Syst. Zool., 36(4):356–367, 1987

A COMMENT ON SHEARING AS A METHOD FOR "SIZE CORRECTION"

F. JAMES ROHLF[1] AND FRED L. BOOKSTEIN[2]

[1]*Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794-5245; and*
[2]*Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan 48109-0406*

Abstract.—The "shear" method of Humphries et al. (1981) is based on a path model intended to explain differences in form by multiple factors: one for size and one or more for shape differences. Its adaptation for "removing" the effects of a within-population size-factor from between-group morphometric analyses is presented in compact matrix form, simplified, and compared to the method of orthogonal projection proposed by Burnaby (1966). While the size-correction methods give similar results for most real data sets, Burnaby's procedure with $k = 1$ (i.e., using a single composite size variable) is recommended for this purpose owing to its geometrical and computational simplicity. An example based on artificial data demonstrates that sheared principal components are not necessarily uncorrelated with size. Path modeling of size and shape together is a different purpose than size-correction, and is better served by a different procedure. [Allometric growth; morphometrics; size; shape; principal components analysis.]

A basic problem in the field of multivariate morphometrics is that of quantifying shape differences among forms separately from size differences. Unfortunately, the terms "size" and "shape" are subject to numerous definitions. In this paper they will be assumed to correspond to general factors (linear combinations of appropriate suites of variables, Bookstein et al., 1985)—rather than single, directly measurable variables such as total length or weight. Jolicoeur et al. (1984) suggest that when one variable, such as body weight, seems to describe size particularly well, other size-related variables can be adjusted by being replaced by their residuals about their regressions on the size variable. Of course, when the correlations among the size-related variables are very high (as is often the case), variables such as total length or weight may be very highly correlated with one's estimate for the general-size factor. But it is still important to make the conceptual distinction between an observed variable and a latent factor score. For instance, residuals from an observed size variable always share a factor representing the measurement error of the size variable ostensibly partialled out (Bookstein et al. 1985:114). These residuals are thus unsatisfactory in principle as shape variables—they share a factor of unique size variance. When one partials out an estimate of size based on averaging over several observed measurements, no matter how well correlated with size, this effect is reduced. General-size factors have large positive correlations with size-related variables in a study. Shape vectors, on the other hand, may have positive, negative, and zero correlations with diverse variables. Although there are many ways in which an estimate of a general-size factor could be defined, it is most commonly taken as the first eigenvector of a within-groups variance-covariance (or correlation) matrix based on log-transformed morphometric variables (Jolicoeur, 1963). The present paper is not concerned with methods for estimating size, but only with a class of adjustments that have been proposed to remove the effects of size variation upon morphometric description.

There are several reasons why one might wish to analyze shape differences among populations separately from any differences there may be in size. Size is labile ontogenetically, of course, but also phylogenetically; we could reasonably wish that its variability not affect our systematic judgments. Should a sample for one population contain more juveniles than another (perhaps because it was taken at a different time of year), then observed differences would not be the same as those found when one compares the populations at another time of year. Humphries et al. (1981) point out that ordinary linear discriminant functions or canonical variates analysis do not circumvent this problem. When samples differ with respect to both size and shape (however defined), discriminant functions and canonical axes must also reflect these differences since these methods construct optimal linear functions for detecting differences among populations based on the assumption that present samples are representative of future ones. While one hopes that one's samples reflect the major trends of covariation in the populations, the observed range of sizes can be expected to vary with the time of the year, nutritional state, etc. The effects of size can mask more subtle, and more biologically interesting, patterns of covariation among suites of variables.

Proper correction for size effects is also important for studies concerned with patterns of covariation among variables. In many morphometric studies the first eigenvector may account for 80% or more of the variation among a set of variables. One can, of course, design multivariate procedures (for instance, the path analysis below) that nevertheless accurately extract factors after the first; but if conventional component analyses or ordination techniques are to be executed instead, one does better at interpretation if size is "removed" beforehand. If the organisms were of the same size (or if the effect of size could be removed statistically) then these factors could now be seen to explain an appreciable proportion of the variance due to all factors other than size.

There has been little agreement as to how one should go about adjusting for the effects of size. Some workers propose simple ratios (division of each variable by the estimate of size, or difference of log-transformed variables). Others suggest regression using size as a covariate, and still others recommend more complex multivariate adjustments. In 1981 Humphries et al. proposed a hybrid computational method, "shearing," for multivariate discrimination by shape in the presence of size variation. In Bookstein et al. (1985) these same authors explained that their method was at its root a path model for the explanation of observed morphometric variables jointly by one size factor together with one or more factors for shape difference between groups after adjustment for size. In this approach, size is defined as the first factor of the observed pooled within-group covariance matrix (usually of log-transformed measurements), and group shape difference is analogously characterized by the first and subsequent factors of the same matrix after the effects of size have been held constant. The predicted value of each morphometric variable takes the form of a linear combination of factor scores constant within groups, added to an allometric contribution explaining a maximum of covariance within group (see Fig. 1).

