Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks
Author(s): F. James Rohlf and Dennis Slice
Source: *Systematic Zoology*, Vol. 39, No. 1 (Mar., 1990), pp. 40-59
Published by: Taylor & Francis, Ltd. for the Society of Systematic Biologists
Stable URL: http://www.jstor.org/stable/2992207
Accessed: 25-05-2016 17:39 UTC

Syst. Zool., 39(1):40–59, 1990

EXTENSIONS OF THE PROCRUSTES METHOD FOR THE OPTIMAL SUPERIMPOSITION OF LANDMARKS

F. JAMES ROHLF AND DENNIS SLICE

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

Abstract.—Superimposition methods for comparing configurations of landmarks in two or more specimens are reviewed. These methods show differences in shape among specimens as residuals after rotation, translation, and scaling them so that they align as well as possible. A new method is presented that generalizes Siegel and Benson's (1982) resistant-fit theta-rho analysis so that more than two objects can be compared at the same time. Both least-squares and resistant-fit approaches are generalized to allow for affine transformations (uniform shape change). The methods are compared, using artificial data and data on 18 landmarks on the wings of 127 species of North American mosquitoes. Graphical techniques are also presented to help summarize the patterns of differences in shape among the objects being compared. [Morphometrics; resistant-fit; least-squares; theta-rho analysis; rotational fit; affine transformations.]

An important problem in morphometrics is that of comparing configurations of landmarks in two or more specimens. Thompson (1917) suggested an elegant approach, using "transformation grids," that depicts the overall form of one organism as a distortion in the shape of a reference organism. The basic idea was to place a Cartesian coordinate grid over the reference organism and then distort the image of the organism (including the grid) in various ways until the form of the second organism was achieved. The differences in shapes of the two organisms are shown by the deviations of the fitted grid (usually bent and stretched in various ways) from the original simple square grid. Thompson (1917) sketched the grids subjectively without an explicit specification of which landmarks were used. Not all landmarks shown in pairs of drawings are located exactly where the superimposed grids would imply they should be. This means that the grids should be more complex than those shown in Thompson (1917) in order to accurately show the differences between two organisms. Bookstein (1978) developed the method of biorthogonal grid analysis which quantifies Thompson's approach and makes it objective. But it is complex and has not been applied very often. A recent break-through (Bookstein, 1989) is the use of methods based on thin-plate splines. It is now easy to display a transformation grid that maps the configuration of landmarks of one organism exactly into those of another.

An alternative approach to fitting a model that completely describes the differences between two organisms is to fit a very simple model that only takes into consideration global parameters such as differences in rotation, translation, and scale. Geometrically, this corresponds to superimposing one organism on top of another so that its landmarks align as well as possible (in some sense) with the positions of the corresponding landmarks on the second. Differences in shape are then shown by differences in positions of corresponding landmarks. Shape differences between two organisms are found by studying these residuals. These methods are the subject of the present paper.

Sneath (1967) investigated the problem of finding the optimal translation, rotation, and size change of one object in order for it to be superimposed on another. The two objects were represented as sets of x,y-coordinates of landmarks. A least-squares criterion was used to measure the goodness of fit of one object to another. Gower (1971) further developed Sneath's (1967) method and expressed the operations in terms of matrix algebra. Siegel and Benson (1982) made the important observation that a

global least-squares criterion usually results in a general lack of fit at most points—even if the forms being compared are identical except for the position of a few landmarks—which makes the direct interpretation of residuals very misleading. They then proposed a "resistant-fit" method, based on a nonparametric analog of least-squares regression, that is better able to reveal differences between two objects when the major differences are mostly in the relative positions of a few landmarks. Olshan et al. (1982) compared the two approaches and demonstrated the usefulness of the resistant-fit method. Gower (1975) generalized the least-squares method so one can compare any number of objects. In this paper we present a similar generalization for the resistant-fit method.

More recently, Goodall and Green (1986) have suggested the use of affine transformations when fitting one set of landmarks to another. This permits dilatations (magnifications and contractions). Affine transformations can be interpreted as uniform shape changes along two orthogonal axes—rather than only simple isometric changes in overall size. Affine transformations have the properties that pairs of vectors that are parallel stay parallel and pairs of orthogonal vectors stay orthogonal after the transformation. In this paper we generalize this technique to handle any number of specimens by modifying Gower's (1975) generalized Procrustes method. We also develop a resistant-fit version.

Of course, neither of these methods can be used on all types of data. In organisms where one has only an outline and few homologous landmarks, other approaches are required. For example, one can fit various functions (such as Fourier series) to the outlines and then use their parameters as characters for multivariate analysis (e.g., Rohlf and Archie, 1984; Ferson et al., 1985).

ALIGNMENT OF TWO OBJECTS

Least-Squares Fit—Orthogonal Procrustes Analysis

Sneath (1967) described a method for superimposing coordinates of one object so that they best match (in a least-squares sense) the configuration of corresponding points in another object. The corresponding points (called "*h*-points" by Sneath, 1967) must be at least "operationally" homologous.

The coordinates of the p points for the two k-dimensional objects can be stored in two p by k matrices, $\mathbf{X}_1$ and $\mathbf{X}_2$. Sneath's (1967) and Gower's (1971) methods consist of translating the coordinates for the two objects to a common center (each object's centroid is translated to the origin), scaling to a common size, and rotating (and possibly reflecting) one object until it aligns with the other object.

Their procedure consists of the following steps:

1. Center the coordinates of one object on the other. This can be done by expressing the coordinates of each object as deviations from their x,y-means by premultiplying $\mathbf{X}_i$ by $\mathbf{I} - \mathbf{P}$, where $\mathbf{I}$ is a $p \times p$ identity matrix and $\mathbf{P}$ is a $p \times p$ matrix with all elements equal to $1/p$.
2. Scale both objects. Sneath (1967) divided the elements of $\mathbf{X}_i$ by the square root of the average squared distance of the p landmarks to their centroid. The points will then have average squared distance to the centroid equal to p. We have found it more convenient to follow Gower's (1971) convention of using the square root of the sum of the squared distances so that after translation and scaling the sum of the squared distances to the centroid will be equal to unity. The transformed coordinates for the ith object will be
$$\mathbf{X}_i' = (\mathbf{I} - \mathbf{P})\mathbf{X}_i / s_i, \qquad (1)$$
where
$$s_i = \sqrt{\mathrm{tr}((\mathbf{I} - \mathbf{P})\mathbf{X}_i\mathbf{X}_i'(\mathbf{I} - \mathbf{P}))}, \qquad (2)$$
the sum of the squared distances of each landmark to the object's centroid.
3. Rotate the second set of points to achieve the best approximation, $\mathbf{X}_2^*$, to the first. This can be expressed most compactly as

$$\mathbf{X}_2^* = \mathbf{X}_2'\mathbf{H}. \quad (3)$$

In two-dimensions, the rotation matrix **H** is

$$\mathbf{H} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \quad (4)$$

Sneath (1967) showed that the optimal rotation angle, θ, can be obtained from the following equation:

$$\tan\theta = -\frac{(X_{21}''X_{12}' - X_{11}''X_{22}')}{X_{11}''X_{21}' + X_{12}''X_{21}'}, \quad (5)$$

where X_{ij}' is the jth column vector of matrix $\mathbf{X}_i'$. Sneath suggests repeating the above steps after changing the signs of the x-coordinates for one of the two objects to see if such a reflection would result in a better fit.

Gower (1971) computed the rotation matrix directly as

$$\mathbf{H} = \mathbf{VSU}^t, \quad (6)$$

where **U** and **V** are such that

$$\mathbf{X}_1''\mathbf{X}_2' = \mathbf{U}\Sigma\mathbf{V}^t \quad (7)$$

and Σ is diagonal, i.e., an Eckart-Young singular-value decomposition of $\mathbf{X}_1''\mathbf{X}_2'$ is performed (Gower, 1975). **S** is a diagonal matrix with $s_{ii} = \pm 1$. The signs are taken to be the same as the corresponding diagonal elements of the Σ matrix. This will automatically take into consideration any reflections that would result in a better fit. Using **S**, rather than Σ, constrains **H** to correspond to a rotation rather than a shear (see below).

Measure the difference between $\mathbf{X}_1'$ and its least-squares approximation, $\mathbf{X}_2^*$. The criterion proposed by both Sneath (1967) and Gower (1971) is a function of the sum of the squared differences in x,y-coordinates between the reference points and the corresponding fitted points.

$$\sum^{n} \Delta_{12}^2 = \mathrm{tr}\left((\mathbf{X}_1' - \mathbf{X}_2^*)(\mathbf{X}_1' - \mathbf{X}_2^*)^t\right) \quad (8)$$

Sneath (1967) suggests using $d_{12} = \sqrt{\sum \Delta_{12}^2/p}$ as a measure of average "taxonomic distance."

