[1]*Dipartimento di Scienze Chimiche e Geologiche, Universita` di Modena e Reggio Emilia, Modena, Italy;* [2]*School of Anatomy, Physiology and Human Biology, The University of Western Australia, Crawley, WA, Australia*

Left, right or both? Estimating and improving accuracy of one-side-only geometric morphometric analyses of cranial variation

ANDREA CARDINI[1,2]

Abstract

Procrustes-based geometric morphometric analyses of bilaterally symmetric structures are often performed using only one side. This is particularly common in studies of cranial variation in mammals and other vertebrates. When one is not interested in quantifying asymmetry, landmarking one side, instead of both, reduces the number of variables as well as the time and costs of data collection. It is assumed that the loss of information in the other half, on which landmarks are not digitized, is negligible, but this has seldom been tested. Using 10 samples of mammalian crania and a total of more than 500 specimens, and five different landmark configurations, I demonstrate that this assumption is indeed easily met for size. For shape, in contrast, one-side landmarking has potentially more severe consequences on the estimates of similarity relationships in a sample. In this respect, microevolutionary analyses of small differences are particularly affected, whereas macroevolutionary studies are fairly robust. In almost all instances, however, a simple preliminary operation improves accuracy by making one-side-only shape data more similar to those obtained by landmarking both sides. The same operation also makes estimates of allometry more accurate and improves the visualization. This operation consists in estimating the missing side by a mirror reflection of bilateral landmarks. In the Supporting Information, I exemplify how this can be easily done using free user-friendly software. I also provide an example data set for readers to repeat and learn the steps of this simple procedure.

Key words: Allometry – asymmetry – mammals – mirror reflection – Procrustes superimposition

Introduction

Morphometricians frequently study only half of bilaterally symmetric structures instead of landmarking and analysing both sides. This often happens in studies of cranial variation and has been common since the early days of Procrustes-based geometric morphometrics (PGMM; Adams et al. 2013; Cardini 2013). In fact, biology offers many examples of object symmetry, where an internal plane separates left and right mirror sides (Klingenberg et al. 2002). If one side is a mirror image of the other, and a researcher is not interested in their presumably small differences (i.e. asymmetries), it seems reasonable to only focus on one or the other half.

Hundreds of PGMM analyses have been performed using one-side-only data (Cardini in press). The expedient of landmarking only one side is especially common in 2D studies, which use pictures, but also occurs in 3D analyses. Mainly, this is a shortcut that speeds up data collection by reducing the number of points being digitized, and this number has in fact become larger in recent years, because of the increasing adoption of semilandmarks methods on curves and surfaces (Gunz and Mitteroecker 2013). Sometimes other reasons are also mentioned to justify one-side-only digitizations. It is said (e.g. Cardini and O'Higgins 2005; Cardini et al. 2005; Yazdi et al. 2014) that they avoid redundancy, because, by measuring only one side, shape coordinates of paired landmarks (i.e. those on both sides) are not uselessly duplicated. Also, especially (Gunz and Mitteroecker 2013; and references therein) but not exclusively (Yazdi et al. 2012) in palaeontology, only one or the other side might have been preserved and therefore working exclusively on that side can be the only option. Finally, there might be specific hypotheses which require to be tested on a single side, which therefore represents the complete landmark configuration in that context. However, I could find no example of this kind and it does seem that most biologists using one-side-only data are in fact interested in whole bilaterally symmetric structures.

Corresponding author: Andrea Cardini (alcardini@gmail.com, cardini@unimo.it)

Leaving aside, for now, the motives behind the common practice of one-side landmarking (OSL), the main question is whether size and shape variation estimated from this type of data is a faithful approximation of the 'real' structure, complete with landmarks on both sides (BSL). In this article, for the sake of brevity, I will refer to the goodness of this approximation as 'accuracy'. Clearly, some information is lost in OSL data, which therefore might be more or less accurate. Also, as a Procrustes superimposition produces different shape spaces for different landmark configurations, OSL and BSL may partly differ because of the relative registration of the specimens in a sample. A difference, which can be obvious even by eye, is that midplane landmarks may have a larger component of variation perpendicular to the midplane itself in OSL compared to BSL data after the Procrustes superimposition. The displacement of anatomical features beyond the apparent 'midplane boundary' looks counter-intuitive and particularly undesirable in the visualization, as exemplified in Fig. 1 using deformation grids. Nevertheless, analytical results crucially depend on the shape distances, which capture differences integrated above a whole landmark configuration. Thus, as long as these relative distances are well approximated in OSL, OSL data will be accurate compared to BSL.

Asymmetry may be another factor influencing differences (or the lack of them) between OSL and BSL. This is because OSL captures only small asymmetries on the midplane but misses all those in the paired landmarks. Asymmetries are in fact often discarded in BSL analyses, when the researcher is not interested in estimating differences between left and right sides (Klingenberg et al. 2002). However, the asymmetric component of shape is part of the variance in the original superimposition and might contribute to the relative alignment of the specimens in a sample. We know that the proportion of asymmetry varies depending on the sample, and taxonomic level, as well as the study structure. For instance, Schaefer et al. (2006) found relatively larger amounts of shape asymmetry in dental arches in a human sample from the Island of Hvar compared to a sample from Zagreb and reported in the latter a proportionally larger (more than twice bigger) directional asymmetry in the upper compared to lower arch, while they found the opposite in the island population.
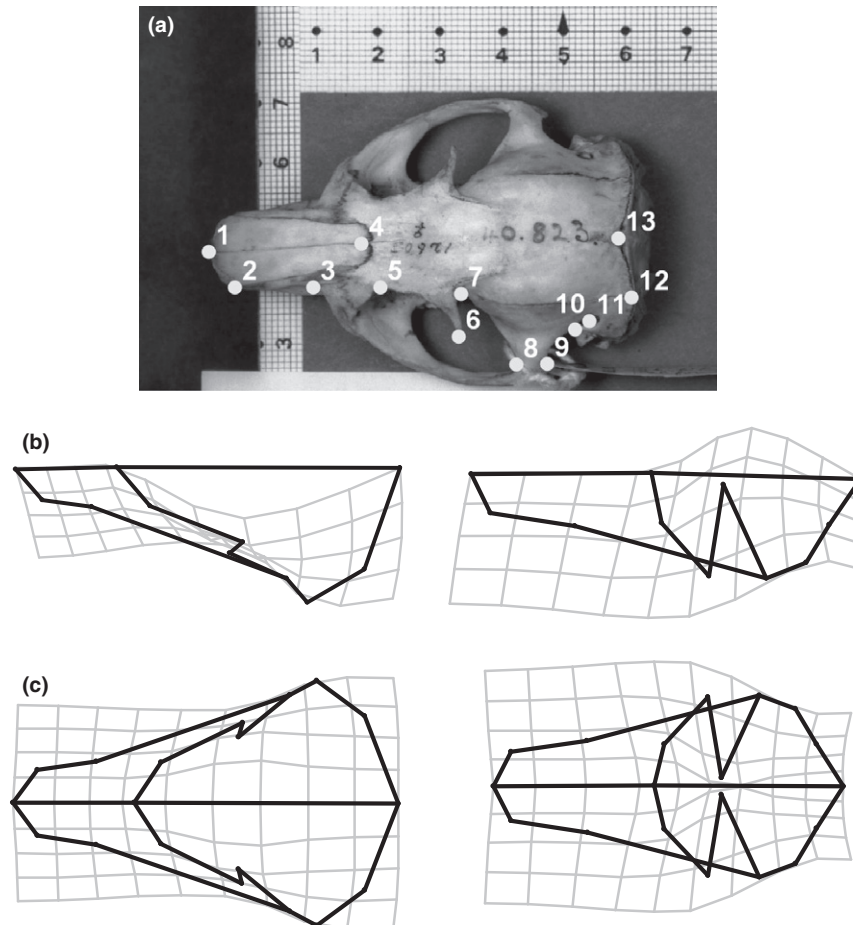
Fig. 1. Ontogenetic allometry in crania of *Marmota flaviventris* illustrated in dorsal view using wireframes and thin-plate spline deformation grids (Klingenberg 2013, and references therein) for opposite extremes of size variation: (a) landmark configuration digitized on the left side only; (b) visualization using only the left side; (c) visualization after mirroring the left side to reconstruct the right side

