MORPHOMETRIC SPACES, SHAPE COMPONENTS AND THE EFFECTS OF LINEAR TRANSFORMATIONS

F. James Rohlf

Department of Ecology and Evolution
State University of New York
Stony Brook, New York 11794-5245

ABSTRACT

The various types of mathematical spaces used in morphometrics are reviewed with emphasis given to the distinctions between the physical space of the organism, shape space and tangent spaces. The effects of linear transformations of the coordinates of the landmarks (on the specimens or the reference) on various statistical analyses and estimates of the uniform and nonaffine components of shape variation are presented. It is shown why statistical analyses of the estimates of the nonaffine components of shape variation are sensitive to affine transformations of the reference or the specimens (which may seem counterintuitive). Some implications of these results on the choices of methods for the study of shape are discussed.

INTRODUCTION

There has been considerable interest in recent years in geometrically based methods for the statistical study of shape variation. A short overview of the new field is given by Rohlf and Marcus (1993) and Rohlf (1990a). A comprehensive overview is given in the text by Bookstein (1991). Rohlf and Bookstein (1990) and Marcus et al. (1993) are proceedings of morphometrics workshops. In order to apply these new methods properly it is necessary that one understand some of the theory behind the methods. This paper discusses some aspects that are often misunderstood.

There has also been important theoretical work on various mathematical spaces that are useful for studying the statistics of shape change (e.g., Goodall, 1991). These spaces start with the assumption that information on shape has been captured by recording the coordinates of the positions of landmarks on an organism. The coordinates are two or three dimensional corresponding to the physical space in which the coordinates of the organism's landmarks were recorded. Although coordinate data contain some information not relevant

Advances in Morphometrics, Edited by L. F. Marcus et al.
Plenum Press, New York, 1996

to the study of shape (e.g., location and orientation of the specimen), such data also contain information sufficient to allow the computation of any possible mathematical index that might be proposed to describe the positions of the landmarks relative to one another. Thus working with coordinates is not a limitation. There has also been important work on the statistical distributions of points in these spaces (e.g., Kent; 1991, Goodall, 1992a).

Several different approaches are being used in biology to study shape. One of the purposes of the present note is to summarize some of the most important terms used in shape statistics in an accessible form to clarify some of the interrelationships among morphometric methods. It is not the intention of this paper to review all of the approaches currently being used (see Rohlf, 1990a, or Rohlf and Marcus, 1993, for more general reviews).

A major purpose of this paper is to point out some of the statistical consequences of different choices of reference configurations and linear transformations of the coordinates of the landmarks of the reference or the original organisms. In order to do this, it is necessary to review briefly some important equations used in the computations for each method. Knowing what manipulations of the data do or do not affect the results helps one interpret the results of a statistical method.

The notation used by other authors was adjusted in order to avoid duplication of symbols (especially the symbols α and $\mathbf{W}$) within this paper and to be compatible with the notation used in Rohlf (1993a). The letter "t" when used as a superscript to a matrix indicates matrix transposition.

MORPHOMETRIC SPACES

Goodall (1991) presents a set of definitions for the various multidimensional spaces that are used in morphometrics. In all of these spaces the ith object (an individual specimen) consists of k coordinates in the physical space of the organism (x,y or x,y,z for two or three dimensions, respectively) at each of p landmarks, corresponding to the ith point in a multidimensional space. Although only some of these spaces will be used in most practical morphometric studies, an understanding of relationships between them and of their geometric properties is needed for understanding the new morphometric methods.

The original coordinates of the landmarks of an object (arbitrarily positioned in the digitizing plane or volume) characterize a **figure space**. This is a space of pk dimensions because the location of each of the p landmarks is described by k coordinates in the k-dimensional physical space of the organism.

If the n objects are translated in the plane or volume so that their centroids are superimposed, then the coordinates of a centered object corresponds to point in a **preform space**. This space is of $pk-k$ dimensions because the k coordinates of the centroid have been fixed for each object.

If the objects are translated and rotated (but not reflected or scaled) so that they superimpose optimally according to some criterion, then the resulting coordinates characterize a **form space** of $pk-k-k(k-1)/2$ dimensions. In two dimensions, one angle of rotation is fixed, and in three dimensions three angles are fixed. Note that Lele (1993) allows reflections in his definition of form space and gives a different number of dimensions (see below).

On the other hand, if one translates the objects to the origin and scales the objects to unit centroid size (the square root of the sum of squared distances of all landmarks to the centroid of the object, Bookstein, 1991) but does not perform any rotations or reflections, the resulting coordinates characterize what is called a **preshape space** of $pk-k-1$ dimensions. Preshapes that differ only by rotation are said to belong to the same fiber (the closed path of preshapes defined by all possible rotations of the object in the figure space).

Finally, if the objects are centered on the origin, scaled to unit centroid size and optimally rotated (but not reflected) so as to minimize the sum of the squared distances between homologous landmarks of a pair of objects (the square of the Procrustes distance between them), then an object corresponds to a point in Kendall's (1984, 1986) **shape space**. This space is of $pk-k-k(k-1)/2-1$ dimensions and is denoted by the symbol Σ_k^p (for p landmarks on a k-dimensional object). Each point in shape space corresponds to a fiber in preshape space. Shape space is non-Euclidean because Procrustes distance is used as the measure of distance between points in this space. The Procrustes distance between two shapes is the minimum distance between any two preshapes on the fibers to which the two shapes belong.