Gower (1975) proposed the use of the distance coefficient

$$m_{12} = \sqrt{\Sigma\, \Delta_{12}^2/2} \quad (9)$$

(called R^2 in Gower, 1971). With the scaling used by Gower (1971, 1975), this reduces to

$$m_{12} = \sqrt{1 - \mathrm{tr}\,\Sigma}. \quad (10)$$

The scaling in step 2 above is important since, otherwise, the distance from the second object to the first, m_{12}, need not result in the same value, m_{21}, one would obtain if one were to fit the first object to the second (Gower, 1971). Sneath (1967) also considered the problem of rotational fits in three-dimensions and allowed for the special case of bilateral symmetry. Gower's (1975) procedure is valid for any number of dimensions.

Orthogonal Resistant-Fit

For least-squares fitting procedures the overall fit can be greatly influenced by the presence of a few outliers (points that fit especially poorly). Sneath (1967) considered more robust fitting methods but concluded that available methods were "not attractive in practice." Bookstein and Sampson (1987) show how one can detect the effects of outliers by simply refitting with various combinations of landmarks left out of the fitting process. Siegel and Benson (1982) developed an automatic procedure based on Siegel's (1982a) method of robust regression using repeated medians that works well in the two-dimensional case.

As in the case of least-squares fit, a scale factor, angle of rotation, and a translation must be estimated. Siegel and Benson's approach is to superimpose the first object using the following equation:

$$\mathbf{X}_2^* = \mathbf{1}_{p\times 1}(\tilde{\alpha}, \tilde{\beta}) + \tilde{\tau}\mathbf{X}_2\tilde{\mathbf{H}}, \quad (11)$$

where $\mathbf{1}_{p\times 1}$ is a column vector with all elements equal to unity, $(\tilde{\alpha}, \tilde{\beta})$ is a translation vector, $\tilde{\tau}$ is a scale factor, and $\tilde{\mathbf{H}}$ is a rotation matrix. Unlike the least-squares procedure described above, the parameters are estimated using procedures that are not based

on any explicit optimization principle—which makes it unclear how the quality of the final results should be judged.

1. An initial least-squares fit of the second object to the first is performed using the procedure described above for least-squares. This takes care of any necessary reflections and ensures we begin close to the eventual solution. This facilitates the robust estimation of the rotation parameter, $\tilde{\theta}$.

2. The scale factor, $\tilde{\tau}$, is computed as the repeated median of the ratios of corresponding distances between landmarks i and j in the results of step 1.

$$\tilde{\tau} = \underset{i}{\text{median}}\left(\underset{j \neq i}{\text{median}}\ (\tau_{ij})\right), \qquad (12)$$

where

$$\tau_{ij} = \sqrt{\frac{(X_{1j} - X_{1i})(X_{1j} - X_{1i})^t}{(X_{2j} - X_{2i})(X_{2j} - X_{2i})^t}}, \qquad (13)$$

and X_{1j} is the jth row vector of the data matrix for object 1.

3. A procedure similar to that used for the scale factor is used to estimate the rotation angle, $\tilde{\theta}$.

$$\tilde{\theta} = \underset{i}{\text{median}}\left(\underset{j \neq i}{\text{median}}\ (\theta_{ij})\right), \qquad (14)$$

The angle, θ_{ij}, indicates the rotation needed to point a vector connecting a pair of points, i, j, in object 2 in the same direction as the corresponding vector in object 1. Siegel and Benson (1982) suggest the use of values for θ_{ij} that lie between $-\pi$ and π since subsequent to the least-squares fitting one would expect the estimates to have an average near 0. If a range such as 0–2π were used then one would have to allow for the fact that $2\pi - \epsilon$ is the same angle as $-\epsilon$ in the determination of the median value. The rotation matrix $\tilde{\mathbf{H}}$ is as defined above for $\mathbf{H}$ but using $\tilde{\theta}$. For three-dimensional objects, rotations in the 1-3 and 2-3 planes would need to be made in an analogous manner.

4. The vector of translation parameters $(\tilde{\alpha}, \tilde{\beta})$ is computed using ordinary medians. Let

$$\mathbf{T} = \mathbf{X}_1 - \tilde{\tau}\mathbf{X}_2\tilde{\mathbf{H}}. \qquad (15)$$

Then

$$\tilde{\alpha} = \underset{i}{\text{median}}\ (t_{i1}),$$

$$\tilde{\beta} = \underset{i}{\text{median}}\ (t_{i2}), \qquad (16)$$

where t_{ij} are the elements of the $p \times 2$ matrix $\mathbf{T}$.

Siegel (1982b) furnished a FORTRAN program to perform the above computations. Siegel and Benson (1982), Siegel (1982b), and Benson et al. (1982) gave a number of examples that demonstrate how objects that differ in the locations of just a few points are aligned correctly using resistant-fit but may show an apparent lack of alignment at most of the points when a least-squares criterion is used—thus masking the nature of the actual differences between objects. Siegel and Benson (1982) showed that, whenever more than $(n + 1)/2$ of the points can be made to fit closely, the resistant-fit method will obtain the "correct" solution. When there are fewer points that fit closely there can be ambiguity—it may be possible to fit different regions of an object separately but not all of them simultaneously. While other robust methods could also be used, the repeated median procedure has the advantage that its "breakdown value" is very high (almost 50% of the data points can be perturbed before the resistant-fit of the unperturbed points is affected). The use of the ordinary median in this algorithm results in a breakdown value of 29% (Siegel and Benson, 1982). By contrast, the least-squares solution can be affected by a change in even a single point.

ALIGNMENT OF MORE THAN TWO OBJECTS

The methods discussed above are all designed to compare only a single pair of objects. A procedure for their use with more than 2 objects is to compute a measure of

distance between each of the $n(n-1)/2$ possible pairs of objects and then analyze the matrix of distances. In the case of least-squares fit one could use the m_{ij} or $\sqrt{\Sigma\Delta_{ij}^2}$ values defined above. Siegel and Benson (1982) did not suggest an appropriate coefficient for use with a resistant-fit but the median or the average of the residual differences could be used. But, unlike the case with least-squares, the difference is not symmetric, that is, the fit of object 1 to object 2 is not the same as the fit of 2 to 1. The distance matrix (perhaps made up of averages of the reciprocal distances) could then be clustered or analyzed by metric or non-metric multidimensional scaling. The disadvantage of this approach is that it gives no direct information about *how* the objects differ. One can get the same numerical value for distance between two objects whether there were many small differences or a few large ones. Also, large differences could be due to either the same or different landmarks in different objects. These cases would be indistinguishable yet their interpretations would be quite different. For these reasons we do not recommend this approach. The methods described below avoid these problems.

Least-Squares Fit—Generalized Orthogonal Procrustes Analysis

Gower (1975) proposed an alternative approach that he called *generalized Procrustes analysis*. In this algorithm, every object's points are translated, scaled, and rotated iteratively until the least-squares fit of all objects to a consensus object is no longer improved. We have simplified the algorithm somewhat by assuming that the scaling step, that he listed as optional, would always be performed. We have also suggested a change in the initial scaling.

1. Center and scale the coordinates for each of the n objects, as in equation 1 above, to yield $\mathbf{X}_i'$. Note that this scaling is different from Gower's (1975) procedure where all $\mathbf{X}_i$'s were multiplied by $\lambda = \sqrt{n/\Sigma_j^n s_j^2}$ (as before, s_j^2 is the sum of the squared distances of the landmarks to the object's centroid). With his procedure the initial estimates tend to be dominated by the largest objects. Our scaling results in a better initial approximation and still satisfies the requirement that $\Sigma_i \text{tr}(\mathbf{X}_i'\mathbf{X}_i'^t) = n$.
2. Set $\mathbf{Y} = \mathbf{X}_1'$ as the initial estimate of the coordinates of the points in the consensus object, then rotate $\mathbf{X}_i'$ to fit matrix $\mathbf{Y}$, for $i = 2, 3, \ldots, n$ using the least-squares fitting procedure described earlier.
3. Compute a new consensus, $\mathbf{Y}$, as the mean of the original $\mathbf{X}_1'$ and the rotated $\mathbf{X}_i'$. Gower (1975) suggested that one update $\mathbf{Y}$ after each object is fitted. We have found it more convenient to update it only at the end.
4. Compute the initial residual sum-of-squares as $S_r = n(1 - \text{tr}(\mathbf{Y}\mathbf{Y}^t))$.
5. Set the individual scale factors, ρ_i, to 1, for $i = 1, \ldots, n$.
6. Rotate each of the $\mathbf{X}_i'$ matrices to fit their mean, $\mathbf{Y}$, using the least-squares procedure. This involves the following steps.

   - Rotate and scale the ith object to the current consensus, $\mathbf{Y}$.

   $$\mathbf{X}_i^* = \rho_i \mathbf{X}_i' \mathbf{H}_i, \tag{17}$$

   where $\mathbf{H}_i$ is computed as in equation 6.

   - Compute a new estimate, $\mathbf{Y}^*$, of the consensus as the mean of the $\mathbf{X}_i^*$.
   - Then let

   $$\mathbf{X}_i^{**} = \frac{\rho_i^*}{\rho_i} \mathbf{X}_i^* \tag{18}$$

   where

   $$\frac{\rho_i^*}{\rho_i} = \sqrt{\frac{\text{tr}(\mathbf{X}_i^*\mathbf{Y}^{*t})}{\text{tr}(\mathbf{X}_i^*\mathbf{X}_i^{*t})\text{tr}(\mathbf{Y}^*\mathbf{Y}^{*t})}}. \tag{19}$$

   The purpose of this scaling is to ensure that $\Sigma_i \text{tr}(\mathbf{X}_i^{**}\mathbf{X}_i^{**t}) = n$ (Gower, 1975).

