AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 134:24–35 (2007)

The Problem of Assessing Landmark Error in Geometric Morphometrics: Theory, Methods, and Modifications

Noreen von Cramon-Taubadel,[1]* Brenda C. Frazier,[2] and Marta Mirazón Lahr[1]

[1]*Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge CB2 1QH, UK*
[2]*Department of Anthropology, Penn State University, PA 16802*

KEY WORDS measurement error; precision; morphometrics; Pinocchio effect

ABSTRACT Geometric morphometric methods rely on the accurate identification and quantification of landmarks on biological specimens. As in any empirical analysis, the assessment of inter- and intra-observer error is desirable. A review of methods currently being employed to assess measurement error in geometric morphometrics was conducted and three general approaches to the problem were identified. One such approach employs Generalized Procrustes Analysis to superimpose repeatedly digitized landmark configurations, thereby establishing whether repeat measures fall within an acceptable range of variation. The potential problem of this error assessment method (the "Pinocchio effect") is demonstrated and its effect on error studies discussed. An alternative approach involves employing Euclidean distances between the configuration centroid and repeat measures of a landmark to assess the relative repeatability of individual landmarks. This method is also potentially problematic as the inherent geometric properties of the specimen can result in misleading estimates of measurement error. A third approach involved the repeated digitization of landmarks with the specimen held in a constant orientation to assess individual landmark precision. This latter approach is an ideal method for assessing individual landmark precision, but is restrictive in that it does not allow for the incorporation of instrumentally defined or Type III landmarks. Hence, a revised method for assessing landmark error is proposed and described with the aid of worked empirical examples. Am J Phys Anthropol 134:24–35, 2007.

Geometric morphometrics (e.g., Bookstein, 1991; Rohlf and Marcus, 1993) are powerful statistical and analytical tools for shape analysis that have become increasingly widespread in anthropological research (Adams et al., 2004). Examples of applications in physical anthropology are numerous and include systematic and taxonomic analyses (Delson et al., 2001; Singleton, 2002; Frost et al., 2003; Guy et al., 2003; Harvati, 2003a,b; Harvati et al., 2004), morphological evolution and phylogenetic assessment (Lockwood et al., 2002, 2004; Bastir and Rosas, 2005; Nicholson and Harvati, 2006), the examination of morphological (dis)similarity to assess patterns of ontogeny and growth (O'Higgins and Jones, 1998; Collard and O'Higgins, 2001; O'Higgins and Collard, 2002; Viðarsdóttir et al., 2002; Bookstein et al., 2003; Cobb and O'Higgins, 2004; Mitteroecker et al., 2004), and studies of population admixture (Martinez-Abadías et al., 2006).

The basis of geometric morphometrics is the identification and quantification of landmarks, defined as "a point of correspondence on an object that matches between and within populations" (Dryden and Mardia, 1998, p 3). The major advantage of landmark methods over more traditional methods reliant on interlandmark distances is the preservation of the full geometry of the specimens under study (see O'Higgins et al., 2001 for a recent review) and the generation of clear graphical outputs of the associated shape changes. However, landmark data have the disadvantage of being more difficult to analyze statistically than traditional morphometric variables due to the problem of registration. Shape differences between individual specimens or groups of specimens are analyzed as the perceived relative displacement of individual landmarks within the configuration. In order for these displacements to be quantified, the individual landmark configurations must first be placed in a common frame of reference such that variation between specimens due to translation, reflection, rotation and scaling is removed. Various methods exist to remove or minimize these differences, including baseline or "two-point" registration to produce Bookstein shape coordinates (Bookstein, 1984; Mardia and Dryden, 1989). Other methods such as Generalized Resistant-fit employ the use of repeated medians to transform the data (Siegel and Benson, 1982; Chapman, 1990; Rohlf and Slice, 1990), while Generalized Procrustes Analysis (GPA) registers landmark configurations by minimizing the sum of the squared distances between corresponding landmarks (Boas, 1905; Chapman, 1990). These latter methods generate Procrustes variables, which can then be subjected to multivariate techniques such as principal components analysis, operating in non-Euclidean shape space (Kendall, 1984; Slice, 2001), the geometric and statistical properties of which are increasingly well understood (e.g., O'Higgins, 2000; Rohlf, 2000; O'Higgins et al., 2001; Slice, 2001). A further set of landmark-based morphometric approaches describe shape differences in

*Correspondence to: Noreen von Cramon-Taubadel, Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, The Henry Wellcome building, Fitzwilliam street, Cambridge CB2 1QH, UK. E-mail: n.voncramon@human-evol.cam.ac.uk

Received 2 September 2006; accepted 26 February 2007

DOI 10.1002/ajpa.20616
Published online 14 May 2007 in Wiley InterScience (www.interscience.wiley.com).

terms of deformations rather than absolute landmark displacement. These include Finite-Element Scaling Analysis (FESA, Cheverud et al., 1983) and Thin-plate Splines (TPS, Bookstein, 1989), the latter of which operates on the basis of deforming a target landmark configuration to match a reference object and then quantifying the distortion (i.e., bending energy) (Richtsmeier et al., 2002).

Geometric morphometric methods are becoming increasingly easy to implement through the proliferation of freely-available software, such as Morpheus (Slice, 1998) and Morphologika (O'Higgins and Jones, 2006). The use of 2D and 3D digitizing equipment (e.g., Summasketch III[1] tablet and Microscribe 3DX™ [2]) have become ever more widespread and the acquisition of landmark data is quicker and easier than traditional measuring methods, such as the use of calipers. However, the preferred use of landmark-based methods over co-ordinate free methods (such as Euclidean Distance Matrix Analysis; Lele and Richtsmeier, 1991) has been criticized on the grounds that the results are more the product of an arbitrarily chosen registration method rather than the result of the biological information inherent in the data (Richtsmeier et al., 2002, 2005).

It is not the purpose of this study to provide a critical argument for or against landmark-based methods of morphometric comparison, but simply to recognize its widespread adoption in anthropology and to address the issue of landmark error assessment. A review of currently used methods for assessing measurement error was conducted and each method was evaluated in terms of its mathematical and practical utility. However, no satisfactory method was found that allowed the calculation of measurement error for configurations composed of both easily identifiable (i.e., Type I) and ambiguous (i.e., Types II and III) landmarks. Here, a solution to this problem is proposed by way of a practical extension and modification of an existing method for calculating individual landmark error.

ERROR ASSESMENT IN MORPHOMETRICS

The need for accurate means of identifying error due to observer differences has long been recognized in quantitative anthropometry (e.g., Davenport et al., 1935; Jamison and Zegura, 1974; Heathcote, 1981). As in any science, rigorous, repeatable procedures are required to ensure that results are reproducible. Ideally, data collected in one lab should be usable by others to corroborate or extend the original analyses. Every investigator interacts with a specimen in slightly different ways, and even multiple interactions of the same observer will vary subtly. The use of different measurement instruments adds a further source of variation in data collection. These differences should be quantified and evaluated to assess their impact on the data being interpreted. Using the appropriate methods, investigators can report the precision, and by extension, the validity of their results.

