

An Integrated Approach for Landmark-Based Resistant Shape Analysis in 3D

Sebastián Torcida · S. Ivan Perez · Paula N. Gonzalez

Received: 25 September 2013 / Accepted: 27 November 2013
© Springer Science+Business Media New York 2013

Abstract The study of shape changes in morphology has seen a significant renovation in the last 20 years, particularly as a consequence of the development of geometric morphometric methods based on Cartesian coordinates of points. In order to extract information about shape differences when Cartesian coordinates are used, it is necessary to establish a common reference frame or system for all specimens to be compared. Therefore, a central issue in coordinate-based methods is which criterion should be used to align these configurations of points, since shape differences highly depend on those alignments. This is usually accomplished by aligning the configurations in a way that the sum of squared distances between coordinates of homologous points (landmarks) is minimized: the least-squares superimposition method. However, it is widely recognized that this method has some limitations when shape differences are not homogeneous across landmarks. Here we present an integrated approach for the resistant shape comparison of 3D landmark sets. It includes a new ordinary resistant Procrustes superimposition and its corresponding generalized resistant Procrustes version. In addition, they are combined with existing resistant multivariate statistical techniques for depicting the results. We demonstrate, by using both simulated and real datasets, that resistant Procrustes better detects and measures localized shape variation whenever present in up to half but one of

the landmarks. The resistant Procrustes results are highly concordant with a priori biological information, and might dramatically improve the quality of inferences on patterns of shape variation.

Keywords Resistance · Repeated medians · 3D landmarks · Procrustes superimposition · Spatial median · Resistant MDS

Introduction

The study of evolutionary and developmental changes in morphology has seen a significant renovation for the last two decades, particularly due to the use of methods based on Cartesian coordinates of points or landmarks which capture the geometric information of phenotypic structures (Rohlf and Marcus 1993; Adams et al. 2004; Mitteroecker and Gunz 2009). Because a large proportion of phenotypic variation among individuals and/or species involves spatial changes in specific anatomical points, landmark-based analysis are shedding light on morphometric aspects not previously addressed (Zelditch et al. 2004; Mitteroecker et al. 2005; Hallgrímsson and Lieberman 2008).

In order to compare all of a set of specimens to extract information about the spatial relationships among landmarks, it is necessary to establish a common reference frame or coordinate system. A central issue in coordinate-based methods is therefore which criterion should be used to align these configurations of points, since shape differences among specimens highly depend on those alignments (Richtsmeier et al. 2002; Perez et al. 2006; Theobald and Wuttke 2006; Catalano and Goloboff 2012). An optimal superimposition exhibiting shape differences can only be achieved when differences due to translation, scaling and

S. Torcida (✉)
Departamento de Matemática, Facultad de Ciencias Exactas,
Campus, UNICEN, 7000 Tandil, Argentina
e-mail: unisebas@gmail.com

S. Ivan Perez · P. N. Gonzalez
División Antropología, Museo de La Plata, Universidad
Nacional de La Plata, CONICET, Paseo del Bosque s/n,
1900 La Plata, Argentina

rotation or reflection have been filtered out, because these transformations do not alter shape. The least-squares (LS) or classical *Procrustes* superimposition has become the preferred alignment criterion within the *morphometric synthesis* (Bookstein 1996; Adams et al. 2004, 2013; Zelditch et al. 2004; Mitteroecker and Gunz 2009). This method minimizes the sum of squared Euclidean distances between the Cartesian coordinates of landmarks after superimposition, and the magnitude of shape differences between any two configurations of landmarks is estimated by this sum of squared Euclidean distances, traditionally named *Procrustes distance*. Procrustes distance is afterwards analyzed by statistical techniques that preserve its geometric properties, such as PCA or multivariate regression (Adams et al. 2004; Zelditch et al. 2004; Mitteroecker and Gunz 2009), and shape changes are finally depicted as relative shifts of landmark positions, or by thin-plate spline interpolating functions that use deformation grids and 3D morphing (Mitteroecker and Gunz 2009; Klingenberg 2013).

Despite being the method commonly used the LS Procrustes alignment has some limitations, particularly when shape differences are not homogeneous across landmarks (Slice 1996; Richtsmeier et al. 2002; Theobald and Wuttke 2006; Van der Linde and Houle 2009). In such cases the resultant superimposition can be somewhat misleading, as relatively large differences in only a few landmarks tend to be spread since the LS Procrustes optimal matching averages the overall lack of fit (Rohlf and Slice 1990). As a consequence, relative shape changes of different parts of a structure or an organism may not be accurately depicted and/or measured, and hence revealed (Fig. 1).

Although several alternatives have been proposed to overcome this drawback (Rohlf 1990; Rohlf and Slice 1990; Zelditch et al. 2004; Van der Linde and Houle 2009), the *robust* or *resistant* Procrustes fit (Siegel and Benson 1982) is probably the most elegant and efficient strategy among them. The resistant alignment between any two configurations of landmarks is intuitively achieved by perfectly superimposing those unchanged landmarks. Then, true shape differences may be clearly exhibited through the lack of fit of the remaining landmarks and their corresponding large residuals (Fig. 1). This parsimonious matching strategy is a realization of the statistical property of *resistance*: whenever a few points from a data set deviate from the trend of shape change for most of points, a resistant procedure does not allow them to have a major influence on the resultant fit. It is worth mentioning that whenever more than half of the landmarks change, the notion of *change* itself -its measurement and/or recognition- becomes ambiguous. The referred strategy is also followed in other biological areas, such as the alignment of DNA sequences (Lemey et al. 2009). The resistant fit (RF)

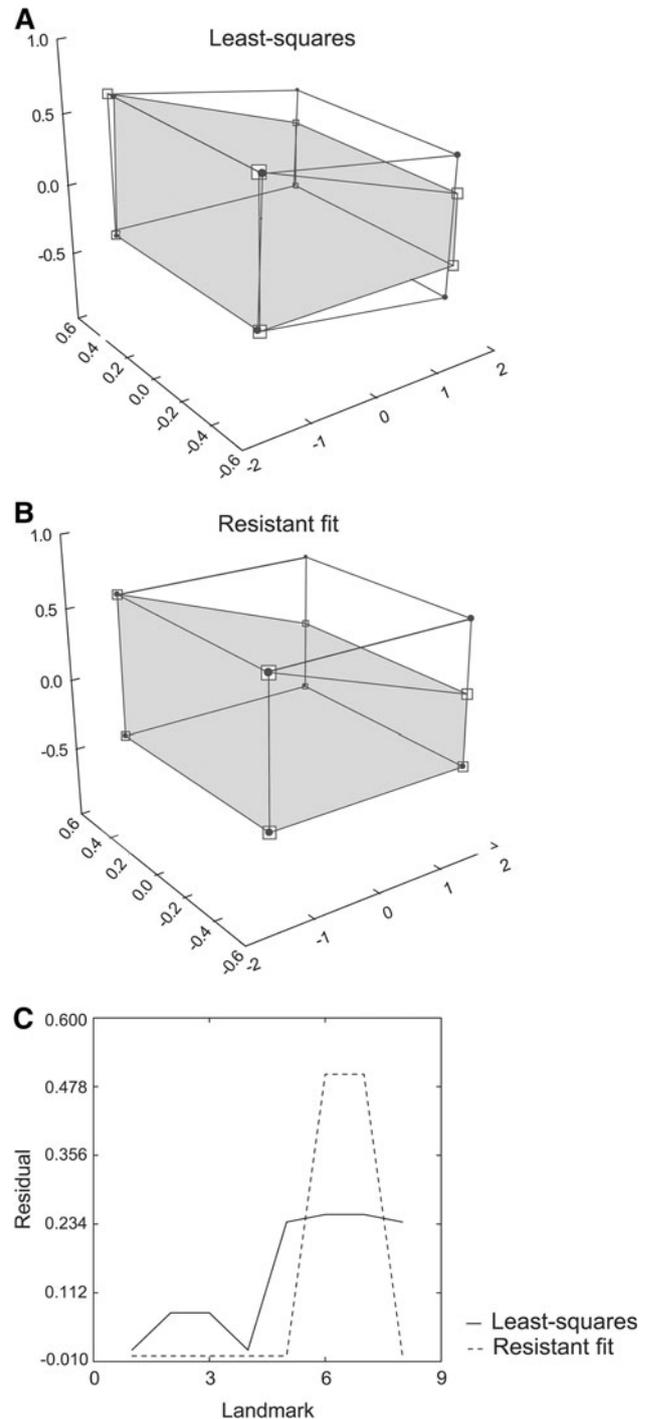


Fig. 1 LS (a) and resistant fit (b) Procrustes superimpositions of the 8 vertices (landmarks) of a standard cube and a similar cube that results from distorting the upper face moving the two top posterior landmarks downwards. Non-squared absolute residuals for each landmark following superimposition are shown in (c)

uses a *repeated-medians* calculation to estimate the alignment parameters: the best possible superimposition for most of the landmarks is therefore obtained, being not affected by huge displacements in just a few points. This

desirable property comes from the fact that if up to half but one values in a set vary, their nested-median remains the same, or resists. Accordingly, shape differences resulting from a resistant superimposition typically reflect more accurately those localized differences among configurations. This facilitates the biological understanding and interpretation of phenotypic variation in many situations (Siegel and Benson 1982; Slice 1996).

