Syst. Zool., 38(2):169–173, 1989

Allometry, Isometry and Shape in Principal Components Analysis

KEITH M. SOMERS

Department of Zoology, The University of Toronto, Toronto, Ontario M5S 1A1, Canada

Morphological variation in size and shape has physiological, ecological and taxonomic significance (e.g., Peters, 1983; Schmidt-Nielsen, 1984; Bookstein et al., 1985). As a result, there is considerable interest in partitioning components of variation associated with size from components associated with shape (e.g., Mosimann and James, 1979; Humphries et al., 1981; Reist, 1985; Sampson and Siegel, 1985; Somers, 1986; Corruccini, 1987).

Size-related variation is often evaluated within a bivariate framework involving logarithmic transformations and regression (e.g., Smith, 1980; Seim and Saether, 1983). As such, log-log regressions with slopes equal to 1.0 indicate bivariate changes in size with no concomitant change in shape (i.e., isometry), whereas slopes differing from 1.0 reflect bivariate changes in shape that are correlated with changes in size (i.e., allometry). This bivariate conceptualization of size and shape has been extrapolated to a multivariate model summarized by principal components analysis (PCA; e.g., see Jolicoeur, 1963). A PCA of a variance-covariance matrix of logarithmically transformed morphometric data generally extracts a first component (i.e., PC1) that summarizes variation associated with size, both isometric and allometric size (e.g., see Jolicoeur and Mosimann, 1960; Mosimann, 1970; Pimentel, 1979:56-63). Because allometric size represents bivariate shape that is correlated with isometric size, PC1 frequently contains information in both size and shape (e.g., Hopkins, 1966; Mosimann, 1970; Jungers and German, 1981; Hills, 1982; Corruccini, 1983).

The multivariate principal-component model also introduced a multidimensional concept of shape independent of size (i.e., the second and subsequent components where size-correlated variation is incorporated into PC1; e.g., Jolicoeur and Mosimann, 1960). The bivariate analogue of multivariate shape is the pattern displayed by the log-log regression residuals. Just as in regression, multivariate shape is contingent on the analytical approach and associated transformations, as well as the spatial pattern of the data. In addition, the study of multivariate shape is further complicated by the problem of deciding how many shape components contain non-trivial information (e.g., Linn, 1968; Horn and Engstrom, 1979; Gauch, 1982; Gibson et al., 1984; Stauffer et al., 1985).

SIZE-CONSTRAINED PCA

In 1986, I proposed a modification to PCA that produced a first axis summarizing variation in isometric size alone (i.e., size-constrained PCA; Somers, 1986). The first axis was specified a priori as an isometric size vector (e.g., see Jolicoeur, 1963) with its associated eigenvalue (e.g., see Morrison, 1976:269). Isometric size was factored from the correlation matrix of log-transformed characters by subtraction (i.e., matrix exhaustion) and the residual matrix was subsequently factored into orthogonal eigenvectors and associated eigenvalues. Unfortunately, size-constrained PCA produces negative eigenvalues when values in the residual matrix exceed ± 1.0 (see Somers, 1986; Rohlf and Bookstein, 1987).

Sundberg (1989) reports that the isometric size 'component' from size-constrained PCA is frequently correlated with the shape components. Two related factors contribute to these correlations. First, the removal of isometric size only eliminates isometric size variation such that correlated variation in allometric size is incorporated into subsequent shape axes (this was the objective of the size-constrained approach; see Somers, 1986). As a result of the correlation between isometric size and allometric size, the isometric size vector will not be orthogonal (i.e., independent) to subsequent eigenvectors, and thus the associated component scores will also be correlated. This type of correlation is expected a priori (Bookstein, pers. comm.); however, the magnitude of such correlations will depend on the relative amount of allometric size variation.

Second, the removal of isometric size also confounds calculation of the size-constrained component scores. In PCA, component scores are calculated by post-multiplying the original data by the eigenvector matrix. However, correlated components can result if the data are not appropriately centered and standardized before multiplication with the eigenvector matrix. That is, the data must conform with the association matrix that was used to generate the eigenvectors. (The consequences of this scaling artifact are rarely discussed, but inter-component correlations are also produced by a number of approaches, including multigroup PCA [e.g., see Pimentel, 1979:81–91] and sheared PCA [Rohlf and Bookstein, 1987].)