Traditional morphometrics

In traditional morphometrics, the issue of error assessment is relatively straightforward, with univariate measurement error being reported since anthropometric studies of the 1920s and 1930s (e.g., Dahlberg, 1926; Davenport et al., 1935). By the 1970s, multivariate analyses were employed (Spielman et al., 1972; Jamison and Zegura, 1974) to assess the effects of more than one investigator collecting data for multivariate population comparisons, using analysis of variance, canonical variates analysis and product-moment correlations. Heathcote (1981) designed an intra- and inter-observer error study of human crania modeled upon Jamison and Zegura's (1974) anthropometric error study to distinguish between systematic and random error, and the inter-observer differences were assessed multivariately. Jamison and Ward (1993) tested the relationship between the overall magnitude of a distance measurement and the error associated with it. Perhaps unsurprisingly, they found that the absolute error associated with any anthropometric measurement is independent of its magnitude. Therefore, the reliability of a measurement will be proportional to its length, while the coefficient of variation (CV) will be inversely proportional to its length. In other words, "bigger is better" (Jamison and Ward, 1993, p 499). White and Folkens (2000, p 307) suggest a protocol for assessing intra- and inter-observer measurement error based on the CV statistic, whereby error levels across a range of measurements can be compared and are deemed acceptable if they fall below a certain percentage (e.g., $\leq$5%).

Geometric morphometrics

Landmark configurations present a different challenge in terms of assessing observer-induced measurement variations. There are several potential sources of error unique to the acquisition of coordinate landmarks. One source of error arises because self-assessment during the digitizing process is difficult. The raw coordinate values for each landmark contain no inherent information about the specimen, with the potential consequence that error induced by the interaction of the researcher and the digitizer goes unnoticed. Moreover, it has long been recognized that not all landmarks are equally identifiable. Bookstein (1991) provides a typology of landmarks that is based on the extent to which landmarks represent a unique and easily identifiable entity (Type I) as opposed to a position in space determined by other factors and cues (Types II and III). Therefore, there is good reason to suspect that configurations employed in morphometric analyses will be composed of landmarks of differing identification precision. This highlights a problem for researchers interested in testing the repeatability of their morphometric measurements, both within and between observers. In traditional morphometrics, the effect of discrepancy in data recording between two observers on overall results can be tested univariately (i.e., on a variable-by-variable basis) as well as multivariately. However, in geometric morphometrics, it is more difficult to test the extent of observer-differences in data acquisition until the configurations of landmarks are compared.

A review of the recent anthropological literature reveals three general approaches to the issue of error assessment in geometric morphometrics. One such approach involves the use of a registration method (such as GPA) to superimpose landmark configurations and, thereafter, establish whether repeat measures fall within an acceptable range of variation (e.g., O'Higgins and Jones, 1998; Lockwood et al., 2002; Viðarsdóttir et al., 2002). A second method described by Singleton (2002)

[1]GTCO CalComp., Scottsdale, Arizona.
[2]Immersion, San Jose, California.

employs Euclidean distances between repeat measures of a landmark and the centroid of the configuration to assess the relative repeatability of individual landmarks. The third approach involves the repeated digitization of landmarks with the specimen held in a constant orientation to assess individual landmark variance due to observer error (e.g., Corner et al., 1992; Valeri et al., 1998). A general consensus on methodologies suitable for assessing landmark error was not discovered, nor was there evidence of debate as to the relative validity of one method over another. Each of these error assessment techniques is reviewed and discussed in turn below.

METHODS FOR ASSESSING ERROR IN GEOMETRIC MORPHOMETRICS

Involving the registration of landmark configurations

Several studies (e.g., O'Higgins and Jones, 1998; Lockwood et al., 2002; Viðarsdóttir et al., 2002; Harmon, 2007) assess the impact of measurement error by employing GPA (Gower, 1975) to place repeated measures of the same individual into a common coordinate system. GPA implements generalized least-squares rotation and translation of individual specimens and scales all specimens to unit centroid size. Following GPA, the resulting shape coordinates are analyzed in the tangent space to the Procrustes shape space (Dryden and Mardia, 1998; Rohlf, 1999; Slice, 2001) to assess the impact of the measurement variability on the results of the study. In one study, the authors conclude that "floating (instrumentally determined) landmarks, such as the center of the articular eminence, are not more erratic than those located on specific bony features..." (Lockwood et al., 2002, p 451). Euclidean distances (in principal component shape space) between repeated measurements of eight individuals and between any two individuals in their complete data set were computed to evaluate the "overall effect of landmark error" (Lockwood et al., 2002, p 451). As in the study of O'Higgins and Jones (1998), the repeats clustered tightly relative to the variation in the sample as whole. This method demonstrated empirically that intra-observer error was unlikely to confound the discrimination of individuals in the sample. Although this approach does not provide any information about the differential precision of any single landmark, it is a pragmatic approach to the overall effect of intra-observer error on "individual specimen affinity" (Lockwood et al., 2002, p 451).

However, the finding that no landmarks appeared to be more "erratic" (Lockwood et al., 2002, p 451) than others is not surprising. The generalized least-squares algorithm used to place all trials in a common coordinate system distributes landmark error randomly across the configuration, thereby minimizing the overall error by reducing the residual variation around imprecise landmarks and increasing the variation around highly precise landmarks. This phenomenon has been dubbed the "Pinocchio effect" (Chapman, 1990). The case study that follows clearly illustrates this effect and demonstrates that using GPA to register configurations does not permit the evaluation of individual landmark precision.

Demonstrating the Pinocchio effect. When raw configurations are superimposed using a least-squares criterion, the variance of individual landmarks is "smeared out over all landmarks" (Zelditch et al., 2004, p 119). This effect is detrimental for studies of error assessment as "the variances of the influential landmarks [are] allocated to other points" (Zelditch et al., 2004, p 119). This may be easily demonstrated via the case study described here involving isometrically scaled geometric shapes. These comprised three foam cubes of dimensions 2.5, 5, and 10 cm^3. Nine landmarks were digitized for each shape consisting of the eight corner points of the cubes (Landmarks 1–8, Fig. 1A) plus the tip of a wooden skewer imbedded in and projecting from one face of the cube (Landmark 9, Fig. 1A). Each sized cube plus skewer configuration (hereafter referred to as "size specimens") was digitized using a Microscribe 3DX™. For the first round of digitizing, each skewer was inserted in the precise center of one cube face such that it projected from the face at 90° with a length directly proportional to the dimensions of that cube (i.e., 2.5, 5, and 10 cm respectively). Right-angled rulers were used to ensure that the skewer was orthogonal to the cube face and the tip was sharpened to a fine point. These "isometric" versions of the size specimen were digitized three times. Thereafter, each size specimen was adjusted three times, where in each case the skewers were removed from the center of the face and randomly repositioned ("adjusted" versions). The skewers remained projecting at the same relative distance and were positioned orthogonally to the cube face as before. This was to ensure that the only difference between the isometric and adjusted versions of each size-specimen was the position of landmark 9 (Fig. 1A). The three adjusted versions of each size specimen were again digitized three times. In sum, each size-specimen ($n = 3$) was digitized three times for its isometric version plus three times for each of its three adjusted versions, yielding a total of 36 geometric shapes. Thereafter, these data were imported into Morphologika (O'Higgins and Jones, 2006) and configurations were registered using GPA. The resultant Procrustes variables were visualized using the "wireframe" feature, which connected all nine landmarks for each geometric shape (see Fig. 1A). Residual landmark standard deviations[3] from an analysis involving only the isometric versions were compared with landmark standard deviations where both the isometric and adjusted shapes were included. The differences in these two sets of standard deviations were compared statistically using a paired samples t-test (2-tailed).

