Title: Assessing the reliability of virtual reconstruction of mandibles

Running title: The reliability of mandibular reconstruction

Authors:

Ricardo Miguel **Godinho**[1]

Corresponding author

ricardomiguelgodinho@gmail.com

1) Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArHEB), University of Algarve, Faculdade das Ciências Humanas e Sociais, Universidade do Algarve, Campus Gambelas, 8005-139, Faro, Portugal

Paul **O'Higgins**[2, 3,]

2) PalaeoHub, Department of Archaeology and Hull York Medical School, University of York, Heslington, York YO10 5DD, UK

3) Centre for Forensic Anthropology, The University of Western Australia, 6009, Australia

Célia **Gonçalves**[1]

1) Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArHEB), University of Algarve, Faculdade das Ciências Humanas e Sociais, Universidade do Algarve, Campus Gambelas, 8005-139, Faro, Portugal

Abstract

Objectives:

Mandibular morphological variation is often used to examine various aspects of human palaeobiology. However, fossil and archaeological skeletal remains are often fragmented/distorted and so are frequently excluded from studies. This leads to decreased sample sizes and, potentially, to biased results. Thus, it is of interest to restore the original anatomy of incomplete/distorted specimens. Thin plate splines (TPS), commonly used in Geometric Morphometrics (GM), offer the prospect of reconstruction of missing parts and particularly of interest here, missing landmarks.

Materials and methods:

Here, the reliability of TPS based mandibular reconstruction is tested. To that end missing landmarks were simulated in originally complete hemimandibles. TPS was then used to restore the location of simulated missing data and the predicted landmarks were compared to the original ones.

Results:

Results show that error varies according to the number and location of estimated landmarks. Notwithstanding, estimation error is usually considerably smaller than the morphological differences between individuals from the same species.

Discussion:

TPS based reconstruction allows fragmentary mandibles to be used in studies of whole mandibular variation, provided the above mentioned caveats are considered.

Keywords

Geometric Morphometrics; Thin Plate Spline; Archaeology; Bioanthropology; Palaeoanthropology

Introduction

Aspects of human mandibular morphological variation have been used to estimate age at death and sex (Buikstra & Ubelaker, 1994; Ferembach, Schwidetzky, & Stloukal, 1980; Franklin, O'Higgins, & Oxnard, 2008; Franklin, Oxnard, O'Higgins, & Dadour, 2007), examine modes of subsistence (Galland, Van Gerven, Von Cramon-Taubadel, & Pinhasi, 2016; Katz, Grote, & Weaver, 2017; May, Sella-Tunis, Pokhojaev, Peled, & Sarig, 2018; Pinhasi, Eshed, & von Cramon-Taubadel, 2015; Pokhojaev, Avni, Sella-Tunis, Sarig, & May, 2019; von Cramon-Taubadel, 2011), analyse patterns of growth and development (Kelly et al., 2017; Singh, 2014; Wellens, Kuijpers-Jagtman, & Halazonetis, 2013) and inter-population differences (Buck & Vidarsdottir, 2004; Crevecoeur, Brooks, Ribot, Cornelissen, & Semal, 2016; Galland et al., 2016; Mounier et al., 2018). Such information is often subsequently used to examine sex-based differences in funerary anthropology (Baine, 2014; Härke & Belinskij, 2014; Kurila, 2015), palaeodemography (Chamberlain, 2006; Séguy & Buchet, 2014), and diet (Pearson, Grove, Özbek, & Hongo, 2013). However, archeological/fossil remains are commonly incomplete, distorted or eroded due to post-depositional processes, precluding their use (Arbour & Brown, 2014; R. M. Godinho & O'Higgins, 2017; Gunz, Mitteroecker, Neubauer, Weber, & Bookstein, 2009; O'Higgins, Fitton, & Godinho, 2019). Such exclusion of specimens may result in small sample sizes and so lead to biased results (Brown & Vavrek, 2015; Cardini & Elton, 2007; Cardini, Seetah, & Barker, 2015). Thus, reconstruction of incomplete specimens may be needed to restore the original morphology of specimens that would otherwise not be available for analyses, to increase sample size with the expectation of more reliable results (Arbour & Brown, 2014; Gunz et al., 2009).

Virtual reconstruction of skeletal/fossil remains is now common (Amano et al., 2015; Attard et al., 2014; Bauer & Harvati, 2015; Stefano Benazzi, Gruppioni, Strait, & Hublin, 2014; S. Benazzi, Stansfield, Milani, & Gruppioni, 2009; Bermúdez de Castro et al., 2015; Ricardo Miguel Godinho, Fitton, et al., 2018; R. M. Godinho & O'Higgins, 2017; Ricardo Miguel Godinho & O'Higgins, 2018; Jiménez-Arenas, Bienvenu, Toro-Moyano, Ponce de León, & Zollikofer, 2019; Rmoutilová et al., 2018; Senck, Bookstein, Benazzi, Kastner, & Weber, 2015) and uses multiple approaches to estimate the original morphology of incomplete specimens (Arbour & Brown, 2014; R. M. Godinho & O'Higgins, 2017; Gunz et al., 2009; Neeser, Ackermann, & Gain, 2009; O'Higgins et al., 2019). Geometric Morphometrics (GM) offers a toolkit that aids in the reconstruction of incomplete specimens (Gunz, Mitteroecker, Bookstein, & Weber, 2004; Gunz et al.,

2009; O'Higgins et al., 2019). To estimate the location of missing anatomical landmarks (LMs), GM approaches use multiple regression or Thin Plate Splines (TPSs; Gunz et al., 2004; Gunz et al., 2009). Multiple regression reconstruction is a statistical method that uses the relationship between different anatomical points to estimate the locations of missing landmarks (Arbour & Brown, 2014; Gunz et al., 2009; Neeser et al., 2009). In contrast, TPS reconstruction is a geometric approach that deforms (i.e., warps) a reference complete specimen to an incomplete target specimen to estimate the locations of the missing landmarks of the incomplete specimen (Arbour & Brown, 2014; Gunz et al., 2009; Neeser et al., 2009). Prior sensitivity analyses comparing these two most common reconstruction methods show that multiple regression may outperform the TPS approach (Gunz et al., 2004; Neeser et al., 2009), however it requires very large reference samples (Gunz et al., 2004; Neeser et al., 2009), ideally from the same population/group as the material being reconstructed. When such large samples are not available TPS may perform better than multivariate regression (Neeser et al., 2009). TPS requires only one reference specimen which can be a real specimen or an estimate of, e.g., the mean derived from the available sample. Further, when dense landmarking using conventional and semi-LMs is carried out, TPS provides very reliable reconstructions (Gunz et al., 2009; Senck et al., 2015).

While such studies compare the performance of those reconstruction methods and show which produce the best estimates (i.e., smaller residuals relative to the original location of the landmarks; Gunz et al. 2004; Neeser et al 2009), most do not demonstrate the impact of including or excluding reconstructed specimens in studies of morphological variation. The few studies that do, use crania and not mandibles (Arbour & Brown, 2014) and so the reliability of mandibular reconstruction and the impact of the inclusion of reconstructed mandibles in studies of mandibular morphological variation is yet to be tested. In this study we perform a sensitivity analysis of the reliability of mandibular reconstruction using the TPS method. This is crucial because the reliability of TPS based reconstructions depends on the geometry of the missing and surrounding regions and so varies between regions and bones with different morphologies (Senck et al., 2015). Further, this study also assesses if the inclusion of reconstructed specimens is beneficial or detrimental in the study of mandibular morphological variation.

Materials and methods

This study uses 11 archaeological mandibles, which were digitised using CT scanning. Of those, 8 originate from different Mesolithic Muge shell-middens (Portugal; Cabeço da Arruda: 3; Moita do Sebastião: 5) and 3 from the Chalcolithic burial site of Monte da Guarita 2 (Portugal). LM coordinates were collected from the mandibles to capture morphology and then subsequently removed and estimated using GM methods (see details below). Hemi-mandibles, and not the full mandibles, were landmarked because archaeological skeletal remains are often incomplete and because this study does not address asymmetry (in which morphological differences between sides are of relevance; e.g. Klingenberg, 2015, Klingenberg et al., 2002). The reliability of the estimation of original landmark locations was also assessed using GM methods (see details below), which are typically used to examine morphological variation and how morphology covaries with other underlying variables (Zelditch, Swiderski, Sheets, & Fink, 2012).

1.1 Digitisation, segmentation and landmarking

The mandibles were CT scanned (Toshiba, 120 kV, voxel size 0.348 * 0.348 * 0.3, revolution time 0.75 sec, spiral pitch factor 0.94) at the Faculty of Veterinary Medicine of the University of Lisbon. Standard protocols, previously described by Godinho and O'Higgins (2017; 2018) and Godinho et al. (2018; 2018), were used for segmentation of the hemi-mandibles in 3D Slicer (Fedorov et al., 2012). After segmentation the mandibles were exported as surface files and loaded into Landmark Editor (Wiley et al., 2005), where coordinates of a set of 21 landmarks (Table 1) were collected.

