Evolutionary Biology
https://doi.org/10.1007/s11692-026-09674-1

RESEARCH ARTICLE

Comparative Assessment of Landmark-Based and Surface-Based Shape Quantification and Analysis in Geometric Morphometrics

Jordyn Eovito[1] · **Elizabeth Glynne**[1] · **Dean C. Adams**[1]

Received: 21 March 2026 / Accepted: 8 July 2026
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2026

Abstract

Evolutionary biologists commonly use landmark-based geometric morphometrics to characterize and compare biological shapes. Advances in three-dimensional imaging hold potential to catapult morphometric studies, but quantifying shape from 3D surfaces efficiently remains challenging. Several surface-only morphometric procedures have been proposed, which align surfaces to one another and use their differences as the basis of shape quantification and analysis. Comparisons of landmark-based and surface-only methods for specific datasets have revealed mixed results, but these studies utilized empirical datasets where the true patterns of shape differences are unknown. In this study we performed a simulation experiment where the true shape differences between objects were known, and evaluated the performance of two landmark-based and four surface-only methodologies under these conditions. Several procedures, including landmarks plus semilandmarks, as well as the surface-only method DAA (and to a lesser extent auto3d), most consistently recovered the known patterns of shape variation. Results from other surface-only procedures (GPSA and pseudoLM) were more variable. Notably, patterns in more complex shapes were better characterized than patterns in simpler shapes, implying that automated surface-only methods were challenged when presented with shapes whose principal axes were similar, or where structures contained few localized shape differences for these methods to properly orient the surfaces. Overall, our findings highlight that while biologically informed landmarking remains imperative as a foundation for morphometric analyses, some surface-only procedures do offer a powerful and efficient means of characterizing shape, and hold promise for morphometric studies.

Keywords Geometric morphometrics · Landmark analysis · Surface-only morphometrics · Shape analysis · 3D morphology

Introduction

The quantitative analysis of biological structures is commonly accomplished using landmark-based geometric morphometric (GM) methods (Adams et al., 2013; Mitteroecker and Schaefer 2022). These approaches leverage the locations of corresponding (homologous) landmark points as a means to characterize shape, after the effects of non-shape information due to position, orientation, and scaling differences have been mathematically held constant. In regions where fewer homologous landmarks can be identified, semilandmarks on curves and surfaces (Bookstein, 1997; Gunz et al., 2005; Gunz & Mitteroecker, 2013) may also be included to describe variation of those anatomical regions, which are frequently used in conjunction with homologous landmark points in shape analysis. In most GM applications, researchers manually place landmarks at equivalent locations across specimens, thereby generating their topological correspondence. Because placing landmarks in the correct location requires expert anatomical knowledge, landmark-based GM approaches are often considered the 'gold standard' of shape analysis (sensu Mulqueeney et al. 2025; Rolfe et al. 2021a), as such knowledge provides the link between quantifiable data and our biological notions of matched homology (see Oxnard & O'Higgins, 2009; Shui et al., 2023).

Over the past decade, increased availability of 3D imaging tools such as computed tomography (CT), X-rays,

✉ Dean C. Adams
dcadams@iastate.edu

[1] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

Published online: 23 July 2026

photogrammetry, and surface scanners, have enabled researchers to generate large databases of surface scans representing tens of thousands of specimens (Boyer et al., 2015; Goswami, 2015). This has spawned interest in 'high density morphometrics' (sensu Goswami et al., 2019) where variation in organismal form is described by hundreds to thousands of landmarks and semilandmarks obtained from hundreds of surface scans (for an early example see Polly, 2008), which may then be subjected to statistical evaluation. However, while technological advances have greatly facilitated our ability to obtain large amounts of primary phenotypic information in the form of surface scans, quantifying landmark data from those scans has become a significant bottleneck. To this end, numerous semiautomated landmarking tools have been developed that take a template of landmarks and semilandmarks obtained from a reference specimen and transfer them to another (target) specimen to identify the corresponding locations of those points (Bardua et al. 2019; Boyer et al. 2015; Gunz et al. 2005; Porto et al. 2021; Rolfe et al. 2021a; Schlager 2017; for earlier approaches using landmarks from photographs see: Rohlf, 2006). These semiautomatic approaches reduce digitizing errors and increase the efficiency of landmark data acquisition, making it possible to obtain GM data on large numbers of specimens.

One potential challenge with landmark-based GM is the fact that some structures have few landmarks that can be reliably identified (Polly, 2008; Shui et al., 2023). Additionally, because shape is described solely from the landmark locations, variation in other anatomical regions remain uncharacterized (Pomidor et al., 2016; Toussaint et al., 2021). For this reason, researchers have devised procedures that utilize the entire 3D surface scan as the basis of shape quantification (Boyer et al. 2015; Polly 2008; Pomidor et al. 2016; Rolfe et al. 2021b; Toussaint et al. 2021). These approaches (hereafter termed 'surface-only' methods) first align the surface scans to one another in some way (discussed below) and then compare a reference (template) scan to all remaining (target) scans to describe the differences between them. For instance, some surface-only procedures define a set of pseudolandmarks on the template surface and find their corresponding locations on the target surfaces (Polly 2008; Boyer et al. 2015; Rolfe et al. 2021b). These pseudolandmarks are then used in GM-based shape analyses using standard analytics. Other approaches use the entire set of vertices describing each surface scan and compare these across surfaces to quantify an overall measure of shape difference (Pomidor et al., 2016) or to obtain a set of variables that describe the deformation of the template surface to each target surface (Toussaint et al., 2021).

An important question when utilizing surface-only procedures is how well they characterize patterns of shape variation as compared with landmark-based methods. Some empirical comparisons between the two have found a high correspondence in the patterns of shape variation captured by both quantification schemes (e.g., Boyer et al., 2015), implying that some surface-only procedures may serve as a reasonable, and more time-efficient, means of characterizing shape. By contrast, other empirical comparisons found weaker associations between patterns from surface-only and landmark-based data (e.g., Pomidor et al., 2016; Toussaint et al., 2021; Mulqueeney et al., 2025) suggesting that some aspects of shape variation captured by one quantification scheme were not represented in the shape variation captured by the other procedure. One possible explanation for this discrepancy is that surface-only methods can be more sensitive to anomalies and surface noise (Boyer et al., 2015; Pomidor et al., 2016), which would be captured in the surface-only data but would have a lesser impact on the landmark dataset. Alternatively, it has also been posited that surface-only procedures can capture patterns of shape variation more completely than do landmarks and semilandmarks, as landmark datasets often have many fewer points of correspondence between objects (Toussaint et al., 2021). Unfortunately, because prior comparisons of landmark-based and surface-only procedures utilized empirical datasets, verifying the extent to which surface-only approaches capture greater – or different – aspects of shape variation as compared with landmarks was not possible, because the actual patterns of shape differences were unknown. Thus, distinguishing between these alternatives based on analyses of empirical data is compromised. By contrast, when using simulations, the degree of shape differences between 3D surface scans is known, and one can therefore effectively compare patterns of shape variation obtained from alternative procedures to determine the extent to which each method captures the known patterns of variation in the data.

In this study we performed a simulation experiment in which we deformed several original reference surfaces by known amounts to obtain samples of surface scans from each; thereby generating samples of surfaces where the true shape differences among objects were known. For each sample of surfaces, we then obtained estimates of shape variation using both landmark-based and surface-only GM procedures. We quantified the shape differences among objects estimated from each approach and compared these to the known input patterns of shape differences generated through simulation. Using this experimental design, we could thus determine the extent to which each digitization procedure characterized the actual (known) differences among objects, and how and why they differed.

Materials and Methods

Simulation Protocol

Initial Objects

We used five initial objects (templates) for our simulations, which were selected to display a range of structural complexity, spanning from geometric forms to anatomically detailed structures with multiple distinct regions (Fig. 1). Our initial objects were a cube, a human patella, a human humerus (right), a human scapula (left), and a mouse skull. Surface scans of skeletal elements were obtained from public repositories (for details see Supplemental Material). Holes in the initial scans were filled using the mesh repair tools in MeshInspector 2.5.2.208 (AMV Consulting LLC, 2026) and the watertight meshes were then down sampled in MeshLab 2023.12 (Cignoni et al., 2008) using quadratic edge collapse decimation to reduce the number of vertices and bring meshes to a comparable resolution, thereby facilitating the subsequent analyses. The surface scan of the cube was generated in R 4.5.1 (R Core Team, 2025) and subdivided in MeshLab, resulting in a scan containing ~ 12,000 faces.