   - Compute the next estimate, $\mathbf{Y}^{**}$, of the consensus as the mean of the

$\mathbf{X}_i^{**}$. $\mathbf{Y}^{**}$ is a weighted mean of the $\mathbf{X}_i^*$ with the ratios ρ_i^*/ρ_i acting as weights—those $\mathbf{X}_i^*$ that agree most with $\mathbf{Y}^*$ are given the highest weights.

- The new residual sum-of squares is

$$S_r^{**} = n\left(1 - \operatorname{tr}(\mathbf{Y}^{**}\mathbf{Y}^{**t})\right). \qquad (20)$$

- In preparation for the next iteration, set $\mathbf{Y} = \mathbf{Y}^{**}$, $\rho_i = \rho_i^*$, and $\mathbf{X}_i' = \mathbf{X}_i^{**}$.

7. If the change in the fit, $S_r - S_r^{**}$, is greater than a specified tolerance, then set $S_r = S_r^{**}$ and repeat step 6 above. Otherwise, the iteration is complete. The method involves a considerable amount of computation but it converges on a solution quite well. Gower suggested a tolerance of 10^{-3} or 10^{-4} and we found that to be satisfactory.

For convenience, the final $\mathbf{Y}$, $\mathbf{H}_i$, and $\mathbf{X}_i$ matrices can be post-multiplied by the normalized eigenvectors of $\mathbf{YY}^t$ (to align the axes with the directions of maximum variance). The final ρ_i gives the scale, or "size," of object i. Since the method is based on a least-squares fit, it may not give the most desirable solution when most of the differences among a set of objects are localized at a few landmark points. See the next section for one solution to this problem.

Generalized Orthogonal Resistant-Fit Analysis

Gower's (1975) approach can be modified to use resistant-fit rather than least-squares procedures for determining the best translation, scaling, and rotation at each step. We use least-squares to achieve an initial alignment and then use the resistant-fit method to obtain the final alignment of each object to a mean *consensus object*. Our method is detailed below.

1. Center and scale the coordinates for each of the n objects to a common size to yield $\mathbf{X}_i'$. Here common size is defined as each object being scaled such that the median squared interlandmark distance is unity, i.e., $\underset{i \neq j}{\text{median}}\,(\mathbf{X}_{1i} - \mathbf{X}_{1j})(\mathbf{X}_{1i} - \mathbf{X}_{1j})^t = 1$, where $\mathbf{X}_{1i}$ is the ith row vector of the coordinates of a landmark of $\mathbf{X}_1$.
2. Set $\mathbf{Y} = \mathbf{X}_1'$ as the initial estimate of the coordinates of the locations of the points in the consensus object, then rotate the $\mathbf{X}_i'$ to fit matrix $\mathbf{Y}$ using the least-squares procedure.
3. Compute a new $\mathbf{Y}$ as the median of the rotated $\mathbf{X}_i'$. For convenience in testing convergence, scale the $\mathbf{Y}$ matrix such that $\underset{i \neq j}{\text{median}}(\mathbf{Y}_i - \mathbf{Y}_j)(\mathbf{Y}_i - \mathbf{Y}_j)^t = 1$, where $\mathbf{Y}_i$ is the ith row of $\mathbf{Y}$.
4. Rotate the current $\mathbf{X}_i'$ matrices to fit the present $\mathbf{Y}$ matrix using the resistant-fit procedure.
   - Compute a new estimate, $\mathbf{X}_i^*$, of the best fit of the ith object to the current consensus object.

$$\mathbf{X}_i^* = \mathbf{1}_{\rho\times 1}(\tilde{\alpha}_i, \tilde{\beta}_i) + \tilde{\tau}_i\mathbf{X}_i'\tilde{\mathbf{H}}_i, \qquad (21)$$

   where $\tilde{\tau}_i$ is the scale factor and $\tilde{\mathbf{H}}_i$ is the rotation matrix from a resistant-fit of $\tilde{\tau}_i\mathbf{X}_i'$ to $\mathbf{Y}$.
   - Compute the new estimate of the consensus, $\mathbf{Y}^*$, as the mean of the $\mathbf{X}_i^*$.
5. Since there is no explicit criterion being optimized, the most appropriate measure of convergence is not known. We use the median difference in $\mathbf{Y}$ matrices in successive iterations (the median difference between $\mathbf{Y}$ and $\mathbf{Y}^*$). While a criterion such as $\operatorname{tr}(\mathbf{Y}^* - \mathbf{Y})(\mathbf{Y}^* - \mathbf{Y})^t$ could be used, the median distance seems more in line with the goals of the robust procedure. If the value of the criterion is greater than a specified tolerance, then set $\mathbf{Y} = \mathbf{Y}^*$, rescale $\mathbf{Y}$ to have a unit squared interlandmark distance, set $\mathbf{X}_i' = \mathbf{X}_i^*$, and repeat step 4 above. Otherwise, the iteration is complete.

One could, if desired, compute the residual sum-of-squares, S_r^*, the sum of the squared deviations of the objects, $\mathbf{X}_i^*$, from their average, $\mathbf{Y}^*$. Note that S_r^* is not

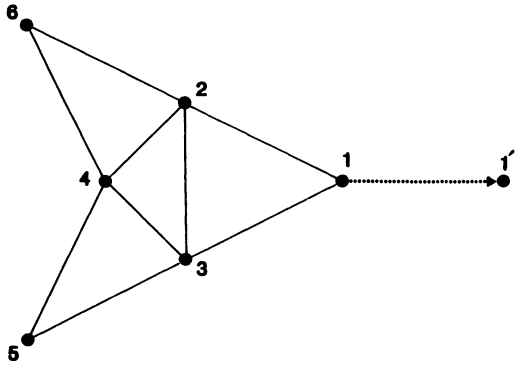

FIG. 1. Diagram showing configuration of landmarks for two artificial "organisms" differing only in the location of landmark number 1.

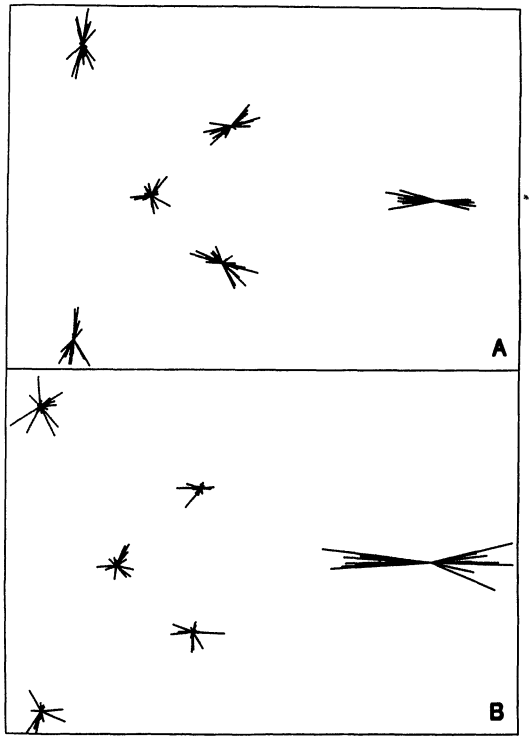

FIG. 2. Results of (A) generalized Procrustes and (B) resistant-fit analyses for a sample of 20 from the population defined in Figure 1. Vectors emanating from the centroid of each landmark show the deviation of each landmark from the consensus configuration.

expected to be minimal since the resistant-fit procedure was not intended to minimize the residual sums of squares. Differences in scaling preclude the direct comparison of the magnitude of the residuals of the two methods. This procedure is not guaranteed to monotonically converge as in the case of the least-squares procedure. In practice, the convergence properties are quite good and appear to improve with increasing sample size and number of landmarks. We have achieved satisfactory results using a convergence tolerance of 10^{-3}.

Comparison of Generalized Orthogonal Least-Squares and Resistant-Fit Results

To illustrate differences in the results obtained by these two techniques, we constructed an artificial data set with six landmarks for each of 20 hypothetical organisms. The basic configuration of the landmarks is shown in Figure 1. Independent, identically distributed random displacements were then added to each landmark, and each configuration was randomly rotated, magnified, and displaced. The data were also subdivided into two groups of 10 configurations. In one of the groups the "nose" of each organism, landmark 1, was shifted away from the other landmarks to simulate the impact of systematic variation (Fig. 1). All configurations were then subjected to the procedures described above.