1.2 Simulating, estimating and assessing the reliability of missing data estimation

After collection of 3D LM coordinates, missing data were simulated (total simulations = 781) by removing known landmarks. To that end, the full sets of coordinates were imported into Microsoft Excel, where selected LMs were deleted and labeled as missing data. Simulation of missing LMs was based on what was observed in 67 incomplete mandibles from 7 prehistoric archaeological sites (Table 2). A tabulation was made of how many and which landmarks were most frequently missing per mandible

and this was used to indicate which landmarks to delete in subsequent simulations. Per complete mandible, 1 (4,8% of the full set of LMs) to a maximum of 11 (52,4% of the full set of LMs) missing LMs were simulated. The chosen maximum number of missing LMs was ~50% of the full set because previous studies show the estimation error of the TPS approach increases non-linearly and dramatically as the number of missing LMs increases (Arbour & Brown, 2014; Neeser et al., 2009). Moreover, our results show that when missing LMs are ~50% of the full set estimation, error approaches or overlaps the morphological differences between specimens. Thus, results of reconstructions of specimens based on ~50% or more are likely not meaningful (see details below). The first set (1 missing LM) simulated the absence of each landmark one by one, and so was performed 21 times per mandible (231 simulations). The other sets (2-11 missing LMs) simulated 5 different combinations of missing LMs per mandible, totaling 550 simulations (11 mandibles x5 combinations x10 sets). When 5 different combinations of missing LMs for a particular number of missing LMs were not observed in the archaeological material, additional combinations were generated randomly (list of combinations in Table 1 of Supplementary Information).

The data with simulated missing LMs were then imported into the R package Geomorph (Adams, Collyer, Otarola-Castillo, & Sherratt, 2015), where the original landmark locations were estimated using the "estimate.missing" TPS function. This function uses the raw LM coordinates of all complete specimens in a given sample to calculate a mean reference specimen (no scaling is carried out). Missing data are then estimated by interpolation using the TPS (Adams et al., 2015). After estimation of incomplete data, Microsoft Excel was used to calculate the residuals (i.e., the pairwise difference between the original and estimated coordinates) of each estimated LM. In this procedure, the absolute distance (in mm.) between the original and corresponding estimated LM coordinates is calculated. This gives an indication of the reliability of missing data estimation and which missing LMs are most difficult to estimate. In the sets with several missing LMs this analysis reports the sum of the pairwise distances between the true and estimated LMs in that set. This allows assessment of overall estimation error and if this increases with the number of missing LMs alone or if it also depends on which LMs are missing.

After these procedures, standard GM analyses ensued in the EVAN Toolbox (Weber & Bookstein, 2012) using all specimens (originally complete and incomplete specimens with estimated missing LMs). Both shape and form analyses were carried out.

In shape analysis, Generalized Procrustes Analysis (GPA) removes the effects of location, rotation and size (Zelditch, Swiderski, Sheets, & Fink, 2004). In form analysis information about size is included by augmenting the shape variables (which are obtained after GPA) with the natural logarithm of centroid size (Mitteroecker, Gunz, Windhager, & Schaefer, 2013). After superimposition using all specimens, Procrustes distances (PRDs) between complete and corresponding specimens with estimated LMs were calculated to examine the impact of missing LM estimation error on GM shape variables. Similarly, form space distances were calculated in the analyses of form. Lastly, the differences between the original complete specimens and the corresponding specimens with estimated missing data were visually assessed in shape and form PCA plots. This allows visualization of the impact of missing data estimation on standard GM morphological analysis.

Results

Our results show (Figure 1) that the LMs which are most difficult to estimate (i.e., those with larger residuals and so more error in the tests with single missing LMs) are the posterior lingual alveolus, gonion and the coronoid process. Such LMs are generally located at anatomical edges and lack nearby LMs (the coronoid process and gonion) or they present variable locations relative to the other LMs (the posterior lingual alveolus).

Estimation error tends to increase (i.e., residuals increase) as the number of missing LMs increases (Figure 2). Yet, sets with larger numbers of missing LMs may yield comparable or lower total error than sets with smaller numbers of missing LMs. For example, in Moita do Sebastião T, the sets missing 3-6 LMs show generally comparable error (below 20 mm., with the exception of an outlier in the set missing 6 LMs; see red box, Fig. 2. highlighting the relevant data). Similarly, the sets missing 4-6 LMs in Moita do Sebastião T3 show generally comparable error. Such cases occur when the sets with fewer missing LMs include one or more LMs that are more difficult to estimate (e.g., gonion or the coronoid process) or when complete anatomical regions are missing (e.g., all the LMs in the mandibular condyle).

The results of this study also show that PRDs (y axis) tend to increase as the number of missing landmarks increases (x axis; Figure 3). Yet, there are cases in which sets with more missing LMs produce comparable or lower Procrustes distances than sets with fewer missing LMs (see Fig. 3, red boxes for Cabeço da Arruda Ar I, Moita do

Sebastião T, Monte da Guarita 2 [134] id 97). Regardless, the impact of landmark estimation error on shape is generally smaller than the shape differences found among individuals. This is shown in Figure 3 by the red dotted lines in each plot that mark the lower limit of the among individual Procrustes distances. Overlap between inter-individual PRDs and PRDs deriving from estimation error occurs only in 5 cases where 11 LMs were estimated.

Consistently, results of the Procrustes form distances analysis (Figure 4) are comparable to those of the shape analysis (Figure 3). Form distances among individuals are almost always larger than those deriving from estimation error. Thus, almost all boxes representing estimation error lie below the dotted line that marks the lower limit of form distances among individuals (Inter-ind. box). The only exceptions are 5 of 11 cases in which 11 LMs were estimated. Form distances arising from estimation error tend to increase as the number of estimated landmarks increases. As with shape (Figure 3), there are cases in which sets with more missing LMs produce comparable or lower form distances than sets with fewer missing LMs (red boxes, Figure 4).

Shape analysis (Figure 5A) also shows that error resulting from landmark estimation is most frequently smaller than the differences found among individuals from the same sample. Thus, reconstructed specimens are frequently closer to the original specimen (specimen in the center of largest concentration of a given colour, which represents one specimen) than to other individuals (represented by the other colours). Figure 5B and C show further detail of the same PCA but showing the scores in that original PCA for just one specimen. The plots show a specimen located close to the centre of morphospace (Figure 5B) and a peripheral one (Figure 5C). In these plots, numbers represent how many missing LMs were estimated. The reconstructions of the specimen closer to the center of morphospace (Figure 5B; Moita do Sebastião T3) show a smaller spread (range of PC scores) than the more peripheral specimen (Figure 5C; Monte da Guarita 2 [134] id97). As expected, smaller error tends to emerge when fewer LMs are estimated (smaller numbers tend to cluster closer to the original specimen, labelled 0 and highlighted with the solid red circles). Yet, there are cases in which specimens with fewer estimated LMs show larger error than in those with more estimated LMs. Thus in Figure 5B, the plot of PC1 vs. PC2, a specimen with 7 missing landmarks is further from the original specimen than one with 11 missing landmarks (this and other examples are highlighted with red dotted circles in the plots of 5B and 5C). The insets in Figure 5A depict mandibular morphological variation along the PC axes. The insets in Figure 5B

and 5C show the morphological difference between the original complete specimen and the target reconstructed specimens. There are few morphological differences in Figures 5B and 5C (mostly imperceptible) when compared to the morphological variation in Figure 5A (visible in, e.g, the width of the ramus, height of the coronoid process, height of the mandibular body).

Form analysis (Figure 6) shows comparable results to shape analysis (Figure 5): errors resulting from estimation are generally smaller than inter-individual distances. Similar to the shape analysis, error of estimation of landmarks in individuals closer to the centre of the morphospace tends to be smaller than that in more peripheral individuals. This is visible in detail in Figure 6B and C. The spread, and so range of PC scores, is smaller in the more central individual (Figure 6B; Moita do Sebastião T3) and larger in the more peripheral individual (Figure 6C; Monte da Guarita 2 [134] id97). Error tends to be larger in sets with more estimated LMs, but sets with fewer estimated LMs may yield comparable or smaller error than sets with more estimated LMs. Such cases are visible in the plots. Thus, as in Figure 5B and C, some mandibles with smaller numbers of estimated landmarks (examples highlighted with red dotted circles) are more distant from the original complete specimen (labeled 0 and highlighted with a solid red circle) than those with larger numbers of estimated landmarks (examples highlighted with red dotted circles). Consistently with the visualization of mandibular shape variation (Figure 5), there are small morphological differences in Figures 6B and 6C (mostly imperceptible) between specimens when compared to the morphological variation in Figure 6A (visible in, e.g, the width of the ramus, height of the coronoid process, height of the mandibular body).

Discussion

This study shows that TPS based estimation of missing LMs is a reliable tool for the reconstruction of incomplete mandibles. This is consistent with other studies showing that TPS based reconstruction results in small errors when compared to inter-individual differences (Gunz et al., 2004; Gunz et al., 2009). Our results also show that including reconstructed mandibles is often preferable to excluding those specimens from morphological analyses. This is not only because excluding incomplete specimens would necessarily result in less information about morphological variation in a given sample due to decreased sample size, but also because differences due to estimation error in a given specimen are almost always smaller than those among specimens, in shape (Figure 3) and

in form space (Figure 4). Average shape PRDs deriving from estimated data range from 12% (in sets missing 1 LM) to 57% (in sets missing 11 LMs) of the average of PRDs between complete specimens. In form space, average PRDs deriving from estimated data range from 10% (in sets missing 1 LM) to 51% (in sets missing 11 LMs) of the average of PRDs between complete specimens. Consistently, the PCA plots (Figure 5-6) show that differences between individuals are frequently larger than differences due to estimation error. Moreover, estimation error induced morphological differences between originally complete reference specimens and reconstructed target specimens are very small (almost imperceptible) when compared to the full range of morphological variation of the full sample (Figure 5-6). These results imply that it is preferable to include specimens with estimated missing data than to exclude them. They are consistent with those of Arbour and Brown (2014), who showed that including crania with estimated missing LMs better reflects the shape variation of a sample than using only complete crania. Yet, there are some caveats when using TPS based missing data estimation that should be stressed.