Figure 1B illustrates the Pinocchio effect using the 36 superimposed geometric shapes following GPA. Although it is clear that the variation around landmark 9 is greatest, it is also obvious that the corners of the original cube (landmarks 1–8) do not completely align once the adjusted shapes are included. The GPA protocol has minimized the overall distance between corresponding landmarks, thereby reducing the effect of the variable landmark 9 and increasing the variability around landmarks 1–8. The centroid sizes for the 36 digitized configurations are given in Table 1. The standard deviations for landmarks 1–8, following GPA of the geometric shapes (Table 2), were used to assess the effects of the imprecise landmark 9 on the entire configuration. The

[3]Landmark standard deviation was calculated as: $\sigma = \sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2+(y_i-\bar{y})^2+(z_i-\bar{z})^2/3N}$ where N is the number of trials (observations) for that landmark.

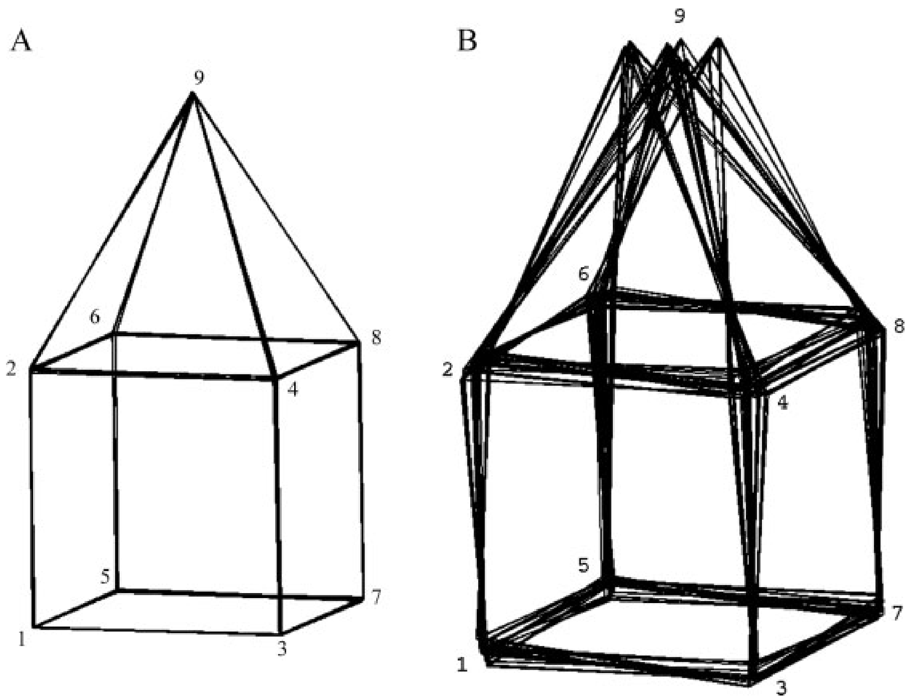

Fig. 1. (**A**) Superimposed isometric versions of three size-specimens following GPA ($n = 9$). (**B**) Pinocchio effect. Superimposed isometric and adjusted versions of three size-specimens following GPA ($n = 36$).

TABLE 1. Centroid sizes (mm) for the repeatedly digitized landmark configurations for isometric and adjusted versions of each of the three size-specimens

	Isometric	Adjusted 1	Adjusted 2	Adjusted 3
Large (10 cm^3 cube)	279.76	280.64	281.41	280.82
	279.02	280.16	281.53	280.73
	278.56	280.70	281.36	280.29
Medium (5 cm^3 cube)	139.42	139.47	138.96	139.74
	139.19	139.56	139.10	139.49
	139.65	139.67	139.04	139.64
Small (2.5 cm^3 cube)	69.25	68.67	68.68	68.82
	69.33	68.59	68.79	68.70
	69.20	69.01	68.98	68.95

standard deviation values for the isometric versions reflect small variation around landmark position due to the interaction of the measurer and the digitizer, and possibly small-scale differences in the manufacture of the foam cubes used to construct the geometric shapes. The adjusted versions (Table 2) show the mean landmark variances of the same landmarks (1–8) following the introduction of the errant landmark 9. The introduction of the errant landmark increases the variation around landmarks 1–8 compared with the isometric versions. The differences in the mean standard deviations around landmarks 1–8 between the isometric and the adjusted versions analysis were found to be significant ($P < 0.0001$, $\alpha = 0.05$).

This case study demonstrates the effect of a single error-prone landmark on an entire configuration. Given that configurations are composed of landmarks with presumably different levels of precision, any approach to error assessment operating on the entire configuration rather than on individual landmarks is likely to yield misleading results. Two further means of assessing error on an individual landmark basis are reviewed below.

TABLE 2. Mean landmark standard deviations (mm) for landmarks 1–8 (see Fig. 1A) following GPA

Landmark	Isometric[a]	Adjusted[b]
1	0.0535	0.1951
2	0.0370	0.1371
3	0.0522	0.1828
4	0.0343	0.1165
5	0.0465	0.1705
6	0.0370	0.1267
7	0.0507	0.1842
8	0.0409	0.1429

[a] Isometric values refer to the configuration with no variant landmark 9 present.
[b] Adjusted values refer to the configuration when errant landmark 9 is included.

Methods involving Euclidean distances

Singleton (2002) presents a method for assessing individual landmark precision employing the Euclidean distance between a given landmark and the configuration

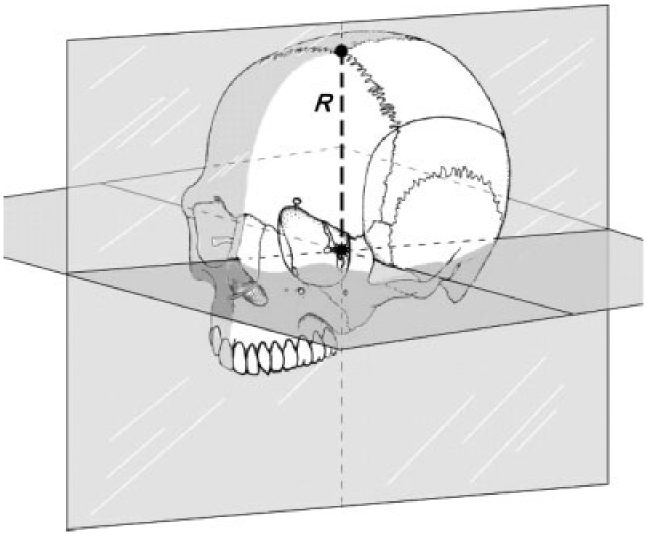

Fig. 2. The centroid radius (R) is the Euclidean distance between a 3D landmark (in this case the anatomical landmark Bregma) and the skull's centroid. Adapted from Franklin (2005).

centroid, a distance referred to here as the centroid radius (Fig. 2). In this case, multiple observations of a single specimen are taken by three observers. The centroid of each landmark configuration is determined, and the centroid radius is computed for each landmark repeat. Mean error and percentage error are calculated for the centroid radius corresponding to each landmark, and the average deviation across landmarks is determined within and between observers. This approach is desirable in that it attempts to assess observer precision on a landmark-by-landmark basis. However, there are two primary reasons why this method may fail to accurately determine landmark precision. Error is calculated on the basis of a distance from landmark to centroid, with the resulting percentage error being highly dependant upon the absolute distances between individual landmarks and the centroid. Landmarks that lie closer to the centroid will exhibit higher relative error and vice versa. This problem becomes more pronounced for smaller configurations (due to the small magnitude of the centroid radii) and those that deviate from a general circular or spherical shape (due to the large variation in centroid radii). The more problematic aspect of this method can be highlighted using Pythagorean geometry (Fig. 3). Landmark error that is collinear with the original centroid vector will be fully detectable (Fig. 3C), but error in any other orientation will be underestimated. For example, given a landmark 30 mm from the centroid, an error of 0.5 mm in any direction perpendicular to the centroid radius will only have an "apparent error" of $30 - \sqrt{(30^2 + 0.5^2)} = 0.004$ mm (Fig. 3A). In such a scenario, the apparent error would be so underestimated as to be functionally negligible. This method may be particularly problematic for morphometric studies in anthropology, given that many surfaces upon which landmarks are located are approximately perpendicular to the centroid radius (e.g., facial anatomy). The problem is exacerbated when the landmarks lie on a surface approximating a sphere (e.g., neurocranial vault) (Fig. 3B). In this case, actual landmark error would go undetected using the centroid radius, as the radius of a sphere is, by definition, of equal length to all points on the surface. In summary, while this method has the advantage of attempting to assess error on an individual landmark basis, it cannot be recommended on the grounds that it is unlikely to correctly estimate actual error.

