YEARBOOK OF PHYSICAL ANTHROPOLOGY 45:63–91 (2002)

The Promise of Geometric Morphometrics

JOAN T. RICHTSMEIER,[1,2,3*] VALERIE BURKE DELEON,[3] AND SUBHASH R. LELE[4]

[1]*Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania 16802*
[2]*Center for Craniofacial Development and Disorders, Johns Hopkins Hospital, Baltimore, Maryland 21205*
[3]*Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*
[4]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada*

KEY WORDS landmark data; shape; size; form; invariance; statistical models

ABSTRACT Nontraditional or geometric morphometric methods have found wide application in the biological sciences, especially in anthropology, a field with a strong history of measurement of biological form. Controversy has arisen over which method is the "best" for quantifying the morphological difference between forms and for making proper statistical statements about the detected differences. This paper explains that many of these arguments are superfluous to the real issues that need to be understood by those wishing to apply morphometric methods to biological data. Validity, the ability of a method to find the correct answer, is rarely discussed and often ignored. We explain why demonstration of validity is a necessary step in the evaluation of methods used in morphometrics.

Focusing specifically on landmark data, we discuss the concepts of size and shape, and reiterate that since no unique definition of size exists, shape can only be recognized with reference to a chosen surrogate for size. We explain why only a limited class of information related to the morphology of an object can be known when landmark data are used. This observation has genuine consequences, as certain morphometric methods are based on models that require specific assumptions, some of which exceed what can be known from landmark data. We show that orientation of an object with reference to other objects in a sample can never be known, because this information is not included in landmark data. Consequently, a descriptor of form difference that contains information on orientation is flawed because that information does not arise from evidence within the data, but instead is a product of a chosen orientation scheme.

To illustrate these points, we apply superimposition, deformation, and linear distance-based morphometric methods to the analysis of a simulated data set for which the true differences are known. This analysis demonstrates the relative efficacy of various methods to reveal the true difference between forms. Our discussion is intended to be fair, but it will be obvious to the reader that we favor a particular approach. Our bias comes from the realization that morphometric methods should operate with a definition of form and form difference consistent with the limited class of information that can be known from landmark data. Answers based on information that can be known from the data are of more use to biological inquiry than those based on unjustifiable assumptions. Yrbk Phys Anthropol 45:63–91, 2002.

Grant sponsor: NIDCR; Grant numbers: 1 P50 DE11131 (Project II), 1 P60 DE13078 (Project VI); Grant sponsor: NSERC; Grant sponsor: NSF; Grant numbers: SBR-929083, SBER0049031.

*Correspondence to: Joan T. Richtsmeier, Department of Anthropology, Pennsylvania State University, 409 Carpenter Bldg., University Park, PA 16802. E-mail: jta10@psu.edu

DOI 10.1002/ajpa.10174
Published online in Wiley InterScience (www.interscience.wiley.com).

TABLE OF CONTENTS

Introduction 64
Conceptual Issues 65
Why measure? 65
Limitations of landmark data 65
What does morphometric data tell us (or, what are size and shape anyway)? 67
The three problems that constrain what we can know: orientation, orientation, and orientation 68
Models and methods in morphometrics 70
Landmark coordinate data and the choice of statistical models 70
Estimation of parameters: what *can* be known and estimated from a sample? 71
What *can* be known about the difference between forms? 74
The relationship of morphometric spaces to parameter estimation and statistical testing 75

Morphometric Methods for Studying Form Difference 77
Superimposition methods 77
Deformation methods 78
Linear distance-based methods 79
Analysis of data by various morphometric methods (or, "one fish, two fish, old fish, new fish") 80
Research design 81
Rationale 81
Generation of forms 81
Generation of landmark data 81
Methods 81
Superimposition approaches 81
Deformation approaches 82
Linear distance-based approaches 82
Results 82
Form 1 → Form 2 comparison 82
Form 1 → Form 3 comparison 84
Form 1 → Form 4 comparison 85
Synopsis 86
Summary and Conclusions 87
Acknowledgments 89
Literature Cited 89
Appendix: Terms and Concepts 91

INTRODUCTION

"The analysis of shape-variation and growth has turned out to be a more difficult problem than one might have thought a decade ago." (Reyment et al., 1984, p. 5.)

The study of shape and shape change is intrinsic to biological anthropology. Before modern statistics was established as a discipline, students of biology were observing shapes and recording metric observations in an attempt to understand the way in which biological forms varied from one another, to establish the correspondence between form and function, and to quantify the description of characteristic traits used in the identification of species. As evolutionary thought entered mainstream biology, quantitative studies were included in the discussion of the natural variation and phylogenetic relationships among species. Later, the long-standing biological interest in the relationship between morphology, development, and phylogeny introduced by von Baer took on a decidedly quantitative approach when rejuvenated by the work of others (Gould, 1977, 1981; Huxley, 1932).

As biological inquiry became more quantitative, a plethora of methods were borrowed from modern statistics, some of which (e.g., significance testing) have become mandatory in published analyses of biological data. Multivariate statistics provided an entirely new collection of tools. Whole series of observations designed to capture the essence of form could be analyzed simultaneously by these methods. Renewed interest in the work of D'Arcy Thompson during the latter half of the 20th century, most notably in the work of Bookstein (1978), steered the focus from multivariate space back to the geometry of biological form. Early geometry-based methodologies proposed by Boas (1905) and Sneath (1967) did not gain the attention they merited until later (for the specifics of Boas' contribution, see Cole, 1996). This movement and the methods developed subsequently comprise the field of *geometric morphometrics*. Defined as the fusion of geometry and biology (Bookstein, 1982), morphometrics deals with the study of form in two- or three-dimensional space. The field flourished from a desire to analyze biological forms in ways that preserved the physical integrity of form in two or three dimensions, and to avoid collapsing the form into a series of linear or angular measures that do not include information pertaining to geometric relationships of the whole (Fig. 1).

In this paper, we look at morphometrics against a background of the philosophy and practical application of various methods, and evaluate the competence of these tools to decipher the morphological complexities that are the product of biological processes. Given landmark data, we differentiate those aspects of form than *can* be known and those that *cannot* be known, and urge that these facts not be ignored. We outline the assumptions underlying certain methods, relating them to the limitations of evidence available from data sets, and emphasize the importance of the assumptions that underlie morphometric methods. Validity and appropriateness, as opposed to convenience and engaging graphics, should guide users in their choice of method and software.

With the ready availability of digitizers, powerful desktop computers, and morphometrics software packages, morphometric analyses can be performed with little comprehension of the mathematical or logical basis of the approach. This is a dangerous situation for any science. It is not our intention to survey the entire field of morphometrics in this paper (for reviews, see Richtsmeier et al., 1992; Rohlf

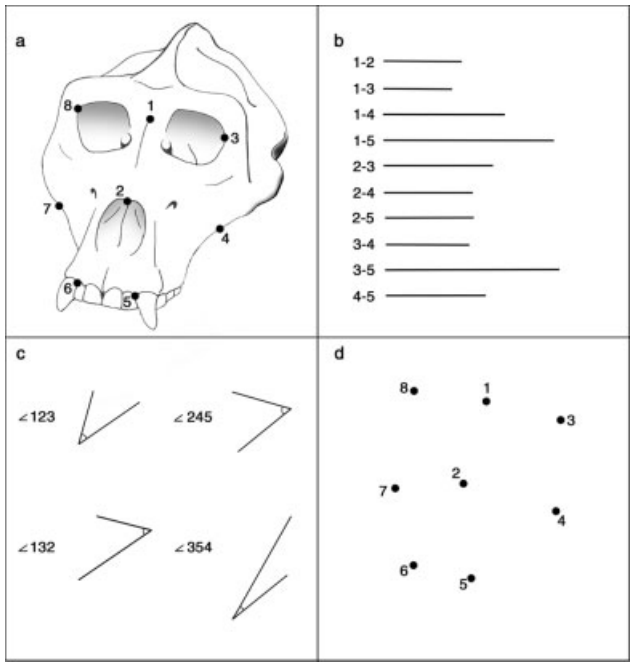

Fig. 1. Measure of a biological form captured by (**a**) the location of landmarks on the form, (**b**) selected individual linear distances between those landmarks, (**c**) selected angles formed by distances between landmarks, and (**d**) landmarks in a coordinate system. Relative location of landmarks (d) maintains the geometry of relative location of landmarks under study, while the inventory of linear distances (b) or angles (c) does not maintain the form, though geometric information can be reconstructed from this subset. Certain subsets of linear distances or angles can provide means to reconstruct an entire form (Lele and Richtsmeier, 2001; Strauss and Bookstein, 1982), but isolated linear distances and angles do not.

and Marcus, 1993). While we have experimented with other morphometric approaches, we have developed and worked extensively with Euclidean distance matrix analysis, and therefore admit a partisan viewpoint. We aim to demonstrate our viewpoint while showing the qualities of all approaches within a broad context of the properties of landmark data. We present a discussion of those items that must be considered when deciding on which morphometric method to use in analysis. To do so, we visit several key issues pertaining to the application of landmark-based morphometric techniques to biological data. We discuss statistical reasoning as it relates to the study of form, and provide an analysis of a simulated data set using several alternative morphometric methods. We provide theoretical and practical reasons for differences in the results given by various methods. The purpose of our presentation is to take a step back to remember the reasons why geometric morphometric techniques were developed, and the nature of the biological phenomena to which they are applied.

CONCEPTUAL ISSUES

Why measure?

The application of quantitative methods to biological data sets is usually done for two distinct purposes that are aimed at different objectives: hypothesis generation, and hypothesis confirmation. In morphometrics, the first approach involves the application of methods to discover new information within the data, and is called "data exploration." In this instance, quantitative methods are applied to a data set, and the investigator looks for patterns within the data that suggest underlying biological processes or effects. This exercise should not, and usually does not, take the form of blind application of methods to data. In the usual situation, a knowledgeable researcher applies methods to explore a "hunch." Although as such it is not a formal hypothesis test, the activity is driven by a hypothesis, or at least by an idea. This approach is often discouraged (especially by the major funding institutions), but is an aspect of pattern recognition, an extremely useful activity that can result in the discovery of complex relationships that might escape a more constrained hypothesis-driven analysis of the data.

The second approach involves more formal hypothesis testing. Here a question is formulated, data are collected that relate directly to the question posed, and comparisons are made with the intent of answering a specific question. Ideas are often formulated and tested as a null hypothesis of no effects and a nonspecific alternative hypothesis of nonzero effects.

Whether null hypotheses are proposed and tested, or alternate approaches are adopted, quantitative data are required to operationalize ideas about the morphology of biological phenomena under study. This, then, is the foremost reason that we measure: to explicitly propose, test, and defend our ideas to our scientific peers. But whether we collect data to test a hypothesis or explore a hunch, the data we collect are chosen with an explicit plan in mind. This, of course, presents the possibility that we neglect some features by targeting certain measurements.

Limitations of landmark data

Capturing geometry by way of landmark data has become rather commonplace. Landmarks are precise locations on biological forms that hold some developmental, functional, structural, or evolutionary significance. Diverse authors have discussed various types of landmarks (Bookstein, 1991; Marcus et al., 1996; Valeri et al., 1998; Lele and Richtsmeier, 2001). Landmark locations are recorded as two- or three-dimensional coordinates resulting in a spatial map of the relative location of the chosen points (Fig. 1d). When the same landmarks are collected on a number of objects, we refer to them as *corresponding* landmarks. The basis for this correspondence may be phylogenetic (these are sometimes called homologous points), structural, developmental, or biomechanical (Lele and Richtsmeier, 2001).

With the transfer of technology from the defense industry to the scientific community during the 1970s and 1980s, it became possible to easily and accurately record the location of points on an object in two- or three-dimensional space. Since biological objects are composed of many structural components whose location can be precisely defined, the identification of landmarks is simple. In craniometry as well as anatomy and paleontology, classical anatomical definitions of biological landmarks already existed. The location of these, as well as additional nontraditional landmarks, could now be accurately obtained as coordinate locations, providing the researcher with a map of the relative location of these points in space. Thus, data summarizing the geometry, or *form*, of biological objects were reliably and quickly recorded. This seemed a great opportunity for biologists, but the challenge remained to develop methods that used these data in their full, geometric configuration.

The use of landmarks has become widespread because landmarks are repeatable, because they provide geometric information in terms of the relative location of points, and because a variety of methods have been developed for their analysis. However, landmark data may not be the appropriate choice for all biological investigations. Salient features of morphology are overlooked when landmark data are used exclusively (Read and Lestrel, 1986). Landmarks do not contain information on the spaces, curves, or surfaces between them. If data concerning regions between landmarks are not part of the data collected, then we cannot expect to obtain verifiable information regarding the aspects of form or form change occurring between landmarks. However, a picture of the relative location of points (Fig. 1d) does not convey much information to the uninitiated without a superimposed outline. This realization has not been missed by the authors of landmark analyses, who often present results superimposed onto an outline of the forms studied, even though the outline is not considered in the analysis. These outlines are meant to provide a context within which to put the landmark-based results.

Clearly, data other than landmarks are available for morphometric analysis, and methods to analyze alternate data types such as outlines have been developed (e.g., Lestrel, 1982, 1989; Lohmann, 1983; Lohmann and Schweitzer, 1990; MacLeod, 1999; MacLeod and Rose, 1993; Read and Lestrel, 1986). These alternate data types are useful and appropriate for specific investigations, but are not considered in this paper. Here, we are concerned exclusively with the analysis of landmark data.

If landmark data are appropriate to the research question posed, then a choice must be made concerning which landmarks to include in analysis. Landmarks should be selected based on the biological questions to be addressed by the data, but measurement error must also be considered in the decision of which landmarks to include. Measurement error must always be evaluated in relation to the specific data collection machine and technique employed (e.g., Corner et al., 1992; Hildebolt and Vannier, 1988; Kohn and Cheverud, 1992; Richtsmeier et al., 1995; Valeri et al., 1998; Williams and Richtsmeier, 2002) and with consideration of the magnitude of the comparisons being made (Kohn and Cheverud, 1992). For example, assessment of the differences between two groups of juveniles requires more precision than comparison of juveniles and adults.

With the adoption of landmark coordinate data comes the harsh reality that there are certain things that *can* be known about the forms under study, and other things that simply *cannot* be known. First, as in all statistical studies, we can never know the true population parameters, but can only estimate these using a sample. Reasons for small sample size are as diverse as the topics studied by research scientists, but are common in anthropological research. Moreover, even with a large sample, we can only know certain aspects of the population mean and variance, and we cannot know others. Variability is particularly difficult to characterize, because each data set is collected in a coordinate system specific to the orientation of the object during data collection. The operations of translation, rotation, and reflection are routinely used to transport all forms into a single coordinate system to estimate variability. These operations do not make variability in landmark location knowable, but make it *appear* that variability can be estimated properly. Regrettably, the estimate of variability that is produced after registration is flawed. This issue is explored fully later in this paper.

Second, we cannot know the whole form when landmark data are used in analysis. The use of landmark data requires that choices be made regarding the exact number, location, and definition of landmarks identified to represent the form. Certain portions of the form may be overrepresented due to a high density of biological features, while other sections remain underrepresented. Practical considerations of time, efficiency, and measurement error, coupled with a need for biological relevance and a balanced representation of components, limit the number and nature of landmarks available for analysis.

Third, lack of a common coordinate system among forms represented by landmark coordinate data means that the difference between forms cannot be realistically measured in the context of a particular coordinate system. Methods that are coordinate system-based require that a particular coordinate system be adopted and, as we will show, this choice can profoundly influence analytical results.

This last constraint is not trivial, and cannot be dismissed by stating that matrix algebra enables easy transformation from one coordinate system to another. The lack of a common coordinate system forms the core of our thesis regarding what can and cannot be known about form and change in form as

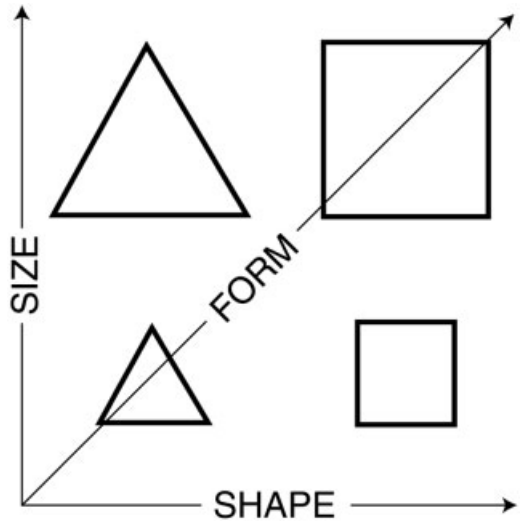

Fig. 2. Change in form is intuitively thought of as the result of a combination of change in size and change in shape. The terms size and shape have no unique definition, and are consequently problematic when used in precise studies of change in form. See Corruccini (1995) for a discussion of the efficacy of various methods in correctly identifying simple shapes.

quantified by the analysis of landmark data. The repercussions of varying coordinate systems among specimens under study are quite complex and serious if one wants to determine the true difference between forms. We will return to this limitation repeatedly in this article.