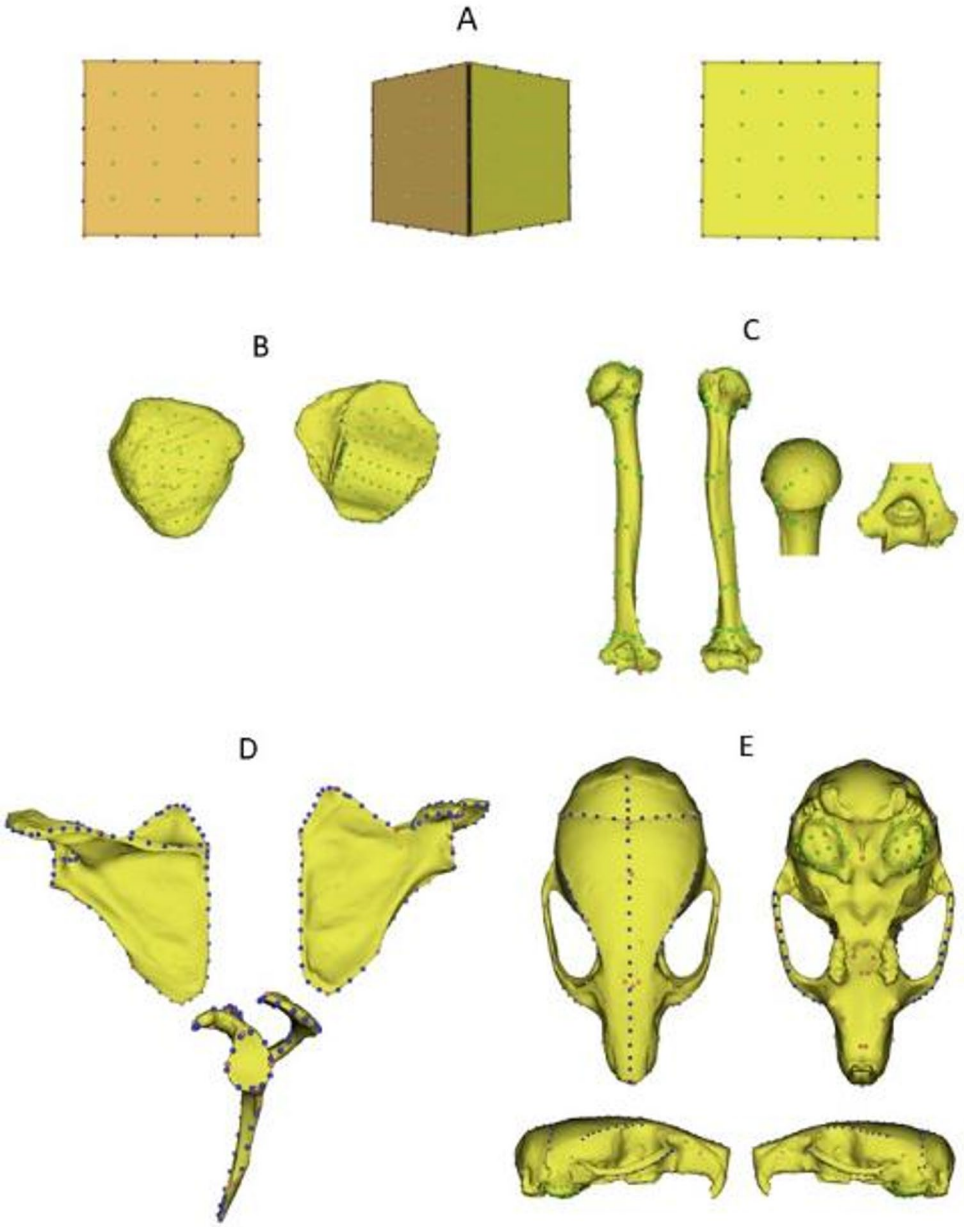


Fig. 1 Initial surface scans of five objects used in this study, represented in order of increasing shape complexity, spanning from geometric forms (a cube) to anatomically detailed structures with multiple distinct regions (a skull). The initial objects are as follows: **A** cube, **B** human patella, **C** human humerus (right), **D** human scapula (left), and **E** mouse skull. The set of landmarks (orange), semilandmarks on curves (blue), and semilandmarks on surfaces (green) used in landmark-based GM shape quantification are shown

Surface Deformation and Sample Generation

For each of our initial reference objects we generated a sample of surface scans with known differences by deforming the reference scan by known amounts along its *x*, *y*, and *z* axes, as well as combinations of those axes. This generated samples of surfaces that differed from the reference by affine transformations of varying degrees. Because affine transformations represent relatively simple shape transformations, the simulations here correspond to a 'best case' scenario for evaluating the ability of each shape quantification procedure for characterizing known patterns of shape variation. Additionally, by transforming the reference object systematically in the different ordinal directions, our simulations enabled us to sample shapes across a broad range shape space using an objective criterion, which is more challenging to accomplish when using non-affine shape transformations.

To simulate surfaces that differed by known amounts we first generated a transformation matrix, **T**, whose rows contained the input parameters describing the degree of affine transformation to be applied to each reference scan. Each row of **T** contained three values, representing the degree of transformation in *x*, *y*, and *z* axes respectively. Parameter values varied systematically between 1.0 and 0.5 (in 0.1 increments) for each axis, as well as for pairs of axes, resulting in a matrix containing 31 rows (Fig. 2A). Next, the coordinates of the vertices of a reference scan (**X**) were assembled into a *p x 3* matrix, where the *p* rows represented the coordinates for each vertex. Then, **X** was transformed by the input parameters in the i^{th} row of **T** via elementwise multiplication (i.e., a Hadamard product) found as $\mathbf{X}_{tr} = 1_p \mathbf{T}_i \odot \mathbf{X}$, which generated a modified scan transformed by the desired amount (Fig. 2B). This procedure was repeated for each of the *i* rows in **T**, resulting in a sample of 31 objects for each initial scan (the original plus 30 transformed objects). Additionally, because transformation by **T** altered both the size and the shape of the surface mesh, we re-scaled each $\mathbf{X}_{tr}$ to its original size by multiplying $\mathbf{X}_{tr}$ by the ratio of the original centroid size relative to the centroid size (CS) of the transformed surface mesh: $\mathbf{X}_{new} = \mathbf{X}_{tr} * (CS_{orig}/CS_{tr})$. The deformed and rescaled surface meshes were then centered using the internal *geomorph* function geomorph:::center, resulting in a sample of surface meshes that were centered, scaled, and positioned similarly (the initial orientation was not altered during the transformation procedure, meaning they were oriented

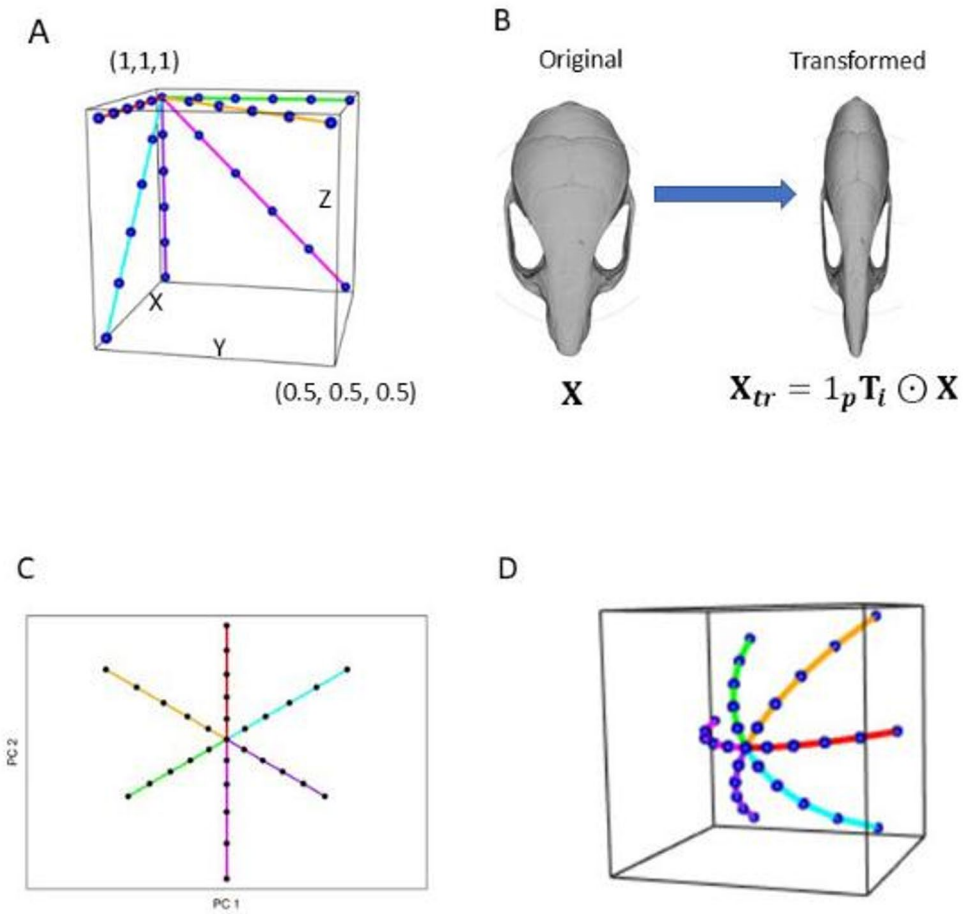


Fig. 2 Visual depiction of the simulation protocol. **A** The set of initial transformation parameters shown along the x, y, z axes, and combinations of those axes. **B** An example of deforming the mouse skull template along the x-axis. **C** Representation of the input shape differences for the simulated objects, once size-standardized, as found along the first two principal components of the deformation matrix, $\mathbf{X}_{tr}$. **D** The same input shape differences represented along the first three PC axes. In panels A, C, and D, colors designate increasing deformations along the same input axis as follows: red (x-axis), green (y-axis), purple (z-axis), yellow (x- and y-axes), blue (x- and z-axes), and magenta (y- and z-axes)

similarly). Finally, known input differences in shape due to the affine transformations were represented by a matrix of Euclidean distances between the rows of **T**, scaled by the size ratio so that only shape was represented (Fig. 2C). The above process was then repeated for each of the five initial reference objects, resulting in 5 datasets, each containing 31 surface scans that differed in shape by known amounts. All computations were performed in R 4.5.1 (R Core Team, 2025) and *geomorph* version 4.0.10 (Adams et al., 2025; Baken et al., 2021).