Methods involving a constant orientation

A third approach to assessing error is outlined by Corner et al. (1992). This technique follows the principle of repeatedly digitizing landmarks on one or more specimens by several observers, while keeping the digitizer and the specimen in a constant orientation relative to each other. Digitizing equipment (such as Microscribe 3DX™) function by referencing the tip of the digitizing stylus to the machine, with the consequence that any movement of the specimen relative to the equipment will register different raw coordinate values between measuring sessions. However, if the specimen and the equipment are not moved between sessions, the raw repeat values of each landmark can be directly compared for variation. Corner et al. (1992) reported the standard deviations of landmark co-ordinates for 11 cranial landmarks, and demonstrated that the error associated with intra- and inter-observer was distributed evenly in all three axes of orientation.

While Corner et al.'s (1992) method is useful for assessing landmark precision on a landmark-by-landmark basis; their approach can only be applied to landmarks taken on an object held static relative to a digitizer, or one that is in a constant frame of reference (such as a CT scan). However, many morphometric studies do not involve data taken from CT scans and include landmarks that require definition relative to cues from other biological features (e.g., anteriormost point). In these cases, the error associated with the procedure used to identify the point on a specimen (e.g., instrumentally defining a point and then marking it in some manner) is intrinsic to the definition of the landmark in space. Therefore, we suggest an extension of the method described by Corner et al. (1992), which provides a means of estimating individual landmark precision, while allowing the freedom of moving the specimen between repeat trials.

PARTIAL SUPERIMPOSITION: A REVISED METHOD FOR ASSESSING LANDMARK PRECISION

Here we propose a methodological extension of the protocol outlined by Corner et al. (1992), designed to overcome the restriction that specimens must not be moved between repeat measuring sessions. The method described here will utilize three landmarks (reference landmarks) upon the specimen to superimpose the repeat trials and thereafter estimate residual error of the nonreference landmarks. Least-squares criteria for removing the effects of rotation will be applied only to these three points, thereby minimizing distortion in the within- and between-observer error estimates for the other landmarks. The general perturbation model (Goodall, 1991; Lele, 1993) underlying the partial superimposition method includes the assumption of isotropic error, where reference landmarks have equal variances and are uncorrelated (Richtsmeier et al., 2005). This approach requires two key assumptions to be made regarding the reference landmarks. Firstly, the error surrounding each reference landmark is spherical in dis-

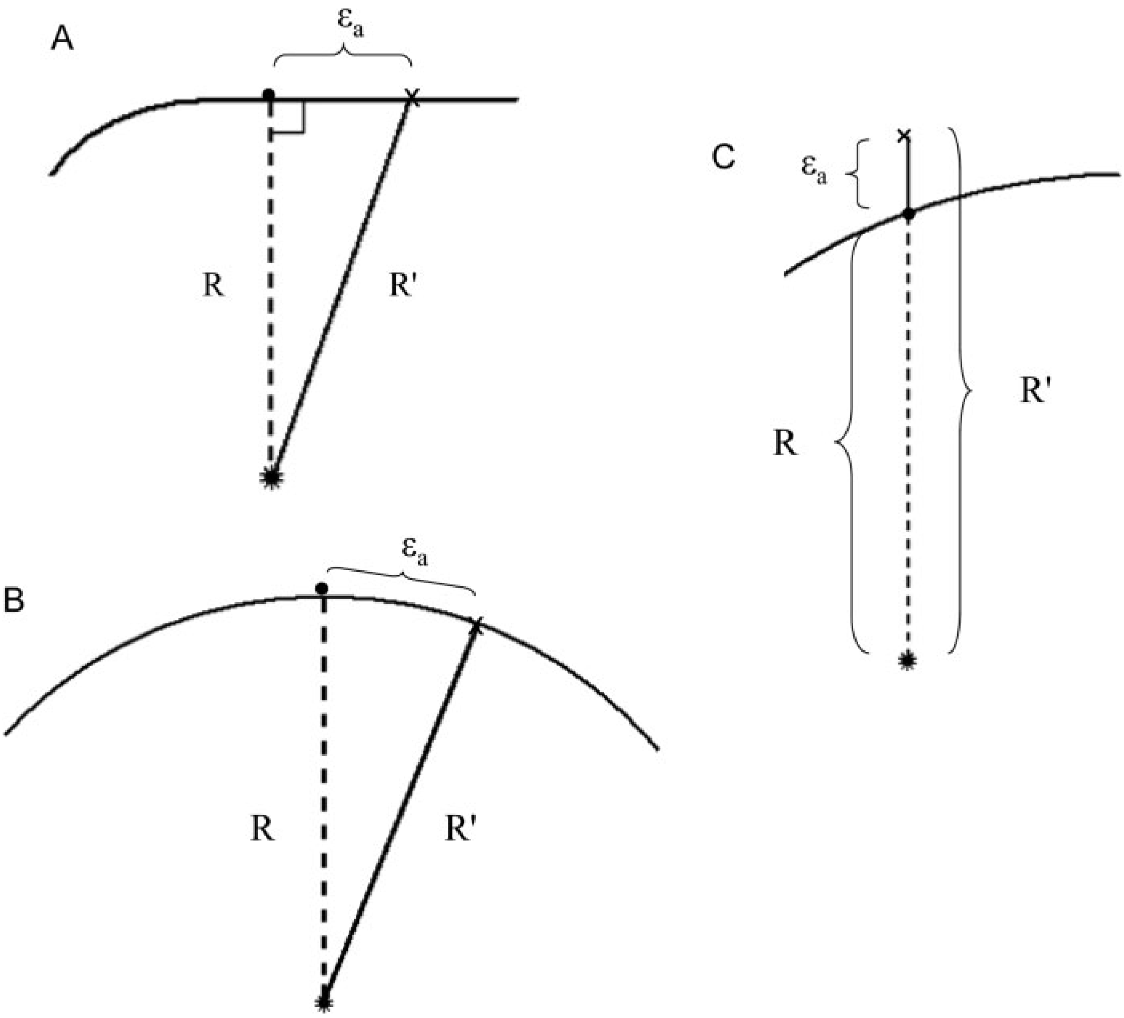

Fig. 3. Potential discrepancies between the calculated error ($R' - R$) and the actual error (ε_a) when using the centroid radius (R) as a means of error assessment. (**A**) Modeling the vicinity of the landmark as a flat surface, error (ε_a) occurs perpendicular to R. Thus, $R' = \sqrt{R^2 + \varepsilon_a^2}$. (**B**) Modeled as a sphere, $R' = R$ and therefore no error along the surface of the specimen is detectable using this method. (**C**) If error occurs normal to the surface, R' and R are collinear and therefore $\varepsilon_a = R' - R$, giving the correct result.

tribution. Additionally, error associated with all three reference landmarks must be considered equal for the purposes of the following calculations. Choosing reference landmarks spread across the configuration will help mitigate the effects of rotation during registration. In some respects this method is similar to the approach taken by Loy et al. (1993), where raw repeats of a 2D landmark configuration are registered using Bookstein's 2-point registration (Bookstein, 1984). Standard deviations were then calculated for the repeatedly measured landmarks not involved in the baseline registration. In contrast to the partial superimposition method proposed here, however, Loy et al. (1993) did not test the level of error inherent in the two landmarks used to align their repeated measures. Measurement variance due to the effect of inter- and intra-observers must first be calculated for reference landmarks using the method proposed by Corner et al. (1992). Thereafter, a single variance value must be chosen to represent the error inherent in the reference landmarks. For a conservative error estimate, we recommend using the square root of the mean of the variances of the reference landmarks.[4] Given these parameters, it is clear that not all points make suitable reference landmarks. They must exhibit low intrinsic variances both within and between observers to avoid the Pinocchio effect during registration. Additionally, points used to superimpose trials must not be collinear and should span the full extent of the configuration rather than cluster together. This method is demonstrated with the following example.

