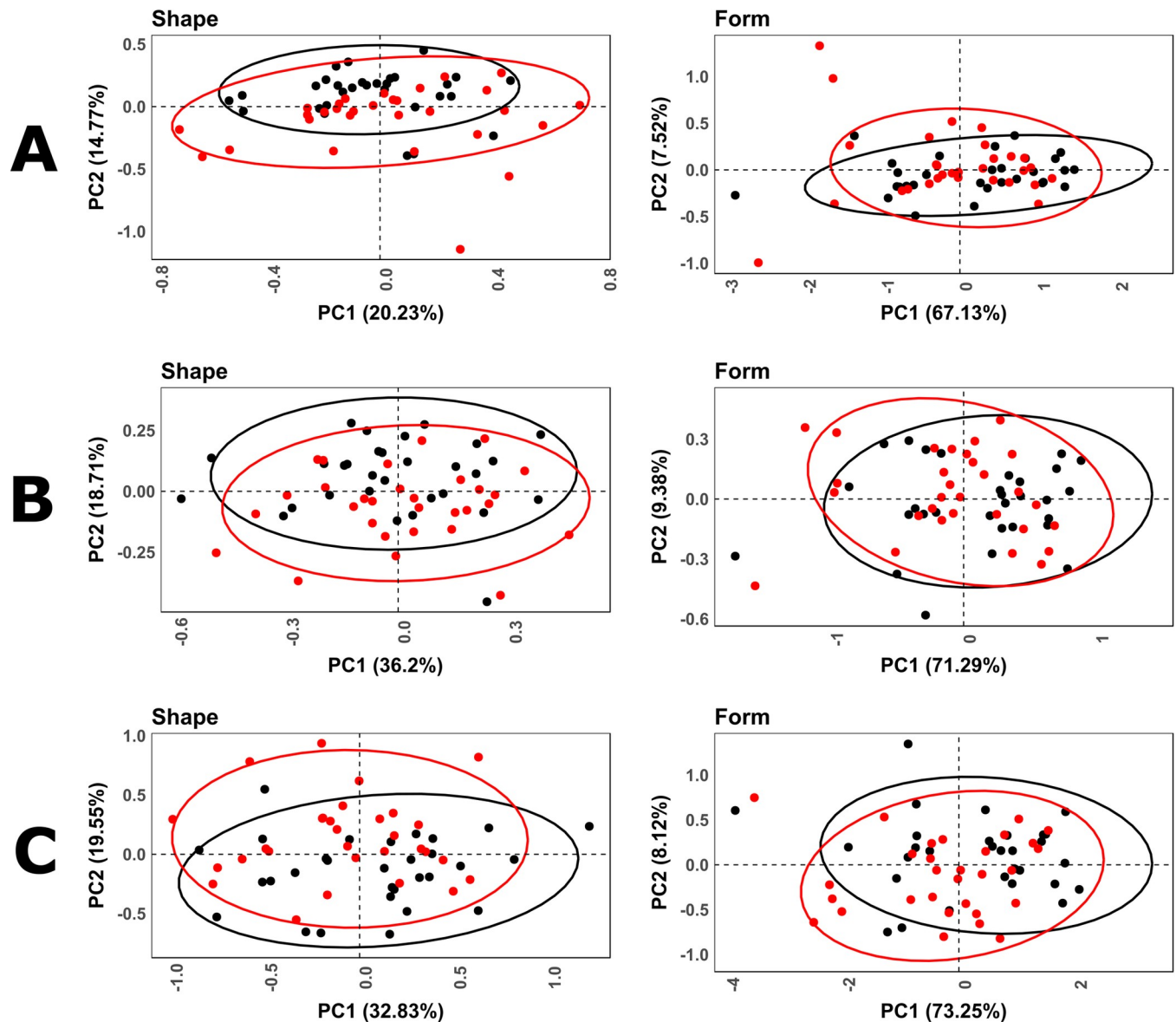


**Fig 9.** Scatter plot of principal components analysis results for partially superimposed landmark coordinates. Ellipses represent 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0240328.g009>

landmark model proposed by Aramendi et al. [55]; (1) experience in GMM is a highly conditioning factor in the amount of error produced within the model and (2) error is observed to increase upon the inclusion of LM14-17.

While the observed error does not necessarily correlate with the type of landmark being used (Kendall's  $\tau = -0.26$ ,  $p = 0.14$ ), Type III landmarks such as LM6-13 tend to present lower degrees of error because their location is influenced by their spatial relationship with Type II LM1-4. If we were to then exclude LM6-13, the weight of this correlation substantially