What does morphometric data tell us (or, what are size and shape anyway)?

Morphometrics, by definition, involves the quantitative study of form. It is intuitively understood that "form" consists of "size" and "shape" (Fig. 2), but the measures we collect to study form contain information pertaining to a combination of size and shape. A great deal of effort has been targeted at developing ways to separate these intertwined components, but such attempts often remove biologically interesting information from the analysis (Oxnard, 1978). The terms size and shape permeate the morphometrics literature, but their definition is not precise. We all have colloquial definitions of "size" and "shape," but can we rewrite these ideas as precise definitions?

A unique and precise definition of *size* does not exist. Any given surrogate for size (e.g., a specific length, area, centroid, or volume) is precisely defined within a study, but these definitions are specific to a given analysis (Corruccini, 1987; Jungers et al., 1995). Shape is defined on the basis of the chosen surrogate for size, so that as the choice of size measure changes, so does the definition of shape. These definitions cannot be applied across all data sets or across all studies.

Before the development of morphometric techniques for data collection and analysis, the quantitative study of size and shape was most closely linked with the field of allometry. In allometric studies, metric data are collected from organisms, and the relationship between size and shape is studied using an allometric equation, e.g., $y = bx^m$, where y is a variable whose increase is considered relative to that of another variable x. The variable x may represent a different dimension of the same organ or a measure of total body size (Gould, 1966). As one of the variables is usually thought a priori to represent size, the focus of allometry is the relationship between size and shape in populations of organisms.

Although this approach has produced our traditional ideas about how size and shape are related, the allometric equation defines shape only in relation to an associated size trajectory (Mosimann, 1979; Mosimann and James, 1979; for an insightful chronology of operational definitions of shape, see Godfrey and Sutherland, 1995). In allometric studies, size connotes magnitude and is often represented by a single surrogate measure (e.g., total length, weight), a linear combination of metrics (e.g., arithmetic mean), or a more complex combination of metrics (e.g., area, volume, geometric mean). None of these measures of size is more intrinsically suitable than any other measure, but the choice of which measure to use is of critical importance in determining the outcome of the analysis (Godfrey and Sutherland, 1995). Shape is not an individual measure that can be directly collected, replicated, or checked by consulting a single measure on any organism, and the exact definition of shape changes with the choice of the surrogate for size. Because this choice exists, the definition of size is nonunique. Though form is undeniably composed of size and shape, neither size nor shape can be defined uniquely (for a discussion on the case of landmark data, see Lele, 1991; for a scale-invariant approach, see Rao, 2000).

An immediate goal in many analyses is to discover a size variable that is statistically independent of "shape." But as Mosimann and James (1979) demonstrated, "shape" as defined in any study can be independent of only the chosen size measure. Consequently, for all but the chosen measure of size, associations exist between size and shape, and the nature of this association will vary depending on the size variable. Statistical independence is specific to the way in which both size and shape are measured or estimated, and does not necessarily translate to independence between our idiomatic ideas of size and shape. In biology there are few examples of organisms that grow or evolve by changing in size while maintaining a constant shape. Dwarf or giant species are rarely perfectly scaled versions of related organisms.

This is not to say that allometric studies are useless or misleading. Quite the contrary: allometric studies can be very powerful and informative. The researcher must remember, however, that "size" and "shape" are defined a priori and may not pertain directly to the geometry of the forms under study (Jungers et al., 1995; Mosimann, 1979; Sprent,

1972). Empiric definitions of size and shape have been implicitly meshed with colloquial meanings of the same terms, and the result is imprecise language, ineffective description, and unclear communication among scientists.

Despite the language of many studies, differences in size can never be removed or eliminated (but measures can be scale-invariant), one can never analyze "pure shape" (but shape with reference to a specified size variable can be defined), and "size" and "shape" are never biologically independent but are instead inextricably interrelated. Only the measures that an investigator *chooses* to represent size and shape may be independent (uncorrelated) in a statistical sense. Neither size nor shape can be uniquely defined.

Some morphometric studies may benefit from an attempt to reduce the influence of size. In a comparison of gorilla and chimp cranial anatomy, most if not all measurements from the gorilla would be greater than those from the chimp. The effect of size would likely overwhelm and therefore obscure the differences in "shape" between these two samples. However, in attempting to reduce the influence of size, we might choose to scale data by the geometric mean of all distances, or by centroid size, or by a single linear distance. This choice will affect the results of any comparison of what we think of as "shape." We emphasize that the researcher must be aware of these consequences. Many morphometric analyses automatically apply a scaling algorithm in calculating mean shapes and in comparing mean shapes from two or more samples. Scaling can be a useful step in certain research situations, but the results of these analyses must be interpreted *relative* to the chosen definition of size.

The three problems that constrain what we can know: orientation, orientation, and orientation

Biological organisms that constitute a group resemble each other to such a degree that we have an instinctive understanding of a typical or "average" form that is representative of all members of the group. We expect some members to be very similar to the average, while others will be less like the average. Since all forms differ from each other in various ways, a scheme for characterizing these differences is needed. It is convenient to organize and specify these differences as divergences from an average form. Individuals within a group can each be represented by landmark data, and a mean form can be estimated from these data. To make clear the problems associated with estimating a mean and variance from landmark data, we introduce the following scenario.

Imagine a machine at your favorite zoo that makes wax galago figurines. The mold in the machine produces identical galago forms. Being an anthropologist, you really like your wax galagos; you buy many of them from the machine, and then try to make each one distinctive by sculpting its eyes, and stretching or compressing its ears and tail a little differently. Your addition of details to each galago represents individual variation in the population of wax galagos. Assume that the mold is equivalent to the average of all these individual galagos (i.e., the mean form). By recording coordinate locations of salient features of the galagos, including the eyes, ears, and tails, variation from the mean can be recorded.

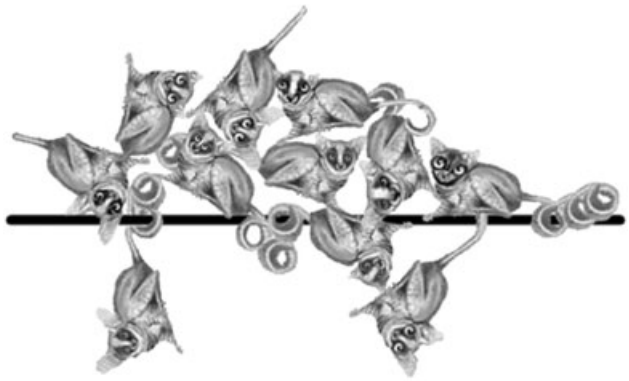

Fig. 3. Set of toy galagos created following guidelines described in the text constitutes our data. Face, ears, and tail of each galago differ from the mold (the mean) in terms of fine distinctions of features (landmark location) and in terms of orientation of each galago with respect to the mold (rotation, translation, and reflection). When describing this group of toy galagos mathematically, an error term characterizes differences in ear, tail, and facial features with respect to the mold. Rotation and translation parameters describe the orientation of each toy as it lies on the shelf with reference to the mold.

After each galago is made from the mold in the machine and modified by you, they are carelessly placed on a storage shelf (Fig. 3). This movement, from the mold to the shelf, consists of rotation and translation. *Rotation* refers to a change in orientation characterized as movement around an axis (Fig. 4). Mathematically, rotation of an object corresponds to multiplication of a landmark coordinate matrix by an orthogonal matrix. Upon rotation, the relative locations of the points representing any single galago remain the same, but the exact coordinates of these landmarks change. *Translation* corresponds to a form sliding in any direction (e.g., along a plane defined by the shelf) while remaining stable in terms of rotations around axes (Fig. 4). Mathematically, translation corresponds to adding a matrix of identical rows to a landmark coordinate matrix. As with rotation, the relative locations of points within any particular form are maintained when a form is translated, but the exact coordinates of the landmarks change. *Reflection* is another type of transformation, although not likely in this particular example. Reflection refers to flipping (or mirroring) an object across one axis or plane. Mathematically, reflection corresponds to multiplying the values of a column of a coordinate matrix by -1.0. In symmetric forms, the left side is a mirror image, or a reflection of the right side. Upon reflection, the relative loca-

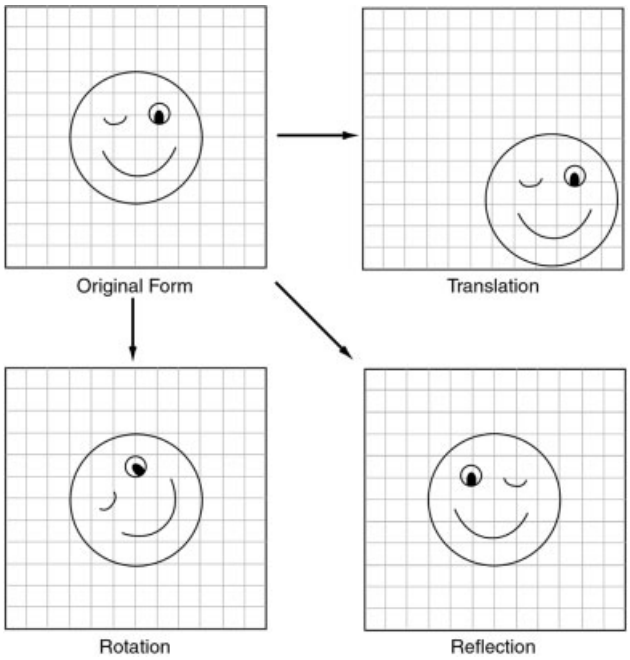

Fig. 4. Graphic demonstration of parameters of rotation, translation, and reflection, each of which corresponds with a particular mathematical operation (see text). Features on original form remain unchanged whether it is rotated, translated, or reflected, and therefore relative location of landmarks representing these features would remain invariant for all four forms. However, rotation, translation, and reflection of the whole form in relation to its original position result in different x and y coordinates for these features in the four forms.

tion of points remains the same, but their value along the axis or plane of reflection is reversed (in other words, if the sagittal plane is the plane of reflection, left becomes right).

Now, assume that the mold in your galago machine breaks and is no longer available. But you want to make more galagos using the same exact mold. Can you recreate the mean form using the galagos that you previously made? Each galago differs from the mold (the mean) in terms of nuances of sculpted detail and in terms of orientation with respect to the original position in the mold in the machine. An error term representing individual variation characterizes differences in sculpted detail with respect to the mold. The parameters of rotation and translation describe where and in what orientation each toy lands on the shelf with reference to its original position in the machine (mean form).

If we knew the exact path that each galago traveled from the machine to its position on the storage shelf, the inverse of these paths could be used to reorient them to their original positions in relation to the mold. We could use the information about each galago in its original position in the mold along with the average configuration of the landmarks to reconstruct the mean form. However, the rotation and translation of each galago away from its original orientation (the path each galago took as it moved from the mold within the machine to the storage shelf) are unknown and unknowable. Because our toy galagos are irregular three-dimensional forms, there is no particular edge, or other outside frame of reference, that can be used to put them together in a way that will necessarily bring us to the original arrangement. You might think you added the least amount of modification to the eyes and so try to align the eyes of all the galagos, but this is a poor substitute for the lost mold. We have an informed idea of what the mean form might look like, but the only data that we have are the galagos that we have created.

This situation is similar to that encountered by a biologist when collecting data from a sample of biological organisms. The mean form and the relationship of the other forms to the mean (the perturbation structure) are unknown. The rotation and translation required to place each form into a similar orientation are unknown and unknowable. Any orientation may be picked to estimate the mean and the variance, but as we will show, the choice of this orientation has implications for the estimation of the mean and variance.

As a biologist you are not really interested in galago molds, but instead in an accurate calculation of a mean form from your data. As a biologist, you are also not particularly interested in the difference in orientation among the forms, but you are interested in how the relative locations of features of each form differ from their arrangement on the mean form. Some biologists might argue that no "true" difference between forms can ever be known, because the true mean landmark configuration of a biological population does not (nor did it ever) have a true orientation. And if you argue that there is no true difference, then organisms cannot be defined by the way a particular landmark has moved relative to its original position (which never existed). But if these arguments are correct, what good are morphometric approaches?

In practice it is unrealistic to try to determine the true orientation of a biological form. For example, one might argue that orientation should be based on the primitive state (e.g., the orientation of the implanted fertilized egg or some other justifiable ontogenetic stage), but that choice would be as arbitrary as adopting the standard anatomical position of the adult as the chosen orientation. Forces of nature (e.g., gravity), locomotion, and evolutionary and ontogenetic change will work to make many chosen orientations impractical, or at least inconvenient when studying varieties of any form. An argument for the adoption of a particular coordinate system on mathematical grounds makes little biological sense. Unfortunately, most morphometric approaches require that an orientation be adopted for analysis. The software program often does this alignment routinely, with no input from the user.

The primary problem facing biologists who study form using morphometric techniques involves the relationship between the coordinate system local to

each object, the coordinate system used to collect landmark coordinate data from each object, and the coordinate system used for analysis. No information is available regarding how these coordinate systems relate to one another. Yet, to ascertain biological variability using the coordinate locations of landmark data, knowledge of the relationship of one form to another in their respective coordinate spaces (each galago in its original position in the mold in the machine) is essential. Even if organisms are not inherently defined in a pre-established, universal three-dimensional (3D) coordinate system, the adoption of a coordinate system is required by certain morphometric methods. If all superimposition schemes gave the same sample estimates, the choice would be trivial. Unfortunately, estimates of the mean and variance change with the coordinate system adopted, as do the results of form comparison.

In reality, genetic and environmental influences combine to affect structures, thereby creating perturbed forms. In geometric morphometrics, we collect landmark data to represent these forms. When observing a group of forms, we think of these individuals and the variation among them by relating them to a typical, representative form that does not exist but that we are able to envision. This is formally done in statistics by calculating an average or mean form, and measuring variation in the sample with reference to an average. To do this requires adoption of a statistical model. We now explore the role that models play in estimation of the mean and the variance.

Models and methods in morphometrics

If landmark data are appropriate for investigation of a specific research question, there are several methods available for their analysis. However, what is often unappreciated is the fact that the statistical model adopted by or incorporated into a method is a critical choice in the analysis of data. Every time you use a statistical package, your choice of a method also includes the specification of a statistical model. Because models are often not provided for the user's inspection and therefore not considered, the user may remain unaware of the implication(s) of the selected model for the analysis of data. Choice of the model and its approximation or "fit" to the reality under study is an important consideration in morphometrics.

In statistical analysis, model specification comes first. A *model*, as used here, is a mathematical construct that attempts to characterize certain aspects of the underlying phenomena (e.g., dimensions, dynamics, properties, interactions). This mathematical construct includes quantities called parameters. Using a specific model, parameters are estimated for each sample. Data per se are not needed to specify the model or characterize the parameters. However, knowledge of particular properties of the phenomena that the data represent can and should be included in proposing a model. Whatever characteristics the scientist deems important or explanatory should be included in the model. To specify a model in morphometrics, good statistical sense and solid knowledge of the phenomena under study are needed. If a model seems appropriate and correct, methods for checking the assumptions of the model can be developed (if they do not already exist) and applied.

For example, in predicting height based on femur length, we might use a simple linear regression model that describes the relationship between height and femur length of an individual. Data exist for femur length and height of individuals in a sample, and are denoted as variables X and Y, respectively. The statistical model used in this situation is: $Y = \beta_0 + \beta_1 X + \epsilon$. This model follows general knowledge of body size estimators, and states that the height of an individual is a linear function of femur length of the individual with the addition of some random variability, ϵ. The parameters of this model are the intercept β_0 and the slope β_1. The term ϵ denotes the variability (sometimes referred to as error) around the mean response, $\beta_0 + \beta_1 X$. Estimates of these parameters from the data provide information about the nature of the relationship between X and Y. The assumptions of the regression model can be checked by analysis of the residuals.

Note that the observations (data) do not enter into the specification of the model. A model is formulated using statistical expertise and intuition, based on whatever previous experience and knowledge the scientist may have. Once a model is formulated, data are used to determine those parameters of the model that are most compatible with the observations. This process is called "estimation of the parameters." In the example above, β_0 and β_1 are estimated from the data. Many methods for estimation of the parameters may exist and may be equally appropriate. A *method* is any technique used in estimating the parameters of the model and in further analysis such as hypothesis testing or calculation of confidence intervals. For example, in the linear regression case, one can use a least squares method, major axis method, or reduced major axis method for estimation of the parameters.

When conducting any scientific study, a model should be specified and understood first. Following the specification of a model, methods should be devised or chosen for estimating the parameters of the model and for conducting any other relevant data analysis. A particular method is judged as correct or incorrect, appropriate or inappropriate, only in relation to its efficacy under a particular model. An evaluation of the conclusions of any scientific analysis of data must consider the accuracy of the method and the validity of the model.

Landmark coordinate data and the choice of statistical models

When estimating statistical parameters for samples of forms represented by landmark data, a par-

ticular perturbation model is commonly adopted. This perturbation model specifies the variability of the sample around the mean, and is used to estimate parameters that describe the relationship of individuals to the mean. We have already shown by way of our galago example that the true relationship of the forms as represented by landmark coordinate data to the mean (mold) and to one another cannot be known. The information is simply not provided in the matrix of landmark coordinate locations. How, then, can a perturbation model for landmark data estimate the variation around the mean in terms of the parameters of rotation and translation, if that information is not contained in the data? To further clarify the reasoning behind the commonly used perturbation model for landmark data and its failure when applied to landmark data sets, basic matrix algebra is used.