Baseline variance

Before choosing suitable reference landmarks, it was necessary to gauge the level of landmark variance that might be expected simply as a result of the interaction between observer and digitizer rather than the result of

[4]Given reference landmarks d, e, f, we recommend calculating the average reference landmark error value ($\bar{\sigma}$) as: $\bar{\sigma} = \sqrt{\sigma_d^2 + \sigma_e^2 + \sigma_f^2/3}$.

TABLE 3. Mean landmark standard deviations (mm) for three paper landmarks digitized by each of three observers a total of six times

Landmarks	A	B	C
Observer 1	0.213	0.227	0.201
Observer 2	0.217	0.204	0.203
Observer 3	0.184	0.273	0.252
Inter-observer	0.208	0.254	0.230

variance in the definition of the point in space. To achieve this, a small empirical experiment was designed whereby three observers repeatedly digitized three points accurately marked on pieces of graph paper. The three points were secured in different locations close to the digitizing equipment: one horizontally placed on the table in front of the digitizer, another vertically placed to the left of the digitizer and the third placed vertically and behind the digitizing equipment. Several hours elapsed between recordings of the points, and each point was recorded six times by each of three researchers. It should be noted that the observers differed in their prior experience with handling the equipment. Observer 1 (NvCT) was an experienced user of the Microscribe digitizer (i.e., had spent over 100 h digitizing); Observer 2 (BCF) had experience in electronically locating landmarks on CT scans rather than using a Microscribe. Observer 3 had no former experience in landmarking of any nature. The results show that, on average, a point depicted with certainty can be repeatedly recorded by an observer to within an accuracy of approximately 0.25 mm (Table 3). It is notable that these results are invariant regardless of whether the landmark was placed vertically or horizontally, or the level of the observer's prior digitizing experience. The accuracy level reported by the manufacturer of the Microscribe 3DX used is ±0.23 mm (Immersion, CA.). Therefore, these results indicate that it is possible to repeatedly capture the spatial position of a clearly defined landmark to over 90% of the potential equipment accuracy. It also highlights the importance of understanding the contribution of inherent equipment imprecision prior to embarking upon an analysis of observer-induced measurement error.

Choosing reference landmarks

The results of this analysis gave a baseline level of acceptable error for highly repeatable landmarks. Hence, we might expect real biological landmarks to exceed this level to a greater or lesser extent. It was decided to test three Type I landmarks (sensu Bookstein, 1991) of a modern human cranium to ascertain the level to which they exceed the observer-induced variance levels found in the paper landmarks. Bregma and Sphenion (left and right) were chosen arbitrarily as potentially suitable reference points since they fulfill the criteria of noncollinearity and extend across the neurocranium (see Fig. 4). Two human crania were tested and the three observers undertook the digitizing process. As before, the specimen was kept in a constant orientation relative to the digitizer with 1 day elapsed between each measurement session. The results (Table 4) show that intra-observer error (standard deviation) of the reference landmarks varied from 0.153–0.339 mm, while inter-observer error ranged between 0.217−0.268 mm. The average error for the three reference landmarks was also calculated (see Table 4). To test whether the landmark variances acquired for

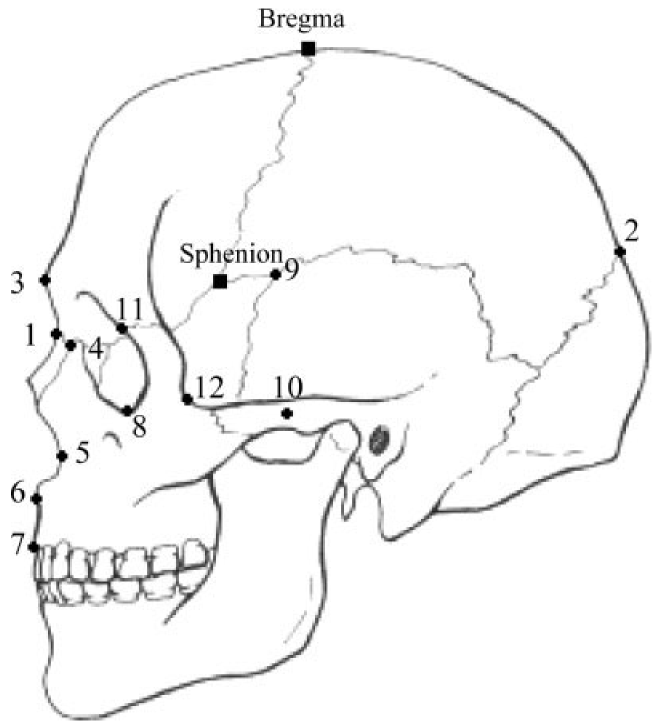

Fig. 4. Configuration of 19 biological nonreference landmarks (circles) and three reference landmarks, Bregma, Sphenion Left and Sphenion Right (squares), tested for measurement precision. See Table 5 for anatomical descriptions.

the "paper landmarks" and the potential reference landmarks differed significantly, a Kruskal–Wallis test was performed. This nonparametric alternative to an ANOVA was chosen because the data violated the assumption of being normally distributed. No significant difference ($\alpha \leq 0.05$) was found between the median values of landmark error of the paper landmarks and those of the biological landmarks in either of the two human crania ($H = 3.722$, $H_c = 3.722$, $P = 0.156$). This result established that the three biological points tested made suitable reference landmarks for further investigation of observer-induced error in a configuration of cranial landmarks.

Full landmark configuration

To demonstrate the proposed partial superimposition method for estimating measurement error, 19 additional (nonreference) cranial landmarks were considered (Table 5, Fig. 4). This configuration contained landmarks of Types I, II and III, following definitions given by Bookstein (1991). Two observers (NvCT and BCF) each digitized the configuration three times with several days lapsing between measuring sessions. In contrast to the method described by Corner et al. (1992), the specimen was moved between sessions, allowing Type III landmarks to be identified and marked afresh prior to each measuring session. The raw data were imported into the cross-platform program Morpheus (Slice, 1998), which was employed to superimpose repeat trials using the three reference landmarks previously tested for precision (shown as squares in Fig. 4). This was achieved by treating each measurement trial as a separate specimen with missing values for all landmarks except its own 19 nonreference landmarks. For example, to calculate intra-

TABLE 4. Mean landmark standard deviation (mm) for three reference landmarks digitized on two[a] modern human crania by each of three observers a total of five times

Landmarks	Bregma	Sphenion left	Sphenion right	Average error[b] ($\bar{\sigma}$)
Observer 1	0.265 (0.289)	0.258 (0.271)	0.252 (0.220)	0.260
Observer 2	0.153 (0.196)	0.189 (0.269)	0.199 (0.256)	0.214
Observer 3	0.263 (0.264)	0.229 (0.272)	0.339 (0.205)	0.265
Inter-observer	0.226 (0.244)	0.217 (0.268)	0.220 (0.227)	0.234

[a] Value for second human cranium in parentheses.
[b] Average error (standard deviation) is calculated across both crania and all three reference landmarks.