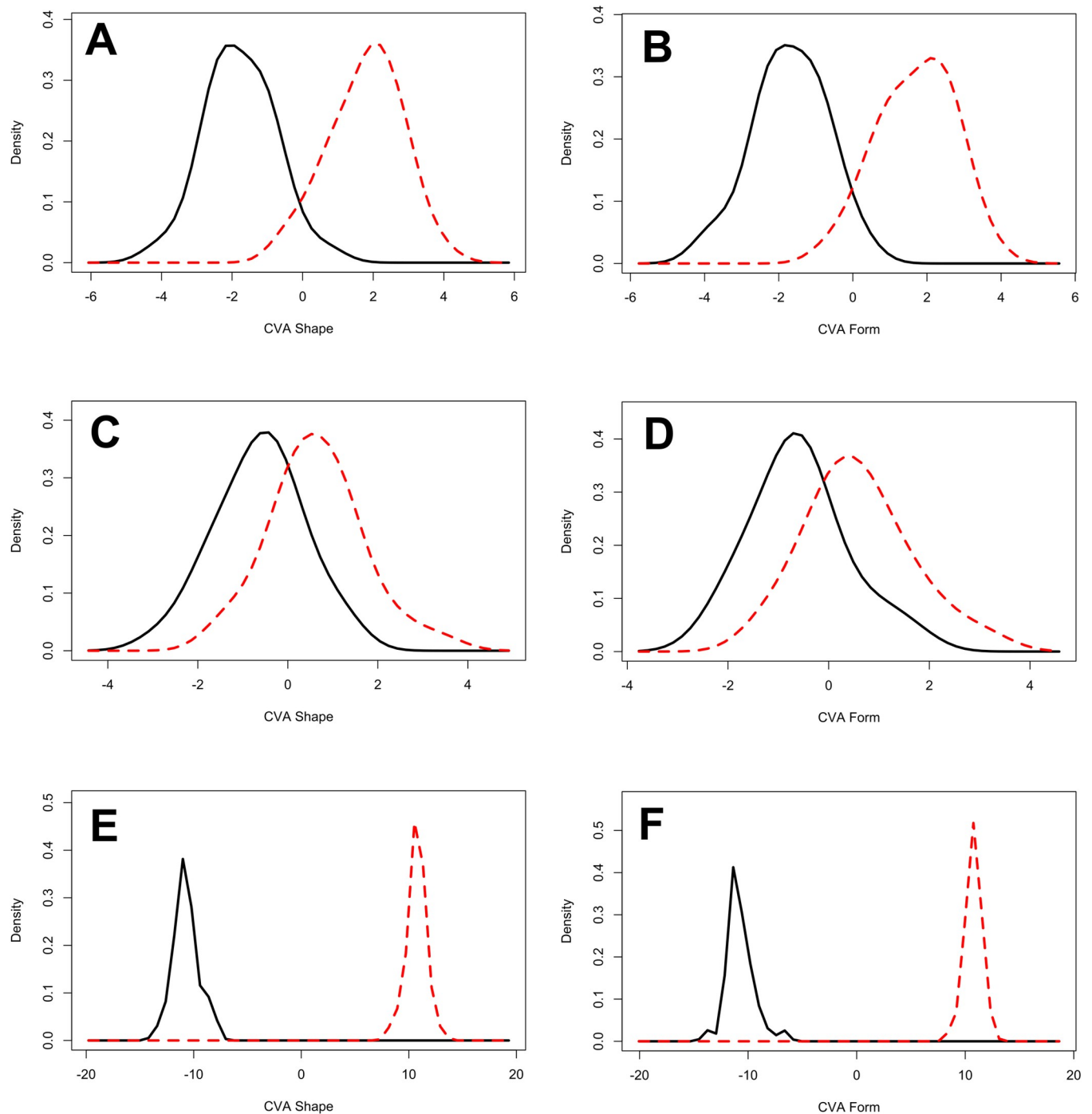
**Fig 10. Scatter plots for principal components analysis results comparing different landmark configurations for (Red) dog and (Black) wolf tooth pits.** (A) 17-Landmark model in (Left) shape and (Right) form. (B) 5-Landmark model in (Left) shape and (Right) form. (C) 30-landmark model in (Left) shape and (Right) form. Ellipses represent 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0240328.g010>

increases ( $\tau = 0.45$ ,  $p = 0.09$ ), arguing LM14-17 to be the greatest source of error. This logically agrees with observations made by multiple authors on the value and accuracy of Type III as opposed to Type II landmarks [60, 61, 72].

From a different perspective, while the relationship between experience and repeatability is relatively clear and logical to assume [62–68], upon analysing RMs alone no reliable  $p$  values can be used to quantify this correlation for this case study ( $\tau = -1$ ). When including the metric errors and overall distances between configurations, however, correlations can be successfully calculated supporting the aforementioned conclusions (Kendall's  $\tau = -0.03$ ,  $p = 0.017$ ).

Further in depth analysis of overall results presents A3 to be the most consistent in placing landmarks regardless of the sample. Nevertheless, other underlying factors may be



**Fig 11.** Distribution density plots for Canonical Variance Analyses in (A, C, E) shape and (B, D, F) form comparing (Red) dog and (Black) wolf tooth pits. (A, B) 17-landmark model. (C, D) 5-Landmark model. (E, F) 30-Landmark model.

<https://doi.org/10.1371/journal.pone.0240328.g011>

conditioning some of the results that are harder to quantify. For example, A1 and A2 took little time when placing landmarks (ca. < 1 min), taking a total of approximately 3 hours to process the entire sample. A3 on the other hand took longer processing each pit (ca. > 1 min), and took over twice as long to process the entire sample (approximately 6.5 hours). While the time

taken to process a model might not be the most important variable, here it can be argued that elements of an observer's personality are likely to take some effect in the results, as more meticulous individuals can be seen to take longer and may handle the sample differently. These observations, however, are more subjective and harder to model, stressing the need for analysts to take exceptional care when processing each individual pit so as to avoid these types of error.

Observations regarding the “*Pinocchio Effect*” present notable differences between fully superimposed, partially superimposed and raw data. With regards to geometric morphometric data, PCA appears to be the test where differences become most notable (S1 Fig of S1 File), with drastic changes to feature space appearing when including the scaling procedure in GPA.

Finally, considering the vast amount of data produced in this study, deductions can be made supporting the decision of removing LM14-17 from future analyses so as to reduce statistical errors. On the other hand, while LM6-13 do not produce significant errors in analyses, under the premise of reducing as much statistical noise as possible that may be dependent on the analyst, we have decided to remove all Type III LMs from the rest of this study as well. According to this decision, the definitive fixed Type II landmarks that will be used in our final model will be LM1-5.

### 3.3. Landmark model comparisons

PCAs (Fig 10A) obtained by the original 17-Landmark model present a non-polarized morphospace with relatively high overlapping of wolf and dog samples in both shape and form space. The defined feature space described by each of these PCAs presents a high number of dimensions in both shape ( $\mathbb{R}^{44}$ ) and form ( $\mathbb{R}^{51}$ ), representing low overall variability in the first 2 dimensions (shape = 35.39%, form = 74.65%). The first 10 dimensions represent a much higher cumulative proportion of variance (shape = 83.78%, form = 94.39%), with MANOVA results highlighting significant differences between carnivore tooth samples in shape ( $p = 0.001$ ), but not in form ( $p = 0.059$ ). CVA results obtained from this model present a much clearer separation between samples (Fig 11A and 11B), with both significant Procrustes ( $D = 0.072$ ,  $p = 0.002$ ) as well as Mahalanobis distances ( $D = 3.55$ ,  $p < 0.0001$ ).

Removal of Type III landmarks with the 5-Landmark model presents a slight increase in overlapping in PCA (Fig 10B) of carnivore samples in a reduced number of dimensions for both shape ( $\mathbb{R}^8$ , PC1&2 = 55.50%) and form ( $\mathbb{R}^{15}$ , PC1&2 = 80.67%). CVA additionally presents a significant reduction in both Procrustes ( $D = 0.064$ ,  $p = 0.04$ ) and Mahalanobis ( $D = 1.26$ ,  $p = 0.009$ ) distances with high overlapping across graphs (Fig 11C and 11D). Likewise, MANOVA tests indicate differentiation between carnivores to be less clear when using only LM1-5 (shape  $p = 0.029$ , form  $p = 0.039$ ).

