



Available online at:

<http://www.italian-journal-of-mammalogy.it/article/view/10993/pdf>

doi:10.4404/hystrix-25.2-10993

Research Article

Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures?

Andrea CARDINI^{a,b,*}^aDipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, I.go S. Eufemia 19, 41121 Modena, Italy^bCentre for Forensic Science, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia**Keywords:**

Analysis of variance
cranium
landmarks
mandible
marmots
measurement error
photos
Procrustes
shape analysis

Article history:

Received: 19 November 2014

Accepted: 29 December 2014

Abstract

Procrustean geometric morphometrics has made large use of 2D images for studying three-dimensional structures such as mammalian bones or arthropod exoskeletona. This type of use of 2D data is still widespread today and will likely remain common for several years due to its simplicity, efficiency and low cost. However, using 2D pictures to measure morphological variation in a 3D object is an approximation that inevitably implies measurement error. Despite this being an obvious problem, which was emphasized since the early days of the first applications of geometric morphometrics to biology, whether 2D is a good proxy for 3D has been a rather neglected topic in the literature until very recently. In this paper, using marmot mandibles and crania as an example, I show how to assess the potentially crucial impact of 'missing the third dimension' in 2D landmarks and suggest a new method to test the accuracy of these data: the method is simple and can be easily performed in a user-friendly free software such as MorphoJ. This test is complimentary to other more exploratory analyses, that can also be performed using free programs and might offer a routine protocol to estimate the goodness of the 2D to 3D approximation in geometric morphometrics. Example data and a fully worked out MorphoJ project are provided for readers to learn how to replicate the analysis.

Acknowledgements

I am deeply grateful to the curators of the United States Natural History Museum, Washington DC, who allowed and help me to study their collection first in 1999 and later in 2004. I am greatly in debt to Ian Dryden (University of Nottingham), whom I repeatedly tormented in early stages of this work, back in 2013, with nonsensical questions. I am also in debt to the email discussion list of morphometricians and especially to a number of its subscribers who provided advice on references and some specific methodological issues: for this help, many thanks to Göran Arnqvist (Uppsala University), Mauro Cavalanti (Ecoinformatics Studio, Rio de Janeiro), Marko Djurakic (University of Novi Sad), Hanneke van Heteren (University of Bonn), Norman MacLeod (Natural History Museum, London), and Dennis Slice (Florida State University, Tallahassee). Finally, I am very happy to acknowledge the fundamental help of David Sheets (Canisius College, Buffalo, and SUNY at Buffalo), Jim Rohlf (Stony Brook University), and (again!) Mauro Cavalanti (Ecoinformatics Studio, Rio de Janeiro), who reviewed the original manuscript: I thank them a lot for many useful suggestions that greatly improved the paper.

This paper is dedicated to the memory of Luigi Cagnolaro (1934–2014), Curator, and later Director, of the Natural History Museum of Milan; honorary President of the Italian Teriological Association (ATI); certainly one of the greatest Italian mammalogists, whose kindness and scientific curiosity I will always remember, and whose work on marmots (among many other mammals he studied) has been inspirational to my own research.

Introduction

Procrustean geometric morphometric (GMM) analyses Adams et al. (2013) of three-dimensional (3D) structures are often performed on two-dimensional (2D) images. For instance, in my personal collection of 76 GMM studies using Procrustes methods, and published in major evolutionary, zoological and anthropological journals in the first nine months of 2014, more than half (57%) are 2D analyses, although virtually all of them concern 3D anatomical structures. This is clearly not an exhaustive assessment of the recent literature, but it helps to appreciate how common 2D analyses are even in the second decade of the 21st century, a time when tools for 3D measurements (3D digitizers, surface scanners etc.) have become cheaper and more accurate than in the early days of GMM. Then, why are 2D analyses still so popular?

Besides a few analytical advantages, such as computational simplicity and an effective visualization of shape change using thin-plate spline (TPS) grids (Klingenberg, 2013, and references therein), 2D data are easy to collect and generally inexpensive. Most of the time,

a tripod or a copy stand and a camera (or a microscope) with good lens allow to acquire accurate pictures. These, together with a protocol to standardize photographic settings (e.g., Cardini and Tongiorgi, 2003), make data collection effective, fast and low cost. 2D landmarks can also be easily digitized on digital images using TPSDig (Rohlf, 2014) and efficiently processed in a variety of free GMM programs (<http://life.bio.sunysb.edu/morph/toc-software.html>). Using the same type of software, the now increasingly popular methods for the analysis of semilandmarks¹ are similarly easily applied to the study of 2D curves. Thus, cost-effectiveness, rapidity of data collection and analytical simplicity contribute to explain the enduring success of 2D landmark analyses. Also, in specific instances, 2D data might be interesting in comparative terms (e.g., Bush et al., 2011, forensic assessment

¹Semilandmarks lack the precise one to one correspondence of anatomical landmarks. For this reason they are generally slid on a curve to improve their *mathematical* correspondence. Sliding is extremely simple for 2D data using a user-friendly free program such as TPSRelw (Rohlf, 2014). Although this is marginal to this paper, to avoid a common confusion on what sliding does, I am using *italics* to stress that this method – described by Bookstein (1997), and recently reviewed by Gunz and Mitteroecker (2013) – is about mathematical/geometric correspondence, which is not generally the same as biological “homology” – Klingenberg (2008); Oxnard and O’Higgins (2011).

* Corresponding author

Email address: cardini@unimo.it (Andrea CARDINI)

of the performance of 2D and 3D analyses of bitemarks) or in themselves (e.g., in a biometric individual identification from pictures), as they may correspond to the reality of the data one has to extract information from.

It is obvious to anyone, however, that whenever a study structure is a 3D object, 2D pictures will inevitably imply a loss of information and a degree of inaccuracy in estimating size and shape. For brevity, this approximation of a three-dimensional object with a two-dimensional picture will be henceforth abbreviated with TTD (Two to Three Dimensional approximation). Fly wings and some plant leaves may produce 2D data with virtually no TTD error, as they are so flat to be almost perfectly two-dimensional. Radiographs are also somewhat special, in that they generally capture a specific 2D aspect of variation in 3D objects (e.g., human cranial midplanes – Bastir et al., 2006); as long as only that given aspect is concerned, radiographs may therefore be a very accurate source of 2D data. The outline of some types of fishes and shells also may introduce very small TTD errors, as long as the points used to measure form are on the same plane (e.g., the midplane of a fish – Cavalcanti et al., 1999 — or the occlusal margin of a bivalve shell – Márquez et al., 2010). However, by far the largest majority of studies on mammal bones, a most common subject in geometric morphometrics (Cardini and Loy, 2013), focus on highly 3D anatomical features. Similarly, the head, pronotum and elytra of beetles (Pizzo et al., 2006) or the carapace of a crab (Rufino et al., 2006) are highly 3D structures, and arthropods are another common study group of morphometricians. Therefore, 2D pictures of many organisms or their organs can only be regarded as a proxy of the real three-dimensional anatomy. This proxy, however, may not be as accurate as one might wish.

Even if the problem is self-evident and clearly important, TTD has, to my knowledge, rarely been mentioned in the GMM literature. Roth's (1993) article on "On three-dimensional morphometrics, and on the identification of landmark points" may be one of the main exceptions. In that paper, Roth (p. 46) asks: "How is it that the two-dimensional representation of three-dimensional objects has become standard practice? Why does this seem natural, and why does it not strike us perpetually as a compromise?". She goes on providing an in depth discussion of the crucial relevance of 3D landmarks in the study of morphological variation. In doing this, she anticipates the potential of photogrammetry and suggests how to minimize TTD errors. However, despite such an important reference dating back from the early days of geometric morphometrics, no one (but see also below) seems to have suggested a method to effectively test whether the inaccuracy in 2D data is really negligible.

Other sources of measurement error, in contrast, such as digitizing, positioning and instrumental errors, are commonly assessed and sometimes tested by comparing the amount of "real" biological variation in a sample to measurement error using an analysis of variance (ANOVA – Arnqvist and Martensson, 1998; Klingenberg et al., 2002; Viscosi and Cardini, 2011). In fact, Arnqvist and Martensson (1998), in their detailed treatment of measurement error in GMM, do not fail to mention also TTD. However, they do not provide a suggestion on how to test it and only this year, more than two decades after Roth's (1993) seminal paper, there seems to be a sudden resurgence in TTD studies. Indeed, at the time when this study was done, there were very few online references on this topic but some useful suggestions on ongoing studies were provided by subscribers of the email morphometric discussion list (MORPHMET — <http://www.morphometrics.org/home/morphmet>). Among others, MacLeod et al. (2014), as well as morphometricians in Serbia (Djurakic, pers. comm.), are working on comparisons of 2D and 3D techniques; also, in the last three months there have been at least two conference abstracts presenting methods to estimate TTD (Close and Friedman, 2014; Watanabe, 2014), none of which, however, provides details on what is actually done. Thus, there might be now several parallel developments on how to explicitly and numerically address the TTD issue, and my hope is that this study will contribute to move another small step forward in that direction.

