

# Morphometric Shape Analysis with Measurement Covariance Estimates

Hossein Ragheb  
hossein.ragheb@manchester.ac.uk  
Neil Thacker  
<http://www.tina-vision.net/~nat/>

Imaging Sciences Research Group,  
School of Cancer and Enabling Studies,  
University of Manchester, UK.

---

## 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. We discuss the properties of our approach and how measurement covariances can be considered characteristic of the local shape. Our formulation 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 fly wing, hand and face data. We use these data to explore possible advantages and disadvantages over the use of standard Procrustes/PCA analysis. In the process we show how appropriate weighting provides more efficient use of the available landmark data <sup>1</sup>.

## 1 Introduction

Linear point distribution models have been used for a variety of tasks, ranging from graphics and tracking, through to scene interpretation. The level of statistical rigor required when generating these models varies, depending upon the application, with the strongest emphasis being on scientific studies of shape, or morphometry. 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 [1], Goodall [2] and Kendall [3]. Later, Bookstein [4] 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. 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 which has provided a framework to allow data to be analysed according to a measurement process. We introduce two alternative techniques below and explain, in the discussion section, the expected benefits for morphometric analysis.

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. 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 which is described with the mathematical formalisms of iso-morphism. Because of such restrictions on the definition of landmarks, semi-landmarks were introduced [5] in order to allow inclusion of other points which are not homologous among the specimens. By this we mean that a unique corresponding location cannot be defined. Measurement at these locations must be augmented by a constraint, such as bending energy [5, 6], in order to recover the information missing due to the nature of local structure. However, it is easy to construct examples where such resolution of corresponding locations is not biologically meaningful.

From a statistical perspective a homology must be augmented by distributions indicative of the extent to which a correspondence can be established. There have been several attempts in the literature to include measurement errors for landmark points. For example, Rohr *et al.* [7] used covariance matrices in a Mahalanobis distance form for non-isotropic data. This was, however, for the purpose of image registration using splines. Fitzpatrick *et al.* [8] worked on the relationship between localisation error and registration error in rigid-body, point-based registration. Chui and Rangarajan [9] proposed a general framework for non-rigid point matching, where outliers are effectively rejected.

Text books [10] 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 [11] suggested a method which assumed a single covariance for all landmark perturbation. It has been noted that such a matrix is inestimable [12]. 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 [13] 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. We will now investigate possible generalisations of Procrustes along these lines, and show one of several ways that such a measurement covariance can be estimated. As a key issue here is the computability of these covariances, the stability of the resulting analysis is investigated using repeat mark-up data for a range of sample sizes.

## 2 The Method

Suppose that there are  $K$  shapes (2D) in our data-set and each shape vector  $\mathbf{w}_k$  contains  $N$  landmark points, i.e.  $\mathbf{w}_k = [w_{1x}, w_{1y}, w_{2x}, w_{2y}, \dots, w_{Nx}, w_{Ny}]$ . 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  $\mathbf{z}_k = [\mathbf{z}_{k1}, \mathbf{z}_{k2}, \dots, \mathbf{z}_{kN}]$  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.

We assume a common, fixed,  $2 \times 2$  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  $2 \times 2$  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. 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. Although transformation of data can be considered as a new space, it can also be interpreted as an affine re-projection. 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. In turn, this allows accurately measured structure to be encoded in the most significant eigen-vectors of the linear model. We hence transform  $\mathbf{z}_k$  to ghost points  $\mathbf{g}_k$  using the matrix  $W$  in the same coordinate system 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  $\mu_j$  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  $\lambda_{jk} = \mathbf{e}_j \cdot \mathbf{g}_k$  can be computed to best approximate  $\mathbf{z}_k$  with the model  $\mathbf{z}'_k$ ;

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

In order to find the best  $R_k, \mathbf{t}_k, s_k$  parameters for each  $k$ , we minimise a Mahalanobis distance (LC method) given by

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

This is simply the Likelihood estimate for the location of the shape given the linear model and the assumed measurement covariances. 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 [14], 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. 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<sup>2</sup>. Using this argument, if we wish to align to the mean shape

<sup>2</sup>Bookstein [15]: "Wherever there is partial registration the true value of a (vector deformation) is inaccessible."