Of the 6 semilandmark models tried and tested, immediate improvements in results were noted with the model employing a 5x5 net of semilandmarks with the 5 fixed Type II landmarks (30 in total, S1 Appendix). PCA results (Fig 10C) display similar graphical results to the original 17-landmark model, with a relatively highly overlapping non-polarized morphospace. Nevertheless, inclusion of semilandmarks display a clear increase in the cumulative degree of variation represented by each graph in both shape ( $\mathbb{R}^{59}$ , PC1&2 = 53.06%, PC1-10 = 92.60%) and form ( $\mathbb{R}^{60}$ , PC1&2 = 81.38%, PC1-10 = 97.91%). These results directly indicate that semilandmark based models are able to increase the chance of finding structure within the data by minimizing bending energy and thus reducing noise [72]. This in turn increases the possibility of finding statistical separation between the samples. These results are also seen through CVA results where clear separations between wolf and dog samples are noticeable (Fig 11E and 11F), with equally as significant Mahalanobis ( $D = 5.88$ ,  $p < 0.0001$ ) and Procrustes ( $D = 0.067$ ,  $p = 0.012$ ) distance values. While CVA results are likely to be overestimate

separation considering sample size and the increased number of variables when including semilandmarks [102, 103], other statistical tests such as those obtained through MANOVA are also able to highlight significant differences between samples in both shape ( $p = 0.004$ ) and form ( $p = 0.007$ ).

PCA results do not vary to a significant degree upon increasing the number of semilandmarks, with the total accumulative variance increasing by only 0.002% across the first 10 PC scores using 36 semilandmarks. Beyond 49 semilandmarks, percentage of represented variance even begins to decrease and MANOVA results also display poorer separation between samples. Likewise, CVA results on the other hand deteriorate with the inclusion of more semilandmarks, with less clear differences between groups. This is to be expected when considering the sample sizes available, and thus number of semilandmarks should be reduced [102, 103]. This concludes that the optimal statistical model for discerning between carnivore agencies remains to be the 30-landmark model presented here (S1 Appendix).

#### 4. Discussion and conclusions

Throughout the years, experimentation has become a fundamental component in archaeological and palaeontological research. Needless to say, with the ever-growing realisation of the impact true analogy has on experimental results, creating parallelisms with the fossil record is often difficult. Product of this is an increasing importance in the availability of experimental samples. This is especially apparent in geographic regions where the present day wild fauna is drastically different to that which existed over 10,000 years ago. In areas such as Europe, access to wild carnivores such as hyenas, lions and leopards is particularly difficult, highlighting the need for collaborative efforts in order to study the archaeological/palaeontological register effectively.

In archaeology/palaeontology, a possible response to this situation has been the collection and pooling of morphometric datasets via online resources [104, 105 *inter alia*], while some datasets are available through supplementary materials [e.g. 56–59]. Nevertheless, pooling of data from different sources is often conflictive, and induced errors are likely to increment if data collection is not strictly controlled [67]. Moreover, in many cases making 3D models readily available often requires a certain degree of computational costs and obtaining permissions to share these files can often be problematic. This latter point is unfortunately due to the additional issue of competitiveness amongst many research teams, making scientific progress much more politically, rather than empirically, oriented.

In fields of data science, the sharing of information can be considered indispensable, contributing to some of the most important advances in applied sciences. This is especially relevant when confronting issues of generalisation in predictive modelling [106]. With the advent of *Domain Adaptation* in computational learning [106, 107: pg. 526–531], the development of artificially intelligent Transfer Learning [108, 109], and the creation of One-Shot/Zero-Shot models [110–113], new possibilities for collaborative efforts in data science are now available to solve many research questions. Similar significant advances through *Multimodal* and *Federated Optimization* strategies present new possibilities for the collaborative integration of different datasets [114–116]. Sharing information and training classification models over multiple datasets in this manner presents the distinct advantage of protecting components of data privacy and overcoming issues presented by dataset centralisation [117].

Through more open-minded collaborations among teams, and greater efforts to make these learning strategies available, it may be possible for archaeological and palaeontological researchers to obtain similar success rates to other fields of science [118–120].

Nevertheless, in order for this to happen, careful attention must be paid to the statistical quality of the datasets being used. While tooth pit analyses have been able to reach > 95%

classification using GMMs [58, 59, 121], each of these studies have the distinct advantage of landmark data being collected by a single experienced individual [personal communication and direct participation]. Furthermore, while the original 17-landmark model was able to successfully draw conclusions from each of these studies (with the inclusion of [58, 59, 121, 122]), the integration of data produced here may increase the precision of these results. The results presented within this research thus represent a means of removing as much analyst-induced subjectivity as possible in the study of carnivore tooth pits. This is achieved through replacing weak landmark points with more precise computational semilandmarks. Nevertheless, as seen here, LM1-5 still present an important margin of error that must be confronted.

It is common knowledge that in-person training is the most valuable tool for obtaining optimal results in any task. Such an observation has been proven true not only for geometric morphometric research [121], but is also applicable to evolutionary theory [123, 124]. Likewise, the descriptive quality of any metric model is fundamental in its reproducibility. Under this premise, so as to ensure reproducibility and accuracy, a visual guide to the proposed methodological approach has been provided and is available at <https://vimeo.com/409256777>. This is further accompanied by a detailed description of the fixed landmarks (LM1-5) employed in this study (S1 Appendix), and with graphical representation in Fig 2D. Needless to say, on a practical note the time and care taken to process a 3D model is clearly an important component to consider. Although subjective, personal observations of the different analysts processing the models revealed clear differences between some individuals based on how thorough they were when handling the 3D models. These differences became increasingly obvious as time went by in each work session. While experience is a clear variable of importance, we strongly advise that even the most experienced observers carry out digitisation procedures as meticulously as possible over numerous digitisation sessions and taking full advantage of the tools in most 3D modelling software (e.g. measurement, zoom, pan, and rotating tools). While it is true that repeatability measures are likely to be higher in computerised procedures [66, 69], which is one of the clearest advantages to the methodological approach presented here, work ethic is still a valuable variable that needs to be controlled at all times.