TABLE 5. Anatomical descriptions of the 19 landmarks tested for measurement precision

Landmarks	Type[a]	Anatomical description
Nasion	I	The juncture of the frontonasal and internasal sutures in the sagittal plane[b]
Lambda	I	The midline point where the sagittal and lambdoidal sutures intersect[b]
Glabella	III	Most anterior midline point on the frontal bone parallel to the Frankfort plane[b]
Infranasion (2[c])	I	The point of intersection of the nasofrontal, nasomaxillary, and maxillofrontal sutures[b]
Alare (2)	II	The most lateral point on the margin of the nasal aperture
Subspinale	II	The midline point at which the inferior edge of the nasal spine becomes the anterior edge of the maxilla[b]
Prosthion	II	The most anterior midline point on the maxillary alveolar process between the two central incisors[b]
Orbitale (2)	II	The most inferior points on the orbital margins
Krotaphion (2)	I	The most posterior extent of the sphenoparietal suture
Zygion (2)	III	The most lateral points on the lateral surface of the zygomatic arches[b]
Frontomalare orbitale (2)	II	The point where the zygomaticofrontal suture crosses the orbital rim/margin[b]
Jugale (2)	II	The point in the depth of the notch between the temporal and frontal process of the zygomatic bone[b]

[a] As defined by Bookstein (1991).
[b] Based on Martin and Saller (1957).
[c] Left and right.

observer error for the configuration of 19 nonreference and three reference landmarks, with three repeats taken of each, the input file must appear to contain a total of 3 (reference) + 3(19 (nonreference)) = 60 landmarks. For the first repeat specimen, the first block of 19 nonreference landmarks will contain values and the following two blocks of landmarks ($n = 38$) are deemed missing. This pattern is repeated for the second and third repeat measuring trials, with the second block of nonreference landmarks being filled for the second repeat and the third block being filled for the third repeat, etc. In each case, the two unfilled blocks of landmarks are deemed missing. GPA is then applied, but only the three reference landmarks are present for each "specimen", allowing the remaining 57 nonreference landmarks to fall into groups of three repeats. These groups of repeated landmarks can then be assessed for measurement variance on an individual landmark basis by calculating the mean landmark standard deviation. For comparison, the entire configuration (19 landmarks plus three superimposition points) was also subjected to full GPA using Morpheus (Slice, 1998), and both intra- and inter-observer error (in the form of landmark standard deviations) were assessed for each landmark.

Results from both of these procedures are reported and compared in Table 6. Partial superimposition yielded mean landmark standard deviations ranging from 0.105–1.02 mm for Observer 1, from 0.36–1.40 mm for Observer 2, from 0.387–1.19 mm across both Observers (Table 6). The slightly elevated values for Observer 2 are not surprising given that Observer 2 was less experienced with the specific landmark set and in using the 3D digitizing equipment than Observer 1. The results for the inter-observer error study did not vary greatly from those of the intra-observer values for Observer 2, implying that variation in Observer 2's ability to repeatedly identify individual points was random rather than systematic. In contrast, employing GPA registration to superimpose the entire configuration yielded different results. Mean landmark standard deviations ranged from 0.13–0.861 mm for Observer 1, from 0.189–0.901 mm for Observer 2, and from 0.284–0.869 mm across both Observers (in parentheses, Table 6). This latter set of results illustrates the Pinocchio effect, in that the range of landmark error estimates is much lower than those produced by the partial superimposition method. Additionally, aside from one landmark (Infranasion L), all inter-observer landmark error estimates for the full GPA method were lower than those obtained for the partial superimposition method, yielding a misleading under-representation of the true imprecision of the landmarks. For example, the full GPA method underestimates the inter-observer error inherent in Jugale R as being almost half that found by the partial superimposition method (only 0.5 mm as opposed to 0.9 mm)

To critically evaluate the partial superimposition method for assessing landmark error, a configuration of 14 Type I landmarks was assessed for intra-observer error using three methods; the Corner et al. (1992) method, partial superimposition as proposed here and full superimposition using GPA. The three landmarks (landmarks 1–3, Fig. 5) with the lowest measurement error (determined using the Corner et al. method) were used for partial superimposition. The results (Fig. 5)

TABLE 6. Mean standard deviation (mm) for 19 cranial non-reference landmarks digitized by each of two observers a total of three times

	Observer 1	Observer 2	Inter-observer
Nasion	0.230[a] (0.248)[b]	0.464 (0.472)	0.460 (0.397)
Lambda	0.260 (0.273)	0.661 (0.189)	0.744 (0.428)
Glabella	0.352 (0.282)	0.360 (0.437)	0.507 (0.443)
Infranasion L	0.105 (0.266)	0.360 (0.473)	0.387 (0.399)
Infranasion R	0.117 (0.188)	0.422 (0.413)	0.426 (0.395)
Alare L	0.975 (0.803)	0.707 (0.753)	0.981 (0.865)
Alare R	1.020 (0.861)	0.384 (0.373)	0.910 (0.821)
Subspinale	0.163 (0.194)	0.599 (0.243)	0.563 (0.305)
Prosthion	0.312 (0.285)	0.816 (0.565)	0.870 (0.546)
Orbitale L	0.315 (0.341)	1.060 (0.813)	0.851 (0.722)
Orbitale R	0.462 (0.441)	1.019 (0.401)	0.873 (0.437)
Krotaphion L	0.198 (0.130)	0.636 (0.273)	0.569 (0.284)
Krotaphion R	0.346 (0.261)	0.880 (0.496)	0.738 (0.427)
Zygion L	0.516 (0.342)	0.756 (0.399)	0.713 (0.393)
Zygion R	0.880 (0.826)	1.400 (0.901)	1.191 (0.869)
Frontomalare orbitale L	0.363 (0.338)	0.854 (0.372)	0.816 (0.452)
Frontomalare orbitale R	0.414 (0.348)	0.993 (0.494)	0.856 (0.482)
Jugale L	0.337 (0.276)	0.887 (0.543)	0.977 (0.818)
Jugale R	0.409 (0.360)	1.127 (0.540)	0.909 (0.512)

[a] First values are the observed error estimates obtained using the partial superimposition method.
[b] Values in parentheses are estimates of landmark error following GPA of the full landmark configuration.

