Tina Memo No. 2012-001
Submitted to the Journal of Evolutionary Biology

Morphometric Shape Analysis with Measurement Covariance Estimates.

Hossein Ragheb, Neil.A.Thacker, Paul Bromiley, Anja Schunke, Diethard Tautz.

Last updated
17 / 4 / 2012

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Abstract

We propose a shape analysis system based upon the description of landmarks with measurement covariance, which extends statistical linear modelling processes to 'pseudo landmarks' for scientific studies. Our formulation is based upon a self consistent approach to the construction of Likelihood based parameter estimation and includes corrections for parameter bias, induced by the degrees of freedom within the linear model. The method has been implemented and tested on measurements from 2D fly wing, 2D mouse mandible and 3D mouse skull data. We use these data to explore possible advantages and disadvantages over the use of standard Procrustes/PCA analysis via a combination of Monte-Carlo studies and quantitative statistical tests. In the process we show how appropriate weighting provides not only greater stability but also more efficient use of the available landmark data [1].

1 Introduction

The understanding of genetic influences has long been associated with morphometric description (shape variation) of species. Though only in the last few decades has it become possible to link quantitative descriptions of shape with associated genetic factors. This process generally involves the construction of a parametric model based upon exemplar biological shape specimens, and the most popular of these are linear models. These are used to quantify and predict the correlations in shape variation between and within species. The objectives of this paper are to improve the statistical efficiency of analysis techniques used in the genetic interpretation of shape variation (morphometrics) and to broaden the scope of problems which can be tackled with shape analysis tools. In particular we believe that much shape data is not suitable for use in current approaches, and 'pseudo-landmarks' (those poorly localised in one direction and the majority of measurements for smooth 3D shape) cannot be appropriately utilised [25, 35].

Over a decade during the 70's, bio-mathematical and biometrical aspects of biological shape studies were treated separately. This early work was later criticised during the 80's by Bookstein [4], Goodall [15] and Kendall [20]. Later, Bookstein [5] worked towards converging notations from Goodall, Kendall and himself, for the biometric analysis of landmark data in a bio-mathematically interpretable framework of shape. As a consequence of these efforts, the standard method for analysis of variation in landmark position is generally regarded as 'Procrustes'. It comprises a least-squares alignment of a set of landmark features to a mean shape, and this is often followed by eigen-vector analysis of the linear correlations in variation around that mean. While the technique is now very popular the approach has several limitations with regard to the types of variation with which it can deal. One of these limitations is due to the assumption associated with taking least-squares differences and eigen-vector summaries of distributions. Though many regard these as simply definitional, and in particular associated with 'shape', any statistical interpretation suggests that data are measures with homogeneous noise. Traditionally, the Mantel test [23], [27] has been used as an alternative to Procrustes distance to compute correlation between distance matrices (usually symmetric). Though many papers have been published in this area, we are aware of no work in this, or any related, area of point distribution modelling that has provided a framework to allow data to be analysed according to a measurement process.

Although landmarks are generally carefully chosen in order to allow accurate measurements of position within the image, problems will occur if 'pseudo-landmarks', measured from smooth curves and only accurately localised in one dimension, are input to the analysis. Landmarks with a high degree of variability can act as outliers in the alignment stage, generating correlated compensating shifts and rotations of the other points. As PCA aims to describe the main sources of variation, high levels of such correlated movement will then necessarily contaminate the extraction of eigen-vectors [19]. This contamination cannot be considered a genuine biological variation, as it has occurred purely due to the uncertainty in the measurement. This in turn follows from the subjective definition of the landmark leading to the view that problems can be avoided via appropriate definition. The mathematical concept of homology (and mapping) underlies many of the considerations behind much theoretical work that is described with the mathematical formalisms of iso-morphism. Because of such restrictions on the definition of landmarks, semi-landmarks were introduced [17] in order to allow inclusion of other points which are not homologous among the specimens. By this we mean that a unique corresponding location can not be defined. Measurement at these locations must be regularised by a constraint, such as bending energy [6, 17], in order to recover the information missing due to the nature of local structure.

[1] This work was performed in collaboration with the Max-Planck Institute for Evolutionary Biology, Plön, Germany.

From a statistical perspective a homology must be augmented by distributions indicative of the extent to which a correspondence can be established. The standard way to deal with inappropriate weighting of data in a least-squares fit is to generalise the least-squares cost to a Mahalanobis distance, computed using measurement covariances. By avoiding the requirement of specifying a unique homologous location, this has the advantage of accommodating varying precision in measured data without having to try to re-create missing data. There have been several attempts in the literature to include measurement errors for landmark points. Fitzpatrick et al. [12] worked on the relationship between localization error and registration error in rigid-body, point-based registration. Chui and Rangarajan [10] proposed a general framework for non-rigid point matching, where outliers are effectively rejected.

Text books [36] state that using weighted Procrustes does not lead to a Kendall's shape space. Claiming that "statistical analysis cannot employ parametric models", they suggested that resampling-based methods must be used instead. Another reason for rejecting the idea of a weighted Procrustes was said to be a "lack of clear criteria for determining appropriate weighting of semi-landmarks". These criticisms can only really be interpreted once a method for weighting is specified. Goodall [16] suggested a method which assumed a single covariance for all landmark perturbation. It has been noted that such a matrix is inestimable [22]. Goodall himself acknowledged that "as a model of measurement error this is a drawback, as the direction of greatest variation may vary considerably between landmarks". Despite this problem, later work [33] generalised this idea to a Bayesian framework. We believe that it makes sense instead to suggest an approach which can support the process of landmark location as measurement, with a covariance describing the localisation of each landmark *separately*. For example, Rohr et al. [31] used covariance matrices in a Mahalanobis distance form for non-isotropic data. This was, however, for the purpose of image registration using splines. The basic concept can be implemented via a standard technique used in pattern recognition, often refferred to as whitening. Recently, the technique has been applied to the within group biological covariances [24], but not to the process of noise on measurements. Here, we shall investigate possible generalisations of Procrustes along these lines, and the different ways such a measurement covariance may be estimated. As a key issue here is the computability of these covariances, the stability of the resulting analysis is an important question for investigation.

There has been an ongoing discussion in the biology literature regarding appropriate ways to deal with non-homologous landmarks (points defined on smooth curves and surfaces) during statistical analysis. For instance, Klingenberg [21] has objected to Polly's conclusions [29] regarding the benefits of existing homology-free approaches. He believes that these approaches all critically depend on some sense of homology since they are not really free of assumptions about the correspondence of parts. Oxnard and O'Higgins [26] have recommended that it is biology that has to inform morphometrics in planning the landmark configuration (mainly mathematical landmarks) in relation to the hypothesis available. The approach to dealing with semi-landmarks in the morphometric analysis of shape currently seems to be divided between two alternatives, both of which aim to adjust the position of these landmarks by optimising a specific metric, before constructing a linear model of variation about the mean. These metrics are bending energy (BE) and Procrustes distance (PD) respectively [25]. Arguments for and against these approaches are based upon specific examples in biology. Although evidence has been reported of utility [8], Slice [32] has stated that the application of the BE approach to biomedical and anthropological problems has been minimal. Vignon and Pierre [35], and Prez et al. [28] have shown concern regarding the observation that different methods for handling semi-landmarks could result in different conclusions in a discriminant analysis study. Gomez-Robles et al. [14] have examined the advantages and disadvantages of different novel methods in geometric morphometric analyses including homology-free approaches, landmark-based approaches, and combinations of both techniques.

Comparison between results from shape analysis and genetics is an important research topic in evolutionary biology. For instance, Frederich et al. [13] have attempted to estimate the statistical correlation between morphological, genetic and geographical distances. We offer an alternative shape analysis method that tackles the existing problem in the literature, so that any such comparison become statisticaly valid and informative.

2 Methods

Suppose that there are K shapes (2D) in our data-set and each shape vector $\mathbf{w}_k$ contains N landmark points, i.e. $\mathbf{w}_k = [w_{1x}, w_{1y}, w_{2x}, w_{2y}, ..., w_{Nx}, w_{Ny}]$. We then apply a scale s_k, a rotation R_k and a translation $\mathbf{t}_k$ to the original data to get an aligned version of the data called $\mathbf{z}_k$, where $\mathbf{z}_k = [\mathbf{z}_{k1}, \mathbf{z}_{k2}, ..., \mathbf{z}_{kN}]$ and $\mathbf{z}_{kn} = s_k R_k(\mathbf{w}_{kn} - \mathbf{t}_k)$.

The mathematical description of the model so far can accommodate any value of scale or orientation for the definition of mean model. We therefore define the orientation of mean shape so that the line between a specified pair of points is horizontal. This also has the benefit that initial estimates of alignment for sample k can be set according to the relative positions of these points. We also use the average distance between these same landmarks to rescale the mean shape at each iteration so that scale remains numerically defined.

For 2D data, we assume a common, fixed, 2x2 covariance for each landmark derived from the measurement process. These are composed into the matrix C. This is a tri-diagonal matrix, the diagonal line of which contains data for individual landmarks. Outside of the 2x2 covariances, the off diagonal elements of C are zero, i.e. there are no correlations between landmarks. The use of a fixed data covariance cancels out when taking the weighted mean, to regenerate the conventional formula for the mean $\mathbf{m} = (1/K)\sum_{k=1}^{K} \mathbf{z}_k$.

The points $\mathbf{z}_k$ do not have uniform independent noise distribution around them, which is one of the assumptions for the PCA to be applicable. However, this property can be obtained via a whitening transformation. Although transformation of data can be considered as a new space, it can also be interpreted as an affine re-projection. Ghost points are accordingly defined in the original coordinate system and, being scaled projections relative to the shape centroid, are an alternative way to summarise the original measurement. The process amplifies the spatial variation in directions which are well measured relative to those which are not so that the resulting locations have isotropic errors (as required). In turn, this allows accurately measured structure to be encoded in the most significant eigen-vectors of the linear model. We transform $\mathbf{z}_k$ to ghost points $\mathbf{g}_k$ using the matrix W so that $\mathbf{g}_k^T = W(\mathbf{z}_k - \mathbf{m})^T$.

By applying singular value decomposition to C^{-1}, i.e. $C^{-1} = U^T V U$, and making it equivalent to $W^T I W$, we find that the required whitening matrix is $W = V^{1/2}U$. Application of PCA to $\mathbf{g}_k$ follows for construction of the shape covariance, giving the eigen-vectors $\mathbf{e}_j$ and values μ_j for the whitened space of ghost points as those which minimise the unexplained variance for fixed $J < N$, where J is the number of eigen-vectors used in the model. Hence, for any specific shape example k, linear factors $\lambda_{jk} = \mathbf{e}_j.\mathbf{g}_k$ can be computed to best approximate $\mathbf{z}_k$ with the model $\mathbf{z}'_k$;

$$F = \sum_{k=1}^{K} \mathbf{g}_k^T \mathbf{g}_k \approx \sum_{j=1}^{J} \mu_j \mathbf{e}_j^T \mathbf{e}_j \quad , \quad \mathbf{z}'_k = \mathbf{m} + W^{-1} \sum_{j=1}^{J} \lambda_{jk} \mathbf{e}_j \tag{1}$$

A genuine Likelihood should be based upon the variation of the data around the assumed model. Failure to do this results in only partial registration and residual distributions which cannot be meaningfully interpreted [2]. Using this argument, if we wish to align to the mean shape we should use a covariance which is consistent with the distribution around the model. In order to find the best $R_k, \mathbf{t}_k, s_k$ parameters for each k, we minimise a Mahalanobis distance given by

$$\log(P_{kz'}) = (\mathbf{z}'_k - \mathbf{z}_k)^T C^{-1} (\mathbf{z}'_k - \mathbf{z}_k) \tag{2}$$

This is simply the Likelihood estimate for the location of the shape given the linear model and the assumed measurement covariances and can be interpretted directly as a χ^2 statistic. By replacing C with I and $\mathbf{z}'_k$ with $\mathbf{m}$ this reduces to the least-squares function for standard Procrustes. We can therefore interpret this as a generalisation of the standard approach. However, we do not wish to generalise further by using for example PPCA [34], as this requires an additional assumption of a Gaussian distribution over derived variables, which is generally invalidated in morphometric data sets.