Shape Quantification

For each sample of surface scans, we quantified shape using six procedures: two landmark-based and four surface-only procedures. Acronyms designating each of the procedures are as follows: fixed landmarks (LM); landmarks and semilandmarks (LM+Semi); surface-only pseudolandmarks via auto3dgm (auto3d); surface-only deformation-based analysis (DAA); surface-only generalized Procrustes surface analysis (GPSA); surface-only pseudolandmarks via pseudolandmark generator (pseudoLM).

Landmark-Based: Fixed Landmarks (LM) Our first data type represented the shape of each surface scan using a set of homologous landmark points. Landmarking schemes largely followed previous studies, and the number of landmarks differed between objects (see Fig. 1). The following number of landmarks were used: cube, $p=8$; patella, $p=4$ (simplified from Peeters et al. 2020); humerus, $p=16$ (sensu Conaway et al., 2018); scapula, $p=15$ (sensu Green, 2013), and skull, $p=51$ (sensu Rolfe et al. 2021a). Landmarks were manually placed on the reference surface scan for each of the five datasets using the markups module in 3D Slicer 5.81 (Fedorov et al., 2012). Next, landmarks were transferred to the remaining surface scans in each sample using ALPACA (Porto et al., 2021) in SlicerMorph (Rolfe et al. 2021b). The initial landmark placements made by ALPACA were then manually adjusted to ensure topological and biological correspondence. Finally, all non-shape variation was removed via a Generalized Procrustes Analysis (GPA: Rohlf and Slice 1990), and the aligned configurations were projected orthogonally to tangent space to generate a set of tangent space coordinates (see Rohlf, 1999; Adams et al., 2013), which were treated as a set of shape variables for downstream analyses. These computations were accomplished using the gpagen function in *geomorph* (Adams et al., 2025).

Landmark-Based: Landmarks and Semilandmarks (LM+Semi) Our second data type represented the shape of each surface scan using both landmarks and semilandmarks. The number of curves and surface patches differed between objects (see Table S2), resulting in the following number of landmarks and semilandmarks for each dataset: cube, 8 landmarks and 144 semilandmarks; patella, 4 landmarks and 183 semilandmarks; humerus, 16 landmarks and 100 semilandmarks; scapula, 15 landmarks and 97 semilandmarks; and skull, 51 landmarks and 130 semilandmarks. Landmark and semilandmark schemes are shown in Fig. 1. For each object, the landmarks from dataset #1 (LM) were uploaded into Checkpoint 2025.07.09.09:08 (Stratovan Corporation, 2025) and matched to the reference surfaced scan. Next, a series of semilandmarks on curves and semilandmarks on surface patches were placed on the reference scan using the curve and patch functions respectively. The full set of landmarks and semilandmarks was then transferred to the remaining surface scans in each sample using the template function, and landmark and semilandmark placement was manually adjusted as needed. To remove non-shape variation a Generalized Procrustes Analysis (GPA: Rohlf and Slice, 1990) was performed using gpagen in *geomorph*, using the minimal Procrustes distance criterion for sliding semilandmarks, and the aligned configurations were projected orthogonally into tangent space. The resulting set of tangent space coordinates were then treated as a set of shape variables for downstream analyses.

Surface-Only: auto3dgm (auto3d) Our third data type represented the shape of each surface scan using pseudolandmarks obtained automatically using auto3dgm (Boyer et al., 2015). This automated approach first aligns the reference (template) surface scan to a target scan using their principal axes and an iterative closest point (ICP) algorithm (sensu Besl & McKay, 1992). ICP algorithms iteratively perform rigid transformations of two objects to identify the set of transformations that minimize the sum of squared distances between them. With auto3d, the template and target scans are first down-sampled to a smaller number of vertices, and their principal axes are identified. Next, the principal axes are aligned and ICP is performed (to minimize Procrustes distance). This is repeated for all eight combinations of principal axes until a global minimal Procrustes distance (and thus overall alignment) is obtained. Next, a sample of evenly distributed pseudolandmarks is identified on the template scan, and these points are then transferred to the target scan based on topological correspondence across aligned surfaces. This process is then repeated for all surface scans, resulting in a set of digitized specimens.

For our datasets we used the auto3dgm module in SlicerMorph (Rolfe et al. 2021b) to characterize shape with this approach. Initially we used 1000 pseudolandmarks to

quantify the surfaces. Shape variables were then obtained from these following a Procrustes superimposition and orthogonal projection to tangent space using gpagen in *geomorph* (see Boyer et al., 2015). Additionally, to determine the sensitivity of our findings based on the density of points, we repeated this procedure using both 100 and 300 pseudolandmarks for comparison.

Surface-Only: Deterministic Atlas Analysis (DAA) Our next data type represented the shape of each surface scan using deformations obtained from Deterministic Atlas Analysis (DAA: Toussaint et al., 2021). This semi-automatic approach uses large deformation diffeomorphic metric mapping (LDDMM: Durrleman et al., 2014) to describe the shape differences between a reference (template) scan and a target scan. Similar to thin-plate spline deformations, LDDMMs deform a reference object to match a target object, and the parameter values required to accomplish this transformation serve as a set of shape variables for further analysis.

With DAA the surface meshes are first aligned to one another; typically by utilizing a set of landmark points which are aligned via GPA (see Toussaint et al., 2021). Next, a mean (template) surface mesh is calculated, and from it an atlas is constructed, which describes the ways in which the template may be deformed into other surfaces. The precision at which the shape deformations will be described is defined by a kernel parameter (kernel width), where smaller values can be used to describe more fine-scale shape deformations, while larger values are used to describe broader-scale shape differences. Based on the kernel width, a set of reference points called 'control points' are placed on the template surface, with more points being distributed across the surface with smaller kernel widths. The control points serve as the locations for subsequent shape comparison, and are automatically placed on the template surface in such a manner that they are more concentrated in regions of higher variability (see Toussaint et al., 2021; Mulqueeney et al., 2025). Then, for each specimen, shape differences between the template surface and the target surface are encoded as directional deformation vectors at the control points (called momenta vectors), which represent the parameter values required to deform the template surface to match that of the target surface. These parameters are then treated as the basis of shape description, much like partial warp scores can be used as a set of shape variables based on the thin-plate spline: (Bookstein, 1991). Finally, because the momenta vectors describe deformations in a nonlinear space, they are subjected to a kernel PCA to obtain a set of shape variables in a linear tangent space for downstream analysis (for further details and a more technical description of DAA see: Durrleman et al., 2014; Toussaint et al., 2021; Mulqueeney et al., 2025).

For our datasets, surface scans were manually scaled and centered during the initial simulation procedure. Additionally, because their orientation was unchanged during transformation no further alignment was required (see above). Next, we obtained a set of control points for the reference scan of each dataset and deformed these onto each target surface using Deformetrica (Bône et al., 2018) to obtain the shape variables (momenta vectors). For each reference object, the kernel width was first determined to optimize the deformation for that particular structure (see Supplemental Material). Because the momenta vectors represented the shape of each object in a nonlinear space (Toussaint et al., 2021; Mulqueeney et al., 2025), kernel PCA (kPCA: Schölkopf et al., 1998; Schölkopf & Smola, 2001) was used to represent the relationships among observations in a linear approximation to this space. We used a Gaussian kPCA with a radius basis function, whose kernel width (σ) was found using an unsupervised spectrum-based approach. Specifically, we computed the eigenvalues of the centered kernel matrix across a range input values for σ, and selected the σ that maximized the ratio:

$$SR = \frac{\sum_{i=1}^{5} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

SR describes the proportion of variance described by the first *k* PC components relative to the total variance (5 in our case), and provides a heuristic describing how well the selected value of σ concentrates the signal in the kPCA. In essence, this approach identified the value of σ that maximized variation in the first *k* PCs (for discussion of similar criteria in related fields see: Ng et al., 2002; von Luxburg, 2007). This approach was implemented in R, using the kpca function in *kernlab* (Karatzoglou et al., 2004) and routines written by one of the authors (DCA).

Surface-Only: Generalized Procrustes Surface Analysis (GPSA) Our fifth data type quantified differences in the shape of surface scans using Generalized Procrustes Surface Analysis (GPSA: Pomidor et al., 2016). This automated approach first aligns the reference (template) scan to each target scan using an iterative closest point (ICP) algorithm. Specifically, vertices on the template scan are associated with their nearest neighbor on the target scan, and a set of rigid transformations are performed iteratively until the sum of squared distance between them is minimized. Importantly, because the two surfaces can have different numbers

of vertices, a modified Procrustes distance (the Procrustes Surface Metric) is used during the ICP procedure:

$$D_{PSM} = \sqrt{\frac{1}{2m_a}\sum_{i=1}^{m_a}(p_{A,i} - q_{B,i})^2 + \frac{1}{2m_b}\sum_{j=1}^{m_b}(p_{A,j} - q_{B,j})^2},$$

where m_a and m_b are the number of vertices in scan A and B respectively, and p and q represent the pair of closest vertices from each. D_{PSM} describes the distance between the vertices in the template scan (A) to the closest points in the target scan (B), and vice versa, weighted by the number of vertices in each (the two surfaces are not constrained to have the same number of vertices, see Pomidor et al., 2016). D_{PSM} is analogous to Procrustes distance, and thus represents the degree of shape difference between two surface scans.

For each of our datasets, we obtained D_{PSM} for each pair of objects using the software GPSA (Pomidor et al., 2016) and assembled a matrix of shape differences for all pairs of specimens. A principal coordinates analysis of D_{PSM} resulted in a set of shape variables for all downstream analyses. Finally, to evaluate the effect of differing numbers of vertices between surface scans, we randomly down sampled each scan by drawing a uniform random number between 0.5 and 0.95 and decimating the scans by that amount. All procedures above were then repeated on these datasets for comparison.