Missing LMs located at anatomical edges, that lack other nearby existing LMs and curve dramatically (e.g., gonion and the coronoid process) are more difficult to estimate and so result in larger error (i.e., residuals). This is because TPS warping is guided by the common existing LMs and depends on the smoothness of the nearby anatomical regions (Gunz et al., 2004; Gunz et al., 2009; Senck et al., 2015). Thus, when there are no nearby existing LMs and/or the region of interest deviates markedly from the locations of adjacent landmarks, larger estimation error emerges. When missing LMs lie outside the convex hull formed by the remaining existing LMs this leads to larger error because missing data estimation extrapolates from the region in which LMs are present during TPS warping (Senck et al., 2015). LMs that present highly variable locations relative to other LMs are also difficult to estimate (Senck et al., 2015). This was the case for the posterior lingual alveolous which, although not located at an anatomical edge, is highly variable relative to other LMs due to the presence/absence of the third molar.

Estimation error tends to be larger in specimens that occupy the periphery of morphospace (i.e., that are located closer to the limits of the PC axes). This is because specimens that are closer to the center of morphospace are morphologically more similar to the mean specimen which was used as the reference specimen for TPS interpolation. Specimens that differ more from the mean (reference), unsurprisingly show larger error (Gunz et al., 2009; Neeser et al., 2009; Senck et al., 2015). Thus, to minimise estimation

error it is preferable to use sample/population specific specimens as the TPS reference (Gunz et al., 2009; Neeser et al., 2009; Senck et al., 2015). In practice this means that when a sample contains complete and incomplete specimens from several morphologically distinct populations it is preferable to use, as the TPS reference for reconstructing any particular mandible, the mean of complete specimens from the same population rather than the mean of all complete specimens from all populations.

As expected, LM residuals are larger in specimens with more missing LMs (Figure 2, PRDs in Figure 3-4 and the PCA plots of Figure 5-6). Yet, there are several cases in which specimens with fewer missing LMs yield comparable or larger error than specimens with more missing LMs. This is the case when the former mainly include missing LMs that are particularly difficult to estimate (e.g., gonion, coronoid process, posterior lingual alveolous) or when all LMs of an entire anatomical region are lacking (e.g., the mandibular condyle). In the latter case, this results in the absence of nearby LMs, sub-optimal warping of the TPS and so in larger error. Further, increased estimation error is to be expected in missing regions that curve dramatically, such as the mandibular condyle, because estimation efficacy relates to the smoothness of surrounding regions (Gunz et al., 2009; Senck et al., 2015).

Considering that estimation error depends on multiple variables, it is not possible to provide specific guidelines to decide if an incomplete specimen should be included (using missing data estimation) in a study that apply to all cases. Notwithstanding, it should be stressed that the decision of including (or excluding) a reconstructed specimen should consider not only the number of missing LMs but also which LMs are missing, and if there are any other nearby LMs to optimize estimation (Gunz et al., 2009; Senck et al., 2015). When all LMs are missing from a given region (e.g., condyle or mid-line) estimation error may be meaningful in shape/form analysis (see details above). This is more likely in intra-specific studies of closely related/morphologically close specimens. In such cases it may be prudent to exclude LMs from those regions or, if those regions are crucial for the study, specimens from the analyses. If many LMs are missing, then estimation error may be greater than inter-individual differences. This was the case in PRDs in some cases when ~50% of the LMs were missing and in PCA with fewer LMs missing. This is, again, most likely in intra-specific studies with samples of closely related specimens. When estimation error induces larger differences than inter-individual distances it may also be prudent to exclude specimens from analyses.

Previous studies show reconstructions using semi-LMs provide better results than those using only conventional LMs because denser anatomical coverage is achieved resulting in smoother reconstructions (Gunz et al., 2009). Despite that, this study only used conventional LMs because archaeological/fossil remains are frequently fragmented/incomplete and so correspondence between regions that do not display conventional LMs may be unreliable in some cases. When correspondence is reliable, the use of both conventional and semi-LMs may provide better results in terms of estimation of the morphology of missing regions (Gunz et al., 2009). Further, when comparing populations which are clearly morphologically distinct, estimation error may be negligible (Senck et al., 2015). Notwithstanding, this study shows that reconstruction based on conventional LMs alone is reliable (provided the above discussed caveats are considered) and so is a useful way of increasing sample size. This is critical because small sample sizes often preclude statistical testing and so the robustness of the results and analysis in a given study. This is clearly illustrated in recent studies showing that using small samples (n) relative to the number of variables (p = number landmarks X number of dimensions) in GM analyses results in p/n ratios that subsequently cause biased results that, e.g., artificially create/increase inter-group differences when using between group PCA (Bookstein, 2017, 2019; Cardini, O'Higgins, & Rohlf, 2019). While GM studies present various p/n ratios (Cardini et al., 2019), most do not approximate recommended ratios because that would require very large samples that are often unavailable (Bookstein, 2017). Thus, estimation of missing landmarks in damaged specimens is one way of increasing sample size and of mitigating limitations imposed by sample size in GM studies where additional, incomplete specimens are available.

Acknowledgements

Ricardo Miguel Godinho is funded by the European Regional Development Fund (FEDER) via the Programa Operacional CRESC Algarve 2020, of Portugal2020 (project ALG-01-0145-FEDER-29680). Célia Gonçalves is funded by the Portuguese Foundation for Science and Technology (FCT; contract reference DL 57/2016/CP1361/CT0029). This research was also partially funded by the Archaeological Institute of America (The Archaeology of Portugal Fellowship). Thanks are also due to Dr. Miguel Ramalho and José António Moita for granting access to the skeletal remains of Muge, António Valera for granting access to the skeletal remains from Monte da Guarita 2 and Sara Ramos for

granting access to the skeletal remains from Monte do Carrascal 2, Prof. Sandra Jesus and Dr. Óscar Gamboa for CT scanning at the Faculty of Veterinary Medicine of the University of Lisbon.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of interest statement

The authors declare no competing interests.

References

Adams, D., Collyer, M., Otarola-Castillo, & Sherratt, E. (2015). Package "geomorph".

Amano, H., Kikuchi, T., Morita, Y., Kondo, O., Suzuki, H., Ponce de León, M. S., . . . Ogihara, N. (2015). Virtual reconstruction of the Neanderthal Amud 1 cranium. *American Journal of Physical Anthropology, 158*(2), 185-197. doi:10.1002/ajpa.22777

Arbour, J. H., & Brown, C. M. (2014). Incomplete specimens in geometric morphometric analyses. *Methods in Ecology and Evolution, 5*(1), 16-26. doi:10.1111/2041-210x.12128

Attard, M. R. G., Parr, W. C. H., Wilson, L. A. B., Archer, M., Hand, S. J., Rogers, T. L., & Wroe, S. (2014). Virtual Reconstruction and Prey Size Preference in the Mid Cenozoic Thylacinid, *Nimbacinus dicksoni* (Thylacinidae, Marsupialia). *PLOS ONE, 9*(4), e93088.

Baine, K. E. (2014). *Mortuary ritual and social change in neolithic and Bronze Age Ireland.* (PhD thesis), University of Iowa,

Bauer, C. C., & Harvati, K. (2015). A virtual reconstruction and comparative analysis of the KNM-ER 42700 cranium. *Anthropologischer Anzeiger, 72*(2), 129-140.

Benazzi, S., Gruppioni, G., Strait, D. S., & Hublin, J.-J. (2014). Technical Note: Virtual reconstruction of KNM-ER 1813 Homo habilis cranium. *American Journal of Physical Anthropology, 153*(1), 154-160. doi:10.1002/ajpa.22376

Benazzi, S., Stansfield, E., Milani, C., & Gruppioni, G. (2009). Geometric morphometric methods for three-dimensional virtual reconstruction of a fragmented cranium: the case of Angelo Poliziano. *International Journal of Legal Medicine, 123*(4), 333-344. doi:10.1007/s00414-009-0339-6

Bermúdez de Castro, J. M., Martín-Francés, L., Modesto-Mata, M., Martínez de Pinillos, M., Martinón-Torres, M., García-Campos, C., & Carretero, J. M. (2015). Virtual reconstruction of the Early Pleistocene mandible ATD6-96 from Gran Dolina-TD6-2 (Sierra De Atapuerca, Spain). *American Journal of Physical Anthropology*, n/a-n/a. doi:10.1002/ajpa.22919

Bookstein, F. L. (2017). A newly noticed formula enforces fundamental limits on Geometric Morphometric analyses. *Evolutionary Biology, 44*(4), 522-541. doi:10.1007/s11692-017-9424-9

Bookstein, F. L. (2019). Pathologies of Between-Groups Principal Components Analysis in Geometric Morphometrics. *bioRxiv*, 627448. doi:10.1101/627448

Brown, C. M., & Vavrek, M. J. (2015). Small sample sizes in the study of ontogenetic allometry; implications for palaeobiology. *PeerJ, 3*, e818. doi:10.7717/peerj.818

Buck, T. J., & Vidarsdottir, U. S. (2004). A proposed method for the identification of race in sub-adult skeletons: a geometric morphometric analysis of mandibular morphology. *Journal of Forensic Science, 49*(6), JFS2004074-2004076.