we should use a covariance  $B$  consistent with the distribution around this mean. It therefore quantifies the entire variation of the sample as opposed to the noise in the measurement. Then the corresponding Likelihood function (LB method) is

$$\log(P_{km}) = (\mathbf{m} - \mathbf{z}_k)^T B^{-1} (\mathbf{m} - \mathbf{z}_k) \quad , \quad B = \frac{1}{K} \sum_{k=1}^K (\mathbf{m} - \mathbf{z}_k)(\mathbf{m} - \mathbf{z}_k)^T \quad (3)$$

This also stabilises the alignment process with respect to poorly localised landmarks, but unlike  $C$ ,  $B$  is independent of eigen-vector estimation. We can therefore iteratively compute the best alignment and  $B$  prior to construction of the linear shape model. We therefore have two alternative methods, LC and LB, to consider. As use of either Eq. (2) or Eq. (3) require 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  $\mathbf{e}_j$ ,  $\mathbf{m}$ ,  $C^{-1}$  (or  $B^{-1}$ ) 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;

$$C = \frac{1}{K} \sum_{k=1}^K (\mathbf{z}'_k - \mathbf{z}_k)(\mathbf{z}'_k - \mathbf{z}_k)^T \quad (4)$$

A statistical test is required to give an indication of the appropriateness of the assumed linear model for fixed  $J$ . For this work, we have used a  $\chi^2$  test based upon the construction of these corrected covariances on one data set and then used for the calculation of  $\chi^2$  for a second set.

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

$$\Delta_n C = \frac{\mathbf{f}_n^T \otimes \mathbf{f}_n}{\mathbf{f}^T C^{-1} \mathbf{f}} \quad , \quad \Delta C = \sum_{n=1}^N \Delta_n C \quad (5)$$

Note that the denominator is the change in  $\chi^2$  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

$$\Delta C = \frac{W^{-1} \mathbf{e}^T \otimes W^{-1} \mathbf{e}}{W^{-1} \mathbf{e}^T C^{-1} W^{-1} \mathbf{e}} = W^{-1} \mathbf{e}^T \otimes W^{-1} \mathbf{e} \quad (6)$$

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  $\Theta_i$  represents one of the

direction vectors of these parameters<sup>3</sup>, to orthogonalise the model, we modify ghost points as follows.

$$\mathbf{g}'_k = \mathbf{g}_k - (\mathbf{g}_k \cdot \hat{Y}_i^T) \hat{Y}_i^T, \quad \hat{Y}_i^T = W \Theta_i^T \quad (7)$$

The new  $\mathbf{g}_k$  is computed iteratively using each  $\Theta_i$  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

$$\Delta C_{\Theta} = \sum_{i=1}^4 \{(\Theta_i^T C' \Theta_i)(\Theta_i^T \times \Theta_i)\} \quad , \quad C' = C + \Delta C + \Delta C_{\Theta} \quad (8)$$

Using the above formula, the contribution to the  $\chi^2$  lost by using a scaling parameter associated with each vector  $\Theta_j$  contributes a value of unity to the  $\chi^2$  for every additional independent degree of freedom, totaling  $J + 4$ . Our method for covariance correction is therefore consistent with an Akaike correction [16]. 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 [17]. The same covariance estimates would also be obtained when using PPCA with an uninformative prior [14].

Note that while for the LC method we apply all the corrections described by Eqs. (6), (7) and (8), for the LB method we can only correct for the linear model (Eq. 6). This is due to the measurement covariances being defined in different co-ordinate frames for the two approaches. As a consequence, for the LB method, it is necessary to use a limit to stop variances becoming too small. If any variance component was found to have been limited (by the methods described above), the local 2x2 diagonal section of  $C$  was rescaled to give a  $\chi^2/DoF = 1$  (degrees of freedom) before testing on the second data 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  $4K$ .

## 4 Experiments

We experiment with several data sets of manual mark-ups. The first set corresponds to left-hand images (40 samples, 56 landmarks) of 4 individual hands with various angles between fingers [18]. The second set corresponds to face images (200 samples, 68 landmarks) taken from video frames of talking faces of 3 individuals [19]. Finally we test the stability of our method using 4 sets of repeat mark-up data (200 samples per set, 15 landmarks) from fly wings [20]. In Fig. 1, for each data set, landmarks are shown superimposed on one of the corresponding image samples. We choose the minimum number of linear components which gives a ratio  $\chi^2/DoF$  about unity. The number is 3 for fly wings, 9 for hands, and 11 for faces. We show the aligned data in Figs. 2 and 3. Here sub-figures a, b and c correspond to Procrustes, LB (Eq. 3) and LC (Eq. 2) methods, respectively.