Let M denote the landmark coordinate matrix corresponding to the mold. If the coordinates of 10 three-dimensional landmarks were collected to represent salient features on the face and body of the galago mold, then M is a matrix with 10 rows, one for each landmark, and three columns, one for each dimension. Since each galago that we have crafted is slightly different than the mold, this variation can be described as error, E_i. The error is different for each galago, as signified by the subscript i. Each galago can be represented by $M + E_i$. Recall that we tossed the galagos onto a shelf and have therefore moved them from their original position with reference to the mold, M. This movement can be described by two additional parameters: rotation and translation. The model that describes the collection of galagos takes the form:

$$X_i = (M + E_i)\Gamma_i + t_i$$

where X_i is the landmark coordinate matrix of the ith galago, M is the mean form or the landmark coordinate matrix representing the mold, E_i is the error or the variation that was added to each individual during sculpting, Γ_i is the rotation, and t_i is the translation that happened when the galago was dropped onto the shelf. This is the statistical model commonly used in morphometrics to describe variation of individuals with reference to a mean form (Bookstein, 1986; Dryden and Mardia, 1998; Goodall, 1991; Lele, 1993).

Let us think further about the parameters of this model for morphometric data sets. X_i represents the appearance of each individual in a sample (X_1 is the first individual, X_2 is the second individual, and so on). The mean form M has a fairly straightforward interpretation. M represents the average appearance of individuals in the population. The variance-covariance of E_i, the random variation about the mean, and the specifics of this parameter refer to perturbations at each landmark along each axis (see Lele and Richtsmeier, 2001). A model may specify that the variation is similar at all the landmarks or different at each landmark, or there might be correlation between these errors along two or more of the axes. Thus, E_i tells us about the pattern of change in the relative location of landmarks within a form with reference to the average configuration of landmarks. Γ_i refers to rotation of the entire form, and t_i refers to translation of the entire form with reference to the location and orientation of the mean, i.e., with reference to the way the mean mold was situated when fashioning the galagos.

Now remember that after sculpting eyes and styling ears and tails, we have all of the galagos but we have lost the mold, M, that was used to create them. Since the core of biology concerns the study of phenotypic variability, there are two parameters that need to be estimated correctly: the mean (or average) which is now lost, and the variability around it. Can we somehow use the galagos that we fashioned (the data) to recreate the mold, at least approximately? Can we also estimate the pattern of variability around the mean form from the data? It is at this juncture, at the point of estimation of parameters for a single sample, that the consequences of tossing the galagos onto the shelf (i.e., the introduction of rotation and translation) loom large with profound consequences.

Estimation of parameters: what *can* be known and estimated from a sample?

Let us consider the number of unknowns in the model, $X_i = (M + E_i)\Gamma_i + t_i$. The mean form, M, and variance-covariance structure of the errors, E_i, are unknown. In addition, the rotation and translation parameters, Γ_i and t_i, are unknown. The parameters M and the variance-covariance structure of E_i are fixed, meaning that they are constant with reference to the sample size. However, the parameters Γ_i and t_i are different for *every specimen*, because each individual has a unique orientation with reference to the mean. This means that we have a total of $(2 + 2n)$ unknowns for the given equation, the 2 referring to M and the variance-covariance of E_i for the sample,[1] and the 2n referring to Γ_i and t_i for *each individual* in the sample of size n. The number of unknowns $(2 + 2n)$ is therefore larger than the sample size (n). A basic tenet of inferential statistics is that one cannot estimate more parameters than the number of observations. Therefore, the parameters Γ_i and t_i are unknown and cannot be estimated.

If these parameters cannot be estimated, then why are they commonly estimated by certain morphometric techniques? Just because parameters are statistically inestimable does not mean a computer program won't estimate something when you ask for it. But just because a parameter is estimated does not mean it is correct. Given the nature of most

[1]Although, M, Γ_i, t_i, etc., are matrices, for the sake of exposition we consider them as single entities. Strictly speaking and in mathematical terms, the number of parameters is of O(n), which is the same order of magnitude as the sample size.

morphometric software, the user can be unaware that the estimates given could be invalid.

In statistical terminology, the rotation and translation parameters that are not of scientific interest are known as *nuisance parameters*. Neyman and Scott (1948) were the first to point out the problem of nuisance parameters. Lele (1993) originally described the consequences of the inclusion of nuisance parameters in morphometrics. The existence of nuisance parameters in the statistical model makes the accurate estimation of parameters of interest difficult. The rather dire penalty for including nuisance parameters in a model is that neither the mean form, M, nor the variance covariance matrix, E_i, can be estimated from the data. Likelihood-based methods that attempt such an estimation do not work properly.

From a scientific point of view, these nuisance parameters are of no real interest because knowledge of the orientation of an object with respect to its original position (as characterized by the rotation and translation parameters) is unimportant. Think of our galagos example: do we really want to know the path that each galago took as it was transferred from the mold to the shelf? We are really only interested in the mean form, M, and the variance-covariance structure of the errors, E_i. If the reader agrees that rotation and translation of the specimens with relation to the original position are uninteresting and unnecessary to the scientific question being asked, then the situation is less dismal. If we focus only on M and the variance-covariance of E_i, then the number of unknowns is fixed and does not change with sample size.

So, how can the mean and variance covariance structure be estimated correctly? To understand how landmark data can be used for statistical analysis of biological forms, we need to carefully consider the mathematical definition of "form:"

The form of an object is that characteristic that remains invariant under any translation, rotation, or reflection of the object.

This definition requires that we use the coordinate data to summarize form without using a coordinate system. Put in the context of our galago example, this definition means that the exact coordinates of the mold of the galago in its original orientation, M, are not important if we are interested in the study of form. Instead, we express the data in such a way that all translated, rotated, and/or reflected versions of M are equivalent, and any of them will suffice for our scientific purpose.

For example, let us say that a form P consists of three two-dimensional landmarks that are arranged on a single sheet of paper. Think of making hundreds of exact copies of that form by simply copying the piece of paper. If you drop that stack of copies onto the floor, the forms remain congruent and differ only on the basis of orientation; i.e., they are rotated, translated, and reflected versions of the original.

Let us assume that besides the actual coordinates of the three landmarks that make up P, we have the following information: together the three landmarks form an equilateral triangle, and the length of one side of P is known. Knowing that the form is an equilateral triangle and knowing the length of one side provide all the information we need to reconstruct P up to translation, rotation, and reflection, i.e., if we are not choosy about the orientation of P. According to the definition of form given above, we do not need the coordinates of the landmarks to define the form. Any particular coordinate-based representation of P is equivalent to all other representations of P expressed in any other coordinate system, as they differ only on the basis of translation, rotation, and/or reflection, and our definition does not include that information. All equilateral triangles with legs of a given length have the same form. Formally, we say that all equilateral triangles with legs of a given length (the various coordinate system-based representations of the form P that fell onto the floor) are invariant under rotation, translation, and reflection, and that they occupy an *orbit* defined by P.

In landmark-based morphometrics, the collection of *all* landmark coordinate matrices that can be obtained by any rotation, reflection, and translation of a given landmark coordinate matrix is the *orbit*. All landmark coordinate matrices within any single orbit characterize exactly the same form, because they differ only on the basis of translation, rotation, or reflection. We cannot know the exact location of any particular landmark coordinate matrix on the identified orbit, but in knowing the orbit, we know the form. The concept of orbit is closely related to the invariance principle, and is used routinely in statistics. Berger (1980) provides a precise definition of orbit in a statistical context.

As an aid to understanding the idea of an orbit and the definition of form with respect to orbit, consider a topographical map of a mountainous area. Each contour on a topographic map corresponds to a surface of constant (equal) elevation. No matter where a point lies in longitude or latitude, elevation remains the same as long as the point stays on the defined contour. In other words, as long as a point remains on a single contour line, elevation is *invariant* with respect to latitude and longitude.

Suppose that a person is hiking on the surface described by the topographic map (Fig. 5). Suppose further that this hiker carries a sensor that sends out a signal that identifies his elevation exactly, but only his elevation. We would like to locate this hiker, but if our data consist only of the signal from his sensor, we can place the hiker onto a contour (the orbit) defined by a particular elevation, but not to any specific location along that contour. This is analogous to the limitations of our knowledge when dealing with landmark coordinate data and asking questions about the parameters. We can know the *form* (as defined previously) of an object precisely, but we

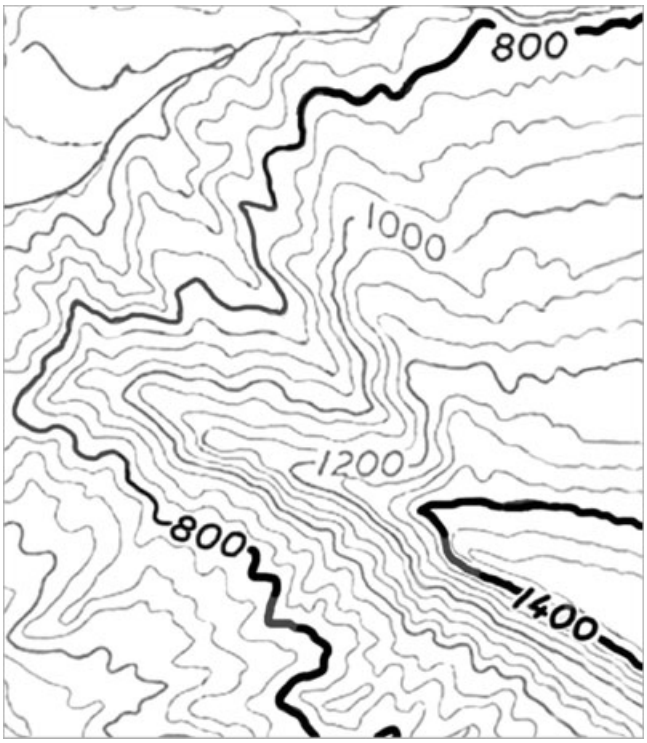

Fig. 5. Each band on this topographic map indicates a region in which elevation is the same. We highlight the distribution of two elevation bands, one at 800 feet and one at 1,400 feet. Contiguous bands indicate changing elevation as measured in vertical feet. Where bands are close together, topography is steep. Where bands are wider, the increase or decrease in elevation is more gradual.

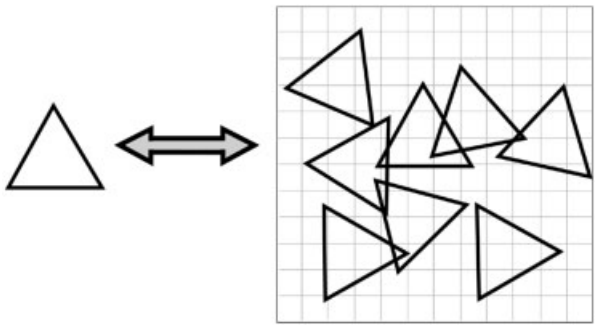

Fig. 6. Form of an object (here a triangle) is that characteristic that remains invariant under any translation, rotation, or reflection of an object. The form, free of a referential coordinate space, is equivalent to the *orbit* (shown at right) of all possible rotations, translations, and reflections of the data.

can never know the exact orientation of a form. All rotated, translated, and reflected versions of any form are geometrically congruent.

Remember that M is a coordinate matrix of two or three dimensions. Consider the infinite number of matrices that can be obtained by rotating, reflecting, and translating M. Since there has been no change in the *relative* locations of landmarks, all of these matrices represent forms identical to M and can be considered equivalent to one another. The collection of all matrices equivalent to M so obtained is the *orbit* defined by M (Fig. 6). Each rotated or translated version of M occupying the orbit defined by M is considered an element of the orbit. The only way to specify the element is to provide its rotation and translation with respect to M. Except for very specific biological questions, it is sufficient to know the orbit of M, but it is not necessary, nor is it possible, to know the exact element of the orbit.

The critical question then becomes: given landmark coordinate data, can we estimate the orbit to which M (the mold) belongs? The answer is, yes. One way to do this is to rewrite the landmark coordinate matrix as a matrix of all possible linear distances among unique pairs of landmarks. Since the matrix of linear distances does not change with rotation, translation, or reflection of the form, by defining the matrix of linear distances between unique pairs of points, we define the orbit to which that form belongs. Unfortunately, only certain features of the variance-covariance matrix associated with the errors E_i can be estimated. The mathematical details of this condition are beyond the scope of this presentation, but those interested may consult Lele and Richtsmeier (2001). We have shown elsewhere that the features of E_i that *can* be estimated are sufficient for conducting null hypothesis testing and for estimating confidence intervals for difference in form (Lele, 1991; Lele and Richtsmeier, 1995).

There are several methods available for the estimation of the orbit of M as well as for the variance-covariance structure of E_i. To make an educated choice of which method to apply to data, several issues should be addressed. Issues that need to be considered when evaluating the validity and performance of methods used for estimating parameters for a single sample include but are not limited to bias, consistency, and efficiency of an estimator. Definitions of these criteria are provided in the Appendix. With these qualities in mind, three of the methods used for estimation in morphometrics are discussed below.

1) Method of moments. The method of moments estimators of the mean form and the variance-covariance structure suggested by Lele (1993) are consistent and simple to compute.
2) Generalized Procrustes. The generalized Procrustes method of estimation, as well as the estimator suggested by Bookstein (1991) based on shape coordinates, yield *in*consistent estimators of the mean form, mean shape, and variance-covariance structure under realistic models (Kent and Mardia, 1997; Lele, 1993).
3) Maximum likelihood. The maximum likelihood estimators suggested by Dryden and Mardia (1998) are consistent. They are a bit more efficient than the method of moments estimators suggested by Lele (1993), but they suffer from computational problems of nonconvergence (see Appendix) of the numerical maximization routine

(Lele and McCulloch, 2002; Lele and Richtsmeier, 2001).

In general, for a given set of data, these methods all give fairly similar mean form and mean shape estimators. However, generalized Procrustes analysis estimators of the variance-covariance matrix are flawed (for an illustration and reasoning, see Lele, 1993). A consequence of incorrect estimation of the variance-covariance matrix is that the statistical inference procedures that utilize generalized Procrustes estimators can produce inaccurate results, since all statistical testing procedures are based on the estimation of the variance-covariance matrix. Walker (2001) recently reiterated the conclusions of Lele (1993) by reporting the inability of Procrustes methods to estimate the correct variance-covariance structure and the associated implications for statistical inference. Although Procrustes may be the most common method of estimation in morphometrics, the variance-covariance structure estimated is wrong. This estimate is crucial for statistical testing and ultimately for making inferences about how or why organisms differ, since these inferences are often based on the results of statistical tests. We propose the use of either the method of moments estimators or the maximum likelihood estimators that are based on the exact shape density.

Estimation of parameters is a critical step in the quantitative description of a single sample. We have presented the galago experiment, mathematical evidence, and the hiker analogy to demonstrate that the mean form and variance estimated for a sample represented by landmark data can only be known *up to* translation, rotation, and scaling. That is, the mean and variance can only be estimated correctly if the nuisance parameters of rotation and translation are not included in the estimation procedure. Below, we discuss the consequences of this finding for the scientific problem of form comparison.

What *can* be known about the difference between forms?

The fact that we can never know the orientation of objects as they relate to one another presents analogous problems for the comparison of samples of forms, and for the testing of statistical hypotheses. To understand the limitation as it relates specifically to determining the difference between forms, let us continue with the hiker example. Suppose we know that our hiker started his hike at a contour associated with an elevation of 800 feet above sea level, but that he is now at a contour associated with an elevation of 1,400 feet. If our information about the hiker remains limited to elevation, we have no information regarding his exact starting or finishing place, and can only know with certainty that the hiker has ascended 600 vertical feet. This result is invariant to the exact location of the hiker on the initial contour and to the location of the hiker on the final contour. Although we would like to locate the hiker and describe the exact path followed to gain this elevation, without further information we cannot. We can assign a "sensible" starting point for the hiker, but this information is not verifiable and may prove wrong or misleading. The description of the exact path and his final location can be provided if, and only if, we know exactly the point where the hiker started and every intermediate point that he traversed as he attained the higher elevation. If our information is limited to elevation of his start and his finish, we simply do not have the information to obtain a description of his path.

This is similar to the information available when forms are compared. We can unambiguously know the orbit of the first form and the orbit of the second form, but we cannot know the exact location of the forms on their respective orbits, nor the exact transformation required to shift from one grbit to another, as there are infinitely many possible transformations. We can assume a location on the initial orbit (i.e., choose an orientation) and assume a location on the second orbit, but neither of these is verifiable, and any specified choice may prove wrong or misleading. Since we can only know the orbit to which a form belongs, the difference between forms can only be studied correctly as the difference between orbits that the forms occupy.