For three points in the plane, shape space, Σ_2^3 can be visualized as the surface of a sphere. Kendall (1986) gives an illustration and refers to it as a "spherical blackboard." Procrustes distances correspond to chord distances in this space. Some workers (e.g., Kendall, 1984) prefer to work with geodesic distances along great circles on the surface of the sphere. It is easy to convert back and forth between these two kinds of distances without any loss of information. In general, for p landmarks in the plane, shape space is the complex (p-2)-torus constructed from the Cartesian product of spheres with radius 1/2. The mathematical properties of this shape space are surprisingly complex given that it is generated as a consequence of pursuing the simple idea of using Procrustes distance between pairs of shapes. Shape spaces for $k = 3$ dimensions are "substantially more complicated" (Goodall, 1992b).

An alternative approach to the study of shape is to approximate the non-Euclidean shape space by a **tangent space** that has a Euclidean geometry. This space is made up of the projections of the objects in shape space onto a linear vector space that is tangent to shape space. The two spaces intersect at a point corresponding to the reference object, which can be an estimated mean (consensus) configuration from a Procrustes analysis. For three landmarks and two dimensions the tangent space is an ordinary plane that touches the surface of a sphere, corresponding to shape space, at the point which corresponds to the reference.

The shape variation captured in the linear tangent space is not limited to just linear changes in shape (uniform shape change, see below), as all shape variation is embedded in the tangent space. There is no loss of information as in a projection into a lower dimensional space. Every point in shape space corresponds to a point in the tangent space (and vice versa). For objects close to the reference, the Euclidean distances between pairs of objects in the tangent space will be good approximations to geodesic distances between points in shape space. Statistical analyses based on variation in the tangent space are expected to reflect accurately variation in shape space for only "small" variation in shape around the reference object. Bookstein (1991) gives some guidelines of what constitutes "small" variation in shape for triangles of landmarks. One can conjecture that when there are more than just a few landmarks, the usual magnitude of biological variation in shape should fill only a relatively small region of shape space and thus standard multivariate statistical analyses (such as described by Marcus, 1990, or Krzanowski, 1988) in the tangent space should be satisfactory.

THE THIN-PLATE SPLINE

Bookstein's (1989) use of thin-plate splines corresponds to one method for visualizing a tangent space for the statistical analysis of shape variation. The thin-plate spline function is a smooth function (it is twice differentiable) that maps all points in the physical space of the reference onto corresponding points in the space of the ith specimen. For completeness, basic aspects will be reviewed here. Bookstein (1991) gives a general description of thin-plate splines. The TPSPLINE program (Rohlf, 1990b) performs the necessary computations and plots the differences between two configurations of landmarks as a transformation grid based on the thin-plate spline.

For two dimensions the thin-plate spline function can be written as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix} + \sum_{j=1}^{p} \mathbf{n}_j \, \mathrm{U}(r_j) \,, \tag{1}$$

where x and y are the coordinates of any point in the space of the reference, x' and y' are the resultant coordinates of a point in the space of the ith specimen, $\mathbf{A}$ is a $k \times (k+1)$ matrix of parameters that specify an affine transformation (translation, rotation, scale and shear), $\mathrm{U}(r_j)$ is the function $r_j^2 \ln r_j^2$, r_j is the distance between a point (x,y) in the space of the reference and the jth landmark in the reference, and the $\mathbf{n}_j$ are k-dimensional vectors of parameters that determine the nonaffine deformation of the space. For three dimensions the function $\mathrm{U}(r_j)$ is defined to be simply r_j. Points corresponding to the landmarks in the reference are mapped exactly to the locations of the homologous landmarks in the ith specimen. Intermediate points in other locations within the physical space of the reference are mapped to some mathematically homologous location within the physical space of the ith specimen. Because the only information used to compute a thin-plate spline is the locations of the landmarks and because this function is based on the physical properties of a thin sheet of metal rather than of biological structures, one should not take the implied homologies of intermediate points too seriously. The details of the transformation grids are very suggestive, but they are just a way of expressing the relative displacements of the landmarks. One of the most important properties is that the thin-plate spline function leads to convenient orthogonal vectors which span the tangent space and thus allows one to visualize all possible differences in shape that affect the positions of the landmarks. It is also very convenient that these vectors correspond to smooth functions. Thus any statistical analysis that expresses its results in terms of linear combinations of variables can be illustrated as a deformation in the style of a Thompson (1917) transformation grid.