This model appears not to have any sort of optimal estimate in convenient form whenever size does not exhaust the within-group covariance structure. The "shear" is an approximate solution in the case that group differences are orthogonal to the first few principal components of within-group variation after the first. In this approximation (Bookstein et al., 1985:sec. 4.2), a size factor having coefficients corresponding to the first within-group principal component is regressed out of several total principal components (eigenvectors computed from the total variance-covariance matrix) following the first. Coefficients of the resulting vectors are interpreted as path coefficients relating the observed morphometric variables to a shape difference factor or factors; and scores, when scattered against size, provide a useful decomposition of observed differences in the original space of the (pooled) principal components. The scatter of shape difference factor I against size approximates a shearing of the scatter of the first two principal components—hence the name of the technique. The scores on the shape factors are

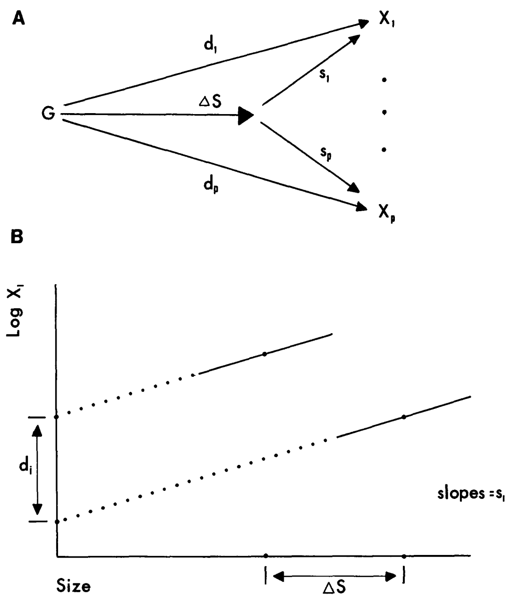

FIG. 1. The shear as a path model. (a) Path diagram, after Bookstein et al. (1985), in the context of two groups only. Observed variables, $X_1, \ldots, X_p$ are modeled as dependent variables regressed upon a group shape difference factor G and a general size factor S. (b) The equivalent analysis of covariance, showing the origin of the path coefficients as regression coefficients of observed data upon factors. ΔS, mean size difference between the groups; s_i, allometric coefficient for the regression of X_i upon S; d_i, regression of X_i upon group controlling for size—one coefficient of the shape difference factor.

not intended to be studied in isolation, but only as augmented by size information and other factors from other analyses.

In fairness to readers who may be puzzled, we should point out that this description does not resemble that of Humphries et al. (1981), where the technique was first introduced—but it describes the same computation. Possibly because that initial rationale was incomplete, the technique does not seem to have been used in very many published studies other than those by members of the University of Michigan Morphometrics Group (who have been very prolific, e.g., Barbour and Chernoff, 1984; Humphries, 1984; Smith and Todd, 1984; Strauss, 1985; Strauss and Fuiman, 1985). In any case, the original published description of the algorithm was not very explicit. With the publication of Bookstein et al. (1985) and their inclusion of a SAS procedure developed by L. Marcus and D. Swofford the shearing method has been made available for wider use. It is also present now in some statistics packages. For this reason, it is important that the claims and properties of this method be reviewed and its advantages and disadvantages be fully understood.

SHEARING

The method of shearing may be considered to incorporate a method of size-correction in the steps leading to generation of the shape-difference factor or factors. The originators had not intended that these steps be considered in isolation, yet they did not caution against this use. In any case, when these steps are considered by themselves, they differ substantially from other approaches to "removing" size, mainly because the resulting factors were intended to be presented only in the context of the path model of joint explanation with size restored (Fig. 1). In the remainder of this paper, we will restate the relevant steps of the "shearing" algorithm, indicate how it relates to other methods of size-correction, and show how its differences follow from its having been constructed for a different purpose.

The Shear Algorithm

A recapitulation of Humphries et al.'s (1981) method is presented here so that it may be compared easily with the present more mathematical version of their algorithm. It is based in part on the SAS procedure given in Appendix A.5 of Bookstein et al. (1985), but the notation was changed to cast the algorithm into a more conventional mathematical style.

1. Combine into a single matrix the measurements on p variables for samples from g groups, each of size n_i. Transform the variables to logarithms and let **X** be the resultant n by p data matrix, where $n = \Sigma n_i$ specimens and p is the number of variables.
2. Compute the total variance-covariance matrix, **T**. This matrix is based on de-

viations of observations from the grand mean vector and corresponds to the total mean-square matrix in multivariate analysis of variance (MANOVA).

3. Perform a principal components analysis of the total variation in the study, i.e., compute the eigenvector matrix, **E**, of **T**. The eigenvectors correspond to the columns of **E** and are assumed to be normalized ($\mathbf{E}^t\mathbf{E} = \mathbf{I}$). Let $\mathbf{E}_1$ denote the first column of this matrix, $\mathbf{E}_j$ the jth, and $\mathbf{E}_{1j}$ a p by 2 matrix with $\mathbf{E}_1$ as column 1 and $\mathbf{E}_j$ as column 2.
4. Construct a group-centered n by p data matrix, $\mathbf{X}_C$. In this matrix the mean for each variable within each of the g samples is zero.
5. Compute the pooled within-group variance-covariance matrix, **W**, based on deviations of observations from the mean vector of their group. Note that in computing this matrix both Humphries et al. (1981) and Bookstein et al. (1985) use $n - 1$ as their divisor, instead of $n - g$, the appropriate degrees of freedom. **W** is the within-group or error mean square matrix of MANOVA.
6. Compute the first eigenvector of **W**. This vector, $\mathbf{F}_1$, is assumed to represent the intra-group size vector.
7. Project the group-centered specimens onto the size vector:

$$\mathbf{Q}_1 = \mathbf{X}_C\mathbf{F}_1, \tag{1}$$

where $\mathbf{Q}_1$ is a column vector of n elements, representing the estimated sizes of the group-centered specimens.

8. Then, to shear a variable (for example $\mathbf{E}_j$, the jth eigenvector of **E**), perform the following steps.
   (a) Project the group-centered data onto $\mathbf{E}_1$ and $\mathbf{E}_j$:

$$\mathbf{P}'_1 = \mathbf{X}_C\mathbf{E}_1$$
$$\mathbf{P}'_j = \mathbf{X}_C\mathbf{E}_j. \tag{2}$$

   (b) Regress $\mathbf{P}'_j$ on $\mathbf{Q}_1$. The slope, α_j, is

$$\alpha_j = (\mathbf{Q}_1^t\mathbf{P}'_j)/(\mathbf{Q}_1^t\mathbf{Q}_1). \tag{3}$$

   The intercept is zero since the data have been centered.