Klingenberg et al. (2002) estimated that asymmetry accounted for almost 18% of the total shape sum of squares in the pharyngeal jaw of a species of cichlids, while in crania of *Macaca mulatta*, Willmore et al. (2005) found less than half (ca. 8%) of that amount.

Regardless of their source, the central question is whether differences between OSL and BSL data appreciably affect results of PGMM studies. On this, there was little information in the literature until very recently, when a first study specifically focusing on this issue has been published (Cardini in press). That paper showed that shapes estimated from OSL may be affected by differences to BSL that can account for up to 10%, and potentially even more, of the total variation in an intraspecific sample. Also OSL allometric variance may differ, with estimates which might be 10–20% smaller or larger of that obtained using BSL. In the same study, it was suggested that by simply reconstructing a missing side by mirror reflection of the available bilateral landmarks accuracy of size, shape and allometry can be improved, as well as the visualization (Cardini in press).

The aim of this article is to expand that first preliminary analysis, which included just two example data sets, by exploring more extensively OSL data in relation to BSL in PGMM. This will be done with a specific focus on cranial variation in mammals using 10 different data sets from five orders (rodents, carnivores, primates, cetarctiodactyls and diprotodont marsupials), a variety of taxonomic levels (from intraspecific to infraordinal) and both 2D and 3D data from a total of five different landmark

configurations. The correlational approach of Cardini (in press) will be used to (1) assess how accurately size, shape and allometry are estimated in OSL; (2) investigate the effect of reconstructing the missing side by mirror reflection; and (3) provide empirical evidence to discuss whether any generalization can be tried to help researchers of mammalian cranial variation deciding whether and how to perform OSL analyses.

Materials and Methods

Data

Landmark configurations are shown in Fig. 2 and sample sizes in Table 1. All specimens are adults. Data came either from published studies (marmots and Marmotini: Cardini et al. 2005; wallaroo and Macropodinae: Milne and O'Higgins 2002; Cardini et al. 2015; Dahlak gazelles: Chiozzi et al. 2014) or public databases (pictures of crania of foxes and other Carnivora: Takahashi et al. 2006 – http://1kai.dokkyomed.ac.jp/mammal/en/mammal.html; 3D landmarks on human crania: Ross 2014 – http://www.3d-id.org/forensic-3d-coordinates).

For each data set, size, shape and the proportion of shape correlated to size (allometry) were estimated, as well as the amount of asymmetric variation (Klingenberg et al. 2002) in the total configuration. It is important to bear in mind that in this methodological study, although for the sake of brevity this is mentioned only here, the asymmetric component of shape variation includes both biological asymmetries and asymmetries due to measurement error. Overall, in each sample, five data sets were compared to the total BSL configuration: the first one is simply the symmetric component of size and shape from that same complete
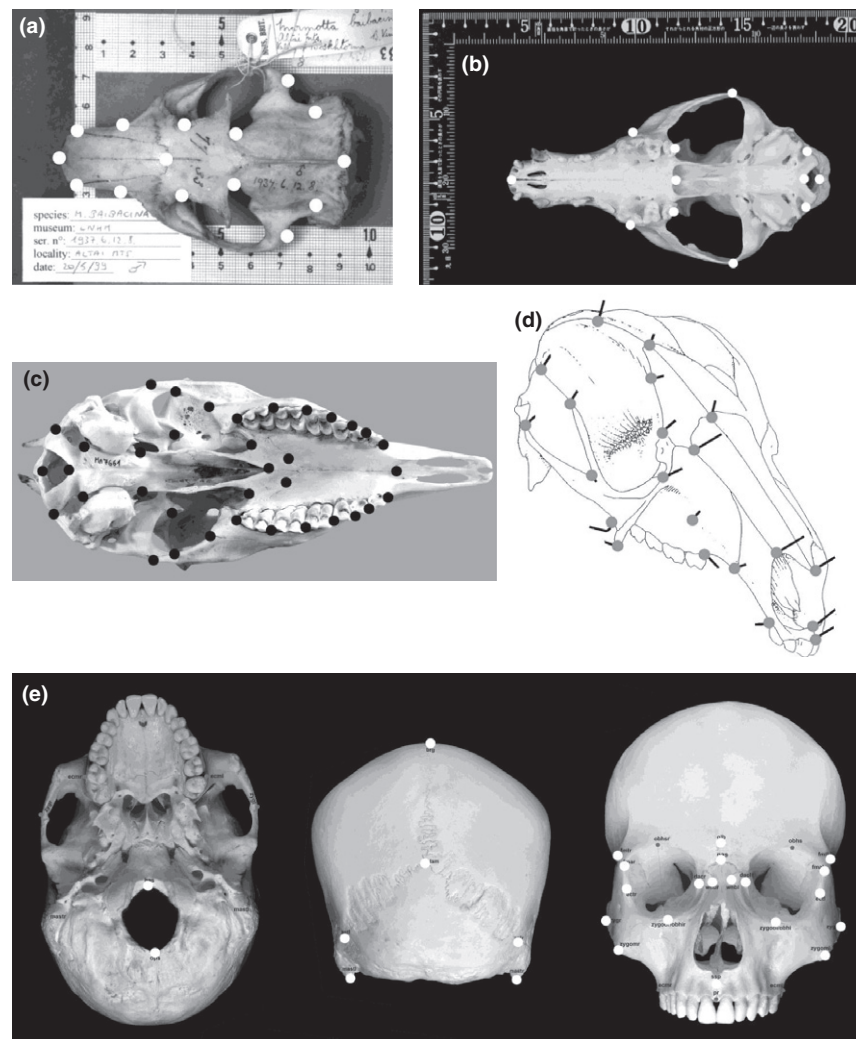
Fig. 2. Landmark configurations: (a) marmots and other Marmotini; (b) foxes and other Carnivora; (c) *Nanger soemmeringii*; (d) kangaroos and walla-roos (modified from Milne and O'Higgins 2002); (e) *Homo sapiens* (modified from the PDF help manual of 3d-id: http://www.3d-id.org/)

configuration; the other four are OSL data sets using the left side only, the left side with the other side reconstructed by mirroring, the right side only and the right side with the other side reconstructed by mirroring.