Buikstra, J., & Ubelaker, D. (1994). *Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History*. Fayetteville: Arkansas Archeological Survey.

Cardini, A., & Elton, S. (2007). Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology, 126*(2), 121-134. doi:10.1007/s00435-007-0036-2

Cardini, A., O'Higgins, P., & Rohlf, F. J. (2019). Seeing distinct groups where there are none: spurious patterns from between-group PCA. *bioRxiv*, 706101. doi:10.1101/706101

Cardini, A., Seetah, K., & Barker, G. (2015). How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and

variances in simple randomized selection experiments. *Zoomorphology, 134*(2), 149-163. doi:10.1007/s00435-015-0253-z

Chamberlain, A. T. (2006). *Demography in archaeology*. New York: Cambridge University Press.

Crevecoeur, I., Brooks, A., Ribot, I., Cornelissen, E., & Semal, P. (2016). Late Stone Age human remains from Ishango (Democratic Republic of Congo): New insights on Late Pleistocene modern human diversity in Africa. *Journal of Human Evolution, 96*, 35-57. doi:http://dx.doi.org/10.1016/j.jhevol.2016.04.003

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., . . . Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging, 30*(9), 1323-1341. doi:10.1016/j.mri.2012.05.001

Ferembach, D., Schwidetzky, I., & Stloukal, M. (1980). Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution, 9*(7), 517-549. doi:https://doi.org/10.1016/0047-2484(80)90061-5

Franklin, D., O'Higgins, P., & Oxnard, C. E. (2008). Sexual dimorphism in the mandible of indigenous South Africans: A geometric morphometric approach. *South African Journal of Science, 104*(3-4), 101-106.

Franklin, D., Oxnard, C. E., O'Higgins, P., & Dadour, I. (2007). Sexual dimorphism in the subadult mandible: Quantification using geometric morphometrics. *Journal of Forensic Sciences, 52*(1), 6-10. doi:DOI 10.1111/j.1556-4029.2006.00311.x

Galland, M., Van Gerven, D. P., Von Cramon-Taubadel, N., & Pinhasi, R. (2016). 11,000 years of craniofacial and mandibular variation in Lower Nubia. *Scientific Reports, 6*, 31040. doi:10.1038/srep31040 http://www.nature.com/articles/srep31040#supplementary-information

Godinho, R. M., Fitton, L. C., Toro-Ibacache, V., Stringer, C. B., Lacruz, R. S., Bromage, T. G., & O'Higgins, P. (2018). The biting performance of Homo sapiens and Homo heidelbergensis. *Journal of Human Evolution, 118*, 56-71. doi:https://doi.org/10.1016/j.jhevol.2018.02.010

Godinho, R. M., & O'Higgins, P. (2017). Virtual reconstruction of cranial remains: the H. Heidelbergensis, Kabwe 1 fossil. In T. Thompson & D. Errickson (Eds.), *Human remains - Another dimension: the application of 3D imaging in funerary context* (pp. 135-147). London: Elsevier.

Godinho, R. M., & O'Higgins, P. (2018). The biomechanical significance of the frontal sinus in Kabwe 1 (Homo heidelbergensis). *Journal of Human Evolution, 114*, 141-153. doi:https://doi.org/10.1016/j.jhevol.2017.10.007

Godinho, R. M., Spikins, P., & O'Higgins, P. (2018). Supraorbital morphology and social dynamics in human evolution. *Nature Ecology & Evolution, 2*(6), 956-961. doi:10.1038/s41559-018-0528-0

Gunz, P., Mitteroecker, P., Bookstein, F., & Weber, G. (2004). *Computer-aided reconstruction of incomplete human crania using statistical and geometrical estimation methods.* Paper presented at the Computer Applications and Quantitative Methods in Archaeology.

Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G. W., & Bookstein, F. L. (2009). Principles for the virtual reconstruction of hominin crania. *Journal of Human Evolution, 57*(1), 48-62. doi:DOI 10.1016/j.jhevol.2009.04.004

Härke, H., & Belinskij, A. (2014). Causes and Contexts of Long-Term Ritual Change. *Death and Changing Rituals: Function and meaning in ancient funerary practices, 7*, 93.

Jiménez-Arenas, J. M., Bienvenu, T., Toro-Moyano, I., Ponce de León, M. S., & Zollikofer, C. P. E. (2019). Virtual reconstruction and re-evaluation of the Neanderthal frontal bone from Carigüela Cave (Granada, Spain). *Quaternary Science Reviews*. doi:https://doi.org/10.1016/j.quascirev.2019.03.014

Katz, D. C., Grote, M. N., & Weaver, T. D. (2017). Changes in human skull morphology across the agricultural transition are consistent with softer diets in preindustrial farming groups. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1702586114

Kelly, M. P., Vorperian, H. K., Wang, Y., Tillman, K. K., Werner, H. M., Chung, M. K., & Gentry, L. R. (2017). Characterizing mandibular growth using three-dimensional imaging techniques and anatomic landmarks. *Archives of Oral Biology, 77*, 27-38. doi:https://doi.org/10.1016/j.archoralbio.2017.01.018

Klingenberg, C. (2015). Analyzing Fluctuating Asymmetry with Geometric Morphometrics: Concepts, Methods, and Applications. *Symmetry, 7*(2), 843.

Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution, 56*(10), 1909-1920. doi:10.1111/j.0014-3820.2002.tb00117.x

Kurila, L. (2015). The Accuracy of the Osteological Sexing of Cremated Human Remains: A Test Based on Grave Goods from East Lithuanian Barrows. *Collegium Antropologicum, 39*(4), 821-828.

May, H., Sella-Tunis, T., Pokhojaev, A., Peled, N., & Sarig, R. (2018). Changes in mandible characteristics during the terminal Pleistocene to Holocene Levant and their association with dietary habits. *Journal of Archaeological Science: Reports*. doi:https://doi.org/10.1016/j.jasrep.2018.03.020

Mitteroecker, P., Gunz, P., Windhager, S., & Schaefer, K. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix-Italian Journal of Mammalogy, 24*(1), 59-66. doi:DOI 10.4404/hystrix-24.1-6369

Mounier, A., Correia, M., Rivera, F., Crivellaro, F., Power, R., Jeffery, J., . . . Mirazón Lahr, M. (2018). Who were the Nataruk people? Mandibular morphology among late Pleistocene and early Holocene fisher-forager

populations of West Turkana (Kenya). *Journal of Human Evolution, 121*, 235-253. doi:https://doi.org/10.1016/j.jhevol.2018.04.013

Neeser, R., Ackermann, R. R., & Gain, J. (2009). Comparing the accuracy and precision of three techniques used for estimating missing landmarks when reconstructing fossil hominin crania. *American Journal of Physical Anthropology, 140*(1), 1-18. doi:10.1002/ajpa.21023

O'Higgins, P., Fitton, L. C., & Godinho, R. M. (2019). Geometric morphometrics and finite elements analysis: Assessing the functional implications of differences in craniofacial form in the hominin fossil record. *Journal of Archaeological Science, 101*, 159-168. doi:https://doi.org/10.1016/j.jas.2017.09.011

Pearson, J., Grove, M., Özbek, M., & Hongo, H. (2013). Food and social complexity at Çayönü Tepesi, southeastern Anatolia: Stable isotope evidence of differentiation in diet according to burial practice and sex in the early Neolithic. *Journal of Anthropological Archaeology, 32*(2), 180-189. doi:https://doi.org/10.1016/j.jaa.2013.01.002

Pinhasi, R., Eshed, V., & von Cramon-Taubadel, N. (2015). Incongruity between Affinity Patterns Based on Mandibular and Lower Dental Dimensions following the Transition to Agriculture in the Near East, Anatolia and Europe. *PLOS ONE, 10*(2), e0117301.

Pokhojaev, A., Avni, H., Sella-Tunis, T., Sarig, R., & May, H. (2019). Changes in human mandibular shape during the Terminal Pleistocene-Holocene Levant. *Scientific Reports, 9*(1), 8799. doi:10.1038/s41598-019-45279-9

Rmoutilová, R., Guyomarc'h, P., Velemínský, P., Šefčáková, A., Samsel, M., Santos, F., . . . Brůžek, J. (2018). Virtual reconstruction of the Upper Palaeolithic skull from Zlatý Kůň, Czech Republic: Sex assessment and morphological affinity. *PLOS ONE, 13*(8), e0201431.

Séguy, I., & Buchet, L. (2014). *Handbook of Palaeodemography*. Dordecht: Springer.

