

Ability of Geometric Morphometric Methods to Estimate a Known Covariance Matrix

JEFFREY A. WALKER

Department of Zoology, Field Museum of Natural History, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA; E-mail: walker@usm.maine.edu

Current Address: Department of Biology, University of Southern Maine, 96 Falmouth St., Portland, ME 04103

Abstract.—Landmark-based morphometric methods must estimate the amounts of translation, rotation, and scaling (or, nuisance) parameters to remove nonshape variation from a set of digitized figures. Errors in estimates of these nuisance variables will be reflected in the covariance structure of the coordinates, such as the residuals from a superimposition, or any linear combination of the coordinates, such as the partial warp and standard uniform scores. A simulation experiment was used to compare the ability of the generalized resistant fit (GRF) and a relative warp analysis (RWA) to estimate known covariance matrices with various correlations and variance structures. Random covariance matrices were perturbed so as to vary the magnitude of the average correlation among coordinates, the number of landmarks with excessive variance, and the magnitude of the excessive variance. The covariance structure was applied to random figures with between 6 and 20 landmarks. The results show the expected performance of GRF and RWA across a broad spectrum of conditions. The performance of both GRF and RWA depended most strongly on the number of landmarks. RWA performance decreased slightly when one or a few landmarks had excessive variance. GRF performance peaked when ~25% of the landmarks had excessive variance. In general, both RWA and GRF performed better at estimating the direction of the first principal axis of the covariance matrix than the structure of the entire covariance matrix. RWA tended to outperform GRF when >~75% of the coordinates had excessive variance. When <75% of the coordinates had excessive variance, the relative performance of RWA and GRF depended on the magnitude of the excessive variance; when the landmarks with excessive variance had standard deviations (σ) $\geq 4\sigma$ minimum, GRF regularly outperformed RWA. [Covariance structure; partial warps; principal components analysis; relative warps analysis; resistant fit; shape; simulation; superimposition.]

Landmark-based geometric morphometrics have been increasingly exploited during the past decade to explore systematic, developmental, ecological, or pathological differences among individuals or populations. To compare shape among digitized configurations of landmarks, or figures, one must estimate and remove from the data variation attributable to the arbitrary position, orientation, and size of the figures. Two frequently used classes of methods that remove nonshape variation from a set of figures are superimposition analysis and thin-plate spline (TPS) decomposition analysis (Rohlf and Bookstein, 1990; Bookstein, 1991; Marcus et al., 1993, 1996). Within superimposition analysis itself, several different methods have been proposed, including the two-point registration (Bookstein, 1986), generalized least-squares (Procrustes) superimposition (Sneath, 1967; Rohlf and Slice, 1990), and various resistant-fit superimpositions (Siegel and Benson, 1982; Rohlf and Slice, 1990; Dryden and Walker, 1999).

Given the availability of different methods, biologists presumably have to choose among methods to analyze their data. But

over the past few years, many morphometricians have reached a consensus that statistics should not be performed on resistant-fit residuals but on the generalized Procrustes residuals (residuals from the generalized least-squares [GLS] superimposition) or on some linear combination of these residuals, such as partial warp scores (Bookstein, 1996a; Rohlf, 1996). This is because a figure superimposed by GLS corresponds to a point in Kendall's shape space, a curved space containing all possible shapes with p landmarks in k dimensions. By this definition of shape, then, the residuals from superimpositions other than GLS contain nonshape (translation, rotation, and scale) variation (Bookstein, 1996a).

However, least-squares superimpositions can sometimes give a misleading visual image of how a pair of figures differ from each other (Siegel and Benson, 1982; Rohlf and Slice, 1990), especially when most of the shape differences between a pair of figures are limited to a small subset of landmarks. The poor estimation of the least-squares superimposition reaches its extreme when all of the shape variation occurs at a single