We can also approach the construction of a Likelihood in an alternative manner [30], this is however beyond the scope of current paper. As use of Eq. (2) requires an initial estimate of the model and transformed data $\mathbf{z}_k$, it makes sense to start from the Procrustes result. To reach the best possible alignment we should iteratively estimate R_k, $\mathbf{t}_k$ and s_k using the assumed $\mathbf{e}_j$, $\mathbf{m}$, C^{-1} and W^{-1}. This gives us a new $\mathbf{z}_k$, and so a new $\mathbf{m}$ and F for construction of $\mathbf{e}_j$. For fixed covariances, convergence can be monitored via construction of the total Likelihood $\log(P) = \sum_{k=1}^{K} \log(P_k)$. One may use the final estimates of $\mathbf{z}_k$ and $\mathbf{z}'_k$ to construct the sample covariance;

$$C' = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{z}'_k - \mathbf{z}_k)(\mathbf{z}'_k - \mathbf{z}_k)^T \tag{3}$$

For a well defined Likelihood method this covariance should be consistent with the assumed distribution C. However, the use of free parameters during alignment and model construction introduce biases which must be addressed in an iterative analysis in order to avoid instabilities, which will now be described.

2.1 Covariance Correction

When attempting to estimate C, the use of free parameters during model fitting reduces the sample covariance obtained from residuals. A possible outcome of this is the over weighting of landmarks leading to a runaway convergence on one landmark, during iterative estimation. For a single scale parameter associated with an approximate

[2] Bookstein [7]: "Wherever there is partial registration the true value of a (vector deformation) is inaccessible."

linear vector $\mathbf{f}$ we can estimate the expected reduction in the covariance for each 2x2 landmark component of the matrix as

$$\Delta_n C \;=\; \frac{\mathbf{f_n}^T \otimes \mathbf{f_n}}{\mathbf{f}^T C^{-1} \mathbf{f}} \qquad , \qquad \Delta C \;=\; \sum_{n=1}^{N} \Delta_n C \tag{4}$$

Note that the denominator is the change in χ^2 expected due to a unit change in $\mathbf{f}$, and $\mathbf{f}_n \;=\; D_n \mathbf{f}$ is the 2D component of $\mathbf{f}$ corresponding to landmark n. For an eigen-vector $\mathbf{e}$ defined in the whitened ghost space, this would suggest a total correction of

$$\Delta C \;=\; \frac{W^{-1}\mathbf{e}^T \otimes W^{-1}\mathbf{e}}{W^{-1}\mathbf{e}^T C^{-1} W^{-1}\mathbf{e}} \;=\; W^{-1}\mathbf{e}^T \otimes W^{-1}\mathbf{e} \tag{5}$$

This is correct for the required tri-diagonal components, and must be added to the sample covariance. The known structure of the covariance can be enforced by zeroing relevant off-diagonal terms. The parameters of the linear model, including scale, rotation, translation and linear model weightings can also be written in this way, but as we now explain, we first need to make the linear model and alignment parameters orthogonal.

Strictly, Kendall's definition of shape explicitly removes aspects of object transformation before model construction. Though in practice estimated linear shape parameters can correlate with transformation parameters, thereby making joint estimation of shape and alignment parameters unstable. Here our aim is to stabilise this process by removing first order correlations from the data covariance F prior to model construction. If Θ_i represents one of the direction vectors of these parameters (with $2N$ elements for 2D data), it follows that direction vectors $\Theta_1 = [1, 0, 1, 0, ...]$ and $\Theta_2 = [0, 1, 0, 1, ...]$ corresponding to translation in x and y directions are orthogonal, i.e. $\Theta_1.\Theta_2 = 0$. Similarly, direction vectors $\Theta_3 = \mathbf{m} = [\mathbf{m}_{1x}, \mathbf{m}_{1y}, \mathbf{m}_{2x}, \mathbf{m}_{2y}, ...]$ and $\Theta_4 = [-\mathbf{m}_{1y}, \mathbf{m}_{1x}, -\mathbf{m}_{2y}, \mathbf{m}_{2x}, ...]$ corresponding to scaling and rotation are orthogonal, and so $\Theta_3.\Theta_4 = 0$, where $\mathbf{m}$ is the vector corresponding to the mean shape.

Hence, to orthogonalise the model, we modify ghost points as follows.

$$\mathbf{g}'_k = \mathbf{g}_k - (\mathbf{g}_k.\hat{Y}_i^T)\hat{Y}_i^T \qquad , \qquad \hat{Y}_i^T = W\hat{\Theta}_i^T \tag{6}$$

where the unit vector $\hat{\Theta}_i$ is the normalised form of Θ_i. The new $\mathbf{g}_k$ is computed iteratively using each $\hat{\Theta}_i$ so that any variation about the mean which could have been described by an alignment parameter is removed from the correlation matrix F prior to model construction.

The corresponding measurement covariance correction term is given by

$$\Delta C_\Theta = \sum_{i=1}^{4} \{(\hat{\Theta}_i^T C'^{-1} \hat{\Theta}_i)^{-1} (\hat{\Theta}_i^T \times \hat{\Theta}_i)\} \qquad , \qquad C = C' + \Delta C + \Delta C_\Theta \tag{7}$$

Using the above formula, the contribution to the χ^2 lost by using a scaling parameter associated with each vector Θ_j contributes a value of unity to the χ^2 for every additional independent degree of freedom,

$$\Delta_j \chi^2 \;=\; (\sqrt{\Theta_j^T C \Theta_j}\Theta_j^T) C^{-1} (\sqrt{\Theta_j^T C \Theta_j}\Theta_j) \;=\; \sqrt{\Theta_j^T C \Theta_j}(\Theta_j^T C^{-1} \Theta_j)\sqrt{\Theta_j^T C \Theta_j} \;=\; 1 \tag{8}$$

totaling $J + 4$. Our method for covariance correction is therefore consistent with a degree of freedom correction as described in conventional analysis approaches [3]. This is also equivalent to using expectation of the likelihood for estimation of the measurement covariance. As a consequence the estimation process can be considered equivalent to the Expectation-Maximisation (EM) algorithm, both in operation and parameter estimates, so that the conventional proof of convergence is applicable [11]. The same covariance estimates would also be obtained when using PPCA with an uninformative prior [34].

2.2 Extension from 2D to 3D

Here we outline the mechanism we use to extend 2D shape rotation analysis, and the extraction of corrected anisotropic measurement covariances, to 3D. The methods are demonstrated in the analysis of 3D mouse mandible data, both as a test of the theory/software implementation and as an illustration of use for the identification of outlier landmarks.

The extension to 3D data is mainly involved with the mechanism of representing and estimating 3D shape rotations. We define a fixed orientation co-ordinate system from a set of 3D data-points based upon a selection of three

landmark points. We then represent a rotation matrix in terms of three separate rotations about the co-ordinate axes. Finally we compute the linear vectors which approximate the first order shifts seen in the 3D points due to these rotations. These are then used in the linearised approximation for sample covariance correction, as described earlier. These extensions are enough to support a quantitative analysis of 3D landmark data, for the estimation of landmark accuracy and identification of outlier data. The mathematical model used is described in detail here and in Appendix A. We provide quantitative tests in the experiments section which demonstrate the numerical stability of the algorithms using Monte-Carlo data.

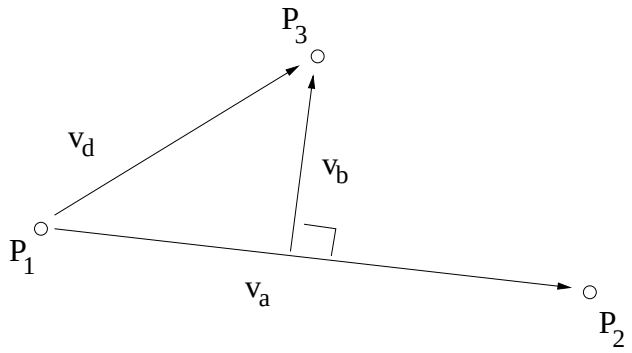

Figure 1: The geometry shows how to define a base-plane in 3D (consistent with the base-line in 2D) using three landmark points.

Rotation Matrix: Our first task is to define a co-ordinate system for a 3D data-set, from which we can define certain basic properties of orientation for the mean shape, and so that individual data samples can be approximately oriented prior to optimisation during linear model construction. In the 2D case this is done by simply defining the line between two landmark points in the mean model as horizontal. In 3D this situation is more complicated as we have the possibility of a rotation matrix which must be defined by a minimum of 3 parameters. In order to stay consistent with the 2D case we define 2 points to establish a horizontal, and then a third to define the vertical relative to the first two.

Given a 3D shape, we take three points p_1, p_2, p_3, with relatively large distance from each other (Fig. 1) to define the orientation plane for the shape. The rotation matrix R^T is hence found based on basic vector calculations (see Appendix A).

Roll, Pitch and Yaw Angles: Given the rotation matrix which brings a dataset into alignment with the preferred co-ordinate system it is now possible to represent the rotation as a sequence of rotations about three orthogonal axes. According to basic 3D rotation formulas, and using α, β, and γ as yaw, pitch, and roll respectively, the 3D rotation matrix is defined as three consecutive rotations around the z, y, and x coordinate axes.

$$R_{xyz} = R_x(\gamma)R_y(\beta)R_z(\alpha) \tag{9}$$

The multiplication of matrices gives the matrix

$$R_{xyz} = \begin{pmatrix} \cos\beta\cos\alpha & -\cos\beta\sin\alpha & \sin\beta \\ \cos\gamma\sin\alpha + \sin\gamma\sin\beta\cos\alpha & \cos\gamma\cos\alpha - \sin\gamma\sin\beta\sin\alpha & -\sin\gamma\cos\beta \\ \sin\gamma\sin\alpha - \cos\gamma\sin\beta\cos\alpha & \sin\gamma\cos\alpha + \cos\gamma\sin\beta\sin\alpha & \cos\gamma\cos\beta \end{pmatrix} \tag{10}$$

By making the rotation matrix R^T equivalent to R_{xyz}, we find the yaw, pitch, and roll angles (see Appendix A). Thus we can convert easily between the rotation matrix and rotation parameters.

Orientation Adjustments: In order to perform shape alignment, we must initialise the rotation angles. This is done by computing the R^T matrix for every original shape in the data set based upon the three identified landmark points and extracting the corresponding α, β, and γ angles. These are then further adjusted during iterative alignment via optimisation of the an-isotropic measurement based Mahlanobis distance. We also need to perform orientation adjustment on the mean shape following every iteration over the set of shape samples. In this case the set of yaw, pitch, and roll angles from the mean shape are subtracted off the corresponding rotation angles for each shape sample. Following convergence of the model the computed mean shape then complies with the orientation constraint defined according to our three point method.

Direction Vectors: In order to correct the covariances due to alignment parameters in 3D, we need the approximate linear direction vectors corresponding to translation, rotation and scale. Computing these for translation

and scale are straightforward. If $\mathbf{m} = [\mathbf{m}_{1x}, \mathbf{m}_{1y}, \mathbf{m}_{1z}, \mathbf{m}_{2x}, \mathbf{m}_{2y}, \mathbf{m}_{2z}, ...]$ is the vector corresponding to the 3D mean shape (with $3N$ elements), then the direction vectors due to translation in x, y and z directions are simply given by $\Theta_1 = [1, 0, 0, 1, 0, 0, ...]$, $\Theta_2 = [0, 1, 0, 0, 1, 0, ...]$ and $\Theta_3 = [0, 0, 1, 0, 0, 1, ...]$. Also, the direction vector due to scaling is $\Theta_4 = \mathbf{m}$.