   (c) Construct an estimate of $\mathbf{Q}_1$ that lies in the plane of $\mathbf{P}'_1$ and $\mathbf{P}'_j$. This can be found using multiple regression. The partial regression coefficient vector is

$$\mathbf{B}_j = (\mathbf{P}'^t_{1j}\mathbf{P}'_{1j})^{-1}\mathbf{P}'^t_{1j}\mathbf{Q}_1 \tag{4}$$

   where $\mathbf{P}'_{1j}$ is a matrix with $\mathbf{P}'_1$ as column 1 and $\mathbf{P}'_j$ as column 2. As before, the intercept is zero.

   (d) Estimate $\mathbf{H}_j$, the size-adjusted $\mathbf{E}_j$, as

$$\mathbf{H}_j = \mathbf{E}_j - \alpha_j\mathbf{E}_{1j}\boldsymbol{B}_j. \tag{5}$$

   (e) Compute the scores for the n specimens on this shape axis, using the raw, *uncentered* data

$$\mathbf{S}_j = \mathbf{X}\mathbf{H}_j. \tag{6}$$

   These scores (projections) may be plotted against the usual projections of the n points onto the first principal component axis

$$\mathbf{P}_1 = \mathbf{X}\mathbf{E}_1. \tag{7}$$

   The result is a *shear* since $\mathbf{E}_1$ and $\mathbf{H}_j$ (or $\mathbf{P}_1$ and $\mathbf{S}_j$) are not the result of a rigid rotation of the original variables.

Notes on the Shearing Algorithm

To clarify the shear method, it is given below using a more compact mathematical expression rather than as a series of algorithmic steps. It is convenient to define an n by n matrix **N** whose elements are all equal to $1/n$. Then the data matrix centered about the grand mean is just $(\mathbf{I} - \mathbf{N})\mathbf{X}$ and the total variance-covariance matrix is

$$\mathbf{T} = \mathbf{X}^t(\mathbf{I} - \mathbf{N})\mathbf{X}/(n - 1). \tag{8}$$

Similarly, one can define an n by n block diagonal matrix **M** with the ith block being an n_i by n_i matrix with all elements equal to $1/n_i$ where n_i is the number of specimens in the ith group. The group-centered data matrix is then $(\mathbf{I} - \mathbf{M})(\mathbf{X})$, and the within-group variance-covariance matrix is

$$\mathbf{W} = \mathbf{X}^t(\mathbf{I} - \mathbf{M})\mathbf{X}/(n - g). \tag{9}$$

This matrix can be expressed in terms of its eigenvectors, **F**, and eigenvalues, Λ:

$$\mathbf{W} = \mathbf{F}\Lambda\mathbf{F}^t. \tag{10}$$

It can easily be shown that

$$\mathbf{Q}_1 = (\mathbf{I} - \mathbf{M})\mathbf{X}\mathbf{F}_1, \tag{11}$$

$$\mathbf{P}'_j = (\mathbf{I} - \mathbf{M})\mathbf{X}\mathbf{E}_j, \tag{12}$$

and

$$\mathbf{W}\mathbf{F}_1 = \lambda_1\mathbf{F}_1. \tag{13}$$

These relationships can be used to show that

$$\alpha_j = \mathbf{F}_1^t\mathbf{E}_j \tag{14}$$

and

$$\mathbf{B}_j = (\mathbf{E}_{1j}^t\mathbf{F}\Lambda\mathbf{F}^t\mathbf{E}_{1j})^{-1}\mathbf{E}_{1j}^t\mathbf{F}_1\lambda_1. \tag{15}$$

Both α_j and $\mathbf{B}_j$ are functions of the cosines of the angles between the within and total eigenvectors.

Let us consider the case in which there are only $p = 2$ variables and we wish to adjust the variable $\mathbf{E}_2$ (i.e., $j = 2$). The calculation of α_j and $\mathbf{B}_j$ as given above can be greatly simplified and combined into one step:

$$\mathbf{H}_2 = \mathbf{E}_2 - (\mathbf{F}_1^t\mathbf{E}_2)\mathbf{F}_1. \tag{16}$$

This is in the form of the standard Gram-Schmidt "sweep" operation of matrix algebra which removes that portion of the variation of $\mathbf{E}_2$ that can be predicted by $\mathbf{F}_1$. The scale factor, $\mathbf{F}_1^t\mathbf{E}_2$, is the fraction of $\mathbf{F}_1$ to be subtracted from $\mathbf{E}_2$. It is also the cosine of the angle between the first within-group eigenvector and the second total eigenvector (since the eigenvectors are normalized). The resultant vector, $\mathbf{H}_2$, is collinear with $\mathbf{F}_2$, the second eigenvector of the pooled within-group variance-covariance matrix. This is not surprising since, in 2-dimensional space, any vector orthogonal to $\mathbf{F}_1$ must be parallel to $\mathbf{F}_2$. Thus, for the case of only 2 variables, one does not have to be concerned with the various regression steps described above—one can simply compute $\mathbf{F}_2$ and interpret it for what it is, the second within-group eigenvector.