The main problem with estimating TTD is that 2D and 3D shape data have different dimensionality and thus occupy different shape spaces.

This precludes any analysis (PCA, cluster analysis, ANOVA etc.) that requires commensurable data. Correlational studies are still possible, but they provide only a partial answer on the impact of TTD, as they simply produce crude estimates of the proportionality of size and shape differences in 2D and 3D datasets. An example of this approach, which I call "traditional", is my own work with Thorington in 2006. In that paper on marmot ontogeny, we explored the impact of TTD by selecting subsets of landmarks in common between a 3D configuration and 2D landmark data on both the dorsal and lateral 2D views of the cranium. Thus, using the same specimens and landmarks, we estimated the correlation between centroid size in the different datasets; we computed the correlation between the matrices of Procrustes shape distances from 2D and 3D data; and we also compared ontogenetic patterns in scatterplots and shape diagrams based on the 2D-3D data. Our conclusion was that, for that specific dataset, correlation was very high for size and fairly high for shape with results generally in good agreement between the more accurate² 3D landmarks and the 2D photos. However, that study does not provide any real test of TTD and most of the assessment of the importance of TTD is left to the researchers' individual judgement.

In this paper, I take our 2006 work as a starting point and, using partly the same data and partly new data on mandibles and ventral crania, I explore multiple methods to carefully assess TTD. In this process, I suggest a new approach that allows to bring 2D and 3D data in a common shape space to perform a direct comparison of their similarities and differences. This new approach is based on simple operations that transform the data so that TTD becomes testable using the same ANOVA model which is routinely employed to test other sources of measurement error in GMM data (Viscosi and Cardini, 2011). In the article, I will explain in details how to replicate the analysis using MorphoJ (Klingenberg, 2011) and a subset of my own data, as an example. MorphoJ is free (http://www.flywings.org.uk/morphoj_page.htm), user-friendly but powerful, and likely represents today the most commonly used program for Procrustes GMM. With this study, and by providing data and a fully worked out MorphoJ project, I wish to further stimulate work on the fundamental but to date neglected subject of whether the widely used 2D data really are a good proxy for their 3D counterparts. Testing this potentially crucial source of error might indeed soon become routine in GMM, as for digitizing and other types of errors.

Material and Methods

Data acquisition

Two structures and a total of four datasets are analysed. The first one is a sample of left hemimandibles (henceforth briefly referred to as mandibles) of 20 adults hoary marmots (*Marmota caligata*), where a configuration of nine landmarks (Fig. 1a; Cardini, 2003, and Cardini et al., 2009 and references therein) was digitized in 2D using either pictures taken with a camera (Cardini and Tongiorgi, 2003) or simply acquired using a flat-bed scanner (Nagorsen and Cardini, 2009). The same nine landmarks were also digitized directly in 3D on the actual bones using a Microscribe 3D digitizer.

The second structure and the other three datasets are from a sample of cranial data on the left half of the skull of 49 yellow-bellied marmots (*Marmota flaviventris*). These are mostly adults but also include a few young. They are the same subset of data used by Cardini and Thorington (2006) to test the congruence of shape distances in 2D and 3D. However, compared to that study, which only used the dorsal (Fig. 1b) and lateral (Fig. 1c) views of the cranium, I have also included here their ventral view (Fig. 1d; Cardini and O'Higgins, 2005). 2D pictures of crania were taken using a camera and the same protocol described for the mandibles (Cardini and Tongiorgi, 2003). Unlike the relatively flat mandibles, crania cannot be imaged using a flat-bed scanner, as the

²Accuracy is rigorously defined as "closeness of agreement between a measured quantity value and a true quantity value of a measurand" (JCGM, 2008, p. 21). In this study, I am informally using accuracy also to refer to the better quantitative description of a 3D structure provided by 3D data compared to 2D measurements with their inevitable loss of information.

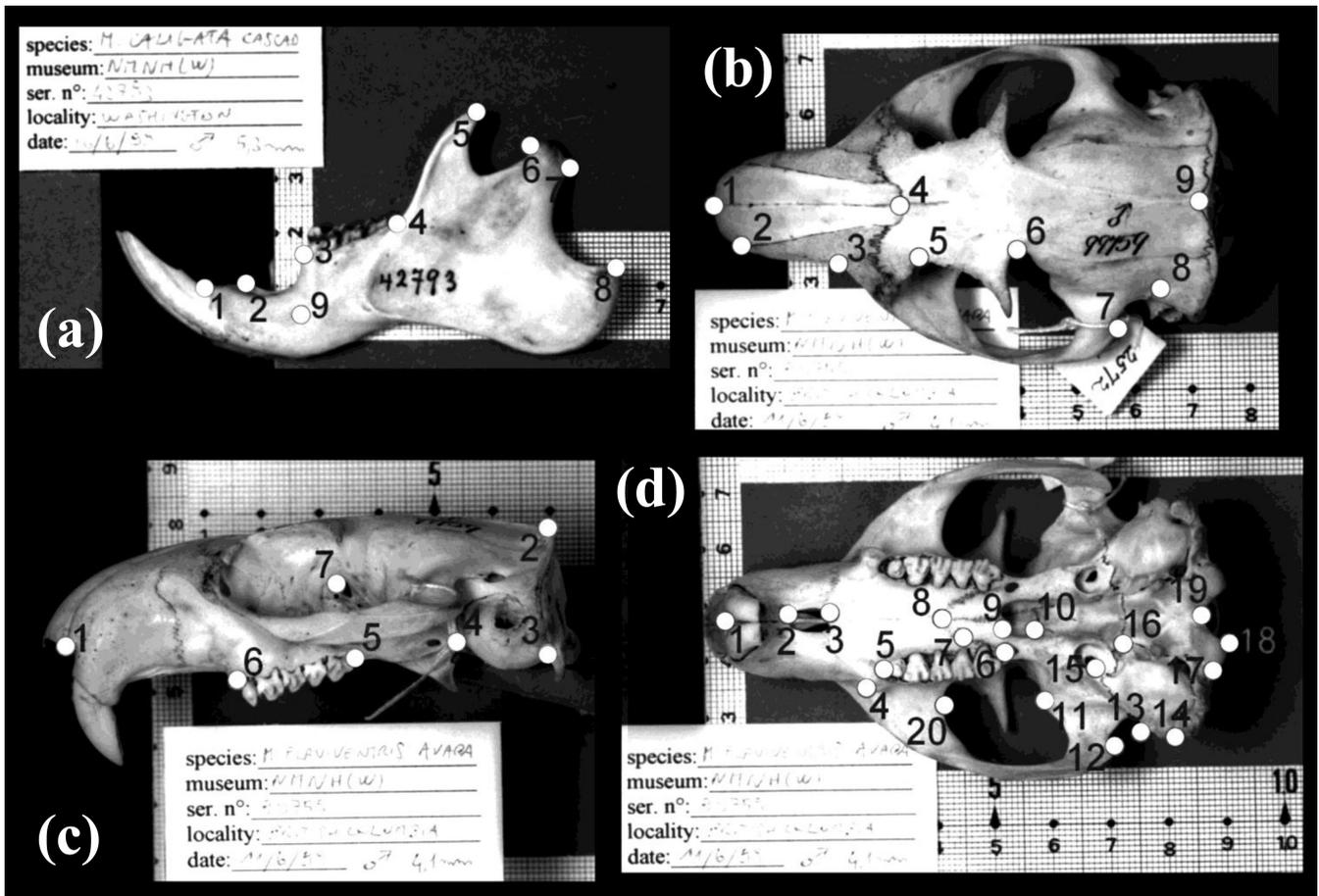


Figure 1 – Landmark configurations on mandibles (N=20; 9 landmarks) (a), and crania (N=49) in (b) dorsal (9 landmarks), (c) side (7 landmarks) and (d) ventral view (20 landmarks).

depth of field of about 1 cm (or less) of a standard scanner is inadequate for highly three-dimensional objects. As for the mandibles, 3D landmarks were also digitized directly on crania using a Microscribe. In each different view, as in Cardini and Thorington (2006), the landmarks shown in Fig. 1b,c,d represent a subset of anatomical points, in common between 2D and 3D data, selected from an originally larger configuration of landmarks.