The $k(k+1)+kp$ parameters needed to transform the coordinates of the landmarks in the reference into those of the ith specimen can be computed as follows:

$$\mathbf{A}_i = \mathbf{X}_i \, \mathbf{L}_q^{-1}$$

$$\mathbf{N}_i = \mathbf{X}_i \, \mathbf{L}_p^{-1\prime} \tag{2}$$

where $\mathbf{X}_i$ is the $k \times p$ matrix of coordinates of the ith specimen; L_p^{-1} is the upper left $p \times p$ block of the inverse of the $\mathbf{L}$ matrix (as defined in Bookstein, 1989, and Rohlf, 1993a) and is called the **bending energy matrix;** L_q^{-1} is the upper right $p \times (k+1)$ block of the inverse of $\mathbf{L}$; and $\mathbf{N}_i$ is a $k \times p$ matrix whose columns correspond to the $\mathbf{n}_j$ vectors in equation (1). Note that, for a fixed reference, the parameters, $\mathbf{A}_i$ and $\mathbf{N}_i$, are simply linear combinations of the coordinates of the ith specimen.

The different parameters correspond to differences between the reference and a given object at different geometric scales. The first $k(k+1)$ parameters, the elements of matrix $\mathbf{A}_i$, describe the differences in terms of an affine transformation. This includes translation, scale, rotation and shearing. The effects of an affine transformation are of infinite scale since the effects of these changes cannot be localized to any particular region of an organism. Only shearing is a component of shape. It corresponds to a uniform stretching or compression and possibly reflection, of an object in a particular direction. Bookstein (1991) refers to this as the **uniform shape component**. A shear can be specified in several ways. One of the easiest to visualize is in terms of **principal strains**—a set of orthogonal directions along which the object is uniformly stretched or compressed by some factor. For two-dimensional data, there is a direction of maximum stretching and a direction of minimum stretching at right angles to it. The ratio of the factors by which the object is stretched in these two directions is a dimensionless quantity called **anisotropy**. See Bookstein (1991) for a more detailed discussion with examples. For three-dimensional data, a shearing of the space can be described as stretching or compression along a set of three orthogonal directions.

The kp nonaffine parameters, $\mathbf{N}_i$, describe the differences between an object and the reference as the sum of nonlinear deformations of an object. These correspond to local regions of expansion, compression, bending, and so on. The subspace defined by these parameters is actually of rank $kp-k-k^2$ because translation, rotation, scaling and shearing have been fixed. This subspace can in turn be decomposed into $p-k-1$ k-dimensional geometrically orthogonal components (called partial warps, Bookstein, 1989, see below) corresponding to deformations along the x, y (and possibly z) coordinate axes at different geometric scales. Small-scale variation corresponds to changes in the relative positions of landmarks that are close together in the reference. Large-scale (but not infinite-scale) variation correspond to classic, evenly graded growth gradients.

The **principal warps** (Bookstein, 1989) are a set of eigenvectors that span the tangent space defined by the nonaffine components of the thin-plate spline. They are computed from the following decomposition of the bending energy matrix:

$$\mathbf{L}_p^{-1} = \mathbf{E}\Lambda\mathbf{E}' \,, \tag{3}$$

where $\mathbf{E}$ is a $p \times p$ matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues (called bending energies, but it is better to interpret these simply as inverse measures of scale). The last $k+1$ eigenvalues are equal to 0 because their corresponding eigenvectors span the part of the tangent space pertaining to affine variation (which is of infinite scale). The principal warps correspond to the first $p-k-1$ column vectors of $\mathbf{E}$. The vectors are orthogonal because they are the eigenvectors of a real symmetric matrix.

One can express the thin-plate spline parameters of each specimen in terms of these principal warps rather than in terms of the original coordinates of its landmarks as in equation (2). This yields what are called **partial warps** (Bookstein, 1989). These are the principal warps applied separately along each coordinate axis and scaled appropriately so as to correspond to the nonaffine part of the thin-plate spline function that transforms the coordinates of the reference, $\mathbf{X}_c$, into those of a particular specimen, $\mathbf{X}_i$. The inner product of each row of $\mathbf{V}_i = \mathbf{X}_i - \mathbf{X}_c$ with the $p-k-1$ columns of $\mathbf{E}$ that correspond to eigenvalues greater than 0 yield what Rohlf (1993a) calls **partial warp scores**. They express the nonaffine shape differences between the reference and the ith specimen in terms of a new set of variables. They are computed as follows:

$$\mathbf{W}_i = \frac{1}{\sqrt{n}}\,\mathbf{V}_i\mathbf{E}\Lambda^{-\alpha/2} \tag{4}$$

where the division by $\sqrt{n}$ and the subtraction of $\mathbf{X}_c$ are for convenience in describing the variation in shape with a sample of n specimens; the multiplication by $\Lambda^{-\alpha/2}$ allows the possibility (depending on the value of α, see below) of different forms of weighting the principal warps according to geometric scale. Note that it is sometimes useful to retain all the columns in matrix $\mathbf{E}$ so that all variation captured by $\mathbf{X}_i$ is preserved in the $\mathbf{W}_i$ matrix. The additional columns of $\mathbf{W}_i$ are not partial warps because they describe affine differences not local deformations. Sometimes other information, such as an estimate of the uniform shape component, is appended to the matrix of partial warp scores.