In Figs. 4-6, we show the error bars computed using the eigen-vectors and values of the 2x2 covariance matrices. 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

<sup>3</sup>Direction vectors  $\Theta_1$  and  $\Theta_2$  corresponding to translation in x and y directions are orthogonal, i.e.  $\Theta_1 \cdot \Theta_2 = 0$ . Similarly, direction vectors  $\Theta_3$  and  $\Theta_4$  corresponding to scaling and rotation are orthogonal, and so  $\Theta_3 \cdot \Theta_4 = 0$ .

method show exactly why we cannot assume isotropic distributions for the data as assumed in Procrustes. Again, sub-figures a, b and c correspond to Procrustes, LB and LC methods.

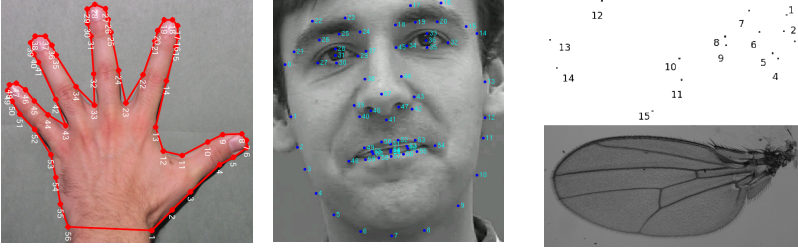


Figure 1: Typical landmarks on sample images from the hand, face and fly wing data.

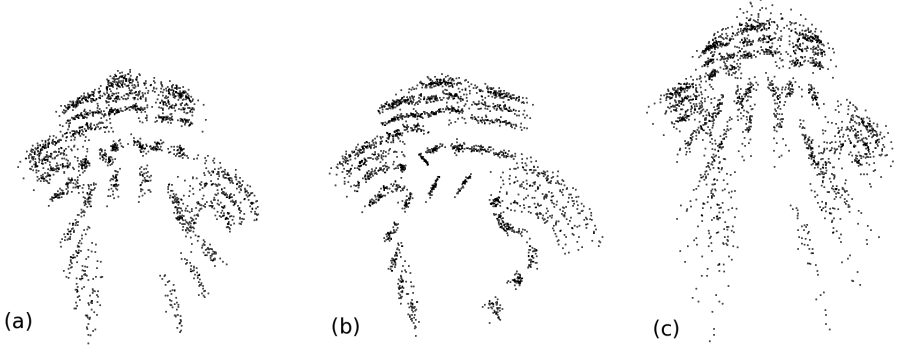


Figure 2: Hand data: aligned points using Procrustes (a), LB (b) and LC (c) methods.

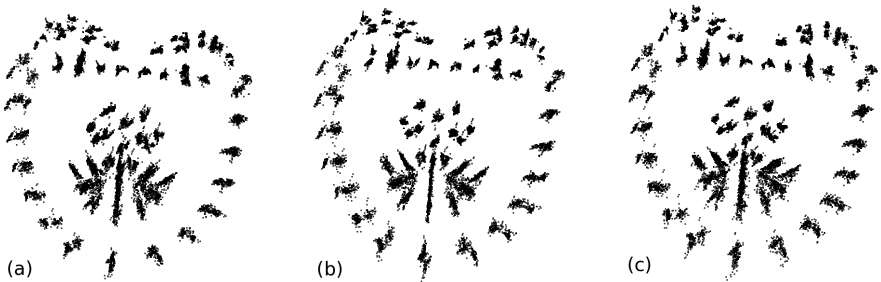


Figure 3: Face data: aligned points using Procrustes (a), LB (b) and LC (c) methods.

It is clear from Figs. 2.a and 2.b that compared to Procrustes, the LB method results in tighter groupings and better use of signal variation directions than noise. However, Fig. 2.c shows that by forcing orthogonality between the alignment and shape parameters, the LC method uses scale as a means to linearise the curvature in variation seen at the finger tips. Although this behaviour could be eliminated by forcing the use of one scale parameter for each hand in the data set, this points up the ambiguity otherwise inherent to Kendall's definition of shape, i.e. any component of biological shape variation which can be interpreted as alignment will be erased from the data before model construction.