All specimens belong to the collection of the United States Natural History Museum in Washington DC. A list of museum catalogue numbers can be made available from the author upon request. As an example, mandible raw data (3D and 2D photos) are made available in `nts` format as supplementary information online. The `nts` format is a commonly used ASCII (i.e., text) format that can be easily imported in programs of the TPS Series (Rohlf, 2014) and IMP Series (Sheets, 2014), Morphologika (O’Higgins and Jones, 2006), MorphoJ (Klingenberg, 2011) and the Geomorph R package (Adams and Otárola-Castillo, 2013). For this dataset, a fully worked out MorphoJ project is also provided for readers to use it as a guideline and learn how to replicate the main steps of the analysis on their own data. For this same didactical reason, I will sometimes include details about how to obtain specific analyses in MorphoJ and, in a few instances, PAST 2.17c (Hammer et al., 2001). For more detailed information on software, however, I refer readers to the extensive help files of those programs. Also, I want to stress that, although I tend to focus mainly (but not exclusively) on user-friendly software to aid beginners, analyses can be as well performed in advanced statistical environments such as R (R Core Team, 2014). For those interested in this, I made available as online supplementary material a simplified R script, which allows to perform the resampling version of the ANOVA model used in the new approach and not yet available in MorphoJ.

Analysis I: comparison of patterns and correlations (“traditional” approaches)

These are largely “traditional” approaches. They are mainly exploratory in nature. 2D and 3D datasets are superimposed separately using a Procrustes fit (Adams et al., 2013) to compute centroid size (henceforth, simply referred to as “size”) and Procrustes shape coordinates (henceforth, simply “shape coordinates”). This means that 2D and 3D shape data belong to different shape spaces each with its own specific Procrustes distance metric.

Patterns are visually compared across dataset (e.g., 2D dorsal view of the cranium vs its 3D counterpart) using box-plots (size) and scatter-plots of the first principal components (PCs) of the shape coordinates. Box-plots can be drawn in PAST after rearranging size data in columns (menu Plot, Barchart/Boxplots). PC scatterplots can be obtained in MorphoJ (menu Preliminaries, Generate covariance matrix, followed by Variation, Principal Component Analysis), as well as in PAST (menu Multivar, Principal Component Analysis using the Var-covar option). Variation at the opposite extremes of a PC axis is visualized using wire-frame shape diagrams (Klingenberg, 2008). These were computed in Morpheus et al. (Slice, 1999), but can be more directly obtained in MorphoJ, as part of the outcome of a PC analysis.

The covariation of 2D and 3D data is assessed using simple Pearson correlations for size and three different approaches for shape. The first of these approaches is a simple matrix correlation between the Procrustes shape distances from 3D and 2D data. Computations are performed in PAST (menu Multivariate, Mantel test using Euclidean distances from shape coordinates exported from MorphoJ³). In doing this,

³To generate the shape coordinates one could perform the Procrustes superimposition directly in PAST and then project the data in the tangent space in the same software. By doing this separately for each of the 2D datasets and the 3D landmarks, results should be identical to those of MorphoJ. However, a bug in the 2.17c version prevents from a correct computation of the 3D shape coordinates. Therefore I suggest to simply import

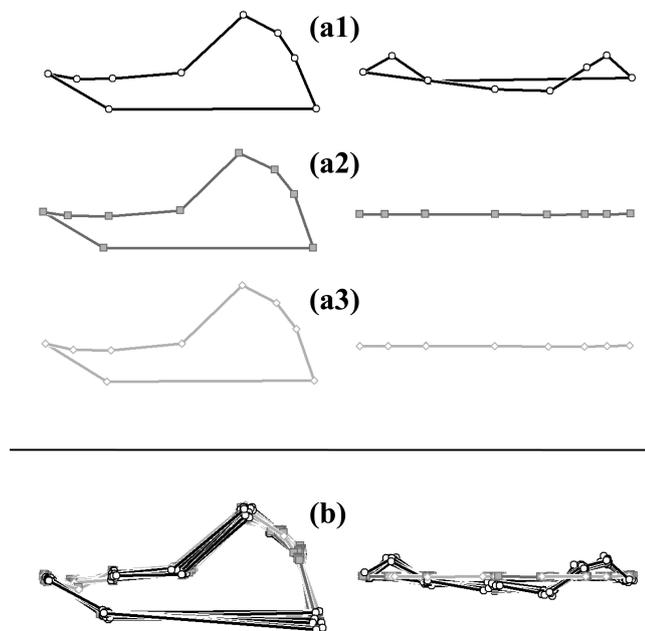


Figure 2 – (a) Mandibles: same individual seen in side (left) and from above (right) views using 3D data (a1), 2D pictures (a2), and 2D scans (a3). (b) Same views showing all specimens after the superimposition in the common shape space.

the shape coordinates from MorphoJ or check if the bug has been fixed in a more recent version of PAST.

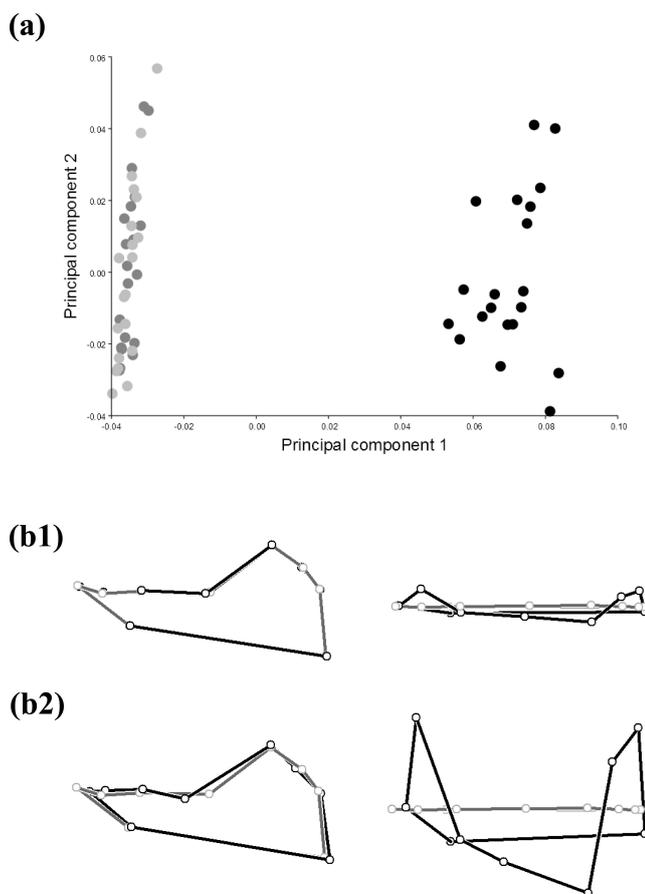


Figure 3 – (a) First two PCs of mandible shape coordinates showing two main clusters, which correspond to 2D data (negative scores, light and dark grey circles) and 3D data (positive scores, black circles); PC1 accounts for 58.9% and PC2 for 12.5% of total shape variance. (b) Shape wireframe diagrams corresponding to opposite extremes of PC1 are shown superimposed below the scatterplot (left, lateral view; right, view from above): grey is the negative extreme (2D data) and black the positive one (3D data); (b1) is without magnification; (b2) is with the positive extreme magnified five times relative to the observed score (i.e., if the observed score was 0.08, the shape being visualized corresponds to its extrapolation for a score of 0.4).

it is important to bear in mind that the test is only performed as a way to obtain the correlation value. The p value itself is meaningless in this specific context and must not be considered, as it is not the aim of the analysis and it would be incorrect because the data are not independent. The same reasoning applies to the RV coefficient (below), which is an overall measure of association between two sets of variables (i.e., an estimate of the strength of the covariance between 3D and 2D), and the test of variances and covariances (further below). Thus, as a second step, the RV coefficient is computed in MorphoJ (menu Covariation-Partial least squares, focusing only on the RV value in the results window). Finally, the similarity in variance-covariance matrices (VCV) from the two types of data is measured using again a simple matrix correlation in MorphoJ (menu Variation, Matrix correlation).