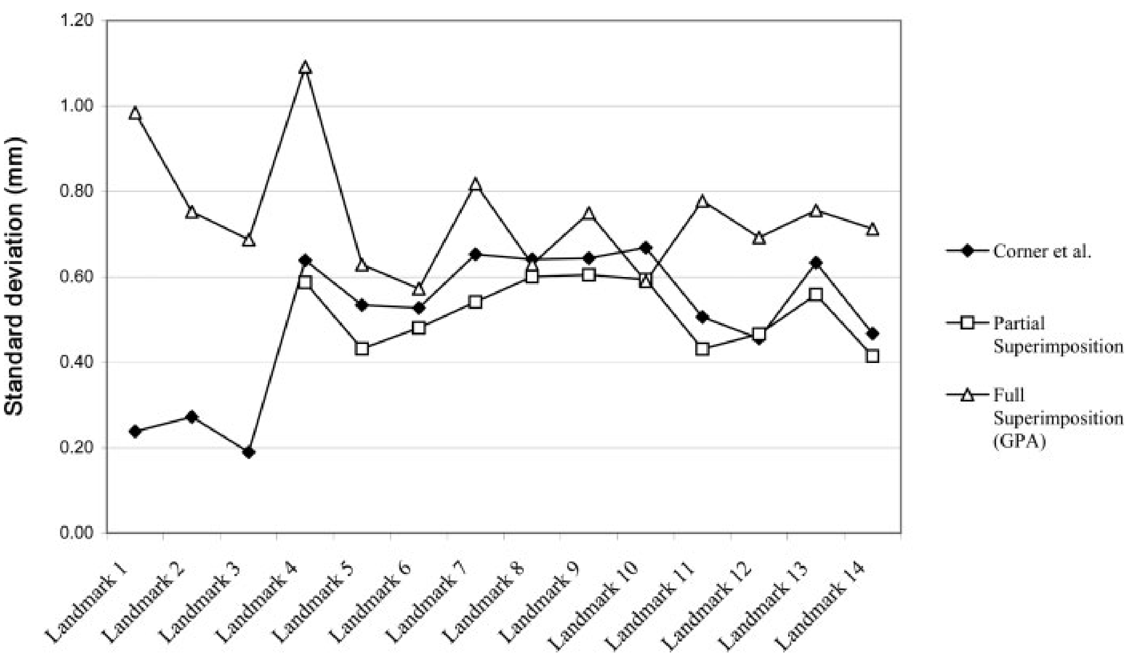

Fig. 5. Comparison of error assessment methods for 14 landmarks. Landmarks 1–3 were found to have the lowest error estimates using the Corner et al. (1992) method and were chosen as reference landmarks for the partial superimposition method. Partial superimposition and the Corner et al. method give the same pattern of results in contrast to the random pattern created as a result of using GPA.

demonstrate that while the partial superimposition method does not yield exactly the same variance estimates as the Corner et al. method, it produces the same pattern of results, in contrast to using GPA, which yields a random pattern of error estimates relative to the Corner et al. method. The latter method for landmark error assessment can successfully be implemented for configurations composed entirely of Type I landmarks or those in a constant coordinate system. However, in the many cases where configurations are composed of different types of landmarks, the partial superimposition method represents a viable means for detecting errant landmarks in the dataset.

DISCUSSION

A method employing partial superimposition for calculating and comparing individual landmark precision has been described and demonstrated employing a configuration of commonly used cranial landmarks. It is advantageous relative to some alternative methods in that it estimates error on an individual landmark basis, thereby mitigating the potentially misleading consequences of the Pinocchio effect. It is a more practical approach than that described by Corner et al. (1992), as many landmarks commonly used in anthropometry require definition relative to other cues (e.g., extrema relative to a plane of orientation) and therefore part of the measurement error resides, at least potentially, in the identification of the point on the specimen. The method described here represents a practical approach to comparing the relative precision of landmarks, where the configuration contains a combination of Type I, II, and III landmarks. Currently no "gold standard" exists amongst users of geometric morphometric methods regarding "acceptable" levels of landmark imprecision. The basis of geometric morphometric analysis is the detection of nonisotropic distributions of specimens in shape space brought about because their configurations deviate from a model of independent isotropic distribution (i.e., they exhibit 'shape' differences) (O'Higgins et al., 2001). Therefore, it is important that analyses are not confounded by nonisotropic measurement-induced variance caused by poorly defined landmarks. The Pinocchio effect is the result of a small number of landmarks with greatly inflated variances affecting the positioning of all other landmarks. Hence, the actual levels of landmark precision are perhaps less important than the fact that all landmarks within a configuration should be approximately equal in terms of precision. This can only be achieved on a landmark-by-landmark basis and should be assessed as such, before testing the overall repeatability of a configuration.

Several general factors must be borne in mind regarding the assessment of landmark precision. One issue relates to the level of accuracy inherent to the equipment employed to acquire the raw data. To compare levels of repeatability across morphometric studies, the baseline accuracy of the digitizing equipment must be reported in tandem. Another factor important in the establishment of a systematic standard for landmark precision is the resolution of the biological data being analyzed. That is, acceptable average levels of repeatability and precision will depend largely on the type of data being compared. In comparative studies involving relatively small-bodied or very similar (e.g., intra-specific) taxa, the effect of measurement error is a greater cause for concern, as data resolution becomes critical for the successful differentiation of specimens within the analysis. Therefore, it could be recommended that error studies in geometric morphometrics be two-fold in their design. Firstly, individual landmark precision should be quantified such that errant landmarks can be identified within the configuration. Subsequently, an assessment of the effect of the configuration variance on the analysis can be made (e.g., as in Lockwood et al., 2002), to ensure that the landmark data acquired are appropriate for the resolution of the research question being addressed.

The partial superimposition method presented here allows the researcher to estimate the measurement error associated with both well-defined and instrumentally defined landmarks. However, it is crucial that reference landmarks used to superimpose repeat trials of a configuration are first tested for measurement error using the Corner et al. (1992) method. The partial superimposition method requires that reference landmarks are uncorrelated and exhibit minimum and approximately equal variances. If these assumptions are fulfilled, then the partial superimposition method will yield measurement error estimates that exhibit the same pattern of results as the Corner et al. method, thereby allowing for the accurate identification of errant landmarks.[5] These errant landmarks can be treated in one of two ways: they can either be discarded from the dataset or they can be more accurately defined in terms of their anatomical or geometric position in space. One of the potential problems in making the transition from traditional morphometrics to landmark-based morphometrics in anthropology is the inadequacy of the anatomical landmark definitions currently available in physical anthropology. Describing a landmark in terms of a junction of sutures may provide an accurate description of an anatomical locality, but may fail to define a single point with sufficient exactness. Therefore, a re-writing of current "landmark" definitions, in a more precise and exacting manner, may be required.

CONCLUSIONS

Having critically examined the methods currently available for assessing measurement error in geometric morphometrics, it was found that no methods were available for assessing individual landmark precision in a context where configurations are composed of both well-defined and instrumentally defined landmarks. Assessing error at the level of the configuration can yield potentially misleading results due to the Pinocchio effect. Here we propose a solution to this problem by presenting an extension of a method described by Corner et al. (1992) that employs partial Procrustes superimposition of landmark configurations, in such a way that individual landmark precision can be estimated. As geometric morphometric methods become more frequently applied to high-resolution shape data, physical anthropologists face the challenge of successfully bridging the gap between anatomical descriptions of "landmarks" and precise definitions of landmarks. Hence, a conservative approach to reporting intra- and inter-observer error levels associated with landmark configurations, such as the one presented here, would appear a desirable research strategy for the future.

ACKNOWLEDGMENTS

We are indebted to S.J. Lycett for his perceptive comments and to R.P. Adams for advice on mathematical and statistical issues. We thank F. Crivellaro for helping to collect some of the data presented in this study. Thanks also to R.A. Foley, L.C. Vinicius and K.E. Willmore for helpful comments during the preparation of this manuscript. Helpful suggestions from Clark Larsen and two anonymous reviewers greatly improved this work. Specimens measured during the course of these investigations are housed in the Duckworth Collection, University of Cambridge. Funding was gratefully received from the Cambridge Gates Trust and St. Johns College, Cambridge (NvCT), the National Science Foundation, USA(BCF) and the Leverhulme Trust (MML).