As anticipated at the beginning of the Introduction, I will use 'accuracy', in this specific context, to mean how well data based on partial information (e.g. left side only – LS) correspond to data using all available information (both sides and all shape components – tBSL). Before detailing methods, I summarize below the main abbreviations in this study:

General ones, which had already been introduced:

PGMM: Procrustes-based geometric morphometrics.
OSL: one-side landmarking.
BSL: both sides landmarking.

Additional abbreviations specific to the type of data set:

tBSL: total configuration using both sides and all size and shape components (i.e. symmetry and asymmetry are both included).

tsBSL: symmetric component of tBSL.

LS and LSM: left-side configuration without paired landmarks on the right side, and the same configuration with the missing right side reconstructed by mirror reflection, respectively.

RS and RSM: right-side configuration without paired landmarks on the right side, and the same configuration with the

missing left side reconstructed by mirror reflection, respectively.

Accuracy assessment (i.e. comparison of results from the analysis of both sides and one side only)

Data sets were compared using correlations. As explained in Cardini (in press), this allows to compare different shape spaces generated by separate Procrustes superimpositions, '. . . a requirement in order to be in the same analytical setting as in the real case of an operator having measured either both sides or just one of the two. . . A common superimposition would have the advantage of bringing all data sets in the same shape space, making them apparently easier to compare. However, to also have the same dimensionality in all sets of data, one would need to estimate the missing side and that, for the specific purpose of this study, should be done in a way such that LS and RS distances are the same as when they are superimposed on their own one at a time. . . Unfortunately, this. . . operation. . . introduces a bias [in LS/RS]'. Thus, accuracy in centroid size (henceforth simply referred to as size) was estimated using Pearson correlations between tBSL and all other data sets. Accuracy in shape was similarly estimated using matrix correlations (i.e. Pearson correlations between vectorized distance matrices) of tBSL specimen pairwise Procrustes distances compared to those from the other five data sets. Finally, to assess the effects of OSL when size and shape are analysed together, the percentage of allometric variance was computed using a

Table 1. Data sets description and parameters estimated. Description: taxon (all adults, both sexes unless specified); sample size (N); type of landmarks (2D or 3D); tBSL number (#) of landmarks (L); and number of midplane (mid.) landmarks. Parameters: percentage of variance accounted for by asymmetry in total as well as after partitioning it between directional (DA) and fluctuating (FA); size (SIZE r) and shape (SHAPE r) correlations between tBSL and either its symmetric component (tsBSL) or one-side (LS or RS) or one-side mirrored (LSM or RSM) configurations; percentage of total variance 'predicted' (i.e. accounted for) using the polynomial model approximation (r-PCA var.); allometric variance

Taxon	N	Data	2D/3D	# L	# Mid. L	Shape asymmetry, %			SIZE r[1]	SHAPE r[1]	r-PCA vari-ance, %	Allom. vari-ance, %
						Total	DA	FA				
Marmota caligata	40	tBSL	2D	15	3	11.8	1.2	10.6	–	–	–	10.6
		tsBSL		15	3	–	–	–	1.000	0.991	87	11.6
		LS		9	3	–	–	–	0.996	0.878	56	7.4
		LSR		15	3	–	–	–	0.998	0.901	61	11.6
		RS		9	3	–	–	–	0.997	0.868	54	5.9
		RSM		15	3	–	–	–	0.998	0.893	59	8.6
Marmota flaviventris	40	tBSL	2D	15	3	8.8	0.4	8.4	–	–	–	17.0
		tsBSL		15	3	–	–	–	1.000	0.990	87	17.8
		LS		9	3	–	–	–	0.998	0.750	35	12.1
		LSR		15	3	–	–	–	0.999	0.812	42	16.1
		RS		9	3	–	–	–	0.998	0.916	61	11.2
		RSM		15	3	–	–	–	0.999	0.906	58	16.1
Marmotini	40	tBSL	2D	15	3	4.1	0.5	3.6	–	–	–	64.1
		tsBSL		15	3	–	–	–	1.000	0.999	95	66.2
		LS		9	3	–	–	–	1.000	0.963	71	66.0
		LSR		15	3	–	–	–	1.000	0.983	81	64.4
		RS		9	3	–	–	–	1.000	0.967	73	65.1
		RSM		15	3	–	–	–	1.000	0.979	79	62.8
Vulpes vulpes	78	tBSL	2D	12	4	19.0	2.7	16.3	–	–	–	4.9
		tsBSL		12	4	–	–	–	1.000	0.971	76	5.5
		LS		4	4	–	–	–	0.997	0.881	54	3.6
		LSR		12	4	–	–	–	0.998	0.894	56	4.6
		RS		4	4	–	–	–	0.997	0.868	53	4.8
		RSM		12	4	–	–	–	0.998	0.880	54	4.8
Carnivora	31	tBSL	2D	12	4	0.8	0.1	0.7	–	–	–	6.4
		tsBSL		12	4	–	–	–	1.000	1.000	100	6.4
		LS		4	4	–	–	–	0.998	0.945	77	5.5
		LSR		12	4	–	–	–	1.000	0.996	94	6.6
		RS		4	4	–	–	–	0.998	0.948	77	5.3
		RSM		12	4	–	–	–	1.000	0.996	94	6.2
Macropus robustus	21	tBSL	3D	36	6	17.1	1.9	15.2	–	–	–	13.4
		tsBSL		36	6	–	–	–	1.000	0.974	68	14.7
		LS		15	6	–	–	–	0.998	0.846	42	10.6
		LSR		36	6	–	–	–	0.999	0.887	47	13.1
		RS		15	6	–	–	–	0.999	0.942	57	10.9
		RSM		36	6	–	–	–	0.999	0.948	59	13.4
Macropodinae	12	tBSL	3D	36	6	2.6	0.4	2.2	–	–	–	32.8
		tsBSL		36	6	–	–	–	1.000	1.000	100	33.3
		LS		15	6	–	–	–	1.000	0.990	91	33.8
		LSR		36	6	–	–	–	1.000	0.996	95	32.3
		RS		15	6	–	–	–	1.000	0.988	90	34.6
		RSM		36	6	–	–	–	1.000	0.995	94	32.7
Homo sapiens (EU-AM)[2]	##	tBSL	3D	27	7	15.4	0.7	14.7	–	–	–	1.3
		tsBSL		27	7	–	–	–	1.000	0.975	75	1.4
		LS		10	7	–	–	–	0.965	0.894	53	1.6
		LSR		27	7	–	–	–	0.984	0.901	54	1.3
		RS		10	7	–	–	–	0.958	0.908	55	1.6
		RSM		27	7	–	–	–	0.985	0.919	58	1.5
H. sapiens (AFR-AM)	##	tBSL	3D	27	7	18.7	2.9	15.8	–	–	–	2.2
		tsBSL		27	7	–	–	–	1.000	0.960	70	2.3
		LS		10	7	–	–	–	0.980	0.869	47	1.9
		LSR		27	7	–	–	–	0.990	0.872	49	2.0
		RS		10	7	–	–	–	0.983	0.884	51	2.4
		RSM		27	7	–	–	–	0.990	0.896	54	2.6
Nanger soemmerringii[3]	21	tBSL	2D	36	4	17.7	1.5	16.2	–	–	–	18.5
		tsBSL		36	4	–	–	–	1.000	0.975	72	21.5
		LS		16	4	–	–	–	0.996	0.913	55	21.5
		LSR		36	4	–	–	–	0.996	0.915	55	20.4
		RS		16	4	–	–	–	0.997	0.810	40	20.6
		RSM		36	4	–	–	–	0.996	0.896	51	17.0

[1]SIZE r is the Pearson correlation for centroid size (simply referred to as size); SHAPE r is the matrix correlation using Procrustes shape distances between all possible pairs of specimens; the only two cases in which LS/RS performed better (closer correspondence to tBSL) than LSM/RSM are emphasized using a light grey background.