Analysis II: comparison of 2D and 3D data in the same shape space and ANOVA test (“new” approach)

The “new” approach requires bringing the 2D and 3D data into the same shape space. This apparently difficult operation can be achieved with a trivial expedient: adding a third zero coordinate to each landmark in the 2D data. This operation is manually performed in a spreadsheet, and allows to superimpose 2D and 3D data together and thus obtain shapes whose differences are measured in the same units of Procrustes shape distance, as they are all in a common shape space.

The common superimposition has no effect on the estimate of centroid size. In fact, if variance in size is significantly larger than differences between 2D and 3D estimates (i.e., the TTD error) can be tested regardless of whether the data are superimposed separately (approach I) or together (approach II). This type of test is done using the same ANOVA model routinely employed for other sources of error (Viscosi and Cardini, 2011). I am performing the ANOVA on size here, as part of the new approach (II), simply because this way all size and shape analyses are performed in parallel in the same section and using the same software (see shape ANOVA, below).

The 2D-3D common Procrustes analysis also has a negligible effect on shape distances within each dataset (2D vs. 3D). This is shown by comparing shape distances within a dataset (e.g., 2D dorsal view) after separate superimpositions (approach I) or a common superimposition (approach II). In all instances, matrix correlations are virtually equal to 1. For example, for the 2D dorsal view, the correlation between shape distances after a 2D superimposition and after a 3D GPA (2D and 3D superimposed together using a zero Z coordinates for 2D) is 0.99992. Because this result is marginal to the main aim, I am briefly mentioning it in the methodological description of the new approach. However, that shape distances are faithfully preserved within type of dataset (2D or 3D) regardless of whether 2D and 3D data are or are not in the same shape space may be specific to the study samples and should be preliminarily checked. This is a recommendation that I suggest to follow as long as we gain more experience on the effects of the common 2D-3D superimposition, but it seems unlikely that results will be very different from mine. This is because, when the tangent space approximation in both 2D and 3D is good, the amount of variation is relatively small and a common superimposition should fit the data about as well as separate ones. In the datasets I used, the tangent space approximation was indeed excellent with almost identical shape distances in the curved Procrustes shape space and in its tangent Euclidean space projection (correlations and slopes of shape distances both virtually equal to one in all 2D and 3D datasets – TPSSmall, Rohlf, 2014).

Data in the common shape space, for instance the 2D and 3D mandibles, should not, nevertheless, be compared directly as they are. This is because 2D and 3D shapes will be far apart in the space simply because of the third Z dimension (“thickness”), which is in fact present only in the real 3D data (Fig. 2 and 3). This difference is a loss of information in the 2D data but may not represent the most problematic source of inaccuracy. This is because, at least when variation is small as in a group of closely related species, it is likely to affect all individuals in a similar fashion. This type of systematic error, or bias, although important in absolute terms, can be controlled for. In fact, for obtaining accurate numerical results in 2D, what is really crucial is that the

relative shape distances among specimens in 2D are similar to those of 3D data. This would imply that, despite the loss of information in the third dimension, 2D landmarks accurately capture the relative shape differences among individuals in the sample.

Thus, the second step of the new approach is to mean center the datasets, which is achieved by removing (manually in a spreadsheet) the 2D as well as the 3D means from the respective shape data. This obviously does not change the relative structure of the similarity relationships within each dataset (3D and 2D), as it is easily verified by observing that matrix correlations of 2D (or 3D) data before and after mean centering are all equal to 1.

After the common superimposition and the mean centering, data can finally be compared as customary (Viscosi and Cardini, 2011): thus, as for size, if the 2D approximation introduces a negligible relative inaccuracy, shape differences among individuals should be significantly larger than the difference between 2D and 3D datasets. This is tested in an ANOVA with individuals as a random factor and TTD as the error term. The test can be performed either using parametric statistics (in MorphoJ or the car R package – Fox and Weisberg, 2011) or using a permutational ANOVA (in the vegan R package – Oksanen et al., 2013 – or in PERMANOVA – Anderson, 2001; Franklin et al., 2013, for an appendix with a tutorial). Actually, with a simple expedient (below), the whole procedure including mean centering and the parametric test (menu Variation, Procrustes ANOVA) can be performed in MorphoJ. This first require mean centering, an operation that I suggested to perform manually in a spreadsheet, even if it can also be achieved directly in MorphoJ by regressing shape data from the common superimposition onto a dummy variable which specifies if data are 2D or 3D. The dummy variable is created as a MorphoJ covariate, where groups are coded with -1 (or zero) for 2D data and 1 for 3D data. The residuals of this regression will be equivalent to the mean-centered data from the procedure described at the beginning of the paragraph. Then, in principle, one should be able to select the regression results in the MorphoJ project and specify the Procrustes ANOVA. The ANOVA simply requires a grouping variable (classifier) that uniquely identifies specimens (e.g., 1 for specimen A in both 2D and 3D data, 2 for specimen B in 2D and 3D data etc.). This classifier is used as the individual factor in the analysis. In practice, however, the current 1.06c version of the software does not allow users to perform ANOVAs on regression results. Unless this option becomes available in the future, the way to get around the problem is simply to export the residuals, re-import them and recreate the individual classifier to finally perform the Procrustes ANOVA⁴.

The ANOVA test offers an objective method to verify if one can be confident that 2D data are accurate in relation to the amount of variation in the study sample. This is because TTD, like all types of measurement error, is never absolute and must be assessed in relation to the differences among the organisms in the study sample (Arnqvist and Martensson, 1998). The ANOVA does provide this objective and essential information but an exclusive focus on statistical significance may be unwise. Results could be highly significant even if TTD is relatively large and 2D data are therefore not as good as it might be desirable. To avoid this pitfall, the magnitude of the effect being tested is also estimated. This is equal to the proportion of variance accounted for by individual variation with the remaining unexplained variance being the TTD error. The proportion is easily computed manually as the ratio between the individual sum of squares (SSQ) and the total SSQ (individual plus error) from the output table of MorphoJ.

Finally, to further increase confidence in the accuracy of 2D data, the mean centred shape variables are subjected to a UPGMA (Unweighted Pair Group Method with Arithmetic mean) cluster analysis using Euclidean distances in PAST (menu Multivar, Cluster analysis, using Eu-

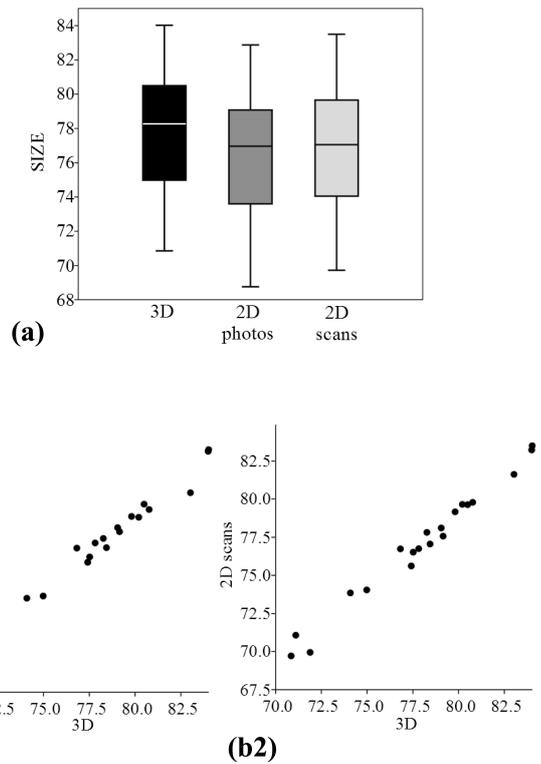


Figure 4 – Mandibles: example of graphical comparison of patterns in 3D and 2D estimates of size. (a) Box-plot of size for 3D, and 2D photos and scans. (b) Scatterplots of size from 2D photos (b1) and scans (b2) compared to 3D data.

clidean distances and paired group algorithm). The resulting phenogram summarizes shape relationships as a tree and generally preserves well small scale distances. In this tree, I counted how many specimens clustered with their 3D counterpart: if TTD is small, the expectation is that 2D and 3D data of most, if not all, specimens will be “sister” in the terminal branches of the tree.