The $\mathbf{W}_i$ matrix for the ith specimen has k rows and either p or $p-k-1$ columns depending on whether the last $k+1$ columns of $\mathbf{E}$ (the principal warps with eigenvalues equal to 0) are retained or not. If they are not retained then this equation can be simplified to

$$\mathbf{W}_i = \frac{1}{\sqrt{n}} \mathbf{X}_i \mathbf{E} \Lambda^{-\alpha/2} \tag{5}$$

because $\mathbf{X}_c\mathbf{E} = 0$ for the first p-k-1 columns of $\mathbf{E}$ (the principal warps were constructed to be orthogonal to the reference configuration). Although $\mathbf{W}_i$ is a linear function of $\mathbf{X}_i$, it specifies the nonlinear deformations needed to transform the physical space of the reference so that the locations of its landmarks are as in specimen i.

ESTIMATION OF THE UNIFORM COMPONENT OF SHAPE VARIATION

There are a number of methods to estimate the uniform component of shape variation. Bookstein and Sampson (1990) and Bookstein (1991) proposed two closely related methods for estimating the uniform component of shape change between a given specimen and reference specimen for two-dimensional data. Both methods use the following equation to compute the vector of parameters:

$$\mathbf{u} = (\mathbf{M}' \mathbf{S}^{-1} \mathbf{M})^{-1} \mathbf{M}' \mathbf{S}^{-1} \Delta\mathbf{V}\,, \tag{6}$$

where

$$\mathbf{M}' = \begin{bmatrix} v_{23} & 0 & v_{24} & 0 & \dots & v_{2k} & 0 \\ 0 & v_{23} & 0 & v_{24} & \dots & 0 & v_{2k} \end{bmatrix}$$

is a matrix of the y-coordinates of the shape coordinates (Bookstein, 1986) of landmarks 3 ... p in the reference (assuming a baseline defined by landmarks 1 and 2). The vector $\Delta\mathbf{V}$ contains the differences in x and y shape coordinates between the given specimen and the reference (strung out as a single column with $2k$–4 rows). The two methods differ in the estimation of the $(2k-4)\times(2k-4)$ variance–covariance matrix $\mathbf{S}$. For the method that minimizes Procrustes distance, $\mathbf{S}$ is the expected covariance among landmarks based on their geometrical position in the reference and assuming only circular digitizing noise with the same variance at each landmark. For their method that minimizes Mahalanobis distance, $\mathbf{S}$ is estimated from the observed variation in a sample of specimens. These methods are implemented in Bookstein's computer program PROJECT (distributed with Rohlf and Bookstein, 1990).

Rohlf (1993a) estimates the uniform shape component based on the affine parameters of the thin-plate spline. This has the advantage of extending the mathematically elegant decomposition of nonaffine differences in shape described in the previous section to include a decomposition of the affine differences into uniform shape change versus nonshape differences (translation, rotation and scale). Unfortunately, as pointed out by Bookstein ("Combining" article, this volume), the thin-plate spline computations weight differences inversely as a function of distances between landmarks in the reference. This means, for example, that shape differences involving a pair of landmarks that are very close together in the reference will be weighted very heavily compared with shape differences involving larger regions on the reference. This weighting seems inappropriate for estimating uniform shape differences (or the other affine components), for which larger scale differences should have the higher weight. The problem can be seen by the counterintuitive superimpositions one sometimes obtains using the "minimum energy superimposition" display option in the TPSRW program (Rohlf, 1992) on data in which some landmarks are very close together in the reference.

It is also possible to estimate the uniform component of shape variation as the "nuisance parameters" of affine superimposition methods. These superimposition techniques are described by Rohlf and Slice (1990a) and by Slice ("Three dimensional" article, this volume). In the least-squares approach (Generalized Affine Least Squares, GALS, superimposition) one estimates translation, rotation, scale and the uniform shape components so as to minimize the residual differences between two specimens (or between a specimen and an average consensus configuration of landmarks).

Bookstein ("Standard formula" article, this volume) proposes a linearized Procrustes method to estimate the uniform shape change component. He derives it to be equivalent to that obtained from GALS by Rohlf and Slice (1990a). For simplicity in notation, it is convenient to translate the reference (Generalized Least Squares, GLS, consensus configuration) to the origin, scale it to unit centroid size and then rotate it to its principal axes. Then in two dimensions $\Sigma xy = 0$, $\Sigma x^2 + \Sigma y^2 = 1$ and $\Sigma x = \Sigma y = 0$, where x and y are the x and y-coordinates of the reference after the centering, scaling and rotation operations described above are performed. Next use orthogonal least-squares Procrustes analysis (Sneath, 1967, Rohlf and Slice, 1990a, Goodall, 1991) to superimpose the ith specimen on the rotated reference and let Δx and Δy denote the differences between the new coordinates of the specimen and those of the rotated reference.

The uniform shape component for the ith object is then estimated as

$$u_1 = (\Sigma\, y_i\, \Delta x_i + \lambda_2\, \Sigma\, (x_i \Delta y_i - y_i \Delta x_i)) \,/ \sqrt{\lambda_1 \lambda_2}$$

$$u_2 = \Sigma\, y_i \Delta y_i \,/\, \sqrt{\lambda_1 \lambda_2} \qquad (7)$$

where $\lambda_1 = \Sigma\, x_i^2$ and $\lambda_2 = \Sigma\, y_i^2$.