For $p > 2$ variables, the computations are more complex (largely due to the step involving multiple regression) but they can still be expressed in a form analogous to a sweep operation. However, $\mathbf{F}_1$ is no longer being swept from $\mathbf{E}_j$. Thus

$$\mathbf{H}_j = \mathbf{E}_j - \alpha_j\mathbf{F}'_{1(j)}, \tag{17}$$

where

$$\begin{aligned}\mathbf{F}'_{1(j)} &= \mathbf{E}_{1j}\mathbf{B}_j \\ &= \mathbf{E}_{1j}(\mathbf{E}_{1j}^t\mathbf{F}\Lambda\mathbf{F}^t\mathbf{E}_{1j})^{-1}\mathbf{E}_{1j}^t\mathbf{F}_1\lambda_1.\end{aligned} \tag{18}$$

The constant, α_j, is the length of the projection of the jth total eigenvector, $\mathbf{E}_j$, onto the size vector, $\mathbf{F}_1$. The "(j)" in the subscript is to indicate that $\mathbf{F}'_1$ is, in part, a function of j. It is not quite consistent to sweep out $\mathbf{F}'_1$ instead of $\mathbf{F}_1$. We will return to this matter in the Discussion.

Another Formulation of the Shear Method

The procedure described above can be interpreted as essentially constructing a vector in the $\mathbf{E}_1$, $\mathbf{E}_j$ plane that is orthogonal to the projection of $\mathbf{F}_1$ onto this plane.

The least squares projection of a size vector, $\mathbf{F}_1$, onto the $\mathbf{E}_1$, $\mathbf{E}_j$ plane (based on the group-centered data) is simply

$$\mathbf{F}''_{1(j)} = (\mathbf{F}_1^t\mathbf{E}_1)\mathbf{E}_1 + (\mathbf{F}_1^t\mathbf{E}_j)\mathbf{E}_j. \tag{19}$$

The "(j)" in the subscript is to indicate that $\mathbf{F}''_1$ is dependent on which vector, j, is being held constant. The adjusted $\mathbf{E}_j$ vector is proportional to

$$\mathbf{H}'_j = (\mathbf{F}_1^t\mathbf{E}_j)\mathbf{E}_1 - (\mathbf{F}_1^t\mathbf{E}_1)\mathbf{E}_j. \tag{20}$$

Figure 2 shows these relations diagrammatically. SAS proc matrix operations that accomplish the above steps are furnished in the appendix. While similar in intent, this method is *not* mathematically equivalent to the method proposed by Humphries et al. (1981). As pointed out below, neither method is recommended if one's purpose is size adjustment alone.

BURNABY'S (1966) METHOD

Another procedure is to use Burnaby's (1966) simple method of sweeping the effect of one or more extraneous variables from the data and then carrying out principal components analysis, discriminant function analysis, etc. on the adjusted data matrix. The resulting axes, clusters, etc. are then based on variation that is orthogonal to the vectors corresponding to the variables being held constant. For example, Burnaby suggested, one could hold constant the first eigenvector of the variance-covariance matrix from each of the g populations. In his example he held two "nuisance factors" constant—$\mathbf{F}_1$ and a phenotypic response to an ecological gradient. We will consider the case in which one holds constant a single vector, $\mathbf{F}_1$.

Several different (but mathematically equivalent) algorithms can be used. Burnaby (1966) suggested the following method, which is quite compact. If $\mathbf{F}_1$ is the p by 1 size-vector that one wishes to correct for, then the matrix

$$\mathbf{L} = \mathbf{I}_p - \mathbf{F}_1(\mathbf{F}_1^t\mathbf{F}_1)^{-1}\mathbf{F}_1^t \qquad (21)$$

projects any vector onto the subspace complementary to $\mathbf{F}_1$. $\mathbf{L}$ is an idempotent symmetric matrix of order p by p and $\mathbf{I}_p$ is the p by p identity matrix. Then the jth shape vector is

$$\mathbf{H}_j'' = \mathbf{L}\mathbf{E}_j. \qquad (22)$$

An adjusted data matrix, $\mathbf{X}'$, can be computed as

$$\mathbf{X}' = \mathbf{X}\mathbf{L}. \qquad (23)$$

SAS proc matrix operations that accomplish the above steps are furnished in the appendix.

The steps are mathematically equivalent to projecting the n specimens onto the within-group eigenvectors (all *min* $(p, n - g)$ of them), replacing the values for the projections onto the first axis with zeros, and then rotating the n specimens back into the original space. They will now have different values since the effects of differences in within-group size have been *completely* removed and the data points now all lie on a hyperplane within the original space. The $\mathbf{X}'$ matrix may then be used in place of $\mathbf{X}$ for the computation of size-adjusted principal components analysis, cluster analysis, discriminant analysis, canonical variates analysis, etc. Rao (1966) and Gower (1976) present various extensions and generalizations of Burnaby's (1966) approach. Reyment and Banfield (1976) furnish an example.

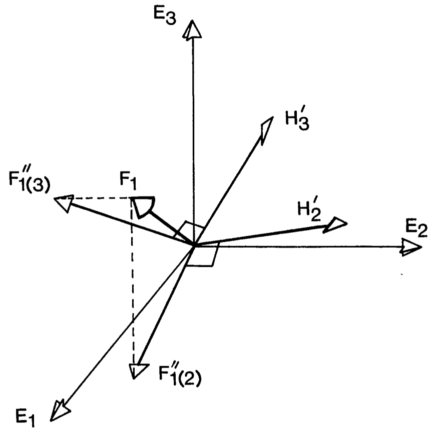

FIG. 2. Geometrical relationships based on the alternative formulation of the shear procedure. Coordinate axes have been rotated to correspond to the total eigenvectors. The first within eigenvector is projected onto the 1,2 and 1,3 planes, and the shape vectors $\mathbf{H}_2'$ and $\mathbf{H}_3'$, respectively, are in these planes and orthogonal to the projections.