For rotation, we compute the direction vector corresponding to each individual rotations R_z, R_y and R_x. For the mean shape $\mathbf{m}$ rotated by the yaw angle α around the z axis, we have $\mathbf{m}' = R_z(\alpha)\mathbf{m}$. The direction vector then would be the first derivatives of the vector $\mathbf{m}'$ with respect to the angle of rotation α, as α becomes very small (see Appendix A). If the coordinates for the mean point j is $[\mathbf{m}_{jx}, \mathbf{m}_{jy}, \mathbf{m}_{jz}]$, then the tangential direction of movement in landmark points due to this rotation is

$$\mathbf{u}_{\alpha \approx 0} = [-\mathbf{m}_{jy}, \mathbf{m}_{jx}, 0] \tag{11}$$

This corresponds to the linearised approximation vector which is used for correction of anisotropic 3D measurement covariances. By applying the same method, one can find the direction vectors due to rotation by the pitch angle β around the y axis and by the roll angle γ around the x axis. Hence we have

$$\mathbf{u}_{\beta \approx 0} = [\mathbf{m}_{jz}, 0, -\mathbf{m}_{jx}] \tag{12}$$

and

$$\mathbf{u}_{\gamma \approx 0} = [0, -\mathbf{m}_{jz}, \mathbf{m}_{jy}] \tag{13}$$

Hence, $\Theta_5 = [-\mathbf{m}_{1y}, \mathbf{m}_{1x}, 0, -\mathbf{m}_{2y}, \mathbf{m}_{2x}, 0, ...]$, $\Theta_6 = [\mathbf{m}_{1z}, 0, -\mathbf{m}_{1x}, \mathbf{m}_{2z}, 0, -\mathbf{m}_{2x}, ...]$ and

$\Theta_7 = [0, -\mathbf{m}_{1z}, \mathbf{m}_{1y}, 0, -\mathbf{m}_{2z}, \mathbf{m}_{2y}, ...]$. The set of vectors Θ_1, Θ_2 and Θ_3 on one hand, and the set of vectors Θ_5, Θ_6 and Θ_7 on the other hand are mutually orthogonal and orthogonal to the vector Θ_1 due to scaling. These direction vectors now constitute the linearised parameterisations needed for corrections to the sample covariance for degree of freedom biases (where i varies between 1 and 7 in Eqs. 6-7).

2.3 Model Selection

A method is needed to select appropriate linear model order based upon the outputs from our analysis. If the linear model is valid then estimated measurement covariances will combine two processes of statistical fluctuation. The first of these will be measurement precision σ_r (our ability to define homologous points reliably), and the second will be due to random (unmodellable) biological variation σ_b. So that the observed statistical variation seen in a given direction v for any landmark σ_v is

$$\sigma_v^2 \;=\; \sigma_r^2 \;+\; \sigma_b^2 \tag{14}$$

Unfortunately we cannot know the expected value of σ_b in advance. However, the first of these terms can be estimated via reproducibility experiments and compared to the measured directional covariances, using the observation that

$$\sigma_v \geq \sigma_r \tag{15}$$

Thus if we observe individual estimates of measurement covariance which begin to surpass the limiting accuracy known to be set by reproducibility tests, then the model must be overfitting the data and therefore has too many parameters. We check that for a given model order this inequality is satisfied within statistical limits by considering the principle axes of each landmark measurement distribution. We use a 1% confidence level to set the hypothesis test for overfitting. This test is expected to be most reliable for the largest variances.

2.4 Monte-Carlo Tests and Outlier Identification

As our method is based on likelihood, we require that the assumed distribution matches the sample covariance or at least the corrected covariance. The standard way to validate this is through generating Monte-Carlo (MC) data using the known distributions.

In what follows we experiment with MC data and display a number of informative scatter plots for two forms of test;

Test A: When applying our method to the MC data, the mean shape, eigen-vectors and measurement covariances used are identical to the ones used when generating the simulated data.

Test B: All parameters are estimated using the MC data in order to compare the measurement covariances estimated using the simulated data with those expected, i.e. the ones assumed when generating the MC data.

For Test A the covariances estimated using our method are expected to be within statistical sampling limits of the ones used when generating the MC data. Failure to do so is taken as an indication of a problem with the data sample (i.e. outliers). Outliers can be identified at early stages of analysis as those points which have the largest normalised residual errors.

We use 2.8 standard deviations of the error on the sample variance (or being allowed to have 1% of data falling outside the limits), where the error on the standard deviation σ is $\sigma/\sqrt{2(K-1)}$ with K being the number of samples. Additional variance will be seen for Test B, where the linear model must also be estimated, so that we can interpret variations beyond the statistical limits as due to instability in linear model construction (specifically the mean and eigen vectors).

2.5 χ^2 Test

A test is needed to confirm the equivalence of measurement covariances computed during repeatability experiments, in order to confirm that our methods generate estimates which are consistent. This can also be done by splitting the data into two separate groups if there are a sufficient number of samples. We perform a modified χ^2 test based upon the construction of corrected covariances on one data set and then used for the calculation of χ^2 for the second set. For large numbers of samples ($K > 30$) the resulting statistic when applied to each 2D landmark is expected to be approximately Gaussian with mean $2K$ and variance $4K$. We set the statistical test for significant difference on the basis of an allowable range of χ^2/DoF corresponding to ± 2.8 $S.D.$, i.e. [0.8, 1.2] for 200 samples. The corresponding plot would confirms the stability of the method if 99% of the χ^2/DoF values fall inside the range expected.

2.6 Fisher Information

Fisher information (FI) is a concept for quantifying the constaint on an estimated value associated with data. It has the useful property that the amount of estimated information is linear in the quantity of data. It is generally defined according to the second derivative of a log-likelihood function, but from the association of this function and the Cramer-Rao Bound (CRB) we can also observe that, for good model fits, it is proportional to the inverse variance. An impirical estimate of the FI contained in data, and associated with a particular model, can therefore be obtained from the residual distributions following parameter estimation.

We use this idea here to summarise the amount of information which has been extracted from data for a specific analysis. As this quantity scales linearly with the quantity of data it allows us to make comparative statements regarding the statistical efficiency associated with the estimation process. For example, if the FI is seen to double on the same dataset when applying an alternative analysis then this is statistically equivalent to having four times as much data to begin with, i.e. a poor analysis method might need a lot more data to reach the same level of statistical equivalence in a hypothesis test than a good method.

3 Experiments

The purpose of the following experiments is to investigate the stability of covariance weighted shape analysis and to compare quatitative performance figures to the standard approach using Procrustes. We have selected several datasets in order to demonstate behaviour with different quantities of data, data dimensionality (i.e. 2D and 3D) and model order.

3.1 Data

We experiment with two 2D data sets of manual mark-ups (Fig. 2). The first data set, called MM1, corresponds to mouse mandible micro-CT images and consists of 337 samples with 14 landmarks per sample. We also have a repeat data set, called MM2, for which same mandible images have been used to mark-up the points. Both data sets MM1 and MM2 were provided by our collaborators [2].

Next, we use some fly wing data in order to test the performance of our method on pseudo landmarks and also to test the statistical stability of our method. There are four original data sets available from left and right wings (L and R) of 200 female flies, called FL1, FL2, FR1 and FR2 [1]. Two images of each wing has been taken from slightly different viewing positions (1 and 2), and used for marking-up in order to perform reproducibility tests [1]. Each of these four data sets has 200 samples with 15 landmarks per sample. Further, as we had access to the fly

wing images, we have added four pseudo landmarks to each sample of the original data set FL1. Once finished, we removed 5% outliers and stored 189 samples with 19 landmarks per sample. This resulting data set, which is called P-FL1, plays an important role in our experiments with pseudo landmarks. In order to be able to test the repeatability with these added pseudo landmarks, we have repeated the marking-up process only for the four new landmark points and using a subset of the left fly wing images.

We also experiment with the mouse skull (MS) 3D data of automatic mark-ups (Fig. 3) for which the micro-CT images and training examples have been provided by our collaborators [2] as a typical 3D data set of their interest in evolutionary biology research. We have used our automatic tool to localise landmarks on these mouse skulls based on few given manual markup examples [9]. The automatic tool also removes outliers and so we do not expect any outlier in this data set. The mark-up data set obtained this way (MS) consists of 42 samples with 50 landmarks per sample. Further, there are two sub-sets of repeat data based on manual mark-ups (on the mouse skulls) each consisting of 12 samples to be used in repeatability tests.

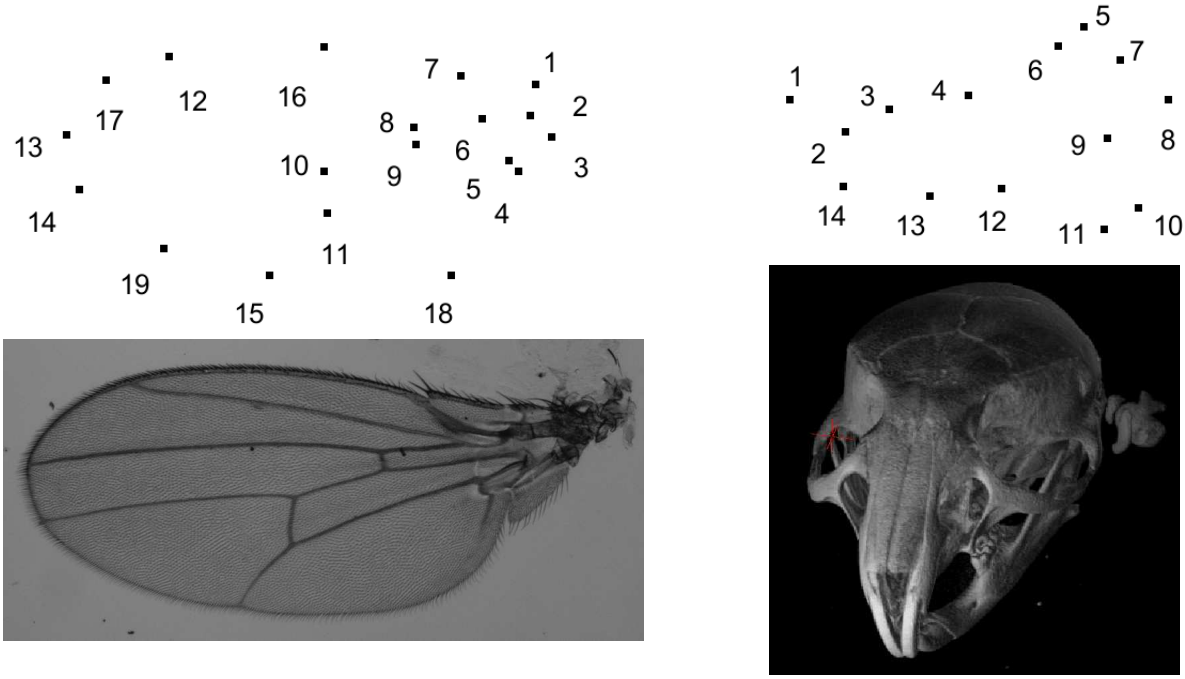

Figure 2: Typical landmarks corresponding to sample images of fly wings (left) and mouse mandibles (right); for the fly wing data, landmarks 1-15 correspond to the original data sets FL1, FL2, FR1 and FR2, while landmarks 16-19 were added later (to FL1) in order to experiment with pseudo landmarks (P-FL1).

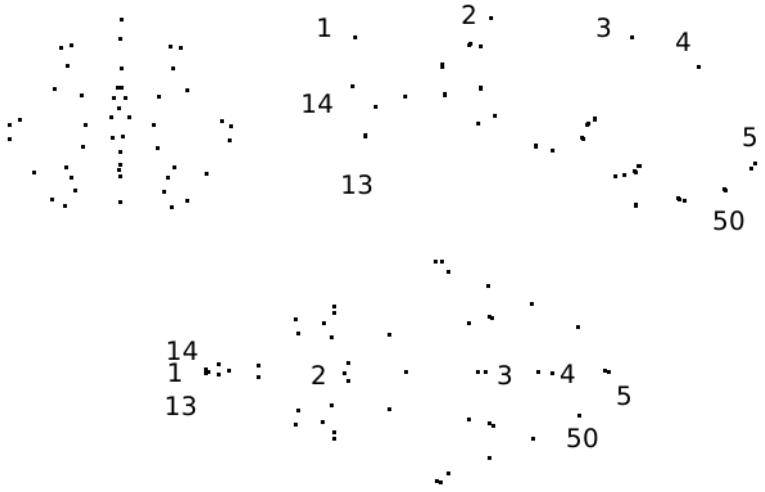

Figure 3: Typical landmarks for a sample shape from the 3D mouse skull (MS) data when projected on the zy (left), xy (middle) and xz (right) planes (the 50 points are too close to display full numbering).