When PCA is based on a variance-covariance matrix, orthogonal component scores will result if the original data are centered and standardized by subtracting the character mean and dividing by N − 1 before post-multiplying by the eigenvector matrix (where N is the number of individuals or observations). Similarly, if the correlation matrix is employed, orthogonal component scores result if the original data are centered and standardized by subtracting the mean and dividing by the standard deviation associated with each character. However, in size-constrained PCA the eigenanalysis operates on a correlation matrix from which the isometric size effects have been subtracted. The usual centering and standardization associated with an analysis of a correlation matrix is inadequate for the size-constrained procedure, and thus the original size-constrained algorithm produces correlated component scores. These correlations include those between isometric size and subsequent shape axes as described by Sundberg (1989), as well as correlations between the shape components.

Thus, the correlations reported by Sundberg are not unexpected due to the correlation between isometric and allometric size, but the magnitude of the correlations is confounded by improper centering and standardization in the original algorithm. Instead of Sundberg's table 1, the effects of the correlation between isometric and allometric size should be evaluated with angular comparisons of the eigenvectors (i.e., the inverse cosine of the product of any two eigenvectors; see Pimentel [1979: 87] for details). In the turtle example presented in Somers (1986; i.e., table 3), the angle between the isometric vector and the second eigenvector is 80.3°, and with the third eigenvector is 88.1° (where 90° implies that the two vectors are orthogonal). Sundberg (1989) reports results for the 24 female turtles alone, and the corresponding angles are 89.6° and 82.0°, respectively. For comparison, the crayfish example (i.e., table 4 in Somers, 1986) produced an isometric vector at 78.0° with the second eigenvector, and at 54.5° with the third. However, in all three examples the angles between the second and third eigenvectors are all 90.0°, as would be expected from any eigenanalysis.

When the original size-constrained algorithm calculated component scores, the orthogonal relationship between the shape eigenvectors was distorted producing correlations between scores for the shape components. This distortion occurred because the raw data were inappropriately scaled. In size-constrained PCA, the raw data must be scaled by subtracting character means and dividing by the standard deviation in the usual fashion, but because the eigenanalysis operates on the residual correlation matrix the observation means must also be subtracted. This latter step was omitted in the original algorithm and is now incorporated into the current version. Using this approach, the shape components will be orthogonal if no negative eigenvalues are produced in the eigenanalysis. However, if negative eigenvalues occur, this modification will not eliminate correlations between the shape components, and thus alternate approaches to remove isometric size may be warranted.

One such alternative incorporates the isometric vector into Burnaby's growth-invariant approach to remove isometric size effects (e.g., Burnaby, 1966; Rohlf and Bookstein, 1987). Instead of subtracting the product of the isometric eigenvector and associated eigenvalue from the character correlation matrix as in size-constrained PCA, the P-by-P correlation matrix (of the P characters) should be pre- and post-multiplied by a P-by-P matrix that is equivalent to an identity matrix minus a P-by-P matrix of 1/P values (Krzanowski, pers. comm; also see Rohlf and Bookstein, 1987:360–362). The resultant eigenvectors, including the isometric vector, are orthogonal and the component scores are also uncorrelated (i.e., the original data are standardized by subtracting character means and dividing by the standard deviations as in a traditional PCA of a correlation matrix). Burnaby's approach has been incorporated into a version of the size-constrained PCA program and the results have been confirmed with several different data sets. Both programs are available upon request, and both will analyze data matrices of up to 30 characters for 500 individuals.

In addition, J. C. Gower (pers. comm.) noted that an isometric size vector can be removed quite simply by another related approach. Specifically, isometric size can be eliminated from a PCA using either a correlation or covariance matrix if the data are doubly centered. Thus, if a data matrix of log-transformed characters is centered by subtracting both row and column means (i.e., much like principal coordinates analysis or reciprocal averaging; e.g., Gower, 1966; Legendre and Legendre, 1983:295–299; Pielou, 1984:176–188), isometric size variation will be removed. But in contrast with traditional or size-constrained PCA, the last eigenvalue from doubly centered PCA equals zero. Thus, isometric size effects can be eliminated by subtracting the row mean from each character for a given observation (i.e., each row). In fact the vector of row means is highly correlated ($r = 0.999$; Somers, unpubl. data) with the first component from the size-constrained approach (as it should be; Bookstein, pers. comm.; also see Mosimann, 1970). Removing size effects with doubly centered PCA contrasts with regression-based procedures (e.g., Wood, 1983; Reist, 1985), but using the vector of row means as a surrogate size measure is not new (e.g., Mosimann, 1970; McGillivary and Johnston, 1987).