Surface-Only: Pseudlandmarks (pseudoLM) Our final data type quantified the shape of each surface scan using pseudolandmarks obtained in SlicerMorph (Rolfe et al. 2021b). With this automated approach, the reference (template) scan is aligned to each of the target scans; first using a global registration consensus algorithm, followed by a local alignment plane-to-point iterative closest point (ICP) algorithm (sensu Rusinkiewicz & Levoy, 2001). Specifically, the surfaces are first down-sampled, and an initial alignment is performed (see Porto et al., 2021). Next, ICP is used to identify the rigid transformations that minimize the sum of squares between surfaces and result in the alignment. Then, a series of pseudolandmarks regularly distributed across the mesh (but outside a user-specified threshold: see Rolfe et al. 2021b) are identified on the template scan, which are subsequently transferred to the target scan based on topological correspondence across aligned surfaces, and used to represent the shape of each object.

For our datasets, we used the PseudoLM module in SlicerMorph (Rolfe et al. 2021b) to identify pseudolandmarks on the reference scan. Next, the ALPACA procedure in SlicerMorph (Porto et al., 2021) was used to align all scans and to transfer the pseudolandmarks from the template scan to each target scan. Initially, we adjusted the spatial tolerance of the algorithm to use approximately 1000 pseudolandmarks to quantify the surfaces. Shape variables were then obtained from these following a Procrustes superimposition and orthogonal projection to tangent space using gpagen in *geomorph*. Additionally, to determine the sensitivity of our findings based on the density of points, we repeated this procedure using approximately 100 and 300 pseudolandmarks for comparison.

Statistical Analysis

For each of our five datasets, we estimated the shape of all surface scans using each of the six shape quantification procedures described above (two landmark-based and four surface-only methods). Then, patterns of shape variation obtained from each procedure were subjected to principal components analysis, and the patterns were visually compared relative to the true patterns of shape differences based on the matrix of input parameters. Additionally, because all shape quantification procedures generated shape variables represented in linear tangent spaces, we calculated the Euclidean distance matrices for the shape data from each procedure (for procedures whose data were represented as landmarks, semilandmarks, or pseudolandmarks subjected to GPA and orthogonal projection, this was mathematically equivalent to partial Procrustes distance: see Rohlf, 1999). We then performed a matrix correlation between the shape distance matrices and the Euclidean distance matrix of the input parameters. For each shape quantification procedure, the mean and standard deviation of the correlations across datasets was calculated. All statistical analyses were performed in R 4.5.1 (R Core Team, 2025).

Results

The true patterns of input shape differences are shown in Figs. 3, 4, 5, 6 and 7. For each of the true input patterns, the first two PC dimensions described shape differences that radiated outward in a starburst pattern from the initial surface scan along six main axes. Each of these axes corresponded with the known input deformations generated along a particular direction (e.g., increasing deformations along the *x*- direction, along the *y*- direction, etc.). When viewed in three dimensions, these deformation axes formed an umbrella pattern (Figs. S1 – S5), where the primary axes (PC1 and PC2) described shape differences between transformation directions (e.g., deformations in the *x*- direction versus the *y*- direction, etc.). Secondarily, PC3 described shape differences within transformation directions (e.g.,

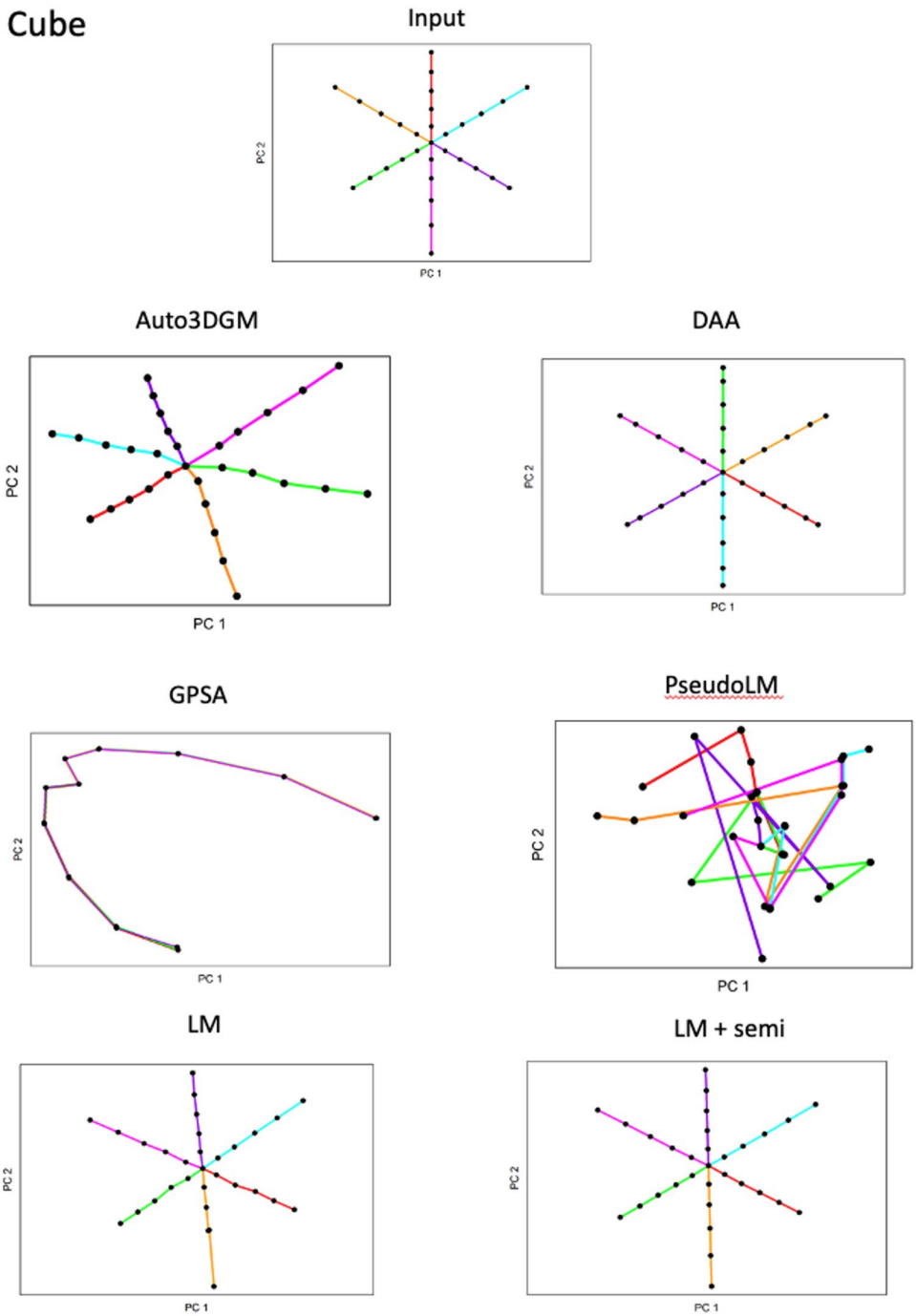


Fig. 3 PC plots of true (input) and estimated shape differences for shape data generated from the cube dataset. Each data type is labelled as per their landmark-based or surface-only procedure

differences between deforming the reference by 0.8 and 0.6 in the *x*- direction).

Next, we compared the true patterns of shape differences to those estimated by each of the six shape quantification procedures: (1) fixed landmarks (LM); (2) landmarks and semi-landmarks (LM+Semi); (3) surface-only pseudolandmarks via auto3dgm (auto3d); (4) surface-only deformation-based analysis (DAA); (5) surface-only generalized Procrustes surface analysis (GPSA); and (6) surface-only pseudolandmarks via pseudolandmark generator (pseudoLM). Correlations between the true shape differences and those estimated by each of the shape quantification procedures are found in Table 1. Based on these patterns, the most consistent methods in estimating the known patterns of shape variation across datasets were LM+Semi and DAA, followed by auto3d. These procedures displayed high correlations between estimated and known shape differences for all of the shapes examined. The remaining approaches (LM, PseudoLM, and GPSA) were more variable and generally displayed lower correlations with the known patterns.