3.2 Model Selection

In Figs. 4, 5 and 6, we plot the eigen-values corresponding to the errors estimated against those computed from the repeat data. These are the magnitudes of the errors in the direction of major eigen-vectors. It can be seen that while for the fly wing data the errors are comparable (Fig. 4), for the mouse mandibles there are several landmarks for which the error estimates are much larger than expected (Fig. 6). We cannot argue for an increased model order as this then reduces other values to well below the observed repeatability (over fitting). As the additional variance seen is due to the inability of the model to predict correlations in the data, our conclusion must be that either this data is not well described by a linear model, or the repeatability estimate systematically underestimates the true accuracy with which points can be meaningfully located. This can happen if local image features (which are themselves not well biologically related to the main structures, such as the brightest pixel) are used to identify locations. The plot for the more complex mouse skull data (Fig. 6) suggests that 14 components is about the number of components needed by the linear model. Hence, when experimenting with the 3D MS data we use 14 model components, while using 6 components with the 2D MM data and 2-3 components with the 2D FL and P-FL data. Here the 1% allowable range is wider than expected. This is because the horizontal axis corresponds to 12 repeat data samples while the vertical axis corresponds to 42 data samples. It is the smaller number of samples which must be used to calculate the statistical range.

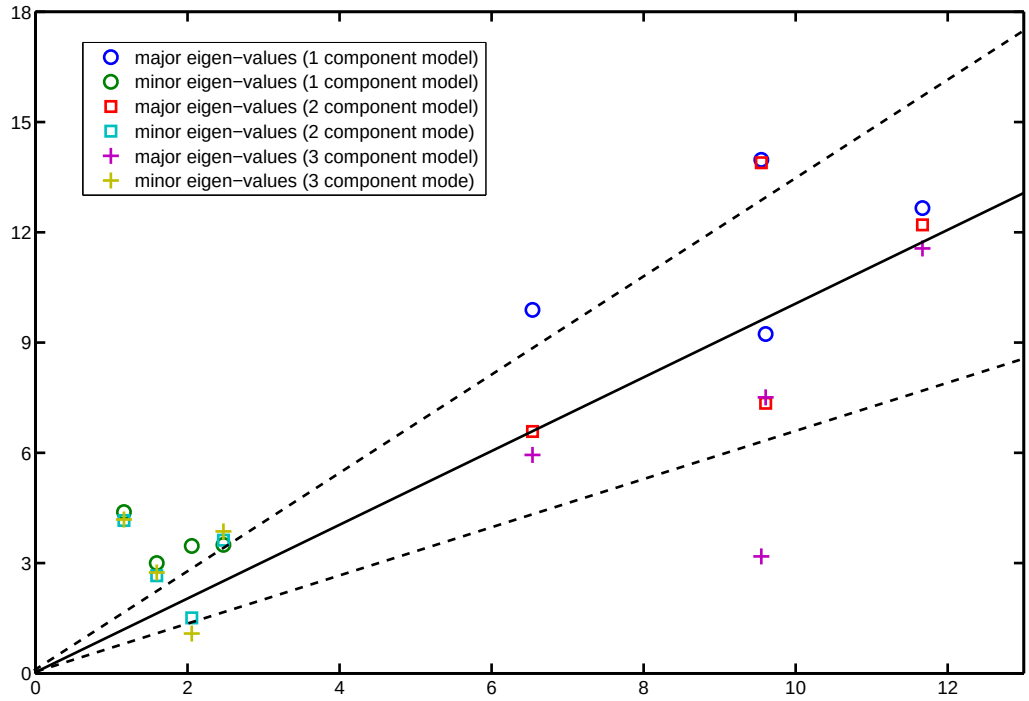

Figure 4: Fly wing data (P-FL1): major eigen-values of the error using our 1, 2 and 3 component models against those computed using a repeatability test on four new pseudo-landmarks placed manually on a subset of data; the two component model gives closest agreement to the expected localisation values; the two dashed lines show the $\pm 2.8\sigma$ range.

3.3 Monte-Carlo Tests

We show the Monte-Carlo plots for the Test A in Figs. 7, 9 and 11, while the Figs. 8, 10 and 12 show those for the Test B. The results for the Test A on 2D data sets (Figs. 7 and 9) indicate very little difference for the low parameter fly wing data, and a more obvious systematic underestimate of covariance (as expected) for the 6 dimensional mouse mandible data.

Further, the results for the Test B (Figs. 8 and 10) indicate that even for the mouse mandible data, the values of covariance are significantly different, due to the amplification of initial estimation bias during the process of iterative linear model estimation. The correction process now removes these instabilities bringing estimated covariances back close to the expected sampling limits and symmetrically around the expected correlation line.

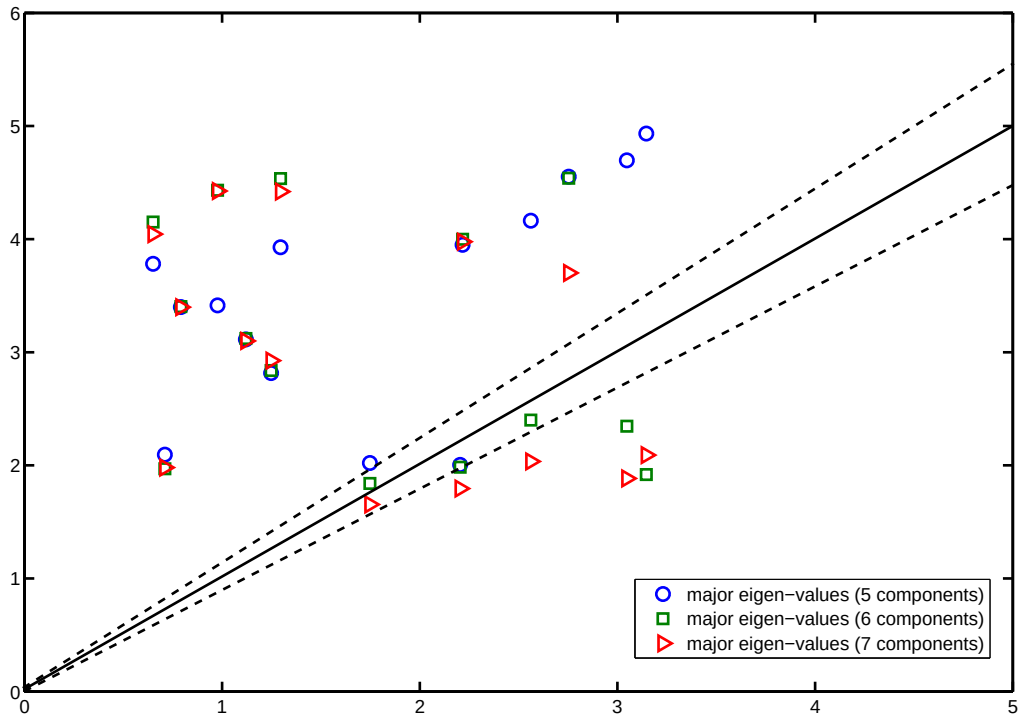

Figure 5: Mouse mandible data: major eigen-values of the error estimated using our 5, 6 and 7 component models on MM1 data against those computed using the corresponding repeatability test (MM1 and MM2); the two dashed lines show the $\pm 2.8\sigma$ range.

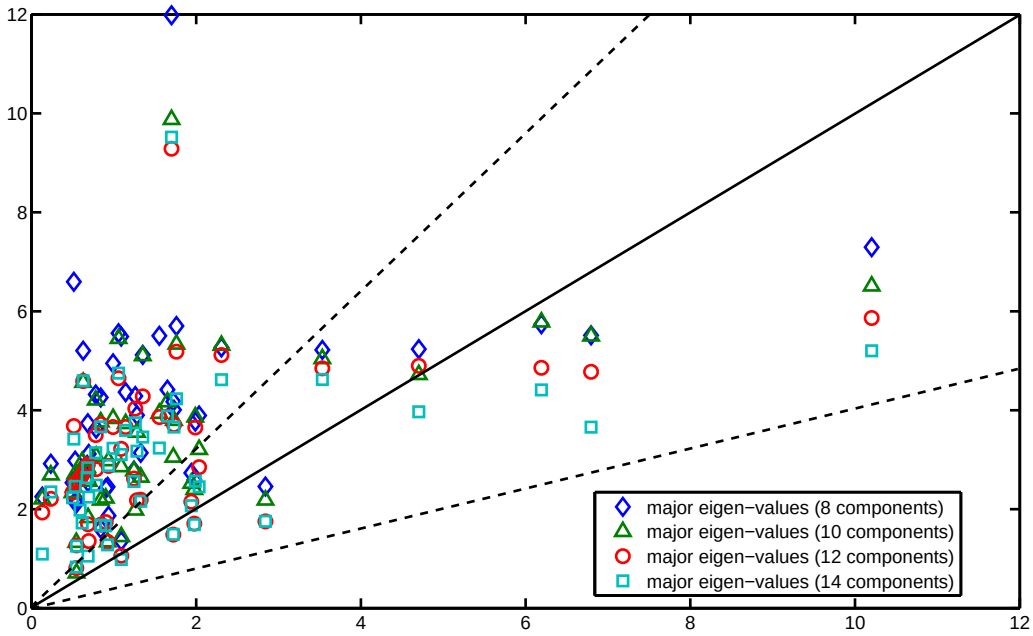

Figure 6: Mouse skull data: major eigen-values of the error estimated using our 8, 10, 12 and 14 component models for the 3D MS data against those computed using the corresponding repeatability test; the two dashed lines show the $\pm 2.8\sigma$ range.

Turning to the 3D MS data, for the Test A (Fig. 11) the eigen-values fall inside the allowable range (dashed lines). However for the Test B (Fig. 12), the eigen-values appear to fall under the lower bound. This is due to insufficient number of data samples leading to over-fitting when estimating the eigen vectors of the linear model. The proportion of under-estimation can be estimated using a proportional correction based upon the number of samples and model complexity $(K - J)/K$. Unlike the earlier biases this under-estimation does not otherwise destabilise the analysis, as the effect generates a common factor on all variance estimates.