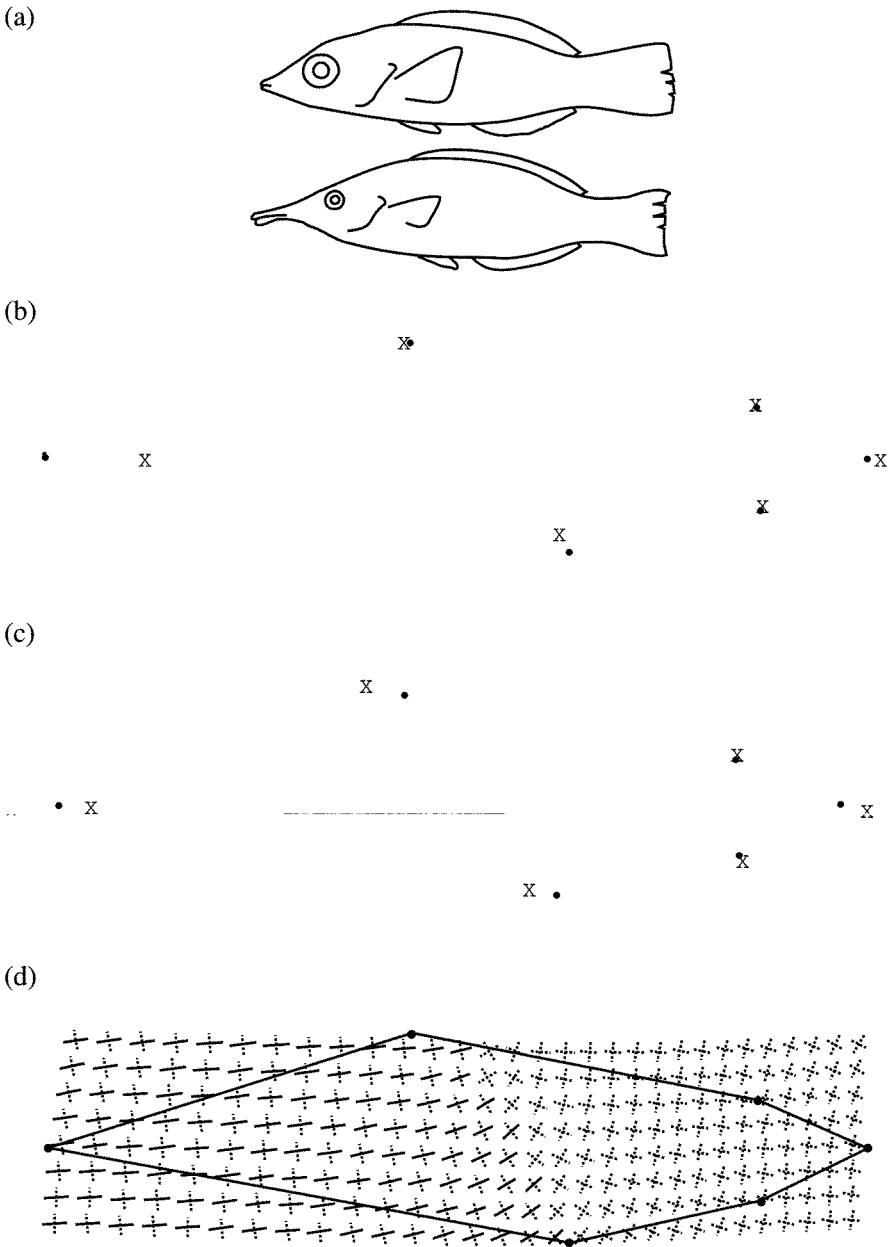


FIGURE 1. Example of the Pinocchio effect. Outlines of (a) juvenile (top) and initial phase (bottom) *Gomphosus varius*, scaled to the same standard length, suggest that most of the shape change during growth occurs at the snout. The (b) resistant-fit superimposition supports this hypothesis but the (c) least-squares and (d) TPS suggest substantial postanal shape change (contraction). The TPS is displayed as a field of strain crosses. The major and minor axes of each cross indicate the major and minor axes of deformation of the equivalent grid cell. Solid lines indicate directions of extension and dashed lines indicate directions of contraction.

landmark, which may be called the Pinocchio effect. For example, during growth of the bird wrasse, *Gomphosus* sp., the snout greatly elongates but the postrostral body shape changes very little. The repeated-medians resistant fit (Siegel and Benson, 1982) cor-

rectly locates most of the shape difference between a juvenile and initial phase bird wrasse at the snout, whereas a least-squares superimposition spreads this difference at the snout across many of the other landmarks (Fig. 1). Note that a TPS deformation

suggests the same shape difference as that implied by the least-squares superimposition (Fig. 1).

Despite its apparent usefulness at locating shape differences, resistant-fit superimpositions have been largely ignored with the acceptance of Kendall's shape space and its relationship to GLS and TPS decomposition (Bookstein, 1996a). Bookstein (1996a) stated that, at best, resistant-fit residuals should be used only to suggest how figures differ and should never be used in any statistical analysis, such as those of Walker (1993, 1996, 1997).

Several statistical methods, however, such as principal components analysis (PCA), are nothing more than sophisticated, exploratory methods to summarize patterns of shape variation that are too difficult to recognize from superimposed figures or a series of TPSs. If a least-squares superimposition can mistakenly identify the location of shape difference between a pair of figures, then a PCA of GLS residuals can potentially result in poor estimates of the principal directions of shape variation. This might occur, for example, if one landmark had much greater variation than other landmarks. Because the combination of partial warp scores (Rohlf, 1993) and standard-formula uniform scores (Bookstein, 1996b) are simply the GLS residuals recombined into fewer variables (Rohlf, 1999), a PCA of the partial warp/uniform scores, or relative warp analysis (RWA), not only should be equivalent to a PCA of the GLS residuals but also should give equally misleading results under certain patterns of variation.

PCA results from superimposition and TPS decomposition methods have been compared with real data sets (Walker, 1996), but their performance with simulated data sets has not been evaluated. The major advantage of simulated data sets is that patterns of differences among figures are known in advance. In this paper, I use a simulation experiment to compare the efficacy of different geometric morphometric methods to reconstruct known patterns of shape variation within a sample of figures.

The general method followed in the simulation was to construct a set of figures with a known covariance matrix, compute a model covariance matrix from the residual variation estimated by a method, and quantitatively compare the pattern of covariances between the model and known covariance

matrices. The efficacy of geometric morphometric methods to estimate a known covariance matrix was explored for two different methods: RWA and the repeated-medians generalized resistant fit (GRF) (Rohlf and Slice, 1990). The effects of four different variables on the efficacy of the methods were also explored: number of landmarks, magnitude of the standardized covariance (correlation) within the covariance matrix, the number of coordinates with excessive variance relative to the variance of other coordinates, and the magnitude of the excessive variance.

METHODS

Construction of the Samples

For each simulation, N figures with p landmarks in k dimensions were constructed from random orthogonal matrices with specified eigenvalues. For all simulations, N and k were held constant at 100 and 2, respectively. p , or landmark count (LC), varied between 6 and 20 in intervals of 2.

A $pk \times pk$ random, orthogonal matrix, E , was computed (Heiberger, 1978; Tanner and Thisted, 1982).

To control the average correlation, σ , within the random covariance matrix, the pk positive eigenvalues, diagonal elements of L , were used to construct an initial covariance matrix, S ,

$$S = ELE^T \quad (1)$$

The general magnitude of the correlation structure of the known covariance matrix was controlled by varying L , with the j th diagonal element (eigenvalue) equal to

$$l_j = (2q + 2)^{\frac{1}{j-1}} + N \left(0, \frac{(2q + 2)^{\frac{1}{j-1}}}{10} \right) \quad (2)$$

where q , or correlation level (CL), varied between 0 and 1. Thus, the first part of the right-hand side varied from 1, 1/2, 1/4, 1/8... at $q = 0$ to 1, 1/4, 1/16, 1/64... at $q = 1$. Ten uniformly spaced values of CL were simulated.