Results and Discussion

Analysis I: comparison of patterns and correlations (“traditional” approaches)

Box-plots of size are similar regardless of whether size is estimated using 3D data or 2D images. This is exemplified for the mandibles in Fig. 4a. Box-plots of cranial data are similarly congruent across data type (2D vs 3D), although congruence seems slightly less in box-plots from the lateral view of the cranium (not shown). In all datasets, average 2D size has a relative deviation from its 3D estimates that is equal to or less than 2% of the 3D mean itself. For instance, using mandible 2D photos, mean size is 76.6 mm with an absolute deviation of 1.4 mm from the 78.0 mm mean estimated from 3D data, and therefore a relative deviation of $1.4/78.0 = 0.017$ or 1.7%. The corresponding relative deviations for the standard deviations are somewhat larger and range from less than 1% (mandible) to almost 6% (ventral cranial data). The correlations of 2D data with 3D ones are about 0.99 or higher in all structures and views (Tab. 1; Fig. 4b).

Table 1 – Separate superimpositions correlational analysis.

view and structure	3D vs. 2D	Size r	Shape matrix r	RV	VCV
mandible (ma)	photos	0.987	0.790	0.873	0.843
mandible (ma)	scans	0.991	0.667	0.713	0.736
dorsal cranium (dc)	photos	0.996	0.686	0.703	0.717
lateral cranium (lc)	photos	0.995	0.639	0.703	0.733
ventral cranium (vc)	photos	0.998	0.517	0.623	0.513

⁴MorphoJ will force the user to re-superimpose the data after re-importing them. However, as residuals already came from superimposed data, this will not make any appreciable difference. Users are suggested to restore size variation (a simple multiplication of shape coordinates by the corresponding centroid size) before reimporting the data. Otherwise, they need to remember that the ANOVA on centroid has to be done on the original data and that any result on size obtained from the mean-centered shape dataset has to be ignored!

Overall, results from the visual inspection of patterns, the descriptive statistics and plots, and the correlations indicate that size is quite accurate in 2D. There is a small error, however, and some parameters such as standard deviations may be more strongly affected by TTD errors. Interestingly, some of the relative deviations of the means were negative (lateral, -1.7%, and ventral views, -0.2%). This means that the 3D mean is slightly smaller than its 2D estimates. As depth in the third dimension is expected to increase linear distances from the centroid in 3D data compared to the corresponding 2D estimates, the overestimate in 2D photos is counter-intuitive. It may be related to inaccuracies in the scale factor used to convert pixels into millimeters. This is plausible because, especially in the lateral view, which has the highest negative mean deviation, the millimeter paper ruler used to scale the images lies further back relative to the plane where most landmarks are. This means that the number of pixels, corresponding to the length measured in TPSDig to scale the data and convert pixels into millimeters, is smaller than it should. That makes the denominator of the ratio used to scale the data smaller and therefore leads to an overestimate of the factor by which coordinates in pixels are multiplied, thus slightly inflating centroid size. This is overall a small error but one that can be avoided with a careful consideration of the effects of apparently small details, such as the positioning of the scale factor, during data collection. Alternatively, if the relative distance of the scale and the landmarks from the camera is known, one could be able to correct for this type of error.

The comparison of patterns (ordinations and shape diagrams) of shape change in 2D and 3D after separate superimpositions are presented and discussed only for mandibles and ventral crania. These datasets respectively have the highest and lowest correlations of 3D vs 2D shape distances (see below). Results for dorsal and side views of the cranium are largely similar to those of the ventral view. Scatterplots of the first two PCs of mandibular shape indicate a degree of congruence between 3D data and 2D photos. This is somewhat less evident for 2D scans, where the scatterplot looks as if the direction of largest variation in the 3D data is rotated of almost 90 degrees (Fig. 5: a1 vs. a3). However, the general pattern of shape variation captured by PC1 in all three sets of mandibular data is similar and concerns changes from a slender and slightly elongated mandible to a more robust, relatively shorter and deeper one (Fig. 5).

It is important to observe that the comparison of patterns using PC1 is meaningful because this dimension explains almost twice the amount of mandibular shape variance explained by PC2, and its orientation should therefore be relatively stable. This is true also for all 3D and 2D cranial datasets (below), except the 2D ventral crania, whose PC1 explains about the same amount of variance as PC2 (ca. 15% and 13% respectively). PC2, however, is on average (in all datasets) just 1.3 times larger, in terms of variance, than PC3. As with PC2 relative to PC3, also other higher order PCs up to the 10th tend to explain only 1.2–1.5 more variance than the next PC. Thus, in contrast to PC1, variation summarized by other PCs may not be strictly comparable, because shape variance becomes almost circular in the subspace of PCs other than PC1 and therefore the orientation those PCs becomes unstable.

The congruence between 3D and 2D data is much less evident in scatterplots and shape diagrams of ventral crania (Fig. 6). As anticipated, shape variation in 2D ventral crania is almost isotropic, thus making more difficult a truly informative comparison of scatterplots and shape changes. Indeed, in general, there seems to be no evident similarity between the scatter of points on PC1-2 from 2D and 3D analyses. Contrary to mandibular data, this lack of congruence is also apparent in the PC1 shape diagrams: the aspects of shape variation captured by this axis are not easily comparable across type of data, and similarities, if present, are not evident by eye. In fact, the shape corresponding to the negative extreme of PC1 in the 3D analysis appears roughly similar to that of the negative extreme in the 2D analysis, but this corresponds to specimen 31 in the 2D dataset and this specimen is actually at the positive extreme of PC1 using 3D landmarks. This indicates that the difference in PC1 is not just a simple matter of flipping the axis, because of the arbitrary sign of PCs: 3D and 2D data do seem to suggest

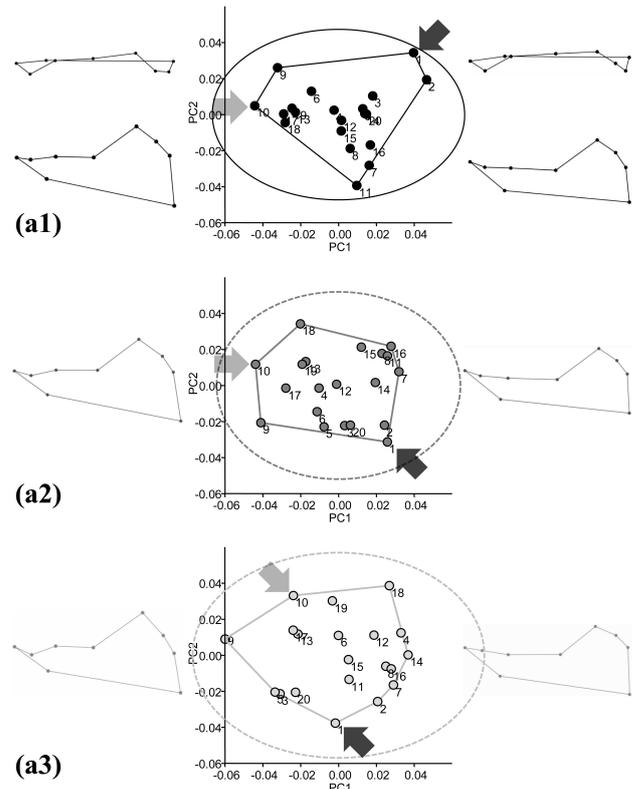


Figure 5 – Scatterplots of the first two PCs of shape after separate superimpositions of mandible 3D and 2D datasets, and visualization of shape variation at opposite extremes of PC1 (for 3D data, the upper diagrams show depth variation, which is missing in 2D data): (a1) 3D (PC1 30.7%; PC2 15.4% of total shape variance); (a2) 2D photos (PC1 34.4%; PC2 22.1%); (a3) 2D scans (PC1 39.4%; PC2 21.7%). In this and the next figure, 95% equal frequency ellipses and convex hulls are shown in each scatterplot; also, arrows are used to emphasize the positions of two specific individuals in scatterplots as an aid to compare directions of change and patterns of differences in 2D and 3D analyses.