Both Figs. 4.b and 4.c show that the largest variances correspond to the landmark points which are not well constrained or unstable to some extent. This is because they are either defined mathematically (based on two neighbouring landmarks) or pseudo landmarks with high variations in one direction. Further, most points on the sides of fingers have their largest

variation in the directions tangent to the boundary curves. From the error bars in Figs. 5.b and 5.c, it is clear that the amount of variations are quite similar on the points which are on symmetrical locations on face samples as expected. Also as expected, with talking faces, the orientation of error bars for points on the symmetry line are approximately vertical.

We do not show the aligned data corresponding to the fly wings as the three methods result in reasonable alignments and differences are not sufficiently clear from the figures. Instead, the error bars shown in Fig. 6 allow us to clearly see the differences. Specifically, it is clear that while in Fig. 6.a error bars corresponding to landmarks 12 and 15 are oriented in the wrong directions, in Figs. 6.b and 6.c these are oriented in the directions we expect based on the structures observed on the fly wings. The rest of the error bars show roughly similar orientations using the three methods. However using LB and LC methods, error bars corresponding to most landmarks exhibit elliptical distributions. We have not listed the absolute values of the 2x2 covariances here. However as expected, the values of standard deviations estimated for these data sets vary between half a pixel and a few pixels (4.4 for fly wings, 8.7 for hands, and 4.0 for faces).

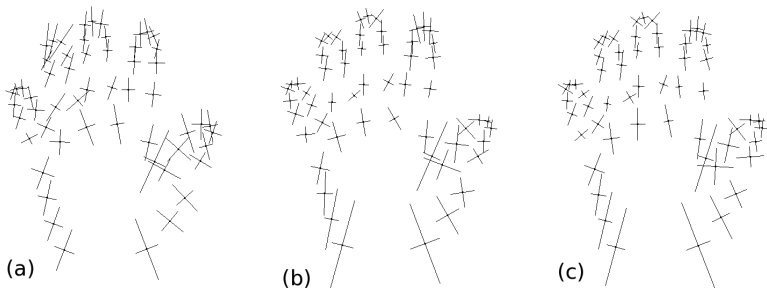


Figure 4: Hand data: error bars ( $\times 20$ ) estimated using Procrustes (a), LB (b) and LC (c) methods.

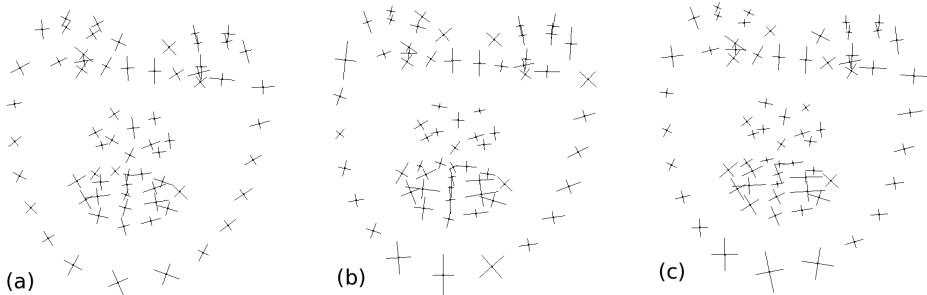


Figure 5: Face data: error bars ( $\times 20$ ) estimated using Procrustes (a), LB (b) and LC (c) methods.



Figure 6: Fly wing data: error bars ( $\times 30$ ) estimated using Procrustes (a), LB (b) and LC (c) methods.

Finally, we investigate the stability of the LB and LC methods. There are four data sets available from left and right wings (L and R) of 200 female flies (L1, L2, R1 and R2). Two images of each wing has been taken from slightly different viewing positions (1 and 2), and used for marking-up in order to do both symmetry and reproducibility tests [20]. We performed a  $\chi^2$  test based upon the construction of corrected covariances on one data set and

then used for the calculation of  $\chi^2$  for a second set. We set the statistical test for significant difference on the basis of an allowable range of  $\chi^2/DoF$  corresponding to  $\pm 4 S.D.$ , i.e.  $[0.72, 1.28]$  for 200 samples. The corresponding plot in Fig. 7 confirms the stability of LB and LC methods, as almost all  $\chi^2/DoF$  values fall in the range expected. Further  $\chi^2$  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.