In this article we first develop an alternative extension of the resistant Procrustes method to handle two configurations of three-dimensional landmarks (i.e., a new ordinary resistant Procrustes superimposition, ORP). This novel formulation avoids the use of skew-symmetric matrices and triplets of points considered in a previous version from Slice (1996) to estimate the rotation matrix. Next, we extend this method to compare more than two configurations of landmarks (i.e., a new generalized resistant Procrustes superimposition, GRP), matching them all iteratively to a pivotal or *resistant consensus* configuration: the configuration whose landmarks are, respectively, the 3D *spatial medians* of the corresponding landmarks; this configuration can be obtained by using a well known algorithm (Weiszfeld 1937). Third, following the resistant superimposition we propose a coherent *distance* to estimate shape differences between configurations; its properties are in turn discussed. Finally, we use this distance as an input for exploratory multivariate analysis based on both the *resistant* version of the *universal* Multidimensional Scaling (Agarwal et al. 2010; rMDS in the reference) and the non-metric Multidimensional Scaling (Taguchi and Oono 2004; nmMDS in the reference), leading to a coherent display of shape differences among configurations. We compare ordinations generated by the rMDS and nmMDS based on the proposed distance with the corresponding ordinations generated by the classical version of the universal Multidimensional Scaling (named fMDS in the reference) and Principal Coordinates based on the Procrustes distance. Examples include both simulated and real data, the latter from phylogenetic and ontogenetic contexts.

Procrustes Superimposition Methods

Procrustes superimposition methods are based upon measurements of n homologous points named landmarks which are identified in every configuration being studied, generally in two or three dimensions. Let $\mathbf{I}_i^{(k)}$ be the Cartesian coordinates of landmark i ($i = 1, \dots, n$) from configuration k ($k = 1, 2$) stored in row i from the n by p ($p = 2$ or 3) matrix \mathbf{X}_k . A combination of an isotropic scale factor $\rho > 0$, a rotation/reflection matrix \mathbf{R} and a (row) translation term \mathbf{t}

are searched such that the transformed landmarks from matrix \mathbf{X}_1 :

$$\rho \mathbf{I}_i^{(1)} \mathbf{R} + \mathbf{t}$$

achieve maximal agreement with corresponding landmarks $\mathbf{I}_i^{(2)}$ from matrix \mathbf{X}_2 . The choice of both the fixed and transformed matrix is arbitrary; for simplicity, it will be assumed in the following that \mathbf{X}_2 will remain fixed while \mathbf{X}_1 will be transformed. Both LS and resistant Procrustes superimpositions are special cases of orthogonal Procrustes analysis; as mentioned, this work focuses on the latter, and the reader may search in the literature for a detailed description about the former (Gower 1970, 1975; Rohlf and Slice 1990).

Resistant Procrustes Superimposition in 3D

Ordinary Resistant Procrustes (ORP) Superimposition

The resistant Procrustes superimposition was originally stated for two-dimensional landmarks configurations (Siegel and Benson 1982), but some steps in that procedure can be carried over to three dimensions with the proper adaptations: this is the case for the scale factor and the translation term estimates. The estimation of the resistant rotation matrix, however, requires a slightly different approach. Following an unpublished paper by Siegel and Pinkerton, Slice (1996) suggested the use of skew-symmetric matrices to parameterize rotations as a preliminary step to apply repeated-medians. Although it has been widely used since for 3D data handling, Slice's formulation turns out to be rather cumbersome and computationally expensive due to the amount of initial estimates of the rotation matrix required before applying a three times-repeated median on all of them. Moreover, no proof of the goodness of the algorithm is given (something Siegel and Benson carefully did in their pioneer work) leaving reasonable doubts on its average performance. These two drawbacks are solved by our new formulation, which is presented next (see Appendix also). A LS Procrustes superimposition is typically assumed to have been previously performed, to take into account eventually needed reflections and to facilitate the estimation of the rotation angle.

The Resistant Rotation

Consider two sets of homologous points in 3D. Unlike in 2D, the alignment of two points in 3D it is not a well defined problem from a mathematical point of view: there are infinitely many solutions, and triplets of points should

instead be considered for a correct formulation. However, it is not possible to exactly map the coordinates of three arbitrary landmarks from one configuration in 3D onto the corresponding landmarks from the other configuration, yet considering rotation, scaling and translation. Overcoming this intrinsic limitation constitutes a major challenge when attempting to extend the resistant Procrustes method from 2D to 3D data.

Every unordered triplet of landmarks from a 3D configuration describes a planar triangle, and three pairs of corresponding landmarks will therefore describe two homologous triangles. From a geometrical point of view, the alignment of a pair of homologous triangles in 3D requires two conditions to be fulfilled:

1. The planes containing the triangles must be parallel, and
2. the alignment of one of the corresponding sides of the triangles it is also necessary.

To estimate the rotation matrix, Slice (1996) considered triplets of landmarks of the form:

$$\{\mathbf{I}_i^{(k)}, \mathbf{I}_j^{(k)}, \mathbf{I}_m^{(k)}\} (k = 1, 2)$$

which have three degrees of freedom and generate $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$ combinations during the estimation process. In contrast, triplets of the form:

$$\{\mathbf{I}_i^{(k)}, \mathbf{I}_j^{(k)}, \mathbf{0}\} (k = 1, 2)$$

are now chosen. They retain only two degrees of freedom as in 2D, and involve significantly fewer combinations of landmarks: $\binom{n}{2} = \frac{n(n-1)}{2}$, resulting in simplicity and computation time saving. Here $\mathbf{0}$ denotes the zero vector, which geometrically refers to the origin of the coordinate system in 3D.

The estimation of the rotation matrix \mathbf{R} may now be approached. For every pair of homologous triplets:

$$\{\mathbf{I}_i^{(k)}, \mathbf{I}_j^{(k)}, \mathbf{0}\} (k = 1, 2)$$

describing homologous triangles whose sides are given by vectors $\mathbf{I}_i^{(k)}$, $\mathbf{I}_j^{(k)}$ and $\mathbf{I}_j^{(k)} - \mathbf{I}_i^{(k)}$, an initial estimate \mathbf{R}_{ij} of the rotation matrix is required to align side $\mathbf{I}_j^{(k)} - \mathbf{I}_i^{(k)}$ from both triangles through the equation:

$$(\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)})\mathbf{R}_{ij} = (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \quad (1)$$

in the view of condition 2) above. (Note that vectors $\mathbf{I}_j^{(k)} - \mathbf{I}_i^{(k)}$ need first to be made of the same length, because lengths are preserved by rotation matrices. Is typical to

make them of unit length -named *unitary* vectors- as the needed operations can always be performed)

Assume for now that landmarks $\mathbf{I}_i^{(k)}$ and $\mathbf{I}_j^{(k)}$ are not collinear in any of the configurations: the area of the triangle they describe is therefore not null, and the spatial orientation of that triangle or the plane containing it can be simply determined by the orientation of a vector orthogonal to them. The *cross product* of two arbitrary vectors lying on a plane produces a vector orthogonal to them and to the plane; thus, any two of the triangle sides will serve to this objective. The choice:

$$(\mathbf{I}_j^{(k)} - \mathbf{I}_i^{(k)}) \times \mathbf{I}_j^{(k)}$$

(where \times is the cross-product of 3D vectors) will aid in solving the matrix equation on \mathbf{R}_{ij} . Now, to make two planes parallel it is equivalent and easier to make its corresponding orthogonal vectors parallel. Thus, by condition 1 above it seems appropriate for rotation matrix initial estimate \mathbf{R}_{ij} to satisfy also equation:

$$\left[(\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)} \right] \mathbf{R}_{ij} = \left[(\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)} \right], \quad (2)$$

where reasoning as before both cross-product vectors have been made unitary. Finally, rotation matrices in 3D are compatible with the cross-product (Gantmacher 1959) in the sense that they preserve orientation and satisfy:

$$(\mathbf{u} \times \mathbf{v})\mathbf{R} = \mathbf{uR} \times \mathbf{vR}$$

for every choice of equal length vectors \mathbf{u} and \mathbf{v} in 3D. Combining this property with Eqs. (1) and (2) above provides a third and final equation:

$$\left\{ (\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \left[(\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)} \right] \right\} \mathbf{R}_{ij} = \left\{ (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \left[(\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)} \right] \right\} \quad (3)$$

where again involved vectors are assumed to be unitary. All these considerations are summarized in the following matrix equation for the initial rotation matrix estimate \mathbf{R}_{ij} :

$$\left[\begin{array}{c} \mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)} \\ (\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)} \\ (\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \left[(\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)} \right] \end{array} \right] \mathbf{R}_{ij} = \left[\begin{array}{c} \mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)} \\ (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)} \\ (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \left[(\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)} \right] \end{array} \right] \quad (4)$$

where matrices between brackets are clearly orthogonal since their rows are pairwise orthogonal and unitary by construction, under the assumption of non-collinearity of

landmarks $\mathbf{I}_i^{(k)}$ and $\mathbf{I}_j^{(k)}$. Solving this equation is immediate by using the orthogonality of the two known matrices:

$$\mathbf{R}_{ij} = \begin{bmatrix} \mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)} \\ (\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)} \\ (\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times [(\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}) \times \mathbf{I}_j^{(1)}] \\ \mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)} \\ (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)} \\ (\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times [(\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}) \times \mathbf{I}_j^{(2)}] \end{bmatrix}' \tag{5}$$

where “'” denotes the matrix transpose. An initial estimate \mathbf{R}_{ij} of the rotation matrix is in this way obtained for every pair of corresponding landmarks $\mathbf{I}_i^{(k)}$ and $\mathbf{I}_j^{(k)}$ ($k = 1, 2$).