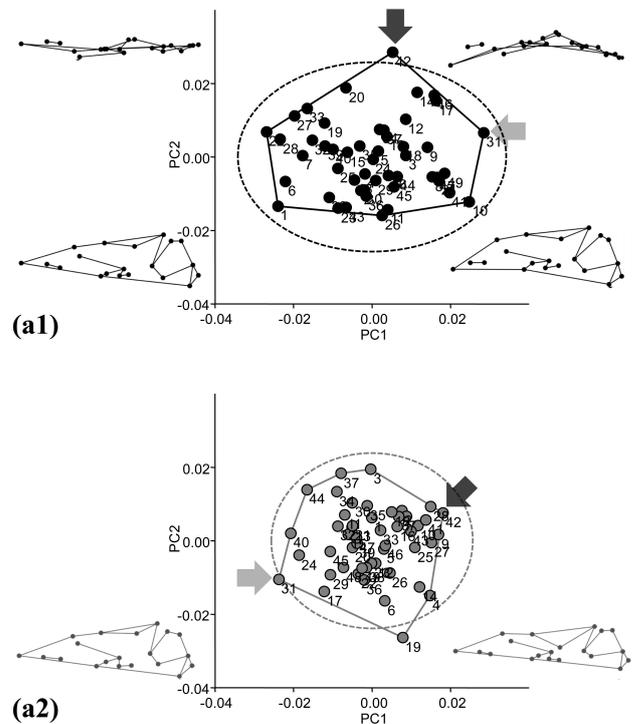


Figure 6 – Scatterplots of the first two PCs of shape after separate superimpositions of ventral crania datasets and visualization of shape variation at opposite extremes of PC1 magnified three times (for 3D data, the upper diagrams show depth variation, which is missing in 2D data): (a1) 3D (PC1 16.0%; PC2 9.6% of total shape variance); (a2) 2D photos (PC1 14.8%; PC2 12.9%).

different patterns of shape variation, and 2D shapes fail to capture any dominant direction of change in ventral crania.

Variation in mandible and cranium depth (Fig. 5–6a1 upper diagrams) cannot be compared, as this is found only in the 3D data. This clearly represents an important limitation to the interpretation of patterns of shape variation. Indeed, even if the 2D approximation of relative shape differences is good and the numerical results accurate, the third physical dimension is not there and any change in that direction will be inevitably missed in the photos.

Unlike the exploration of patterns from the first PCs presented above, all further analyses more rigorously assess the congruence of 2D and 3D data in the full shape spaces. Correlations of shape distances, *RV* coefficients and correlations of VCV produce similar results (Tab. 1). In all instances, using any of the three statistics, the highest congruence is between 3D data and 2D photos of mandibles (ca. 0.8 or higher) and the lowest congruence is found in the ventral cranium (ca. 0.6). All other datasets have intermediate degrees of similarity (ca. 0.7).

Measuring the strength of association between 3D and 2D data using correlational approaches is a step forward to numerically assess TTD. However, one is left with crude numbers that have to be subjectively interpreted to decide whether 2D are or are not adequately accurate as a proxy for 3D structures. To overcome this limitation, the next section presents and discusses the results of the new approach.

Analysis II: comparison of 2D and 3D data in the same shape space and ANOVA test (“new” approach)

Results of the ANOVA for size are presented in Tab. 2. This shows that in all structures and views, individual size variation is significantly larger than TTD error and always explains more than 95% of variance in the data. This is consistent with the high-correlations and the similarities in the box-plots reported in the previous section. The ANOVA thus strengthens the conclusion that estimates of size in 2D are generally very accurate and closely mirror 3D estimates. In fact, size data could also be mean-centered, using the same method and reasoning as for shape (i.e., that relative accuracy in results is more important than absolute accuracy). However, size variation was (and probably is in most datasets) so large compared to the total TTD error that its operation is unlikely to make a difference.

Mean-centred shape data after a common superimposition also show that individual variation is highly significant relative to TTD error in all structures and views, and it explains between 80% and 90% of total variance (Tab. 3). This suggests that TTD error is negligible not only in relatively flat mandibles, but also in highly three-dimensional crania. The error, however, is about 10% in mandibles (slightly more than 10% in the less accurate scans) but appreciably larger in crania, where it accounts for about 20% of total variance.

Phenograms of mean centred shapes (shown in Fig. 7 for mandibles and ventral crania only) show that in 85.0% of cases the 2D mandibular data of a specimen clusters together with the corresponding 3D description of the same individual. This is indicative of differences due to TTD smaller than differences among individuals for the large majority of specimens, which is consistent with the ANOVA results. However, only 59.2%, 55.1% and 36.7% of respectively ventral, dorsal and lateral cranial data performed similarly well. Thus, despite highly significant ANOVAs, cranial data are clearly less well approximated in 2D than mandibles. This is expected, as crania are more complex and three-dimensional than mandibles. It is, in contrast, less obvious why, despite the slightly larger error in ventral crania, their phenogram is more accurate in terms of clustering together 2D and 3D replicas. It is hard to say why this happens but one might be tempted to speculate that it is related to the larger number of landmarks in the ventral cranium, which reduce the probability that 2D and 3D data are mismatched.

Conclusions and recommendations

All approaches are in good agreement in suggesting that TTD is small for size and generally negligible. Also, the battery of analyses performed in this study consistently indicates that 2D marmot mandibular data are generally accurate and, for shape, photos perform slightly

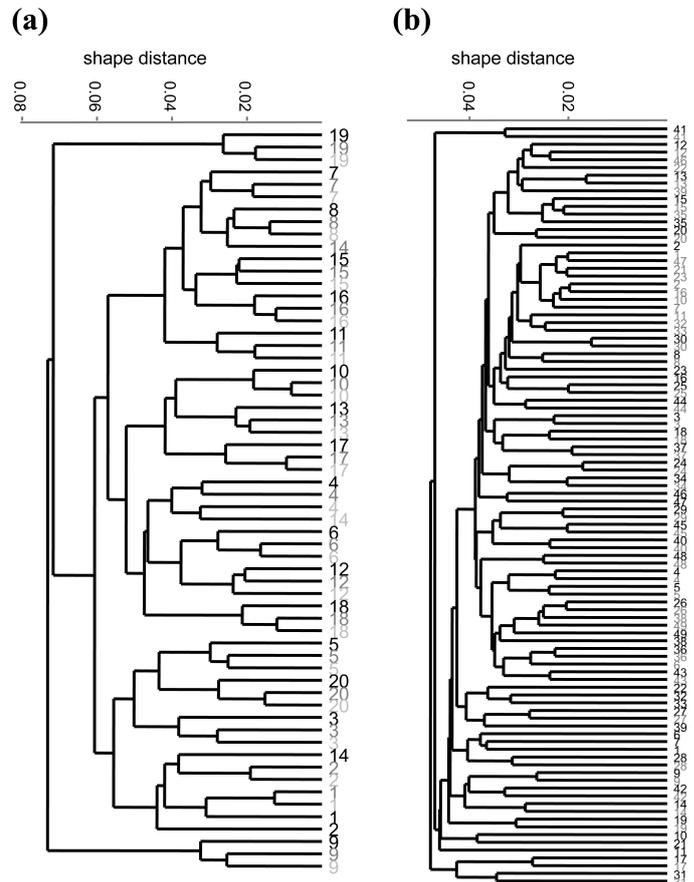


Figure 7 – Phenograms (UPGMA, Euclidean distances) of (a) mandibles and (b) ventral crania using mean-centered data after a common superimposition: numbers indicating specimen identity are shown in black for 3D (both mandibles and crania), grey for 2D photos (both datasets) and light grey for 2D scans (only mandibles).

better (higher correlations of 3D-2D data; smaller TTD error in the ANOVA) than scans. Cranial data, in contrast, provide less clear conclusions. The ANOVA does indicate that the TTD error is negligible relative to cranial sample variation. However, TTD accounts for about 20% of total variance. To help scaling this percentage in relation to observed variability in nature, one could consider that Nagorsen and Cardini (2009) reported that interspecific differences in mandibular shape among species of North American marmots account for about 13% to 36% of total shape variance. On average, therefore, TTD error in cranial data may be about as large as differences among closely related species (assuming that interspecific differences in crania mirror those seen in mandibles). Patterns, correlational analyses and phenograms all suggested that the TTD error is indeed quite large in crania and the congruence of 2D and 3D data modest. Thus, it may be possible to use 2D photos to study marmot crania. However, as they are highly three-dimensional, there will be a degree of inaccuracy, which makes the 2D approximation likely unsuitable for intra-specific analyses. Even when 2D cranial studies are at supra-specific level, they might require cautious interpretations and a clear warning to readers about potential inaccuracies. This is something that clearly applies to my earlier studies using 2D cranial data on marmots (references in Cardini et al., 2007). In those studies I often did not have the 3D data to test TTD and certainly did not know back then how to do it convincingly. I must therefore now issue a “post-publication” warning that all those analyses should be interpreted with a good degree of caution. This warning mainly concerns inter-specific relationships between less distinctive species, as large differences such as the remarkable divergence in the youngest living marmot population, that of the highly endangered Vancouver Island marmot (*Marmota vancouverensis*), have been confirmed in multiple studies including several 2D analyses of mandibles (Cardini, 2003; Nagorsen and Cardini, 2009; Cardini et al., 2009) as well as 3D studies of cranial data (Cardini et al., 2007).