If one wished to retain some size information in the analysis (as in the shear method), then one could append the column vector $\mathbf{X}\mathbf{E}_1$ to the $\mathbf{X}'$ matrix. Alternatively, one could use $\mathbf{X}\mathbf{F}_1$ (using within-group size), which seems a bit more logical. In the latter case, however, distances among the points would then be *identical* to what they were originally and thus a cluster analysis or discriminant analysis of the "adjusted" data would yield results identical to those obtained using the original data. Thus no correction to the interpoint distances will have been accomplished. Since $\mathbf{E}_1$ and $\mathbf{F}_1$ are usually highly correlated, the use of $\mathbf{E}_1$ will usually have a similar result. Of course, the coefficients themselves will have been changed. These were the major interest in Bookstein et al. (1985), not size-correction itself. Their simulations showed that their method was able to recover the original coefficients in their path model. It would be interesting to extend their simulations to see how well the use of Burnaby's method would enable the coefficients to be estimated.

The idea behind the approach of Somers (1986) is closely related. He proposed that one compute the eigenvectors of a corre-

TABLE 1. Artificial data used to illustrate differences among the various methods for size adjustment. Observations were drawn from two log-normally distributed populations.

Group	Variables 1	2	3	4
1	109.8	48.7	128.1	17.7
	14.9	31.7	54.7	8.6
	23.7	25.4	201.5	20.0
	14.1	35.3	213.9	34.0
	41.2	35.8	132.0	31.3
	18.8	20.3	67.2	8.0
	47.6	16.6	234.8	29.7
	21.3	4.5	106.6	25.6
	60.0	20.7	174.0	38.9
	33.9	18.8	98.5	11.4
2	18.4	8.5	3.1	15.3
	16.8	9.0	2.7	12.6
	45.7	17.3	3.2	27.6
	32.0	7.5	3.3	11.5
	75.1	11.5	2.0	16.1
	18.9	7.9	1.5	13.6
	49.5	68.7	4.6	26.2
	15.8	26.1	1.7	9.3
	7.9	22.3	2.4	17.0
	11.1	9.9	3.2	30.0

lation matrix from which the effects of an isometric size vector, $\mathbf{V}_1$, has been removed. In his formulation, the reduced correlation matrix is

$$\mathbf{R}_1 = \mathbf{R} - \frac{1}{p}\mathbf{V}_1^t\mathbf{R}\mathbf{V}_1 \tag{24}$$

where

$$\mathbf{V}_1^t = (p^{-0.5}, p^{-0.5}, \ldots, p^{-0.5}). \tag{25}$$

But this reduces simply to

$$\mathbf{R}_1 = \mathbf{R} - \bar{R}. \tag{26}$$

A factorization of $\mathbf{R}_1$ will yield vectors orthogonal to $\mathbf{V}_1$ only if the average correlation in each row (and column) of $\mathbf{R}$ is the same. This residual matrix has some unfortunate properties—"correlations" in the residual matrix may fall outside of the range $-1 \leq r \leq 1$ and, as Somers (1986) points out, some of the resulting eigenvalues may be negative. One could use $\mathbf{V}_1$ rather than $\mathbf{F}_1$ in Burnaby's (1966) method. Then the eigenvectors of

$$\mathbf{R}_1' = \mathbf{LRL} \tag{27}$$

would be orthogonal to $\mathbf{V}_1$ as Somers (1986) intended and correlations computed from the residual matrix, $\mathbf{R}_1'$, will fall within the expected range.

EXAMPLE

The bivariate test data given in Bookstein et al. (1985:105) cannot be used to illustrate the differences among the procedures described above since they give equivalent results when there are only 2 variables.

Table 1 furnishes a simple artificial dataset. There are $n_i = 10$ specimens in $g = 2$ groups for which $p = 4$ variables have been recorded. The data are intended only to illustrate some of the numerical properties of the methods. They are not intended to be entirely realistic (whatever that means). This dataset was constructed so as to emphasize the differences among the methods for size-correction. The results given below are all based on log-transformed data. Table 2 gives $\mathbf{E}$, the eigenvectors of the total variance-covariance matrix, and $\mathbf{F}_1$, the first eigenvector of the within-group variance-covariance matrix. The percentages of the total variance explained by the total eigenvectors are: 80.0%, 9.5%, 6.8%, and 3.7%. The variances explained by the within-group eigenvectors are: 45.9%,

TABLE 2. Normalized eigenvectors, $\mathbf{E}$ and $\mathbf{F}_1$, of variance-covariance matrices based on the log transforms of the data from Table 1.

Variables	Total eigenvectors 1	2	3	4	Within eigenvector
1	0.10487	0.73147	0.61281	−0.28005	0.69483
2	0.13366	0.64085	−0.75183	0.07871	0.56532
3	0.98204	−0.17791	0.01659	−0.06063	0.30372
4	0.08210	0.15041	0.24277	0.95483	0.32461

TABLE 3. Sheared and Burnaby adjusted axes premultiplied by the transpose of the matrix of normalized total eigenvectors, E, in Table 2.

Total eigen-vectors	Sheared axes		Burnaby adjusted axes			
	2	3	1	2	3	4
1	−0.51911	−0.10103	0.77595	−0.40959	−0.04005	−0.06695
2	0.28997	0.00000	−0.40959	0.25122	−0.07322	−0.12239
3	0.00000	1.00078	−0.04005	−0.07322	0.99284	−0.01197
4	−0.00000	0.00000	−0.06695	−0.12239	−0.01197	0.97999

28.8%, 21.8%, and 3.5%. While there is a within-group size component, it is not as strong as it often is in real morphometric data.

Tables 3 and 4 give Humphries et al.'s sheared axes, Burnaby's adjusted axes, and the alternative sheared axes—all premultiplied by the transpose of the matrix **E** of normalized eigenvectors. This rotation by **E** yields coordinates of the vectors relative to the total eigenvectors used as coordinate axes. This makes it easier to visualize the geometric relationships among the vectors. One can see that $\mathbf{H}_2$ is orthogonal to $\mathbf{E}_3$ and $\mathbf{E}_4$ as is $\mathbf{H}'_2$. Burnaby's axes are not (nor were they intended to be).