Senck, S., Bookstein, F. L., Benazzi, S., Kastner, J., & Weber, G. W. (2015). Virtual Reconstruction of Modern and Fossil Hominoid Crania: Consequences of Reference Sample Choice. *The Anatomical Record, 298*(5), 827-841. doi:10.1002/ar.23104

Singh, N. (2014). Ontogenetic Study of Allometric Variation in Homo and Pan Mandibles. *The Anatomical Record, 297*(2), 261-272. doi:10.1002/ar.22843

von Cramon-Taubadel, N. (2011). Global human mandibular variation reflects differences in agricultural and hunter-gatherer subsistence strategies. *Proceedings of the National Academy of Sciences, 108*(49), 19546-19551. doi:10.1073/pnas.1113050108

Weber, G., & Bookstein, F. (2012). Manuals for the EVAN Toolbox: No.2.

Wellens, H. L. L., Kuijpers-Jagtman, A. M., & Halazonetis, D. J. (2013). Geometric morphometric analysis of craniofacial variation, ontogeny and modularity in a cross-sectional sample of modern humans. *Journal of Anatomy, 222*(4), 397-409. doi:doi:10.1111/joa.12027

Wiley, D. F., Amenta, N., Alcantara, D. A., Ghosh, D., Kil, Y. J., Delson, E., . . . Hamann, B. (2005). Evolutionary morphing. In *Proceedings of the IEEE Visualization 2005* (pp. 431-438). Piscataway, NJ.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric Morphometrics For Biologists: A Primer*. New York: Elsevier.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2012). *Geometric Morphometrics For Biologists: A Primer*. New York: Elsevier.

# LM	LM name	Landmark definition
1	Gnathion	Midline of the inferior border of the mandible
2	Infradentale	On the anterior alveolar ridge, between anterior incisors
3	Linguale	Genial tubercle: in case of a single tubercle, on its tip; in case of two, midpoint between them
4	Orale, mandible	On the posterior alveolar ridge between the anterior incisors
5	Pogonion	Most anterior point of mandibular symphysis
6	C-P3	On the anterior alveolar ridge between canine and first premolar
7	P4-M1	On the anterior alveolar ridge between second premolar and first molar
8	M1-M2	On the anterior alveolar ridge between first and second molar teeth
9	Mental Foramen Anterior	Anterior point of mental foramen
10	Ramus Root	On the anterior rim of the ramus (placed on the level of the alveolar ridge)
11	Gonion	A point on the projection of the bisection of the mandibular angle
12	Condyle, lateral	From a superior view, the lateral point on the condyle
13	Condyle, midpoint	From a superior view, a point in the center of the condyle
14	Condyle, medial	From a superior view, the medial point on the condyle
15	Sigmoid Notch	Mandible is positioned in the mandibular plane in a lateral view, then the lowest point of the mandibular notch is marked
16	Coronoid process	Tip of the coronoid process
17	Mandibular Foramen, inferior	Most inferior point of the mandibular foramen
18	Alveolous, lingual posterior	From a superior view, the most posterior point on the lingual alveolar process
19	Condyle, anterior	A point on the antero-superior aspect of the mandibular notch (on the condyle)
20	Condyle, posterior	The center of the condyle from a posterior view
21	Ramus, posterior	Posteriormost point of the ramus that is line with the ramus root

Table 1: Mandibular landmarks used in this study.

Site	Chronology	Number of mandibles
Cabeço da Amoreira	Mesolithic	1
Cabeço da Arruda	Mesolithic	22
Cova da Onça	Mesolithic	2
Moita do Sebastião	Mesolithic	21
Casa da Moura	Neolithic	8
Monte do Carrascal 2	Chalcolithic	7
Monte da Guarita 2: 6	Chalcolithic	6
Total		67

Table 2: Mandibles used to estimate most frequently missing landmarks.

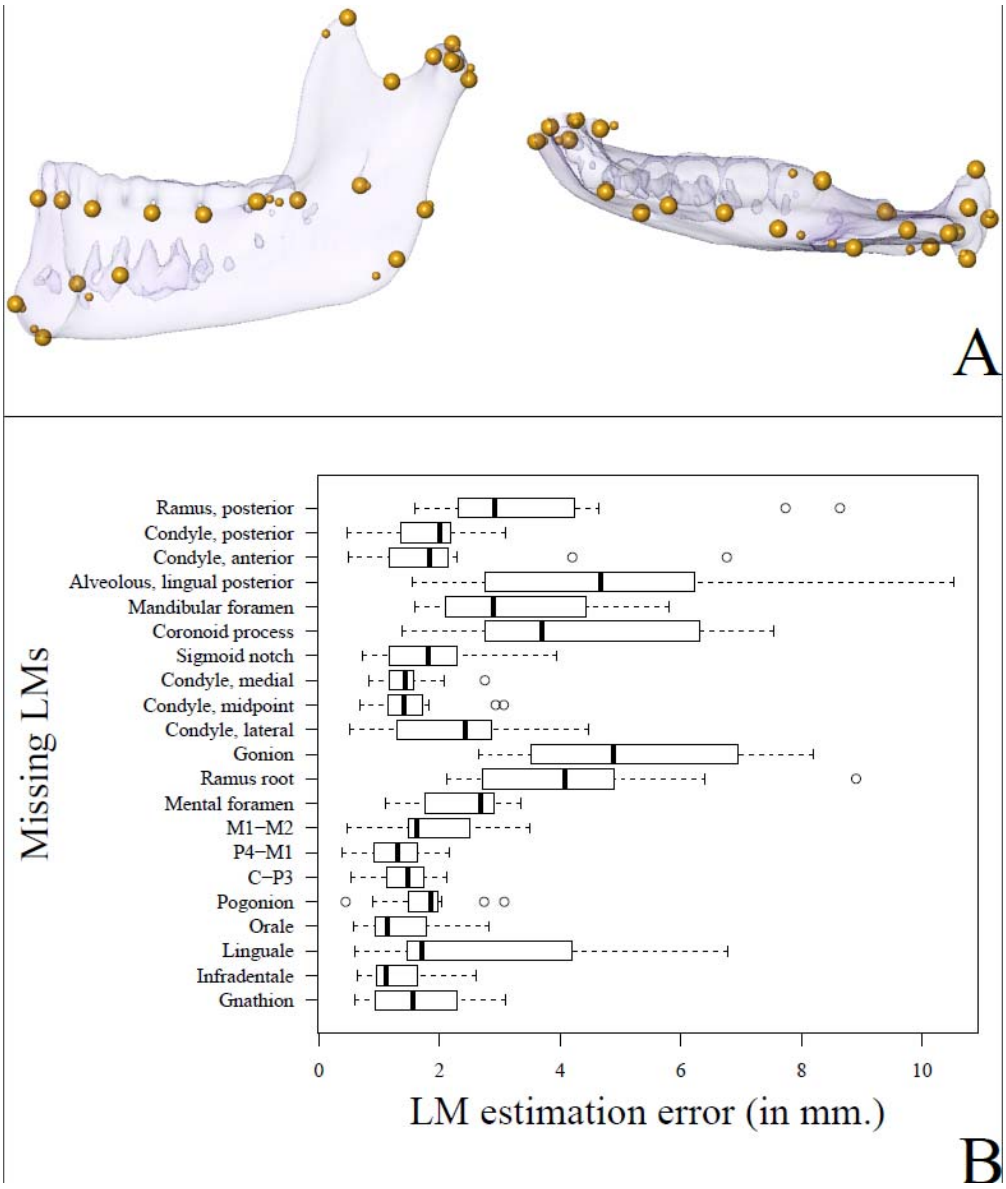


Figure 1: Residuals of landmark coordinates per landmark from the tests in which single missing landmarks are estimated. (A) shows an example specimen (Moita do Sebastião 3) with the locations of the original (large spheres) and estimated (small spheres) landmarks; when error is small, small spheres lie hidden from view, within large ones. (B) quantifies the estimation error (residuals) for each landmark.