Focusing specifically on the present study, human-induced landmark errors of the new proposed model can be conclusively defined at  $0.139 \pm 0.092 \in \{0.002:0.586\}$  mm (Median  $\pm \sqrt{\text{BWMV}} \in \{\text{Min:Max}\}$ ). This presents a total drop in error of 164  $\mu\text{m}$  from the original model which is extrapolated mainly from the statistically noisy landmarks LM14-17. As can be seen by the error range intervals, differences between samples are highly skewed, supporting our use for a robust statistical approach in defining the final model [83]. While the impact this margin of error has on statistical results is arguably not likely to affect the model's ability to separate between carnivore samples (as seen through MANOVA results), it is important to highlight the power that *form* has over *shape* in geometric morphometric analysis [47].

While Procrustes analyses are a valuable tool for most applied geometric applications [35, 65, 125, 126], it has been seen throughout our results how full Procrustes superimposition produces a notable statistical distortion that cannot be ignored in taphonomic GMM. This is direct evidence of the power the “*Pinocchio effect*” can have on geometric morphometric data (Current study: S1 Fig of S1 File; [65, 82]). Nevertheless, recent research into the power of computational learning algorithms highlight the value of *form* data in the predictive modelling of tooth pits [59]. This supports the removal of scaling processes for tooth pit classification tasks, additionally providing an interesting reflection on *form* and *shape* theory in modern morphometrics [127]. While it is also arguable that the increased presence of a *Pinocchio effect* is product of the microscopic scale being used [64, 66, 70], we insist on our recommendations that *form* data be used when working with decentralized pooled data for classification tasks on tooth pits.

Finally, the variability present in canid tooth marks is of increasing interest, considering the statistical noise they are frequently noted to produce [28, 56–59]. As seen here both qualitatively and quantitatively, wolf tooth marks in general are harder to process and frequently produce an increase in error (Deviations from absolute error =  $\{+4 \mu\text{m} : +14 \mu\text{m}\}$ , Table 1). This is primarily due to their more superficial nature and great variability which is an interesting topic to note and should be explored in detail in the future [58], till then, caution is certainly advised when working with this animal.

In conclusion, this study presents a detailed revision of the current methodological approach available for discerning carnivore agencies via tooth pit morphology. Through this methodological update, we present a means of reducing human induced error in taphonomic geometric morphometric data collection; facilitating the pooling of inter-institutional datasets for the training of classification models. While the theoretical and statistical reflections presented here are extensive, if these protocols are strictly followed and their true implications considered, tooth pit morphological analyses can have a promising future in applied scientific research. Under this premise, we predict lines of investigation of this type will be an encouraging development relevant in palaeontological, archaeological, forensic and even modern-day ecological sciences.

## Supporting information

### S1 Appendix.

(PDF)

### S1 File.

(PDF)

## Acknowledgments

Firstly, we greatly appreciate the meticulous reviews and comments made by one anonymous reviewer, Dr. Ros Coard and the editorial staff of PLoS ONE. We truly believe that their comments and suggestions have improved the quality of the present paper. Next, we would like to thank the members and staff of the Musée de l'Homme in Paris, including Florent Detroit, Antoine Balzeau and a special thanks to Martin Friess for their introduction to topics discussed and presented within this paper. Next, we would like to thank the helpful comments by Abel Moclán, Isabel Cáceres and Toni Pineda on earlier version of these studies. We would also like to acknowledge the help and support provided by the staff of the Cábarceno Natural Park and Santiago Borragan for providing the samples used in our experiments. Finally, we would also like to thank the TIDOP Group from the Department of Cartographic and Land Engineering of the Higher Polytechnics School of Avila, University of Salamanca, for the use of their tools and facilities.

## Author Contributions

**Conceptualization:** Lloyd A. Courtenay, Rosa Huguet.

**Data curation:** Lloyd A. Courtenay.

**Formal analysis:** Lloyd A. Courtenay.

**Funding acquisition:** Diego González-Aguilera.

**Investigation:** Lloyd A. Courtenay.

**Methodology:** Lloyd A. Courtenay, Darío Herranz-Rodrigo.



**Project administration:** Rosa Huguet, Diego González-Aguilera, José Yravedra.

**Resources:** Miguel Ángel Maté-González, José Yravedra.

**Software:** Lloyd A. Courtenay.

**Supervision:** Rosa Huguet, Diego González-Aguilera, José Yravedra.

**Validation:** Lloyd A. Courtenay, Darío Herranz-Rodrigo.

**Visualization:** Lloyd A. Courtenay.

**Writing – original draft:** Lloyd A. Courtenay.

**Writing – review & editing:** Lloyd A. Courtenay, Darío Herranz-Rodrigo, Rosa Huguet, Diego González-Aguilera, José Yravedra.

## References

1. Haynes G. Evidence of carnivore gnawing on pleistocene and recent mammalian bones. *Paleobiol.* 1980; 6(3), 341–351.
2. Haynes G. A guide for differentiating mammalian carnivore taxa responsible for gnaw damage to herbivore limb bones. *Paleobiol.* 1983; 9(2), 164–172.
3. Binford LR. *Bones: Ancient Men and Modern Myths.* New York: Academic Press Inc; 1981.
4. Brain CK. *Hunters or the Hunted? An Introduction to African Cave Taphonomy.* Chicago: University of Chicago Press; 1981.
5. Blumenschine RJ. Percussion marks, tooth marks and the experimental determinations of the timing of hominid and carnivore access to long bones at FLK Zinjanthropus, Olduvai Gorge, Tanzania. *J. Hum. Evol.* 1995; 29, 21–51.
6. Johnson DL, Haynes CV. Camels as taphonomic agents. *Quat. Res.* 1985; 24, 365–366.
7. Cáceres I, Esteban-Nadal M, Bennàsar M, Fernández-Jalvo Y. Was it the Deer or the Fox? *J. Archaeol. Sci.* 2011; 38, 2767–2774.
8. Hutson JM, Burke CC, Haynes G. Osteophagia and bone modifications by giraffe and other large ungulates. *J. Archaeol. Sci.* 2013; 40, 4139–4149.
9. Domínguez-Solera SD, Domínguez-Rodrigo M. A Taphonomic study of bone modification and of tooth-mark patterns on long limb bone portions by Suids. *Internat. J. Osteoarchaeol.* 2009; 19, 345–363.
10. Saladié P, Huguet R, Díez C, Rodríguez-Hidalgo A, Carbonell E. Taphonomic modifications produced by modern brown bears (*Ursus arctos*). *Internat. J. Osteoarchaeol.* 2011; 23(1), 13–33.
11. Backwell LR, Parkinson AH, Roberts EM, d'Errico F, Huchet JB. Criteria for identifying bone modification by termites in the fossil record. *Palaeogeog., Palaeoclimatol., Palaeoecol.* 2012; 337, 72–87.
12. Sanders WJ, Trapani J, Mitani JC. Taphonomic aspects of crow hawk-eagle predation on monkeys. *J. Human Evol.* 2003; 44, 87–105
13. Lloveras L, Cosso A, Solé J, Claramunt-López B, Nadal J. Taphonomic signature of Golden eagles (*Aquila chrysaetos*) on bone prey remains, *Historical Biology.* 2017 <https://doi.org/10.1080/08912963.2017.1319830>
14. Singer R. The “Bone Tools” from Hopefield. *Amer. Anthropol.* 1956; 58, 1127–1134.
15. Peterhans JCK, Singer R. Taphonomy of a lair near the peers (or skildegat) cave in Fish Hoek, Western Cape Province, South Africa. *S. Afr. Archaeol. Bull.* 2006; 61(183), 2–18.
16. O'Regan HJ, Kuman K, Clarke RJ. The likely accumulators of bones: Five cape porcupine den assemblages and the role of porcupines in the Post-Member 6 Infill at Sterkfontein, S. Afr. *J. Taphonomy.* 2011; 9(2), 69–87
17. Ames JA, Morejohn GV. Evidence of white shark, *Carcharodon carcharius*, attacks on sea otters, *Enhydra lutris.* *California Fish and Game.* 1980; 66, 196–209.
18. Deméré TA, Cerutti RA. A Pliocene shark attack on a cetotheriid wale. *J. Palaeontol.* 1982; 56, 1480–1482.