According to matrix theory results (Gantmacher 1959), every 3D rotation matrix can be expressed in a rather simple or *canonical* way by means of both an unitary (row) *eigenvector* \mathbf{v} and an angle θ . The eigenvector \mathbf{v} expresses the direction of the rotation *axis*, a straight line left fixed by the rotation, while the angle θ measures the amount of circular displacement around the axis, in radians. A correspondence between 3D rotation matrices \mathbf{R} and pairs of the form (\mathbf{v}, θ) arises in this way naturally. Moreover, this link can be made unique by selecting a fixed range for angle values θ and establishing some criterion to pick vector \mathbf{v} among the two possible and opposite unitary eigenvectors associated to a straight line in three dimensions. The range $[-\pi, \pi]$ is a natural choice for θ if a preliminary LS Procrustes superimposition has been performed, and the uniqueness of vector \mathbf{v} can be achieved if, for instance, its first coefficient is chosen to be positive. This bijective relationship can be expressed by the notation $\mathbf{R}_{(\mathbf{v}, \theta)}$ that will be adopted in the following. Once the initial rotation matrix estimate \mathbf{R}_{ij} is obtained, the associated rotation axis \mathbf{v}_{ij} and rotation angle θ_{ij} can be computed by standard matrix decomposition techniques. The *repeated-median* is then applied componentwise to all these vectors:

$$\text{med}_i (\text{med}_{j \neq i} \mathbf{v}_{ij}) = \tilde{\mathbf{v}} \tag{6}$$

and to their corresponding angles:

$$\text{med}_i (\text{med}_{j \neq i} \theta_{ij}) = \tilde{\theta} \tag{7}$$

producing axis and angle final estimates $(\tilde{\mathbf{v}}, \tilde{\theta})$ on which the *final* rotation matrix estimate $\tilde{\mathbf{R}}_{(\tilde{\mathbf{v}}, \tilde{\theta})} = \tilde{\mathbf{R}}$ is based.

If the assumption of linear independence failed to be true for an arbitrary pair of landmarks $\mathbf{I}_i^{(k)}$ and $\mathbf{I}_j^{(k)}$ in any of the configurations, two situations might be given: they could be collinear in one configuration but not in the other, or they could be collinear in both. In any case, the

collapsed area of at least one of the two triangles described by the triplet $\{\mathbf{I}_i^{(k)}, \mathbf{I}_j^{(k)}, \mathbf{0}\}$ would make impossible the desired alignment. The parsimony principle would suggest in this case not to rotate, and to set by default a trivial initial estimate:

$$\mathbf{R}_{ij} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ (the identity matrix)}$$

of the rotation matrix, associated to rotation angle $\theta = 0$ and to rotation axis:

$$\mathbf{v}_{ij} = [1 \ 0 \ 0].$$

The Scale Factor and the Translation Term

To complete the algorithm, the estimation of both the scale factor $\rho > 0$ and the translation term \mathbf{t} is considered. As mentioned before, straightforward extension from their 2D analogues takes place.

For every pair of different landmarks $\mathbf{I}_i^{(k)}$ and $\mathbf{I}_j^{(k)}$ ($1 \leq i \neq j \leq n$), an initial estimate ρ_{ij} of ρ is defined as the scale factor by which the segment connecting these two points in configuration \mathbf{X}_1 is made of the same length as the corresponding segment in \mathbf{X}_2 :

$$\rho_{ij} = \frac{\|\mathbf{I}_j^{(2)} - \mathbf{I}_i^{(2)}\|}{\|\mathbf{I}_j^{(1)} - \mathbf{I}_i^{(1)}\|}. \tag{8}$$

The final scale estimate is naturally obtained by taking the doubly-repeated median:

$$\text{med}_i (\text{med}_{j \neq i} \rho_{ij}) = \tilde{\rho}. \tag{9}$$

Once the final rotation matrix and scale factor estimates $\tilde{\mathbf{R}}$ and $\tilde{\rho}$ have been independently obtained, the translation term \mathbf{t} is initially estimated by every row \mathbf{t}_i ($i = 1, \dots, n$) of the residual matrix:

$$\mathbf{X}_2 - \tilde{\rho} \mathbf{X}_1 \tilde{\mathbf{R}} = \mathbf{T} = \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_n \end{bmatrix} \tag{10}$$

The single componentwise median along each column of matrix \mathbf{T} produces the final estimate:

$$\text{med}_i \mathbf{t}_i = \tilde{\mathbf{t}} \tag{11}$$

of the translation term.

Generalized Resistant Procrustes (GRP) Superimposition

In order to generalize the proposed ORP method for optimal superimposition of $r > 2$ configurations of landmarks

$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$ in an iterative procedure, both the notion of a consensus configuration and a stopping criterion are needed.

Two alternatives for a consensus configuration \mathbf{Y} seem to be in line with the appealing features of the resistant method: the *componentwise median configuration* of the $n \times 3$ matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$ (Rohlf and Slice 1990), or their *spatial median*. For the latter, the consensus configuration \mathbf{Y} is chosen to be the matrix whose rows or landmarks are, respectively, the spatial medians (Weiszfeld 1937) of the corresponding rows or landmarks from matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$. Unlike the componentwise median, the spatial median configuration \mathbf{Y} is *rotationally invariant* and seems thus to be more appropriate in the context of shape analysis.

In any case, the monotonically convergence of an iterative procedure is not guaranteed as no explicit criterion is being optimized. The median *residual* (i.e., the median of the non squared Euclidean distances between corresponding landmarks) in the consensus configuration \mathbf{Y} for successive iterations:

$$\text{med}_i \|\mathbf{y}_i - \tilde{\mathbf{y}}_i\|$$

can be used as the stopping criterion (Rohlf and Slice 1990), and this is our choice. In addition, it was combined with an efficiency-based configuration update criterion to aid in reaching convergence: each configuration \mathbf{X}_k ($k = 1, 2, \dots, r$) is transformed *only* when its median residual after transformation is confirmed to decrease, meaning that after transformation the lower 50 % of its residuals will be spread over a shorter range (see steps 5 and 6 below).

The algorithm steps to obtain the new GRP superimposition on r arbitrary $n \times 3$ matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$ are as follows:

1. Perform a preliminary LS generalized Procrustes superimposition. This achieves an initial alignment that takes into account reflections if needed, and usually enables a better estimation of the resistant rotation angle.
2. To aid in reaching convergence, scale each of the configuration matrices \mathbf{X}_k ($k = 1, 2, \dots, r$) to a common size, producing either a *unit median inter-landmark distance*:

$$\text{med}_{j \neq i} \|\mathbf{l}_j^{(k)} - \mathbf{l}_i^{(k)}\| = 1$$

or a *unit median landmark length*:

$$\text{med}_i \|\mathbf{l}_i^{(k)}\| = 1.$$

3. Set \mathbf{Y} , the initial consensus matrix, as the matrix whose i th row \mathbf{y}_i is the spatial median of the corresponding

rows from current matrices \mathbf{X}_k ($k = 1, 2, \dots, r$) by using the Weiszfeld algorithm.

4. For every $k = 1, \dots, r$ evaluate the initial residuals:

$$\|\mathbf{l}_i^{(k)} - \mathbf{y}_i\| \quad (i = 1, \dots, n)$$

between all pairs of corresponding landmarks from current matrix \mathbf{X}_k and current consensus matrix \mathbf{Y} , respectively, and next evaluate the corresponding *median residual*:

$$\text{med}_i \|\mathbf{l}_i^{(k)} - \mathbf{y}_i\|.$$

(Note that whenever more than half of the landmarks from \mathbf{X}_k can be perfectly superimposed to the corresponding landmarks from consensus \mathbf{Y} , this median residual will be zero).

5. For every $k = 1, \dots, r$ perform a *tentative* ORP superimposition of current matrix \mathbf{X}_k to current consensus matrix \mathbf{Y} , obtaining the corresponding new or *fitted* matrix $\tilde{\mathbf{X}}_k$. Compute the new residuals:

$$\|\tilde{\mathbf{l}}_i^{(k)} - \mathbf{y}_i\| \quad (i = 1, \dots, n)$$

and the corresponding new median residual:

$$\text{med}_i \|\tilde{\mathbf{l}}_i^{(k)} - \mathbf{y}_i\|.$$

6. If the new median residual is lower than the previous one, confirm the tentative fit and replace the current matrix \mathbf{X}_k by the corresponding fitted matrix: $\mathbf{X}_k = \tilde{\mathbf{X}}_k$. Otherwise, leave matrix \mathbf{X}_k unchanged. Calculate afterwards the new consensus matrix $\tilde{\mathbf{Y}}$ as the spatial median from the updated matrices.
7. If the median difference between the new and the previous consensus matrices:

$$\text{med}_i \|\mathbf{y}_i - \tilde{\mathbf{y}}_i\|$$

is greater than a specified tolerance, set $\mathbf{Y} = \tilde{\mathbf{Y}}$ and repeat from step 5 above. Otherwise, the iteration is finished.

Resistant distance and exploratory multivariate analysis

Once the iterative procedure is completed and a GRP superimposition of configurations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$ has been obtained, a *resistant distance* d_{kj} is computed for each pair of configurations $\mathbf{X}_k, \mathbf{X}_j$ as the total sum of residuals through all pairs of corresponding landmarks:

$$d_{kj} = \text{distance}(\mathbf{X}_k, \mathbf{X}_j) = \sum_{i=1}^n \left\| \mathbf{I}_i^{(k)} - \mathbf{I}_i^{(j)} \right\| \quad (1 \leq k, j \leq r) \quad (12)$$

In the context of morphometries, this measure has been frequently ruled out because it is less tractable from a mathematical point of view: it is neither differentiable, nor directly associated to an inner product which would link distances with lengths and angles. Unlike Procrustes distance, however, this distance is less affected by the presence of huge displacements in just a few landmarks, being in this way an appropriate tool for measuring those shape differences detected by a RF. In addition, whenever more than 50 % of the corresponding landmarks of matrices $\mathbf{X}_k, \mathbf{X}_j$ can be superimposed perfectly, the RF performs that superimposition and makes each of the corresponding terms in (12) be equal to 0. All these reasons have made the proposed distance a reasonable choice to be included in our approach.