To compare results of the doubly centered PCA, I contrasted the first two shape eigenvectors of doubly centered, traditional, and Burnaby's size-constrained methods using non-centered Procrustes analysis (e.g., see Krzanowski, 1979, 1982). Based on results from four data sets, the shape eigenvectors from Burnaby's and the log-correlation approaches were most similar in every case (i.e., using Gower's residual sum of squares; see Gower, 1975). But shape eigenvectors from the doubly centered PCA most closely resembled eigenvectors from the log-covariance approach twice and the log-correlation approach twice. The fact that shape eigenvectors from the doubly centered PCA failed to consistently resemble the results from any one method may reflect differences in the data sets, but more likely the shape eigenvectors are quite variable for each data set (i.e., have large standard errors; see Gibson et al., 1984). Obviously these are preliminary comparisons and the utility of doubly centered PCA in morphometrics remains to be established (e.g., Corruccini, 1987; Rohlf and Bookstein, 1987).

SUMMARY

Sundberg (1989) identifies a problem of inter-component correlation in the original version of size-constrained PCA. Such correlations were produced by two factors: (1) the correlation between isometric and allometric size, and (2) the improper scaling of the raw data to calculate component scores. A correction is implemented in the current version of size-constrained PCA, but unfortunately, this correction fails if negative eigenvalues are encountered in the eigenanalysis. As a result, two alternate

methods to extract isometric size vectors are described. One approach incorporates an isometric vector into Burnaby's growth-invariant method (Burnaby, 1966; Rohlf and Bookstein, 1987). The second simply involves doubly centering the log-transformed data prior to PCA. The vector of observation means that is subtracted from the data in doubly centered PCA is almost perfectly correlated with the isometric size component from the size-constrained PCA, but associated shape components may vary. Detailed comparisons of the various methods that purport to isolate patterns of multivariate size and shape are warranted (e.g., see Corruccini, 1987; Rohlf and Bookstein, 1987).

ACKNOWLEDGMENTS

I thank P. Sundberg for identifying the problem in size-constrained PCA, and I thank W. J. Krzanowski and J. C. Gower for suggesting solutions to this problem. Discussions with M. D. Dennison, D. L. Fuller, R. H. Green, R. I. C. Hansell, D. A. Jackson, P. Legendre, L. Orlóci, R. A. Pimentel, M. S. Ridgway, J. D. Rising and J. D. Smith were most helpful in formulating this response. I thank M. E. Douglas, F. L. Bookstein and two anonymous reviewers for helpful comments on this note. I also acknowledge financial support from an NSERC PDF and an NSERC Operating Grant to H. H. Harvey.

REFERENCES

BOOKSTEIN, F. L., B. 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.

CORRUCCINI, R. S. 1983. Principal components for allometric analysis. Am. J. Phys. Anthropol. 60:451–453.

CORRUCCINI, R. S. 1987. Shape in morphometrics: Comparative analyses. Am. J. Phys. Anthropol., 73: 289–303.

GAUCH, H. G. 1982. Noise reduction by eigenvector ordinations. Ecology, 63:1643–1649.

GIBSON, A. R., A. J. BAKER, AND A. MOEED. 1984. Morphometric variation in introduced populations of the common myna (*Acridotheres tristis*): An application of the jackknife to principal component analysis. Syst. Zool., 33:408–421.

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

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

HILLS, M. 1982. Bivariate versus multivariate allometry: A note on a paper by Jungers and German. Am. J. Phys. Anthropol. 59:321–322.

HOPKINS, J. W. 1966. Some considerations in multivariate allometry. Biometrics, 22:747–760.

HORN, J. L., AND R. ENGSTROM. 1979. Cattell's scree test in relation to Bartlett's Chi-square test and other observations in the number of factors problem. Multivar. Behav. Res., 14:283–300.

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. Syst. Zool., 30:291–308.

JOLICOEUR, P. 1963. The multivariate generalization of the allometry equation. Biometrics, 19:497–499.

JOLICOEUR, P., AND J. E. MOSIMANN. 1960. Size and shape variation in the painted turtle. A principal component analysis. Growth, 24:339–354.

JUNGERS, W. L., AND R. Z. GERMAN. 1981. Ontogenetic and interspecific allometry in nonhuman primates: Bivariate versus multivariate analysis. Am. J. Phys. Anthropol., 55:195–202.