Figure 2A shows the results of the generalized least-squares procedure. No striking difference in the amount of variability about any of the landmarks is evident. But, landmark 4 seems to have a more or less circular scatter of fitted landmarks around it, while variation at the others seems to have distinct patterns of orientation. These patterns are a result of the mixture of long- and short-nosed objects. The difference between the two groups is the displacement at landmark 1. This displacement, if correctly represented, would substantially inflate the least-squares goodness of fit criterion. As a result, the least-squares method tends to distribute these differences over the entire configuration. Here, this is ef-

fected through the increased influence (in the long-nosed organisms) of landmark 1 on the configuration centroid location and scaling parameters. The long nose pulls the centroid more anteriorly, at the same time inflating the size reduction necessary to minimize the least-squares fit to the consensus configuration. The converse is true for the short-nosed specimens. These factors result in a relatively smaller, more posteriorly placed "body" (landmarks 2–5) in the long-nosed objects when the configurations are scaled and their centroids superimposed. Had the differences in the position of landmark 1 between the groups not been aligned with respect to the initial configuration centroid, a rotational component would also be present.

As expected, resistant-fit (Fig. 2B) more accurately exposes the known patterns of interlandmark variation. Using medians and repeated-medians, the excessively large, or small, parameter estimates associated with fitting landmark 1 are ignored in favor of those based on the other landmarks. The result is a more or less uniform fit of landmarks 2–5 and the allocation of most variation to landmark 1 and oriented primarily in the direction of the difference in the location of this landmark in the two groups.

ALLOWING FOR UNIFORM SHAPE CHANGE

The methods described above can be further generalized to incorporate differential contractions or stretching along a pair of orthogonal axes. Bookstein (1985) referred to this as the homogeneous component of shape change. Bookstein and Sampson (1987) called this a *uniform shape change* (in contrast to a local change) and showed how its contribution can be estimated and tested for statistically. By considering this additional operation (and, therefore, adding additional parameters) the degree of fit can be improved. If one object differs from another by having been uniformly stretched along one direction (Fig. 3) then the methods described above should indicate this by showing a consistent nonrandom pattern of directions in the residual vectors. The landmarks near each end of the axis along which the stretching took place should have vectors pointing in opposite directions. From this distinctive pattern one should be able to infer the magnitude and orientation of any stretching or contracting of an object. Figure 1C in Sinervo and McEdward (1988) is an example based on sea urchin larvae. There are two advantages to incorporating dilatations in the fitting process. First, it is more convenient if the direction and magnitude of the stretching or contraction are computed rather than having to be inferred. Second, it is easier to detect local displacements after effects of uniform changes have been removed from the data. Generalization to include affine transformations in the least-squares and resistant-fit methods are described below.

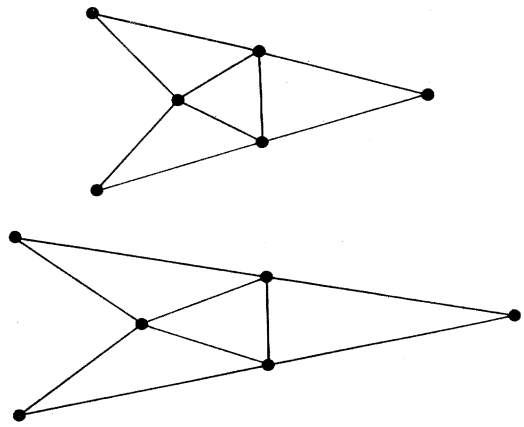

FIG. 3. Affine deformation. The configurations are identical except that the lower one was uniformly stretched in the horizontal direction.

Least-Squares—Oblique Procrustes Analysis

Goodall and Green (1986) suggested application of affine transformations to the morphometric problem of mapping one set of landmarks onto another. The least-squares solution to this problem was developed by Hurley and Cattell (1962). Cattell and Khanna (1977) and Gower (1984) reviewed the general problem of transforming one set of coordinates into another

when orthogonality constraints are not imposed.

As before, the coordinates of the two specimens should first be translated to place their centroids at the origin (using equation 1). It is not really necessary to scale, however, since that will be done as part of the fitting procedure. The matrix of coordinates of the landmarks of the second object, after rotation and dilatation to fit the first object, is given by

$$\mathbf{X}_2^* = \mathbf{X}_2'\mathbf{H}^*, \tag{22}$$

where the affine transformation matrix is

$$\mathbf{H}^* = (\mathbf{X}_2''\mathbf{X}_2')^{-1}\mathbf{X}_2''\mathbf{X}_1'. \tag{23}$$

This is the equation for the least-squares estimates of the partial regression coefficients in a multivariate multiple regression analysis. Note that there is no explicit scaling step since this is taken care of by the multiplication by the inverse matrix. As before, the transformation matrix, $\mathbf{H}^*$, can also be expressed in terms of a singular value decomposition,

$$\mathbf{H}^* = \mathbf{U}\Sigma\mathbf{V}^t \tag{24}$$

$$= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}. \tag{25}$$

The diagonal matrix of eigenvalues, Σ, is used rather than $\mathbf{S}$ as in equation 6 ($\mathbf{S}$ was constrained to have diagonal elements equal to ± 1). Multiplication of $\mathbf{X}_1'$ by $\mathbf{H}^*$ corresponds to the rotation of the landmarks through the angle $-\theta$ to align the deformation axes with the coordinate axes, followed by the multiplication of the resulting landmark coordinates by the scaling factors p and q in the x and y directions respectively, and, finally, the rotation of the configuration through the angle ψ, which realigns the fitted object with the reference.

From these parameters one can determine the net change in the orientation of the fitted object as $\theta - \psi$. However, θ is a function of the placement of the fitted object on the coordinate system and ψ will be a function of the placement of the reference object. Unless both objects were initially placed in some meaningful orientation, neither these angles nor their difference will be biologically interesting by themselves. For a fixed, or common, reference, however, differences in ψ between different fitted configurations will identify differences in the direction of their shape change with respect to the reference.

This can be shown graphically for each configuration as a *strain cross*—a pair of orthogonal axes with lengths proportional to the two eigenvalues and oriented so that the longer axis is parallel to the direction, ψ, of maximum stretching. If the eigenvalues are identical then the change is isotropic and the angles are not uniquely defined. In that case Gower's method, described above, is appropriate and can be used to rigidly align one object with another. Goodall and Green (1986) gave equations for the relationships between the canonical parameters, the strain cross parameters, and the singular value decomposition (equation 24).

To generalize the procedure to handle more than two specimens, one can substitute $\mathbf{H}^*$ for $\mathbf{H}$ in equation 17 above for Gower's generalized Procrustes procedure. This substitution simplifies the computation of the scale factors, ρ_i. Equation 19 becomes

$$\frac{\rho_i^*}{\rho_i} = \sqrt{\frac{\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{*t})}{\mathrm{tr}(\mathbf{Y}\mathbf{Y}^t)\mathrm{tr}(\mathbf{Y}^*\mathbf{Y}^{*t})}}, \tag{26}$$

where $\mathbf{Y}$ is the estimate of the consensus configuration from the previous iteration. The importance of this result is that the scale factors no longer depend upon i and thus $\mathbf{Y}^*$ can be taken as a simple average of the $\mathbf{X}_i^*$. The iterative process corresponds to the equation

$$\mathbf{Y}^* = \frac{1}{n}\left(\sum_i^n \mathbf{X}_i^*\right) = \frac{1}{n}\left(\sum_i^n \mathbf{X}_i\mathbf{H}_i^*\right) = \frac{1}{n}\left(\sum_i^n \mathbf{X}_i(\mathbf{X}_i^t\mathbf{X}_i)^{-1}\mathbf{X}_i^t\right)\mathbf{Y}. \tag{27}$$

The solution of the iterative process (the configuration such that $\mathbf{Y}^* = \mathbf{Y}$) is the matrix of eigenvectors of

$$\mathbf{A} = \frac{1}{n}\sum_{i}^{n} \mathbf{X}_i(\mathbf{X}_i^t\mathbf{X}_i)^{-1}\mathbf{X}_i^t. \quad (28)$$

This procedure "works" in that it rotates and scales all objects so that they are in mutual alignment and the consensus configuration corresponds to the means for each landmark. But the consensus object may not resemble the original objects—even when the fit is perfect. The $\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ operation in equation 28 transforms each object so that the variance in the bivariate distribution of landmarks is the same in all directions in the plane for each object. The consensus configuration is computed by superimposing these transformed objects. What needs to be done next is to transform the consensus configuration so that it has an aspect ratio more like that of an average object. One method is to fit $\mathbf{Y}^*$ (the eigenvectors of $\mathbf{A}$) to each of the $\mathbf{X}_i$ and then use a consensus of these as the final configuration. Let

$$\mathbf{Y}_i^* = \mathbf{Y}^*\mathbf{H}_i \quad (29)$$

where $\mathbf{Y}^*$ is the matrix of eigenvectors of matrix $\mathbf{A}$ and $\mathbf{H}_i$ is the affine transformation matrix $(\mathbf{Y}^{*t}\mathbf{Y}^*)^{-1}\mathbf{Y}^{*t}\mathbf{X}_i$ that fits $\mathbf{Y}^*$ to $\mathbf{X}_i$. In scalar product form this is (after some simplification)

$$\mathbf{Y}_i\mathbf{Y}_i^{*t} = \mathbf{Y}^*\mathbf{Y}^{*t}\mathbf{X}_i\mathbf{X}_i^t\mathbf{Y}^*\mathbf{Y}^{*t}. \quad (30)$$

A scalar products matrix is a linear function of the squared distances between the landmarks (see, for example, Gower, 1984). These matrices can then be averaged over the n objects to yield

$$\begin{aligned}\mathbf{C} &= \frac{1}{n}\sum_{i} \mathbf{Y}_i^*\mathbf{Y}_i^{*t} \\ &= \mathbf{Y}^*\mathbf{Y}^{*t}\left(\frac{1}{n}\sum_{i} \mathbf{X}_i\mathbf{X}_i^t\right)\mathbf{Y}^*\mathbf{Y}^{*t} \\ &= \mathbf{A}\left(\frac{1}{n}\sum_{i} \mathbf{X}_i\mathbf{X}_i^t\right)\mathbf{A}. \quad (31)\end{aligned}$$

Somewhat in analogy to Gower's (1966) principal coordinates analysis, the first two eigenvectors, $\mathbf{Y}^{**}$, of $\mathbf{C}$ can be used as our consensus configuration. It is simply the consensus scaled to have the same shape as the average original object. Finally, each object, $\mathbf{X}_i$, can then be fit to the consensus and a plot produce showing the relationship between each object and the consensus configuration.