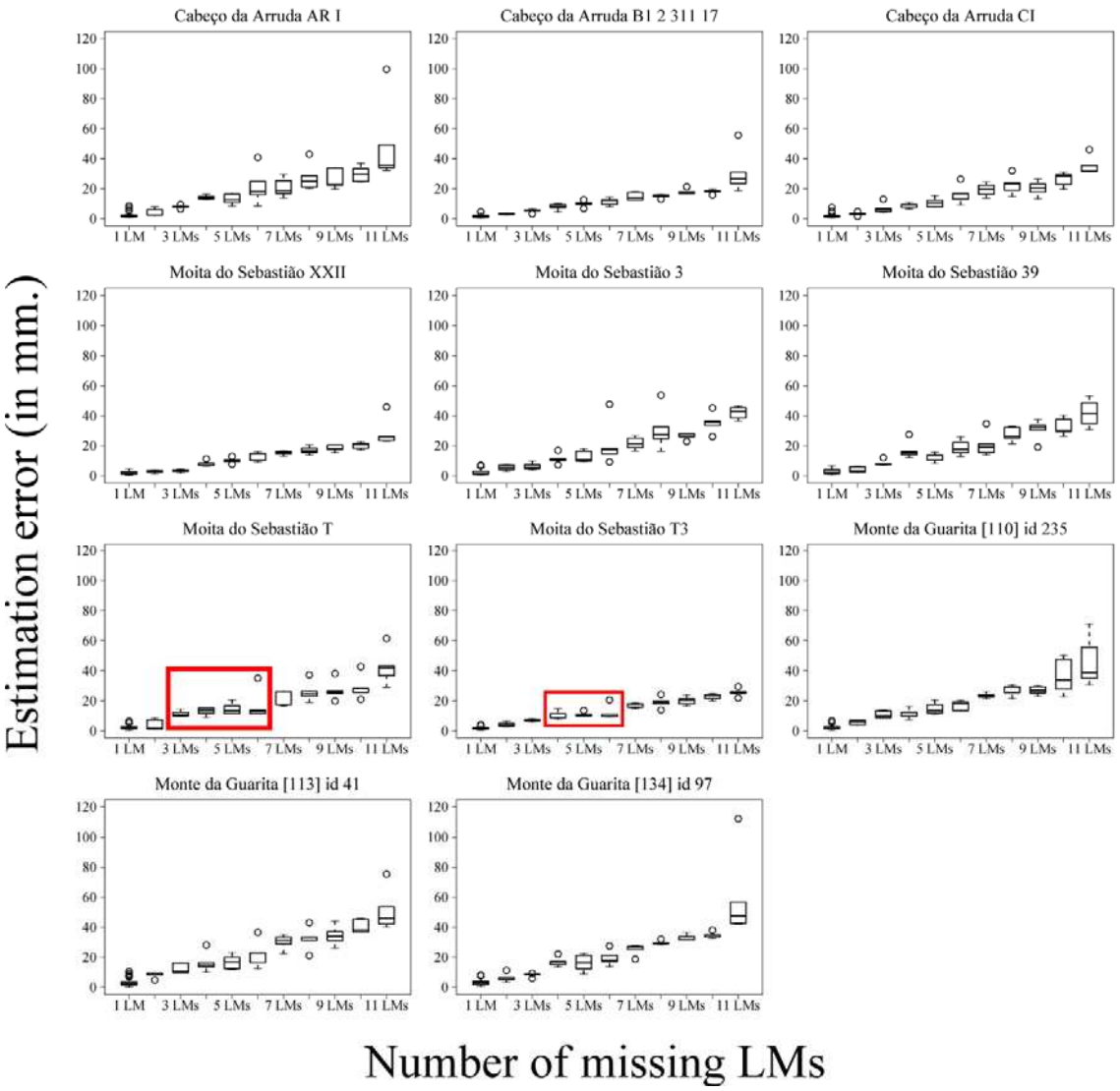


Figure 2: Residuals of landmark coordinates per set of missing landmarks. Number of missing landmarks on the x axis and total error of estimation (mm.) on the y axis. The title of each boxplot provides the original museum identification of each specimen. The red boxes in Moita do Sebastião T and Moita do Sebastião T3 highlight examples in which sets missing more LMs produce comparable or lower error (i.e., residuals) to sets missing fewer LMs.

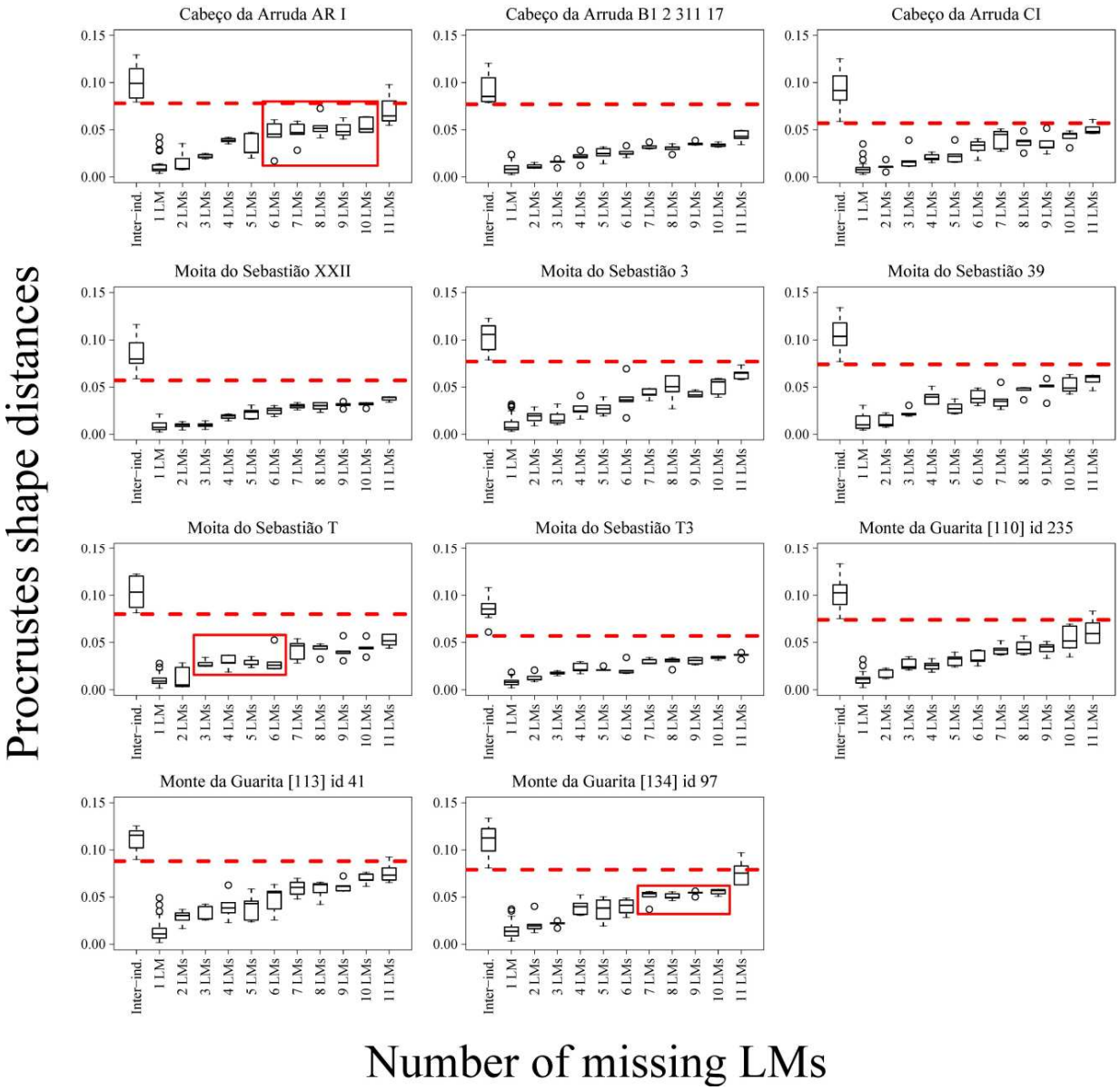


Figure 3: Missing data estimation error (assessed using Procrustes shape distances; PRDs). Procrustes shape distances (y axis) tend to increase as the number of missing landmarks (x axis) increases. Each box presents one specimen, which is identified in the title with the original museum identification. The inter-individual (Inter-ind.) box in each boxplot displays the pairwise PRDs among this and the other original specimens. The other boxes (1 LM, 2 LMs, 3LMs,...) in each plot show the pairwise PRDs between specimens with estimated missing LMs and the original, complete specimen. The red dotted lines mark the lower limit of PRDs found among individuals (Inter-ind.) against which the impact of missing data estimation error can be assessed. The red boxes in Cabeço da Arruda AR I, Moita do Sebastião T and Monte da Guarita [134] id 97 highlight examples in which sets missing more LMs produce comparable or lower error (i.e., residuals) than sets missing less LMs.