[2]European American, males only

[3]Dahlak population

multivariate regression of shape coordinates onto size in each data set and sample. The resulting percentages were compared with the one estimated in tBSL.

Accuracy in shape was further investigated using two types of exploratory approaches as in Cardini (in press). First, a way to crudely estimate of how much variance is retained in shape spaces based on a limited amount of the original tBSL total information was devised. Accuracy between total and partial information was again quantified using correlations of shape distances. Thus, shape variation was summarized by different numbers (first one, first two, first three etc.) of principal components (PCs) of tBSL shape coordinates. For each set of tBSL PCs, the corresponding pairwise Euclidean shape distances were computed and their correlation (r) to tBSL Procrustes shape distances calculated. For instance, in the hoary marmot (*Marmota caligata*) data set, matrix correlations of specimen pairwise Euclidean shape distances based on tBSL PCs and the corresponding Procrustes shape distances in the total tBSL shape space were computed and plotted against the variance explained by those same PCs (Fig. 3). In this sample, using only tBSL PC1, the correlation between shape distances was 0.670 and the variance explained was 24.8%; using the first two PCs, r increased to 0.789 and variance explained to 38.6% etc. Assuming that the same relationship between explained variance and r holds as a proxy for when tBSL is compared to tsBSL and OSL shape data, one can make a very approximate estimate of how much tBSL shape variance those data might account for. A fifth-order polynomial regression was used to model the relationship (Fig. 3). Thus, for instance, in the case of hoary marmots LS shape data, their correlation to tBSL was 0.878, which, using the polynomial approximation for the relationship between tBSL PCs and total shape, corresponded to 56.1% of tBSL variance. This type of estimate is a very crude one and likely underestimates variances. This means that it might make OSL and tsBSL data look less accurate than they really are (Cardini in press). However, it provides a very rough scale to interpret the magnitude of the correlations between shape spaces in terms of variance explained. For the sake of brevity, henceforth, I will loosely refer to these estimates as variances 'predicted' by the r-PCA polynomial model.

The second approach has the same aim as the previous one, and it is more accurate but can only be applied to compare data with the same identical landmark configuration. This means that it cannot be used to compare tBSL with LS/RS. It was used, however, to further assess the accuracy of left and right OSL data after the reconstruction of the missing side by mirror reflection. Thus, tBSL, LSM and RSM were superimposed together and analysed in a common shape space. As it can be easily checked using matrix correlations for shape distances, this operation leaves similarity relationships within each of the three sets of data virtually identical to those obtained when they were superimposed separately. For this reason, findings from the common shape space also apply to data analysed one at a time. In this common shape space, variation was partitioned using sum of squares between two components: the
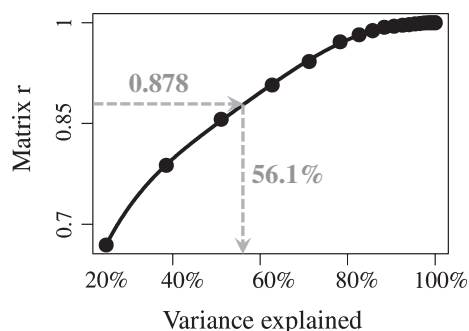

Fig. 3. Scatterplot of matrix correlation *versus* cumulative variance explained in the PCA of the tBSL data set of *Marmota caligata*. Correlations are computed between pairwise shape distances based on all shape variables (i.e. Procrustes shape distances) and Euclidean shape distances based on the first one, two, three etc. PCs of the shape coordinates. The solid black line represents the fifth-order polynomial fitted to the data (r-PCA polynomial model), and the grey dotted lines exemplifies how to infer the ca. 56% variance corresponding to a correlation of 0.878, such as the one between LS shape data and tBSL in this sample.

individual variation (averaged across data sets) and the differences among the same individuals in the three sets of data. If the latter ones are small, that means that LSM and RSM are accurate. Also, a cluster analysis (Unweighted Pair Group Method with Arithmetic Mean (UPGMA) – on the matrix of Euclidean shape distances) was performed including all three types (tBSL, LSM and RSM) of data. The expectation is that 'replicas' (i.e. tBSL, LSM and RSM of an individual) should cluster together if LSM and RSM shape estimates are very similar to tBSL.

Software

Analyses were performed in: TPSUTIL (Rohlf 2015): data manipulations; MORPHOJ (Klingenberg 2011): superimposition, separation of symmetric and asymmetric components of shape, and computations of sum of squares of different data sets; NTSYSPC 2.3 (Rohlf 2013): superimposition and multivariate regressions for estimating allometry; PAST (Hammer et al. 2001): cluster analysis; R (R Core Team 2014): correlational analysis of size and shape, PCA and polynomial regression. A detailed step by step tutorial on how to estimate a missing side by mirror reflection is provided in the Supporting Information, which also includes references to the software and an example data set. An alternative way to perform the mirror reflection is explained in Cardini et al. (2010).

Results

Size

Results are shown in Fig. 4 and Tables 1–2. Size (Fig. 4a) was very accurate in all data sets, regardless of side or mirroring. In particular, size estimated in tsBSL was virtually identical to tBSL size, and in OSL data sets, the correlation with tBSL was always larger than 0.95 (Table 1). LSM and RSM had higher correlations than LS/RS in all but one case, in which RS had $r = 0.997$ compared to $r = 0.996$ in all other OSL data sets. Overall, the inaccuracy in OSL size data was minimal and especially LSM and RSM provided estimates very similar to tBSL and tsBSL.

Shape

Shape (Table 1) asymmetry varied across data sets and accounted for between ca. 1% (Carnivora) and almost 20% (foxes, humans and wallaroos) of total variance. Directional asymmetry was generally small (<3%) with the African American sample, as well as the foxes, showing the highest values. Fluctuating asymmetry was on average 10 times larger than directional asymmetry (ranging from 5 – Macropodinae and Afro-Americans – to 21 times – *Marmota flaviventris* and European Americans), thus clearly accounting for most asymmetric variance.

Symmetric shape (stBSL) was always very highly correlated ($r \geq 0.960$) to total shape (tBSL) with the highest correlations corresponding to the data sets with the smallest total asymmetry. For stBSL, variances 'predicted' by r-PCA polynomial models ranged between ca. 70% and 100% (Fig. 4c, Table 1). In all data sets (tsBSL and OSL), and in all samples, polynomial curves had minimal deviations to the datapoints (e.g. Fig. 3).