In interpreting the magnitude of the TTD error in all datasets one should also bear in mind that its estimate in this study also includes positioning and digitization errors. Potentially these components could be individually quantified by repeating each step of the data collection at least twice. For instance, for the 3D data, crania or mandible should be repositioned and digitized twice in each position. For readers interested in this, Arnqvist and Martensson (1998) provide an almost exhaustive list of potential sources of errors and Viscosi and Cardini (2011) further discuss this point and exemplify how to partition different sources of variance in a multivariate ANOVA.

The issue of the relative importance of different sources of error was raised also by a reviewer, who remarked that “it would be really interesting to see a comparison of TDD error relative to measurement error, which could be obtained by repeated measurements of specimens . . . , as a way of understanding the relative magnitudes of these two types of error”. As anticipated, this is an important suggestion and something that can be easily incorporated in the study design. MorphoJ, for instance, allows to partition two different sources of error: one could use TTD as the first one and positioning and digitization error (POSDIG) as the second one. This could not be done in this study, however, as replicas were not available within each dataset (2D or 3D).

It seems unlikely that POSDIG is a source of measurement error as large as that due to the 2D flattening of a 3D structure. This is not only a reasonable expectation, unless one has a problematic protocol for data collection which leads to a very big POSDIG. It is also suggested by the mandible analysis. In this analysis, 2D replicas are in fact available, as both photos and scans were used. Thus, these data provide an estimate of POSDIG at least in 2D. This estimate is inflated, because it contains a third source of error, which is the different device (i.e., a camera or a flat-bed scanner) used to obtain the images. Nevertheless, the phenogram in Fig. 7a shows that in 16 out of 20 individuals (i.e., 80% of the total), 2D replicas cluster together as “sisters”, while specimens measured using 3D landmarks are more distant in all but three cases. Indeed, if we assume, with a degree of approximation, that sum of squares are comparable in the TTD analysis and in a separate analysis (not shown) I did to estimate digitization error (DIG) by re-digitizing both mandible photos and scans, TTD (or, better, TTD minus DIG) is 3–4 times larger than DIG for shape and 6–60 times bigger for size.

Similar exploratory comparisons of TTD and POSDIG (or just DIG) could not be done for cranial data. However, in cranial data, TTD was relatively larger than in mandibles, whereas POSDIG is likely to be the same. Thus, it does seem that, when 2D images of 3D structures are analysed, TTD is the main source of error, although other sources contribute and the different magnitudes could be easily estimated, as long as replicas are available within both 2D and 3D data.

There is another aspect of this study, and more precisely of the ANOVA, on which I would like to recommend a degree of caution. This is, however, not specific to this type of analysis and actually has a broader relevance in all tests using Procrustes shape data or more generally data from which some information has been removed. Indeed, this is always the case with Procrustes shape coordinates, as they are obtained

by standardizing size and minimizing positional (translation and rotation) differences. This introduces a redundancy in the shape coordinates so that the real number of informative dimensions (the degrees of freedom — df) is actually less than the number of variables. If the statistical software used for a parametric test cannot “see” the redundancy, it might incorrectly compute df and therefore produce unreliable *p* values. A simple way to avoid this potential issue is to perform resampling tests using shape distances, as in the permutational version of the ANOVA employed in this study. Another option is to perform parametric tests on the matrix of all PCs with non-zero eigenvalues (i.e., variances). This is exactly the same information as in the original shape coordinates but, by removing PCs with zero variance, uninformative dimensions are discarded and df correctly correspond to the number of variables. This option, unfortunately, is not available in MorphoJ Procrustes ANOVA, which only allows to perform the test using shape coordinates. However, because MorphoJ is designed to analyse Procrustes shape data, it will likely compute df and *p* values correctly. Users should, nevertheless, quickly check that this is the case. This is easily done by comparing df computed in the shape Procrustes ANOVA in MorphoJ, and divided by either sample size *N* (error term) or *N* – 1 (individual term), to the number of PCs with non-zero eigenvalues for the same data: these numbers should be exactly the same. If not, at least for the isotropic model, one can manually correct the computations of the F ratio and look for its correct *p* value in a published F ratio table. For instance, in the mandible dataset, a PCA of the regression residuals produces 20 PCs with non-zero eigenvalues; the ANOVA indicates 380 and 400 df respectively for the individual and error terms, which, divided by 19 and 20 respectively, gives 20, as expected if all computations are correct.

To summarize, TTD can and should be assessed at least preliminary in a 2D study of 3D structures. This recommendation does not add anything to what Roth (1993) had already rightly and convincingly argued more than 20 years ago, an advice I myself neglected to follow. However, now the understanding of Procrustes shape data is much deeper than in the early days of GMM applications to biology (Adams et al., 2013). Thus, I have presented a variety of approaches to assess TTD and suggested a way to bring 2D and 3D data into the same shape space for a more powerful comparison, which includes a statistical test. These methods are fairly simple and at least in their parametric version can be easily performed largely in MorphoJ. In practice, one simply needs at least a subsample, representative of the variation expected in a study, for which both 2D and 3D data are available. Data availability could be the main obstacle to a routine application of TTD assessment. Indeed, despite technological progresses, 3D data are still more expensive and less easy to access especially in developing countries and countries with a chronic lack of funding for basic research. However, 3D data can now be acquired using inexpensive photogrammetry. This is nicely exemplified by the simplified but pioneering protocol of Fadda et al. (1997), and is becoming more popular, accurate and achievable thanks to cheap digital cameras and new software (Falkingham, 2012; Katz and Friess, 2014). Alternatively, one might be able to reconstruct, with

Table 2 – ANOVA testing inter-individual variation vs TTD differences in size.

3D vs. 2D			SSQ	df	MSQ	F	R ² *	P(10000 perm.)	P(param.)
ma	photos	individuals	562.3	19	29.5942	26.546	96.2%	0.0001	0.0001
		error	22.3	20	1.1148	0.038			
ma	scans	individuals	571.8	19	30.0969	54.373	98.1%	0.0001	0.0001
		error	11.1	20	0.5535	0.019			
dc	photos	individuals	3221.6	48	67.116	40.432	97.5%	0.0001	0.0001
		error	81.3	49	1.66	0.025			
lc	photos	individuals	2181.6	48	45.45	49.696	98.0%	0.0001	0.0001
		error	44.8	49	0.915	0.020			
vc	photos	individuals	4808.6	48	100.178	473.760	99.8%	0.0001	0.0001
		error	10.4	49	0.211	0.002			

* Expressed as percentage of variance explained by individuals.

Table 3 – Common shape space-mean centred data ANOVA testing inter-individual variation vs. TTD differences in shape.

	3D vs. 2D		SSQ	df	MSQ	F	R ² *	P(10000 perm.)	P(param.)
ma	photos**	individuals	0.060583	380	0.000159	9.418	90.0%	0.0001	0.0001
		error	0.006771	400	0.000017				
ma	scans**	individuals	0.065236	380	0.000172	7.350	87.5%	0.0001	0.0001
		error	0.009343	400	0.000023				
dc	photos**	individuals	0.077482	960	0.000081	4.720	82.2%	0.0001	0.0001
		error	0.016757	980	0.000017				
lc	photos**	individuals	0.073507	672	0.000109	4.456	81.4%	0.0001	0.0001
		error	0.016839	686	0.000025				
vc	photos	individuals	0.066257	2544	0.000026	4.132	80.2%	0.0001	0.0001
		error	0.016370	2597	0.000006				

* Expressed as percentage of variance explained by individuals.