Table 5 gives the correlations among the $\mathbf{H}_j$, $\mathbf{H}'_j$, and $\mathbf{H}''_j$ for $j = 2$ and 3 (4 is not included to save space). One can see that the correlations among the alternative shape axes are quite high. On real datasets they are apt to be even higher. Thus the decision about which method should be used is not likely to be resolved by empirical studies—one should decide in principle what one means by size adjustment. Note also that the pairs of adjusted eigenvectors are no longer orthogonal and thus eigenvectors would have to be computed again if one wished to have a set of orthogonal axes. The correlations between size, $\mathbf{F}_1$, and the shape vectors $\mathbf{H}_2$ and $\mathbf{H}_3$ are 0.00875 and 0.03665 (computed as cosines of the angles between vectors). While these correlations are quite small, they demonstrate a lack of independence between size and the shape vectors (if one were to try harder, other examples could no doubt be found with larger correlations). On the other hand, the correlations of size with Burnaby's adjusted axes and with the alternative shared axes are zero.

DISCUSSION

As we noted in the discussion of step 8(c) of the original algorithm, the method of shearing sweeps out a projection, $\mathbf{F}'_{1(j)}$, of size rather than size, $\mathbf{F}_1$, itself. Neither Humphries et al. (1981) nor Bookstein et al. (1985) explain this recommendation.

In another perplexity, the originators recommend that the sheared principal component axes be obtained by shearing each one (after the first) separately rather than simultaneously by using a method based on a multiple regression analysis using all the factors at once. For example, when shearing $\mathbf{E}_2$ one projects into the $\mathbf{E}_1$, $\mathbf{E}_2$ plane but when shearing $\mathbf{E}_3$ one projects into the $\mathbf{E}_1$, $\mathbf{E}_3$ plane. Thus the various sheared axes are, in a sense, not really in the same subspace. Neither of these maneuvers seems reasonable in the context of size-correction; rather, it flows from the original purpose of shearing, which is the estimation of a more complex path model (Fig. 3) explaining all covariances among a vector of size measures by a combination of multiple shape difference factors together with size. This model, as we said, appears difficult to estimate in any convenient fashion. This model could be estimated rather formally by a least-squares

TABLE 4. Sheared axes (based on the reformulation presented in the present paper) premultiplied by the transpose of the matrix of normalized total eigenvectors, E, in Table 2.

Total eigen-vec-tors	Reformulated sheared axes			Size
	2	3	4	
1	−0.42103	−0.17323	−0.27432	0.47334
2	0.23031	0.00000	0.00000	0.86532
3	0.00000	0.96903	0.00000	0.08462
4	−0.00000	0.00000	0.91803	0.14144

TABLE 5. Matrix of correlations among size-adjusted axes from Tables 3 and 4. "Ref." corresponds to the reformulation of the sheared axes as presented in this paper.

	Sheared axes		Ref. sheared axes		Burnaby axes		"Size"
	H_2	H_3	H'_2	H'_3	H''_2	H''_3	F_1
$\mathbf{H}_2$	1.00000	0.08769	0.99996	0.15363	0.95787	−0.00074	0.00875
$\mathbf{H}_3$	0.08769	1.00000	0.08812	0.99709	−0.06327	0.99541	0.03665
$\mathbf{H}'_2$	0.99996	0.08812	1.00000	0.15439	0.95748	0.00000	0.00000
$\mathbf{H}'_3$	0.15363	0.997090	0.15439	1.00000	0.00000	0.98794	0.00000
$\mathbf{H}''_2$	0.95787	−0.06327	0.95748	0.00000	1.00000	−0.14661	−0.00000
$\mathbf{H}''_3$	−0.00074	0.99541	0.00000	0.98794	−0.14661	1.00000	0.00000
$\mathbf{F}_1$	0.00875	0.03665	0.00000	0.00000	−0.00000	0.00000	1.00000

fit of a suitably constrained covariance structure to an augmented data matrix (group membership dummy variables appended to the list of morphometric variables). This would extend the least-squares spirit of Sewall Wright's original path-analytic modeling. The shear method is a simplification of such a procedure that assumes each between-group PC_j after the first is a surrogate for the $(j-1)$st factor $\mathbf{H}_{j-1}$ of group shape difference independent of all the other $\mathbf{H}_k$s except for their joint confounding by size. (This is clearly untenable when extended to all the subsequent PCs, but it is realistic for the first few, as long as the PCs fail to tap "true" factor-based within-group covariation.) Separately for each subscript $j > 1$, the shear takes the space of variation spanned by PC_1 and PC_j to approximate the space spanned by the size factor S and the shape difference factor $\mathbf{H}_{j-1}$. There is no equivalent to this purpose in the context of "size correction" and it is unnecessary outside the context of path-modeling.

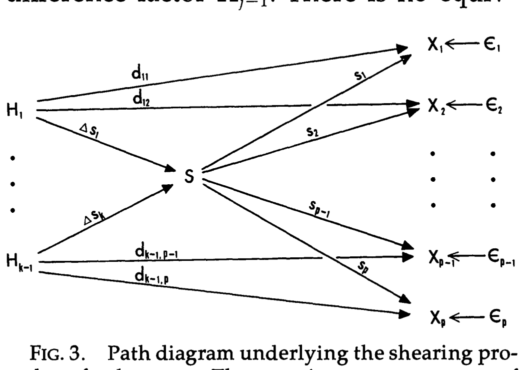

FIG. 3. Path diagram underlying the shearing procedure for k groups. The covariances among a set of observed (or log-transformed) variables are explained by variation in between-group factors together with a within-group size factor differing between groups in both mean and variance. There is no apparent analytical solution for this method.