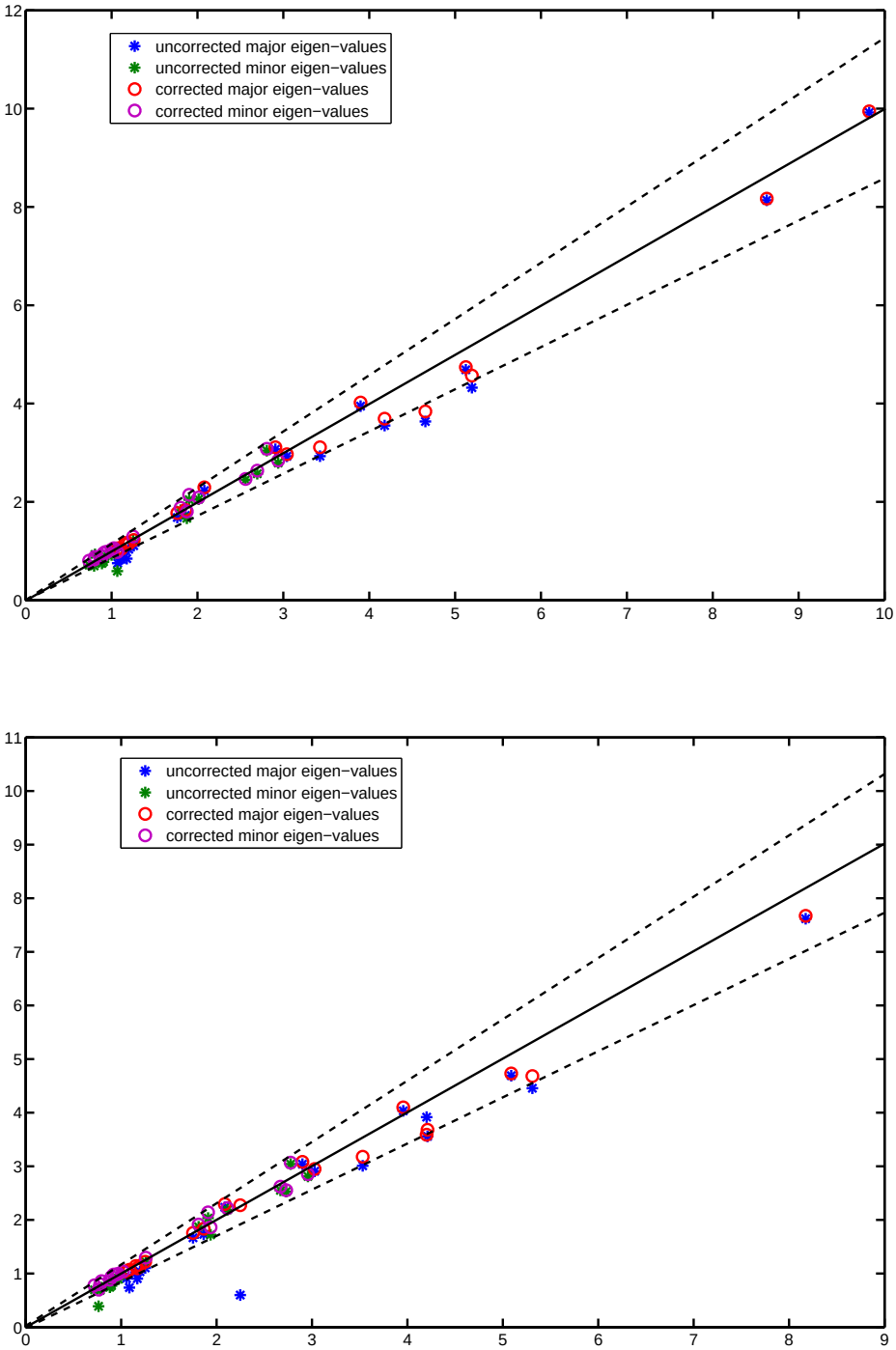

Figure 7: Fly wing data (P-FL1): error eigen-values estimated using the Monte-Carlo data (where mean shape, eigen-vectors, and measurement covariances are identical to the model which generated the simulated data) against the expected ones (Test A); using 2 (top) and 3 (bottom) model components; there is only marginal evidence of estimation bias before correction; the two dashed lines show the $\pm 2.8\sigma$ range.

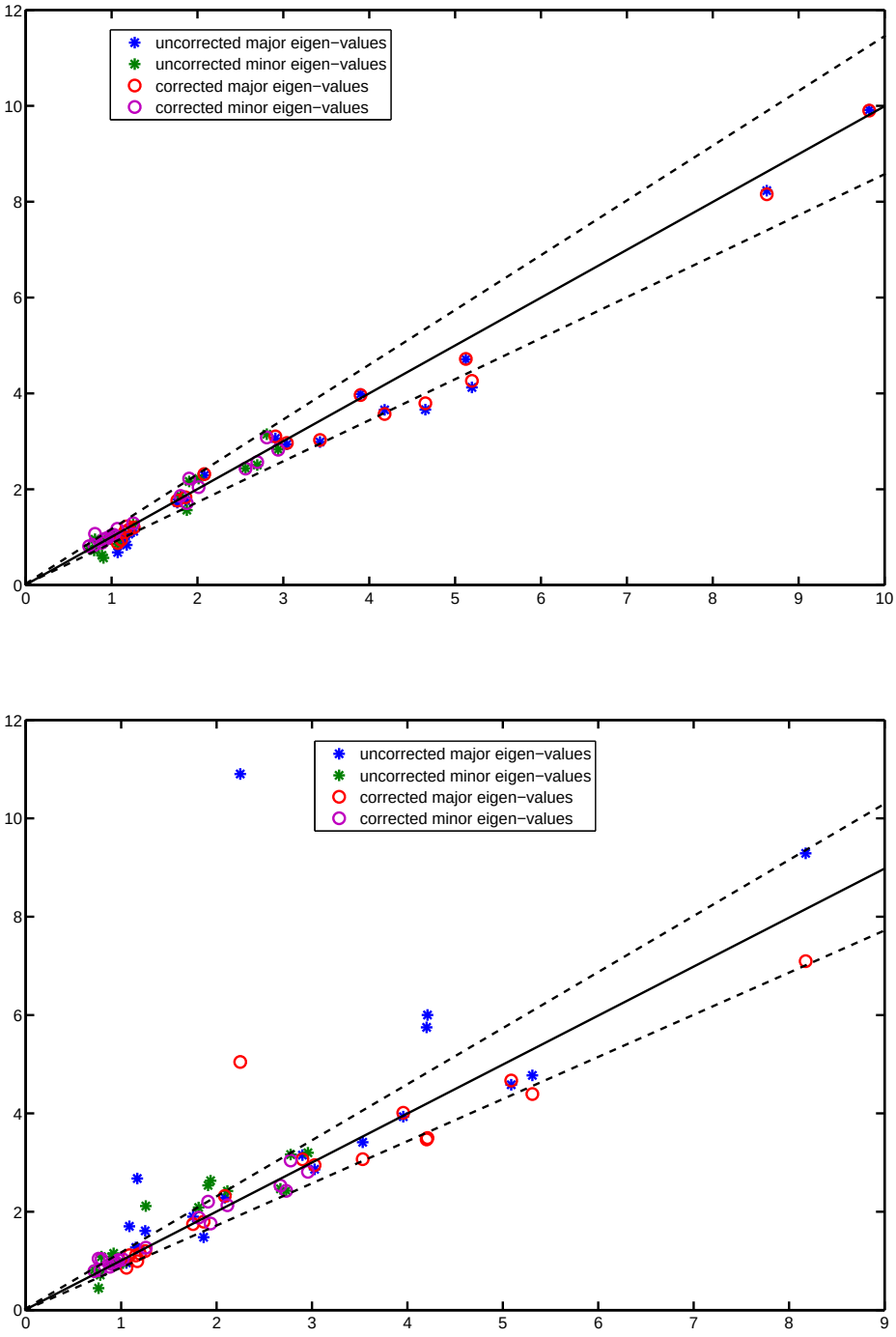

Figure 8: Fly wing data (P-FL1): error eigen-values estimated using the Monte-Carlo data against the expected ones (estimated using the original data) which were used when generating the simulated data; independent models (Test B); using 2 (top) and 3 (bottom) model components; there is evidence of parameter estimation instability which has been reduced by covariance correction; the two dashed lines show the $\pm 2.8\sigma$ range.

3.4 Shape Analysis

In Figs. 13, 14 and 16, we show the anisotropic error bars computed using the eigen vectors and values of the 2x2 covariance matrices for the 2D data sets. All error bars are rescaled for visualisation purposes (see captions). Error bars for each landmark show the extent of an elliptical (non-isotropic) distribution around the corresponding point in the mean shape. Such distributions estimated using our method show exactly why we cannot assume isotropic

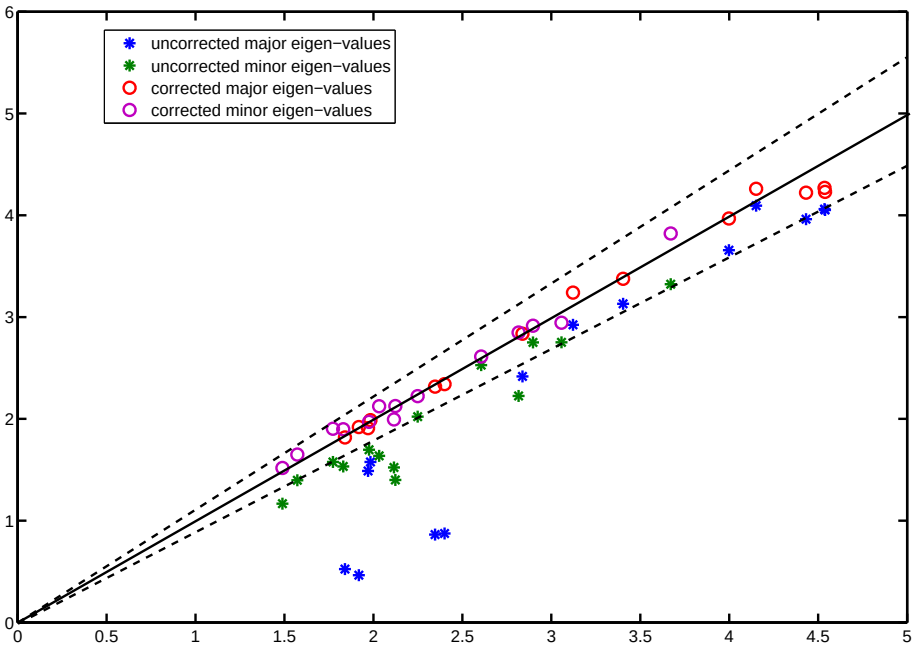

Figure 9: Mouse mandible data (MM1): error eigen-values estimated using the Monte-Carlo data (where mean shape, eigen-vectors, and measurement covariances are identical to the 6-component model which generated the simulated data) against the expected ones (Test A); for this number of parameters there is now evidence of a systematic underestimate of covariance; the two dashed lines show the $\pm 2.8\sigma$ range.

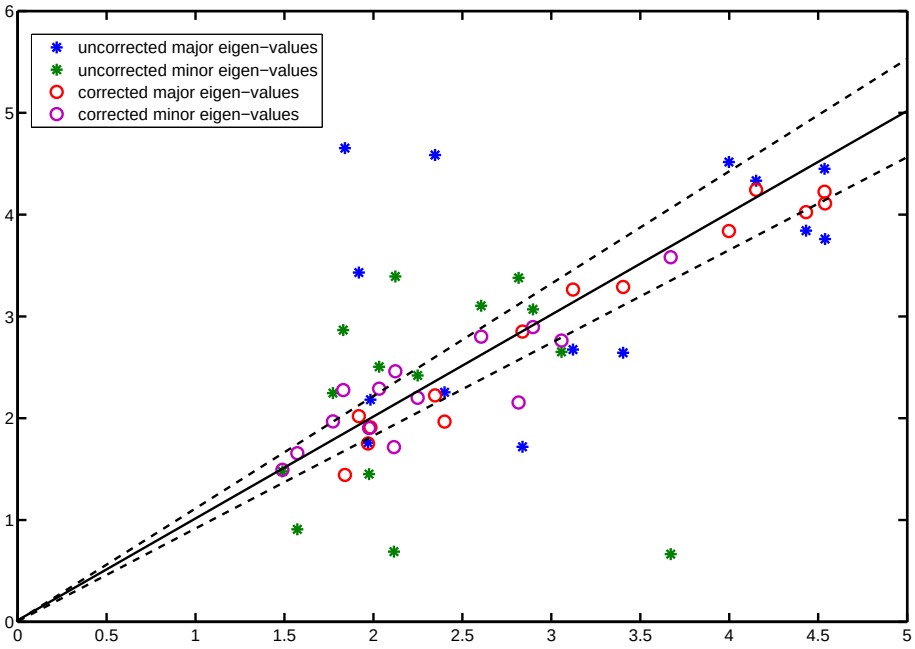

Figure 10: Mouse mandible data (MM1): error eigen-values estimated using the Monte-Carlo data against the expected ones (estimated using the original data) which were used when generating the simulated data; independent 6-component models (Test B); for this number of linear model components there is considerable error in the uncorrected estimates; the two dashed lines show the $\pm 2.8\sigma$ range.

distributions for the data as assumed in Procrustes. The P-FL1 data used in Fig. 13 consists of 19 landmarks (15 + 4) while the FL1 data used in Fig. 14 consists of 15 common landmarks only. Using these plots, one can see

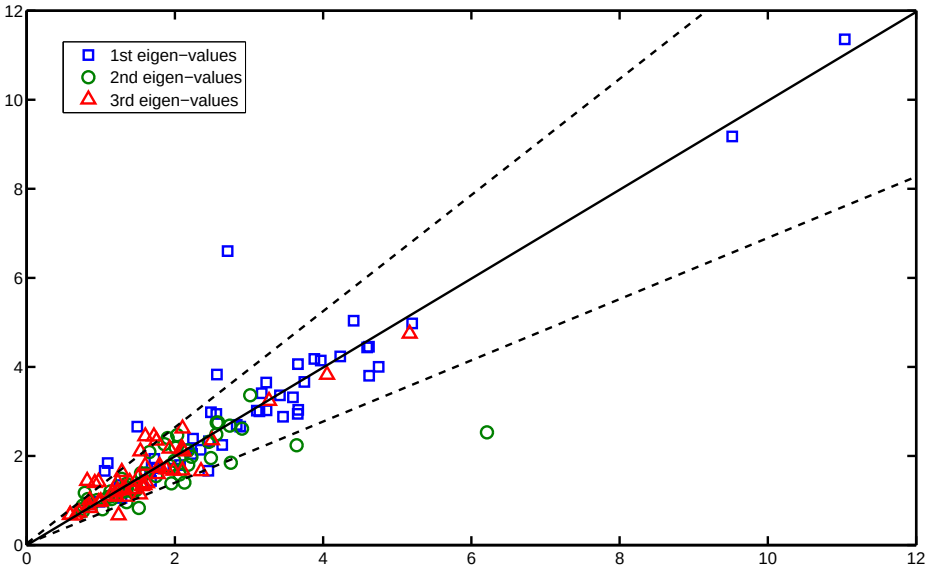

Figure 11: Mouse skull data (MS): error eigen-values estimated using the Monte-Carlo data (where mean shape, eigen-vectors, and measurement covariances are identical to the 14-component model which generated the simulated data) against the expected ones (Test A); the two dashed lines show the $\pm 2.8\sigma$ range.