In two dimensions we can use u_1 and u_2 as uniform shape descriptors for statistical analyses. For a fixed reference, u_1 and u_2 are linear combinations of Δx's and Δy's. They will approximately follow a bivariate normal distribution if the deviations of the specimens from the reference are normally distributed landmark by landmark. The geometric meaning of u_1 and u_2 can be visualized by using the following equation. It describes the effect of a uniform shape transformation on the coordinates of landmarks in the reference.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & u_1 \\ 0 & 1 + u_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad (8)$$

where x and y are, as above, the centered, scaled and rotated coordinates of the reference and x' and y' are the coordinates of the reference after the shearing operation defined by the parameters u_1 and u_2 is applied.

RELATIVE WARPS ANALYSIS

In order to study variation in shape of a sample of n specimens, it is convenient to collect the $\mathbf{W}_i$ (partial warp scores, equation (4 above) into a single matrix $\mathbf{W}$. This new matrix has n rows corresponding to the n specimens and columns corresponding to the partial warp scores for each specimen strung out as a single row (e. g., all of the x projections followed by the y projections and, possibly, the z projections). There will be $k(p-k-1)$ columns (or more if additional columns are appended). This can be done by extension of equation (4) to

$$\mathbf{W} = \frac{1}{\sqrt{n}} \mathbf{V} \left(\mathbf{I}_k \otimes \mathbf{E}\Lambda^{-\alpha/2}\right) \tag{9}$$

where the symbol $\otimes$ denotes a direct (Kronecker) product of two matrices, $\mathbf{I}_k$ is the $k \times k$ identity matrix, and $\mathbf{V}$ is an $n \times kp$ matrix of the differences between each object and the reference strung out along each row (all of the differences in x-coordinates followed by those for the y-coordinates and, possibly, the z-coordinates). The Λ and $\mathbf{E}$ matrices and the scalar α are as defined in equation (4).

Relative warps analysis (Bookstein, 1989, 1991, Rohlf, 1993a) is a principal components analysis (PCA) of variation in shape in the tangent space as described by the **W** matrix. If $\alpha \neq 0$ then it is a weighted PCA with the weights a function of the geometric scale of each component of shape variation. If $\alpha = 0$ then relative warp analysis is consistent with the Procrustes metric for measuring the amount of difference between different shapes. Even though the principal warps are geometrically orthogonal, projections of a sample of specimens onto the principal warps are usually correlated. The purpose of a PCA is to describe the major trends of shape variation in as few statistically orthogonal dimensions as possible.

The principal component vectors of the **W** matrix are called **relative warps**. They are often expressed in terms of displacements of landmarks. They can also be visualized as a transformation grid showing deformations of the physical space of the reference configuration; this is done by expressing each relative warp as a thin-plate spline (examples are given in Rohlf, 1993a, and Rohlf and Marcus, 1993).

EFFECTS OF DIFFERENT CHOICES OF A REFERENCE

Different choices of a reference result in different physical distances between the landmarks in the reference. This leads to different **E** and **V** matrices and thus to different **W** matrices. However, if α is set equal to 0 in equation (9), as suggested by Rohlf (1993a) for exploratory studies, then it is possible to retain all p principal warps in matrix **E**. The **E** matrix will be an orthonormal matrix and different choices of a reference will result in **E** matrices that differ only by a rigid rotation. Thus any analyses of the distance relationships between objects (e.g., cluster analyses or principal coordinates analyses of distance matrices) based on the resulting **W** matrices will differ only as a result of the effect of subtracting off a different reference in computing the **V** matrix. Otherwise, the results will be identical because distances are invariant under rigid rotations.

If, as is usually the case, only the first p-k-1 principal warps are used (those with $\lambda_i > 0$) then the **W** matrix describes just the nonaffine part of the variation in shape among the objects (the part describable by local deformations). This can be visualized as a projection from the kp-dimensional tangent space (whose actual rank is $kp-k-k-k(k-1)/2-1$) into the $k(p-k-1)$-dimensional subspace of the tangent space that corresponds to local deformations. Different choices of a reference result in different principal warps and hence different partitionings of the tangent space into affine versus nonaffine components. Different partitionings lead to different distances among the objects. Distances in the subspace corresponding to nonaffine shape variation will be equal to or less than those obtained when all kp dimensions are used. If $\alpha > 0$ then different choices of a reference also affects what one considers to be large- versus small-scale differences in shape. This is because scale is measured by the reciprocals of the eigenvalues, λ_i. Thus, statistical analyses based on the partial warp scores (equation 9) will give numerically different results if different references are used. Analyses based on similar references should, however, give similar results and very similar statistical conclusions.

EFFECTS OF LINEAR TRANSFORMATIONS OF THE REFERENCE

If we consider only the nonaffine components of the thin-plate spline (i.e., matrix **E** is $p\times(p-k-1)$) then we can use equation (5) to compute the partial warp scores for specimen i. The effects of rotation, scaling and shearing of the reference can be represented as a premultiplication of $\mathbf{X}_c$ by a $k\times k$ matrix **T**. If **T** is orthonormal then using $\mathbf{TX}_c$ as the reference (but keeping the original $\mathbf{X}_i$) will have no effect on the $\mathbf{W}_i$ because the bending energy matrix is invariant to rotation of the reference. If **T** is not orthonormal then using $\mathbf{TX}_c$ shears the reference, which requires a recomputation of the bending energy matrix because distances between landmarks in the reference will change. This leads to new matrices **E** and Λ, which result in new $\mathbf{W}_i$ matrices that will differ in complex ways from those obtained originally. Thus an affine transformation of the reference affects the estimates of the non-affine components of the $\mathbf{W}_i$ through the effect of the transformation on the bending energy matrix.