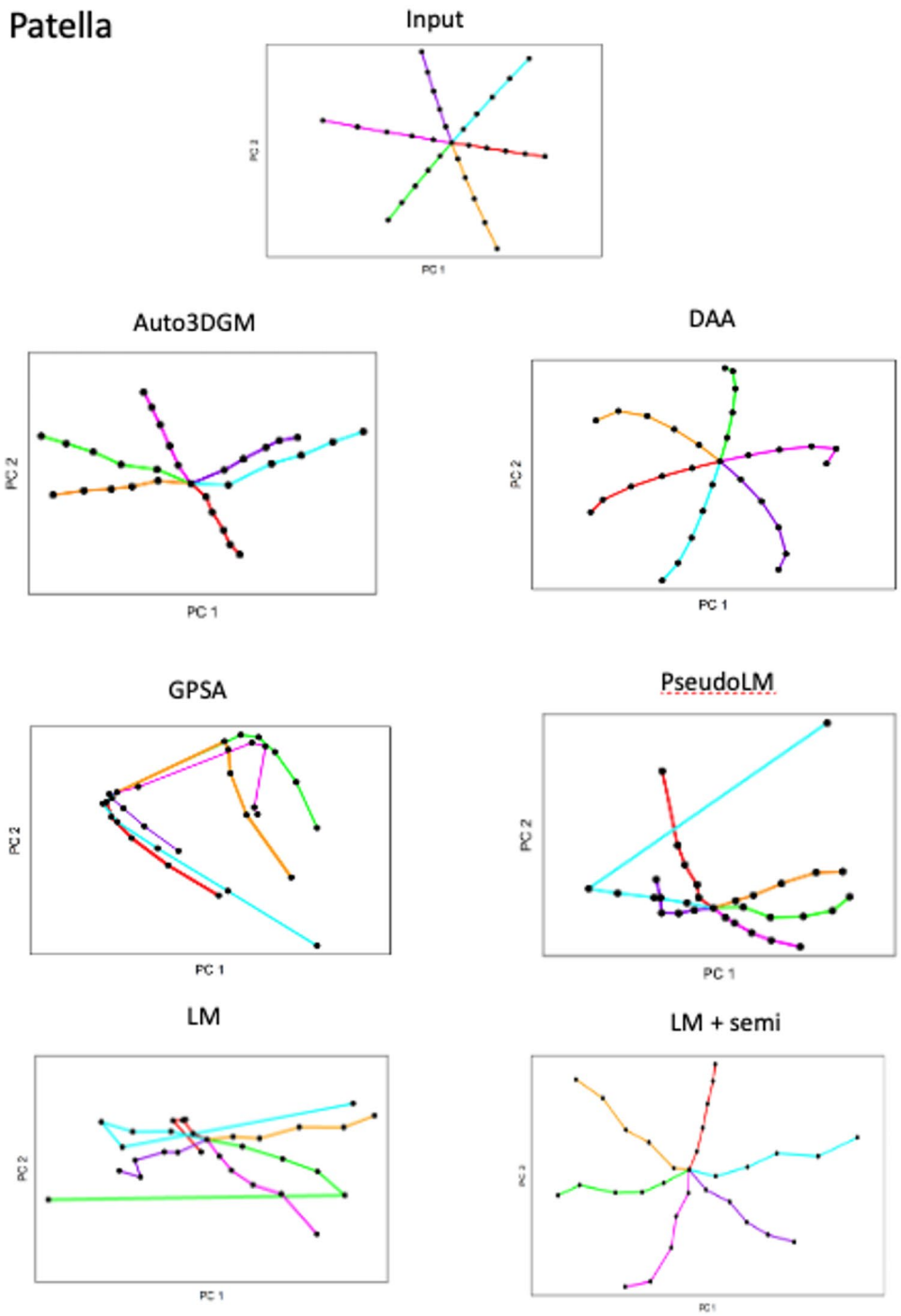


Fig. 4 PC plots of true (input) and estimated shape differences for shape data generated from the human patella dataset. Each data type is labelled as per their landmark-based or surface-only procedure

For instance, when using fixed landmarks only (LM), high correlations similar to those found using LM+Semi were exhibited for all datasets except the patella, which had a low correlation with known patterns, resulting in a reduced average correlation and higher variance across datasets. It should be noted however that the patella contained only 4 landmarks, which limited shape representation via LM for this dataset. Similarly, both GPSA and pseudoLM had difficulty representing shape variation from simple structures (cube and patella), but performed better on more complex skeletal elements, although their results remained less faithful than those of other procedures.

Examining patterns across datasets (Table 1), all shape quantification procedures performed well in capturing shape variation in more complex structures (e.g., mouse skulls), where correlations of estimated shape differences with the true patterns were quite high. By contrast, the fully automated surface-only procedures (auto3d, GPSA,

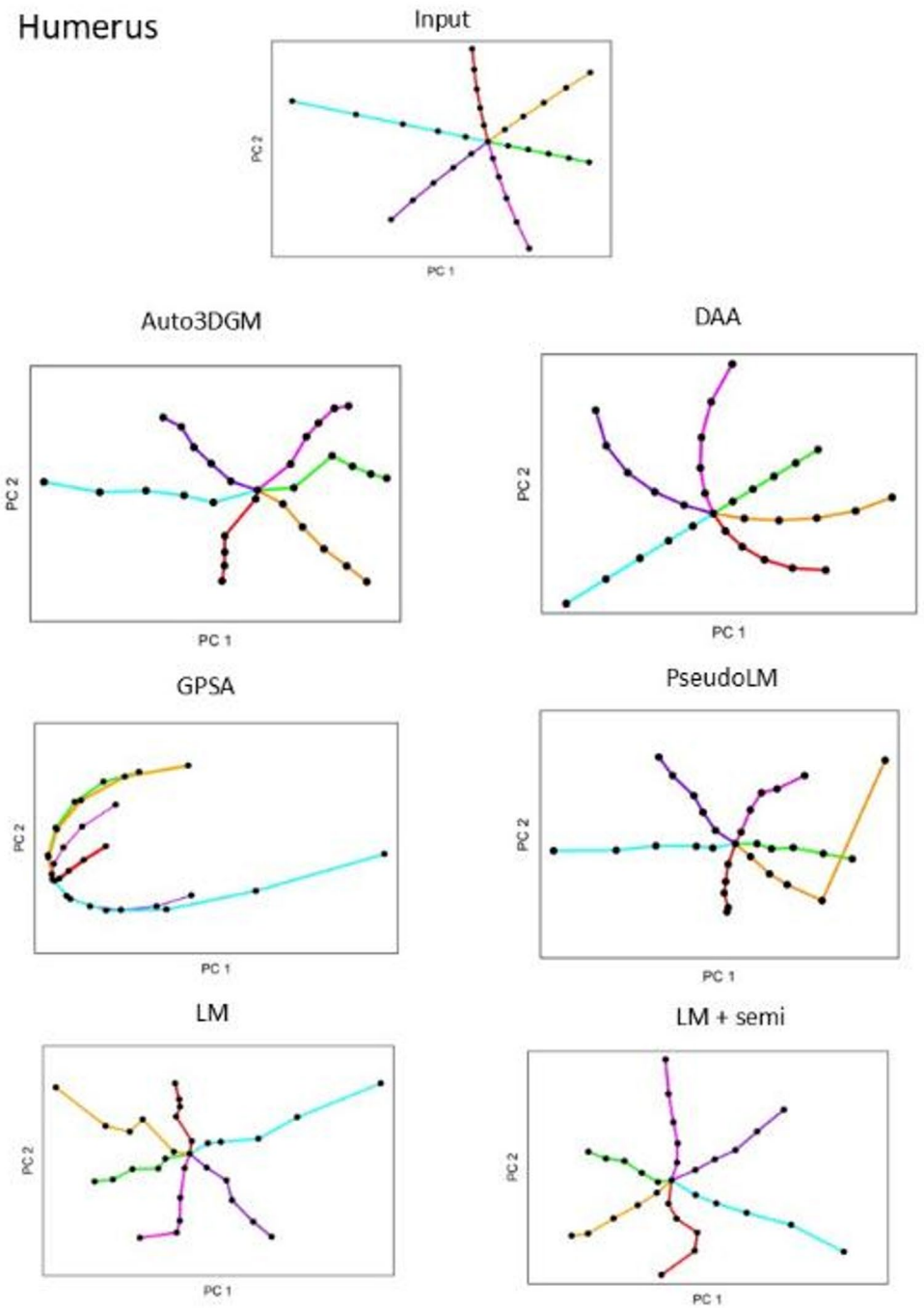


Fig. 5 PC plots of true (input) and estimated shape differences for shape data generated from the human humerus dataset. Each data type is labelled as per their landmark-based or surface-only procedure

and pseudoLM) performed the most poorly on the simplest structures (e.g., cubes and patella), suggesting that these procedures may be challenged with shapes whose primary axes are harder to distinguish from one another. Predictably, when using auto3d, increasing the number of pseudolandmarks also increased the correlation between the estimated and true patterns of shape variation (Table S3). The one exception to this trend was with the scapula dataset, where the correlation remained relatively stable across all pseudolandmark levels. By contrast, pseudoLM did not display an appreciable increase in performance with increasing numbers of pseudolandmarks (Table S3). Finally, for GPSA, evaluating surfaces containing differing numbers of vertices did not substantially affect the ability of the procedure to capture trends of shape differences, implying that the weighted Procrustes metric was robust to differences in surface mesh vertex number (Table S4).

When estimated patterns of shape variation were viewed along the first two principal components (Figs. 3, 4, 5, 6 and 7), they corresponded well with the trends described