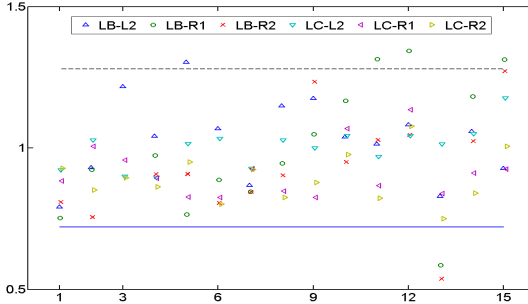


Figure 7: The  $\chi^2/DoF$  ratios when LB and LC methods are applied to 3 sets of repeat fly wing data (L2, R1, R2) using fixed covariances (Figs. 6.b and 6.c) estimated earlier from L1 set.

In Table 1, we list the Fisher Information (FI) value for every method and data set studied here. The LC method gives FI values almost twice those obtained using Procrustes. 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 half as many landmarks from the outset. In principle, we would argue that of the two alternative techniques (LB and LC), the LC method makes best use of the data. However, the LB method would also solve the major problems associated with using pseudo landmarks, and be closer to the methods currently used by the morphometric community.

FI value	Procrustes	LC	LB
Hand data	44.2	91.7	76.0
Face data	75.0	134.4	111.2
Fly wing data	20.2	28.5	48.1

Table 1: Fisher information (FI) value: the sum of inverse variances (from all landmarks); listed for the three methods and three data sets studied.

## 5 Discussion

The aim of this work was to investigate ways of incorporating poorly localised landmarks (such as pseudo landmarks) into Procrustes. The resulting approach deals with linear modelling of shape by treating the data generation process as originating from both linear correlating factors and random fluctuations. This latter process being inherent to the way that corresponding points have been defined and measured (for *any* definition of feature correspondence). The overall analysis estimates both the linear correlations and the distribution of the statistical fluctuations. By performing the analysis in this way we ensure that the assumptions regarding data behaviour are applied in a self-consistent manner during estimation of the linear model.

The expected advantages of this approach are: 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; e) finally, generalisation of the approach would seem to be possible which would support the analysis of curves.

We will now consider the objections to using a weighting process mentioned in the introduction. 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 Kendall's statistics as suggested in [10]. 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. Ghost points are just a new way to define landmark location relative to local shape structure. As a consequence, normalisation of scale in the ghost space  $\mathbf{g} + W\mathbf{m}$  will result in an equivalent behaviour suitable for Kendall's approach. We can re-project scaled shapes onto the tangent space defined in the transformed ghost space if we wish, in order to remove local curvature arising from scale normalisation <sup>4</sup>.

Far from there being no objective way to define these covariances [10, 12], there are at least three; we can estimate them directly from repeatability of measurements; they can be directly estimated via conventional statistical means when using Likelihood based landmark location (Cramer-Rao Bound (CRB)); or they can be estimated as the unexplained stochastic variation (residuals) in fitted data [21]. 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 unmodelled) behaviour of the biology itself. This suggests that valid estimates cannot be obtained without either knowledge of the model order or constraints on the expected ranges of values (i.e. the standard model selection problem). 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. The orientations and dominant directions of the error bars obtained (e.g. in Figs. 6.b and 6.c) agree well with what we expect. This is an indication that the observed variances are being driven by the spatial information encoded in local structure.

Use of a least-squares measure (which assumes isotropic errors) does not define a pure measurement of biology, as we should not interpret a fixed difference in the millimetre position of a landmark as a quantitative biological measure, invariant to its location in the organism. For example, well measured structures at small scales can have important correlations with variations in near-by structures at larger scales. A Procrustes based approach is certain to mask small scale effects with any large scale variation. Whereas covariance weighted analysis would identify these correlations if the measurement is appropriate. Also, re-weighting of data using a covariance is statistically equivalent to modifying the information available by changing the specified set of landmarks. If we object to using measurement covariances then we should also object to the arbitrary initial choice of landmarks. When viewed like this there should be no objection to defining measured locations so that their statistical properties accord with the way we intend to analyse the data (i.e. PCA). We need to be careful in both cases that we do not over-interpret the results of any analysis and any intrinsic relevance to biology.

---

<sup>4</sup>We note this while at the same time pointing out that few researchers seem to use Kendall's methods in practice.

## 6 Conclusions

In this work we have associated the problems of working with pseudo landmarks in biological (and anthropological) 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.