Translation of the reference by a k-dimensional vector **t** will have no effect on the $\mathbf{W}_i$ because

$$\left(\mathbf{t}\mathbf{1}_p^t + \mathrm{X}_c\right)\mathbf{E} = \mathbf{X}_c\,\mathbf{E} \tag{10}$$

Translation, rotation and scaling the reference will not affect the estimate of the anisotropy obtained from the linearized Procrustes method, equation (7), because the method eliminates the effects of these transformations initially. Bookstein and Sampson's (1990) and Bookstein's (1991) methods for estimating the uniform shape component by minimizing either Procrustes or Mahalanobis distances are not influenced by the alignment of the reference because this method is based on the use of Bookstein shape coordinates.

EFFECTS OF LINEAR TRANSFORMATIONS OF THE OBJECTS

Because a superimposition method such as a generalized Procrustes analysis is usually used to construct the reference as an average of the objects after alignment, a shearing of the objects leads to a change in the estimate of the reference (the consequences of which are described above). In this section we will, however, treat the reference as fixed.

Based on the relationship given in equation (5), the effect on $\mathbf{W}_i$ of a linear transformation of the ith object can be expressed as

$$\frac{1}{\sqrt{n}}\mathbf{TX}_i\mathbf{E}\Lambda^{-\alpha/2} = \mathbf{TW}_i \tag{11}$$

where **T** is a $k\times k$ transformation matrix. Thus any effects of the premultiplication by **T** (scaling, rotation, or shearing) are passed directly on to the matrix of partial warp scores. This affects both the affine and the nonaffine components. If **T** is an orthonormal matrix then the space defined by the partial warp scores will be rigidly rotated and the distances between the objects will be unchanged. Thus an ordination of the objects produced in a relative warps analysis will be unchanged if **T** is orthonormal. If **T** is not orthonormal then the results of a relative warps analysis will be affected because the specimens will have changed shape as a result of being sheared. On the other hand, statistical analyses such as MANOVA, canonical variates analysis, multiple regression analysis or the computation of Mahalanobis generalized distances are invariant to the effects of premultiplication by **T** as long as it is of full rank, k.

If instead of a constant matrix **T** one multiplies the ith specimen by $\mathbf{T}_i$ then the ith transformation effects will also be passed along to the corresponding rows of the matrix of partial warp scores. This will change the results that one expects to obtain in an ordination of the specimens or any other multivariate analysis applied to the **W** matrix. Therefore the specimens should be aligned in some consistent way (e.g., by an orthogonal Procrustes analysis) before the matrix of partial warp scores is computed. The use of superimposition programs such as GRF (Rohlf and Slice, 1990b), GRF-ND (Slice, 1993b), or Morphometrika (Walker, 1994) is recommended in order to align the specimens before using programs for relative warp analysis (e.g., TPSRW, Rohlf, 1992, or NTSYS-pc, Rohlf, 1993b), partial warp regression analysis (TPSREGR, Rohlf, 1993c) or for the estimation of the uniform component (Bookstein, "Standard formula" article, this volume). Different alignments of the specimens (even when the same reference is used) will lead to numerically different results for analyses based on the **W** matrix. Empirically (J. A. Walker, personal communication), the differences are usually minor and the biological and statistical conclusions unaffected between the usual methods of least-squares and resistant fitting.

The nonaffine components of **W** are not influenced by a translation of the objects by a vector **t** (or even $\mathbf{t}_i$) due to the same type of relationship shown in equation (10).

Translation, rotation and scaling of the specimens will not affect the estimates of anisotropy obtained from the linearized Procrustes method, equation, because that method first aligns the specimens to the reference by means of an orthogonal Procrustes analysis. Shearing the original specimens will of course affect any estimate of the uniform shape component. The estimated direction of the principal strain for each specimen will be affected because it is expressed relative to the initial orientation of each object. Bookstein and Sampson's (1990) and Bookstein's (1991) methods for estimating the uniform shape component by minimizing either Procrustes or Mahalanobis distances will not be affected because those methods are based upon the use of shape coordinates.

EFFECTS OF DIFFERENT CHOICES OF VALUES FOR α

The effect of different choices of values for α in the computation of the partial warp scores, **W**, is to determine how the matrix of eigenvalues, Λ, of the bending energy matrix is used to weight the principal warps. Because the diagonal matrix $\Lambda^{-\alpha/2}$ is the right-hand factor in equation (9), its effect is just to weight the principal warps. Although α is a continuous variable, it is most commonly set to either 0 (to weight the principal warps equally) or 1 (to weight them by the reciprocals of the square roots of their eigenvalues so that large scale features in the reference are given greater weight).