In other words, the shear method presumes that within-group correlations of the PCs merely express covariances "equal and opposite" to those induced by mean size as it varies separately from group to group in shape difference space. The appropriate regression estimate of any single observed variable X_i—the equation yielding the path coefficients in Figure 3—follows from the path diagram. It is the summation of many independent ancovas, each like that in Figure 1b. The logic of multiple regression does not appear here at all—the correction for size of each PC_j after the first is a matter of PC_1 and PC_j alone. The sweeping of within-group Size out of the successive principal components after the first may be computed in one single operation as long as the residuals are treated as distinct path coefficients apposite to distinct shape difference factors regardless of their observable covariances.

As a further consequence of the difference in purpose between path-modeling and size-correction, while Humphries et al. (1981) state that $\mathbf{H}_j$ is uncorrelated with intra-group size $(\mathbf{F}_1)$, it need not be true for all datasets since the effects of intragroup size, $\mathbf{F}_1$, are not being removed from $\mathbf{E}_j$. Only a function of it, $\mathbf{F}'_{1(j)}$, is being removed. As demonstrated in the examples section, the correlation between $\mathbf{H}_j$ and $\mathbf{F}_1$ need not be zero. The correlation will be zero only if $\mathbf{F}_1$ and $\mathbf{E}_j$ happen to be orthogonal. In practice, the correlation will often be rather small since $\mathbf{E}_1$ and $\mathbf{F}_1$ are often similar (and thus $\mathbf{F}_1$ should be nearly orthogonal to $\mathbf{E}_j$ because $\mathbf{E}_1$ is orthogonal to $\mathbf{E}_j$). The magnitude of the deviation from orthogonality clearly depends upon the datasets being analyzed. Bookstein et al. (1985:104) point out that the sheared $\mathbf{H}_{j-1}$ is an approximation to a size-orthogonal factor only when $\mathbf{F}_1$ lies in the plane of $\mathbf{E}_1$ and $\mathbf{E}_j$. Their condition for orthogonality does not quite agree with the one given here. Bookstein et al. (1985) suggest that the researcher plot $\mathbf{S}_j$ against $\mathbf{P}_1$ to verify that within-group size has been removed from the vector $\mathbf{E}_j$. But it would be better if $\mathbf{P}_1$ were replaced by $\mathbf{Q}_1$ for this purpose since $\mathbf{P}_1$ is a function of the total not the within-group variance-covariance matrix.

As shown above, the eigenvectors are no longer orthogonal after they have been adjusted. Thus one must then analyze the adjusted data using principal components analysis, cluster analysis, factor analysis, or canonical variates analysis (see Gower, 1976), etc.

Humphries et al. (1981) rule out Burnaby's technique because its generalization, adjusting for distinct (non-parallel) size vectors in the different groups, is likely to make corrections that are undesirable. For example, since two non-parallel size vectors define a plane, Burnaby's generalization would consider all variation in this plane to be due to size and would project all points onto the subspace orthogonal to this plane. Indeed, in the case of $p = 2$ variables, all of the variation would be considered owing to size differences. None of the techniques under discussion here include explicit procedures for dealing with non-parallel size vectors in different groups. In the special case of nonlinear allometry (correlation of mean size with allometric slopes), the path model underlying the shear method seems most easily extended by the use of a nonlinear regression on size.

Bookstein et al. (1985:103) make an additional objection to Burnaby's method. They state that Burnaby's method "has coefficients that are not loadings and that can therefore not be compared among themselves." What was meant was that there is no path model to suggest that the adjusted axes should correspond to biologically meaningful underlying factors—they are simply adjusted axes. This objection, then, is just a restatement of the difference in purpose between path modeling and size-correction. One technique emphasizes the coefficients of factors of form approximately constant within groups, while the other emphasizes the ordination of taxa (as by distances between them) somewhat independently of the space in which this ordination is achieved.

Subsequent multivariate analyses may require that the adjusted data conform to the usual assumptions of multivariate normality and homogeneity of variance-covariance matrices in order for conventional significance tests to be valid. In particular, Burnaby (1966) warns that the vector $\mathbf{F}_1$ should be estimated independently of the rest of the data. This implies, for example, that the first eigenvector of the within-groups variance-covariance matrix cannot be used if conventional significance tests are to be used. Neither he nor subsequent workers indicate how serious this problem is or offer alternative procedures. This is unfortunate since $\mathbf{F}_1$ is very commonly used to estimate the general size factor. One obvious solution would be to estimate size by using a small set of size-related variables that are not to be included in the subsequent analyses. Variables which are known (from previous work) to be very high correlated with size could be used with little loss of information in the main study since they would have very little variance after the effects of size had been removed. Another possibility is to use an isometric size vector defined a priori (e.g., Somers, 1986).

In practice, for many datasets the various methods described above will give very similar results since the correlations among $\mathbf{E}_1$, $\mathbf{F}_1$, and size-related characters are often very high. Of the methods considered in this paper, only the method due to Burnaby (1966) simply and directly constructs variables orthogonal to a variable that is considered to represent size and would therefore seem to be the method of choice for this purpose. Other purposes, such as the explanation of covariances among size variables, require other methods.