Let us assume that we know there is a long trail that includes several sets of steps between the elevations of 800 and 1,400 feet (Fig. 7). The most parsimonious conclusion to our question of the way in which the hiker moved from 800 to 1,400 feet is that the hiker used this path to gain 600 feet in elevation. But perhaps the path was designed for average climbers, and this particular hiker prefers a challenge. The hiker forges trails up embankments and scales nearly vertical walls to reach 1,400 feet. Unfortunately, we do not have this information, and our assumption of parsimony forces our interpretation of the hiker's route onto the easiest trail. If the hiker is actually a world-class climber, the assumption of parsimony provides us with incorrect information regarding the hiker's beginning and ending points and how he got from the starting point to the ending point. As scientists, our choice is either to: 1) use only the information that is known and limit our answer to what we are sure of, or 2) include assumptions that are not testable (and may be wrong), but that provide a more "complete" or "satisfying" answer. Inclusion of information other than what is unambiguously known may provide a seemingly more complete, but potentially erroneous answer.

As morphometricians using landmark data, we are forced to establish the true form difference using a coordinate system because that is the nature of the data that have been collected. Whether or not a pre-established universal 3D coordinate system exists in nature is a subject for philosophical debate, but whether we can know the true difference between biological forms represented by landmark data is not. If we cannot know the true difference

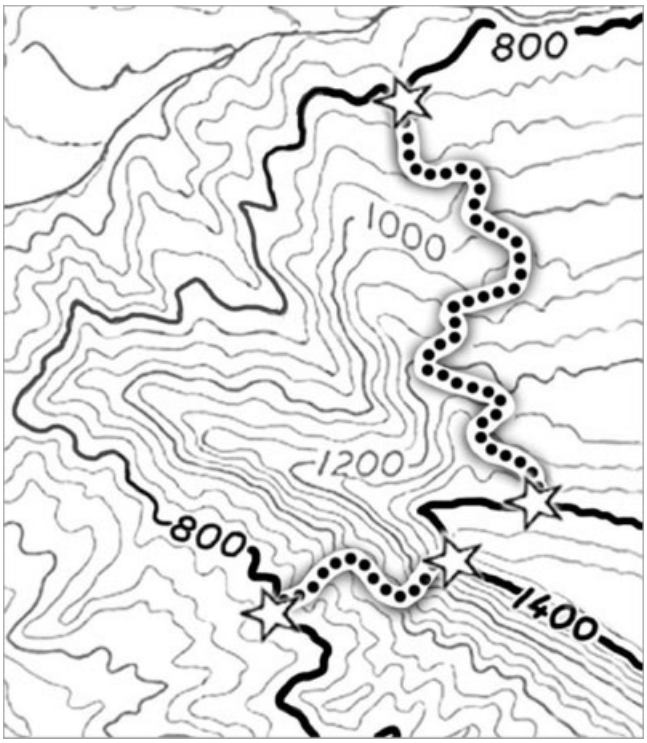

Fig. 7. Two of the infinite possible routes that could be taken from elevation of 800 to 1,400 feet by our hiker are shown as dotted lines on this topographic map. Stars designate the two possible starting and ending positions of each path. One path is long and gentle; the other is short and steep. To determine the hiker's route, we can assume parsimony and conclude that he took the longer, undemanding, but well-used path indicated. This assumption forces a choice in the hiker's exact starting and ending location, although it is equally probable that he took another route with a different starting and ending position, as indicated on the map. This situation mimics what is done when superimposition or deformation methods are used to determine difference in forms. Since landmark data do not provide information about orientation of forms, an orientation must be chosen for the initial form. Other forms adopt this orientation and are "fit" to the beginning form. "Fit" is determined by choice of minimization criteria (the path). Data do not to indicate which fitting criterion should be employed; the choice is arbitrary and has direct implications for the results of any analysis.

between forms using certain approaches, but only multiple answers that differ on the basis of the chosen alignment criteria, why waste our time with morphometrics? Scientific inquiry is aimed at discovering the truth. If forms exist, a true form difference exists, and there are ways to determine it using morphometric approaches.

We showed in a previous example that the determination of the locus of craniofacial deformity in Apert syndrome, a genetically transmitted condition that causes syndactyly of the hands and feet and severe craniofacial malformations, shifts depending on the edge chosen for superimposition (Richtsmeier and Cheverud, 1986; Lele and Richtsmeier, 2001, Fig. 4.3). It is on the basis of these types of studies that surgeons plan reconstructive surgery for children with Apert syndrome. If one superimposition scheme identifies the loci of the deformity at the base of the posterior cranial fossa and the lower face, while another shows the upper face to be primarily affected, on what structures does the surgeon operate? What information informs the surgeon which superimposition is the true one, which structures are truly deformed, and how he should plan his surgery? The arbitrary choice of a superimposition scheme should not influence the surgical procedure.

Dissimilar outcomes from the application of differing registration systems are not different descriptions of the same answers; they are different answers to the same question. The need for determining the true form difference is made clear from this example, but are questions relating to evolution, ontogeny, or function any less worthy of a valid answer? The existence of a true coordinate system has no bearing on whether or not we can accurately determine the true difference between forms. All that is needed is a method that uses only the information that can be known from the landmark data. When using landmark data, the information that is unambiguously known consists of the identification of the orbit on which the forms lie. An invariant descriptor of form change utilizes only this unambiguous information and describes the difference between forms in terms of the orbits that the forms occupy.

The relationship of morphometric spaces to parameter estimation and statistical testing

The past 10 years of morphometric research have resulted in the conceptualization of a surprising number of different "spaces" that are used for the analysis of morphology. Lively discussion has centered on the positive and negative aspects of these spaces that include, among others, Kendall's shape space (Kendall, 1994), Kent's tangent shape space (Dryden and Mardia, 1998), and form space (Lele and Richtsmeier, 2001; Richtsmeier and Lele, 1993). What are these spaces, and how do they relate to the space of coordinate matrices and orbits that we discussed above? Is any particular space statistically better than any other space? How can the claims about the statistical superiority of one space be evaluated?

Previously, we discussed the concepts of equivalent forms and the orbit defined by these equivalent forms. A maximal invariant statistic is any function that maps the entire orbit (i.e., the infinite collection of all equivalent forms, or all possible rotations and translations of a given form) onto a single point. One such function is the *form matrix* used in Euclidean distance matrix analysis (EDMA) (Lele and Richtsmeier, 2001). Another such function is *shape coordinates*, used by Dryden and Mardia (1998). These are just two of all the possible functions that share this property. Many similar functions can be constructed. A particular range, or a set of values that the function can take, identifies each function. The range is the defining feature of the space. The choice of the function then determines the characteristics of the shape space or form space obtained. The char-

acteristics of the space determine the appropriate statistical approaches for analysis within the chosen shape space or form space.

All maximal invariants are statistically equivalent and have identical distributions.[2] In morphometrics, the key concept is the orbit. In any given space, a maximal invariant is described by a given function. The function (and by extension, the space) that is chosen to represent the orbit is not critical. An orbit defined by two different maximal invariants (in two different spaces) is still the identical orbit; it does not matter which maximal invariant is chosen. It follows that *all* shape spaces (as long as they are maximal invariants) are statistically equivalent, and *all* form spaces (as long as they are also maximal invariants) are statistically equivalent. Some spaces might be mathematically more convenient than others, but statistical inferences (such as maximum likelihood estimation) conducted in one space are by definition identical to similar inferences conducted in any other space. The ability to correctly identify the parameters depends on identification of the orbit and not on the use of a particular maximal invariant.

Why then do investigators make claims that certain methods operate "better" than others in particular spaces? If all the functions are similar, how can it matter which function is chosen for analysis? Let us consider the problem of classification as an example. Suppose we use a particular dissimilarity measure such as Mahalanobis distance for classification when studying forms in a particular form/shape space. This distance measure is only appropriate if the form or shape space is a Euclidean space, e.g., if the space is ordinary three-dimensional space. If it is not a Euclidean space, then Mahalanobis distance is an inappropriate choice, and another distance measure should be used. Most of the shape and form spaces referred to in the literature are not Euclidean spaces. It is possible therefore that an investigator could get very odd results if Mahalanobis distance were used for classification in one of the many non-Euclidean spaces. The important thing to remember is that any shape space can be chosen for analysis, but any metric or statistic used within a space must be compatible with the characteristics of that space. In short, it does not matter if a space is Euclidean or non-Euclidean. What matters is that the method chosen to analyze phenomena within a space is appropriate to that space.

When studying form and the difference in form using Kendall's shape space, all forms are first aligned using Kendall shape coordinates. Once registered, each form can be represented as a single point on the sphere that is Kendall's shape space. Procrustes distances are used as a metric on this space. Rohlf (2000) asserts that similar objects have low variance and thus will be tightly clustered in Kendall's shape space. However, we emphasize here that variability as measured in Kendall's shape space is not the same as the variability indicated by the variance-covariance parameter in the perturbation model.

When forms are tightly clustered in Kendall's shape space, a space can be defined on a plane tangent to Kendall's shape space; this plane is called Kent's tangent space. Once defined, forms can be projected to Kent's tangent space. This is a critical projection, because unlike most shape spaces, Kent's tangent space is a Euclidean space and therefore the use of standard multivariate statistical analysis tools is permitted.

As noted above, the assumption of small within-sample variance is commonly used as the condition for using Kent's tangent shape space. Since biological variation is the single most important issue that we study, it seems counterintuitive to assume small variance to enable the use of a specific method and shape space. Moreover, if we assume small variance but variance is in fact large and irregular over an organism (see Lele and Richtsmeier, 1990), Kent's tangent shape space is unsuitable for the study of form difference.

The arrangement of forms in Kendall's shape space and the spread of the same forms in Kent's tangent space appear different. Similarly, the arrangement of the same forms in the EDMA-based form space will look quite different from the arrangement seen in Kendall's shape space. This, of course, does not mean that either Kendall's shape space or Kent's tangent space or the EDMA-based form space is wrong. The range that defines each of these spaces is different, so we expect that the same forms projected into different spaces will take on a different arrangement. Additionally, since the range of the various spaces is different, different metrics are required to measure form difference in each space. In other words, a metric appropriate for measuring the difference between forms, or variability among forms, in one space is most likely inappropriate for use in another space.

Rohlf (2000) concluded that form space is erroneous because the classification obtained when using this space differs from the classification obtained using Kendall's shape space. Contrary to his conclusions, the difficulty reported by Rolhf (2000) does not arise due to a problem inherent to any particular shape space. Instead, it arises because the statistical technique applied is not compatible with the characteristics of the particular space. This problem is closely related to the reason why the standard formula for calculating variance cannot be used in all shape spaces. It was established that the usual variance calculation cannot be applied to the shape space used in Procrustes analysis (Lele, 1993; Lele

[2]All maximal invariants are statistically equivalent. The distributions of all maximal invariants are identical, except for a constant multiplier corresponding to the Jacobian of the transformation between invariants.

and Richtsmeier, 2001; Walker, 2001). Similarly, the problem described by Rohlf (2000) lies in the inappropriate application of a valid statistical technique rather than the inadequacy of a particular shape space for conducting statistical inference. Only when appropriate statistical techniques are applied within a particular space can they be evaluated on the basis of validity, power, and effect size. Short definitions of these concepts are given in the Appendix.

Morphometric methods have provided new ways to think about and to measure form and form change. The user needs to take care that the metrics used are appropriate to the method. When a metric is used, it is desirable to obtain a statistical evaluation of the measured difference. However, if the method applied is not valid, it makes little sense to discuss the statistical properties of the estimates or of the results. Some of the available morphometric methods are discussed next.

MORPHOMETRIC METHODS FOR STUDYING FORM DIFFERENCE

One can study the differences between forms using any one of the methods that belong to three broad classes: superimposition methods, deformation methods, and methods based on linear distances. We examine these methods in detail below, and assume that the mean form is obtained using either the method of moments or the method of maximum likelihood for each of the two samples under study. We then analyze a hypothetical data set, using methods from each of these classes.

Superimposition methods

Superimposition methods involve the arrangement of landmark data from two forms into the same coordinate space, according to a specified rule. One form is designated the "reference" form, and the other is designated the "target" form. Form change is determined by the displacement of landmarks in the target form from the corresponding landmarks in the reference form. The researcher chooses a particular rule of superimposition. For example, in clinical cephalometric x-ray studies, a registration-based method of superimposition is often used, where tracings of x-rays are superimposed on the landmark sella (i.e., translated, so that sella, the centers of the pituitary fossa on each form overlay each other exactly) and registered (i.e., rotated to a standardized position) along the line sella-nasion. Historically, this superimposition was chosen based on the assumptions that sella is the most stable landmark in the cranium, and that the anterior cranial fossa is relatively stable. By choosing this superimposition rule, however, any true displacement of the landmark sella is concealed, displacement at the nasion will be constrained, and those changes will be attributed to other landmarks (Lele, 1991; Moyers and Bookstein, 1979; Richtsmeier and Cheverud, 1986).

All superimposition techniques (e.g., Procrustean approaches, Bookstein's edge matching, and roentgenographic cephalometry) involve three steps:

1) Fix one of the mean forms in a particular orientation and call it the reference object.
2) Translate and rotate the other mean form so that it matches the reference object according to some criterion.
3) Study the magnitude and direction of difference between forms at each landmark.

Different criteria for matching provide different superimpositions. For example, the least squares criterion (where the forms are superimposed so that the sum of the squared distances between corresponding landmarks on the two forms are minimized) leads to a generalized Procrustes superimposition. Matching a specific edge leads to the superimposition used in roentgenographic cephalometry (Broadbent et al., 1975) and to that obtained using Bookstein's edge-matching approach (Bookstein, 1982). It is commonly stated that differences in rotation, translation, and scaling of forms are *removed* or eliminated by superimposition. However, in all superimposition approaches, the alignment of forms is based on the estimation of the parameters of rotation, translation, and scaling. Superimposition does not remove differences due to these parameters; instead, it incorporates this registration into the definition of form and the difference between forms for any particular analysis. The nuisance parameters are not eliminated once estimated. They are fixed arbitrarily and then ignored.

If the scientific inferences obtained using different superimposition schemes were identical to each other, then the use of varying superimpositions would not pose any real problem. However, as is clear from our analysis of simulated data sets (see below) and various other sources (e.g., Lele, 1991, 1999; Lele and Richtsmeier, 2001; Richtsmeier, 1987; Richtsmeier and Cheverud, 1986; Rohlf and Slice, 1990; Siegel and Benson, 1982), different superimposition schemes give different results that yield different scientific inferences. How does an investigator choose which superimposition to use?

Siegel and Benson (1982) suggested that a generalized Procrustes fitting criterion should be used when the difference between forms is spread evenly over the objects, while a robust superimposition technique should be used if the difference between forms is concentrated in a circumscribed region (for a similar comparison of superimposition techniques, see Rohlf and Slice, 1990). Could data collected for scientific purposes ever instruct us in choosing one superimposition scheme over another?

Think of comparing two samples using two different fitting criteria, e.g., generalized Procrustes analysis, and robust or resistant fit analysis. The superimposition obtained by the use of each fitting criterion can each be thought of as a hypothesis

pertaining to the difference between the forms. Critical to any scientific hypothesis is its potential for falsification (Popper, 1959). Platt (1964) argued that an alternate characteristic, the possibility of choosing between competing hypotheses, is the hallmark of scientific method (see also Chamberlain, 1965). To fulfill this characteristic, it is required that, given enough data, the hypothesis that most closely approximates the truth can be selected from two competing hypotheses. If this cannot be done for a pair of hypotheses, then the two hypotheses are considered indistinguishable, given the data. Since no information is incorporated in landmark data to inform the investigator of which superimposition scheme to use in analysis, even infinite data sets will not allow us to properly choose among fitting criteria.

Obviously, if the two hypotheses (in our case, the result of the application of two superimposition schemes to the same data set) lead to identical scientific inferences, then the inability to statistically distinguish between the hypotheses does not present a problem. However, if the scientific inferences and decisions made on the basis of one hypothesis as opposed to the other are different, the inability to distinguish the correct hypothesis from competing hypotheses has far-reaching consequences. If even an infinite amount of data cannot guide us in determining the appropriateness of our conclusions, then we are scientifically bankrupt. One solution is to address only those scientific hypotheses that can be distinguished given the type of information that is realistically available (for a more detailed discussion of this idea, see Lele and McCulloch, 2002). Alternatively, the scientist is left to either arbitrarily choose a hypothesis from the set of indistinguishable hypotheses, or choose one on the basis of assumptions that cannot be demonstrated. This seems rather dissatisfying.

In statistics, situations that lead to indistinguishable hypotheses occur due to "nonidentifiability of the models." A classic example is that of factor analysis, where it has been shown that data can be rotated to obtain any factor loadings desired by the research scientist (Kowalski, 1972). No amount of data can reveal the set of factor loadings that is the most appropriate. Practitioners of factor analysis routinely impose conditions on the factor loading such as Varimax (SPSS, Inc., 1998). These side conditions are purely a choice of the experimenter, and the conclusions are a function of this choice. Though it seems more trouble than it is worth, one could conceivably decide on a final answer and then choose an appropriate side condition (superimposition scheme) that supports a favored conclusion!