Oblique Resistant-Fit Analysis

Affine transformations can also be incorporated into the resistant and generalized resistant-fitting procedures. A set of three, noncollinear landmarks is both the minimum set for which an affine transformation can be determined and the maximum that guarantees a perfect fit. In the context of resistant-fitting, this suggests the examination of all possible triplets of landmarks. But such an approach dramatically increass the computation necessary to determine the fitting parameters. For example, a configuration of only ten landmarks requires the examination of $P(10, 3)$, or 720, triangles of landmarks per object at each iteration. Such a procedure does, however, have the property of being insensitive (resistant) to extreme displacements of some of the landmarks.

Goodall and Green (1986) provide a clear discussion of the details of determining the parameters for the affine fitting of one triangle of landmarks to another. What is required is the determination of a transformation matrix, $\mathbf{H}$, such that

$$\mathbf{XH} = \mathbf{Y}, \quad (32)$$

where $\mathbf{X}$ is a 3×2 matrix of coordinates (with the rows corresponding to the three points to be fitted), $\mathbf{H}$ is a 2×2 transformation matrix, and $\mathbf{Y}$ is the matrix of the coordinates of the vertices of the reference triangle. The elements of $\mathbf{H}$ can be directly determined by first centering both triangles at the origin and then forming reduced matrices, $\mathbf{X}'$ and $\mathbf{Y}'$, consisting of the centered coordinates of two of the landmarks. Only two landmarks need be considered because by centering both triangles at the origin, the coordinates of any two points determine those of the third. $\mathbf{H}$ can be determined as

$$\mathbf{H} = \mathbf{X}'^{-1}\mathbf{Y}'. \quad (33)$$

A singular value decomposition can be performed to obtain

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^t \quad (34)$$

and thereby determine the values of p, q, θ, and ψ as in equation 25.

For configurations with more than three landmarks, one can calculate the affine transformation matrix (or the strain cross parameters) necessary to perfectly fit each triplet of landmarks in a configuration to the corresponding points on a reference object. Goodall (1983) makes brief mention of this possibility but feels the benefits of the resistant method's insensitivity to local landmark shifts do not justify its application to data based on relatively few landmarks. He points out that, if such an approach were to be adopted, the repeated median estimates of the elements **H** should yield estimates closer to the least-squares results than those derived from medians of the strain parameters, p, q, θ, and ψ. This is due, at least in part, to properties of the joint density of the affine transformation parameters (Goodall, 1983).

In addition to any statistical benefits of using repeated median estimates of the elements of **H**, there are also computational advantages since it is not necessary to compute the singular value decomposition of the transformation matrix for each triplet of landmarks in each iteration. We use this approach below. Our procedure for generalized resistant-fitting allowing for uniform shape change is as follows:

1. Begin with a least-squares fit of all of the objects to the first, as described in previous sections. This allows for any necessary reflections and brings the objects close to the desired location and orientation.
2. For each object, i, set the associated transformation matrix, $\mathbf{H}_i$, equal to **I**, the identity matrix. Calculate a new **Y** as the median of the $\mathbf{X}_i$'s. Scale **Y** and each $\mathbf{X}_i$ to have a unit median squared interlandmark distance.
3. Examine all possible triplets of landmarks (j, k, l) for $j = 1 \ldots p$, $k = 1 \ldots p$ and $k \neq j$, and $l = 1 \ldots p$ and $l \neq k$. If the points to be fitted are not collinear then for a given pair (j, k) compute, for each l, the elements of the transformation matrix $\mathbf{H}^*_{i(jkl)}$ required to perfectly fit the triangle of landmarks in $\mathbf{X}_i$ to the corresponding landmarks of the reference as described above. For a given j, the elements of $\mathbf{H}^*_{i(jk)}$ associated with each k are taken as the separately determined median of the corresponding elements of the $\mathbf{H}^*_{i(jkl)}$ matrices. Then for each j the elements of $\mathbf{H}^*_{i(j)}$ are taken as the median values of the corresponding elements of the $\mathbf{H}^*_{i(jk)}$ matrices. And finally, the elements of the transformation matrix for the entire configuration, $\mathbf{H}^*_i$, are taken to be the medians of each of the elements of the $\mathbf{H}^*_{i(j)}$ matrices. Once these values have been determined, transform the entire configuration to $\mathbf{X}^*_i = \mathbf{X}_i\mathbf{H}^*$.
4. Determine the translation term as in the orthogonal generalized resistant-fit and translate each $\mathbf{X}^*_i$.
5. Set the current estimate of $\mathbf{H}_i = \mathbf{H}_i\mathbf{H}^*_i$ and set $\mathbf{X}_i = \mathbf{X}^*_i$.
6. Compute the new consensus configuration, $\mathbf{Y}^*$, and calculate the criterion as in ordinary generalized resistant-fit. If the criterion exceeds a specified tolerance, set $\mathbf{Y} = \mathbf{Y}^*$, rescale **Y**, and repeat the above procedure beginning with step 3. Otherwise, the iteration is complete.
7. As a final step, perform an SVD of the transformation matrix, $\mathbf{H}_i$, for each object to determine the values of p_i, q_i, θ_i, and ψ_i.

The residual sum of squares, S_r, can then be computed. However, affine transformation distorts the space in which the landmarks were originally recorded, and the comparison of the residual sum of squares from this fitting technique with those of others is not likely to be meaningful.

Comparison of Generalized Affine Least-Squares and Resistant-Fit Results

When objects differ due to affine transformations these differences can dominate

the results of orthogonal fitting procedures, thus obscuring other potentially interesting types of variation. This is illustrated in Figure 4 where the results of the orthogonal fitting of 10 long- and 10 short-nosed specimens of our hypothetical organisms are shown. The basic configuration consisted of the same 6 landmarks discussed earlier, except that in addition to scaling, rotation, and translation, each set of landmark coordinates was subjected to affine deformation of random magnitude along the long axis of the body as illustrated in Figure 3.

The results of neither orthogonal technique is very informative given our knowledge of the original data. As before, the least-squares method (Fig. 4A) has distributed differences due to nose length over the rest of the body—perhaps to an even greater extent because the direction of the affine deformations reinforced the differences in nose length. The overall pattern of the variation appears different as well, particularly at landmarks not directly on the axis of deformation. This is due to the differential scaling of longer versus shorter objects.

The orthogonal resistant-fit results (Fig. 4B) are better in that variability is more localized at landmark 1, but a similar nonrandom pattern of variability is evident. In this case, it is because the parameter estimates based on line segments connecting landmarks that are aligned with the major and minor axes of the deformation will tend to yield extreme values and thus be ignored in favor of more oblique segments. The outcome of this will be that, with respect to the primary axes of deformation, shorter configurations will have their displacements more toward the center of the consensus while relatively longer configurations will have more distal displacements.