The number of landmarks with excessive variance was controlled by modifying the variances (and associated covariances) within S . First, the mean standard deviation (square root of a diagonal element of S), σ , was computed. Second, S was transformed

into a correlation matrix. Third, $(1 - u)(p - 1) + 1$ randomly chosen landmarks were assigned excessive variance, where u , or Pinocchio level (PL), varied between 0 and 1. Ten uniformly spaced values of PL were simulated, except when $LC < 10$, in which case p values were simulated. The excessive variance was randomly partitioned among the x and y coordinates of the landmark, $\hat{s}_x^2 = r(v\sigma)^2$ and $\hat{s}_y^2 = (2 - r)(v\sigma)^2$, where r is a uniform random number between 0 and 2, and v , or variance magnitude (VM), was set to 2, 4, or 6. Fourth, the coordinates of the remaining landmarks were assigned the variance σ^2 . The total root variance at a landmark (square root of the sum of the coordinate variances) with excessive variance was, therefore, 2, 4, or 6 times the total root variance of landmarks without excessive variance. Fifth, the vector of coordinate variances was rescaled to again have mean σ^2 . Finally, the correlation matrix was transformed back into a covariance matrix, given the new set of perturbed variances. This new covariance matrix, \tilde{S} , was decomposed into its eigenvectors and eigenvalues by

$$\tilde{S} = \tilde{E}\tilde{L}\tilde{E}^T \tag{3}$$

Note that at $u = 0$, all p landmarks were given excessive variance, which, because of the rescaling to the original σ^2 , is equivalent to none of the landmarks having excessive variance. As u increased, fewer variables were given excessive variance; at $u = 1$, only a single variable was given excessive variance. To illustrate the effect of VM, Figure 2 shows three samples of figures with the same mean figure but with different of VM values applied to a single landmark.

A random mean figure, \bar{x} , was constructed from an array of pk elements picked randomly from a uniform distribution with values between -20 and 20 . \bar{x} is a vector with the first p elements arbitrarily assigned to the x coordinates and the last p elements arbitrarily assigned to the y coordinates. A matrix, Y , was constructed with pk columns of N random normal deviates, $N(0, I_j)$, where I_j is the j th diagonal element of \tilde{L} . The N figures with the population covariance matrix, \tilde{S} , were computed by

$$X = \bar{X} + Y\tilde{E}^T \tag{4}$$

where X is the matrix of figures, and \bar{X} is a matrix with \bar{x} in each row. The entire set

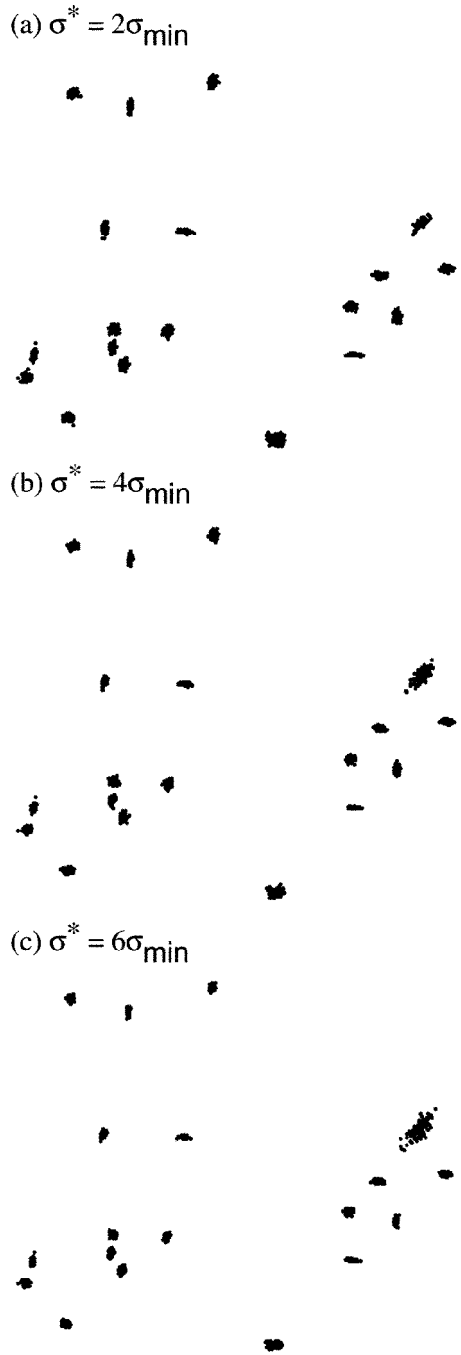


FIGURE 2. Comparison of three samples, each with only a single coordinate having excessive variance. The total root VM of the landmark with excessive variance is (a) 2x, (b) 4x, and (c) 6x that of the other landmarks.

of N figures was rotated so that the principal axes of \bar{x} were aligned with the x and y Cartesian axes. This last rotation was necessary for computation of the standard formula uniform scores (Bookstein, 1996b).

The known, or reference, covariance matrix, \hat{S} , was computed from the rotated figures.

One hundred random covariance matrices were constructed for each combination of LC (8 values), CL (6–10 values), PL (10 values), and VM (3 values)—for a total of 222,000 random matrices. Because CL had little effect (see below), results for all values of CL were pooled, resulting in 1,000 covariance matrices within each combination of LC \times PL \times VM.

Estimation of Model Covariances

Model covariances were estimated from superimposed figures and from TPS decompositions. For the superimposition analysis, nuisance variables that scaled all figures to a common size, rotated all figures to a common orientation, and translated all figures to a common alignment were estimated by the repeated-medians GRF (Rohlf and Slice, 1990; Slice, 1996). The figures were fit to the mean figure, as suggested by Walker (1997), and not the median figure, as originally implemented in Rohlf and Slice (1990). After the superimposition, the model covariance matrix among the coordinates, S_{GRF} , was computed for comparison with \hat{S} .