19. Njau JK, Blumenschine RJ. A diagnosis of crocodile feeding traces on larger mammal bone, with fossil examples from the Plio-Pleistocene Olduvai Basin, Tanzania. *J. Hum. Evol.* 2006; 50(2), 142–162 <https://doi.org/10.1016/j.jhevol.2005.08.008> PMID: 16263152
20. D'Amore DC, Blumenschine RJ. Komodo monitor (*Varanus komodensis*) feeding behaviour and dental function reflected through tooth marks on bone surfaces as the application of ziphodont paleobiology. *Paleobiol.* 2009; 35, 525–552.
21. Erickson GM, Olson KH. Bite marks attributable to *Tyrannosaurus rex*: Preliminary description and implications. *J. Vertebr. Paleontol.* 1996; 16(1), 175–178.
22. Rogers RR, Krause DW, Rogers KC. Cannibalism in the Madagascan dinosaur *Majungatholus atopus*. *Nature.* 2003; 422, 515–518. <https://doi.org/10.1038/nature01532> PMID: 12673249
23. Longrich NR, Horner JR, Erickson GM, Currie PJ. Cannibalism in *Tyrannosaurus rex*. *PLoS ONE.* 2010; 5(10).
24. Fernández-Jalvo Y, Andrews P. When humans chew bones. *J. Hum. Evol.* 2011; 60, 117–123. <https://doi.org/10.1016/j.jhevol.2010.08.003> PMID: 20951407
25. Saladié P, Rodríguez-Hidalgo A, Díez C, Martín-Rodríguez P, Carbonell E. Range of bone modifications by human chewing. *J. Archaeol. Sci.* 2012; 40(1), 380–397.
26. Domínguez-Rodrigo M, Piqueras A. The use of tooth pits to identify carnivore taxa in tooth-marked archaeofaunas and their relevance to reconstruct hominid carcass processing behaviours. *J. Archaeol. Sci.* 2003; 30, 1385–1391.
27. Delaney-Rivera C, Plummer TW, Hodgson JA, Forrest F, Hertel F, Oliver JS. Pits and pitfalls: taxonomic variability and patterning in tooth mark dimensions. *J. Archaeol. Sci.* 2009; 36(11), 2597–2608.
28. Yravedra J, Lagos L, Bárcena F. A taphonomic study of wild wolf (*Canis lupus*) modification of horse bones in Northwestern Spain. *J. Taphonomy.* 2011; 9(1), 37–65.
29. Andrés M, Gidna AO, Yravedra J, Domínguez-Rodrigo M. A study of dimensional differences of tooth marks (pits and scores) on bones modified by small and large carnivores. *Archaeol. Anthropol Sci.* 2012; 4(3), 209–219.
30. Yravedra J. A taphonomic perspective on the origins of the faunal remains from Amalda Cave (Spain). *J. Taphonomy.* 2011; 8(4), 301–334.
31. Domínguez-Rodrigo M, Barba R, Egeland CP. *Deconstructing Olduvai: A taphonomic study of the Bed I sites.* Netherlands: Springer; 2007.
32. Saladié P, Fernández P, Rodríguez-Hidalgo A, Huguet R, Pineda A, Cáceres I, et al. The TD6.3 faunal assemblage of the Gran Dolina site (Atapuerca, Spain): a late Early Pleistocene hyena den. *Hist. Blo.* 2019; 31(6), 665–683.
33. Rodríguez-Hidalgo A, Saladié P, Ollé A, Arsuaga JL, Bermúdez de Castro JM, Carbonell E. Human predatory behavior and the social implications of communal hunting based on evidence from the TD10.2 bison bone bed at Gran Dolina (Atapuerca, Spain). *J. Hum. Evol.* 2017; 105, 89–122. <https://doi.org/10.1016/j.jhevol.2017.01.007> PMID: 28366202
34. Pineda A, Saladié P. The Middle Pleistocene site of Torralba (Soria, Spain): a taphonomic view of the Marquis of Cerralbo and Howell faunal collections. *Archaeol Anthropol Sci.* 2019; 11, 2539–2556.
35. Bookstein FL. *Morphometric tools for landmark data.* Cambridge: Cambridge University Press; 1991.
36. Rohlf FJ, Marcus LR. A revolution in morphometrics. *Trends in Ecol. Evol.* 1993; 8, 129–132.
37. O'Higgins P. The study of morphological variation in the hominid fossil record: Biology, landmarks and geometry. *J. Anat.* 2000; 197, 103–120. <https://doi.org/10.1046/j.1469-7580.2000.19710103.x> PMID: 10999273
38. Singleton M. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). *J. Hum. Evol.* 2002; 42, 547–578. <https://doi.org/10.1006/jhev.2001.0539> PMID: 11969297
39. Guy F, Brunet M, Schmittbuhl M, Viriot L. New approaches in hominoid taxonomy: morphometrics. *Amer. J. Phy. Anthropol.* 2003; 121, 198–218.
40. Baylac M, Villemant C, Simbolotti G. Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biol. J. Linnean Soc.* 2003; 80(1), 89–98.
41. Lockwood CA, Lynch JM, Kimbel WH. Quantifying temporal bone morphology of great apes and humans: an approach using geometric morphometrics. *J. Anat.* 2002; 201, 447–464. <https://doi.org/10.1046/j.1469-7580.2002.00122.x> PMID: 12489757
42. Bastir M, Rosas A. Hierarchical nature of morphological integration and modularity in the human posterior face. *Amer. J. Phy. Anthropol.* 2005; 128, 26–34.
43. Galland M, Friess M. Three-Dimensional Geometric Morphometrics view of the cranial shape variation and population history in the New World. *Amer. J. Phy. Anthropol.* 2016; 28, 646–661.