OSL shape data showed a much larger variation with r ranging from almost 1 to 0.75. The corresponding variances 'predicted' using the r-PCA model varied between 35% and 95%. In all samples except one, LSM and RSM performed better than LS and RS: average, minimum and maximum r were 0.923, 0.812, 0.996 for LSM/RSM and 0.901, 0.750 and 0.990 for LS/RS; average, minimum and maximum variances 'predicted' by r-PCA polynomials were, respectively, 65%, 42%, 95% for LSM/RSM and 60%, 35% and 91% for LS/RS.

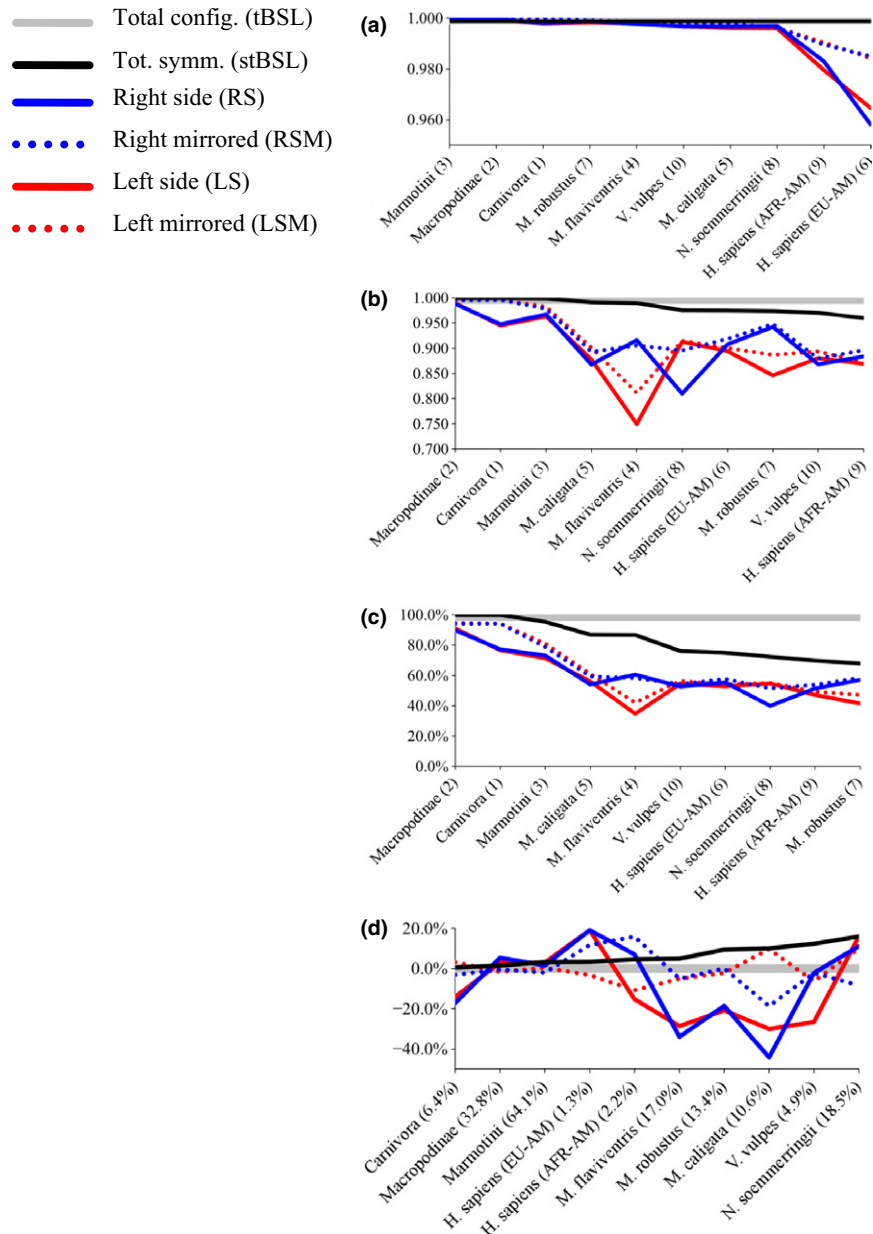When LSM and RSM were compared to tBSL in a common shape space, the variance explained by individuals ranged

Fig. 4. Accuracy of different data sets compared to the total configuration (tBSL). The grey line represents the highest accuracy, that is size or shape differences identical to those of the total configuration; samples are ordered from the highest to the lowest accuracy in the symmetric data set (tsBSL). (a) Pearson correlations of centroid size; (b) matrix correlations of pairwise Procrustes shape distances; (c) r-PCA polynomial approximations of 'predicted' shape variance; (d) percentage expressing how much allometric variance (R2, reported by each sample name) was over- or underestimated compared to its estimate in the total data set (i.e. the difference between R2 based, for instance, on LS and R2 based on tBSL divided by tBSL R2 itself). In (a-b-c), numbers in parentheses by the sample names correspond to rank ordering from samples with less asymmetry to those with more (e.g. Carnivora ranks, (1) with less than 1% of asymmetric shape variance, followed by the Macropodinae (2) with ca. 3% and by the Marmotini (3) with about 4% etc.); these ranks may help to spot any trend related to the amount of asymmetry in the data. LS, left side only

between ca. 83% and 99% with an average of 89%. This indicates that after mirroring the missing side, inaccuracies accounted for ca. 10% of total variance and in any case for never more than 20%. In the cluster analysis, on average almost 90% of the times LSM and RSM clustered together with tBSL according to individuals. The only exception was the sample of foxes, which had a low percentage (53%) of correct clusters.

Allometry

Allometry varied widely across samples, with humans having only ca. 1–2% of variance explained by size and the Marmotini ca. 60–80% (Table 1). In LS and RS, allometry was underestimated about as many times as it was overestimated; however, the underestimates were larger in magnitude so that the overall average was about 8% smaller relative to allometric variance in tBSL (Fig. 4d). tsBSL, in contrast, consistently overestimated allometry (+7% on average relative to tBSL). In LSM and RSM, underestimates of allometry were more common than overestimates (13 times of 20), but the error was almost always smaller than in LS/RS and overall unbiased on average (ca. <1% than in tBSL). LS and RS almost always provided estimates of allometry less accurate (±16–18% average relative error) than tsBSL, LSM and RSM, which in turn were similarly accurate (±5–7% average relative error). Contrary

Table 2. Comparison of tBSL, LSM and RSM in a common shape space. (a) Sum of squares (SS) and percentages of shape variance explained (R2) by individuals (averaged across the three types of data) and residual variation (i.e. differences among tBSL, LSM and RSM). (b) Percentage of individuals in which tBSL, LSM and RSM data 'correctly' cluster together in a phenogram (i.e. those with negligible differences regardless of the type of data).

| Taxon | (a) Partitioning of sum of squares (SS)[1] | | | (b) Cluster analysis % 'correct' |
	Effect	R2, %	SS, %	
Marmota caligata	Individual	88.0	0.100940	92.5
	Residual		0.013820	
Marmota flaviventris	Individual	90.7	0.114094	95.0
	Residual		0.011673	
Marmotini	Individual	95.9	0.325971	96.0
	Residual		0.013852	
V. vulpes	Individual	83.1	0.136107	52.6
	Residual		0.027760	
Carnivora	Individual	99.0	2.881367	94.1
	Residual		0.028389	
M. robustus	Individual	83.6	0.222557	95.2
	Residual		0.043735	
Macropodinae	Individual	97.0	0.225707	100.0
	Residual		0.006999	
Homo sapiens (EU-AM)	Individual	86.0	1.111066	89.1
	Residual		0.181244	
H. sapiens (AFR-AM)	Individual	83.1	1.226596	87.1
	Residual		0.249777	
Nanger soemmerringii	Individual	83.7	0.034477	90.5
	Residual		0.006699	

[1]These estimates are based on the standard ANOVA model for measurement error (Viscosi and Cardini 2011), but they are only employed to compute SS, as significance cannot be tested because the three configurations do not originate from independent sets of data.

to size and shape, in which tsBSL was always the most accurate data set, only 3/10 of the times the symmetric data set provided estimates of allometry closest to tBSL, with LSM and RSM being the most accurate all other times.