It may seem confusing that the figures were superimposed by using a method based on repeated medians but the mean figure was used as the consensus. Two different types of distributions are important. The distribution of differences in the angle or length of corresponding chords between a pair of figures (the variables used to find the rotation and scaling parameters) will have distinct outliers if shape differences are limited to one or a few of the landmarks. The use of the repeated median of this distribution estimates an affine parameter that is robust to these outliers. The second type of distribution, that of a coordinate for a sample of superimposed figures, is independent of the distributions of angle and chord length differences. That is, even if one landmark has excessive variation, the variation around this landmark may have a normal (gaussian) distribution. Hence, the choice of using a mean or median consensus figure should be based only on the shape of the distributions of superimposed coordinates and not on the decision to use least-squares or repeated medians to estimate the affine parameters. Be-

cause I sampled from normal distributions in this simulation, I used the mean figure as the consensus.

For the TPS decomposition, the set of principal warps and standard-formula uniform coefficients were computed from the mean figure after the GLS superimposition. Principal warps, E_w , were computed from the orthogonal decomposition of the bending energy matrix of the mean figure (Bookstein, 1991). The two columns of coefficients, u_1 and u_2 , which give the standard formula for the uniform component of shape (Bookstein, 1996b), were appended to the principal warps, resulting in a matrix with the form

$$E_{w,u} = \begin{bmatrix} E_w & 0 \\ & u_1 & u_2 \\ 0 & & E_w \end{bmatrix} \quad (5)$$

The projection of X_{GLS} onto $E_{w,u}$ results in the matrix of partial warp and uniform scores, W . The covariance matrix, S_w , of these scores cannot be compared with \hat{S} because the variables are in different spaces. The covariance matrix of W in the space of the original coordinates, which can be compared with \hat{S} , is

$$S_{RWA} = E_{w,u} R D R^T E_{w,u}^T \quad (6)$$

where R and D , the relative warps and relative warp eigenvalues, are found in the usual way,

$$S_w = R D R^T \quad (7)$$

Comparisons of Covariance Structure

The similarities between \hat{S} on the one hand, and S_{GRF} and S_{RWA} on the other, were compared in two ways. First, the overall similarity was compared by using a matrix correlation, which is simply the Pearson product-moment correlation between the off-diagonal elements of each matrix. The directions of the first eigenvector of \hat{S} with the first eigenvectors of S_{GRF} and S_{RWA} were compared by using a simple dot product, the cosine of the angle between the two vectors. The matrix correlations are denoted R_{GRF} and R_{RWA} , and the first eigenvector correlations are $r1_{GRF}$ and $r1_{RWA}$.

A matrix correlation between the model and known covariance matrix is a statistic that measures the similarity in the shape of

the covariance structure. A bootstrap was used to test if the correlation was less than expected by randomly sampling from the known population. For each combination from a subset of combinations of LC, PL, and VM, a bootstrapped pseudosample was constructed by choosing, with replacement, N figures from the sample. The covariance matrix was estimated by using the GRF and RWA methods, and the matrix correlations with two different reference covariance matrices were computed: R_{known} , the correlation between the resampled covariance matrix and the known covariance matrix, \hat{S} , and R_{obs} , the correlation between the resampled covariance matrix and the corresponding observed covariance matrix, S_{GRF} or S_{RWA} . The resampling was repeated 100 times and the lower 5th percentile of the distribution of R_{known} and R_{obs} was computed. The lower confidence interval for R_{obs} reflects only sampling error whereas that for R_{known} reflects error in both sampling and estimation of nuisance parameter (or model). Because of the small number of pseudosamples, the lower 5th percentile will be a relatively rough approximation of the true percentile.

The simulation was coded in Pascal by using Code Warrior Academic Pro Release 3 on an Apple Macintosh G3 computer. All code is freely available from the author upon request.

RESULTS

Pooling across all factors, gave median matrix correlations for the GRF and RWA estimates of \hat{S} , of 0.87 and 0.84, respectively. The median first eigenvector correlations were 0.94 and 0.92, respectively.

CL proved an effective method for controlling the magnitude of the average absolute correlation, $\bar{\rho}$, within the known covariance matrix \hat{S} , ($\bar{\rho} = 0.417 + 0.013q$, $R^2 = 0.52$). Pooling across LC, PL, and VM, R_{RWA} increased ($P < 0.001$), whereas R_{GRF} decreased with ($P < 0.001$). Nevertheless, $\bar{\rho}$ explained very little of the variance (RWA, $R^2 = 0.0008$; GRF, $R^2 = 0.002$); accordingly all subsequent analyses are based on the data pooled across CL.

LC had a strong but distinctly nonlinear effect on matrix correlations. Pooling across CL, PL, and VM, R_{RWA} and R_{GRF} increased with LC, although the rate of increase decreased with p . Quadratic regres-

sions for both the RWA and GRF models had significant linear and quadratic components (all $P < 0.0001$). The regression coefficients were remarkably similar between the RWA and GRF models, a pattern that masked the differences between the two methods that were due to the effects of PL and VM (see below).

PL and VM had significant effects on matrix correlations after pooling results for all values of LC and CL. R_{RWA} increased with PL at the lowest VM but decreased with PL at the two higher settings of VM ($2\sigma_{\text{min}}$: slope = 0.004, $P < 0.0001$, $R^2 = 0.008$; $4\sigma_{\text{min}}$: slope = -0.0014, $P < 0.0001$, $R^2 = 0.0007$; $6\sigma_{\text{min}}$: slope = -0.0068, $P < 0.0001$, $R^2 = 0.016$). In contrast, R_{GRF}