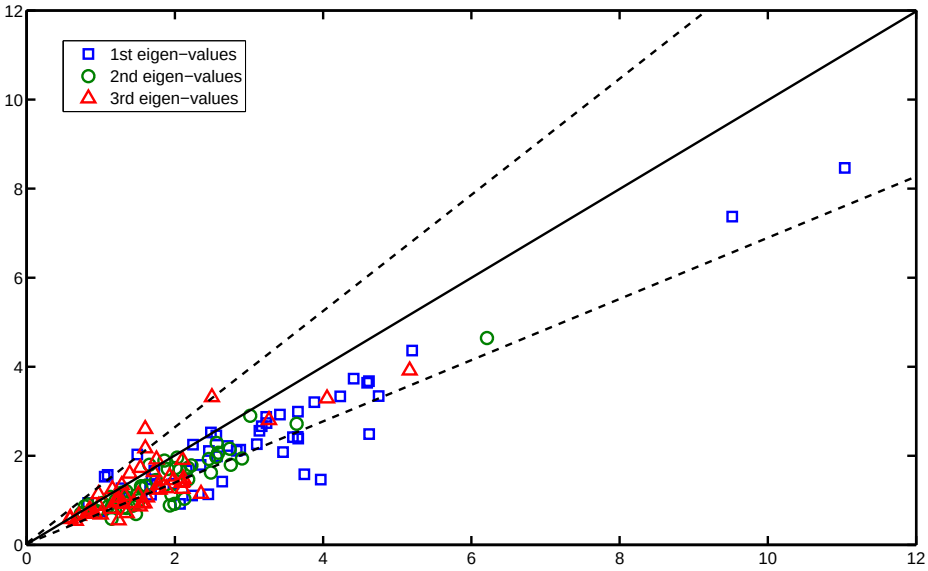

Figure 12: Mouse skull data (MS): error eigen-values estimated using the Monte-Carlo data against the expected ones (estimated using the original data) which were used when generating the simulated data; independent 14-component models (Test B); the two dashed lines show the $\pm 2.8\sigma$ range.

that the psuedo landmarks have anisotropic covariances which match the expected localisation stability. Also we can see how after adding the 4 pseudo landmarks the anisotropic errors estimated using our method remain stable, while with Procrustes some change both in orientation and in size (e.g. points 11 and 15). These error bars are shown again in Figs. 15 and 17 with the corresponding aligned data superimposed. In these figures, the extent of error distributions illustrated by the error bars are not expected to match to those illustrated by the alignment, as general biological shape variation and measurement error are independent processes. Although, localisation is determined by local shape characteristics and measurement accuracy plays a role in the overall distribition of landmarks around the mean shape. As a consequence poorly measured landmarks may have a variation about the mean which is dominated by noise.

Turning to 3D data, we have shown the anisotropic error bars estimated using our method in Fig. 16 and the corresponding aligned data in Fig. 17. In order to display our 3D results we have used their projections on three 2D planes. In both figures, from left ro right, we show the projected results on zy, xy and xz planes. Using the mouse skull volume shown in Fig. 2, one can match these viewing planes (zy, xy and xz) to the coronal, sagittal and transverse planes respectively. In Fig. 16, one can observe that the largest error bars are for the landmarks 3 and 1. This is consistent with the data clouds corresponding these landmarks in Fig. 17 where in each case few points stand away from the main cloud. It is clear from the zy and xz projection planes that expected symmetry is achieved to a large extent in orientation and size for most corresponding error bars. In order to compare the amount of errors estimated using our method to those suggested by the repeat data, one should revisit Fig. 6. The figure suggests comparable error estimations.

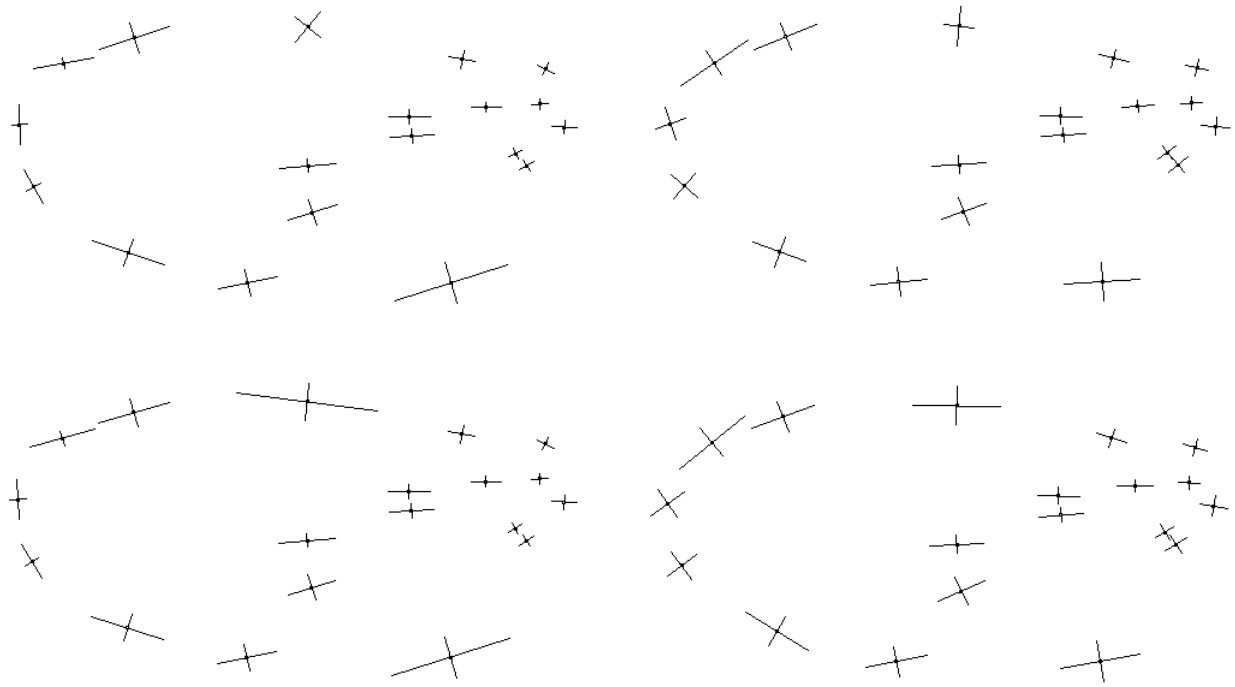

Figure 13: Fly wing data (P-FL1): error bars (×20) estimated using our method (left), and computed from the residuals left using Procrustes (right); using 3 (top) and 2 (bottom) model components.

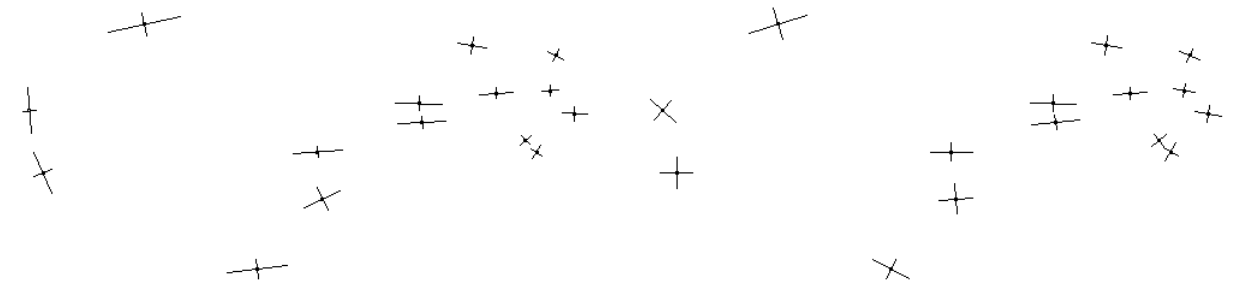

Figure 14: Fly wing data (FL1): error bars (×20) estimated using our method (left), and computed from the residuals left using Procrustes (right); 2-component models.

Here we turn to comparing our method to Procrustes in a more quantitative manner. The inconsistency observed earlier in error bars corresponding to the residuals left after applying Procrustes to the fly wing data (Figs. 13 and 14) is displayed more clearly using a scatter plot in Fig. 20. Here we plot the eigen-values corresponding to the 15 common landmarks after the 4 pseudo landmarks are added against those without any additional landmarks. We plot this for both the likelihood and Procrustes methods. It is clear here that there are departures from the permitted scatter region when Procrustes is used. This indicates a significant change in the unexplained variance

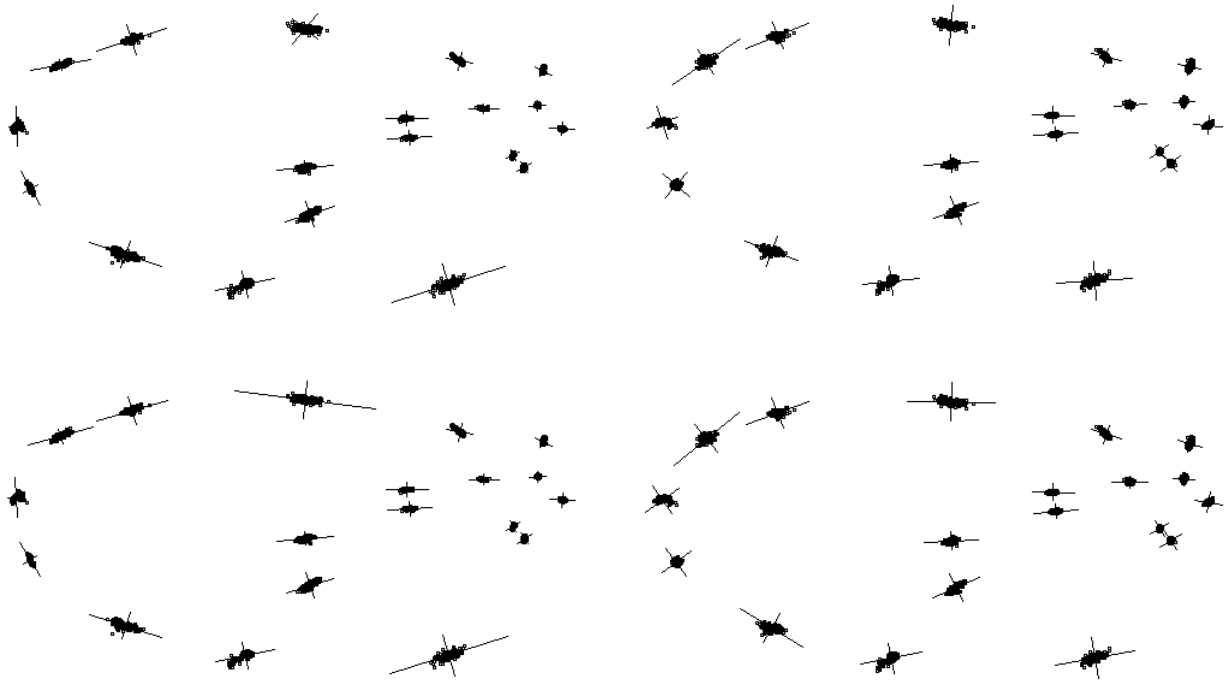

Figure 15: Fly wing data (P-FL1): error bars (×20) estimated using our method (left), and computed from the residuals left using Procrustes (right), with the corresponding aligned data superimposed; using 3 (top) and 2 (bottom) model components.