In order to visualize shape differences following a generalized Procrustes superimposition, the *universal multidimensional scaling* framework (Agarwal et al. 2010) was adopted. This versatile algorithm enjoys convergence and accuracy, allowing several input distances to be combined with different cost functions to produce its outputs. The classical version fMDS, where the generated distances are (non-squared) Euclidean and the cost function is the sum of squares between the original and the generated distances, was used for low-dimensional visualization of the generalized LS Procrustes results. Correspondingly, the resistant version rMDS, where the generated distances are (non-squared) Euclidean and the cost function is the sum of absolute differences between original and generated distances, was used for visualization of the GRP results. The rMDS takes as input: a reference set of 3D points or configurations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r$; the corresponding $r \times r$ distance or similarity matrix $\mathbf{D} = [d_{kj}]$ (in our case, the matrix of the resistant distances defined in (12)) and a parameter m (typically, $m = 2$) expressing the dimension in which those distances in \mathbf{D} are to be depicted. The rMDS approach then obtains, through an iterative procedure, a corresponding set of points $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_r$ in the m -dimensional space such that the sum of absolute non-squared differences (cost function) between the original distances and the (non squared) Euclidean distances between the generated points is minimized:

$$\min_{\mathbf{Z}_1, \dots, \mathbf{Z}_r \in \mathbb{R}^m} \sum_{1 \leq k, j \leq r} |d_{kj} - \|\mathbf{Z}_k - \mathbf{Z}_j\|| \quad (13)$$

This resistant version rMDS was first suggested in (Cayton and Dasgupta 2006); it produces a lower dimensional representation of points that best recovers the original distances for most of them, being less sensitive than a LS

cost function to the presence of outliers. Once again, this is the typical purpose of all resistant tools.

For comparative reasons we also employed a *principal coordinates* analysis (PCo; Davis 1986) in order to visualize the Procrustes distance, and a *non-metric multidimensional scaling* (nmMDS; Taguchi and Oono 2004) to visualize the resistant distance. PCo analysis finds the eigenvalues and eigenvectors of the matrix containing the Procrustes distances between all pairs of landmark configurations (Davis 1986). Then, the eigenvectors or PCo scores are used to plot the differences among these configurations. The nmMDS analysis, in turn, attempts to place the resistant distances between all pairs of landmark configurations in a two or three-dimensional coordinate system such the rank of the distances is preserved (Taguchi and Oono 2004). The PCo and nmMDS analyses were performed in PAST (Hammer et al. 2001).

Examples

All simulated and real examples were analyzed through Scilab written routines; pseudocode versions of them are available under request. A flexible and integrated software toolbox including all mentioned techniques is under preparation.

Simulated Data

Different simulation experiments were designed to evaluate the performance of the resistant Procrustes superimposition under the effect of an increasing amount of variation in a growing subset of landmarks. Based on a single skull that was extracted from a real sample of 152 *Alouatta caraya* specimens, 5 data sets consisting of 10 simulated primate skulls each were computer-generated. The target skull consisted in 35 cranial landmarks in 3D grouped in face, vault and base.

For the first set, 5 landmarks among a total of 26 from the face and cranial vault (Fig. 2) were randomly chosen and 10 individuals were simulated; as the variability of the base was known to be lower in the dataset, landmarks from this structure were not modified. In all of simulated individuals, the same 5 chosen landmarks were perturbed by a centered normally distributed noise whose standard deviation was set to be half the distance departing that landmark from its nearest, in order to obtain 95 % of simulated landmarks not altering this minimal distance. A similar procedure was performed by perturbing 10, 15, 20 and 25 randomly chosen landmarks among the 26 landmarks from the face and cranial vault. For simplicity, landmarks from the cranial base were constrained to present zero variation in all the simulations. The pattern of shape variation was in

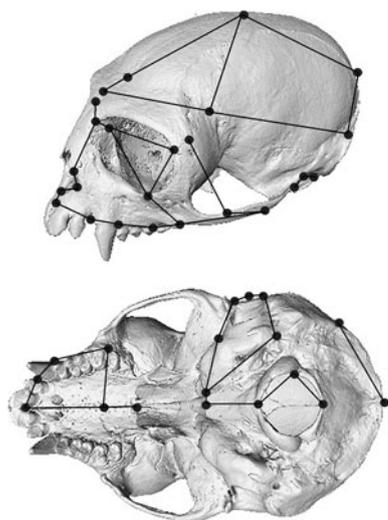


Fig. 2 Cranial landmarks recorded from New World monkeys using a 3D Microscribe G2X digitizer. Wireframes exhibited next in Figs. 10 and 12 are also displayed

this way designed to reproduce the known changes of these structures during the ontogeny of mammalian skulls, where the cranial base reaches adult proportions at a younger age than the face and vault (Cheverud 1995; Sperber 2001).

Two main conclusions were drawn from these simulations. First, following the resistant Procrustes superimposition the residual between each perturbed landmark from any of the generated skulls and the corresponding landmark from the target (real *Alouatta caraya*) skull was always larger than the same residual calculated following a LS Procrustes superimposition (Fig. 3). This result confirmed the specific ability of the RF, when compared to LS, to more effectively detect and depict the true localized shape differences.

Second, the proportion of the total shape distance (considering the Procrustes distance after a LS fit, and the proposed resistant distance after a resistant superimposition) between the simulated skulls and the target skull that was accumulated solely by those perturbed landmarks (the ones capturing the true shape differences by design), the RF exhibited much higher accuracy whenever up to half of the landmarks were perturbed: averages between 87 and 95 % of the total resistant distance were associated to the true shape differences; instead, averages between 60 and 78 % of the total Procrustes distances were associated to true shape differences (Fig. 4). This confirmed that a resistant superimposition enables a more accurate measurement of shape differences: almost all the shape distance was concentrated on those landmarks where localized deformation had truly taken place. The LS fit, on the other hand, added artificial variation in many landmarks not associated to those partial deformations. Whenever more than half of the landmarks were perturbed (20 and 25

landmarks in our simulations, not anymore a pattern of local deformation) the percentage of total shape distance recovered by them was similar for both methods (see also Fig. 4).

Ontogeny of Human Mandible

The postnatal ontogeny of human mandible is an example of moderate shape variation in which change is not expected to be particularly concentrated in a few landmarks. An ontogenetic series including individuals of both sexes with ages ranging from 7 to 45 years was analyzed. These specimens belong to the collection of Identified Skeletons of the University of Coimbra (Rocha 1995). In order to describe shape changes throughout ontogeny, 16 landmarks in 3D were digitized using a Microscribe G2X (Fig. 5). Both generalized LS and resistant Procrustes superimpositions were performed using this dataset. The resulting distances between individuals were represented in a low-dimensional space by means of the classical (fMDS) and the resistant (rMDS) versions of the universal MDS, and by the PCo and nmMDS implemented in PAST, as well.

Ordinations of specimens along the fMDS and nmMDS axes (Fig. 6) were analogous to those obtained by PCo and rMDS, respectively, and therefore are not shown. Both the LS (Fig. 6a) and resistant Procrustes (Fig. 6b) superimpositions exhibited shape differences between the distributions of adults and subadults. The correlation between the respective distance matrices was high and significant (Spearman correlation $R = 0.94$; Pearson correlation $r = 0.96$; regression fit $r^2 = 92\%$; Fig. 7a). An additional superimposition by both methods of two extreme configurations also showed the same pattern of shape change with age (Fig. 8); the alveolar region became relatively shorter and the angle between the alveolar region and the ascending ramus was narrower in older individuals (note that visualization techniques other than wireframes are also available, and they could be used to explore the pattern of shape changes depicted by each method of superimposition; see Klingenberg 2013). There were also striking changes in the posterior side of the ascending ramus: adults have the condylar process placed upward and the angular process placed downward compared to the morphology of young individuals.

Cranial Ontogeny in Species of New World Monkeys

A third example focused on shape changes along postnatal ontogeny of skulls from three platyrrhine species: *Cebus apella*, *Callithrix jacchus* and *Alouatta caraya*. These species exhibit a relatively large variation which is expected to be concentrated in particular landmarks. We