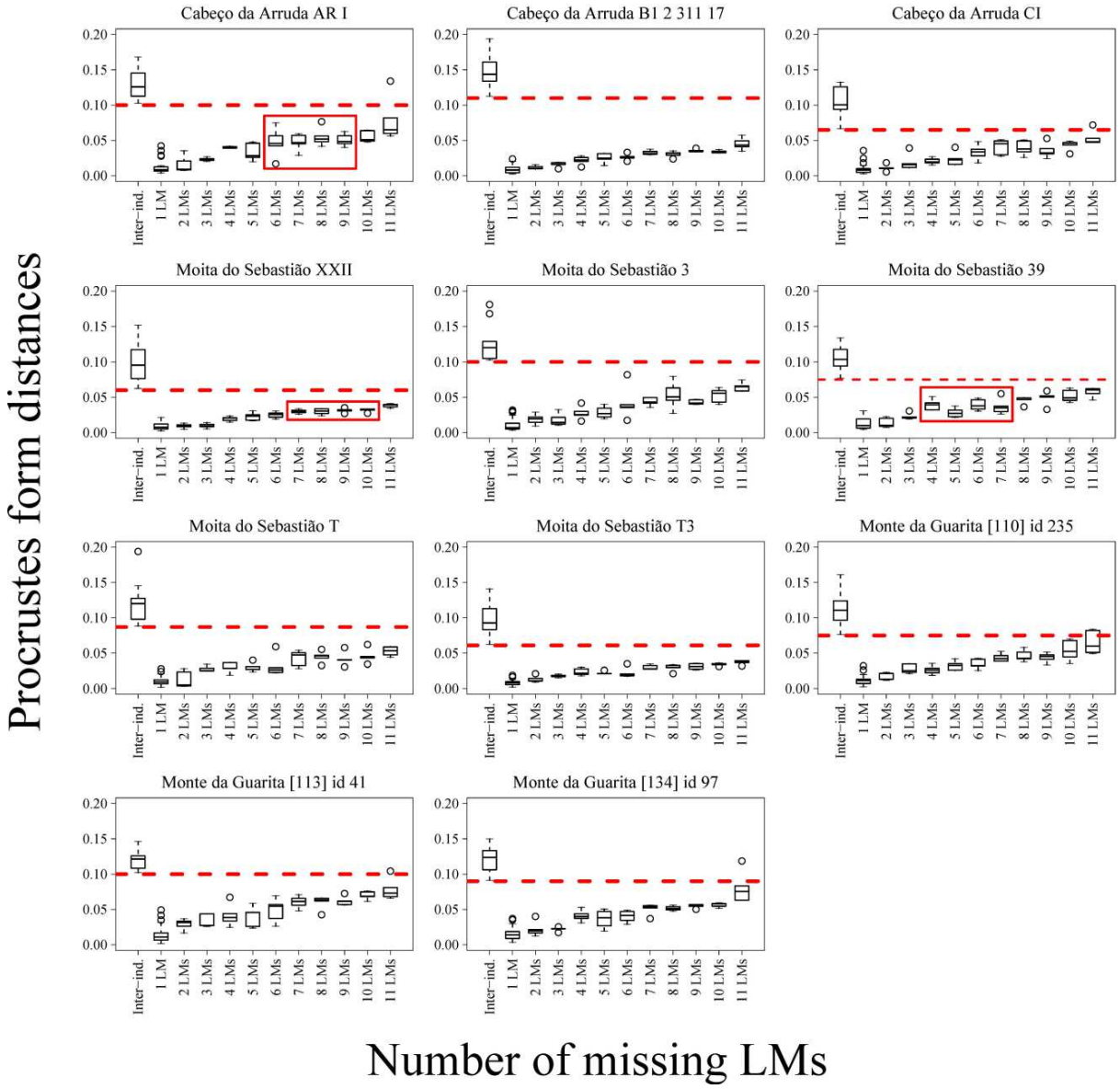
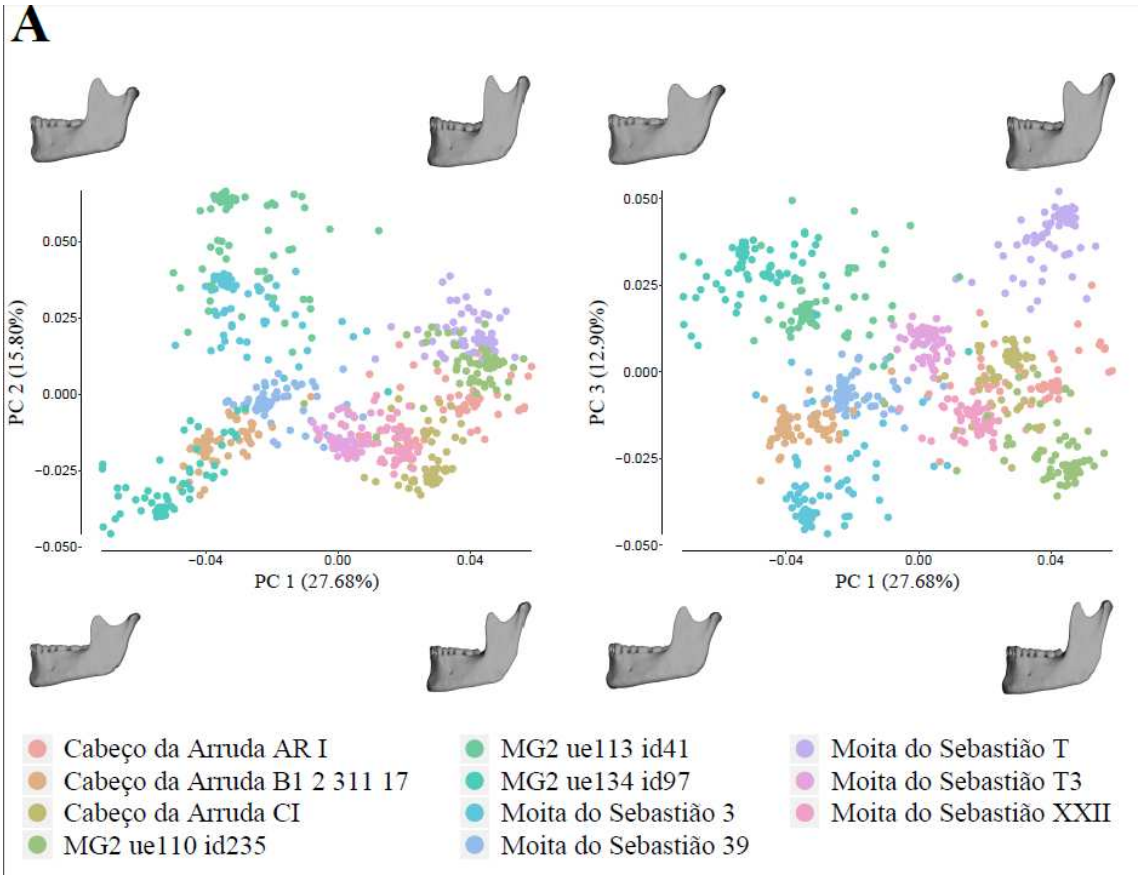


Figure 4: Missing data estimation error (assessed using Procrustes form distances). Procrustes form distances (y axis) tend to increase as the number of missing landmarks (x axis) increases. Each box presents one specimen, which is identified in the title with the original museum identification. The inter-individual (Inter-ind.) box in each boxplot displays the pairwise form distances among this and the other original specimens. The other boxes (1 LM, 2 LMs, 3LMs,...) in each plot show the pairwise form distances between specimens with estimated missing LMs and the original, complete specimen. The red dotted lines mark the lower limit of form distances found among individuals (Inter-ind.) against which the impact of missing data estimation error can be assessed. The red boxes in Cabeço da Arruda AR I, Moita do Sebastião XXII and Moita do Sebastião 39 highlight examples in which sets missing more LMs produce comparable or lower error (i.e., residuals) than sets missing fewer LMs.


A
PC 2 (15.80%)
PC 1 (27.68%)
PC 3 (12.90%)
PC 1 (27.68%)
Cabeço da Arruda AR I
Cabeço da Arruda B1 2 311 17
Cabeço da Arruda CI
MG2 ue110 id235
MG2 ue113 id41
MG2 ue134 id97
Moita do Sebastião 3
Moita do Sebastião 39
Moita do Sebastião T
Moita do Sebastião T3
Moita do Sebastião XXII

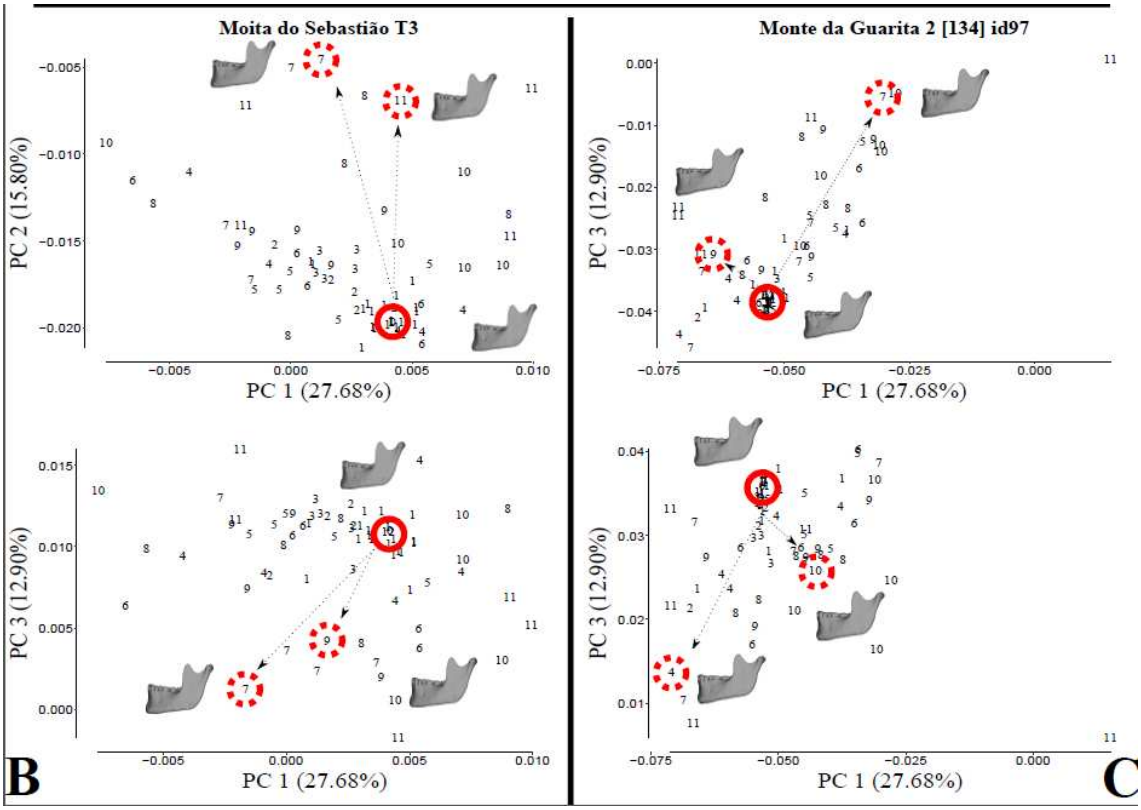

Moita do Sebastião T3
Monte da Guarita 2 [134] id97
PC 2 (15.80%)
PC 3 (12.90%)
PC 1 (27.68%)
B
C

Figure 5: Shape PCA plot of original complete specimens and incomplete specimens with estimated LMs. (A) shows the plot of PC 1 and 2, and PC 1 and 3 in which different specimens (original and estimated) are represented by different colours. Insets of mandibles depict warping of mandibular morphology at the extremes of the PC axes. (B) and (C) plot the results of the same PCA but showing only one specimen in each (Moita do Sebastião T3 and Monte da Magoita 2 [134] id 97, respectively). This is to provide a better understanding of how estimation error impacts PCA. The plots display the original complete specimens (0, highlighted with the solid red circles) and other specimens are numbered according to how many LMs were estimated. The dashed red circles highlight cases in which the estimation of fewer missing LMs results in larger error than the estimation of more missing LMs in the PCA. Insets of mandibles depict warping of mandibular morphology from the reference original complete specimen to the target incomplete mandible with estimated missing data.