Incorporating affine transformations into these fitting techniques effectively eliminates these problems (see Fig. 5). Here, in addition to the residuals, the reciprocal of the length of the major arm of the strain cross for each configuration is shown in the upper right corner superimposed upon

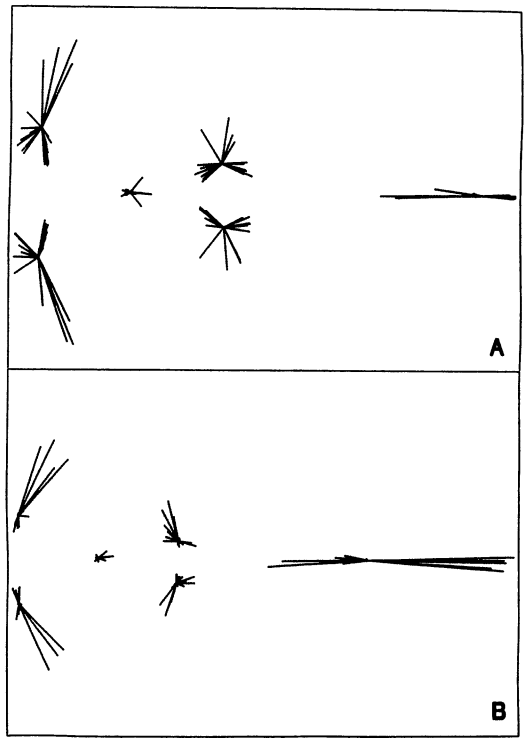

FIG. 4. Results of generalized, orthogonal (A) least-squares and (B) resistant-fit analyses of two populations of "organisms" subject to affine deformations. Groups differ in the relative position of landmark 1 and each configuration was subjected to an affine transformation of random magnitude in the direction of the long axis of the "body."

a unit circle. The strain cross lengths have been further scaled to have a constant, unit product (see next section for discussion). This scaling serves to emphasize the direction of variation since otherwise short strain cross arms, representing diminution in one direction, would be observed by longer arms in the plot. As far as shape is concerned, diminution in one direction is the same as expansion in the perpendicular direction. Both affine techniques correctly identify the primary axis of elongation. The least-squares procedure again tends to spread the lack of fit over the entire form, while resistant-fit correctly identifies most of the variation as being due to differences in the location of landmark 1. That the affine methods produce superior results in

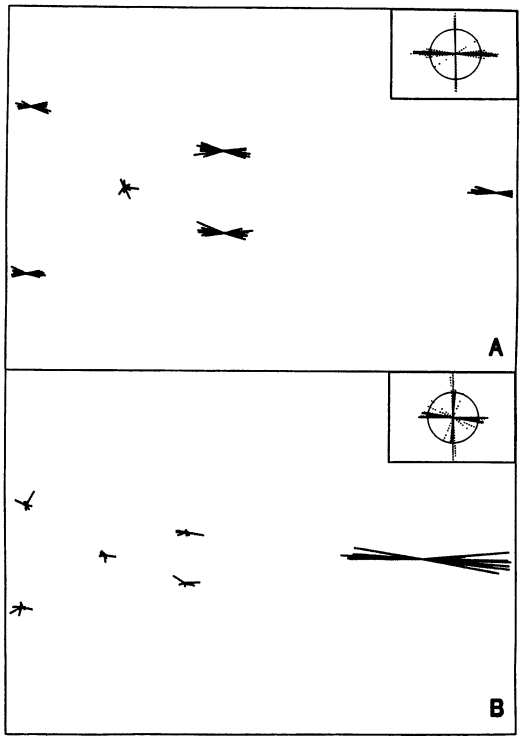

FIG. 5. Results of generalized, affine (A) least-squares and (B) resistant-fit analyses using the same data shown in Figure 4. Upper right corner in each shows the inverse of the scaled minor strain cross for each species.

this case should not be surprising as we have constructed the test data so as to maximize the contrast between the affine and orthogonal methods. However, even had there been no underlying affine deformations to account for, affine methods could still have been used. In this situation, strain cross parameters would have merely indicated isometric shape change, and their direction been indeterminate (computationally this is not a problem). It is worth noting, though, that should localized differences in landmark position be present in the data, this will also influence the least-squares strain cross estimates. This is evident in a comparison of Figures 4A and 5A. Subject to the affine least-squares procedure, the nonrandom pattern of variation seen in Figure 4A has been eliminated. This pattern, which was also described earlier in results from data not subject to affine deformation, has been absorbed to a great extent into the affine transformation parameters.

PRESENTATION OF RESULTS

Having superimposed a group of objects using one of the above techniques, one must next examine and analyze the results of the fitting. To this end it is desirable to summarize the results in ways that allow one to check the data for the presence of suspicious observations and interpret the differences and similarities of the objects being studied. These goals suggest the use of graphical techniques, of which the simplest and most direct is to plot and examine the residuals from the fitting process.

All the above procedures ultimately involve the fitting of each object, using some criteria, to a single configuration of landmarks. In generalized methods this is the consensus configuration, which is defined by the centroids of the individual landmark locations calculated from all objects under consideration. Otherwise, it is a specified reference configuration. In either case, this configuration can be plotted and deviations in the landmark locations of the objects shown as residual vectors from each reference landmark to the corresponding landmarks for each object. This has been the approach used by most authors concerned with the general class of techniques discussed here (Sneath, 1967; Siegel and Benson, 1982; Olshan et al., 1982; Goodall and Green, 1986; Sinervo and McEdward, 1988). Residual vectors can be inspected for extremely deviant observations that might indicate measurement or transcription errors or genuinely unique observations. The patterns of variation can suggest relationships between different parts of the object. This might, for example, indicate the correlated displacement of sets of landmarks in some objects or suggest a more general difference between objects, say affine deformations, that could be accounted for by the use of an alternative fitting technique.

With more than a few objects, one's abil-

ity to identify detailed patterns in the results diminishes, and techniques for summarizing residual variability become desirable. One approach is to treat fitted landmark locations as two-dimensional scatter plots around each reference landmark. The overall pattern of variation about each landmark can then be summarized by plotting the first and second principal component axes of the fitted landmark locations with respect to the reference landmark. These axes are the eigenvectors of the variance-covariance matrix of object landmark locations expressed as deviations from the reference landmark. Their associated eigenvalues indicate the magnitude of variation in each direction. Figure 6A shows such a plot for the results discussed in the previous section.

The plots of the PCAs can then be examined for indications of eccentricity in the scatter of residuals about the landmarks. Should major differences be indicated, the PCAs themselves would reveal the preferred direction of variability. Such a situation might arise if a particular landmark were physically constrained, say at the boundary between distinct regions of the object. In this case, variation in the location of that landmark between objects might be directed primarily along that boundary. Also, systematic differences between objects might create overall patterns of directed variation, for example, objects differing in shape due primarily to affine deformations.

An extension of the use of PCA plots is the construction of constant frequency ellipses (also shown in Fig. 6A). The major and minor axes of these ellipses are constructed using the eigenvectors and eigenvalues of the within landmark variance-covariance matrix. The same types of information as found in the plots of the PCAs can be seen in the plots of such ellipses. We have found it to be most informative to combine the two representations by plotting the PCAs scaled to have lengths of one standard deviation in their respective directions and overlay this with a two-standard deviation equal frequency ellipse. The PCA axes seem to give a better representation of the directionality and eccentricity of the scatter in the two directions, while the ellipses make it easier to visualize the magnitude of variation around a landmark.

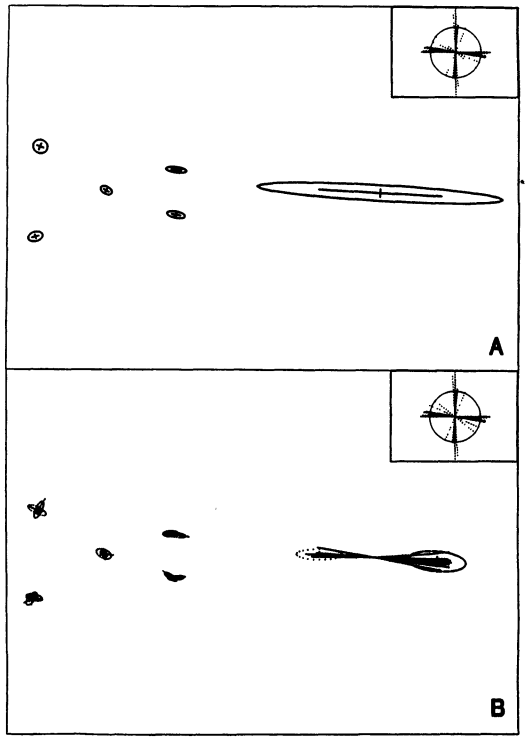

FIG. 6. Alternative representations of the results of the generalized, affine resistant-fitting of data used for Figure 5. A. Principal component axes and constant frequency ellipses. B. Residual vectors, PCAs, and constant frequency ellipses shown separately for long- and short-"nosed" groups. Upper right corner in each shows the inverse of the scaled minor strain cross for each object.

The real efficacy of these ways of summarizing scatter about individual landmarks is dependent upon the actual shape of the scatter on which they are based. For statistical consideration, the residuals should ideally be bivariate normal about each landmark. Since this is not necessarily the case, we emphasize the importance of considering individual residuals, as well as PCA axes and equal frequency ellipses, as part of the interpretive process.

Finally, the preceding discussion of data presentation has been oriented toward ex-

amination of a homogeneous, or purportedly homogeneous, sample of object configurations. It is often the case that one is dealing with replicated sets of objects, for example, data on members of different populations of a species or several species in the same genus. In this case, one is likely to be interested in differences between operationally defined groups. The use of such replicated observations is easily incorporated into the fitting techniques. One can treat the entire sample as homogeneous for the purposes of fitting and then distinguish between the groups at the time of analysis. In the case of residuals plots, one need only distinguish between the groups by some easily recognizable symbolism, e.g., color or linetype. Strong associations within groups might then be identified at specific landmarks, as shown in Figure 6B, where the long- and short-nosed specimens are distinguished in the results of the generalized affine resistant-fitting discussed earlier. Distinction of groups with the PCA or equal frequency ellipse representations is somewhat more complicated but potentially more informative. Here the reference is again either a consensus, or specified, configuration, and one is interested in how individual groups differ in their fit to the reference. After fitting, PCAs and equal frequency ellipses can be calculated for each group with respect to the centroid of that group's residuals about each landmark (also shown in Fig. 6B). These can be plotted and the groups compared as to the relative locations of their centroids and their patterns of within-group variation.