This example demonstrates that the property of nonidentifiabilty is not unique to morphometrics. It also illustrates the problem associated with the choice of superimposition schemes. It can be shown in a mathematically precise fashion (Lele and Richtsmeier, 2001; Lele and McCulloch, 2002) that no amount of data can ever guide us in choosing between different superimposition techniques. The choice of minimizing the sum of squared distances (least squares fitting), minimizing the sum of the distances (robust fitting), or matching a prespecified edge (Bookstein's edge matching, roentgenographic cephalometry) is as arbitrary as the choice of rotation parameters in factor analysis (e.g., Varimax, Quartimax, Parsimax). Conclusions drawn from superimposition schemes are, therefore, a function of the choice of the superimposition criterion, and this may or may not correspond with the data.

Despite these drawbacks, many researchers find superimposition methods useful, in part because they produce clear graphic output. The adoption of a coordinate system and the placement of all forms into this coordinate system allow form or shape differences to be illustrated as absolute displacements of landmarks. Programs employing superimposition methods are freely available at the SUNY Stony Brook Morphometrics website, http://life.bio.sunysb.du/morph/.

Deformation methods

Deformation methods take the area or volume of a reference form and deform it to correspond with that of the target form. Sir D'Arcy Thompson's work (Thompson, 1992) is the earliest best-known example of the use of a deformation technique for the demonstration of the difference between forms. Thompson traces the application of the principle of coordinates to the study of proportion by Dürer (1613; as cited in Thompson,1992). D'Arcy Thompson created "transformation grids" where an orthogonal, two-dimensional grid was placed over one form, and the grid was transformed to correspond to the morphology of the second form. The change in the grid described the difference in forms. Regional changes in form were relatively simple to interpret from the graphic product of the transformed grid. Unfortunately, Thompson's work referred more to outlines than to landmarks, and he did not propose any quantitative method for creating these grids.

The main idea behind deformation approaches is straightforward:

1) Choose one of the forms as the reference object.
2) Deform the target object so that after deformation it matches the reference object exactly.
3) Study the deformation to learn about the difference in the forms.

Deformation techniques that have been used to study morphological differences between forms under the label of geometric morphometric methods include finite-element scaling analysis (Cheverud et al., 1983), adopted for biological applications from the field of engineering, and thin-plate splines (Bookstein, 1989), developed originally in the field of approximations theory. In finite-element scaling analysis, the scientist is required to subdivide the landmarks located on an object into groups that

form elements. A homology function (Lewis et al., 1980) maps the location of landmarks from the initial to the target form, and maps the location of *all* mathematically homologous points internal to each finite element in the initial form to a corresponding location in the target form. The homology function used in finite-element scaling analysis is rather general, though in theory this function can be designed to model material properties or other relevant features of the organism.

Thin-plate splines is a deformation technique that uses chosen functions to map the relative location of points in the initial configuration to their corresponding locations in the target form exactly. The functions are also used to predict how points that lie in those areas between landmarks in the initial form are arranged on the target form. The function that is chosen to obtain this map is designed to satisfy a particular smoothness criterion, such that a given quantity (e.g., bending energy) is minimized. To imagine this process, think of placing a continuous and bendable surface (or plate) over the area or volume encompassed by the landmarks. This plate is then deformed in such a way that: 1) corresponding landmarks in the two objects are mapped to one another exactly; and 2) the quantity of a specific parameter, often bending energy, within the function is minimized. This means that only the minimal amount of energy required to bend the plate to conform to the target object is used.

A closer look at deformation approaches reveals that they possess some of the same unfortunate characteristics as superimposition techniques. In finite-element scaling analysis, the homology function determines the deformation obtained and analytical output changes when the element design is changed (Richtsmeier et al., 1990). In thin-plate splines, the mapping of points from initial (one coordinate system) to final configuration (another coordinate system) depends on the interpolation function used, and the mapping of those points that lie between landmarks depends on the nature of this function. If the function changes (i.e., something other than bending energy is minimized), the deformation or bending of the plate is done differently. This results in varying graphical output for the same comparison using differing interpolation functions (Fig. 8).

Figure 8 shows the different results obtained when different deformations are used to compare the same two forms defined by a set of landmarks. Since the choice of a particular deformation affects the results obtained, we need to determine whether the data can help us choose between competing deformations. The answer is the same as it was in the case of superimposition: no amount of landmark data can lead us to choose between alternative deformations. Consequently, the scientific conclusions drawn from the graphics obtained by deformation approaches can be due more to the choice of the deformation than to the information in the data.

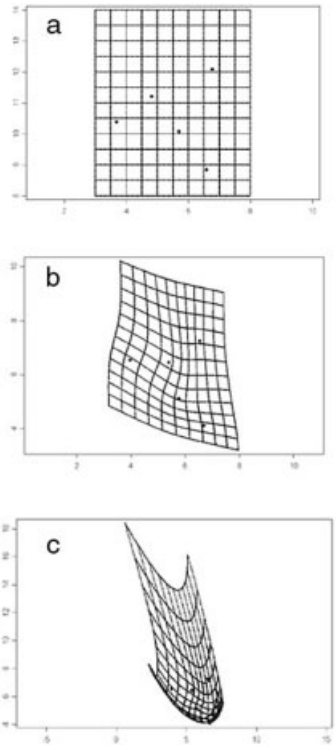

Fig. 8. Deformation of a group of landmarks, using the thin-plate spline. Coordinates for landmarks of the initial configuration and target configuration were taken from Bookstein (1991, p. 320). **a:** Original square grid shows arrangement of landmarks before any deformation. **b:** First deformation shows effect obtained when minimum bending energy is used as the interpolation function (Bookstein, 1991). **c:** Second deformation shows the effect obtained when the approach of Rohlf (1993a) is used, that implements weights inversely proportional to squared distances between landmarks. Scientific inference would be different for two deformations that describe the difference between identical data sets.

The drawbacks of these assumptions notwithstanding, many researchers take advantage of the attractive graphic output of these methods. Some of the more elaborate programs were developed for commercial profit and are prohibitively expensive, but software utilizing deformation methods is freely available on the Internet. Programs implementing the thin-plate spline method are available at http://life.bio.sunysb.edu/morph/. A finite-element scaling analysis program (FIESCA) is available at http://oshima.anthro.psu.edu.

Linear distance-based methods

Linear distance-based methods compare linear distances that connect landmark pairs in one form with the corresponding linear distances in another form, and provide information pertaining to the difference in length of these linear distances. By comparing linear distances rather than landmark coordinate data, these methods require no a priori assumption: no rule of superimposition; no discretization of the form into smaller units that comprise a finite element model; and no adoption of arbitrary rules such as minimum bending energy used with

the thin-plate spline or the homology function in finite-element scaling analysis. This characteristic makes linear distance-based methods preferable in our estimation.

The main idea behind this approach is simple:

1) For each specimen, rewrite the landmark coordinate matrix as a matrix of the linear distances between all unique pairs of landmarks (the "form matrix"). There is a one-to-one correspondence between a form matrix and the orbit defined by a form. For morphometric analyses using samples, a mean form matrix is estimated for each sample. Estimating the mean form matrix is the same as estimating the orbit to which the mean form belongs. The mean form matrix remains invariant under any translation, rotation, or reflection of the object.
2) Compare each like-element (linear distance) of the form matrices as a ratio, or an absolute difference or some other metric.
3) Study the matrix of linear distance comparisons to determine the difference in the forms and the difference for specific linear distances.

Examples of linear distance-based methods are Euclidean distance matrix analysis (or EDMA, pronounced ed·ma, Lele and Richtsmeier, 2001) and its variations as suggested by Rao and Suryawanshi (1996). EDMA (referred to as "the temptingly simple method of cross ratios" by Bookstein, 1991) was introduced relatively recently, but represents an old idea: measuring all possible chords between landmarks just as if these linear distances had been painstakingly recorded using calipers. The analysis is simple and not radically different from traditional morphometric approaches. The important distinction is that results from EDMA are based on what can be known about the *form* of an object and about the difference in form between objects, and takes into account the nonestimability of rotation and translation parameters. The descriptors used in EDMA are invariant to the identification of the particular elements of the two orbits corresponding to the two forms. In other words, the form matrix does not change, no matter *where* the form is or *how* it is oriented. However, many discussions of the drawbacks of EDMA point out that it does not provide the alluring graphics available from other morphometric techniques. These critics do not understand that these graphic displays require adoption of one or more a priori assumptions, thereby influencing the results (or the display of results).

Nonparametric statistical methods developed for EDMA provide summary statistics of differences between forms as well as confidence intervals for individual linear distances. The influence of individual landmarks is contained in the linear distance output. Software for automated graphical output is under development and will be available with the WinEDMA software (©Theodore M. Cole III (Cole, 2002); http://oshima.anthro.psu.edu/). Currently the graphical method used to study the analytical output is to draw lines of different colors, weights, or patterns to depict the various magnitudes of differences for specific linear distances (e.g., Aldridge et al., 2002; DeLeon et al., 2001) and the adoption of more standard data displays (Cole and Richtsmeier, 1998).

It has also been noted that the information contained within the form matrix is redundant, because all chords between all possible landmarks can be calculated if only a small number of chords are known. In short, a form can be reconstructed from a subset of the complete catalogue of linear distances, so it seems that a reduction in linear distances might streamline EDMA. But how does a researcher make the choice of those linear distances to exclude? The arbitrary choice of a subset of linear distances could accentuate the influence of certain linear distances in the comparison of forms, while masking the influence of others. The data do not provide any indication of the subset of linear distances that would be most informative in analysis. This issue is discussed more thoroughly by Lele (1991).

In summary, our discussion of these major classes of morphometric approaches is based on limitations of what can be known about form and form difference, given landmark data. When there are infinite choices for the orientation of the objects being compared, the parameters of rotation and translation are mathematically unknown and unknowable. The information that can be known about a set of forms represented by landmark data is limited. Superimposition and deformation approaches provide persuasive graphical output, but each analysis represents only one of an infinite number of possible interpolations of the difference between forms. No amount of data can tell us which one of these interpolations is the "true" one. The lack of elaborate graphical output makes EDMA seem less satisfying, but the approach is purposefully limited to those classes of information that *can* be known from landmark data, and does not include unverifiable assumptions.

Analysis of data by various morphometric methods (or, "one fish, two fish, old fish, new fish")

We provide here sample applications of some of the landmark-based morphometric methods discussed above. These methods, representing superimposition, deformation, and linear distance-based approaches, are applied to the same data sets, in an attempt to demonstrate the relative merit of these approaches. We begin by providing the rationale and research design for this study. Next, we introduce and briefly describe the analytical programs used. Finally, we present the results of the various analyses of the same data sets, and discuss the positive and negative aspects of each approach.

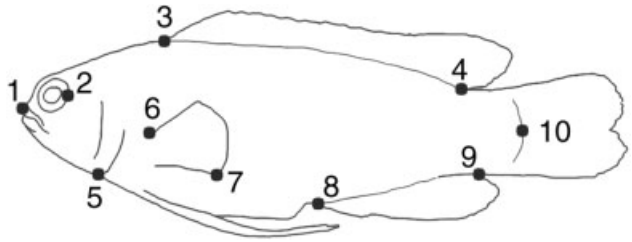

Fig. 9. This two-dimensional biological form, *Gramma loreto*, was used as the basis for hypothetical data sets. Ten biologically relevant landmarks were identified, and coordinate landmark data were collected. They are: 1, mouth; 2, eye; 3, anterior dorsal fin; 4, caudal dorsal fin; 5, operculum; 6, base of pectoral fin; 7, free edge of pectoral fin; 8, anterior ventral fin; 9, caudal ventral fin; and 10, tailfin.

Research design

Rationale. There are now many nontraditional morphometric methods that can be used to quantitatively describe the difference in form between two objects or two samples of objects. A number of morphometric studies provide analyses of data sets using more than one method, in an attempt to see whether the methods provide consistent results (e.g., Richtsmeier and Lele, 1990; Atchley et al., 1992; Corruccini, 1987), or to take advantage of what seem to be the best features of various approaches (e.g., Lague, 2002). But if different methods yield different results, how can we assess the relative merits of each method? How do we know which answer(s) is valid? In order to evaluate the relative strengths and weaknesses of morphometric methods, we need to apply them to a set of data where specific differences in morphology are known a priori.

Here we use the morphometric methods described above to analyze an artificial data set based on a biological form with 10 landmarks (Fig. 9). We modified this original form to create three artificial forms representing samples that differ from the original one in specific ways. Why do we use simulated data to evaluate methods, while at the same time stressing the importance of science when applying morphometric methods? Simulated data are used in order to test whether the methods under consideration provide valid estimates of the parameters and accurate descriptions of form difference. Only when methods are shown to be valid and are able to discern specific, defined modifications in form, can the usefulness of the method for a particular biological problem be evaluated appropriately (e.g., determining which method has the most statistical power or the most informative graphical displays). In our opinion, if the method does not pass the first test (i.e., if the method cannot discern defined modifications in form), then there is little reason to determine whether the format in which these answers are given is "useful," "interpretable," or "informative." Moreover, arguments about the power and robustness of statistical tests developed for certain methods are pointless if the estimated parameters are not valid.

Generation of forms. Three artificial forms were created from an initial true form, each representing specific, defined, and biologically plausible changes in regional morphology. The initial form (Form 1) was digitized from a two-dimensional photograph of the fish *Gramma loreto*. Ten landmarks were chosen on the original form (Form 1) to create a region of high landmark density (Landmarks 1–3 and 5–7), and a separate region of low landmark density (Landmarks 4 and 8–10; Fig. 9). Form 2 was generated by altering Form 1 to represent localized change in a region of high landmark density (Fig. 10). Form 3 was generated by altering Form 1 to represent localized change in a region of low landmark density (Fig. 12). Form 4 was generated by altering Form 1 to create a generalized, overall change in morphology, including a slight increase in size (Fig. 13). Three comparisons were made, Forms 2–4 being compared to the initial Form 1, using each of the morphometric methods described below.

We have defined specific anatomical differences between these forms, and refer to them as "true" differences. We describe these differences in terms of the displacement of specific landmarks in specific directions. However, in most biological applications, one form does not "morph" into another form; the forms are simply different, and we aim to accurately describe this difference. Consequently, the coordinate system that is internal to one form has no true relationship to the internal coordinate system of the other form. But in adopting landmark data, we can only describe the displacement of landmarks *relative to one another*. From the relative displacements, we can attempt to discern the underlying anatomical differences. In this study, by using artificial data sets, we define the anatomical differences first, and can therefore specify the "true" landmark displacements. We use various morphometric methods to discern these displacements.

Generation of landmark data. Two-dimensional coordinate landmark data were collected from Form 1 and from Forms 2–4 using the program Scion Image (©1998 Scion Corp.). Ten biologically relevant landmarks were chosen (Fig. 9). The data sets were organized and put into appropriate format for each of the morphometric programs utilized.

Methods

Superimposition approaches. We used two methods of Procrustean superimposition in this study. The first of these, a generalized least-squares algorithm, calculates transformations and minimizes the sum of the squared differences between corresponding landmarks on the forms being superimposed (Boas, 1905; Chapman, 1990; Sneath, 1967). The second method, a generalized resistant-fit algorithm (Chapman, 1990; Rohlf and Slice, 1990; Siegel

and Benson, 1982), uses repeated medians to calculate the transformations, and attributes differences to a small number of landmarks instead of spreading the difference over the whole object as in the least-squares approach. Both of the superimposition methods utilized, least-squares and resistant-fit, were applications of the RFTRA package of morphometric software by Ralph Chapman (©1989 Smithsonian Institution).

Deformation approaches. Finite-element scaling analysis (FESA) and thin-plate splines (TPS) were chosen to illustrate the use of two deformation-based methods. FESA was performed using the FIESCA (version 3.1) software (Morris, 1989; available for download at http://oshima.anthro.psu.edu). We used two distinct finite-element designs to compare the Reference and Target forms, in order to illustrate the effect that the finite-element design has on the results of FESA. In the first model, elements consisted of seven triangles and one quadrilateral, and each element represented a functional region. In the second model, elements consisted of 10 triangles and represented alternative functional regions. TPS was performed using Splus© software (Splus, 2000). As discussed previously, a model for the bending properties of the thin-plate spline must be chosen. We adopted the minimum bending energy model.

Linear distance-based approaches. EDMA form comparisons were done using WinEDMA software (©2001 T.M. Cole, III; available for download from http://oshima.anthro.psu.edu). We used two metrics in EDMA for comparing forms: relative differences and arithmetic differences. EDMA (relative differences) (Lele and Richtsmeier, 2001) calculates mean forms for each comparison and produces form difference matrices, each element of which is a ratio with a linear distance of Form 1 in the numerator and the same linear distance from the other form in the denominator. Ratios are reported for every interlandmark distance.

EDMA (arithmetic differences) (Lele and Richtsmeier, 2001) also calculates mean forms for each comparison. In contrast to EDMA (relative differences), however, it produces arithmetic difference matrices, each element of which is the arithmetic difference between a linear distance of Form 1 and the homologous linear distance from the relevant comparative form (e.g., Form 2). Differences are reported for every interlandmark distance.