KRZANOWSKI, W. J. 1979. Between-group comparisons of principal components. J. Am. Statist. Assoc., 74:703–707; and 1981. J. Am. Statist. Assoc., 76:1022.

KRZANOWSKI, W. J. 1982. Between-group comparison of principal components—Some sampling results. J. Statist. Comput. Simul., 15:141–154.

LEGENDRE, L., AND P. LEGENDRE. 1983. Numerical ecology. Elsevier Sci. Publ. Co., New York.

LINN, R. L. 1968. A Monte Carlo approach to the number of factors problem. Psychometrika, 33:37–71.

MCGILLIVARY, W. B., AND R. F. JOHNSTON. 1987. Differences in sexual size dimorphism and body proportions between adult and subadult house sparrows in North America. Auk, 104:681–687.

MORRISON, D. F. 1976. Multivariate statistical methods. Second edition. McGraw-Hill Book Co., New York.

MOSIMANN, J. E. 1970. Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. J. Am. Statist. Assoc., 65:930–945.

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

PETERS, R. H. 1983. The ecological implications of body size. Cambridge Univ. Press, New York. 324 pp.

PIELOU, E. C. 1984. The interpretation of ecological data. A primer on classification and ordination. J. Wiley and Sons, Inc., New York.

PIMENTEL, R. A. 1979. Morphometrics: The multivariate analysis of biological data. Kendall/Hunt Publishing Co., Dubuque, Iowa.

REIST, J. D. 1985. An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. Can. J. Zool., 63:1429–1439.

ROHLF, F. J., AND F. L. BOOKSTEIN. 1987. A comment on shearing as a method for "size correction". Syst. Zool., 36:356–367.

SAMPSON, P. D., AND A. F. SIEGEL. 1985. The measure of "size" independent of "shape" for multivariate lognormal populations. J. Am. Statist. Assoc., 80: 910–914.

SCHMIDT-NIELSEN, K. 1984. Scaling: Why is animal size so important? Cambridge Univ. Press, New York. 241 pp.

SEIM, E., AND B.-E. SAETHER. 1983. On rethinking allometry: Which regression model to use? J. Theor. Biol., 104:161–168.

SMITH, R. J. 1980. Rethinking allometry. J. Theor. Biol. 87:97–111.

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

STAUFFER, D. F., E. O. GARTON, AND R. K. STEINHORST. 1985. A comparison of principal components from real and random data. Ecology, 66:1693–1698.

SUNDBERG, P. 1989. Shape and size-constrained principal components analysis. Syst. Zool., 38:166–168.

WOOD, D. S. 1983. Character transformations in phenetic studies using continuous morphometric variables. Syst. Zool., 32:125–131.

Received 29 July 1988, accepted 19 October 1988

Syst. Zool., 38(2):173–180, 1989

"Size and Shape": A Comment on Semantics

FRED L. BOOKSTEIN

Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan 48109

As may be apparent from the preceding interchange between Drs. Sundberg and Somers, considerable semantic difficulties are associated with the notions of "size," "shape," and their relationships, both in the systematics literature proper and in biometrics more generally. The terms have inconsistent sets of meanings separately; furthermore, the copula "and"—as in "size *and* shape"—carries a further assortment of connotations all its own, connotations likewise inconsistent from context to context. In this note I would like to sort out the various meanings of "size," "and," and "shape" under five rubrics. When applied to the same data, these approaches sort the diverse phenomena of our morphometric explanations—size variation, shape difference, allometry—into incompatible subsets. Honest but very irritating discrepancies can creep in if the investigator intentionally discards part of this information, as in projecting onto a subspace of "only shape." Although, at root, these approaches embody mere differences of algebraic formula, they tend to be couched in arguments about biological process instead; and so we go on generating more heat than light.

I will characterize the five techniques as they apply to data sets of homologously measured lengths. The same algorithms can be applied, of course, to other sorts of data as well, but the interpretations of "size" and "shape" naturally may be different. For the sake of diagrammatic clarity, I shall present each of the five main terminologies as it applies to a very simple data set of two measured lengths, A and B. (I most definitely recommend *against* doing morphometric analysis on such paltry data, however! The minimum unit of morphometric analysis is a triangle of landmarks.) The variables A and B are taken to be measured lengths of variances σ_A^2 and σ_B^2 with covariance σ_{AB}. All bivariate associations are taken as covariances, not correlations. (Before or after log transformation, the scale of morphometric variables is crucial to their analysis.)