When objects are superimposed using the affine procedures, there is additional information available concerning shape differences between the configurations. This information is in the magnitude and direction of uniform shape differences and is contained in the strain cross parameters, p, q, θ, and ψ, derived from the affine transformation matrix. As was mentioned earlier, the value of θ will be a function of the initial orientation of the individual objects at the time the landmark locations were recorded. Only in special cases will this be interesting. The other angular measure, ψ, on the other hand, is of potential interest in that it indicates the direction of linear shape differences between the objects with respect to the common reference. The greatest expansion of an object due to the affine fitting will be in this direction and the least in the perpendicular direction.

This information can be graphically represented by overlaying plots of the strain crosses for each object. These can be shown as orthogonal axes scaled to lengths proportional to the magnification factors, p, and q. These crosses are oriented so that the longest axis is at an angle, ψ, to the horizonal. In practice, it is probably more useful to scale the axes to the inverse of the expansion factors. Thus, when presented alongside the fitted configuration it is possible to visualize the deformation necessary to restore each object to its original shape. It can also be useful to superimpose these strain crosses on a unit circle so that one can determine whether the deformations represented contractions or expansions of the original object.

One shortcoming of this representation is that there is no real distinction between effecting a shape change by absolute expansion in one direction versus absolute contraction in the perpendicular direction when constraints on size are ignored. In the generalized fitting procedures, however, a common size constraint is imposed, and objects are fit to an average form which results in the absolute expansion of some objects being balanced by contraction in others. When the strain cross arms, or their inverses, based on such a procedure are plotted it is quite easy for those associated with objects requiring expansion to obscure those of other objects requiring contraction. To address this we scale the expansion factors, p and q, to have a constant, unit product and plot only the longer axis (Fig. 6A, B, upper right). This representation emphasizes the primary directions of deformation and contains all relevant shape change information.

APPLICATION TO MOSQUITO WINGS

To illustrate the application of these techniques to real data, we present an analysis of the coordinates of 18 landmarks on

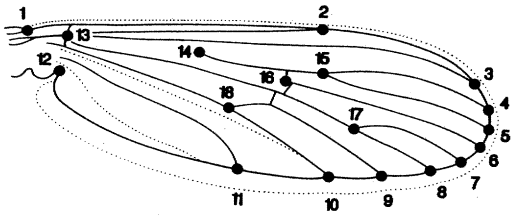

FIG. 7. Diagram of the wing venation of a mosquito (Culicidae). Dotted line around outside shows usual extent of fringe.

the wings of 127 species of North American mosquitoes. The landmarks are the intersections of wing veins with the wing margin, intersections of cross veins with major veins, vein branch points, the apparent anterior attachment of the wing to the body, and the notch between the posterior wing margin and the alula (Fig. 7). Representative drawings of females from Carpenter and LaCasse (1955) were xerographically enlarged 2.14 times. This is the same source of data used by Rohlf and Archie (1984). A Graf/Bar® acoustic digitizer was then used to record coordinates of landmarks on right wings from enlarged drawings. Each wing was digitized five times and the coordinates averaged after generalized least-squares fitting of the replicates. Relatively little measurement error was found. The mean square error was about 10^{-3} (compared to a mean square residual of 0.54 among the 127 species of mosquitoes). The final data set consisted of the average coordinates of 18 landmarks for 127 species.

The results of generalized least-squares and resistant-fitting of the mosquito wings are shown in Figure 8A and B. To contrast the results of the two methods the two-standard deviation ellipses and principal component axes from both methods are superimposed and presented in Figure 8C. The results from the two methods are very similar, however resistant-fitting assigns considerably less variation to the landmarks associated with the tip of the wing and much more to the three landmarks near the body—in contrast to the least-squares results.

Figure 9A and B show the results of generalized affine least-squares and resistant-

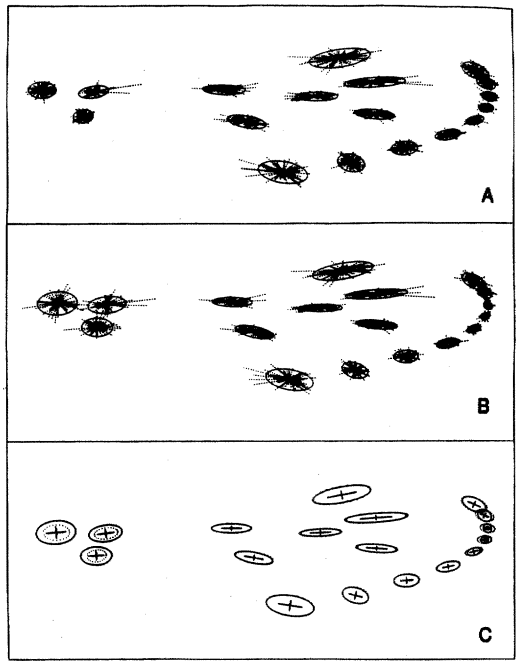

FIG. 8. Results of generalized (A) least-squares and (B) resistant-fit analyses for the wings of 127 species of North American mosquitoes. C. Principal components axes and equal frequency ellipses from both analyses superimposed for comparison (generalized Procrustes results shown with dotted lines).

fit procedures, respectively, with the scaled major strain cross arms for each species. The results are similar to those of the orthogonal procedures, except for a general reduction in variability at all landmarks. This is expected due to the inclusion of the extra affine transformation parameters in the fitting model. Examination of the strain cross arms gives little indication of a directional component to the shape differences between mosquito species in either magnitude or direction of difference. While more specific patterns might be identified in a more detailed comparison of individual species or genera, as a whole these results simply reaffirm those of the orthogonal methods.

There is relatively less variability around landmarks on the tip than at other parts of the wing. This suggests greater limitations on variability near the tip possibly related to aerodynamic constraints. While this is purely speculative, such constraints would

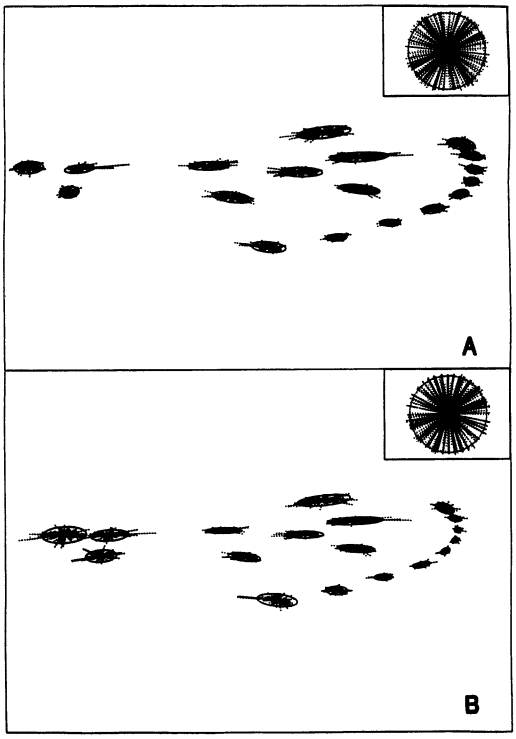

FIG. 9. Results of generalized affine (A) least-squares and (B) resistant-fit analysis of the wings of 127 species of North American mosquitoes. Upper right corner in each shows the inverse of the scaled minor strain cross for each species.

be consistent with the general importance of wing tip shape in vortex shedding (Vogel, 1981). While the difference in wing tip variability following least-squares and resistant-fit is quite clear, it might have been overlooked had the results of the two techniques not been compared. The least-squares fit, however, indicates that, despite the close proximity of the landmarks, much variability in the region is possible. All of the methods also suggest similar amounts and patterns of variability for the landmarks in the interior of the wing. There is a strong tendency for the major axis of variation to be directed parallel to the vein.

Two of the landmarks (2 and 11) located along the margin also show major variation along a preferred axis in each of the analyses. These appear to indicate that most of the variability across species, in these landmarks, is due to variation in the point of intersection of the veins along the margin of the wing. The minor axis of variability seems to be due to the anterior positioning of the front and posterior positioning of the hind wing margins relative to the other landmarks. Should these differences be positively correlated across species, it would indicate the differences between species with long slender wings and those with shorter, more robust ones. The angle of this major axis of variation in landmark 2 compared to the long axis of the wing implies that species with front wing margins more anteriorly located relative to other landmarks also have the intersection of the first longitudinal vein and the front margin more distally located. The converse holds for landmark 11.

DISCUSSION

All of the methods discussed above superimpose data representing the relative locations of distinct, presumably homologous landmarks so that differences in shape can be detected by studying residuals. The methods differ in the properties they optimize or their resistance to certain perturbations in the data. As a whole they possess a number of properties to recommend them for the analysis of shape differences among organisms. For example, they indicate quite graphically the differences in the relative physical locations of the landmarks. The results provide a summary that implies all of the possible interconnections between the data under consideration, not just between selected pairs of points as is the case with the analysis of distance measurements. In principle, all of the information that might be obtained from a traditional analysis is recoverable from the results since the original coordinates can be reconstructed. But information (such as the decomposition of shape differences into allometric and non-allometric components) from some types of analyses are not directly provided.