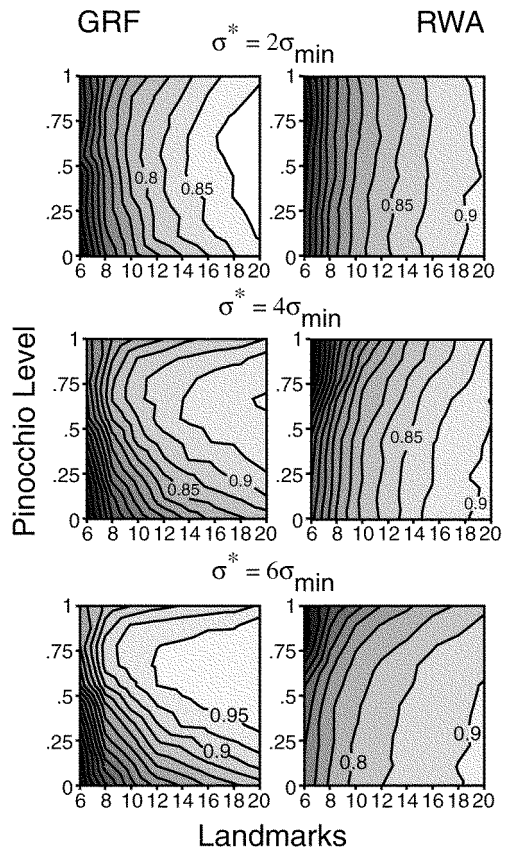


FIGURE 3. Ability of GRF and RWA to reconstruct the known covariance matrix. The height of the surface is the median of the 1,000 matrix correlations between the model covariance matrix and the known covariance matrix for each combination of LC \times PL. The total root VM of the landmarks with excessive variance, relative to that of the other coordinates, is 2 \times for the surfaces in the top panel, 4 \times for those in the middle panel, and 6 \times for those in the bottom panel.

increased with PL for each value of VM ($2\sigma_{\min}$: slope = 0.0076, $P < 0.0001$, $R^2 = 0.026$; $4\sigma_{\min}$: slope = 0.016, $P < 0.0001$, $R^2 = 0.12$; $6\sigma_{\min}$: slope = 0.02, $P < 0.0001$, $R^2 = 0.19$). Note that PL had a much stronger effect on R_{GRF} than R_{RWA} .

The simultaneous effects of PL and LC on the median matrix correlation, for each combination of VM \times MODEL, are illustrated with contour plots (Fig. 3). With VM set at $2\sigma_{\min}$, R_{RWA} and R_{GRF} increased sharply with LC but changed little with PL. At all combinations of PL and LC, the median R_{RWA} was greater than the median R_{GRF} .

At $4\sigma_{\min}$, PL had much more influence on matrix correlations than at $2\sigma_{\min}$, especially for R_{GRF} . R_{GRF} increased to a peak around PL = 0.75. At PL $> \sim 0.4$, R_{RWA} decreased slightly with increasing PL. RWA performed better than GRF at values of PL $> \sim 0.25$ but worse than GRF at values $> \sim 0.25$.

The patterns of R_{GRF} and R_{RWA} observed at $6\sigma_{\min}$ are simply an exaggeration of those at $4\sigma_{\min}$. In general, however, R_{GRF} was substantially greater than R_{RWA} when PL was $> \sim 0.4$. Even at very low LC ($p < 10$), R_{GRF} was > 0.8 for samples in which most of the shape differences were confined to a few landmarks.

GRF and RWA performance were directly compared throughout the PL \times LC space by using the percentage of trials in which R_{GRF} was greater than R_{RWA} (Fig. 4). When VM was $2\sigma_{\min}$, GRF fared poorly relative to RWA everywhere in the space. At $4\sigma_{\min}$ and $6\sigma_{\min}$, GRF performed better than RWA over much of the space. At these high VM values, RWA performed better than GRF ($> 50\%$ of the time) only when excessive variance was given to $> 75\%$ of the landmarks.

The simultaneous effects of PL and LC on the median first eigenvector correlation

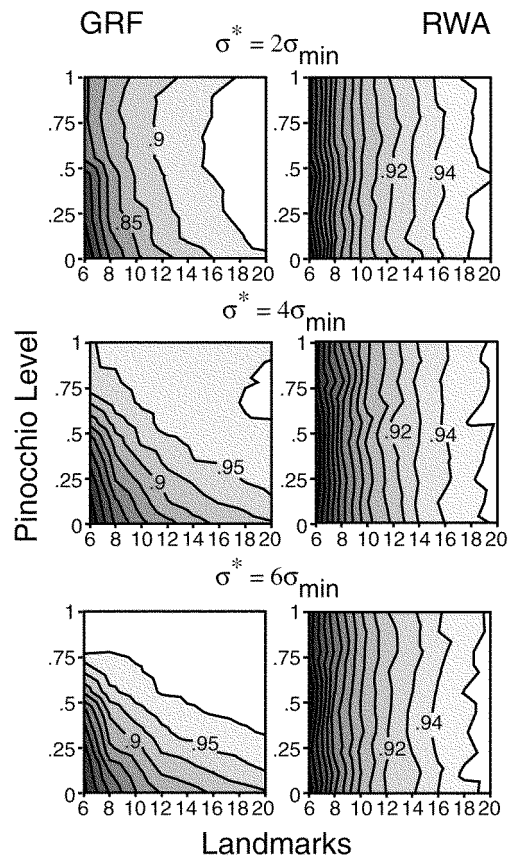


FIGURE 5. Ability of GRF and RWA to estimate the direction of the principal axis of variation. The height of the surface is the median of the 1,000 first eigenvector correlations between the model vector and the known vector for each combination of LC \times PL. The total root VM of the landmarks with excessive variance, relative to that of the other coordinates, is $2\times$ for the surfaces in the top panel, $4\times$ for those in the middle panel, and $6\times$ for those in the bottom panel.

show patterns similar to the effects on the matrix correlation (Fig. 5). In general, however, the first eigenvector correlations, $r1_{GRF}$ and $r1_{RWA}$, were greater than the corresponding

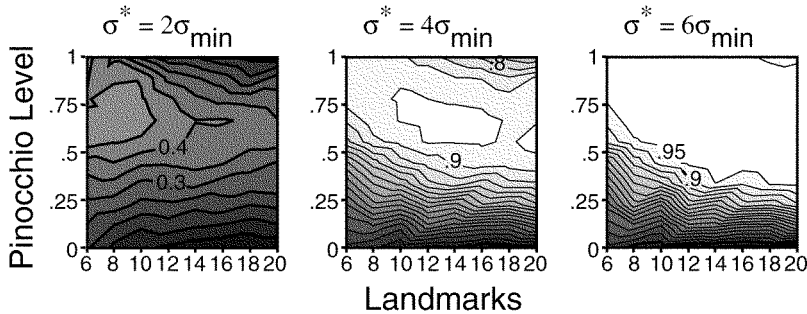


FIGURE 4. Direct comparison of GRF and RWA across the PL \times LC parameter space. The height of the surface is the percentage of times, in 1,000 cases, that $R_{GRF} > R_{RWA}$.