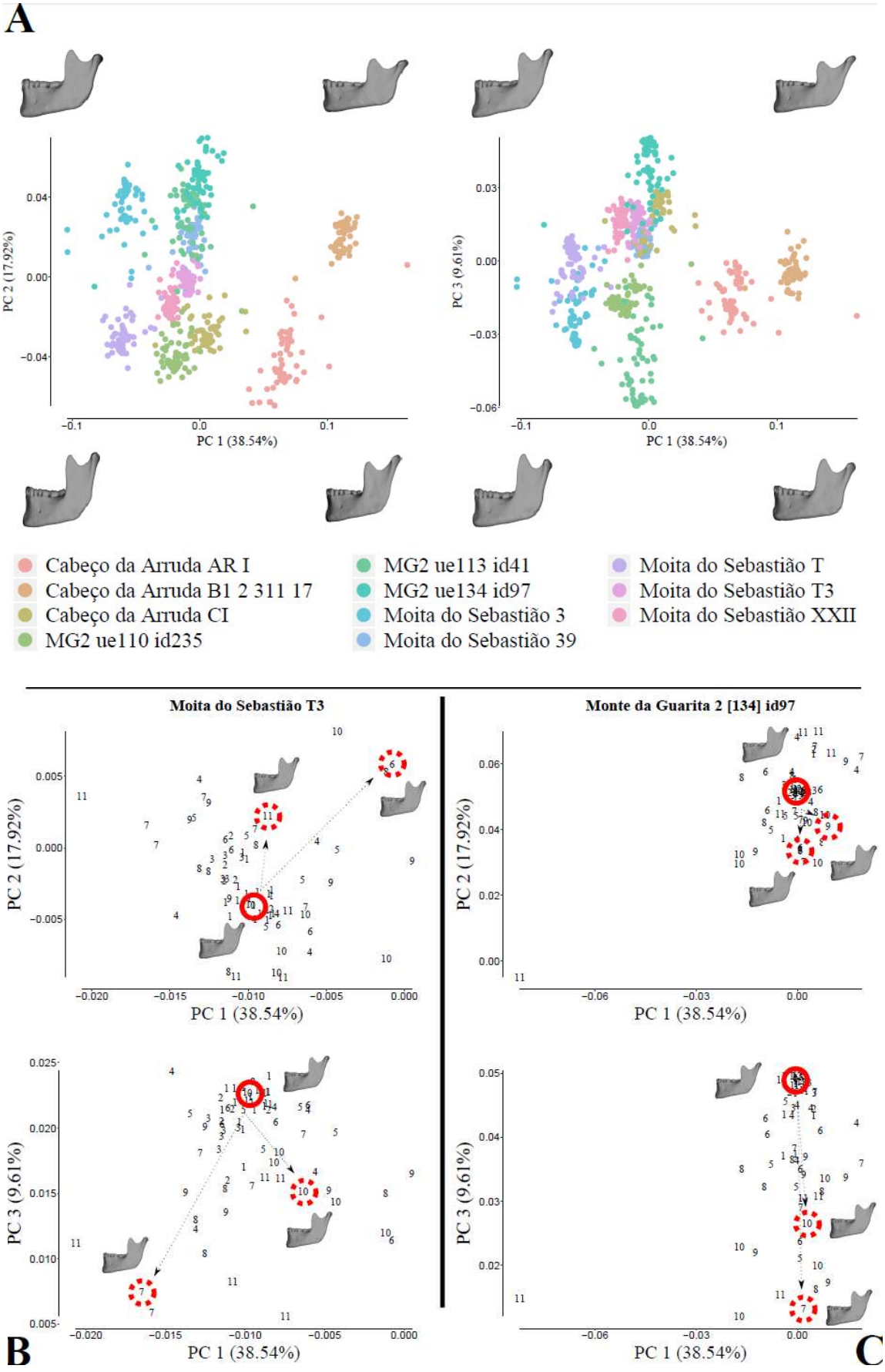

A
PC 2 (17.92%)
PC 1 (38.54%)
PC 3 (9.61%)
PC 1 (38.54%)
Cabeço da Arruda AR I
Cabeço da Arruda B1 2 311 17
Cabeço da Arruda CI
MG2 ue110 id235
MG2 ue113 id41
MG2 ue134 id97
Moita do Sebastião 3
Moita do Sebastião 39
Moita do Sebastião T
Moita do Sebastião T3
Moita do Sebastião XXII
Moita do Sebastião T3
Monte da Guarita 2 [134] id97
PC 2 (17.92%)
PC 1 (38.54%)
PC 3 (9.61%)
PC 1 (38.54%)
B
C

Figure 6: Form PCA plot of original complete specimens and incomplete specimens with estimated LMs. (A) shows the plot of PC 1 and 2, and PC 1 and 3 in which different specimens (original and estimated) are represented by different colours. Insets of mandibles depict warping of mandibular morphology at the extremes of the PC axes. (B) and (C) plot the results of the same PCA in 2D, but with only one specimen each (Moita do Sebastião T3 and Monte da Guarita 2 [134] id 97, respectively). This is to provide a better understanding of how estimation error impacts PCA. The plots display the original complete specimens (0, highlighted with the solid red circles) and other specimens are numbered according to how many LMs were estimated. The dashed red circles highlight cases in which the estimation of fewer missing LMs results in larger error than the estimation of more missing LMs in the PCA. Insets of mandibles depict warping of mandibular morphology from the reference original complete specimen to the target incomplete mandible with estimated missing data.

Table S1

	Identification of missing landmarks (number not in bold) in each set of simulations (numbers in bold)										
	2	3	4	5	6	7	8	9	10	11
1	1, 2	2, 5, 4	1, 3, 5, 17	12, 13, 4, 19, 20	11, 12, 13, 14, 19, 20	15, 13, 11, 7, 21, 5, 6	2, 4, 12, 13, 14, 19, 20, 21	13, 18, 15, 8, 21, 12, 9, 20, 10	1, 2, 4, 6, 7, 8, 10, 17, 18, 20	3, 5, 8, 12, 13, 14, 15, 16, 17, 19, 20
2	2, 3	1, 2, 4	2, 3, 4, 11	1, 2, 4, 6, 20	1, 13, 21, 18, 8, 10	9, 16, 11, 3, 15, 2, 14	14, 3, 21, 1, 11, 12, 2, 17	20, 21, 3, 19, 10, 9, 17, 12, 4	6, 17, 10, 21, 15, 13, 4, 8, 5, 20	14, 7, 5, 21, 18, 11, 13, 20, 17, 2, 3
3	2, 4	2, 4, 8	11, 12, 14, 19	4, 8, 6, 10, 16	19, 16, 6, 4, 1, 14	12, 16, 18, 15, 10, 13, 11	13, 12, 20, 17, 10, 21, 11, 7	10, 4, 15, 2, 14, 12, 1, 11, 7	4, 17, 5, 19, 3, 12, 16, 10, 11, 21	17, 6, 5, 16, 15, 7, 20, 10, 14, 9, 11
4	2, 11	2, 4, 12	12, 13, 14, 20	12, 4, 10, 6, 21	6, 21, 5, 15, 7, 4	5, 16, 21, 4, 14, 1, 10	15, 17, 20, 9, 4, 19, 18, 16	20, 10, 14, 4, 8, 5, 7, 3, 11	9, 8, 18, 12, 10, 13, 20, 2, 4, 19	11, 12, 4, 17, 18, 8, 6, 16, 21, 14, 10
5	3, 5	2, 4, 16	4, 21, 17, 8	9, 4, 3, 21, 2	1, 18, 6, 9, 7, 3	4, 8, 1, 6, 3, 10, 2	6, 12, 3, 9, 7, 1, 2, 21	18, 2, 16, 17, 14, 10, 1, 5, 3	20, 19, 3, 7, 16, 6, 2, 13, 21, 17	14, 9, 15, 4, 12, 11, 19, 18, 17, 10, 2
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										