Results

For all comparisons made in this study, each method produced output according to the convention for that method and the available software. The RFTRA package of morphometric software produced a graphic output for both least-squares and resistant-fit superimposition, with lines showing displacement of each landmark in the Target form relative to its original position in the Reference form. FIESCA (FESA deformation) produced both numeric (not reproduced here) and graphic output, with ellipses at each node (landmark) showing the degree of "compression" or "stretching" at that node. Splus (TPS algorithm) produced a graphic output consisting of a deformed grid and landmark locations representing the Target form. WinEDMA (EDMA linear-distance matrices) produced a numeric data output (not reproduced here) that was manually transformed into a graphic display. In the following examples, we illustrated the lines that represent the more salient features of the differences between forms (see Fig. 11 for illustration criteria).

Form 1 → Form 2 comparison. The actual form change generated here was the caudal displacement of the ventral operculum (Landmark 5) and the dorsal displacement of the pectoral fin base (Landmark 6). This represents a biologically plausible form difference that might be related to adaptive changes in the pectoral girdle of the fish. The two fishes are identical in form, except for the changes in the pectoral girdle evidenced by changes local to these two landmarks.

The results of this comparison, the true changes, and the graphic display of results from each method are shown in Figure 10. The generalized least-squares (GLS) and generalized resistant-fit (GRF) superimposition algorithms both discerned the true form change accurately. In this example, there is no change at 8 of 10 landmarks, meaning that these 8 landmarks can be superimposed exactly. When the changing landmarks are located in a region of high landmark density and only two landmarks are displaced, these superimposition methods produce a graphic output that is very easy to interpret and correctly displays the true form difference.

Both FESA models showed stretching at six nodes (Landmarks 1– 3 and 5–7), although only two nodes (Landmarks 5 and 6) were actually displaced. This discrepancy results from the generalization of local deformation to all nodes connected to the element that contains Landmarks 5 and 6, whether or not they have been displaced. The results of FESA Model 1 correctly show a dorso-ventral stretching of the element bordered by the operculum and the base and free edge of the pectoral fin (Landmarks 5, 6, and 7, respectively), and a dorso-ventral compression of the region between the cranial edge of the dorsal fin (Landmark 3) and base of the pectoral fin (Landmark 6). The combination of this information might be correctly interpreted as the dorsal displacement of the base of the pectoral fin, but there is no unique interpretation for this result. Whatever the interpretation, the magnitude of displacement of the base of the pectoral fin obscures the caudal displacement of the operculum (Landmark 5), which shares an element edge with the base of the pectoral fin.

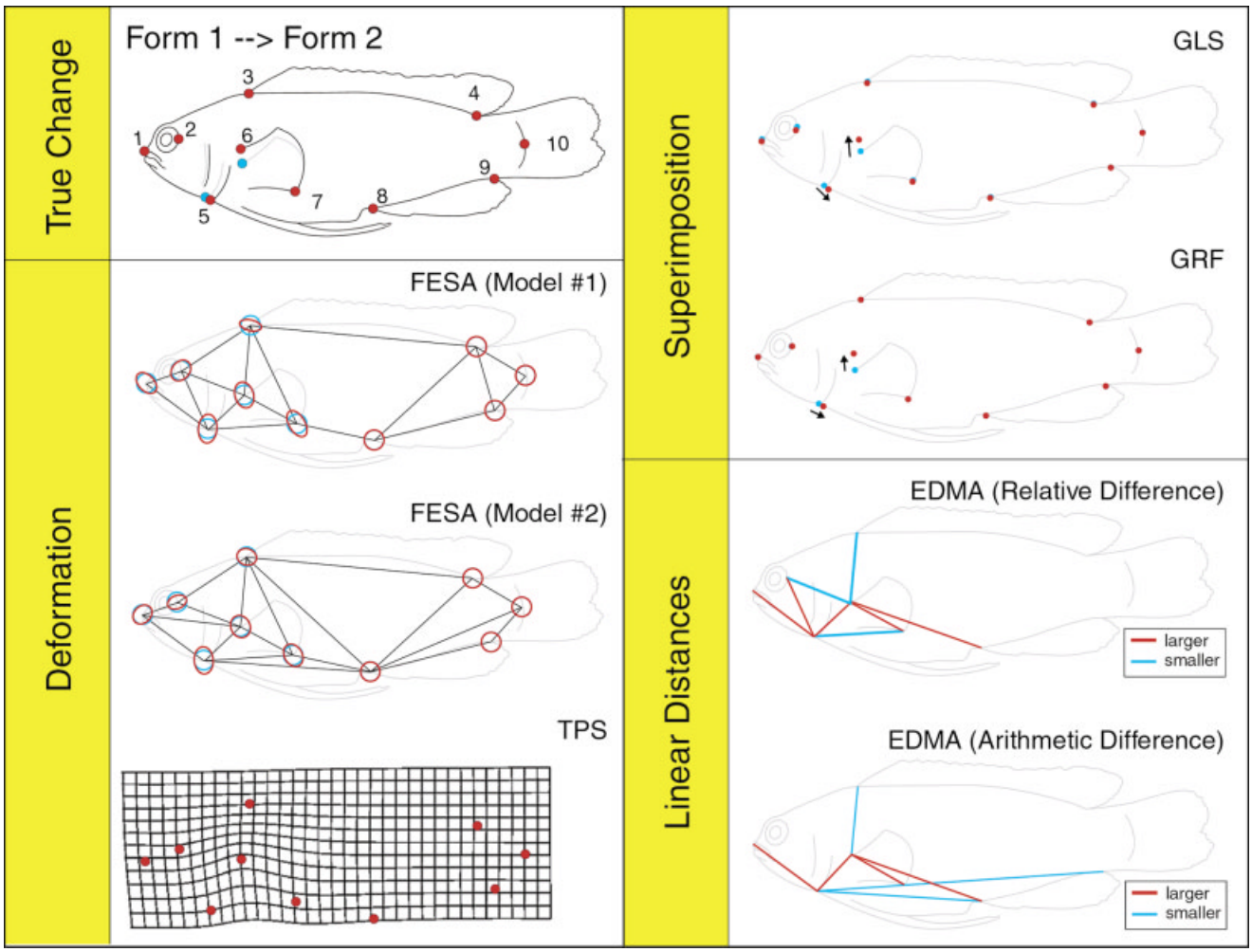

Fig. 10. Comparison of Form 1 to Form 2 by deformation, superimposition, and linear distance-based methods. True change, shown in upper left, involves dorsal displacement of pectoral fin base (Landmark 6) and caudal displacement of operculum (Landmark 5). Reference form (Form 1) is shown in gray, with original landmark locations in blue, while target form (Form 2) is indicated in black, with landmarks shown in red. For each method, we modified graphic output of the computer software by adding fish outline and color for clarity. Arrows were added to superimposition outputs to clarify direction of landmark displacements. EDMA results are illustrated manually, and in this case represent linear distances that are more than 5% different (for EDMA (relative differences)) or more than 2 mm different (for EDMA (arithmetic differences)). Red lines show those linear distances that are relatively larger in target form, while blue lines indicate those distances that are relatively smaller in target form.

FESA Model 2 highlights dorso-ventral compression at the eye (Landmark 2) and dorso-ventral stretching at the ventral operculum (Landmark 5). These results indicate a dorsal displacement of one or both of the landmarks at the mouth (Landmark 1) and the pectoral fin base (Landmark 6). However, this model was unable to localize the displacement specifically to the pectoral fin base (Landmark 6).

TPS results correctly highlight the dorsal displacement of the pectoral fin base (Landmark 6) and indicate a caudal displacement of the operculum (Landmark 5). The anterior portion of the grid is also skewed slightly in an inferior and caudal orientation, suggesting an inferior and caudal rotation of the skull (Landmarks 1, 2, and 5), which is not seen in the true form change.

EDMA (Relative Difference) results correctly suggest that displacement occurs primarily at the operculum (Landmark 5) and the pectoral fin base (Landmark 6). These results demonstrate an important detail in interpreting EDMA results. One must note whether linear distances that are notably different between forms all share a common landmark as an endpoint, because this situation indicates that the landmark common to all these linear distances has been displaced, while the other landmarks remain stable relative to one another. The quantitative results of this analysis generally indicated that Forms 1 and 2 are of the same size, as only linear distances that were more than 5% different are shown in Figure 10.

EDMA (Arithmetic Difference) results were similar to those from EDMA (Relative Difference). They indicated the dorsal displacement of the pectoral fin base (Landmark 6) and a caudal displacement of the operculum (Landmark 5). The quantitative results

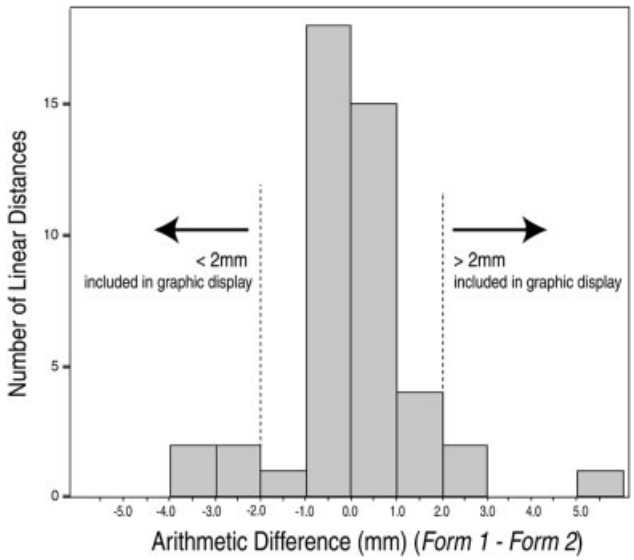

Fig. 11. Density distribution of arithmetic differences between Form 1 and Form 2. WinEDMA output provides an arithmetic difference for every possible interlandmark linear distance. Confidence intervals are available to denote distances that are significantly different between forms, but when sample size prohibits statistical testing, the researcher must arbitrarily choose a threshold for extreme values to include in a graphic display of differences in form. In this case, a threshold of 2-mm difference between like linear distances was used. When choosing an appropriate threshold, the researcher may be guided by natural breaks in the density distribution.

of the EDMA (Arithmetic Difference) analysis also indicated that Forms 1 and 2 were generally of the same size, as only linear distances that were more than 2 mm different are shown in Figure 10. This provides a good example of the manual interpretation currently required for graphic display of EDMA output. WinEDMA produces a matrix of differences for every single interlandmark linear distance, and the researcher must choose a threshold for purposes of illustration. Usually, the threshold selected represents a natural break in the data (Fig. 11).

Form 1 → Form 3 comparison. The actual form change generated here was a cranio-dorsal displacement of both the caudal edge of the dorsal fin (Landmark 4) and the tailfin (Landmark 10). Landmarks 4 and 10 are related to the positioning of the dorsal caudal vertebrae. A form difference of this sort might be found where the only developmental inconsistency among forms is an alteration in these caudal vertebrae. The results of this comparison for each of the methods utilized are presented in Figure 12.

GLS superimposition did not accurately show the form changes occurring in the tail region. The superimposition methods used here include a step in which each specimen is scaled, in an attempt to standardize by size and identify differences in form that are not simply due to differences in scale. Consequently, the results are intended to be interpreted as differences in *shape*. In this comparison, shape changes in the tail were distributed across four of the caudal landmarks. While this result indicates that the "shape" of the ventral fin relative to the other landmarks has changed, an interpretation of these results describing actual changes local to landmarks in the ventral fin would be inaccurate.

In contrast, GRF superimposition displayed the landmark displacements accurately. This particular type of form change, where a small subset of landmarks is displaced and the remainder of the form remains unchanged, is the type of change most suited for GRF analysis. The problem is that we rarely know beforehand that this is the "type" of form change that has occurred, and therefore do not have the necessary information to make this choice intelligently.

FESA Model 1 results correctly showed stretching in the most caudal element along a dorso-caudal to ventro-cranial axis, and compression on a more cranio-caudal axis. A similar pattern is seen in the next most caudal element (bounded by the cranial and caudal points on the ventral fin and the caudal edge of the dorsal fin; Landmarks 8, 9, and 4, respectively). However, it is impossible to determine which of the landmarks are actually displaced from the FESA results, as the deformation local to Landmarks 4 and 10 is generalized to the whole element and to a landmark in an adjoining element (Landmark 8). FESA Model 2 shows a similar effect at the caudal point on the ventral fin (Landmark 9), where there was no change, with less pronounced effects at the other nodes. These results correctly suggest the anterior displacement of the tailfin (Landmark 10). In contrast, the deformation apparent at the caudal ventral fin (Landmark 9) would likely be interpreted as ventro-caudal displacement, although this landmark did not move in the true form change.

TPS correctly indicates the difference in shape as consisting primarily of a cranial and slightly dorsal displacement of the caudal dorsal fin and the tailfin (Landmarks 4 and 10, respectively). However, these results also suggest an inferior displacement of the caudal ventral fin (Landmark 9); an expansion between the cranial and caudal edges of the ventral fin (Landmarks 8 and 9) is also suggested. As with the superimposition methods, these results indicate *shape* differences, and must be interpreted and described as differences in shape rather than form. Descriptions of actual differences in ventral fin morphology would be incorrect in this case.

Results of both EDMA analyses correctly indicate that the caudal edge of the dorsal fin and the tailfin are displaced anteriorly. This example again underscores the need to look for linear distances that share a common landmark as an endpoint to enable valid interpretation of results. Every linear distance shown has as an endpoint on either the caudal dorsal fin or the tailfin (Landmarks 4 and 10, respectively), indicating that those landmarks have been displaced. Careless interpretation of these results might identify an overall shortening of the specimen on an cranio-caudal axis. Consideration of the quantitative results of these analyses again indicates

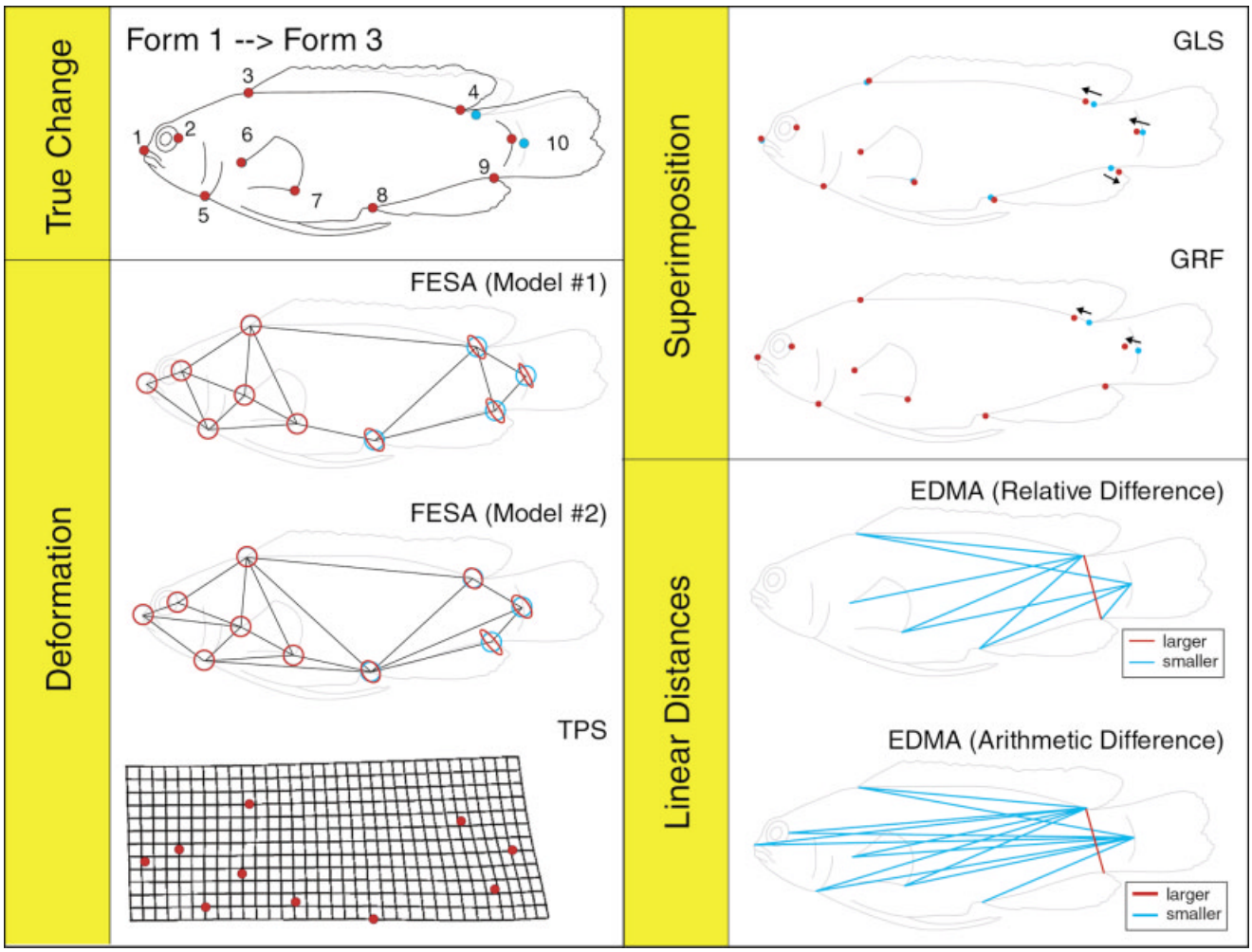

Fig. 12. Comparison of Form 1 to Form 3 by deformation, superimposition, and linear distance-based methods. True change, shown in upper left, primarily involves cranial displacement of posterior dorsal fin and tailfin (Landmarks 4 and 10, respectively). Reference form (Form 1) is shown in gray, with original landmark locations in blue, while target form (Form 3) is indicated in black, with landmarks shown in red. For each method, we modified graphic output of computer software by adding fish outline and color for clarity. Arrows were added to superimposition outputs, to make clear the direction of landmark displacements. EDMA results represent linear distances that are more than 5% different (for EDMA (relative differences)) or more than 3 mm different (for EDMA (arithmetic differences)). Red lines show those linear distances that are relatively larger in target form, while blue lines indicate those distances that are relatively smaller in target form.

that most interlandmark linear distances remain unchanged between Form 1 and Form 3. Only those linear distances that were more than 5% different (for EDMA (Relative Differences)) or more than 3 mm different (for EDMA (Arithmetic Differences)) are shown in Figure 12.