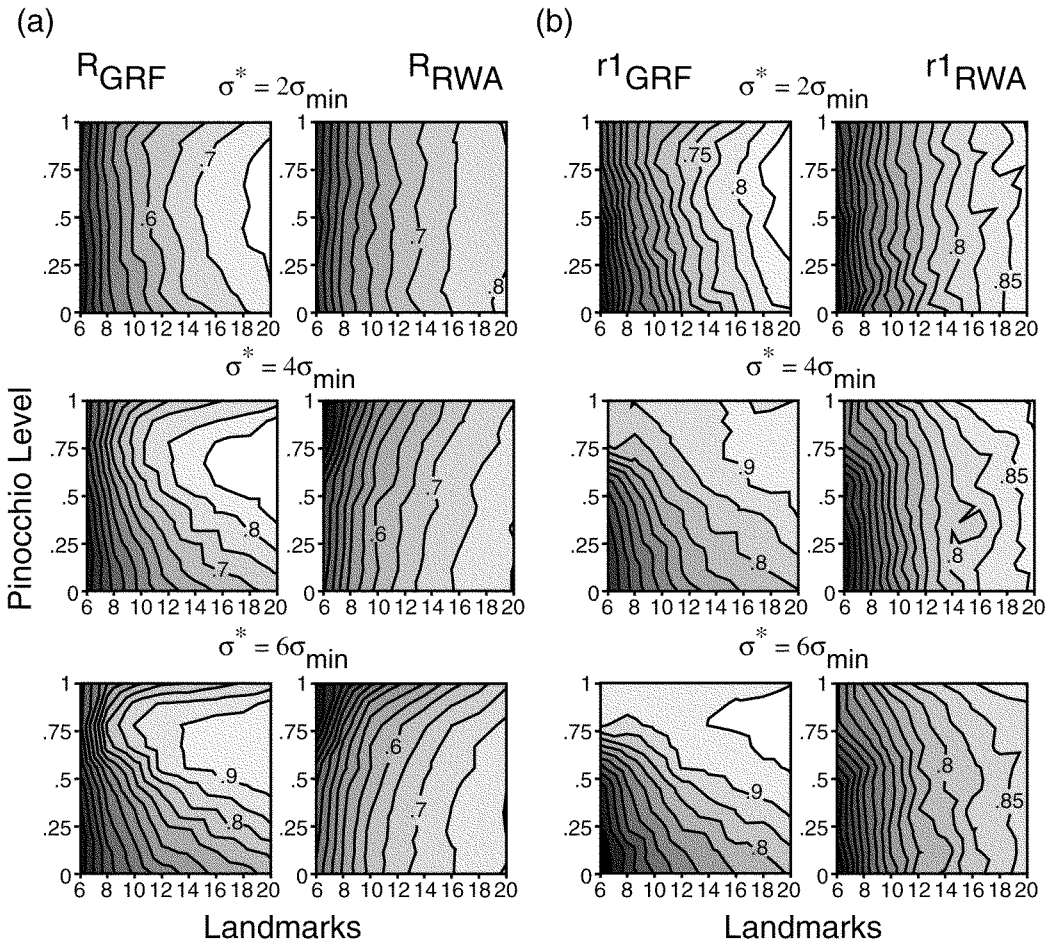


FIGURE 6. Lower confidence limit on the ability of GRF and RWA to reconstruct the known covariance matrix and the known major axis of variation. The height of the surface is the lower 5th percentile of the 1,000 matrix correlations (a) or first eigenvector correlations (b) between the model covariance matrix and the known covariance matrix for each combination of $LC \times PL$. The total root VM of the landmarks with excessive variance, relative to that of the other coordinates, is $2 \times$ for the surfaces in the top panel, $4 \times$ for those in the middle panel, and $6 \times$ for those in the bottom panel.

R_{GRF} and R_{RWA} . Additionally, although PL had a small effect on R_{RWA} , especially at high VM, the magnitude of $r1_{RWA}$ was largely independent of PL and VM.

The lower 5th percentiles of the R_{GRF} , R_{RWA} , $r1_{GRF}$, and $r1_{RWA}$ distributions, as a function of PL and LC, are very similar to the patterns for the medians (Fig. 6). These percentiles show that both GRF and RWA can frequently result in very poor estimates (<0.5) of the known covariance matrix when $LC \leq 8$. GRF and RWA estimates of the known major axis of variation are, in general, much better than estimates of the known covariance matrix. Indeed, when excessive

variance of at least $4\sigma_{min}$ occurs at fewer than 25% of the landmarks, GRF estimates of the first eigenvector are $>0.8 > 95\%$ of the time, even when only six to eight landmarks were measured.

The lower 5th percentile of the matrix correlation between the pseudosample covariance matrix and the observed covariance matrix, $R_{5,obs}$, was generally >0.95 regardless of LC (Fig. 7). In contrast, the lower 5th percentile of the matrix correlation between the pseudosample covariance matrix and the known covariance matrix, $R_{5,known}$, was substantially <0.95 and strongly dependent on LC.