Multivariate statistical analyses using **W** as a "data matrix" will give numerically different results depending upon the value for α unless the analysis is invariant to scale. Statistical analyses that are invariant to scale (such as MANOVA, canonical variates analysis, discriminant functions and multiple regression analysis) will not be affected by different choices of values for α. On the other hand, principal components analysis is sensitive to the scaling of variables and thus the results of a relative warps analysis will depend on the value of α (Rohlf, 1993a).

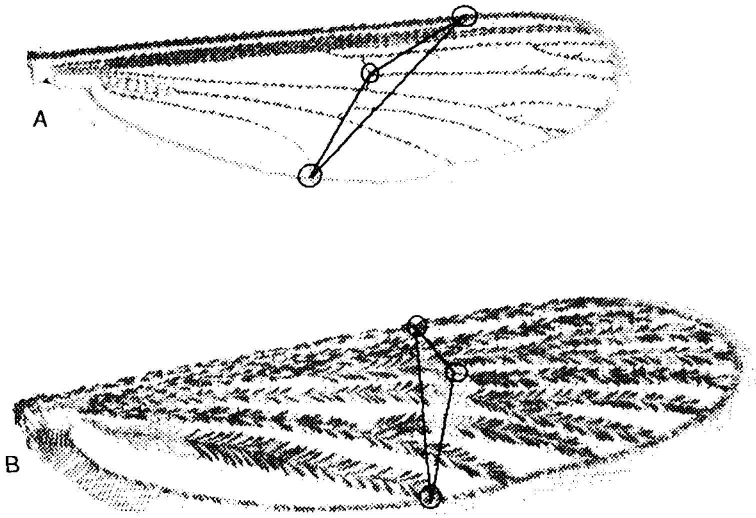

Figure 1. An example of a set of three landmarks whose relative positions in two species differ in part by a reflection. A. *Uranotaenia lowii*. B. *Orthopodomyia signifera*. Images of mosquito wings were scanned from illustrations in Carpenter and LaCasse (1955).

DISCUSSION

As mentioned above, Lele (1993) gives a different definition for form space. He uses a different definition because his methods are based on matrices of Euclidean distances among landmarks and cannot distinguish between a configuration of landmarks and its mirror image. Although invariance to reflection may sometimes be a convenience when working with incomplete specimens of bilaterally symmetric organisms, it can also be a serious limitation. Certain types of differences in shape will not be detectable, and nonsensical shape comparisons can result if reflections are ignored (see Fig. 1 for an example). Although it is much easier to find examples based on just three landmarks, such as shown in Fig. 1, examples can be constructed involving any number of landmarks. Differences due to reflection must be considered shape differences unless one knows that the organism itself has been reflected. Lele (1993) also states that form space is of $p(p-1)/2$ dimensions—because that is how many unique elements there are in a $p \times p$ distance matrix. However, these distances are partially redundant. The space corresponding to all possible physically realizable configurations of landmarks can occupy at most a pk-dimensional space. As discussed above, by fixing differences due to translation and rotation, the dimensionality is reduced to $pk-k-k(k-1)/2$.

It is interesting to compare statistical analyses based on principal warp scores with those based on the residuals from a superimposition analysis. If the reference used in equation (4) is the consensus configuration from an orthogonal generalized Procrustes analysis and if the $\mathbf{X}_i$ are the coordinates of the specimens after alignment with the reference (as is the recommended procedure, F. L. Bookstein, personal communication), then the elements of $\mathbf{V}_i$ are the residuals that are examined (usually graphically in a Procrustes analysis) to look for differences in shape. The square root of the sum of the squared elements of $\mathbf{V}_i$ is the Procrustes distance between the ith specimen and the reference. If $\alpha = 0$ and if all p principal warps are retained in matrix **E** in equation (4), then the $\mathbf{W}_i$ matrices are just a constant times a rigid rotation of $\mathbf{V}_i$. The $\mathbf{V}_i$ and $\mathbf{W}_i$ matrices contain the same information, and standard multivariate statistical analyses based on them will give equivalent if not identical results. Slice (1993a) gives examples of statistical analyses based on the coordinates of specimens aligned using superimposition methods. Slice (1993a) also contrasts the results with those based on partial warp scores.

Lele (1993) makes much of a supposed inconsistency of superimposition methods to estimate an average configuration of landmarks. Unfortunately, his analysis is based on the average of coordinates of landmarks in his definition of form space (specimens not size standardized), which does not correspond to any of the published methods. His example of a dataset for which Procrustes analysis yields an inconsistent estimate has a very extreme covariance structure. Interestingly, his own method, EDMA, also yields inconsistent estimates for that case (C. R. Goodall, personal communication). Procrustes analysis has been shown to yield consistent estimates for the case of circular isotropic error (Kent and Mardia, in prep). This paper is mostly concerned with shape statistics based on projections onto the tangent space (e.g., analyses based on partial warp scores) rather than operating directly in shape space. This is because the tangent space has a Euclidean metric and thus conventional multivariate statistical methods can be applied. For small variation in shape the approximation of shape space by a tangent space is very good. The issue of consistency does not seem critical for multivariate statistical analyses in the tangent space because small differences in the reference result in only slightly different orthogonal matrices to be used to determine the projection into the tangent space (Rohlf, in prep). Some researchers (e.g., Goodall, 1992a; Kent, 1991) are, however, working on the development of specialized techniques that operate in shape space. Such work is important because it will provide new methods capable of exact analyses when variation in shape is not small.