Discussion

Symmetric component only

Before focusing on OSL data, I am going to briefly discuss whether discarding asymmetric variation from complete data with landmarks on both sides produces appreciable differences. This means that results are based on BSL but only use symmetric variance (tsBSL). Although this is a topic tangential to the main question on OSL accuracy, it is relevant because it is another common operation in PGMM, when a researcher is not interested in asymmetric variation, which is the same type of situation leading to OSL analyses. Also, the discussion of tsBSL provides an example that helps to stress the limits of the r-PCA model.

The analysis of the 10 mammalian cranial samples showed that, with size, estimates are virtually identical to those using both symmetric and asymmetric variation. Interestingly, also with shape, and even when asymmetry was large (up to almost 1/5 of total variance in some samples), the symmetric component of shape provided estimates of relative shape differences that very closely mirrored those including both symmetric and asymmetric components. Thus, at least in mammalian crania, whenever one is not focusing on asymmetric variation, results from the analysis of the symmetric component of shapes will be likely almost identical to those using total shape. As Klingenberg et al. (2002) suggested, asymmetry can thus optionally be discarded, which

generally also helps producing a nicer visualization (Klingenberg 2013), without losing any important information. However, an interesting exception to this recommendation might happen in the case of allometry, because it was found that tsBSL data consistently positively biased the estimates of allometric variance. The reason for this is unclear but one might speculate whether averaging sides increases covariance, including its size-related component. Regardless of the explanation, overall, the average overestimate was small in absolute terms and still fairly modest when scaled to the value of allometric variance obtained from tBSL (close to a 20% positive bias in only one sample and much less in most of the others). Thus, if this is generalizable, even with allometry, it does not seem that discarding asymmetry in tBSL makes a really appreciable difference in most cases.

This concise discussion of tsBSL results provides also the opportunity for making a few considerations on the r-PCA model. This was used to obtain a crude approximation of the variance accounted for by the discrepancy between tsBSL and tBSL. The r-PCA model was clearly unnecessary for tsBSL, because one can obtain the exact variance estimate using the conventional PGMM method developed to separate symmetric and asymmetric shape (Klingenberg et al. 2002, and references therein). The only reason to use also the r-PCA model was that it is interesting to compare the correct and approximate estimates. This shows that variances accounted for by tsBSL using the r-PCA model underestimate the real amount of symmetric variance, thus overestimating the differences between tsBSL and tBSL. Yet, on average, the difference was only 5%. In contrast, if the same comparison is made for LSM/RSM using the estimates obtained by partitioning sum of squares in a common shape space, the r-PCA approximation again overestimates differences, but it does so in a much more pronounced way with an average bias of about 25%. This is consistent with previous findings (Cardini in press) and confirms that, as anticipated in the methods section, the r-PCA model is a very crude one, which can only provide a very rough 'scale' to appreciate OSL inaccuracies.

Difficulties of comparing all data within the same morphospace

Another preliminary issue, which was briefly mentioned in the methods section but needs some more discussion, is why the comparison of OSL and other data sets was not performed in a common data space. This is possibly easier for centroid size but does not add much to the results of the correlational analyses, which already convincingly show that size is pretty accurate in all data sets, including LS/RS. In fact, with centroid size, one could mean centre (i.e. subtract the mean) the data within each data set. This preliminary operation is unnecessary for LSM and RSM (which have the same landmark configuration as tBSL) but it is important for LS and RS. Indeed, because LS and RS have less landmarks, their centroid sizes will inevitably be smaller than in tBSL. As this is a strong bias, it may not matter when one is interested in the pattern of relative size differences in a sample (i.e. the fact that specimens show the same relative amount of interindividual differences as in tBSL). By mean centring the data, one removes that bias and makes size comparable across different landmark configurations, and therefore potentially suitable for an analysis of variance to test whether differences between replicas of each individual (i.e. size estimated using tBSL and, for instance, LS) are negligible relative to differences among specimens in the sample.

There are two complications, however, even with size. First of all, despite mean-centring and thus removing the bias in mean differences, there could still be a second type of bias: this is because not only the mean but also the variance of the complete

configurations will likely be larger because of the larger number of landmarks. The practical effect of this second type of bias might be generally smaller but nevertheless affect the comparison of size relationships captured by OSL and tBSL. Likely, it would make results more conservative (i.e. such that they would suggest differences between replicated estimates of an individual size slightly larger than real, thus inflating the inaccuracy of OSL data). To make this point clearer, one can imagine a purely hypothetical extreme case in which size differences among specimens in tBSL are exactly q times larger than among the same specimens in LS (or RS). Thus, in relative terms, differences between sizes of specimens in tBSL and those of specimens in LS would be identical. However, an analysis of variance would still show a difference between tBSL and LS size data simply because of the different variance in the two data sets. Fortunately, for size, this issue may also be easily addressed, as suggested by a reviewer, using z-scores instead of simple mean-centred data.

The second complication, in contrast, is more serious and is that data are not repeated measurements on the same individuals. They are exactly the same measurements with OSL data simply using only part of the total BSL information: this creates a type of non-independence that does not allow to use the standard analysis of variance (Viscosi and Cardini 2011) commonly used to test measurement error. In fact, as anticipated, because size data showed such high correlations, the whole issue of statistically testing the magnitude of OSL differences to BSL has not real practical relevance and the evidence is already strong enough to conclude that size is well approximated in all instances.

For shape, however, and especially so in intraspecific samples, there seem to be potentially important differences between OSL and BSL. This makes it more interesting to see whether one could do better than a simple correlational approach to compare the data. Even just being able to assess if OSL and BSL shapes of the same individual cluster together in a phenogram would provide quantitative clues to better understand whether OSL shape distances accurately capture BSL similarity relationships. Indeed, this was done for LSM and RSM, because they share the same landmarks as tBSL and because, even after a common superimposition, their within-data set shape distances were virtually identical to those from separate superimpositions of the three sets of data. After a common superimposition, sum of squares can be partitioned to estimate inaccuracy. This was done using an analysis of variance with individual as a random factor (Viscosi and Cardini 2011), but without performing any statistical test, as LSM/RSM are in fact non-independent subsets of tBSL (same considerations as above, for size). There might be *ad hoc* tests which overcome this difficulty, but the simple assessment of whether OSL shapes are nearest neighbours of the corresponding tBSL (which should be apparent in the shape distance-based phenogram) might already provide convincing evidence on the accuracy (most data sets), or inaccuracy (foxes), of OSL data.