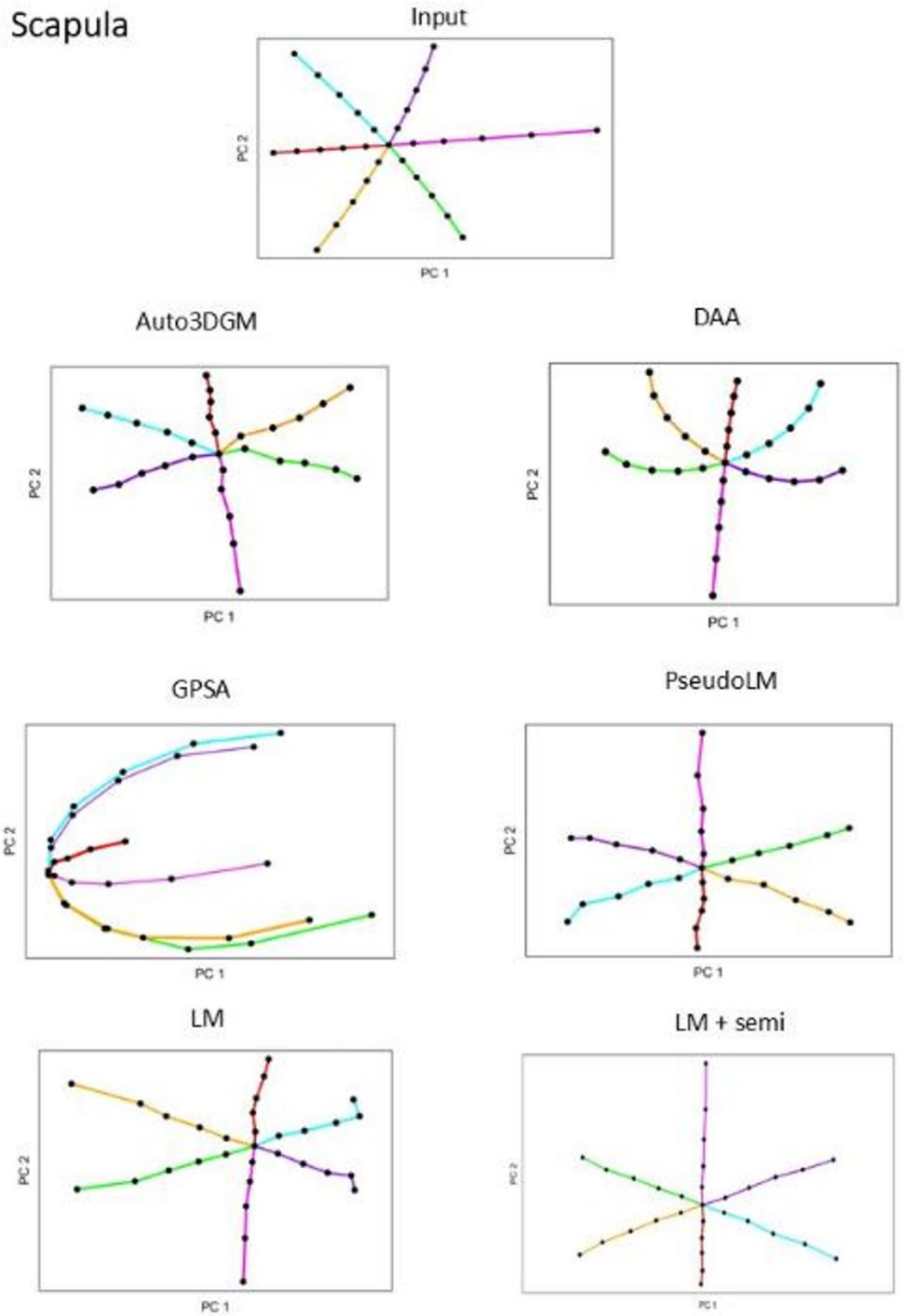


Fig. 6 PC plots of true (input) and estimated shape differences for shape data generated from the human scapula dataset. Each data type is labelled as per their landmark-based or surface-only procedure

via matrix correlation. Specifically, shape quantification procedures exhibiting high correlations with the true patterns tended to display starburst patterns that radiated outward from the initial surface scan, thereby most faithfully representing the true input patterns of shape difference. This was particularly the case for LM+Semi, DAA, and auto3d (Figs. 3, 4, 5, 6 and 7). Likewise, with the exception of the patella dataset, shape patterns obtained using fixed landmarks only (LM) also closely mirrored the starburst pattern in morphospace. For more complex structures (e.g. the scapula, the skull, and to a lesser extent the humerus), pseudoLM displayed starburst patterns in morphospace that generally mirrored the true patterns, indicating that this procedure captured trends of shape variation in these datasets. However, for simple shapes (e.g. the cube, patella), the true patterns were not well represented. Shape variation

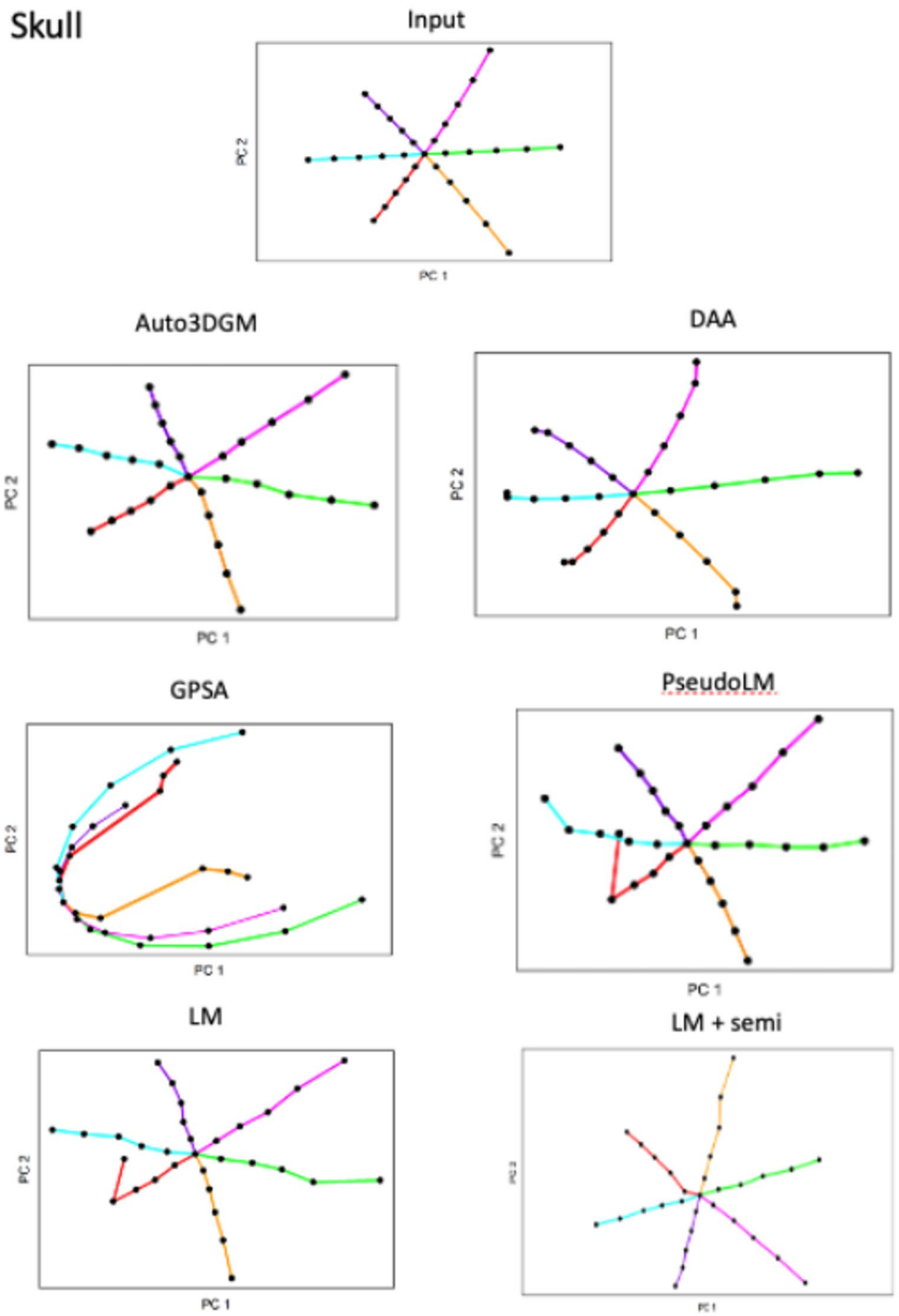


Fig. 7 PC plots of true (input) and estimated shape differences for shape data generated from the mouse skull dataset. Each data type is labelled as per their landmark-based or surface-only procedure

in the cube dataset was particularly poorly characterized with pseudoLM, with the different transformation axes overlapping widely (Fig. 3). This implied that input shape differences generated along differing transformation axes could not be distinguished from one another. Thus, with pseudoLM, deforming a cube to a rectangular prism in the *x*–direction or in the *y*–direction were considered to be the same shape.

Finally, using GPSA, patterns in PC1 and PC2 were notably curved for all datasets (Figs. 3, 4, 5, 6 and 7). Thus, the primary trends characterized by this procedure more closely aligned with differences within transformation directions, rather than the differences between them, in contrast with the input data. This implied that the estimated shape differences between deforming the reference scan by 0.8 and 0.6 units in the *x*-direction displayed greater perceived levels of

Table 1 Matrix correlations between the matrix of estimated shape differences from each data type versus the distance matrix of known input shape differences

DataSet	Landmarks	Landmarks + Semilandmarks	Auto3DGM	DAA	GPSA	PseudoLM
Cube	0.99	0.99	0.92	0.99	0.40	0.68
Patella	0.59	0.92	0.83	0.96	0.62	0.78
Humerus	0.96	0.96	0.96	0.94	0.88	0.89
Scapula	0.80	0.87	0.86	0.92	0.74	0.83
Skull	0.94	0.99	0.97	0.96	0.84	0.93
μ	0.86	0.95	0.91	0.95	0.70	0.82
σ	0.16	0.05	0.06	0.03	0.19	0.10

The mean and standard deviation across datasets is shown

shape difference than did deforming the reference by 0.6 in the *x*- versus 0.6 in the *y*-direction. Thus, GPSA exhibited a distortion of the true input patterns that accentuated smaller shape differences relative to larger shape differences. Additionally, for simpler structures (cube and patella), the different transformation axes overlapped widely. This was particularly evident in the cube dataset, where the different transformation axes overlapped into two distinct sets (Fig. 3; Fig S1): one containing the transformations along single axes (*x*-, *y*-, and *z*-), and another containing the transformations along pairs of axes (*xy*-, *xz*-, and *yz*-). Thus, GPSA could identify differences in transformations along a single axis versus along pairs of axes, but was incapable of distinguishing rectangular prisms deformed along different axes, and could not distinguish parallelepipeds deformed in different planes.