ACKNOWLEDGMENTS

The stimulating questions by Diego Rasskin, Stefan Revets, Jeff Walker and others at the NATO Advanced Study Institute "Advances in Morphometrics" at Il Ciocco, Italy, were very helpful in motivating the present note. Their persistence was very much appreciated. Fred Bookstein, Colin Goodall, Dennis Slice and Jeff Walker read a version of the manuscript and contributed many helpful suggestions that greatly improved this paper. This paper was supported by a grant (DEB 93-17572) from the Systematic Biology Program of the National Science Foundation.

This paper is contribution number 911 from the Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook.

REFERENCES

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

Bookstein, F. L. 1986. Size and shape spaces for landmark data in two dimensions (with discussion and rejoinder). Statistical Science 1:181–242.

Bookstein, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence 11:567–585.

Bookstein, F. L. 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge University Press: Cambridge.

Bookstein, F. L., and P. Sampson. 1990. Statistical models for geometric components of shape change. Communications in Statistics: Theory and Methods 19:1939–1972

Carpenter, S. J. and W. J. LaCasse. 1955. Mosquitoes of North America (North of Mexico). University California Press: Berkeley.

Goodall, C. R. 1991. Procrustes methods in the statistical analysis of shape (with discussion and rejoinder). Journal of the Royal Statistical Society, Series B, 53:285–339.

Goodall, C. R. 1992a. Dynamic graphics in non-Euclidean spaces: The visualization and statistical analysis of shape. 1992 Proceedings of the Statistical Graphics Section American Statistical Association; Alexandria, Virginia.

Goodall, C. R. 1992b. Shape and image analysis for industry and medicine. Short course, XIIth Leeds workshop. University of Leeds, Leeds, UK.

Kendall, D. G. 1984. Shape-manifolds, procrustean metrics and complex projective spaces. Bulletin of the London Mathematical Society 16:81–121.

Kendall, D. G. 1986. Comment on F. L. Bookstein, Size and shape spaces for landmark data in two dimensions. Statistical Science 1:222–226.

Kent, J. T. 1991. The complex Bingham distribution and shape analysis. Technical report STAT 92/04, Department of Statistics, University of Leeds.

Krzanowski, W. J. 1988. Principles of multivariate analysis: A user's perspective. Oxford University Press: Oxford.

Lele, S. 1993. Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference. Mathematical Geology 25:573–602.

Marcus, L. F. 1990. Traditional morphometrics. Pages 77–122 *in* F. J. Rohlf and F. L. Bookstein, (eds.). Proceedings of the Michigan morphometrics workshop. University of Michigan Museum of Zoology Special Publication 2.

Marcus, L. F., E. Bello, and A. Garcia-Valdecasas (eds.). 1993. Contributions to morphometrics. Monografias del Museo Nacional de Ciencias Naturales. 8, Madrid.

Rohlf, F. J. 1990a. Morphometrics. Annual Review of Ecology and Systematics 12,:299–316.

Rohlf, F. J. 1990b. TPSPLINE: A program to compare two shapes using a thin-plate spline. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York, 11794.

Rohlf, F. J. 1992. TPSRW: A program for relative warps analysis. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794.

Rohlf, F. J. 1993a. Relative warp analysis and an example of its application to mosquito wings. Pages 131–159 *in* L. F. Marcus, L. F., E. Bello, and A. Garcia-Valdecasas (eds.). 1993. Contributions to morphometrics. Monografias del Museo Nacional de Ciencias Naturales . 8, Madrid.

Rohlf, F. J. 1993b. NTSYS-pc, version 1.80. Exeter Software: Setauket, New York.

Rohlf, F. J. 1993c. TPSREGR: a program for regression of partial warp scores. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York, 11794.

Rohlf, F. J., and F. L. Bookstein, (eds.), 1990. Proceedings of the Michigan morphometrics workshop. University of Michigan Museum of Zoology Special Publication 2.

Rohlf, F. J., and L. F. Marcus 1993 A revolution in morphometrics. Trends in Ecology and Evolution. 8:129–132.

Rohlf, F. J., and D. Slice. 1990a. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39:40–59.

Rohlf, F. J., and D. Slice. 1990b. GRF: A program for generalized rotational fitting. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York, 11794.

Slice, D. E. 1993a. Extensions, comparisons, and applications of superimposition methods for morphometric analysis. Ph.D. dissertation: State University of New York at Stony Brook.

Slice, D. E. 1993b. GRF-ND: Generalized rotational fitting of n-dimensional data. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York, 11794.

Sneath, P. H. A. 1967. Trend-surface analysis of transformation grids. Journal of Zoology 151:65–122.

Thompson, D. W. 1917. On growth and form. Cambridge University Press: London.

Walker, J. A. 1994. Morphometrika. Geometric morphometrics for the Macintosh. Department of Anatomical Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794.