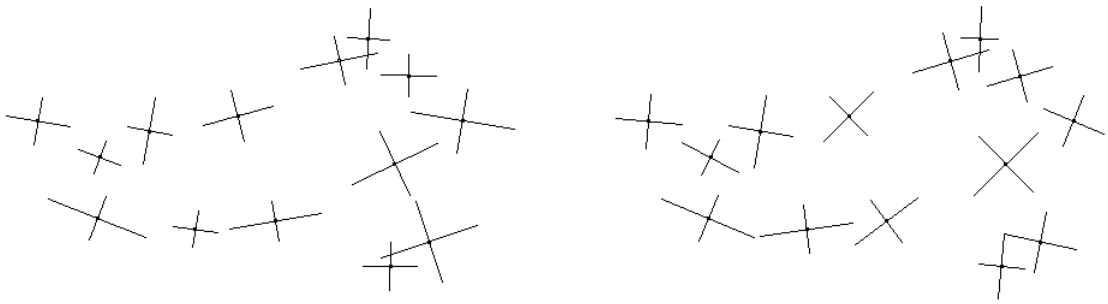

Figure 16: Mouse mandible data (MM1): error bars (×20) estimated using our method (left), and computed from the residuals left using Procrustes (right); 6-component models.

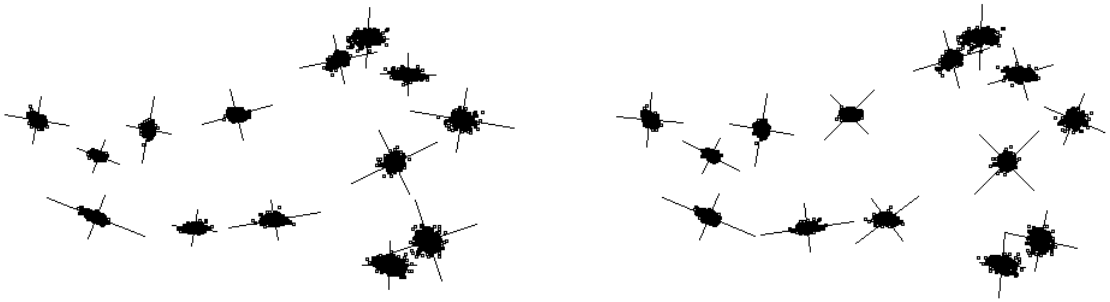

Figure 17: Mouse mandible data (MM1): error bars (×20) estimated using our method (left), and computed from the residuals left using Procrustes (right), with the corresponding aligned data superimposed; 6-component models.

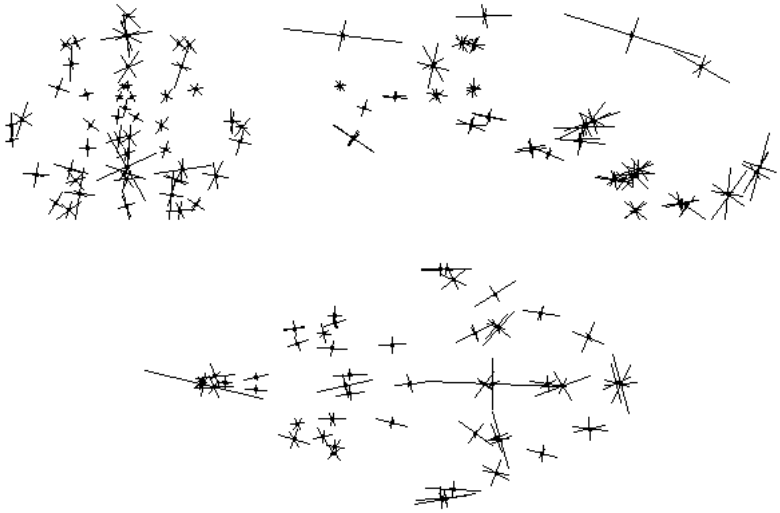

Figure 18: Mouse skull data: error bars (×30) estimated using a 14-component model based on our covariance-based method; projection planes: zy (left), xy (middle) and xz (right).

Figure 19: Mouse skull data: aligned data obtained using a 14-component model based on our covariance-based method; projection planes: zy (left), xy (middle) and xz (right).

following linear model construction, which itself implies differences in the linear model itself, i.e. the Procrustes model is unstable following the addition of poorly measured landmarks.

Further, we performed a χ^2 test based upon the construction of corrected covariances on one data set (FL1/FR1) and then used for the calculation of χ^2 for a second data set (FL2/FR2). The corresponding plot for χ^2 test in Fig. 21 confirms the stability of our method, as all χ^2/DoF values fall in the range expected. Further χ^2 tests (not shown here) with different numbers of data samples and combinations of data sets indicate the appropriateness of the assumed linear model for the fixed number of components.

In Table 1, we list the Fisher Information (FI) value for the two methods and the three data sets studied. Again, for Procrustes, variances used to compute the FI value are obtained from the residuls left after alignment between the data and the simulated linear model. .... Our method which is based on likelihood and measurement covariance gives FI values roughly between two and four times those obtained using Procrustes. The largest difference corresponds to the 3D MS data with 14 model components. As FI is proportional to the quantity of data, this demonstrates that the changes away from the isotropic assumption inherent to Procrustes/PCA has a significant effect on the efficacy of the model, equivalent to having defined only a third as many landmarks from the outset. We can also see in this Table the effect of adding 4 pseudo landmarks to the 15 original landmarks. The numbers in the parentheses show the contributions from the 4 added landmarks to the total FI values. The reason for the decrease in the FI values after adding 4 landmarks is that we are using the same number of degrees of freedom to describe correlations between more points. Also these values are computed for uncorrected covariance estimates, because correction is not available when using Procrustes.

Finally, we have studied the plots (not shown here) corresponding to the cumulative content based on the PCA analysis of the linear model used for Procrustes alignment. It appears that while for fly wing data 3 components can account for about 70% of the data, for mouse mandibles 6 components are needed to account for the similar 70% amount of data. In both cases, the model order preferred by our analysis corresponds to approximately 60% of the observed variance, significantly less than the heuristic limit of 90% used by some researchers.

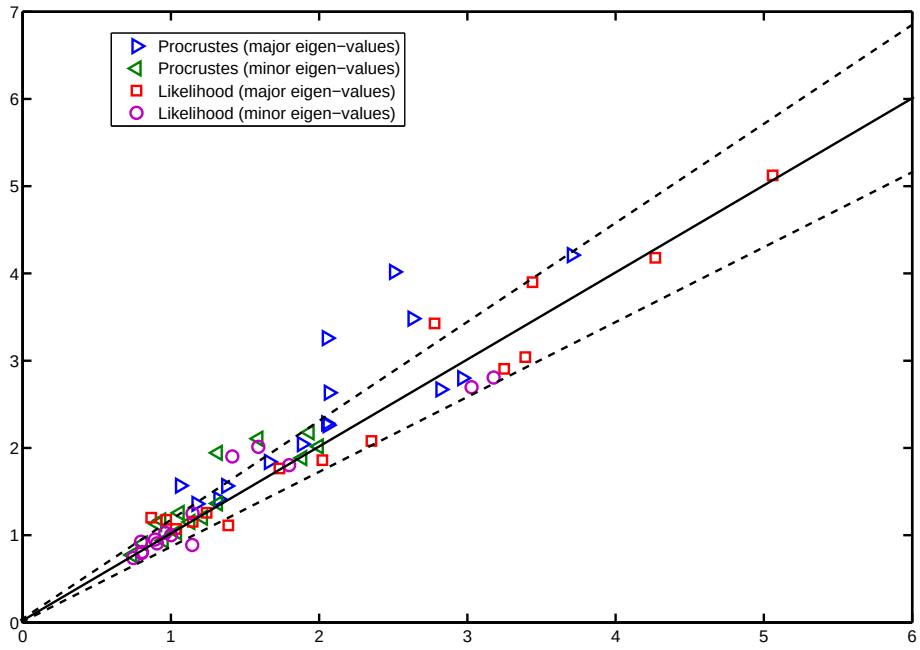

Figure 20: Fly wing data: error eigen values estimated using the likelihood method and those computed using the residuals after Procrustes alignment, when each method is applied to P-FL1 (4 pseudo landmarks (16-19) added to FL1) against those when applied to FL1; the plot is for the 15 common landmarks (1-15); the Procrustes results have many points which are well beyond the expected limits, suggesting that model parameters are not consistently determined upon inclusion of pseudo-landmarks; the two dashed lines show the $\pm 2.8\sigma$ range.

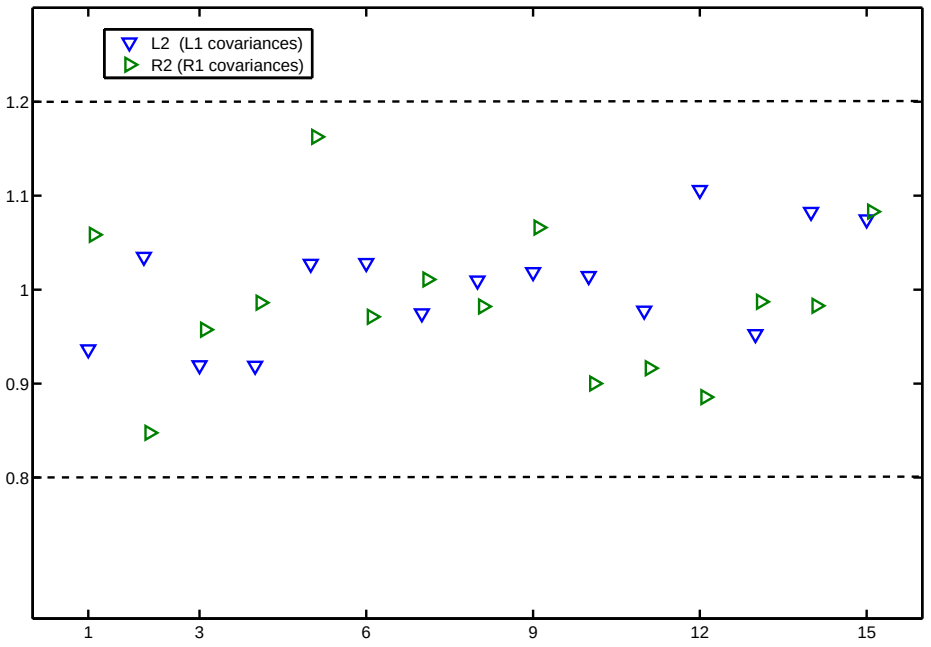

Figure 21: Fly wing data (15 original landmarks): the χ^2/DoF ratios when our method is applied to two sets of repeat fly wing data, FL2 and FR2, using a 3-component model and fixed covariances (Fig. 13: top-left) estimated earlier from the FL1 and FR1 data sets respectively; the two dashed lines show the $\pm 2.8\sigma$ range.

Table 1: Fisher information (FI) values: listed for the Procrustes and our method when applied to the fly wing data (P-FL1), mouse mandible data (MM1) and 3D mouse skull data (MS).