Discussion

Investigations of biological shape fundamentally lie at the intersection of geometry and homology, as the geometry of anatomical objects must be captured in a way that respects biological correspondence. Geometric morphometric (GM) methods accomplish this by using sets of landmarks — points grounded in topological equivalence and biological homology (Shui et al., 2023) — to render shape mathematically tractable, and to generate multivariate representations of shape suitable for hypothesis testing. Recent advances in 3D imaging have transformed this research paradigm, facilitating the rapid acquisition of high-resolution surface meshes and opening the door to "big data phenomics" (Goswami et al., 2019). However, as image acquisition has accelerated, the principal challenge has shifted from one of imaging to one of quantification (Frank et al., 2021). This necessity to characterize shape efficiently across large image databases has spurred the development of surface-only GM approaches, which treat entire surfaces as corresponding entities, rather than their particular anatomical components (see Boyer et al., 2015). Surface-only GM methods hold great potential to better automate the digitization process, thereby enabling researchers to obtain ever-larger samples of morphometric data. However, they also shift the gaze of GM studies away from explicitly defined homology and towards overall geometric similarity of entire surfaces. Thus, an important, but open, question is whether a purely geometric correspondence alone can recover biologically structured patterns of shape variation.

Prior empirical comparisons between landmark-based and surface-only approaches have yielded mixed results. Some studies found high associations of patterns obtained from the two (Boyer et al., 2015) while others found more disparate patterns (e.g., Pomidor et al., 2016; Toussaint et al., 2021; Mulqueeney et al., 2025). However, because the true patterns of shape differences are unknowable in empirical data, evaluating the efficacy of the different procedures, and attributing underlying causes to their differences, is challenging. To circumvent this, we conducted a simulation experiment in which the true shape differences between objects were known, and evaluated the performance of both landmark-based and surface-only methodologies under these conditions. This framework enabled us to identify not only which procedures performed well in capturing known patterns of shape variation, but also how and under what conditions they preserved — or distorted — the structure of those shape differences. Our findings thus clarify the strengths and limitations of alternative shape quantification approaches, and provide a principled basis to future empiricists for selecting among them.

First, among the procedures examined here, both LM+Semi and DAA, and to a lesser extent auto3d (particularly at higher pseudolandmark densities), most consistently recovered the known patterns of shape variation across datasets. Generally, these approaches exhibited high correlations with the true input shape differences, and produced morphospaces whose estimated patterns best matched the known input starburst pattern, in which deformation axes radiated outward from the reference along distinct directions (Figs. 3, 4, 5, 6 and 7). In particular, these methods capably distinguished shape differences between axes of

deformation (i.e., deformations along the *x*-axis versus along the *y*-axis), as well as shape differences found within axes (e.g., differing magnitudes of deformation in the same direction). The fact that landmarks plus semilandmarks most accurately represented the true patterns of shape differences underscores the value of biological homology in morphometrics, and the importance of utilizing anatomical correspondence guided by biological knowledge to inform shape characterization. Additionally, both DAA and auto3d displayed near-equivalent performance to LM+Semi for most datasets. In particular, DAA's diffeomorphic framework, with adaptive control points concentrated in regions of variability (see Mulqueeney et al., 2025), appears well suited to encoding structured shape differences. Likewise, the use of higher numbers of pseudolandmarks with auto3d appears to achieve a similar result. Taken together, these findings suggest that some surface-only GM procedures, as well as landmark-based GM methods, can accurately recover known patterns of shape variation in morphological data.

Second, our simulations revealed that quantifying shape with fixed landmarks alone was sufficient in many circumstances, with LM exhibiting nearly as high a performance as LM+Semi for most datasets. However, there was a notable reduction in performance with the patella dataset, where only four anatomical landmarks were available. This finding underscored that when few landmarks are available, the use of sparse landmark configurations can lead to an incomplete representation of shape (see discussion in Shui et al., 2023). Nonetheless, when landmarks were augmented with semilandmarks placed on anatomical curves and specific surfaces, we found that shape patterns were more faithfully represented; indicating that landmark-based GM approaches were capable of capturing known patterns of shape differences when present.

By contrast, GPSA displayed the poorest performance of the methods examined, exhibiting the lowest correlations between estimated and true patterns of shape variation. Further, the trends in morphospace estimated by GPSA were distinct and consistently curved (Figs. 3, 4, 5, 6 and 7), emphasizing differences within deformation directions rather than between them. This resulted in a distortion of the true structure of variation, effectively accentuating smaller magnitude changes while compressing larger directional differences. Additionally, for simple shapes, deformations along distinct axes of the original object resulted in patterns in morphospace that were completely overlapping, suggesting that GPSA was incapable of distinguishing deformations along different directions of those objects. This shape mis-alignment is likely due to the ICP algorithm employed by GPSA, which aligns surfaces by minimizing the summed squared distance between the point clouds of vertices, but without regard to discrete anatomical components on them. As such, GPSA would optimally align a rectangular prism distorted in the *x*-direction and a rectangular prism distorted in the *y*-direction by orienting their primary axes in the same direction, even though these are two different deformations of the original cube. We suggest that future versions of GPSA could be improved by incorporating a set of fixed landmarks that can be used in conjunction with ICP to provide a mesh alignment that takes landmark position into consideration, similar to procedures used in the DAA pipeline.

Scrutinizing the surface-only methods that utilize pseudolandmarks (auto3d and pseudoLM), there were some differences worth noting. With auto3d, increasing the number of pseudolandmarks improved the correlation between the estimated and true patterns of shape variation. This was in accord with expectation, as more pseudolandmarks have the potential to more comprehensively characterize the shape of the surface mesh (Boyer et al., 2015). However, a similar pattern was not observed with pseudoLM, as increasing the number of pseudolandmarks did not result in improved performance. An obvious question is, why might this be the case? Algorithmically, there are two major differences between the procedures: the mode of surface alignment and the way in which pseudolandmarks are obtained. With respect to alignment, auto3d uses both the principal axes of the surface scans and an ICP algorithm to find the combination of PC directions that maximizes correspondence between scans (Boyer et al., 2015). Contrastingly, pseudoLM utilizes a global registration algorithm followed by a local alignment via plane-to-point ICP (Rolfe et al. 2021b). With respect to pseudolandmarks, auto3d transfers a set of evenly distributed points from the reference to the template, while pseudoLM transfers a set of points that are more randomly distributed, but that are defined to fall beyond a minimum threshold distance to one another. We suggest that one or both of these algorithmic differences results in the performance discrepancies seen between the two procedures. Future research should investigate this possibility to identify potential algorithmic enhancements of pseudoLM.

One goal with surface-only GM procedures is to enhance efficiency of data acquisition, so that shape data may be obtained from large numbers of specimens. To that end, most surface-only GM procedures require little user-input. For instance, auto3d and pseudoLM require the user to specify the number of desired pseudolandmarks or the minimum threshold distance between points, while current implementations of GPSA are fully automated. By contrast, DAA requires many user-defined optimization parameters, perhaps the most important being the kernel width for optimizing the deformation of the reference shape (see Methods). However, finding the appropriate kernel width for a specific

dataset requires numerous rounds of trial and error, as each reference shape requires tuning the kernel width appropriately. Further, DAA has several other independent steps (including the pre-data preparation), making it a lengthy and challenging pipeline for inexperienced researchers or first-time users. In our work we used unsupervised learning to improve one step of the pipeline (kPCA: see Methods). We recommend that similar forms of automation for the selection of the kernel width parameter for a given reference scan would streamline analytics and make a more user-friendly pipeline.

Our study also revealed that the performance of the different approaches varied substantially as a function of structural complexity. Here, all methods performed reasonably well on anatomically complex objects, including the mouse skull, humerus, and scapula. By contrast, several surface-only methods (particularly GPSA and pseudoLM) struggled with simple geometries such as the cube and patella. One explanation for this result is that the more automated methods may fail to distinguish deformations occurring along distinct axes, as no landmarks are used to anchor the deformation relative to the overall shape. This would then result in substantial overlap of patterns in morphospace, as found here. By contrast, such conflation of pattern would not be expected with more complex anatomical structures, as the distinct features exhibited in different regions of the structure would effectively act as homologous points that stabilize surface alignments during registration, thereby enabling deformations in distinct directions to be identified. As a simple example, shape changes in the zygomatic arches of a mouse skull could easily be distinguished from expansions in the brain case or the rostrum. This finding also reveals that the relative performance of surface-only GM approaches – and their performance relative to landmark-based methods – may be context dependent and influenced by the complexity of the geometric structure of the objects under scrutiny.

The quantitative analysis of anatomical shapes lies at the heart of much of biological research. Landmark-based GM approaches have proven especially powerful in this regard, as they maintain the geometry of shape throughout the analysis via the use of anatomical points grounded in biologically-informed landmarking (Bookstein, 1991). Our results reaffirm the strength of this framework, demonstrating that methods incorporating explicit anatomical guidance – specifically using landmarks combined with semilandmarks –most consistently recovered known patterns of shape variation. Additionally, we found that some procedures that utilize entire surfaces as the basis of shape comparison (particularly DAA, and to a lesser extent auto3d) attained similar levels of performance. As such, these procedures complement and extend the standard GM toolkit for capturing patterns of high-dimensional shape variation. As these techniques continue to mature, we believe that they hold considerable promise for accelerating morphometric workflows while preserving the biological rigor that has defined the field. Thus, the results of our study suggest that surface-only GM methods, when carefully implemented, hold substantial promise, and offer a powerful and efficient means of characterizing shape, particularly for large datasets, and for structures where few homologous landmarks can be reliably identified.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11692-026-09674-1.

Acknowledgements The authors thank R. Minjares-Grajeda, A. Toth, and R. Weaver for comments on earlier versions of this manuscript. Alessandro Urciuoli kindly assisted with the implementation of DAA.