It should be emphasized that the analysis given above assumes that it makes sense to define size as the first eigenvector of the within-group variance-covariance matrix for log-transformed variables. Using this eigenvector naturally leads one to the use of vectors orthogonal to this vector being considered shape vectors. This may not be the best strategy for actual studies (especially if the sample of organisms does not differ very much in size). The first within-group eigenvector need not have anything to do with size. It simply gives the direction in which there is maximal variation within the populations. This is usually strongly related to size; but if similar-sized individuals are used then the first eigenvector could represent some other factor. For such cases, Mosimann and James' (1979) procedure—first producing geometrically meaningful definitions of shape (proportionality) and size and then studying their relationships to each other and to time—can lead to more appropriate interpretations. Notions of growth allometry that are embodied in coefficients of size by the eigenvector approach are encoded instead in size-shape correlations by the geometric approach. The biological context of these two interpretations is, of course, the same. Bookstein et al. (1985:appendix A4) also present methods that have this property whenever landmark location data are available (rather than mere measured distances, the starting point for the more conventional multivariate approaches to morphometrics discussed in this paper). See also Bookstein (1986).

ACKNOWLEDGMENTS

We would like to thank Scott Ferson, John Gower, Steve Hartman, Pierre Jolicoeur, Les Marcus, Jim Mosimann, and Richard Reyment for critical readings of the manuscript. The manuscript was begun while FJR was a visitor of the Statistics Department of Rothamsted Experimental Station. Their generosity in the use of the Rothamsted facilities is very much appreciated. FJR would also like to thank Dr. M. Bell, J. Baumgartner, Dr. A. Kence, and other members of the morphometrics seminar group at Stony Brook for their comments, copies of data sets, and general encouragement. Karen Rohlf prepared the illustrations.

This work was supported, in part, by the National Science Foundation under grants DEB 8202269 and BSR 8306004 to FJR and by the National Institutes of Health under grant GM-37251 to FLB. The computations and preparation of the manuscript were aided by the use of an IBM 4361 computing system made available in part by a grant (BSR 8318792) from the National Science Foundation. This is contribution number 643 from the Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook.

REFERENCES

BARBOUR, C. D., AND B. CHERNOFF. 1984. Comparative morphology and morphometrics of the pescados blancos (genus *Chirostoma*) from lake Chapala, Mexico. Pages 111–127 *in* Evolution of fish species flocks (A. Echelle and I. Kornfield, eds.). Univ. Maine at Orono Press, Orono.

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

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

BURNABY, T. P. 1966. Growth-invariant discriminant functions and generalized distances. Biometrics, 22: 96–110.

GOWER, J. C. 1976. Growth-free canonical variates and generalized inverses. Bull. Geol. Instit. Univ. Uppsala, N.S., 7:1–10.

HUMPHRIES, J. M. 1984. *Cyprinodon verecundus*, n. sp., a fifth species of pupfish from Laguna Chichancanab. Copeia, 1984:58–68.

HUMPHRIES, J. M., F. L. BOOKSTEIN, B. CHERNOFF, G. R. SMITH, R. L. ELDER, AND S. G. POSS. 1981. Multivariate discrimination by shape in relation to size. Systematic Zool., 30:291–308.

JOLICOEUR, P. 1963. The degree of generality of robustness in *Martes americana*. Growth, 27:1–27.

JOLICOEUR, P., P. PIRLOT, G. BARON, AND H. STEPHAN. 1984. Brain structure and correlation patterns in Insectivora, Chiroptera, and primates. Systematic Zool., 33:14–29.

MOSIMANN, J. E., AND F. C. JAMES. 1979. New statistical methods for allometry with application to Florida red-winged blackbirds. Evolution, 33:444–459.

RAO, C. R. 1966. Discriminant function between composite hypotheses and related problems. Biometrika, 53:339–345.

REYMENT, R. A., AND C. F. BANFIELD. 1976. Growth-free canonical variates applied to fossil foraminifers. Bull. Geol. Instit. Univ. Uppsala, N.S., 7:11–21.

SMITH, G. R., AND T. N. TODD. 1984. Evolution of species flocks of fishes in north temperate lakes. Pages 45–68 *in* Evolution of fish species flocks (A.

Echelle and I. Kornfield, eds.). Univ. Maine at Orono Press, Orono.

SOMERS, K. M. 1986. Multivariate allometry and removal of size with principal components analysis. Systematic Zool., 35:359–368.

STRAUSS, R. E. 1985. Evolutionary allometry and variation in body form in the South American catfish genus *Corydoras* (Callichthyidae). Systematic Zool., 34:381–396.

STRAUSS, R. E., AND L. A. FUIMAN. 1985. Quantitative comparisons of body form and allometry in larval and adult Pacific sculpins (Teleostei: Cottidae). Can. J. Zool., 63:1582–1589.

Received 10 September 1986; accepted 15 September 1987.

APPENDIX

SAS Procedures

The following is a SAS Proc matrix program to compute the reformulation of the shear procedure as described in the manuscript. It can be added to the end of the SAS procedure given in Bookstein et al. (1985:243–245). Alternatively, it can be added after the "EIGEN AP EP QP;" statement and the rest of the lines in their procedure can be discarded.

```
* reformulation of the shear method;
shraxes = e(,1);                          /* first axis is not sheared */
f1e1 = ep(,1)'*e(,1);
do j = 2 to &retain;
  f1ej = ep(,1)'*e(,j);
  fpp = f1e1*e(,1) + f1ej*e(,j);
  sclfact = f1e1*inv(fpp'*fpp);           /* scale factor */
  hshr = (f1ej*e(,1) - f1e1*e(,j))*sclfact;
  shraxes = shraxes || hshr;              /* append the jth sheared axis */ end;
```

For Burnaby's (1966) method the following statements can be added at the end of the SAS procedure given in Bookstein et al. (1985:243–245). Alternatively, they can be added after the "EIGEN AP EP QP;" statement and the rest of the lines in their procedure can be discarded.

```
                          /* construct an orthogonal operator*/
L = I.(&nvar) - ep(,1)*inv(ep(,1)'*ep(,1))*ep(,1)';
HPP = L*e;                /* Burnaby adjusted vectors */
XP = x*L;                 /* adjusted data */
```