FI value	Procrustes	Likelihood
Mouse 3D skulls (14-component model)	23.62	111.88
Mouse mandibles (6-component model)	6.60	19.55
Fly wings (2-component model) 15 points	17.36	29.68
Fly wings (2-component model) 15 (+4) points	13.81 (+0.88)	25.46 (+2.47)

3.5 Discussions and Conclusions

From a more philosophical standpoint we can consider what we are doing when we identify landmark locations and attempt to compare them between sets (shapes). We do not expect that biology manipulates the locations of our chosen landmarks directly, they simply appear to move around as the net effect of distributed developmental influences. Recent considerations of biology have introduced the phrase palimpsest [18], as an analogy with repeatedly erasing and rewriting text in ancient parchment, to describe the way that structures develop. Notice that the initial choice of landmarks is subjective, not only in terms of the features selected but also how we chose to define their locations. Our concept of a landmark being at a specific location is purely based upon a pre-conception of co-ordinates, both for origin and specification of axes. This is driven by anatomical naming conventions, which encourage researcher to define approximate regions as localisable objects at ever finer scales in order to support correspondance matching. Issues arising due to semi-landmarks are then the inconvenient aspects of this problem which cannot be re-defined away. A landmark is the result of a localisation procedure (partly influenced by multiple biological considerations) which has an associated positional uncertainty. In this work we have associated the problems of working with pseudo-landmarks in biological shape analysis as being a consequence of the statistical assumptions implicit to analysis techniques such as Procrustes/PCA. We have implemented a new method which takes appropriate account of measurement and landmark localisation stability in order to obtain a new form of analysis which is consistent with a Likelihood based definition of the alignment and model building tasks. This method can be equivalently interpreted as a redefinition of the landmark location as ghost points.

The conventional interpretation of Procrustes is that the resulting linear model is a pure shape description which can be directly associated with biological processes. Some may argue that extending the approach to weight data, even to accommodate pseudo-landmarks, breaks with this tradition. However, it is our belief that any distinction between the original landmark and a ghost point, as locations which are somehow true measurements of biology in one case but not the other is arbitrary. Use of a least squares measure (which assumes isotropic errors) does not introduce some absolute measurement of biology. Both approaches need to be calibrated using known samples with identifiable biological cause in order to make any scientific interpretation. Also, reweighting of data using a covariance is statistically equivalent to modifying the information available by changing the specified set of landmarks.

Now that we have a specific definition for how to weight landmark data, we can see that using ghost points does not invalidate use of Kendalls statistics as suggested in [36]. The use of these approaches follows due to scale normalisation of the shape data, it is not an intrinsic property of the use of the original landmarks co-ordinates per-se. We can also reproject scaled (whitened) shapes onto the tangent space defined in the transformed ghost space if we wish, in order to remove local curvature arising from scale normalisation.

Far from there being no objective way to define these covariances [36], [22], there are at least three; a) we can estimate them directly from repeatability of measurements; b) they can be directly estimated via conventional statistical means when using Likelihood based landmark location (CRB); 3) they can be estimated as the unexplained stochastic variation (residuals) in fitted data. For the latter, when estimated using residuals of the fitted shape model, we will see contributions additional to the measurement process, this is the stochastic (therefore unmodellable) behaviour of the biology itself. The CRB estimated by the second approach, (suitable for automated measurement systems) is defined only by measurement, and is therefore also a bound on the minimum observable variation possible for each landmark. Our results indicate that measurement covariances can be reliably estimated in our data for sample sizes at least as small as 40. However, we would not expect it to be possible to use a single measurement covariance for all landmarks [22], nor to estimate this covariance without making appropriate adjustments for estimation bias.

In summary, a linear model constructed for describing shape variation in a normalised space is a low dimensional description of the measurements supporting a lower limit approximation of statistical difference. It only has

biological meaning when locations in the resulting space are calibrated with other biological knowledge. This is no less true for Procrustes/PCA analysis as for the more sophisticated approach suggested in this document. We should therefore be able to use this new approach to resolve problems with analysis of psuedo-landmarks, without destroying any innate biological validity of a shape space derived using Procrustes. Our interpretation of the resulting linear model should simply be this; the locations we intend to define as landmark positions are those points in the image plane which can be projected from observed structure with homogenous measurement noise (ghost points). The model constructed is then a summary of correlations which can be inferred from the set of measurements we have chosen to take.

Our result indicate that the new method summarises the information content of the measured data better (improved FI scores), and with more stability than Procrustes/PCA (consistent models are generated following the addition of new points). The theoretical advantages of this approach are several; a) it will lead to a consistent way of incorporating all forms of landmark into our analysis; b) it removes the instabilities inherent in the analysis due to poorly determined points; c) it affords the application of an eigen vector analysis statistical rigour; d) it offers the possibility of interpreting the linear modelling process as a statistical approximation, with consequent interpretations of the requirement for the number of linear model components; d) finally, generalisation of the approach would seem to be possible which would support the analysis of curves.

We have also demonstrated how linear model order selection can be performed by comparing baseline reproducibility errors with those estimated from the model. Finally, we have shown how the use of repeated analysis on matched samples can be used to confirm the stability of the estimated anisotropic error. We believe that these tools are sufficient to allow use of this technique in biological studies.

The methods described in this paper are freely available from the Tina web site [9] via the automatic 3D landmarking tool, as a system for quality assessment and validation of output data.

Appendix A

Rotation Matrix: Based on the geometry shown in Fig. 1, we first calculate the vectors $\hat{v}_a$, $\hat{v}_b$ and $\hat{v}_c$.

$$\hat{v}_a = \frac{p_2 - p_1}{\|p_2 - p_1\|}$$

$$v_b = (p_3 - p_1) - [(p_3 - p1).\hat{v}_a]\hat{v}_a$$

$$\hat{v}_c = \hat{v}_a \times \frac{v_b}{\|v_b\|}$$

The rotation matrix R is hence given by:

$$R^T = \begin{pmatrix} \hat{v}_{ax} & \hat{v}_{ay} & \hat{v}_{az} \\ \hat{v}_{bx} & \hat{v}_{by} & \hat{v}_{bz} \\ \hat{v}_{cx} & \hat{v}_{cy} & \hat{v}_{cz} \end{pmatrix}$$

Roll, Pitch and Yaw Angles: The rotation matrices about the x, y and z axes are given by:

$$R_x(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}$$

$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\beta \end{pmatrix}$$

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Hence by enforcing $R^T = R_x(\gamma)R_y(\beta)R_z(\alpha)$, we find the rotation angles

$$\begin{aligned}\alpha = &\quad -\tan^{-1}(R^T[1][2]/R^T[1][1])\\ \beta = &\quad -\sin^{-1}(R^T[1][3])\\ \gamma = &\quad -\tan^{-1}(R^T[2][3]/R^T[3][3])\end{aligned}$$

where $R^T[i][j]$ refers to the element of the matrix R^T at row i and column j.

Direction Vectors: At each landmark point j with $\mathbf{m}_{jx}$, $\mathbf{m}_{jy}$ and $\mathbf{m}_{jz}$ as the mean coordinates we can write

$$\mathbf{m}' = \begin{pmatrix} \cos\alpha\mathbf{m}_{jx} - \sin\alpha\mathbf{m}_{jy} \\ \sin\alpha\mathbf{m}_{jx} + \cos\alpha\mathbf{m}_{jy} \\ \mathbf{m}_{jz} \end{pmatrix}$$

The first derivatives of this vector with respect to α gives

$$\mathbf{u}_\alpha = \begin{pmatrix} -\sin\alpha\mathbf{m}_{jx} - \cos\alpha\mathbf{m}_{jy} \\ \cos\alpha\mathbf{m}_{jx} - \sin\alpha\mathbf{m}_{jy} \\ 0 \end{pmatrix}$$

References

[1] Fly wing data. http://www.flywings.org.uk.

[2] Max-Planck Institute for Evolutionary Biology, Plön, Germany. http://www.evolbio.mpg.de/english/index.html.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, 19(6):716–723, 1974.

[4] F. L. Bookstein. Tensor biometrics for changes in cranial shape. *Annals of Human Biology*, 11:413–437, 1984.

[5] F. L. Bookstein. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1:181–242, 1986.

[6] F. L. Bookstein. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(2):225–243, 1997.

[7] F. L. Bookstein. Registration error and functional image analysis. In *Workshop on Biomedical Statistics*, Leeds, UK., 2001.

[8] F. L. Bookstein, D. E. Slice, P. Gunz, and P. Mitteroecker. *Anthropology Takes Control of Morphometrics*. Institute for Anthropology, University of Vienna, Vienna, Austria, 2004.

[9] P. A. Bromiley, H. Ragheb, and N. A. Thacker. The TINA geometric morphometrics toolkit. http://www.tina-vision.net/docs/memos/2010-007.pdf.

[10] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. Roy. Statist. Soc. B (Methodological)*, 39(1):1–38, 1977.

[12] J. M. Fitzpatrick, J. B. West, and C. R. Maurer. Predicting error in rigid-body point-based registration. *IEEE Trans. Medical Imaging*, 17(5):694–702, 1998.

[13] B. Frederich, S. Y. V. Liu, and C. F. Dai. Morphological and genetic divergences in a coral reef damselfish, pomacentrus coelestis. *Evolutionary Biology*, 39(1), 2012.

[14] A. Gomez-Robles, A. J. Olejniczak, M. Martinon-Torres, L. Prado-Simon, and J. M. Bermudez de Castro. Evolutionary novelties and losses in geometric morphometrics: a practical approach through Hominin Molar morphology. *Evolution*, 65(6):1772–1790, 2011.

[15] C. R. Goodall. *The Statistical Analysis of Growth in Two Dimensions*. Department of Statistics, Harvard University, USA, 1983.

[16] C. R. Goodall. Procrustes methds in the statistical analysis of shape. *J. Royal Statistical Society, Series B (Methodological)*, 53(2):285–339, 1991.

[17] P. Gunz, P. Mitteroecker, and F. L. Bookstein. Semilandmarks in three dimensions. *Modern Morphometrics in Physical Anthropology*, pages 73–98, 2005.

[18] B. Halgrimsson, D. E. Lieberman, N. M. Young, T. Parsons, and S. Wat. Evolution of covariance in mammalian skull. In *B.K. Hall & D.E. Lieberman (eds.), Novartis Foundation Symposium - Tinkering: The Microevolution of Development*, volume 284, pages 164–184. Wiley-Liss, New York, 2007.

[19] M. Hubert, P. J. Rousseeuw, and K. Branden. ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1):64–79, 2005.

[20] D. G. Kendall. Shape-manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[21] C. P. Klingenberg. Novelty and homology-free morphometrics: What's in a name? *Evolutionary Biology*, 35(3):186–190, 2008.

[22] S. Lele and J. T. Richtsmeier. Statistical models in morphometric: Are they realistic? *Systematic Zoology*, 39(1):60–69, 1990.

[23] N. A. Mantel. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209–220, 1967.

[24] P. Mitteroecker and F. Bukstein. Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, 38(1):100–114, 2011.

[25] P. Mitteroecker and P. Gunz. Advances in geometric morphometrics. *Evolutionary Biology*, 36(2):235–247, 2009.

[26] C. Oxnard and P. O'Higgins. Biology clearly needs morphometrics. does morphometrics need biology? *Biological Theory*, 4(1):84–97, 2009.

[27] P. R. Peres-Neto and D. A. Jackson. How well do multivariate data sets match? the advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, 129:169–178, 2001.

[28] S. I. Perez, V. Bernal, and P. N. Gonzalez. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *J. Anatomy*, 208(6):769–784, 2006.

[29] P. D. Polly. Developmental dynamics and G-matrices: Can morphometric spaces be used to model phenotypic evolution? *Evolutionary Biology*, 35(2):83–96, 2008.

[30] H. Ragheb and N. A. Thacker. Morphometric shape analysis with measurement covariance estimates. In *Proc. British Machine Vision Conference*, Dundee, UK., 2011.

[31] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Trans. Medical Imaging*, 20(6):526–534, 2001.

[32] D. E. Slice. Geometric morphometrics. *Annual Review of Anthropology*, 36:261–281, 2007.

[33] D. L. Theobald and D. S. Wuttke. Empirical Bayes hierarchical models for regularizing maximum Likelihood estimation in the matrix Gaussian Procrustes problem. *Proc. National Academy of Sciences of the United States of America*, 103(49):18521–18527, 2006.

[34] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *J. Roy. Statist. Soc. B*, 61(3):611–622, 1999.

[35] M. Vignon and P. Sasal. The use of geometric morphometrics in understanding shape variability of sclerotized Haptoral structures of monogeneans (Platyhelminthes) with insights into biogeographic variability. *Parasitology International*, 59(2):183–191, 2010.

[36] M. L. Zelditch, D. L. Swiderski, H. D. Sheets, and W. L. Fink. *Geometric Morphometrics for Biologists, A Primer*. Elsevier Academic Press, New York, 2004.