In use, results should be no less meaningful as a starting point for biological interpretation than those of Procrustes. Though, by adhering more closely to a Likelihood interpretation, they should provide a better summary of the measured data, demonstrated on data sets here to be equivalent to doubling the quantity of data. Our results also indicate that measurement covariances can be reliably estimated in our data for small sample sizes ( $\approx 40$ ), contradicting the common view that this should be impossible.

Although our model has been defined for independent landmarks, it would be possible to define the covariance matrix  $C$  so as to include (off diagonal) measurement correlations along extended strings, so supporting general shape analysis. The theory presented here can thus be considered as an extension to both Procrustes based shape analysis and active shape models, opening up a way for the more general analysis of curves and surfaces.

## References

- [1] Bookstein F. L., Tensor Biometrics for Changes in Cranial Shape, *Ann. Human Biol.*, 11:413-437, 1984.
- [2] Goodall C. R., The Statistical Analysis of Growth in Two Dimensions, Doctoral dissertation, *Department of Statistics, Harvard University*, 1983.
- [3] Kendall D. G., Shape-Manifolds, Procrustean Metrics, and Complex Projective Spaces, *Bull. London Math. Soc.*, 16:81-121, 1984.
- [4] Bookstein F. L., Size and Shape Spaces for Landmark Data in Two Dimensions, *Statist. Sci.*, 1:181-242, 1986.
- [5] Gunz P., Mitteroecker P. and Bookstein F. L., Semilandmarks in Three Dimensions, in *Modern Morphometrics: Physical Anthropology*, Kluwer Academic/Plenum Publishers, N.Y., 73-98, 2005.
- [6] Bookstein F. L., Landmark Methods for Forms Without Landmarks: Morphometrics of Group Differences in Outline Shape, *Med. Image. Anal.*, 1(3):225-243, 1997.
- [7] Rohr K., Stiehl H. S., Sprengel R., Buzug T. M., Weese J. and Kuhn M. H., Landmark-Based Elastic Registration using Approximating Thin-Plate Splines, *IEEE T. Med. Imag.*, 20(6):526-534, 2001.
- [8] West J. B., Fitzpatrick J. M. and Maurer C. R., Predicting Error in Rigid-Body Point-Based Registration, *IEEE Trans. Medical Imaging*, 17(5):694-702, 1998.
- [9] Chui H. and Rangarajan A., A New Point Matching Algorithm for Non-Rigid Registration, *Computer Vision and Image Understanding*, 89(2-3):114-141, 2003.
- [10] Zelditch M. L., Swiderski D. L., Sheets H. D. and Fink W. L., *Geometric Morphometrics for Biologists A Primer*, Elsevier Academic Press, New York, 2004.

- 
- [11] Goodall C. R., Procrustes Methods in the Statistical Analysis of Shape, *J. Roy. Stat. Soc. Series B (Methodological)*, 53(2):285-339, 1991.
- [12] Lele S. and Richtsmeier J. T., Statistical Models in Morphometric: Are They Realistic? *Syst. Zool.*, 39(1):60-69, 1990.
- [13] Theobald D. L. and Wuttke D. S., Empirical Bayes Hierarchical Models for Regularizing Maximum Likelihood Estimation in the Matrix Gaussian Procrustes Problem, *Proc. Nat. Acad. Sci. USA*, 103(49):18521-18527, 2006.
- [14] Tipping M. E. and Bishop C. M., Probabilistic Principal Component Analysis, *J. R. Statist. Soc. B*, 61(3):611-622, 1999.
- [15] Bookstein F. L., Registration Error and Functional Image Analysis, *Workshop on Biomedical Statistics*, Leeds, 2001.
- [16] Akaike H., A New Look at the Statistical Model Identification, *IEEE Trans. Automatic Control*, 19(6):716-723, 1974.
- [17] Dempster A. P., Laird N. M. and Rubin D. B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Statist. Soc. B (Methodological)*, 39(1):1-38, 1977.
- [18] Hand Data, <http://www2.imm.dtu.dk/~aam/>.
- [19] Face Data,  
[www.isbe.man.ac.uk/~bim/data/talking\\_face/talking\\_face.html](http://www.isbe.man.ac.uk/~bim/data/talking_face/talking_face.html).
- [20] Fly Wing Data, <http://www.flywings.org.uk>.
- [21] Ragheb H. and Thacker N. A., Quantitative Localisation of Manually Defined Landmarks, *Proc. Medical Image Understanding and Analysis*, London, UK, 2011.