44. Gunz P, Neubauer S, Falk D, Tafforeau P, Le Cabec A, Smith TM, et al. *Australopithecus afarensis* endocasts suggest ape-like brain organization and prolonged brain growth. *Sci. Adv.* 2020; 6(14), eaaz4729.
45. García-Medrano P, Ollé A, Ashton N, Roberts MB. The Mental Template in Handaxe Manufacture: New Insights into Acheulean Lithic Technological Behavior at Boxgrove, Sussex, UK. *J. Archaeol. Meth. Theor.* 2018; 26(1), 396–422.
46. Lycett SJ, Cramon-Taubadel N, Foley RA. A Crossbeam Co-Ordinate Caliper for the Morphometric Analysis of Lithic Nuclei: a Description, Test and Empirical Examples of Application. *J. Archaeol. Sci.* 2006; 33, 847–861.
47. Dryden IL, Mardia KV. *Statistical Shape Analysis*. New York: John Wiley and Sons; 1998.
48. Bookstein FL. Principal Warps: Thin Plate Spline and the Decomposition of Deformations. *Transactions on Pattern Anal. Mach. Intell.* 1989; 11(6), 567–585.
49. Richtsmeier JT, Lele SR, Cole TM. Landmark morphometrics and the analysis of variation. In: Hallgrímsson B. and Hall B.K. (Eds.) *Variation*. Boston: Elsevier Academic Press. 2005; 49–68.
50. Maté-González MÁ, Yravedra J, González-Aguilera D, Palomeque-González JF, Domínguez-Rodrigo M. Micro-photogrammetric characterization of cut marks on bones. *J. Archaeol. Sci.* 2015; 62, 128–142.
51. Courtenay LA, Yravedra J, Maté-González MÁ, Aramendi J, González-Aguilera D. 3D Analysis of Cut Marks using a New Geometric Morphometric Methodological Approach. *J. Archaeol. Anthropol. Sci.* 2019; 11, 651–665.
52. Yravedra J, Diez-Martín F, Egeland CP, Maté-González MÁ, Palomeque-González JF, Arriaza MC, et al. FLK-West (Lower Bed II, Olduvai Gorge, Tanzania): a new early Acheulean site with evidence for human exploitation of fauna. *Boreas.* 2017; 46(3), 486–502.
53. Komo L, Grassberger M. Experimental Sharp Force injuries to ribs: multimodal morphological and geometric morphometric analyses using micro-CT, macro photography and SEM. *Forensic Sci. Int.* 2018; 288, 189–200 <https://doi.org/10.1016/j.forsciint.2018.04.048> PMID: 29758447
54. Kieser J, Bernal V, Gonzalez P, Birch W, Turmaine M, Ichim I. Analysis of experimental cranial skin wounding from screwdriver trauma. *Int. J. Legal Med.* 2008; 122, 179–187 <https://doi.org/10.1007/s00414-007-0187-1> PMID: 17701196
55. Aramendi J, Maté-González MÁ, Yravedra J, Ortega MC, Arriaza MC, González-Aguilera D, et al. Discerning carnivore agency through the three-dimensional study of tooth pits: Revisiting crocodile feeding behaviour at FLK- Zinj and FLK NN3 (Olduvai Gorge, Tanzania). *Palaeogeog., Palaeoclimatol., Palaeoecol.* 2017; 488, 93–102.
56. Yravedra J, García-Vargas H, Maté-González MÁ, Aramendi J, Palomeque-González JF, Vallés-Iriso J, et al. The use of Micro-Photogrammetry and Geometric Morphometrics for identifying carnivore agency in bone assemblages. *J. Archaeol. Sci. Rep.* 2017; 14, 106–115.
57. Yravedra J, Aramendi J, Maté-González M.Á, Courtenay LA, González-Aguilera D. Differentiating Percussion Pits and Carnivore Tooth Pits using 3D Reconstructions and Geometric Morphometrics. *PLoS ONE.* 2018; 13(3), e0194324. <https://doi.org/10.1371/journal.pone.0194324> PMID: 29590164
58. Yravedra J, Maté-González MÁ, Courtenay LA, González-Aguilera D, Fernández-Fernández M. The use of canid tooth marks on bone for the identification of livestock predation. *Sci. Rptrs.* 2019; 9, 16301.
59. Courtenay LA, Yravedra J, Huguet R, Aramendi J, Maté-González MÁ, González-Aguilera D et al. Combining Machine Learning Algorithms and Geometric Morphometrics: a Study of Carnivore Tooth Marks. *Palaeogeog., Palaeoclimatol., Palaeoecol.* 2019; 522, 28–29
60. Valeri CJ, Cole TM, Lele S, Richtsmeier JT. Capturing data from three-dimensional surfaces using fuzzy landmarks. *Amer. J. Phy. Anthropol.* 1998; 107, 113–124.
61. Sholts SB, Flores L, Walker PL, Wärländer SKTS. Comparison of Coordinate Measurement Precision of Different Landmark Types on Human Crania using a 3D Laser Scanner and a 3D Digitiser: Implications for Applications of Digital Morphometrics. *Internat. J. Osteoarchaeology.* 2011; 21, 535–543.
62. Yezerniac SM, Lougheed SC, Handford P. Measurement error and morphometric studies: statistical power and observer experience. *Systemat. Bol.* 1992; 41(4), 471–482.
63. Arnqvist G, Martensson T. Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae.* 1998; 44(1–2), 73–96.
64. Blackwell GL, Bassett SM, Dickman CR. Measurement error associated with external measurements commonly used in small-mammal studies. *J. Mammology.* 2006; 87(2), 216–223.