Form 1 → Form 4 comparison. The true form change generated here a priori was a generalized, overall expansion of most landmarks away from the center, including in particular the ventral displacement of the cranial point on the ventral fin (Landmark 8) and caudal displacement of the caudal point on the dorsal fin (Landmark 4) and tailfin (Landmark 10). This difference in form might arise from a general increase in size, accompanied by related, positively allometric increases in pectoral fin ray length, rib length, and caudal vertebrae depth.

We note here again that in most biological applications, landmark displacement can only be described relative to the other landmarks. If Form 1 and Form 4 lay side-by-side on the counter at the fish market, could we say whether landmarks 4 and 10 were caudally displaced in Form 4, or whether all the other landmarks were anteriorly displaced? Probably not, because based on the landmark data, we can only describe relative displacement. But for our hypothetical forms, we create discrete morphological changes while holding the rest of the form constant. Therefore, we know and can therefore describe absolute displacement of specific landmarks in specific directions. In this case, we designed the experiment so that the caudal vertebrae are deeper in Form 4, and Landmarks 4 and 10 are caudally displaced.

The results of this comparison using the various morphometric methods are shown in Figure 13. GLS and GRF superimposition algorithms produced similar results. Both superimposition methods gave estimates of the true displacement of cranially located landmarks in the area of high landmark density that were in agreement with the true change. However, they both underestimated the caudal displacement of the tailfin and caudal dorsal fin (Landmarks 10 and 4, respectively), and showed a cranial displacement of the caudal edge of the ventral fin (Landmark 9) when the true displacement of this landmark was caudal and ventral. These results are affected by the scaling factor employed by these methods to "shrink" Form 4 to fit Form 1, and the apparent differences must be interpreted as changes in *shape*.

FESA results suggest that most of the deformation between Forms 1 and 4 occurred in the tail region, which is accurate. In FESA Model 1, the ellipses indicated that principal stretching in the tail region was on a dorso-caudal to ventro-cranial axis. In contrast, FESA Model 2 shows little deformation local to the caudal dorsal fin (Landmark 4), which could be interpreted as indicating little or no change at that landmark. It also shows exaggerated compression local to the caudal ventral fin (Landmark 9). This may be due to the ventral and caudal displacement of the other two nodes of that element in Model 2 (Landmarks 8 and 10). Similarly, note the different orientations of the ellipses in Models 1 and 2 at the free edge of the pectoral fin (Landmark 7). The deformation reflected in the ellipse at any node is computed and displayed as a composite effect from all bordering elements. These examples demonstrate the impact of element design on FESA results.

TPS results correctly indicate an overall increase in size from Form 1 to Form 2, and suggest a craniocaudal stretching. Anterior displacements of the mouth (Landmark 1) and particularly the operculum (Landmark 5) are indicated. The caudal dorsal fin and the tailfin (Landmarks 4 and 10, respectively) are displaced caudally, the cranial edge of the ventral fin (Landmark 8) is displaced inferiorly, and the grid suggests compression between the cranial and caudal ventral fin (Landmarks 8 and 9, respectively). These results also correctly show an increase in the size of the triangle bounded by the operculum and the pectoral fin (Landmarks 5–7).

EDMA (Relative Differences) results correctly indicate that Form 4 was generally larger than Form 1, because most interlandmark distance ratios for this comparison are less than one (i.e., most linear distances were smaller in Form 1 than in Form 4). Ratios that are much less than one are shown in Figure 13 (indicating linear distances that are over 5% larger in Form 4). Other linear distances had ratios closer to one, indicating that there was not as much difference in those linear distances. On careful inspection of these results, a number of patterns are discernible. First, the caudal dorsal fin and the tailfin (Landmarks 4 and 10, respectively) remain equidistant, and yet together move caudally (relative to the other landmarks). Other predominant points of change include the ventro-caudal displacement of the free edge of the pectoral fin (Landmark 7) and the ventro-caudal displacement of the cranial edge of the ventral fin (Landmark 8).

The EDMA (Arithmetic Differences) analysis also demonstrated that Form 4 was generally larger than Form 1, because almost all linear distances were larger in Form 4. Only those linear distances that showed more than a 7-mm difference between Form 1 and Form 4 are shown in Figure 10. These results highlight the separation of the caudal dorsal fin, the tailfin, and the anterior ventral fin (Landmarks 4, 10, and 8, respectively) from more anterior landmarks (Landmarks 1–3 and 5). By looking at the pattern of differences between these landmarks and those around them in the quantitative results of the analysis, one may clarify the specifics of displacement. For example, there is little change (less than 1.5 mm) in the linear distances that define the triangle bounded by Landmarks 1, 2, and 5, indicating that the skull is being displaced as a unit (i.e., relations within the skull remain the same, but move anteriorly relative to the rest of the body).

Synopsis

When forms are different, all the methods presented here show that some difference exists, but to aid biological inquiry, methods should correctly localize and characterize these differences. GLS and GRF superimposition applications both produced clear, unambiguous graphic outputs automatically. They appear easy to interpret, but the results are shown in terms of absolute landmark displacements, and the researcher must limit interpretation to relative landmark displacement. In addition, results are affected by the choice of superimposition algorithm (GLS or GRF).

FESA is a deformation method and requires no predetermined rule of superimposition. However, as shown, deformation local to each landmark is influenced by measures of differences local to landmarks that are part of the same element or that are joined by a side shared between elements. In addition, changes in the finite-element model used can affect the results produced. TPS applications produce graphic output in the form of a deformation grid, which is pleasing to the eye and reminiscent of D'Arcy Thompson's deformation grids. However, like many of the other methods discussed here, TPS requires the researcher to make a priori assumptions about deformational changes, in our example by adopting the minimum bending energy rule.

Finally, EDMA provides a method of analysis that requires no a priori assumptions. The results of EDMA analysis are presented as matrices and simple graphs (although additional graphic programs are currently in progress). These results are not as

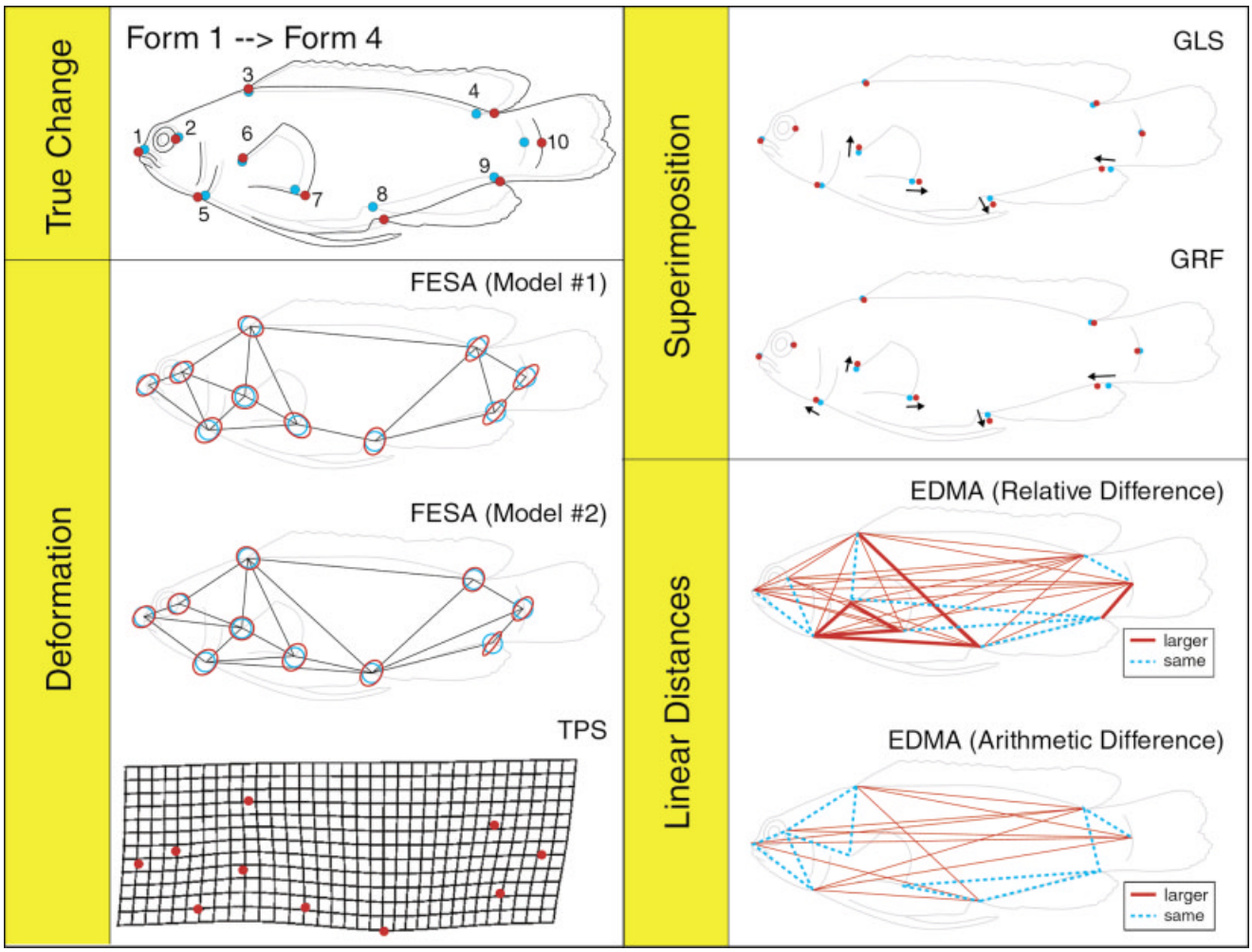

Fig. 13. Comparison of Form 1 to Form 4 by deformation, superimposition, and linear distance-based methods. True change, shown in upper left, involved generalized increase in size due to change at all landmarks. Reference form (Form 1) is shown in gray, with original landmark locations in blue, while target form (Form 4) is indicated in black, with landmarks shown in red. For each method, we modified graphic output of computer software by adding fish outline and color for clarity. Arrows were added to superimposition outputs to make clear the direction of landmark displacements. Linear distances shown in lightly weighted red lines indicate measures that were 5–10% larger in target form (for EDMA (relative differences)) or more than 7 mm larger in the target form (for EDMA (arithmetic differences)). Heavily weighted red lines indicate linear distances that are more than 10% larger in the target form. Since no linear distances were smaller in target form by more than 2% (for EDMA (relative differences)) or by 2 mm (for EDMA (arithmetic differences)), dashed blue lines are used to indicate distances that are similar in Forms 1 and 4. Similarities between forms are often as important as differences when attempting to understand processes at work.

clear as the graphic output of the superimposition methods or TPS, but they have the advantage of incontrovertibly conveying valid representations of change in form as characterized by landmark data.

Unless the local differences between forms are correctly characterized, morphometric methods offer little of use to biologists. Since knowledge of the biology of organisms can tie precise anatomical locations to information concerning developmental, evolutionary, and biomechanical processes, localization of form difference is critical. Though essential, the valid definition of local differences between forms is only the beginning. The design of additional experiments is required to provide an explanation of the biological basis of the results of any valid morphometric analysis. Morphometric methods can only define the difference between forms; they cannot explain these differences.

SUMMARY AND CONCLUSIONS

"The theme must be accurate and fruitful; it must not be twisted to encompass more than it can explain; it must not claim exclusive rights as a unifying approach." (Gould, 1966, p. 588.)

If we lived in a perfect world, there would be a single morphometric method for comparing forms using landmark data. That method would be able to localize differences using valid, uniformly most powerful statistical tests based on statistical models whose estimates are unbiased, consistent, and efficient, with great graphical abilities. Alas, we do not

live in a perfect world. Only certain aspects of these desired features are obtainable from methods currently available for studying form difference using landmark data.

Biology is messy, and biological forms are characterized by changeable degrees of variability. To attempt to understand the biological processes responsible for the variation we see in phenotypes, quantification is necessary. Landmark data have become a reliable, verifiable, and accepted means of recording and comparing two- and three-dimensional forms. However, these data are limited in several ways. For example, when landmark data are collected, no verifiable information regarding the surfaces that lie between the landmarks is retrievable from analysis of the data. The problems involved in quantification of biological form and change in form are complicated, and we need to realize that the data we collect to represent biological forms are limited.

In order to solve the problems associated with the proper study of form change, explicit and precise definitions of any terms that enter into the day-to-day vocabulary of a science are required. We have shown that no unique definitions exist for the terms size and shape, and so as currently used, the terms can be ambiguous. When forms differ greatly in size, they are commonly scaled to adjust for these differences so that the information relating to scale does not obscure other information intrinsic to the comparison. When forms are scaled, it is often said that shapes, rather than forms, are being compared. It has been argued that Procrustes methods were developed for comparison of shapes, while EDMA was developed to study differences in form, and accordingly the results of these two methods cannot be compared. Given what is known about size, shape, and nuisance parameters, is this point valid? We have defined the concept of orbit and established that the comparison of forms requires the comparison of orbits. Scaling adjusts the size of the form, but it does not change a "form" into a "shape." Scaling simply changes the units used to express the form. In terms of orbits, adjusting for differences in scale changes the characterization (extent) of the orbit, but it does not specify the location of any object within the orbit. Consequently, what we found for form comparison pertains equally to shape comparison: we can only reliably compare orbits. The problem of nonidentifiability and the related issues that were raised in the context of superimposition and deformation methods apply equally, whether one is comparing forms or shapes (scaled forms).

Nuisance parameters, as defined by Neyman and Scott (1948), affect our ability to make inferential statements about the form difference discovered by the application of morphometric methods. The nature of landmark data requires the specification of a coordinate system during data collection. We have shown that each form has its own unique but unknowable orientation with relation to the mean form and with relation to any other form or sample of forms. A coordinate system is required for collection of landmark data. Any coordinate system will suffice, however, and further statistical analysis should be invariant to this arbitrary choice. Since the relative orientation of forms is unknowable, our analysis should not depend on an arbitrary choice of orientation.

Deformation approaches provide different inferences based on the choice of the deformation function. Similarly, the inferences provided by superimposition methods change if the minimization criteria are modified. Unfortunately, even an infinite sample of landmark data cannot determine which deformation function or which superimposition scheme is valid. There is a way to avoid these choices, and that is simply to adopt one of the many methods that do not require these choices. These include EDMA, old-fashioned multivariate morphometrics, and other related methods. The results obtained do not provide enticing graphics, because arbitrary choices are required to provide graphical displays of the type provided by deformation and superimposition approaches. The linear distance-based methods utilize landmark data without specifying a coordinate system, and base their inferences only on the information that can be obtained from the data.

Although null hypothesis testing has become a hallmark of biological research, alternate statistical approaches are being proposed, and the field of biology should make use of these methods. For morphometric analyses, in addition to null hypothesis testing, we should be thinking in terms of *how much* form difference (effect size) and *where* the differences are (localization), and provide confidence intervals for that information. Only those hypotheses that can be distinguished on the basis of information available from landmark data should be considered.

The concept of morphometric spaces has caused some confusion for users. The critical issue to understand about the various spaces is the importance of choosing a metric that is appropriate to the space. This can be a difficult task for the uninitiated, and if you need guidance, get it; the literature holds examples of even the experienced falling into this trap. It is important to remember that most forms that we study actually exist in old-fashioned, three-dimensional Euclidean space, though there are relatively flat organisms (e.g., many plants, flatfishes, and flatworms) or parts of organisms (e.g., fly wings) that can be adequately described using two-dimensional data. Approaching problems in three dimensions is straightforward in EDMA and finite-element scaling analysis, but is a bit more complicated in the other approaches.