Author Contributions DCA conceived of the project. DCA, JE, and EG performed simulations and analyses. DCA, JE, and EG wrote the main manuscript.JE prepared all figures and tables. All authors reviewed the manuscript.

Data Availability All primary data and R-scripts for this manuscript may be found on Github (https://github.com/deanadams/ShapeQuantificationComparison).

Declarations

Competing Interests The authors declare no competing interests.

References

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2013). A field comes of age: Geometric morphometrics in the 21st century. *Hystrix the Italian Journal of Mammalogy*, *24*(1), 7–14. https://doi.org/10.4404/hystrix-24.1-6283

Baken, E. K., Collyer, M. L., Kaliontzopoulou, A., & Adams, D. C. (2021). geomorph v4.0 and gmShiny: Enhanced analytics and a new graphical interface for a comprehensive morphometric experience. *Methods in Ecology and Evolution*, *12*(12), 2355–2363. https://doi.org/10.1111/2041-210X.13723

Bardua, C., Felice, R. N., Watanabe, A., Fabre, A. C., & Goswami, A. (2019). A Practical Guide to Sliding and Surface Semilandmarks in Morphometric Analyses. *Integrative Organismal Biology*, *1*(1), obz016. https://doi.org/10.1093/iob/obz016

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256. https://doi.org/10.1109/34.121791

Bône, A., Louis, M., Martin, B., & Durrleman, S. (2018). Deformetrica 4: An Open-Source Software for Statistical Shape Analysis. In M. Reuter, C. Wachinger, H. Lombaert, B. Paniagua, M. Lüthi, & B. Egger (Eds.), *Shape in Medical Imaging* (Vol. Vol. 11167, pp. 3–13). Springer International Publishing. https://doi.org/10.1007/978-3-030-04747-4_1

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511573064

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape.

Medical Image Analysis, *1*(3), 225–243. https://doi.org/10.1016/S1361-8415(97)85012-8

Boyer, D. M., Puente, J., Gladman, J. T., Glynn, C., Mukherjee, S., Yapuncich, G. S., & Daubechies, I. (2015). A New Fully Automated Approach for Aligning and Comparing Shapes. *The Anatomical Record*, *298*(1), 249–276. https://doi.org/10.1002/ar.23084

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., & Ranzuglia, G. (2008). MeshLab: an Open-Source Mesh Processing Tool. *Eurographics Italian Chapter Conference*. The Eurographics Association. https://doi.org/10.2312/LOCALCHAPTEREVENTS/ITALCHAP/ITALIANCHAPCONF2008/129-136

Conaway, M. A., Schroeder, L., & Von Cramon-Taubadel, N. (2018). Morphological integration of anatomical, developmental, and functional postcranial modules in the crab-eating macaque (*Macaca fascicularis*). *American Journal of Physical Anthropology*, *166*(3), 661–670. https://doi.org/10.1002/ajpa.23456

Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., & Trouvé, A. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *Neuroimage*, *101*, 35–49. https://doi.org/10.1016/j.neuroimage.2014.06.043

Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling* (pp. 145–152). Presented at the Third International Conference on 3-D Digital Imaging and Modeling, Quebec City, Que., Canada: IEEE Comput. Soc. https://doi.org/10.1109/IM.2001.924423

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

Frank, L. R., Rowe, T. B., Boyer, D. M., Witmer, L. M., & Galinsky, V. L. (2021). Unveiling the third dimension in morphometry with automated quantitative volumetric computations. *Scientific Reports*, *11*(1), 14438. https://doi.org/10.1038/s41598-021-93490-4

Adams, D. C., Collyer, M., Kaliontzopoulou, A., & Baken, E. (2025). *Geomorph: Software for geometric morphometric analyses. (Version 4.0.10)*. R package.

Goswami, A., Watanabe, A., Felice, R. N., Bardua, C., Fabre, A. C., & Polly, P. D. (2019). High-Density Morphometric Analysis of Shape and Integration: The Good, the Bad, and the Not-Really-a-Problem. *Integrative and Comparative Biology*, *59*(3), 669–683. https://doi.org/10.1093/icb/icz120

Green, D. J. (2013). Ontogeny of the hominoid scapula: The influence of locomotion on morphology. *American Journal of Physical Anthropology*, *152*(2), 239–260. https://doi.org/10.1002/ajpa.22353

Gunz, P., & Mitteroecker, P. (2013). Semilandmarks: A Method For Quanitfying Curves and Surfaces. *Hystrix the Italian Journal of Mammalogy*, *24*(1). https://doi.org/10.4404/hystrix-24.1-6292

Gunz, P., Mitteroecker, P., & Bookstein, F. L. (2005). Semilandmarks in Three Dimensions. In D. E. Slice (Ed.), *Modern Morphometrics in Physical Anthropology* (pp. 73–98). Kluwer Academic Publishers-Plenum. https://doi.org/10.1007/0-387-27614-9_3

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab - an S4 package for kernel methods in R. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v011.i09

Mitteroecker, P., & Schäfer, K. (2022). Thirty years of geometric morphometrics: Achievements, challenges, and the ongoing quest for biological meaningfulness. *American Journal of Biological Anthropology*, *178*(S74), 181–210. https://doi.org/10.1002/ajpa.24531

Mulqueeney, J. M., Ezard, T. H. G., & Goswami, A. (2025). Assessing the application of landmark-free morphometrics to macroevolutionary analyses. *BMC Ecology and Evolution*, *25*(1), 38. https://doi.org/10.1186/s12862-025-02377-9

Ng, A. Y., Michael, I. J., & Weiss, Y. (2002). On Spectral Clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems (NIPS)*, *14*.

Oxnard, C., & O'Higgins, P. (2009). Biology Clearly Needs Morphometrics. *Does Morphometrics Need Biology? Biological Theory*, *4*(1), 84–97. https://doi.org/10.1162/biot.2009.4.1.84

Peeters, W., Van Haver, A., Van den Wyngaert, S., et al. (2020). A landmark-based 3D analysis reveals a narrower tibial plateau and patella in trochleardysplastic knees. *Knee Surgery, Sports Traumatology, Arthroscopy*, *28*, 2224–2232. https://doi.org/10.1007/s00167-019-05802-x

Polly, P. D. (2008). Adaptive Zones and the Pinniped Ankle: A Three-Dimensional Quantitative Analysis of Carnivoran Tarsal Evolution. In E. J. Sargis, & M. Dagosto (Eds.), *Mammalian Evolutionary Morphology* (pp. 167–196). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6997-0_9

Pomidor, B. J., Makedonska, J., & Slice, D. E. (2016). A Landmark-Free Method for Three-Dimensional Shape Analysis. *PLOS ONE*, *11*(3), e0150368. https://doi.org/10.1371/journal.pone.0150368

Porto, A., Rolfe, S., & Maga, A. M. (2021). ALPACA: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. *Methods in Ecology and Evolution*, *12*(11), 2129–2144. https://doi.org/10.1111/2041-210X.13689

Rohlf, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification*, *16*, 197–223. https://doi.org/10.1007/s003579900054

Rohlf, F. J., & Slice, D. E. (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Biology*, *39*, 40–59. https://doi.org/10.2307/2992207

Rohlf, F. J. (2006). *tpsDig. Digitize Landmarks and Outlines, Version 2.05. Stony Brook*. Department of Ecology and Evolution, State University of New York.

Rolfe, S., Davis, C., & Maga, A. M. (2021a). Comparing semi-landmarking approaches for analyzing three-dimensional cranial morphology. *American Journal of Physical Anthropology*, *175*(1), 227–237. https://doi.org/10.1002/ajpa.24214

Rolfe, S., Pieper, S., Porto, A., Diamond, K., Winchester, J., Shan, S., et al. (2021b). SlicerMorph: An open and extensible platform to retrieve, visualize and analyse 3D morphology. *Methods in Ecology and Evolution*, *12*(10), 1816–1825. https://doi.org/10.1111/2041-210X.13669

Schlager, S. (2017). Morpho and Rvcg – Shape Analysis in R. In *Statistical Shape and Deformation Analysis* (pp. 217–256). Elsevier. https://doi.org/10.1016/B978-0-12-810493-4.00011-0

Schölkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press. https://doi.org/10.7551/mitpress/4175.001.0001

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, *10*(5), 1299–1319. https://doi.org/10.1162/089976698300017467

Shui, W., Profico, A., & O'Higgins, P. (2023). A Comparison of Semilandmarking Approaches in the Visualisation of Shape Differences. *Animals*, *13*(3), 1179. https://doi.org/10.3390/ani13030385

Toussaint, N., Redhead, Y., Vidal-García, M., Vercio, L., Liu, L., Fisher, W., E. M. C., et al. (2021). A landmark-free morphometrics pipeline for high-resolution phenotyping: Application to a mouse model of Down syndrome. *Development*, *148*(18), dev188631. https://doi.org/10.1242/dev.188631

AMV Consulting LLC (2026). MeshInspector (v. 2.5.7). Bothell, WA. https://meshinspector.com/

Goswami, A. (2015). Phenome10K: a free online repository for 3-D scans of biological and palaeontological specimens. https://www.phenome10k.org

R Core Team (2025). R: A Language and Environmentfor Statistical Computing. Vienna, Austria. https://www.R-project.org/

Stratovan Corporation (2025). Stratovan Checkpoint. Stratovan Corporation. https://www.stratovan.com/products/checkpoint

von Luxburg, U. (2007). A Tutorial on Spectral Clustering. https://doi.org/10.48550/ARXIV.0711.0189

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.