65. Cramon-Taubadel N, Frazier BC, Lahr MM. The problem of assessing landmark error in geometric morphometrics: theory methods and modifications. *Amer. J. Phy. Anthropol.* 2017; 134, 24–35.
66. Muñoz-Muñoz F, Perpiñán D. Measurement error in morphometric studies: comparison between manual and computerized methods. *Annales Zoologici Fennici.* 2010; 47(1), 46–56.
67. Robinson C, Terhune CE. Error in geometric morphometric data collection: Combining data from multiple sources. *Amer. J. Phy. Anthropol.* 2017; 164(1), 62–75.
68. Shearer BM, Cooke SB, Halenar LB, Reber SL, Plummer JE, Delson E, et al. Evaluating causes of error in landmark-based data collection using scanners. *PLoS ONE.* 2017; 12(11), e0187452. <https://doi.org/10.1371/journal.pone.0187452> PMID: 29099867
69. Daboul A, Ivanovska T, Bülow R, Biffar R.; Cardini A. (2018) Procrustes-based geometric morphometrics on MRI images: An example of inter-operator bias in 3D landmarks and its impact on big datasets, *PLoS ONE.* 13(5):e0197675 <https://doi.org/10.1371/journal.pone.0197675> PMID: 29787586
70. Jamison PL, Ward RE. Brief Communication: Measurement Size Precision and Reliability in Craniofacial Anthropometry: Bigger is Better. *Amer. J. Phy. Anthropol.* 1993; 90, 495–500.
71. Gunz P, Mitteroecker P, Bookstein FL, Weber GW. Computer aided reconstruction of incomplete human crania using statistical and geometrical estimation methods. *Enter the past: computer applications and quantitative methods in archeology.* BAR Internat. Series. 2004; 1227, 96e98.
72. Gunz P, Mitteroecker P, Bookstein FL. Semilandmarks in three dimensions in: Slice DE (Ed). *Modern Morphometrics in Physical Anthropology.* New York: Plenum Publishers. 2005; 73–98.
73. Gunz P, Mitteroecker P, Neubauer S, Weber GW, Bookstein FL. Principles for the Virtual Reconstruction of Hominin Crania. *J. Hum. Evol.* 2009; 57, 48–62. <https://doi.org/10.1016/j.jhevol.2009.04.004> PMID: 19482335
74. Raina, R.; Ng, A.Y.; Killer, D. (2006) Constructing Informative Priors using Transfer Learning, Twentythird International Conference on Machine Learning. DOI: [10.1145/1143844.1143934](https://doi.org/10.1145/1143844.1143934)
75. Gidna A.; Yravedra J.; Domínguez-Rodrigo M. (2013) A cautionary note on the use of captive carnivores to model wild predator behavior: a comparison of bone modification patterns on long bones by captive and wild lions, *Journal of Archaeological Science.* 40:1903–1910
76. Mencha J.A.; Kreger M.D. (1996) Ethical and welfare issues associated with keeping wild mammals in captivity. In: Kleiman D.G.; Allen M.E.; Thompson K.V.; Lumpkin S. (Eds.) *Wild Mammals in Captivity: Principles and Techniques.* Chicago: The University of Chicago Press. 5–15
77. Maté-González MA, Aramendi J, Yravedra J, González-Aguilera D. Statistical Comparison between Low-Cost Methods for 3D Characterization of Cut-Marks on Bones. *Remot. Sens.* 2017; 9(9), 873.
78. Wiley DF, Amenta N, Alcantara DA, Ghosh D, Kil YJ, Delson E, et al. Evolutionary Morphing. *Proceedings of the IEEE Visualization 2005 (VIS'05).* 2005; 431–438.
79. O'Higgins P, Jones N. Facial Growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. *J. Anatomy.* 1998; 193, 251–272.
80. Viðarsdóttir US, O'Higgins P, Stringer C. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. *J. Anatomy.* 2002; 201, 211–229.
81. Corner BD, Lele S, Richtsmeier JT. Measuring precision of three-dimensional landmark data. *J. Quant. Anthropol.* 1992; 3, 347–359.
82. Chapman RE. Conventional Procrustes Approaches In: FJ Rohlf and FL Bookstein (Eds). *Proceedings of the Michigan Morphometrics Workshop.* Ann Arbor: University of Michigan Museum of Zoology. 1990; 2, 251–267.
83. Höhle J, Höhle M. Accuracy assessment of digital elevation models by means of robust statistical methods. *ISPRS J. Photogramm. Remot. Sens.* 2009; 64, 398–406.
84. Cichosz P. (2015) *Data Mining Algorithms: Explained using R.* Chichester: John Wiley & Sons
85. Rodríguez-González P, Garcia-Gago J, Gomez-Lahoz J, González-Aguilera D. Confronting passive and active sensors with Non-Gaussian statistics. *Sens.* 2014; 14, 13759–13777.
86. Hasan A, Pilesjö P, Persson A. The use of LIDAR as a data source for digital elevation models—a study of the relationship between the accuracy of digital elevation models and topographical attributes in northern peatlands. *Hydrol. E. Syst. Sci. Discuss.* 2011; 8(3), 5497–5522.
87. Herrero-Huerta M, Lindenbergh R, Rodríguez-González P. Automatic tree parameter extraction by a mobile LiDAR System in an urban context. *PLoS ONE.* 2018; 13(4), e0196004. <https://doi.org/10.1371/journal.pone.0196004> PMID: 29689076
88. Ariza-López FJ, Rodríguez-Avi J, González-Aguilera D, Rodríguez-González P. A new method for positional accuracy control for non-normal errors applied to airborne laser scanning data. *Appl. Sci.* 2019; 9, 1–18.