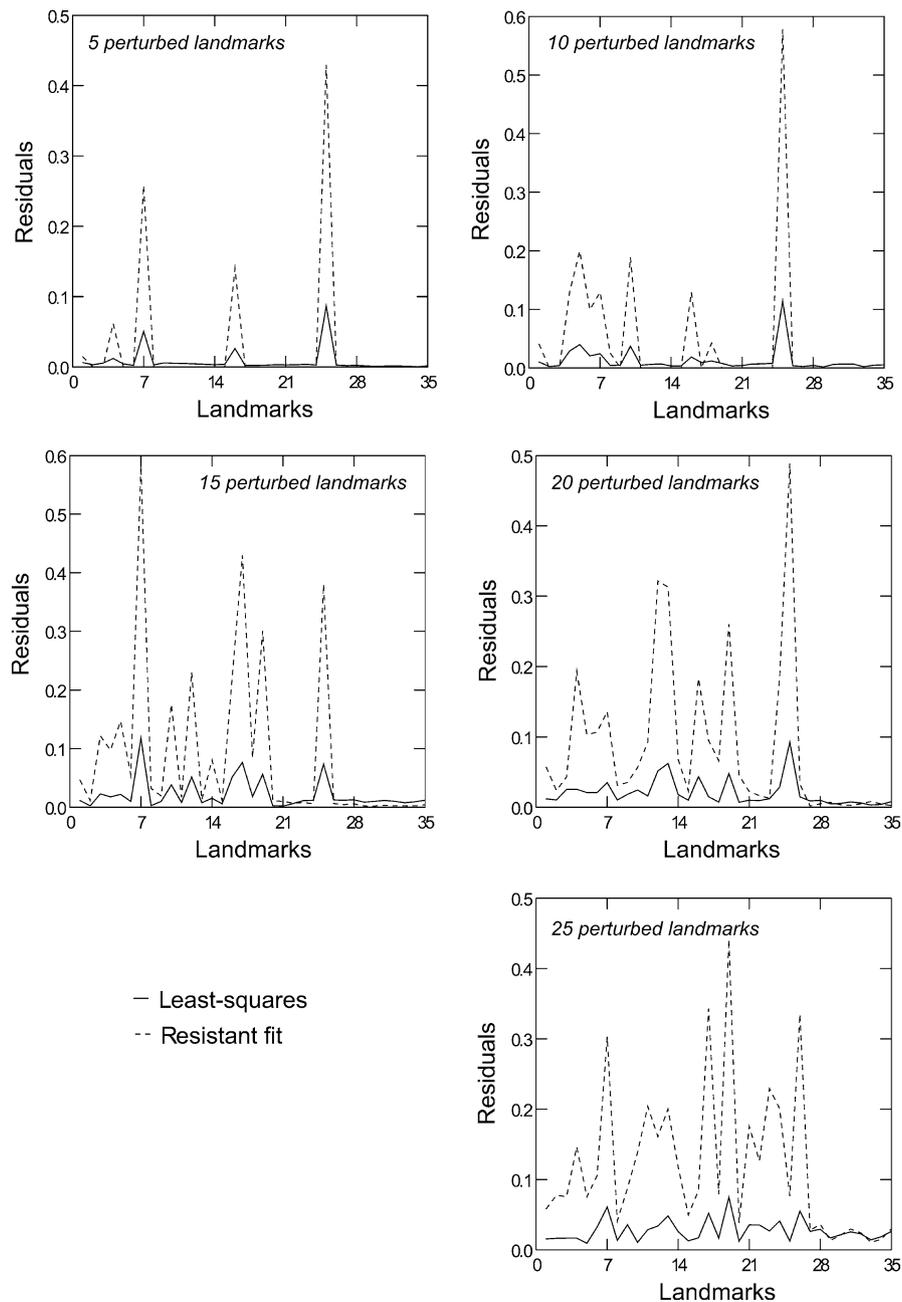


Fig. 3 Plots showing the non-squared absolute residuals between target and selected simulated skulls for each landmark following LS and resistant fit (RF) Procrustes superimpositions

analysed the three ontogenetic series including individuals of both sexes and different ages. These specimens are housed in the Museo Argentino de Ciencias Naturales “Bernardino Rivadavia” (Argentina), in the Museo de Ciencias Naturales de La Plata (Argentina) and in the Museu Nacional de Rio de Janeiro (Brazil). In order to describe shape changes throughout ontogeny, 35 landmarks in 3D were digitized using a Microscribe G2X (Fig. 2). The

generalized versions of both LS and resistant Procrustes superimpositions were performed, and the resulting distances were plotted by using universal fMDS and nmMDS.

The scatter-plot of LS vs resistant distances (Fig. 7b) for all pairs of individuals from the three species, and the ordination of specimens along the MDS axes based on LS and resistant distances were different (Fig. 9). Using LS and fMDS, the ontogenetic trajectories of the three species

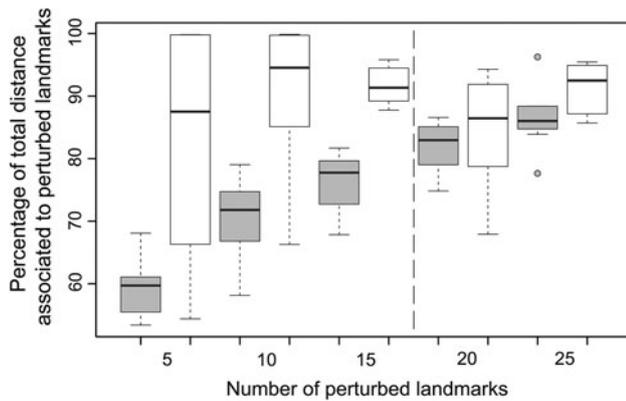


Fig. 4 Percentages of LS (gray) and resistant (white) distances between target and simulated skulls due to true shape differences. Dotted line separates simulated datasets with less (left) and more (right) than 50 % of landmarks perturbed

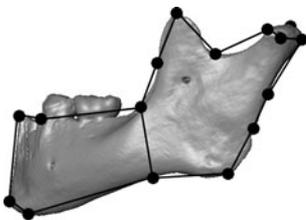


Fig. 5 Landmarks recorded from human mandible using a 3D Microscribe G2X digitizer. Wireframes exhibited next in Fig. 8 are also displayed

shared the same direction, and the *Alouatta* trajectory was an extension of the trajectory followed by *Callithrix* (Fig. 9a). Moreover, the trajectories of *Cebus* and *Alouatta* had a similar extension. Conversely, when the specimens were analysed through the resistant approach the ontogenetic trajectory of *Alouatta* exhibited differences in orientation and extension with respect to the remaining species (Fig. 9b). The scatter-plot in Fig. 7b clearly indicates that the two distance matrices differ considerably, showing a Spearman correlation of 0.81, a Pearson correlation of 0.75 and a regression fit of merely 56 %. The superimposition of two extreme configurations of *Alouatta* by both methods (Fig. 10) exhibited differences in the pattern of shape change with age, as expected. The LS method distributed the variation evenly across the skull, while the resistant method suggested that most of variation was concentrated on the facial region. In this way, the resistant Procrustes superimposition better represented the pattern of primate cranial growth during post-natal ontogeny, which is characterized by the extended growth of facial structures and the associated allometric shape changes (Cheverud 1995; Hallgrímsson and Lieberman 2008).

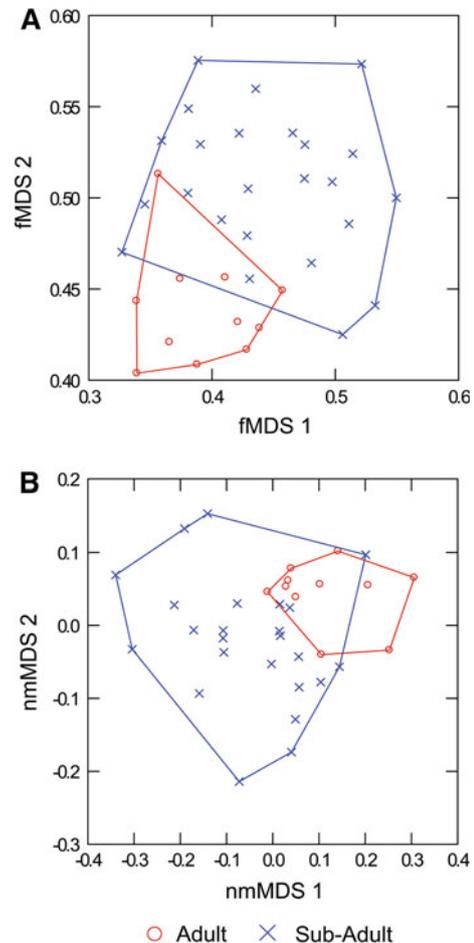


Fig. 6 MDS ordinations showing adult and sub-adult mandibles following generalized LS (a) and resistant (b) Procrustes superimpositions. Dots represent individuals

Cranial Variation among Species of New World Monkeys

New World monkeys were also used to investigate the inter-specific pattern of shape variation at macroevolutionary scales. They are an excellent system for exploring the goodness of the resistant superimposition, due to the large variation in cranial shape (Perez et al. 2011). The variation among 29 species belonging to 5 main clades (*Aotus*, *Cebinae*, *Atelidae*, *Pitheciidae*, *Callithrichinae*) was analyzed; the sample included 221 adult individuals from both sexes (see Perez et al. 2011 for more details on the sample composition). These specimens are deposited in the Museo Argentino de Ciencias Naturales ‘‘Bernardino Rivadavia’’ (Argentina), in the Museo de Ciencias Naturales de La Plata (Argentina), in the Museu Nacional de Rio de Janeiro (Brazil) and in the Museu de Zoologia of Universidade de Sao Paulo (Brazil). In order to describe

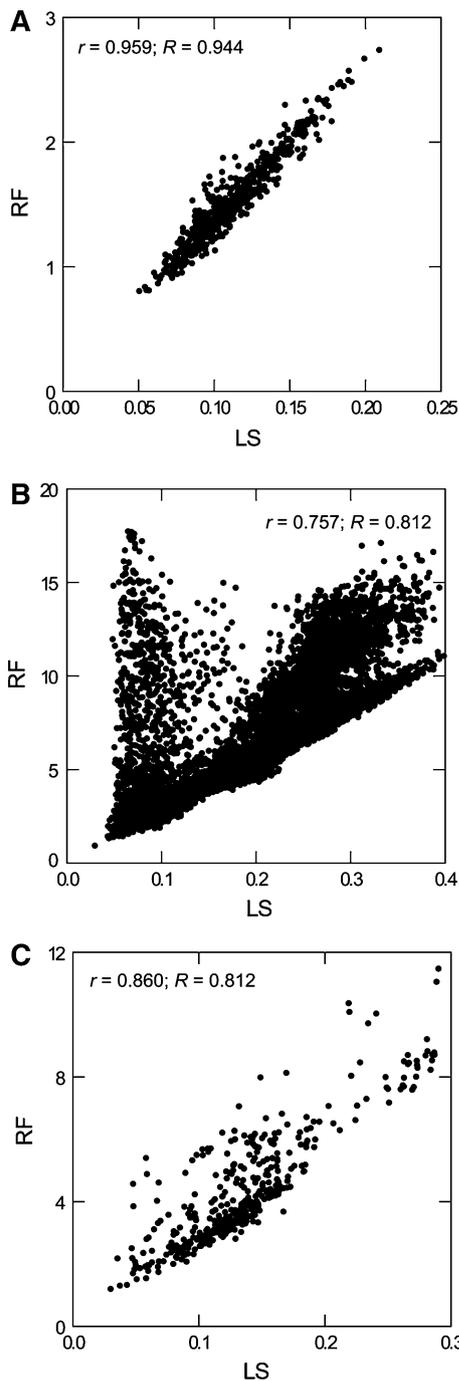


Fig. 7 Scatter-plot of estimated LS versus resistant shape distance matrices for the mandible (a), cranial ontogeny (b) and cranial interspecific (c) datasets. Dots represent the distance values; Spearman rank (R) and Pearson (r) correlation coefficients are indicated

shape changes in 3D, the same set of 35 cranial landmarks considered for the ontogenetic analysis were used (Fig. 2).