** MorphoJ provides also parametric tests using the Pillai's trace and the multivariate approach described by Klingenberg et al. (2002), which does not require isotropic variation, when sample size is large enough relative to the number of shape variables. This was the case in all datasets except the ventral cranium. For these data, Pillai's trace was always larger than 9 and $p < 0.0001$.

some effort, 3D data using linear distances (Monteiro et al., 1997) and techniques related to the truss method (Strauss and Bookstein, 1982), which is implemented in Morphueus et al. (Slice, 1999). ☞

References

- Adams D.C., Otárola-Castillo E., 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4: 393–399.
- Adams D., Rohlf F., Slice D., 2013. A field comes of age: geometric morphometrics in the 21st century. *Hystrix* 24(1): 7–14. doi:10.4404/hystrix-24.1-6283
- Anderson M.J., 2001. Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 626–639.
- Arnqvist G., Martensson T., 1998. Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae* 44: 73–96.
- Bastir M., Rosas A., O'Higgins P., 2006. Craniofacial levels and the morphological maturation of the human skull. *Journal of Anatomy* 209: 637–654.
- Bookstein F.L., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis* 1: 225–243.
- Bush M.A., Bush P.J., Sheets H.D., 2011. Similarity and match rates of the human dentition in three dimensions: relevance to bitemark analysis. *International Journal of Legal Medicine* 125: 779–784.
- Cardini A., 2003. The Geometry of the Marmot (Rodentia: Sciuridae) Mandible: Phylogeny and Patterns of Morphological Evolution. *Systematic Biology* 52: 186–205.
- Cardini A., Elton S., 2007. Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology* 126: 121–134.
- Cardini A., Loy A., 2013. On growth and form in the “computer era”: from geometric to biological morphometrics. *Hystrix* 24(1): 1–5. doi:10.4404/hystrix-24.1-8749
- Cardini A., O'Higgins P., 2005. Post-natal ontogeny of the mandible and ventral cranium in *Marmota* species (Rodentia, Sciuridae): allometry and phylogeny. *Zoomorphology* 124: 189–203.
- Cardini A., Thorington R.W., Jr., 2006. Postnatal ontogeny of marmot (Rodentia, Sciuridae) crania: allometric trajectories and species divergence. *Journal of Mammalogy* 87: 201–215.
- Cardini A., Tongiorgi P., 2003. Yellow-bellied marmots (*Marmota flaviventris*) “in the shape space” (Rodentia, Sciuridae): sexual dimorphism, growth and allometry of the mandible. *Zoomorphology* 122: 11–23.
- Cardini A., Thorington R.W., Jr., Polly P.D., 2007. Evolutionary acceleration in the most endangered mammal of Canada: speciation and divergence in the Vancouver Island marmot (Rodentia, Sciuridae). *Journal of Evolutionary Biology* 20: 1833–1846.
- Cardini A., Nagorsen D., O'Higgins P., Polly P.D., Thorington R.W., Tongiorgi P., 2009. Detecting biological distinctiveness using geometric morphometrics: an example case from the Vancouver Island marmot. *Ethology Ecology and Evolution* 21: 209–223.
- Cavalcanti M.J., Monteiro L.R., Lopes P., 1999. Landmark-based morphometric analysis in selected species of serranid fishes (Perciformes: Teleostei). *Zoological Studies* 38: 287–294.
- Close R., Friedman M., 2014. Probing the third dimension: are morphospaces derived from 2D and 3D fossil fish crania congruent? Abstracts of the 62nd Symposium on Vertebrate Palaeontology and Comparative Anatomy and 23rd Symposium on Palaeontological Preparation and Curation with the Geological Curators' Group, York, UK. Available from http://www.svpca.org/years/2014_york/abstracts.pdf
- Fadda C., Faggiani F., Corti M., 1997. A portable device for the three dimensional landmark collection of skeletal elements of small mammals. *Mammalia* 61: 622–627.
- Falkingham P.L., 2012. Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaentologia Electronica* 15: 15 pp.
- Fox J., Weisberg S., 2011. *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA.
- Franklin D., Cardini A., Flavel A., Kuliukas A., Marks M.K., Hart R., Oxnard C., O'Higgins P., 2013. Concordance of traditional osteometric and volume-rendered M3CT interlandmark cranial measurements. *International Journal of Legal Medicine* 127: 505–520.
- Gunz P., Mitteroecker P., 2013. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix* 24(1): 103–109. doi:10.4404/hystrix-24.1-6292
- Hammer O., Harper D., Ryan P., 2001. PAST: Palaeontological statistics software package for education and data analysis. *Paleontol. Electron.* 4(1): 1–9.
- JCGM (Joint Committee for Guides in Metrology), 2008. International Vocabulary of Metrology—Basic and General Concepts and Associated Terms. JCGM 200:2008. Available from http://www.bipm.org/utls/common/documents/jcgm/JCGM_200_2008.pdf
- Katz D., Friess M., 2014. 3D from standard digital photography of human crania — a preliminary assessment. *American Journal of Physical Anthropology* 154: 152–158.
- Klingenberg C.P., 2008. Novelty and “Homology-free” Morphometrics: What's in a Name? *Evolutionary Biology* 35: 186–190.
- Klingenberg C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* 11: 353–357.
- Klingenberg C.P., 2013. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *Hystrix* 24(1): 15–24. doi:10.4404/hystrix-24.1-7691
- Klingenberg C.P., Barluenga M., Meyer A., 2002. Shape Analysis of Symmetric Structures: Quantifying Variation among Individuals and Asymmetry. *Evolution* 56: 1909–1920.
- MacLeod N., Steart D., Pearce L., Bartleet-Cross C., Nedza C., Frost A., Rose K., 2014. New morphologies for old: application of three dimensional surface analysis and computer vision/machine learning techniques to comparative anatomy with special emphasis on vertebrate paleontology. Abstracts of the 74th Annual Meeting of the Society of Vertebrate Paleontology, Berlin. Available from <http://vertpaleo.org/CMSPages/GetFile.aspx?nodeguid=b51eb899-273b-4edb-ae8a-ae067050a3a8>
- Márquez F., Amoroso R., Sainz M.F.G., Van der Molen S., 2010. Shell morphology changes in the scallop *Aequipecten tehuelchus* during its life span: a geometric morphometric approach. *Aquatic Biology* 11(2): 149–155.
- Monteiro L., Cavalcanti M., Sommer H., 1997. Comparative ontogenetic shape changes in the skull of Caiman species (Crocodylia, Alligatoridae). *Journal of Morphology* 231: 53–62.
- Nagorsen D.W., Cardini A., 2009. Tempo and mode of evolutionary divergence in modern and Holocene Vancouver Island marmots (*Marmota vancouverensis*) (Mammalia, Rodentia). *Journal of Zoological Systematics and Evolutionary Research* 47: 258–267.
- O'Higgins P., Jones N., 2006. Morphologika. Tools for statistical shape analysis. Hull York Medical School. Available from <http://sites.google.com/site/hymsfme/resources>
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H., 2013. *vegan: Community Ecology Package*. R package version 2.0-10. Available from <http://CRAN.R-project.org/package=vegan>
- Oxnard C., O'Higgins P., 2011. Biology Clearly Needs Morphometrics. Does Morphometrics Need Biology? *Biological Theory* 4: 84–97.
- Pizzo A., Mercurio D., Palestini C., Roggero A., Rolando A., 2006. Male differentiation patterns in two polyphenic sister species of the genus *Onthophagus* Latreille, 1802 (Coleoptera: Scarabaeidae): a geometric morphometric approach. *Journal of Zoological Systematics and Evolutionary Research* 44: 54–62.
- R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>
- Rohlf F., 2014. Tps Series. Department of Ecology and Evolution, State University of New York, Stony Brook, New York. Available from <http://life.bio.sunysb.edu/morph/index.html>
- Roth V., 1993. On three-dimensional morphometrics, and on the identification of landmark points. *Contributions to Morphometrics*. Museo Nacional de Ciencias Naturales, Madrid: 41–61.
- Rufino M.M., Abelló P., Yule A.B., 2006. Geographic and gender shape differences in the carapace of *Liocarcinus depurator* (Brachyura: Portunidae) using geometric morphometrics and the influence of a digitizing method. *Journal of Zoology* 269: 458–465.
- Sheets H.D., 2014. IMP Series. IMP-integrated morphometrics package. Department of Physics, Canisius College, Buffalo, NY. Available from <http://www.canisius.edu/~sheets/> and <http://www3.canisius.edu/~sheets/IMP%208.htm>
- Slice D.E., 1999. Morphueus et al. *Ecology and Evolution*. State University of New York, Stony Brook.
- Strauss R.E., Bookstein F.L., 1982. The Truss: Body Form Reconstructions in Morphometrics. *Systematic Biology* 31: 113–135.
- Viscosi V., Cardini A., 2011. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. *PLoS ONE* 6(10): e25630. doi:10.1371/journal.pone.0025630
- Watanabe A., 2014. R2D3 and C3PO: A companion tool for testing the fidelity of two-dimensional geometric morphometric data in a three-dimensional world of vertebrate crania. Abstracts of the 74th Annual Meeting of the Society of Vertebrate Paleontology, Berlin. Available from <http://vertpaleo.org/CMSPages/GetFile.aspx?nodeguid=b51eb899-273b-4edb-ae8a-ae067050a3a8>