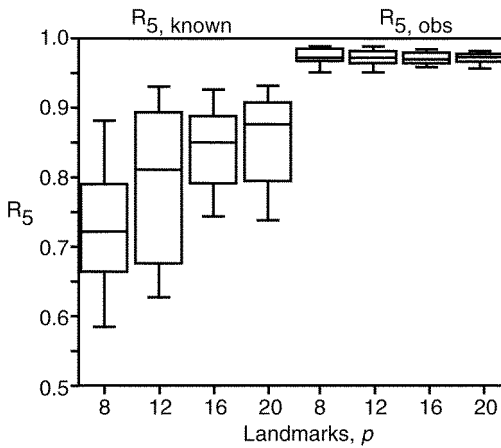


FIGURE 7. Comparison of error in the estimation of the covariance structure that is due to sampling and error that is due to the combination of nuisance parameter estimation and sampling. $R_{5,known}$ is the lower 5th percentile of the distribution of matrix correlations between a pseudosample and the known covariance matrix; it reflects both sampling and model error. $R_{5,obs}$ is the lower 5th percentile of the distribution of matrix correlations between a pseudosample and the observed covariance matrix; it reflects only sampling error.

DISCUSSION

Two features of this simulation need emphasizing. First, samples of figures with realistic patterns of intercoordinate covariance were modeled. This is in contrast to a more simple model with a uniform, independent, circular distribution around each landmark (resulting in population correlations of zero among coordinates). Second, the shape of the simulated figures and the magnitude of correlation among coordinates were random, which makes the results as general as possible. It would be of interest, but beyond the scope of this paper, to explore the cases in which RWA and GRF fared poorly, although this might reveal patterns for predicting relative performance.

The simulation provided three important results. First, it indicated in which parts of the $LC \times PL \times VM$ space one can express confidence or caution about interpreting patterns of covariation in landmark data.

Second, it showed a strong influence of the number of landmarks, p , on the efficacy of RWA and GRF to reconstruct known covariance matrices (see Slice, 1993, for a similar LC effect in a different context). When figures were measured with only a few landmarks (6–10), both RWA and GRF performed rela-

tively poorly, although GRF performed well if there was a strong Pinocchio effect. The influence of the number of landmarks on RWA and GRF performance decreased with p ; at $> \sim 16$ –18 landmarks, performance increased very little with increasing p .

Third, the relative performance of RWA and GRF was found to be highly dependent on the location in the $PL \times VM \times VM$ parameter space. When the landmarks with excessive variance had total root variance $\leq 2\sigma_{min}$, RWA generally outperformed GRF regardless of PL. When $< \sim 75\%$ of the landmarks had excessive variance $\geq 4\sigma_{min}$, GRF generally outperformed RWA. Given that the breakdown value of the repeated medians regression is $\sim 50\%$ (Siegel, 1982), it is interesting that GRF outperforms a least-squares method when $> 50\%$ of the landmarks are given excessive variance. Note also that GRF does not perform best when excessive variance is limited to a single landmark but instead when excessive variance occurs at $\sim 25\%$ of the landmarks.

The response of first eigenvector correlations differed slightly from that for the matrix correlations. For the RWA estimates, the high first eigenvector correlations but low matrix correlations when both PL and VM were high may reflect the model's ability to reconstruct conspicuous patterns, resulting from the high Pinocchio effect, at the expense of more subtle patterns. The generally high value of the lower 5th percentile for the GRF estimates of the first eigenvector, even at small p and only moderate PL, indicate the ability of this method to reconstruct the major patterns of variation when variation among landmarks is not uniform.

The simulation results are useful for suggesting the appropriate morphometrics model to use, given the number of measured landmarks and some knowledge of the variation among coordinate variances. Although one can never know the true coordinate variances, a GRF should give reasonable approximations of the relative variation among the variances. After a GRF superimposition, if the coordinate standard deviations are largely uniform ($\sigma_{max} \leq 2\sigma_{min}$), shape variation should be analyzed by using partial warp and uniform scores. If $< \sim 75\%$ of the landmarks have large total root variances relative to the smallest total root variances ($\geq 4\sigma_{min}$), shape variation should be analyzed with the residuals from a GRF or some

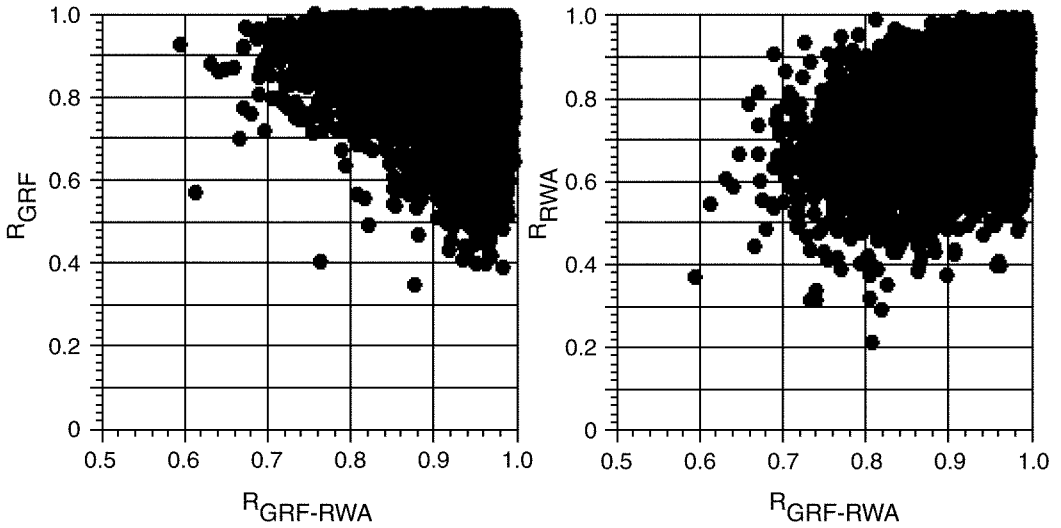


FIGURE 8. Performance of GRF or RWA as a function of $R_{GRF-RWA}$, the matrix correlation between GRF and RWA estimates of the known covariance matrix. This example is for $p = 14$ and is pooled across CL, PL, and VM.

other resistant-fit procedure (e.g., Dryden and Walker, 1999).