The coordinates of landmarks within each species were superimposed first by LS, and consensus configurations for each species were then estimated. These consensus configurations were in turn aligned through a second LS fit,

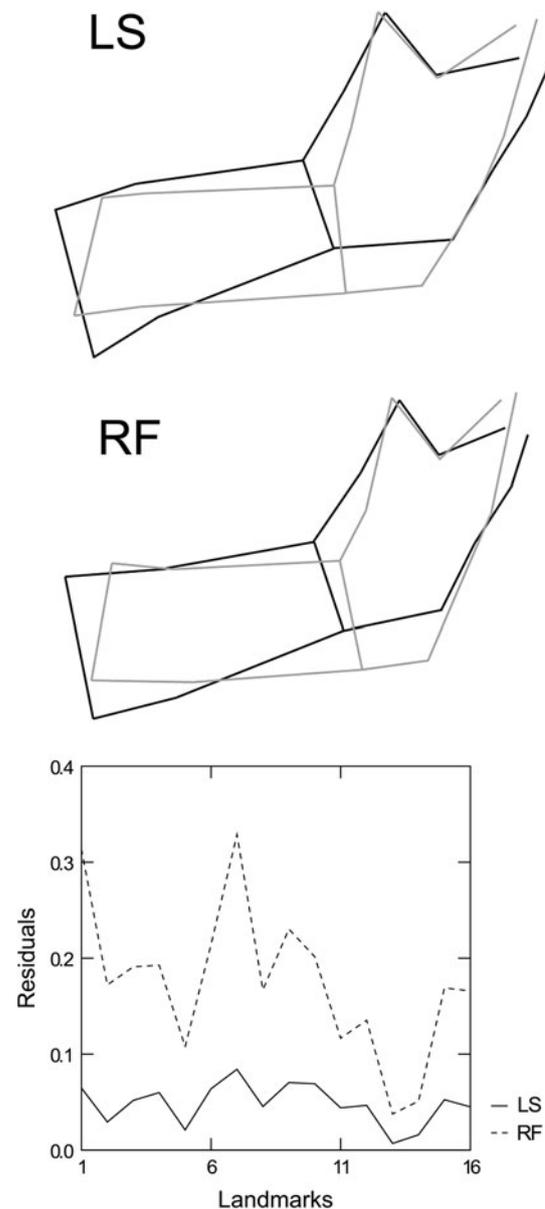


Fig. 8 Wireframes showing mandible shape changes between adult (gray line) and sub-adult (black line) individuals resulting from LS and resistant fit (RF) Procrustes superimpositions. Non-squared absolute residuals for each landmark are also shown

and the obtained distances between the superimposed consensus were represented using fMDS and PCo. When using the resistant approach, a GRP fit followed the LS superimposition within each species, and a resistant consensus for each species was in this way obtained. Afterwards, the LS plus a posterior GRP fit were applied to all the resistant consensus, and the resulting distances were depicted using both the rMDS and nmMDS.

The results showed that the ordination of the species consensus depends both on the method used to superimpose and represent the corresponding distances. While LS seemed to cluster the species by clade (Fig. 11a), the RF

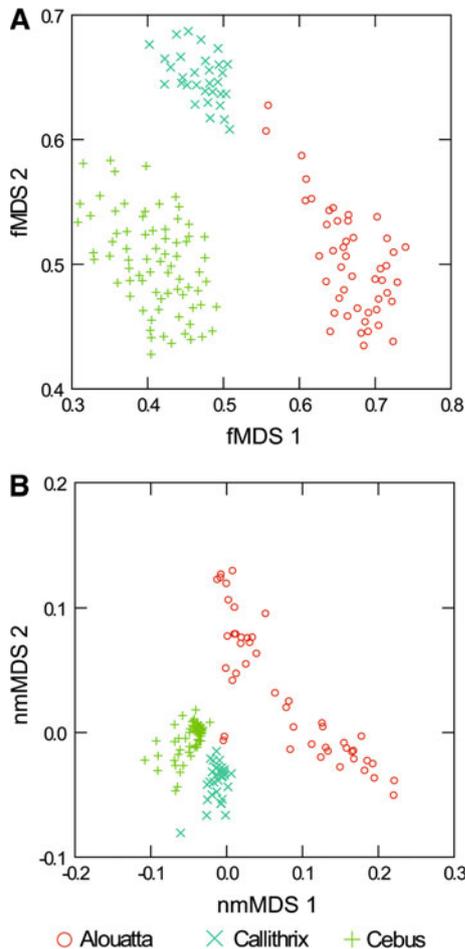


Fig. 9 MDS ordinations displaying ontogenetic trajectories of three platyrrhine species following generalized LS (a) and resistant (b) superimpositions. *Dots* represent specimens from each species

showed greater resemblance among three genera (*Ateles*, *Chiropotes* and *Cebus*) that exhibit convergent morphologies (Fig. 11b), mainly related to a higher degree of encephalization (Hartwig et al. 2011). The scatter-plot in the Fig. 7c showed that both distance matrices differ, but less than the cranial ontogenetic matrices: a Spearman correlation of 0.81, a Pearson correlation of 0.86 and a regression fit of 74 % were obtained. In order to visualize the patterns of shape change, we additionally superimposed by both methods two consensus configurations representing extreme species along the MDS axes (Fig. 12). The results resembled those obtained for the ontogenetic analysis, where the resistant method displayed the largest variation in the facial and some vault landmarks.

Discussion

Both LS and resistant Procrustes fits are based on landmarks: specific points capturing the geometry of those

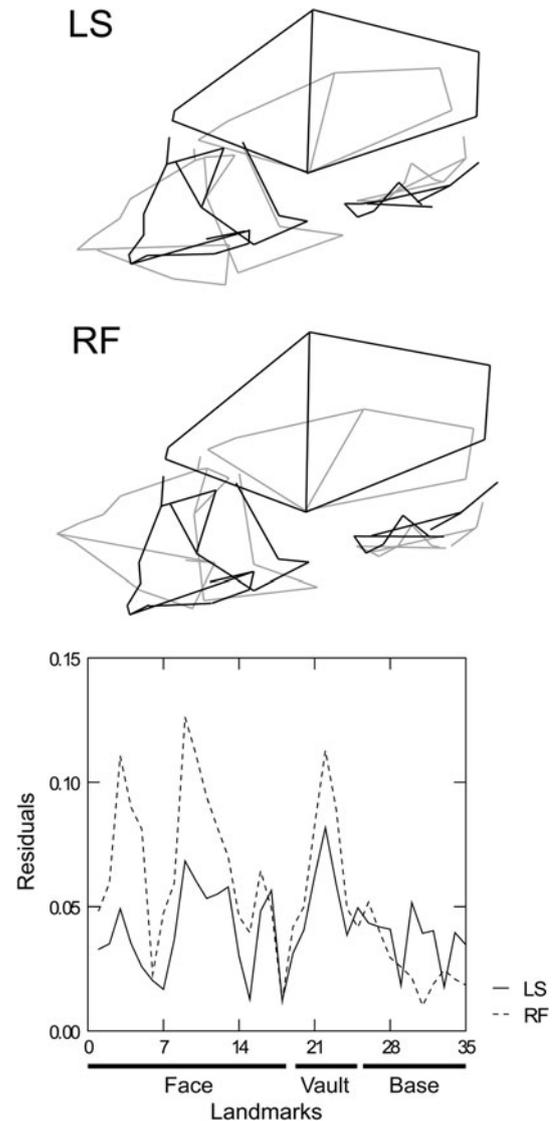


Fig. 10 Wireframes depicting cranial shape changes between adult (gray line) and sub-adult (black line) extreme primates resulting from LS and resistant fit (RF) Procrustes superimpositions. Non-squared absolute residuals for each landmark are also shown

structures being studied. By using homologous points, landmarks enable a rather complete structural understanding of shape variation patterns. The purpose of this work is to offer an integrated resistant approach for landmark-based shape comparisons in 3D: we present both a new ORP method and a corresponding GRP extension for the resistant superimposition of two or more than two configurations of landmarks, respectively. In the process, a resistant consensus configuration and a corresponding resistant distance are also presented.