89. Rodríguez-Martín M, Rodríguez-González P, Ruiz de Oña Crespo E, González-Aguilera D. Validation of portable mobile mapping system for inspection tasks in thermal and fluid-mechanical facilities. *Remot. Sens.* 2019; 11, 1–19.
90. Heathcote GM. The magnitude and consequences of measurement error in human craniometry. *Canad. Rev. Phy. Anthropol.* 1981; 3, 18–40.
91. Hanihara T, Dodo Y, Kondo O, Nara T, Doi N, Sensui N. Intra- and Interobserver errors in facial flatness measurements. *Anthropol. Sci.* 1999; 107(1), 25–39.
92. Vargha A, Delaney HD. The Kruskal-Wallis test and stochastic homogeneity. *J. Educat. Behavioral Stats.* 1998; 23, 170–192.
93. Wilcox RR. *Introduction to Robust Estimation and Hypothesis Testing.* San Diego: Elsevier; 2005.
94. Tomarken AJ, Serlin RC. Comparison of ANOVA Alternatives under Variance Heterogeneity and Specific Noncentrality Structures. *Quant. Meth. Psychol.* 1986; 99(1), 90–99.
95. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 1952; 47(260), 583–621
96. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd Internat. Conference on Knowledge Discovery and Data Mining München Germany.* 1996. pp. 226–231.
97. Satopa V, Albrecht J, Irwin D, Raghavan B. Finding a “Kneede” in a Haystack: Detecting Knee Points in System Behavior. *IEEE 31st Conference on Distributed Computing Systems Workshops.* 2011. pp. 166–171.
98. Fukunaga K, Hostetler L. The estimation of the gradient of a density function with applications in pattern recognition. *IEEE Trans. Info. Theory.* 1975; 21(1), 32–40.
99. Cheng Y. Mean shift mode seeking and clustering. *IEEE Transactions on Pattern Anal. Mach. Intell.* 1995; 17(8), 790–799.
100. Comaniciu D, Meer P. Mean Shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Anal. Mach. Intell.* 2002; 24(5), 603–619.
101. Bookstein FL. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Img. Anal.* 1996; 1, 225–243.
102. Albrecht GH. Assessing the affinities of fossils using canonical variates and generalized distances. 1992; 7(4), 46–69.
103. Kovarovic K, Aiello LC, Cardini A, Lockwood CA. Discriminant function analyses in archaeology: are classification rates too good to be true? 2011; 38(11), 3006–3018. <https://doi.org/10.1016/j.jas.2011.06.028>
104. Adams JW, Olah A, McCurry MR, Potze S. Surface model and tomographic archive of fossil primate and other mammal holotype and paratype specimens of the Ditsong National Museum of Natural History Pretoria South Africa. *PLoS ONE.* 2015; 10, e0139800. <https://doi.org/10.1371/journal.pone.0139800> PMID: 26441324
105. Copes LE, Lucas LM, Thostenson JO, Hoekstra HE, Boyer DM. A collection of non-human primate computed tomography scans housed in MorphoSource a repository for 3D data. *Sci. Data.* 2016; 3(1), 1–8.
106. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* Cambridge: MIT Press; 2016.
107. Pratt LY. Discriminability-Based Transfer between Neural Networks. *Neural Info. Process. Syst.* 1993; 5, 204–211.
108. Pan SJ, Yang Q. A survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering.* 2010; 22(10), 1345–1359.
109. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv. Neural Info. Process. Syst.* 2014. pp. 33220–3328.
110. Fei-Fei L, Fergus R, Perona P. One-Shot Learning for Object Categories. *IEEE Transactions on Pattern Anal. Mach. Intell.* 2006; 28(4), 594–611.
111. Larochelle H, Bengio Y. Classification using discriminative restricted Boltzmann machines. *Proceedings of the 25th Internat. Conference on Mach. Learn.* 2008. pp. 536–543.
112. Palatucci M, Pomerleau D, Hinton G, Mitchell TM. Zero-Shot learning with semantic output codes. *Neural Inf. Process. Syst.* 2009; 22, 1410–1418.
113. Socher R, Ganjoo M, Sridhar H, Bastani O, Manning CD, Ng AY. Zero-Shot Learning through Cross-Modal Transfer. *Neural Inf. Process. Syst.* 2013. pp. 935–943.
114. Srivastava N, Salakhutdinov R. Multimodal Learning with Deep Boltzmann Machines. *J. Mach. Learn. Res.* 2014; 15, 2949–2980.

115. Konečný J, McMahan HB, Ramage D. Federated Optimization: Distributed Optimization Beyond the Datacenter. *Neural Info. Process. Sys.* arXiv: 151103575v1. 2015.
116. Konečný J, McMahan HB, Yu FX, Suresh AT, Bacon D, Richtárik P. Federated Learning: Strategies for Improving Communication Efficiency. *Neural Info. Process. Sys.* arXiv: 161005492v2. 2017.
117. McMahan B, Moore E, Ramage D, Hampson S, Agüera B. Communication-efficient Learning of Deep Networks from Decentralized Data. *Proceedings of the 20th Internat. Conference on Art. Intell. Statistics.* arXiv: 160205629v3. 2017.
118. AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: impact of cross-institutional training and testing. *Med. Phy.* 2018; 45(3), 1150–1158.
119. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed Deep Learning networks among institutions for medical imaging. *J. the Amer. Med. Infor. Association.* 2018; 25(8), 945–954.
120. Balachandar N, Chang K, Kalpathy-Cramer J, Rubin DL. Accounting for data variability in multi-institutional distributed deep learning for medical imaging. *J. Amer. Infor. Association.* 2020. pp. 1–9.
121. Aramendi J, Arriaza MC, Yravedra J, Maté-González MÁ, Ortega MC, Courtenay LA, et al. Who ate OH80 (Olduvai Gorge, Tanzania)? A geometric-morphometric analysis of Surface bone modifications of a *Paranthropus boisei* skeleton. *Quat. Internat.* 2019; 517, 118–130.
122. Arriaza MC, Aramendi J, Maté-González MÁ, Yravedra J, Stratford D. Characterising leopard as taphonomic agent through the use of micro-photogrammetric reconstruction of tooth marks and pit to score ration. *Hist. Blo.* 2019. pp. 1–10.
123. Lombao D, Guardiola M, Mosquera M. Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. *Sci. Rep.* 2017; 7(1), 1–14. <https://doi.org/10.1038/s41598-016-0028-x> PMID: 28127051
124. Pető R, Elekes F, Oláh K, Király I. Learning how to use a tool: mutually exclusive tool-function mappings are selectively acquired from linguistic in-group models. *J. Exp. Child Psychol.* 2018; 171, 99–112. <https://doi.org/10.1016/j.jecp.2018.02.007> PMID: 29567562
125. Crosilla F, Beinat A. Use of Generalised Procrustes Analysis for the photogrammetric block adjustment by independent models. *ISPRS J. Photogramm. Remot. Sens.* 2002; 56(3), 195–209.
126. González-Aguilera D, Gómez-Lahoz J, Muñoz-Nieto Á, Herrero-Pascual J. Monitoring the health of an emblematic monument from terrestrial laser scanner. *Nondestructive Test. Eval.* 2008; 23(4), 301–315.
127. Mitteroecker P, Gunz P, Windhager S, Schaefer K. A brief review of shape form and allometry in geometric morphometrics with application to human facial morphology. *Hystrix Italian J. Mammalogy.* 2013; 24(1), 59–66.