One might wonder if there are ways of bringing also LS/RS together with tBSL in a common shape space to more accurately estimate differences. Cardini (in press) suggested one such approach, which seems intuitive but unfortunately unlikely to work. As an example, I will focus on LS, but this would be the same for RS. The simple idea is to treat bilateral landmarks, which are absent in LS, as missing and a program such as Morpheus et al. (Slice, 1999) to superimpose data together with tBSL. In the same software, missing landmarks can then be replaced with the sample mean (i.e. the mean coordinates of right-side landmarks in tBSL). This expedient produces data of (apparently) identical dimensionality and brings LS and tBSL together in the same shape space. Furthermore, within-data set,

shape distances are virtually the same as with separate superimpositions, which makes results (again, apparently) comparable to the real case scenario of an operator analysing only one side. However, because in the whole set of LS specimens one side was in fact missing, and thus unused in the superimposition, LS shape will consistently differ from tBSL. Thus, depending on the strength of this bias, LS may cluster more or less distantly from tBSL. Preliminary observations (A. Cardini, unpublished) suggest that this type of bias may not be easily removed by a simple mean-centring and will thus prevent a direct comparison of the data sets.

Accuracy of size, shape and allometry in one-side-only data

The assessment of inaccuracies from PGMM estimates of cranial size and shape in mammals using OSL data shows results which largely support the previous analysis based on just two samples (Cardini in press). Relative size differences are very accurately estimated in all sets of data regardless of whether only one side was used and the missing landmarks were or were not reconstructed by mirroring. Shape relationships can be also accurate especially when sample variance is large, as it happens in a macro-evolutionary analysis, as exemplified by the carnivores, kangaroos and marmotine samples. However, in micro-evolutionary analyses of within-species variation, OSL and BSL shape data are less congruent and inaccuracies can account for an appreciable proportion of variance. The r-PCA model overestimates this loss of information but the more accurately partitioning sum of squares of LSM and RSM indicates that, at least in these data sets, inaccuracies in within-species samples account for about 15% of total tBSL shape variance. This is clearly much less than using the r-PCA approximation but it is still a fairly large amount. Besides, as it only takes into account the data whose missing sides have been reconstructed and which were almost always the most accurate of the OSL data, the variance related to LS/RS inaccuracies will be potentially even larger than 15%.

Overall, as with size, shape relationships were consistently and virtually always more accurately estimated using LSM and RSM than using simply LS and RS. This is clearly indicated by LSM and RSM higher matrix correlations to the tBSL shape data (with one exception out of 20 comparisons), as well as by the corresponding shape estimates which accounted for up to 18% more variance, with an average of +5% according to the r-PCA approximation, compared to LS and RS.

Indeed, that, of 40 pairwise comparisons of LS/RS *versus* LSM/RSM, 38 showed that the size and shape are more accurate in the latter is probably the most interesting finding of this study. Thus, it does seem that, as Cardini (in press) suggested, the simple operation of estimating the missing side by mirror reflection might make results from the analysis of OSL data more similar to BSL. This is even more apparent in the estimates of allometry. LSM and RSM provided estimates which not only are closer to tBSL than those from LS/RS, and about as accurate as those of tsBSL, but also are unbiased. Besides, as anticipated in the Introduction, the visualization is improved in OSL analyses with mirror-reflected sides and generally becomes almost identical to that from analyses of the symmetric component of shape, as exemplified in Fig. 5 for the Dahlak gazelles.

Regardless of the type of data, accuracy of OSL size and shape, tends to be higher in supraspecific samples. This is clearly a consequence of the proportionally larger variation among specimens above the species level that reduces the proportion of asymmetry and the relative impact of BSL and OSL differences. That the degree of asymmetry crucially affects the accuracy of OSL data is almost tautological. However, Cardini (in press)
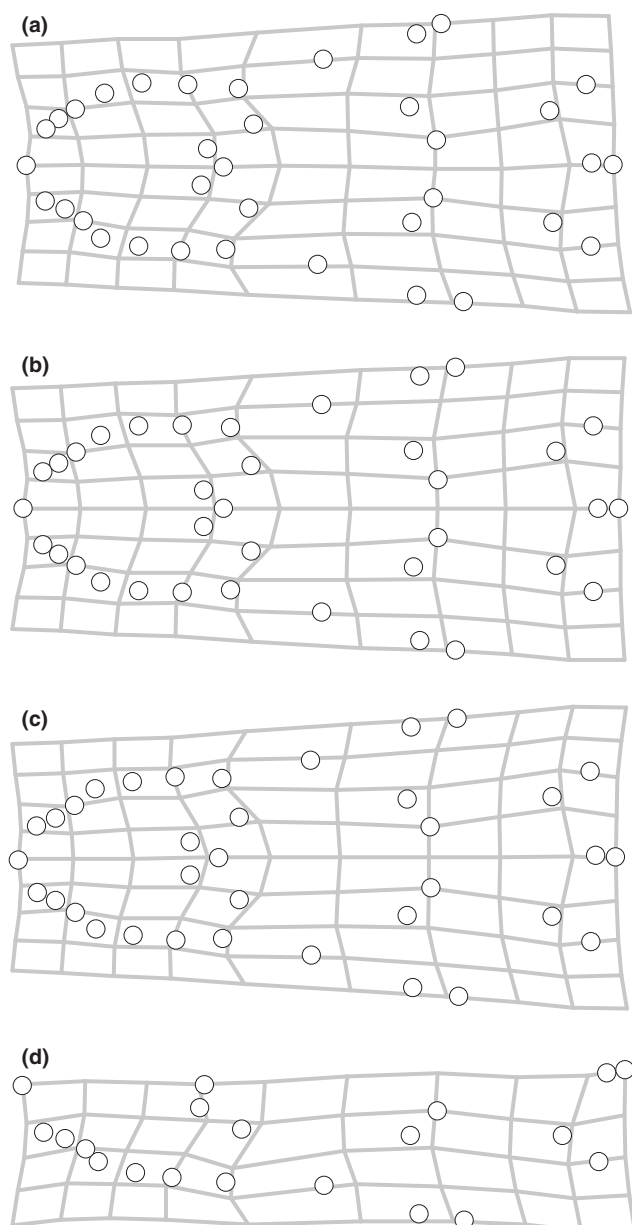
Fig. 5. Thin-plate spline (Klingenberg 2013, and references therein) visualization of the same extreme of PC1 in the sample of Dahlak gazelles: (a) tBSL; (b) tsBSL; (c) LSM; (d) LS. tBSL, total configuration using both sides and all size and shape components; tsBSL, symmetric component of tBSL; LS, left side only

suggested that the effect of fluctuating asymmetry may be particularly important and, if unusually large, that alone could even make LSM/RSM results less accurate than those from LS/RS.