In terms of the ORP method, the new algorithm differs from Slice's (1996) mainly in two features: (1) Slice's formulation is now greatly simplified by using pairs instead of triplets of homologous points, which results in significant

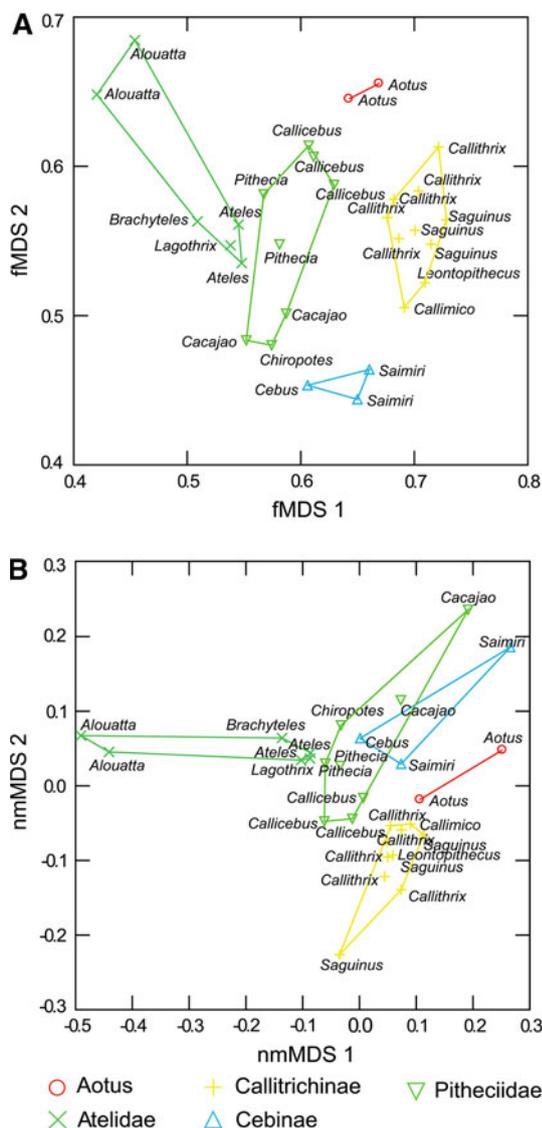


Fig. 11 MDS ordinations of the platyrrhine species following LS (a) and resistant (b) generalized Procrustes superimpositions. *Dots* represent consensus configurations for each species

computation time saving; (2) an optimality theorem is given (see Appendix), establishing the optimal performance of this algorithm whenever localized shape variation or partial deformation takes place. It must be acknowledged that, in comparison to LS, the repeated-medians calculation is computationally more expensive: obtaining the median of n values takes an average time proportional to n (the number of landmarks in this case) because each value has to be examined. Slice's RF requires a processing time proportional to n^3 , while the new algorithm maintains the processing cost at the same level of a two-dimensional superimposition: that is, proportional to n^2 .

The new GRP, in turn, replaces the componentwise median (Rohlf and Slice 1990; Slice 1996) by a consensus

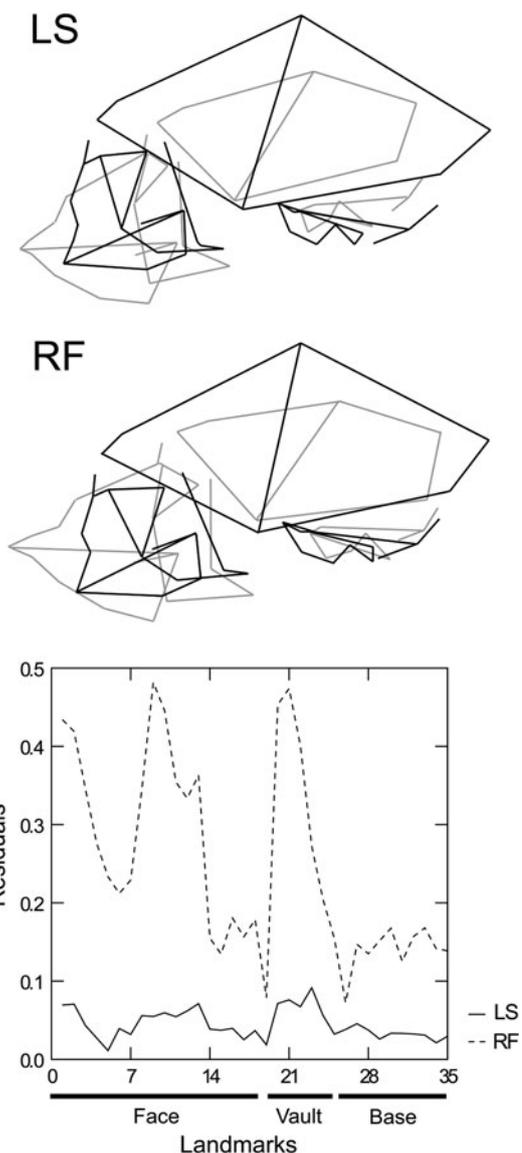


Fig. 12 Wireframes showing cranial shape changes between extreme primate species resulting from LS and resistant fit (RF) Procrustes superimposition (LS). Non-squared absolute residuals for each landmark are also shown

configuration whose landmarks (rows) are, respectively, the spatial median of corresponding landmarks. An appealing feature of this choice is that, unlike the componentwise median, the spatial median is rotationally (and translationally) invariant, just as shape is. A resistant distance is also presented: the overall sum of non-squared Euclidean distances across all pairs of corresponding landmarks. Although not explicitly minimized, the corresponding terms in this distance become zero whenever more than 50 % of the landmarks can be perfectly superimposed.

We have shown through the simulations that the resistant superimposition better detects and measures localized

shape variation: shape change located in up to 50 % of the landmarks. In particular, we demonstrated the ability of the resistant superimposition, when compared to LS, to highlight localized shaped change, producing in general greater residuals for those landmarks capturing partial deformation and lower residuals for those landmarks that did not change. Also, we showed that following a resistant superimposition the proportion of the total shape distance accumulated by those landmarks truly capturing the shape change was above 85 % and typically much higher than its LS analogue, confirming both the accuracy and the efficiency of the resistant approach in measuring real shape variation. Similar results were obtained by Walker (2000) when analyzing the results of generalized LS and resistant Procrustes alignments for estimating known covariance matrices: the latter performed better when less than 25 % of the landmarks had excessive variance, while both methods had a similar performance when more than 75 % of the landmarks had excessive variance.

The analysis of the three real data sets produced dissimilar results. Firstly, the mandibles example showed that when there is small and homogeneously distributed shape change, superimpositions by LS and RFs do not greatly differ and obtain a similar pattern of shape change. This example represents a real case in which more than half of the landmarks were perturbed, or changed along phylogenetic or ontogenetic evolution. Conversely, the phylogenetic and ontogenetic cranial data presented a very different scenario: whenever a moderate-to-great non-homogeneously distributed shape change was suspected, superimpositions by LS and RF methods revealed different patterns of shape change.

The LS fit has been typically favoured as “the” Procrustes method for optimal superimpositions mainly because it is based on the Euclidean distance, the distance we all are used to. Besides, the sum of squared differences it is mathematically more tractable than many other alternative distances: it is differentiable, and has a direct link to the inner product which enables the measurement of vector lengths and angles. Additionally, it has been suggested that LS has a theoretical advantage because it is placed in the geometrical theory of shape from Kendall (1984; Zelditch et al. 2004). Slice (2001), however, showed that LS methods used in biological studies are *only an approximation* to Kendall’s shape space. Lastly, the LS superimposition is considered the only way to obtain the so-called *shape variables* following landmark digitization (Zelditch et al. 2004; Mitteroecker and Gunz 2009). It seems thus reasonable to conclude that the adoption of the LS superimposition as a standard is more a matter of a consensus relying on acquaintance and mathematical or theoretical

convenience than a decision grounded on biological reasons.

A relevant feature of the LS Procrustes method is that it spreads landmark’s variability homogeneously among all of them (Richtsmeier et al. 2002). This is a drawback, only admissible if it is known in advanced that the variation in each point is isotropic (Bookstein 1991) as we suspected was the case for the late ontogeny of the mandibles example. However, many morphometric studies expect shape variability among specimens and/or species to be placed at specific points from structures (Cheverud 1995; Slice 1996; Zelditch et al. 2004; Hallgrímsson and Lieberman 2008). Recently, Van der Linde and Houle (2009) have proposed a modification of the traditional LS Procrustes superimposition based on prior biological knowledge about the variation in form on those structures under study. The method progressively discards landmarks from a dataset if a generalized LS Procrustes superimposition (GLS) excluding those landmarks results in a significant reduction in the Procrustes residuals. This sort of alternative superimposition method, just as the resistant approach proposed here, would therefore be preferable over GLS whenever local shape changes and non-isotropic variation is expected.

Due to its mathematical formulation, the RF typically requires a preliminary LS superimposition to perform reflections, if necessary. The subsequent RF not only does not worsen the results: most of the time, it gives insightful information on where shape differences are specifically placed. When local shape changes do not take place, a RF superimposition does not greatly differ from that obtained by LS.

In the view of the previous considerations, morphometric studies may face the question: should shape differences between two objects, following an estimated optimal superimposition, be depicted (and therefore perceived) as homogeneously distributed, when on the basis of complementary information patterns of localized shape variation would be expected? A quantitative answer can be approached. The breakdown value (Donoho and Huber 1983; Hampel et al. 1986) of an estimation method is a measure describing the percentage of data that can be arbitrarily changed or perturbed without modifying the resultant estimate. LS superimposition breakdown value is 0 %, since a single change in data produces a different estimate. The repeated-medians superimposition breakdown value is, instead, nearly 50 % (the maximum possible; Siegel 1982) as the estimate remains the same even if up to half but one of the points vary. To put it clearly: in the context of shape analysis, if differences between two configurations of landmarks were placed, say, in a single

landmark, the LS Procrustes superimposition would produce different fits depending on the particular location of that landmark, compelling the remaining $n-1$ points that did not change (an undisputed majority) to fall apart in order to reduce the sum of squared differences. As a consequence, artificial shape variation is often introduced in most of the landmarks when using LS, which in turn may mask the real shape differences.