Bookstein et al. (1985) emphasize that the distance measures used in morphometrics should represent actual physical distances on the objects. The fitting meth-

ods discussed here and methods of assessing shape deformations such as biorthogonal grids, shape coordinates (Bookstein et al., 1985), or thin-plate splines give results that can be mapped directly onto a representative form of the organisms being studied. This allows for an intuitive, visual representation that is easily translatable into physical differences between organisms. Another advantage to the superimposition methods described in this paper is that they explicitly limit their results to the landmark locations themselves. Biorthogonal grids and thin-plate splines display interpolated shape changes in interlandmark areas. Since there is no information in the data pertaining to shape changes in these areas and there is no guarantee of a uniform shape change field in the intervening area, such results could be misinterpreted. Also, the relative conceptual and mathematical simplicity of the procedures described here allow for a more complete understanding of their properties and application by a broader range of users.

Each fitting method defines "best superimposition" differently. One must, therefore, decide how objects should be compared. The least-squares procedures superimpose the objects so that there is a minimum sum of squared distances between corresponding landmarks on different organisms. This criterion allows for the development of the direct solution for two objects and iterative procedures for n objects for any number of dimensions. The disadvantage of this approach is that if localized differences between forms are present then these differences are distributed over the entire configuration of landmarks. In many cases such local differences are likely to be of the greatest interest. Resistant-fit methods are robust to this type of localized variation in up to fifty percent of the landmarks. The resistant-fit approach is computationally much more expensive since, for each fitting, parameters must be estimated for all possible landmark pairs (or triplets in the case of affine fitting). In most studies in which these techniques are likely to be employed, the investigator will probably be interested in the precise nature of the differences between forms in the relative location of landmarks. There will also be no a priori reason to dismiss the possibility of localized variation in some of the landmarks. Indeed, the absence of such variation would eliminate the need for most such analyses. These considerations suggest the use of resistant-fit procedures.

In addition to the decision of whether to use least-squares or resistant-fit techniques, one has the choice of whether or not to incorporate an affine transformation into the fitting procedure. In general, its use seems desirable since the removal of any uniform shape differences will increase the likelihood of detecting more subtle local patterns. The primary disadvantage of including the affine transformation in the analysis is the increased computer time necessary to compute the necessary parameters. This is not such a problem for least-squares methods since the direct determination of the eigenvalues of **A** (equation 28) bypasses the need to iteratively fit each configuration to a consensus. In some cases this may actually reduce the computation time required. This is not the case in the resistant-fit procedure where the increase in computation time over an orthogonal fitting can be substantial.

Another important aspect to be considered in the selection of methods as well as in the interpretation of the results is the effect of configuration geometry. Geometry affects the fitting methods in different ways. Points farther away from the centroid tend to dominate estimates of scaling, translation and rotation in the least-squares procedures just as points farther from the mean dominate the determination of the slope in a linear regression. With resistant-fit methods, on the other hand, these points virtually are ignored and generally are allocated increasing variability with increasing distance from the main constellation of landmarks. Such consideration might account for part of the differential allocation of variability to wingtip versus near body landmarks in the analysis of North American mosquito wings presented above.

We are currently investigating these problems and hope to develop corrections for these artifacts.

The affine resistant-fit technique seems to be the most generally desirable approach. Its insensitivity to local variation in the determination of overall scaling, translation, and rotation parameters and the isolation of affine transformational differences into separate, interpretable parameters provide the greatest possibility for detecting important differences between sets of landmarks. On the other hand, this approach is computationally the most expensive and its results the most affected by landmark geometry. The former is less of a problem since data sets up to 150 objects and 25 landmarks can be processed using our program GRF for the IBM PC microcomputer. The latter is currently being studied. At this point it seems advisable that one try both least-squares and resistant-fit methods and with both affine and orthogonal transformations—and compare the results.

ACKNOWLEDGMENTS

The manuscript benefited greatly from discussions with numerous people when it was presented at the International Conference on Numerical Taxonomy in 1987. Discussions at the Morphometrics Workshop at the University of Michigan, May 1988, were also very helpful. Fred Bookstein's critical comments (especially regarding uniform shape change) on several versions of the manuscript resulted in important extensions to its scope. R. E. Strauss and S. Hartman also provided helpful reviews of the manuscript. We would also like to thank Richard Jensen for his helpful comments on the manuscript and on our program, GRF, that implements the methods described in this paper. That program, for the IBM PC, is available upon request.

This work was supported, in part, by the National Science Foundation under grant BSR 8306004 to F.J.R. The computations were aided by the use of an IBM 4361 computing system made available in part by grant BSR 8318792 from the National Science Foundation. This is contribution number 743 from the Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook.

REFERENCES

BENSON, R. H., R. E. CHAPMAN, AND F. SIEGEL. 1982. On the measurement of morphology and its change. Paleobiology, 8:328–339.

BOOKSTEIN, F. L. 1978. The measurement of biological shape and shape change. Lecture notes in biomathematics. Volume 24. Springer-Verlag, New York. 191 pp.

BOOKSTEIN, F. L. 1985. Transformations of quadrilaterals, tensor fields, and morphogenesis. Pages 221–265 *in* Mathematical essays on growth and the emergence of form (P. L. Antonelli, ed.). Univer. Alberta Press, Alberta.

BOOKSTEIN, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. I.E.E.E. Trans. Pattern Anal. Mach. Intelligence (in press).

BOOKSTEIN, F. L., AND P. D. SAMPSON. 1987. Statistical models for geometric components of shape change. Pages 18–30 *in* Proceedings of the Section on Statistical Graphics, San Francisco, August 1987. American Statistical Association, Alexandria, Virginia.

BOOKSTEIN, F. L., B. C. CHERNOFF, R. L. ELDER, J. M. HUMPHRIES, G. R. SMITH, AND R. E. STRAUSS. 1985. Morphometrics in evolutionary biology. Special Publ. 15, Academy of Natural Sciences of Philadelphia, Philadelphia. 277 pp.

CARPENTER, S. J., AND W. J. LACASSE. 1955. Mosquitoes of North America (North of Mexico). Univ. California Press, Los Angeles. 487 pp.

CATTELL, R. B., AND D. K. KHANNA. 1977. Principles and procedures for unique rotation in factor analysis. Chapter 9 *in* Mathematical methods for digital computers. Volume III (K. Einstein and A. Ralston, eds.). Wiley-Interscience, New York.

FERSON, S., F. J. ROHLF, AND R. K. KOEHN. 1985. Measuring shape variation among two-dimensional outlines. Systematic Zool., 34:59–68.

GOODALL, C. R. 1983. The statistical analysis of growth in two dimensions. Ph.D. Dissertation, Harvard Univ., Cambridge, Massachusetts. 150 pp.

GOODALL, C. R., AND P. B. GREEN. 1986. Quantitative analysis of surface growth. Bot. Gaz., 147:1–15.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53:325–338.

GOWER, J. C. 1971. Statistical methods of comparing different multivariate analyses of the same data. Pages 138–149 *in* Mathematics in the archaeological and historical sciences (F. R. Hodson, D. G. Kendall, and P. Tautu, eds.). Edinburgh Univ. Press, Edinburgh.

GOWER, J. C. 1975. Generalized Procrustes analysis. Psychometrika, 40:33–51.

GOWER, J. 1984. Multivariate analysis: Ordination, multidimensional scaling and allied topics. Pages 727–781 *in* Handbook of applicable mathematics. Volume VI: Statistics (E. Lloyd, ed.). Wiley, New York.

HURLEY, J. R., AND R. B. CATTELL. 1962. The Procrustes program: Producing direct rotation to test an hypothesized factor structure. Behavioural Sci., 7:258–262.

OLSHAN, A. F., A. F. SIEGEL, AND D. R. SWINDLER. 1982. Robust and least-squares orthogonal mapping: Methods for the study of cephalofacial form and growth. Amer. Jour. Physical Anthropology, 59:131–137.

ROHLF, F. J., AND J. ARCHIE. 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). Systematic Zool., 33:302–317.

SIEGEL, A. F. 1982a. Robust regression using repeated medians. Biometrika, 69:242–244.

SIEGEL, A. F. 1982b. Geometric data analysis: An interactive graphics program for shape comparisons. Pages 103–122 *in* In Modern data analysis (R. L. Launer and A. F. Siegel, eds.). Academic Press, New York.

SIEGEL, A. F., AND R. H. BENSON. 1982. A robust comparison of biological shapes. Biometrics, 38:341–350.

SINERVO, B., AND L. R. MCEDWARD. 1988. Developmental consequences of an evolutionary change in egg size: An experimental test. Evolution, 42:885–899.

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

THOMPSON, D. W. 1917. On growth and form. Cambridge, London. 793 pp.

VOGEL, S. 1981. Life in moving fluids: The physical biology of flow. Princeton Univ. Press, Princeton. 352 pp.

Received 1 March 1989; accepted 26 November 1989