As mentioned, when LSM, RSM and tBSL were compared in the same shape space, their differences accounted for less than 20% of total variance. This relatively small error (which is in fact slightly overestimated by analysing together LSM-RSM and tBSL) is congruent with the results of the cluster analysis. In the phenogram, the percentage of individuals whose OSL replicas both clustered with tBSL was generally very high (>90% of average). Thus, despite missing information in OSL data, relationships among specimens were faithfully captured, in relative terms, in LSM and RSM and therefore very similar to those obtained by measuring both sides (tBSL). Foxes (53% accuracy),

however, were an exception, as they performed remarkably worse than all other samples (93% average accuracy excluding foxes). This was unexpected, as their matrix correlations to tBSL data were not the lowest overall. The reasons for the particularly poor accuracy in the phenogram of foxes are unclear. It could be that this sample, of unknown provenance, had relatively smaller interindividual differences (e.g. if part of the sample was from an inbred captive population) and part of the differences were random noise in a small landmark configuration with at least a few low precision landmarks (e.g. those on the zygomatic arch). In general, however, foxes were consistent with the observation that samples with less variation (microevolutionary analyses) are more strongly affected by inaccuracies in OSL data.

Conclusions

In conclusion, it seems advisable that, unless asymmetry is very large (which could be estimated, preliminary to a main study, using a subsample or based on the literature), whenever one performs OSL analyses, first the missing side is reconstructed and then all analyses, including the visualization, are conducted on LSM (or RSM) data. One may argue that a better suggestion is to simply measure both sides (BSL). Indeed, this would also apply to studies of structures with matching symmetry (such as left and right hands or left and right marmot hemimandibles etc.), as by landmarking both sides and averaging them within individuals, measurements would more accurately capture morphological variation. However, Cardini (in press) argued, using a real example from a large data collection, that landmarking just one side does save time, which often translates in larger samples at lower costs. When large samples are crucial, as in most taxonomic studies, and data collection is expensive, as when using 3D data from trips to many museums, OSL might be justifiable and help to almost double the number of specimens measured in a given amount of time (Cardini in press).

More research is required to allow strong generalizations to be made. Both empirical studies on other groups and structures, as well as theoretical analyses and simulations, might help to better understand the advantages and disadvantages of OSL analyses and the different factors affecting accuracy. Besides those already mentioned and briefly discussed here and in Cardini (in press), other potential factors, which might contribute to how small or large differences between OSL and BSL data are, likely include the number and type of landmarks. Future studies might also confirm that, as suggested by one of the reviewers, 'one of the main causes, [why] Procrustes analysis of OSL, compared to BSL must lead to different results, is that the centre of rotations (centroid) around which the shapes are rotated to minimize Procrustes distances differ significantly for partial shapes; thus, the relative positions of the rotated landmarks will differ inevitably. [That implies that] LSM and RSM are probably a good approximation of this centroid, with the resulting alignments, and the resulting shape space, being quite similar'.

For the time being, using a much larger number of specimens and samples from five mammalian orders, a variety of taxonomic levels and both 2D and 3D data from a total of five different landmark configurations, this study corroborates all the main conclusions of Cardini (in press) on OSL data: size is always accurate; shape is accurate in macroevolutionary samples but much less so in intraspecific microevolutionary analyses; reconstructing the missing side by mirror reflection generally increases accuracy of size, shape and especially allometry, while also improving the visualization; asymmetry is likely the most important factor to take into account in order to decide whether OSL data are appropriate and mirror reflection might help.

Acknowledgements

References

Adams DC, Rohlf FJ, Slice D (2013) A field comes of age: geometric morphometrics in the 21st century. Hystrix It J Mammal 24:7–14.

Cardini A (2013) GEOMETRIC MORPHOMETRICS, in Biological Science Fundamental and Systematics, Eds. UNESCO-EOLSS Joint Committee, in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO. Eolss Publishers, Oxford, UK: http://www.eolss.net.

Cardini A (In press) Lost in the other half: improving accuracy in geometric morphometric analyses of one side of bilaterally symmetric structures. Syst Biol. http://sysbio.oxfordjournals.org/content/early/2016/06/24/sysbio.syw043.abstract?keytype=ref&ijkey=zE9qSbgc7NXGCwM.

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, Hoffmann RS, Thorington RW (2005) Morphological evolution in marmots (Rodentia, Sciuridae): size and shape of the dorsal and lateral surfaces of the cranium. J Zool Syst Evol Res 43:258–268.

Cardini A, Filho JAFD, Polly PD, Elton S (2010) Biogeographic analysis using geometric morphometrics: clines in skull size and shape in a widespread African arboreal monkey. In: Elewa AMT (ed). Morphometrics for Non-Morphometricians. Springer, Berlin, Heidelberg, pp 191–217.

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 134:1–15.

Chiozzi G, Bardelli G, Ricci M, De Marchi G, Cardini A (2014) Just another island dwarf? Phenotypic distinctiveness in the poorly known Soemmerring's Gazelle, Nanger soemmerringii (Cetartiodactyla: Bovidae), of Dahlak Kebir Island. Biol J Linn Soc 111:603–620.

Gunz P, Mitteroecker P (2013) Semilandmarks: a method for quantifying curves and surfaces. Hystrix It J Mammal 24:103–109.

Hammer O, Harper D, Ryan P (2001) PAST: Paleontological statistics software package for education and data analysis. Paleontol Electron 4:1–9.

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. Mol Ecol Resour 11:353–357.

Klingenberg CP (2013) Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. Hystrix It J Mammal 24:15–24.

Klingenberg CP, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution 56:1909–1920.

Milne N, O'Higgins P (2002) Inter-specific variation in Macropus crania: form, function and phylogeny. J Zool 256:523–535.

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rohlf FJ (2013) NTSYSpc: Numerical Taxonomy System, ver. 2.3. Exeter Publishing, Ltd, Setauket, New York.

Rohlf FJ (2015) The Tps series of software. Hystrix It J Mammal 26:9–12.

Ross AH (2014) Forensic 3D Database, http://www.3d-id.org/forensic-3d-coordinates.

Schaefer K, Lauc T, Mitteroecker P, Gunz P, Bookstein FL (2006) Dental arch asymmetry in an isolated Adriatic community. Am J Phys Anthropol 129:132–142.

Slice DE (1999) Morpheus et al. Ecology and Evolution. State University of New York, Stony Brook.

Takahashi H, Yamashita M, Shigehara N (2006) Cranial photographs of mammals on the web: the Mammalian Crania Photographic Archive (MCPA2) and a comparison of bone image databases. Anthropol Sci 114:217–222.

Viscosi V, Cardini A (2011) Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. PLoS One 6: e25630.

Willmore KE, Klingenberg CP, Hallgrimsson B (2005) The relationship between fluctuating asymmetry and environmental variance in rhesus macaque skulls. Evolution 59:898–909.

Yazdi FT, Adriaens D, Darvish J (2012) Geographic pattern of cranial differentiation in the Asian Midday Jird Meriones meridianus (Rodentia: Muridae: Gerbillinae) and its taxonomic implications. J J Zool Syst Evol Res 50:157–164.

Yazdi FT, Colangelo P, Adriaens D (2014) Testing a long-standing hypothesis on the relation between the auditory bulla size and environmental conditions: a case study in two jird species (Muridae: Meriones libycus and M. crassus). Mammalia 79:185–200.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1 Guidelines to reconstruct a missing side by mirror reflection, using free user-friendly software.