The only error in the results from this simulation were due to misestimation of the nuisance parameters. By randomly sampling from the reference data sets, however, an estimate of expected sampling error could be computed. Application of the bootstrap to estimate the confidence intervals of PCA, regression, or canonical variates analysis (CVA) coefficients of geometric morphometric shape variables (e.g., Walker, 1997) is not unlike the application of the bootstrap to phylogenetic trees. That is, the bootstrap method attempts to estimate only the error from sampling and not the error resulting from a violation of the model’s methodology (parsimony, least-squares, and so forth). The comparisons of $R_{5,obs}$, $R_{5,known}$, and the model error illustrated in Figure 7 show that model error overwhelms sampling error, especially when only a few landmarks are measured. The large, combined model-and-sampling error ($R_{5,known}$) when $< \sim 14$ landmarks were measured suggests that extreme caution should be used when interpreting patterns of shape variation, such as principal axes or multivariate regressions, when p is small.

Other features of the data may give hints on the efficacy of the RWA or GRF solution. For example, if RWA and GRF give similar results, then one might reasonably assume that the results of either are good estimates

of the actual pattern of shape variation. In other words, does a high matrix correlation between S_{GRF} and S_{RWA} , or between R_{RWA} and R_{GRF} , indicate a close estimate of either to \hat{S} ? The answer, in general, is no. Regardless of LC, GRF was more likely to result in worse estimates of \hat{S} as the correlation between S_{RWA} and S_{GRF} ($R_{RWA-GRF}$) increased (Fig. 8). In contrast, R_{RWA} increased with $R_{RWA-GRF}$ for all LC, but low values of R_{RWA} (< 0.8) were common even when $R_{RWA-GRF}$ approached 1.0 (Fig. 8). These results reflect that as S_{GRF} and S_{RWA} become more similar, there is less of a Pinocchio effect.

The results from these analyses argue against the blind use of RWA or of any other analysis of partial warp and standard uniform scores, including multivariate regression or the interpretation of canonical variates. Instead, biologists should first explore patterns of variation among the landmarks of GRF superimposed figures to evaluate whether an analysis of partial warp/ uniform scores or GRF coordinates is indicated.

ACKNOWLEDGMENTS

I thank F. James Rohlf and two anonymous reviewers for improving the manuscript.

REFERENCES

BOOKSTEIN, F. L. 1986. Size and shape spaces for landmark data in two dimensions (with discussion). *Stat. Sci.* 1:181–242.

- BOOKSTEIN, F. L. 1991. Morphometric tools for landmark data. Geometry and biology. Cambridge University Press, Cambridge.
- BOOKSTEIN, F. L. 1996a. Combining the tools of geometric morphometrics. Pages 131–151 in *Advances in morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. Naylor, and D. E. Slice, eds.). Plenum, New York.
- BOOKSTEIN, F. L. 1996b. Standard formula for the uniform shape component in landmark data. Pages 153–168 in *Advances in morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. Naylor, and D. E. Slice, eds.). Plenum, New York.
- DRYDEN, I. L., AND G. WALKER. 1999. Highly resistant regression and object matching. *Biometrics* 55:820–825.
- HEIBERGER, R. M. 1978. Generation of random orthogonal matrices. *Appl. Stat.* 27:199–206.
- MARCUS, L. F., E. BELLO, AND A. GARCIA-VALDECASAS (eds.). 1993. *Contributions to morphometrics*. Museo Nacional de Ciencias Naturales, Madrid.
- MARCUS, L. F., M. CORTI, A. LOY, G. NAYLOR, AND D. E. SLICE (eds.). 1996. *Advances in morphometrics*. Plenum, New York.
- ROHLF, F. J. 1996. Morphometric spaces, shape components and the effects of linear transformations. Pages 117–129 in *Advances in morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. Naylor, and D. E. Slice, eds.). Plenum, New York.
- ROHLF, F. J. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. *J. Class.* 16:197–223.
- ROHLF, F. J., AND F. L. BOOKSTEIN (eds.). 1990. *Proceedings of the Michigan Morphometrics Workshop*. University of Michigan Museum of Zoology, Ann Arbor, Michigan.
- ROHLF, F. J., AND D. SLICE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* 39:40–59.
- SIEGEL, A. F. 1982. Robust regression using repeated medians. *Biometrika* 69:242–244.
- SIEGEL, A. F., AND R. H. BENSON. 1982. A robust comparison of biological shapes. *Biometrics* 38:341–350.
- SLICE, D. E. 1993. *Extensions, comparisons, and applications of superimposition methods for morphometric analysis*. Ph.D. Thesis. State University of New York at Stony Brook.
- SLICE, D. E. 1996. Three-dimensional generalized resistant fitting and the comparison of least-squares and resistant fit residuals. Pages 179–199 in *Advances in morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. Naylor, and D. E. Slice, eds.). Plenum, New York.
- SNEATH, P. H. A. 1967. Trend-surface analysis of transformation grids. *J. Zool. London* 151:65–122.
- TANNER, M. A., AND R. A. THISTED. 1982. A remark on AS127. Generation of random orthogonal matrices. *Appl. Stat.* 31:190–192.
- WALKER, J. 1996. Principal components of shape variation within an endemic radiation of threespine stickleback. Pages 321–334 in *Advances in morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. Naylor, and D. E. Slice, eds.). Plenum, New York.
- WALKER, J. A. 1993. Ontogenetic allometry of threespine stickleback body form using landmark-based morphometrics. Pages 193–214 in *Contributions to morphometrics* (L. F. Marcus, E. Bello, and A. García-Valdecasas, eds.). Museo Nacional de Ciencias Naturales, Madrid.
- WALKER, J. A. 1997. Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (*Gasterosteidae*) body shape. *Biol. J. Linn. Soc.* 61:3–50.

Received 29 June 1999; accepted 6 October 1999

Associate Editor: G. Naylor