The goal of this paper was to discuss the current state of morphometrics. Our conclusion is that we have a long way to go before the promise of morphometrics is fulfilled. We focused our argument on what *can be known* about form and form difference, given landmark data. Discussions of statistical

power in current morphometric approaches are futile if the methods being discussed do not provide correct and verifiable answers. Biologists need to be mindful of the assumptions of their chosen method, and morphometricians must consider the fundamental purpose of studying biological form and form change when proposing new methods. The nature of morphological diversity, the production of diversity through development, and its evolution over time hold the great questions of modern biology. To be of any use in deciphering the major trends in morphological diversity, morphometrics must offer models and methods that enable precise definition of the parameters that bear directly on the production and nature of this diversity. Morphometric methods are simply tools to help define the difference between forms. The results of a valid morphometric analysis are not *the* answer, but should be used to design further studies that probe the processes working to produce the differences revealed.

ACKNOWLEDGMENTS

We thank Erin Lindsay for assisting in drafting Figures 1, 3, 5, and 7, and Kristina Aldridge for assistance with Figures 10, 12, and 13 and the cover illustration. Kristina Aldridge and Mary Silcox read various drafts of the manuscript and provided thoughtful comments. We thank Alan Walker for his constant support and urgings to communicate our findings and our methods to a larger audience, and John Fleagle for initiating our interest in writing this review. We thank Christopher Ruff, Robert Corruccini, Ralph Chapman, and an anonymous reviewer for their remarks that helped us to clarify certain aspects of our argument. References in this paper are meant to be a guide to the potential literature, and are not meant to be an exhaustive bibliography. An online bibliography and glossary of morphometrics as well can be found at http://life.bio.sunysb.edu/morph/. The original collaborative work by J.T.R. and S.R.L. that led to the development of EDMA, the ©WinEDMA software written by Theodore Cole III, and many of the topics developed in this paper were funded by NSF grant SBR-929083. The most current version of ©WinEDMA software and the user's guide can be found at Dr. Cole's website: http://c.faculty.umkc. edu/colet/. Step-by-step EDMA instructions for the analysis of example data sets are presented in Lele and Richtsmeier (2001). These data sets and the ©WinEDMA programs can be found at the Richtsmeier laboratory website: http://oshima.anthro.psu.edu/.

LITERATURE CITED

Agresti A. 1989. Categorical data analysis. New York: John Wiley and Sons.

Aldridge KA, Marsh JL, Gorier D, Richtsmeir JT. 2002. Central nervous system phenotypes in craniosynostosis. J Anat 201:31–39.

Atchley WR, Cowley D, Vogl C, McLellan T. 1992. Evolutionary divergence, shape change, and genetic correlation structure in the rodent mandible. Syst Biol 41:196–221.

Berger J. 1980. Statistical decision theory. New York: Springer-Verlag.

Boas F. 1905. The horizontal plane of the skull and the general problem of the comparison of variable forms. Science 21:862–863.

Bookstein F. 1978. The measurement of biological shape and shape change. New York: Springer-Verlag.

Bookstein FL. 1982. Foundations of morphometrics. Annu Rev Ecol Syst 13:451–470.

Bookstein F. 1986. Size and shape spaces for landmark data in two dimensions. Stat Sci 1:181–242.

Bookstein F. 1989. Principal warps: thin plate splines and the decomposition of deformations. IEEE Trans Pattern Anal 11: 567–585.

Bookstein F. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press.

Broadbent BS, Broadbent BJ, Golden W. 1975. Bolton standards of dentofacial developmental growth. St. Louis: Mosby.

Casella G, Berger RL. 1990. Statistical inference. Belmont, CA: Duxbury Press.

Chamberlain T. 1965. The method of multiple working hypotheses. Science 148:754–759.

Chapman R. 1990. Conventional Procrustes approaches. In: Rohlf FJ, Bookstein F, editors. Proceedings of the Michigan Morphometrics Workshop. Ann Arbor: University of Michigan Museum of Zoology. p 251–267.

Cheverud J, Lewis J, Bachrach W, Lew W. 1983. The measurement of form and variation in form: an application of three-dimensional quantitative morphology by finite-element methods. Am J Phys Anthropol 62:151–165.

Cole TM III. 1996. Historical note: early anthropological contributions to "geometric morphometrics." Am J Phys Anthropol 101:291–296.

Cole TM III. 2002. WinEDMA: Windows-based software for Euclidean distance matrix analysis. Kansas City: Department of Basic Medical Science, University of Missouri at Kansas City.

Cole TM III, Richtsmeier JT. 1998. A simple method for visualization of influential landmarks when using Euclidean distance matrix analysis. Am J Phys Anthropol 107:273–283.

Corner BD, Lele S, Richtsmeier JT . 1992. Measuring precision of three-dimensional landmark data. Quantitative Anthropol 3:347–359.

Corruccini R. 1987. Shape in morphometrics: comparative analyses. Am J Phys Anthropol 73:289–303.

Corruccini R. 1995. Of ratios and rationality. Am J Phys Anthropol 96:189–191.

DeLeon V, Zumpano M, Richtsmeier J. 2001. The effect of neurocranial surgery on basicranial morphology in isolated sagittal craniosynostosis. Cleft Palate Craniofac J 38:134–146.

Dryden I, Mardia K. 1998. Statistical shape analysis. Chichester: John Wiley and Sons.

Dürer A. 1613. Les quatres livres d'Albert Durer de la propotion des parties et pourtraicts des corps humains. Arnheim.

Godfrey L, Sutherland M. 1995. What's growth got to do with it? Process and product in the evolution of ontogeny. J Hum Evol 29:405–431.

Goodall C. 1991. Procrustes methods in the statistical analysis of shape. J R Stat Soc B 53:285–339.

Gould SJ. 1966. Allometry and size in ontogeny and phylogeny. Biol Rev 41:587–640.

Gould SJ. 1977. Ontogeny and phylogeny. Cambridge, MA: Harvard University Press.

Gould SJ. 1981. The mismeasure of man. New York: W.W. Norton.

Hildebolt C, Vannier M. 1988. 3-D measurement accuracy of skull surface landmarks. Am J Phys Anthropol 76:497–504.

Huxley J. 1932. Problems of relative growth. London: Methuen.

Jungers W, Falsetti A, Wall C. 1995. Shape, relative size, and size-adjustments in morphometrics. Yrbk Phys Anthropol 38: 137–161.

Kendall J. 1994. Shape manifolds, Procrustean metrics and complex projective spaces. Bull Lond Math Soc 16:81–121.
Kent J, Mardia K. 1997. Consistency of Procrustes estimators. J R Stat Soc B 59:281–290.
Kohn L, Cheverud J. 1992. Calibration, validation, and evaluation of scanning systems: anthropometric imaging system repeatability. In: Vannier MW, Yates RE, Whitestone JJ, editors. Electronic imaging of the Human Body Workshop. Dayton, OH: Crew System Ergonomics Information Center. p 114–123.
Kowalski CJ. 1972. A commentary on the use of multivariate statistical methods in anthropometric research. Am J Phys Anthropol 36:119–132.
Lague MR. 2002. Another look at shape variation in the distal femur of *Australopithecus afarensis*: implications for taxonomic and functional diversity at Hadar. J Hum Evol 42:609–626.
Lele S. 1991. Some comments on coordinate free and scale invariant methods in morphometrics. Am J Phys Anthropol 85:407–418.
Lele S. 1993. Euclidean distance matrix analysis (EDMA) of landmarks data: estimation of mean form and mean form difference. Math Geol 25:573–602.
Lele S. 1999. Invariance and morphometrics: a critical appraisal of statistical techniques for landmark data. In: Chaplain M, Singh G, McLachlan J, editors. On growth and form: spatiotemporal pattern formation in biology. Chichester: John Wiley and Sons. p 325–336.
Lele S, McCulloch C. 2002. Invariance, identifiability and morphometrics. J Am Stat Assoc 971:796–806.
Lele S, Richtsmeier JT. 1990. Statistical models in morphometrics—are they realistic? Syst Zool 39:60–69.
Lele S, Richtsmeier JT. 1991. Euclidean distance matrix analysis: a coordinate free approach to comparing biological shapes using landmark data. Am J Phys Anthropol 98:73–86.
Lele S, Richtsmeier J. 1995. Estimating confidence intervals for the comparison of forms. Am J Phys Anthropol 98:73–86.
Lele S, Richtsmeier J. 2001. An invariant approach to the statistical analysis of shapes. Boca Raton: Chapman & Hall/CRC.
Lestrel P. 1982. A Fourier analytic procedure to describe complex morphological shape. In: Dixon A, Sarnat B, editors. Factors and mechanisms influencing bone growth. Los Angeles: Alan R. Liss. p 393–409.
Lestrel P. 1989. Method for analyzing complex two-dimensional forms: elliptical Fourier functions. Am J Hum Biol 1:149–164.
Lewis J, Lew W, Zimmerman J. 1980. A nonhomogeneous anthropometric scaling method based on finite element principles. J Biomech 13:815–824.
Lohmann GP. 1983. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. Math Geol 15:659–672.
Lohmann GP, Schweitzer PN . 1990. On eigenshape analysis. In: Rohlf FJ, Bookstein FL, editors. Proceedings of the Michigan Morphometrics Workshop. Ann Arbor: University of Michigan Museum of Zoology. p 147–166.
MacLeod N. 1999. Generalizing and extending the eigenshape method of shape visualization and analysis. Paleobiology 25: 107–138.
MacLeod N, Rose K. 1993. Inferring locomotor behavior in Paleogene mammals via eigenshape analysis. Am J Sci 239:300–355.
Marcus LF, Corti M, Loy A, Naylor GJP, Slice D, editors. 1996. Advances in morphometrics. NATO ASI series. New York: Plenum Press.
Morris G. 1989. FIESCA: software for thE application of FESA in biological research. Baltimore: Department of Civil Engineering, Johns Hopkins University.
Mosimann J. 1979. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. J Am Stat Assoc 63:930–978.
Mosimann J, James F. 1979. New statistical methods for allometry with application to Florida red-winged blackbirds. Evolution 33:444–459.
Moyers RE, Bookstein FL. 1979. The inappropriateness of conventional cephalometrics. Am J Orthod 75:599–617.
Neyman J, Scott E. 1948. Consistent estimates based on partially consistent observations. Econometrica 16:1–32.
Oxnard C. 1978. One biologist's view of morphometrics. Annu Rev Ecol Syst 9:219–241.
Platt J. 1964. Strong inference. Science 146:347–353.
Popper K. 1959. The logic of scientific discovery. London: Hutchinson.
Rao C. 2000. A note on statistical analysis of shape through triangulation of landmarks. Proc Natl Acad Sci USA 97:2995–2998.
Rao C, Suryawanshi S. 1996. Statistical analysis of shape of objects based on landmark data. Proc Natl Acad Sci USA 93: 12132–12136.
Read D, Lestrel P. 1986. Comment on uses of homologous-point measures in systematics: a reply to Bookstein et al. Syst Zool 35:241–253.
Reyment R, Blackith R, Campbell N. 1984. Multivariate morphometrics. London: Academic Press.
Richtsmeier J. 1987. A comparative study of normal, Crouzon and Apert craniofacial morphology using finite-element scaling analysis. Am J Phys Anthropol 74:473–493.
Richtsmeier J, Cheverud J. 1986. Finite element scaling analysis of normal growth of the human craniofacial complex. J Craniofac Genet Dev Biol 6:289–323.
Richtsmeier J, Lele S. 1990. Analysis of craniofacial growth in Crouzon syndrome using landmark data. J Craniofac Genet Dev Biol 10:39–62.
Richtsmeier JT, Lele S. 1993. A coordinate-free approach to the analysis of growth patterns: models and theoretical considerations. Biol Rev 68:381–411.
Richtsmeier J, Morris G, Marsh J, Vannier M. 1990. The biological implications of varying element design in finite-element scaling analyses of growth. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 12. Philadelphia: IEEE. p 387–388.
Richtsmeier JT, Cheverud JM, Lele S. 1992. Advances in anthropological morphometrics. Annu Rev Anthropol 21:231–253.
Richtsmeier J, Paik C, Elfert P, Cole TM I, Dahlman H. 1995. Precision, repeatability and validation of the localization of cranial landmarks using computed tomography scans. Cleft Palate Craniofac J 32:217–227.
Rohlf F. 1993. Morphometric spaces, shape components and the effects of linear transformations. In: Marcos LF, Corti M, Loy A, Naylor G, Slice DE, editors. NATO Advanced Studies Institute on Morphometrics. New York: Plenum Press. p 117–129.
Rohlf F. 2000. On the use of shape spaces to compare morphometric methods. Hystrix Ital J Mammol 11:9–25.
Rohlf F. 2002. Consistency and bias in morphometric methods. Am J Phys Anthropol [Suppl] 34:133 [abstract].
Rohlf F, Marcus L. 1993. A revolution in morphometrics. TREE 8:129–132.
Rohlf F, Slice D. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool 39:40–59.
Siegel A, Benson R. 1982. A robust comparison of biological shapes. Biometrics 38:341–350.
Sneath PHA. 1967. Trend-surface analysis of transformation grids. J Zool Lond 151:65–122.
Splus. 2000. Statistics software. Seattle: MathSoft, Inc.
Sprent P. 1972. The mathematics of size and shape. Biometrics 28:23–37.
SPSS, Inc. 1998. SYSTAT Statistics. Chicago: SPSS, Inc.
Strauss R, Bookstein FL. 1982. The truss: body form reconstructions in morphometrics. Syst Zool 31:113–135.
Thompson D. 1992. On growth and form: the complete revised edition. New York: Dover.
Valeri C, Cole T III, Lele S, Richtsmeier J. 1998. Capturing data from three dimensional surfaces using fuzzy landmarks. Am J Phys Anthropol 107:113–124.
Walker J. 2001. Ability of geometric morphometric methods to estimate a known covariance matrix. Syst Biol 49:686–696.
Williams F, Richtsmeier J. 2002. Testing the efficacy of image analysis against digitized mandibular landmarks. Clin Anat (in press).

APPENDIX: TERMS AND CONCEPTS

Bias of an estimator: Bias is measured as the average difference between an estimator and the true value of the parameter that it tries to estimate, for finite samples. If this difference is nonzero, the estimator is biased.

Consistency of an estimator: An estimator is considered to be *consistent* if the estimator converges to the true value of the parameter as the sample size increases. Estimators that do not converge to the true value as the sample size increases are *inconsistent* estimators of the given parameter. It seems natural that as sample size increases, the estimation of certain population quantities should improve, becoming more and more representative of the true value. When this is not the case, it is due to inconsistency of the estimators. Consistency is generally considered an essential property of any estimator. Also, an estimator cannot be *efficient* if it is *inconsistent*.

Effect size and confidence intervals: In most practical situations, simple testing for the presence of an effect is not enough. An estimator of the magnitude of the effect (*effect size*) and the uncertainty associated with that estimator is necessary. Confidence intervals provide this information. This is one of the reasons why most statisticians prefer reporting confidence intervals for the difference in means, rather than simply testing whether or not the difference in the means is zero (Agresti, 1989).

Efficiency of an estimator: An estimator is considered *efficient* if it has the smallest (asymptotic) variance among all consistent estimators.

Euclidean space: For our purposes (but not technically), *Euclidean space* is ordinary two- or three-dimensional space and their higher-order analogues.

Maximum likelihood: This is the value of the parameter that makes the observed data most likely (for details, see Casella and Berger, 1990).

Method: A *method* is any technique used in estimating the parameters of a model (see below) and in further analysis such as hypothesis testing, pattern recognition, or calculation of confidence intervals.

Method of moments: This is the value of the parameter that equates the sample moments to the population moments (for details, see Casella and Berger, 1990).

Model: A *model*, as used in this paper, is a mathematical construct that attempts to characterize certain aspects of the underlying phenomena (e.g., dimensions, dynamics, properties, or interactions). This mathematical construct includes quantities called parameters that are estimated for each sample under consideration.

Nonconvergence: By *nonconvergence* in this instance, we mean that the optimization algorithm of specific computer routines is unable to find the maximum.

Non-Euclidean space: Spaces that are not Euclidean. For example, a space defined by the surface of a sphere is a *non-Euclidean space*.

Power of a statistical test: The *power* of a statistical test corresponds to the probability of rejecting a null hypothesis when it is false. A uniformly most powerful (UMP) test is a test that has most power among all valid tests.

Shape: According to the *Oxford English Dictionary* (compact edition, 1971), *shape* is "external form or contour; that quality of material object (or geometrical figure) which depends on constant relations of position and proportionate distance among all the points composing its outline or its external surface." Shape of a form and the definition of shape can change when a different size measure is used to standardize the forms under study.

Size: According to the *Oxford English Dictionary* (compact edition, 1971), *size* is "the magnitude, bulk, bigness, or dimensions of anything." Different surrogates can be chosen as measures for size. This choice affects the comparison of size of forms, and the operational definition of shape as the latter definition is dependent on the chosen surrogate for size.

Validity of a statistical test: A statistical test is considered *valid* provided the true probability of type I error (the probability of rejecting a hypothesis when it is true) is equal to the specified probability of the type I error. Tests must be valid before one can compare their powers. For example, the usual two-sample t-test that assumes equal variances in the two populations is invalid if the population variances are not equal. It would make little sense to compare powers of two statistical approaches if one of them is invalid.