The resistant Procrustes fit is instead designed to add no artificial variation when a relative displacement is present not only in one but in up to half but one of the landmarks. Based on parsimony, its results tend to be typically more in agreement with biological foundations; the adoption of procedures incorporating these biological assumptions might dramatically improve the quality of inferences on shape variation patterns.

In the context of shape analysis, the need of superimposition methods not only mathematically sound but also and perhaps mainly biologically meaningful has been previously pointed out (Richtsmeier et al. 2002; Van der Linde and Houle 2009; Catalano and Goloboff 2012). Since no consensus has been reached yet, methodological contributions in the near future will keep on defying the goodness of traditional LS techniques, aiming at the same time to establish the performance or improvements that alternative methods can bring on solving specific biological problems.

Acknowledgments We want to thank Fundación Antorchas for partially funding Mr. Torcida for this research. We are grateful to Dennis Slice, Dean Adams and Sergio F. dos Reis for their helpful comments on early versions of this manuscript.

Appendix: Optimality of the Presented Resistant Method (Overall Performance When Shape Variation is Located in Half but One of the Landmarks)

The 3D version of the resistant method presented in this work achieves the best possible superimposition whenever shape differences are located in up to 50 % of the landmarks. This is proved next.

Theorem *Suppose there exist parameters $\rho > 0$ (an isotropic scaling), $\mathbf{R}_{(\mathbf{v}, \theta)}$ (a 3D rotation matrix with associated rotation angle θ and rotation axis \mathbf{v} , respectively) and \mathbf{t} (a translation vector) such that the equation:*

$$\mathbf{l}_i^{(2)} = \rho \mathbf{l}_i^{(1)} \mathbf{R}_{(\mathbf{v}, \theta)} + \mathbf{t} \tag{14}$$

holds for more than $\frac{n+1}{2}$ different landmarks, the repeated-medians estimates (6), (7), (9) and the single median (11) are exactly those values; that is,

$$\text{med}_i \left(\text{med}_{j \neq i} \rho_{ij} \right) = \tilde{\rho},$$

$$\text{med}_i \left(\text{med}_{j \neq i} \mathbf{v}_{ij} \right) = \tilde{\mathbf{v}},$$

$$\text{med}_i \left(\text{med}_{j \neq i} \theta_{ij} \right) = \tilde{\theta},$$

and

$$\text{med}_i \mathbf{t}_i = \tilde{\mathbf{t}}.$$

Proof The result is showed only for the rotation matrix parameters, because they use the specific formulation for the 3D case. The proof for the remaining parameters is analogous.

Whenever more than $\frac{n+1}{2}$ in a set of values are the same, the median is that repeated value. Then, when landmarks $\mathbf{l}_i^{(k)}$ and $\mathbf{l}_j^{(k)}$ satisfy (14):

$$\mathbf{l}_j^{(2)} - \mathbf{l}_i^{(2)} = \rho (\mathbf{l}_j^{(2)} - \mathbf{l}_i^{(2)}) \mathbf{R}_{(\mathbf{v}, \theta)} + \mathbf{t}, \tag{15}$$

and setting for the moment $\rho = 1$ (the scale factor can be estimated afterwards, without loss of generality) the rotation matrix \mathbf{R}_{ij} satisfies equations (1), (2), (3) and consequently (4), producing:

$$\mathbf{R}_{ij} = \mathbf{R}_{(\mathbf{v}_{ij}, \theta_{ij})} = \mathbf{R}_{(\mathbf{v}, \theta)}.$$

From this, whenever landmark $\mathbf{l}_i^{(k)}$ satisfies (14), more than half of the remaining landmarks $\mathbf{l}_j^{(k)}$ ($j \neq i$) also satisfy (14) and therefore:

$$\text{med}_{j \neq i} \mathbf{v}_{ij} = \mathbf{v}, \text{ (componentwise median)}$$

$$\text{med}_{j \neq i} \theta_{ij} = \theta,$$

which leads to:

$$\text{med}_i \left(\text{med}_{j \neq i} \mathbf{v}_{ij} \right) = \tilde{\mathbf{v}} = \mathbf{v},$$

$$\text{med}_i \left(\text{med}_{j \neq i} \theta_{ij} \right) = \tilde{\theta} = \theta,$$

and finally:

$$\tilde{\mathbf{R}}_{(\tilde{\mathbf{v}}, \tilde{\theta})} = \mathbf{R}_{(\mathbf{v}, \theta)},$$

which concludes the proof.

References

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: 10 years of progress following the ‘revolution’. *Italian Journal of Zoology*, 71, 5–16.

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2013). A field comes of age: Geometric morphometries in the twenty first century. *Hystrix*, 24(1), 7–14.

- Agarwal, A., Phillips, J. M., & Venkatasubramanian, S. (2010). Universal multidimensional scaling. In *KDD'10: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1149–1158). New York: ACM.
- Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. New York: Cambridge University Press.
- Bookstein, F. L. (1996). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, 58(2), 313–365.
- Catalano, S. A., & Goloboff, P. A. (2012). Simultaneously mapping and superimposing landmark configurations with parsimony as optimality criterion. *Systematical Biology*, 61(3), 392–400.
- Cayton, L., & Dasgupta, S. (2006). Robust Euclidean embedding. In *ICML'06: Proceedings of the 23rd international conference on machine learning* (pp. 169–176). New York, ACM.
- Cheverud, J. (1995). Morphological integration in the saddleback tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist*, 145(1), 63–89.
- Davis, J. C. (1986). *Statistics and data analysis in geology*. New York: Wiley.
- Donoho, D. L., & Huber, P. J. (1983). The notion of breakdown point. In P. J. Bickel et al. (Eds.), *A Festschrift for Erich L. Lehmann* (pp. 157–184). Belmont, Wadsworth.
- Gantmacher, F. R. (1959). *The theory of matrices, Vols. 1 and 2*. New York, Chelsea.
- Gower, J. C. (1970). Statistical methods of comparing different multivariate analyses of the same data. In F.R. Hodson et al. (Eds.), *Mathematics in the archaeological and historical sciences* (pp. 138–149). Edinburgh, Edinburgh University Press.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33–51.
- Hallgrímsson, B., & Lieberman, D. E. (2008). Mouse models and the evolutionary developmental biology of the skull. *Integrative and Comparative Biology*, 48(3), 373–384. doi:10.1093/icb/icn076.
- Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaentologia Electronica* 4(1), art. 4, 9.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley.
- Hartwig, W., Rosenberger, A. L., Norconk, M. A., & Owl, M. Y. (2011). Relative brain size, gut size, and evolution in New World monkeys. *Anatomical Record (Hoboken)*, 294(12), 2207–2221.
- Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin London Mathematical Society*,. doi:10.1112/blms/16.2.81.
- Klingenberg, C. P. (2013). Visualizations in geometric morphometrics: How to read and how to make graphs showing shape changes. *Hystrix*, 24(1), 15–24.
- Lemey, M., Salemi, M., & Vandamme, A. M. (2009). *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. Cambridge: Cambridge University Press.
- Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*, 36, 235–247.
- Mitteroecker, P., Gunz, P., & Bookstein, F. L. (2005). Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. *Evolution and Development*, 7(3), 244–258.
- Perez, S. I., Bernal, V., & Gonzalez, P. N. (2006). Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy*, 208(6), 769–784.
- Perez, S. I., Klaczko, J., Rocatti, G., & dos Reis, S. F. (2011). Patterns of cranial shape diversification during the phylogenetic branching process of New World monkeys (Primates: Platyrrhini). *Journal of Evolutionary Biology*, 24(8), 1826–1835.
- Richtsmeier, J. T., DeLeon, V. B., & Lele, S. R. (2002). The promise of geometric morphometrics. *American Journal of Physical Anthropology, Supplement* 35, 63–91.
- Rocha, M. A. (1995). Les collections ostéologiques humaines identifiées du Musée Anthropologique de l'Université de Coimbra. *Antropologia Portuguesa*, 13, 7–38.
- Rohlf, F. J. (1990). Rotational fit (Procrustes) methods. In F.J. Rohlf et al., (Eds.) *Proceedings Michigan morphometrics workshop* (pp. 227–236). Special publication no. 2, Museum of Zoology, Michigan, University of Michigan.
- Rohlf, F. J., & Marcus, L. F. (1993). A revolution in morphometrics. *Tree*, 8, 129–132.
- Rohlf, F. J., & Slice, D. E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1), 40–59.
- Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika*, 69(1), 242–244.
- Siegel, A. F., & Benson, R. H. (1982). A robust comparison of biological shapes. *Biometrics*, 38, 341–350.
- Slice, D. E. (1996). Three-dimensional generalized resistant fitting and the comparison of least-squares and resistant fit residuals. In L. F. Marcus et al., (Eds.), *Advances in morphometrics* (pp. 179–199). New York, Plenum Press.
- Slice, D. E. (2001). Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. *Systematic Biology*, 50(1), 141–149.
- Sperber, G. H. (2001). *Craniofacial development*. Hamilton: BC Decker Inc.
- Taguchi, Y. H., & Oono, Y. (2004). Novel non-metric MDS algorithm with confidence level test. <http://www.granular.com/MDS/src/paper.pdf>.
- Theobald, D. L., & Wuttke, D. S. (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 18521–18527.
- Van der Linde, K., & Houle, D. (2009). Inferring the nature of allometry from geometric data. *Evolutionary Biology*, 36, 311–322.
- Walker, J. A. (2000). Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology*, 49(4), 686–696.
- Weiszfeld, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal*, 43, 355–386.
- Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric morphometric for biologists: A primer*. London: Academic Press.