LITERATURE CITED

Adams DC, Rohlf FJ, Slice DE. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. Ital J Zool 71:5–16.

[5]We emphasize that this partial superimposition method yields only an estimate of error for the nonreference landmarks. Our purpose here is to provide a simple, practical technique for use by morphometricians. Clearly, more sophisticated statistical work may improve upon this approximation method.

Bastir M, Rosas A. 2005. Hierarchical nature of morphological integration and modularity in the human posterior face. Am J Phys Anthropol 128:26–34.
Boas F. 1905. The horizontal plane of the skull and the general problem of the comparison of variable forms. Science 21:862–863.
Bookstein FL. 1984. A statistical method for biological shape comparisons. J Theor Biol 107:475–520.
Bookstein FL. 1989. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans Pattern Anal Mach Intell 11:567–585.
Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press.
Bookstein FL, Gunz P, Mitteroecker P, Prossinger H, Schaefer K, Seidler H. 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J Hum Evol 44:167–187.
Chapman R. 1990. Conventional Procrustes approaches. In: Rohlf FJ, Bookstein FL, editors. Proceedings of the Michigan Morphometrics Workshop (Special publication No. 2). Ann Arbor: University of Michigan Museum of Zoology. p 251–267.
Cheverud JM, 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.
Cobb SN, O'Higgins P. 2004. Hominins do not share a common postnatal facial ontogenetic shape trajectory. J Exp Zool (Mol Dev Evol) 302B:302–321.
Collard M, O'Higgins P. 2001. Ontogeny and homoplasy in the papionin monkey face. Evol Dev 3:322–331.
Corner BD, Lele S, Richtsmeier JT. 1992. Measuring precision of three-dimensional landmark data. J Quant Anthropol 3:347–359.
Dahlberg G. 1926. Twin births and twins from the hereditary point of view. Stockholm: Bokförlags A.B. Tidens Tryckeri.
Davenport CB, Steggerda M, Drager W. 1935. Critical examination of physical anthropometry on the living. Proc Am Acad Arts Sci 69:265–285.
Delson E, Harvati K, Reddy D, Marcus LF, Mowbray KM, Sawyer GJ, Jacob T, Marquez S. 2001. The Sambungmacan 3 *Homo erectus* calvaria: a comparative morphometric and morphological analysis. Anat Rec 262:380–397.
Dryden IL, Mardia KV. 1998. Statistical shape analysis. London: Wiley.
Franklin D. 2005. Cranial variation in southern Africa. Ph.D. dissertation. Perth: University of Western Australia.
Frost SR, Marcus LF, Bookstein FL, Reddy DP, Delson E. 2003. Cranial allometry, phylogeography, and systematics of large-bodies Papionins (Primates: cercopithecinae) inferred from geometric morphometric analysis of landmark data. Anat Rec A 275:1048–1072.
Goodall C. 1991. Procrustes methods in the statistical analysis of shape. J R Stat Soc B 53:285–339.
Gower JC. 1975. Generalized Procrustes analysis. Psychometrika 40:33–50.
Guy F, Brunet M, Schmittbuhl M, Viriot L. 2003. New approaches in hominoid taxonomy: morphometrics. Am J Phys Anthropol 121:198–218.
Harmon EH. 2007. The shape of the hominoid proximal femur: a geometric morphometric analysis. J Anat 210:170–185.
Harvati K. 2003a. Quantitative analysis of Neanderthal temporal bone morphology using three-dimensional geometric morphometrics. Am J Phys Anthropol 120:323–338.
Harvati K. 2003b. The Neanderthal taxonomic position: models of intra- and inter- specific craniofacial variation. J Hum Evol 44:107–132.
Harvati K, Frost SR, McNulty KP. 2004. Neanderthal taxonomy reconsidered: implications of 3D primate models of intra- and interspecific differences. Proc Natl Acad Sci USA 101:1147–1152.
Heathcote GM. 1981. The magnitude and consequences of measurement error in human craniometry. Can Rev Phys Anthropol 3:18–40.
Jamison PL, Ward RE. 1993. Brief Communication: measurement size, precision, and reliability in craniofacial anthropometry: bigger is better. Am J Phys Anthropol 90:495–500.
Jamison PL, Zegura SL. 1974. A univariate and multivariate examination of measurement error in anthrometry. Am J Phys Anthropol 40:197–204.
Kendall DG. 1984. Shape manifolds, procrustean metrics and complex projective spaces. Bull Lond Math Soc 16:18–121.
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, Richtsmeier JT. 1991. Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. Am J Phys Anthropol 86:415–427.
Lockwood CA, Kimbel WH, Lynch JM. 2004. Morphometrics and hominoid phylogeny: support for the chimpanzee-human clade and differentiation among great ape subspecies. Proc Natl Acad Sci 101:4356–4360.
Lockwood CA, Lynch JM, Kimbel WH. 2002. Quantifying temporal bone morphology of great apes and humans: an approach using geometric morphometrics. J Anat 201:447–464.
Loy A, Corti M, Marcus LF. 1993. Landmark data: size and shape analysis in systematics. A case study on old world Talpidae (Mammalia, Insectivora). In: Marcus LF, Bello E, Garcia-Valdecasas A, editors. Contributions to morphometrics. Madrid: Museo Nacional de Ciencias Naturales. p 215–240.
Mardia KV, Dryden IL. 1989. The statistical analysis of shape data. Biometrika 76:271–281.
Martinez-Abadias N, Gonzalez-Jose R, Gonzalez-Martin A, Van der Molen S, Talavera A, Hernandez P, Hernandez M. 2006. Phenotypic evolution of human craniofacial morphology after admixture: a geometric morphometrics approach. Am J Phys Anthropol 29:387–398.
Martin R, Saller K. 1957. Lehrbuch der anthropologie. Stuttgart: Fisher.
Mitteroecker P, Gunz P, Bernhard M, Schaefer K, Bookstein FL. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J Hum Evol 46:679–698.
Nicholson E, Havarti K. 2006. Quantitative analysis of human mandibular shape using three-dimensional geometric morphometrics. Am J Phys Anthropol 131:368–383.
O'Higgins P. 2000. The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. J Anat 197:103–120.
O'Higgins P, Chadfield P, Jones N. 2001. Facial growth and the ontogeny of morphological variation within and between the primates *Cebus apella* and *Cercocebus torquatus*. J Zool Soc Lond 254:337–357.
O'Higgins P, Collard M. 2002. Sexual dimorphism and facial growth in papionin monkeys. J Zool Soc Lond 257:255–272.
O'Higgins P, Jones N. 1998. Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. J Anat 193:251–272.
O'Higgins P, Jones N. 2006. Tools for statistical shape analysis. York: Hull York Medical School. http://www.york.ac.uk/res/fme/resources/software.htm
Richtsmeier JT, Deleon VB, Lele S. 2002. The promise of geometric morphometrics. Yearbook Phys Anthropol 45:63–91.
Richtsmeier JT, Lele SR, Cole TM III. 2005. Landmark morphometrics and the analysis of variation. In: Hallgrimsson B, Hall BK, editors. Variation. Boston: Elsevier Academic Press. p 49–68.
Rohlf FJ. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J Classification 16:197–223.
Rohlf FJ. 2000. Statistical power comparisons among alternative morphometric methods. Am J Phys Anthropol 111:463–478.
Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. TREE 8:129–132.
Rohlf FJ, Slice DE. 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.
Singleton M. 2002. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). J Hum Evol 42:547–578.

Slice DE. 1998. Morpheus et al.: software for morphometric research. Revision 01-30-98. New York: Department of Ecology and Evolution, State University of New York, Stony Brook.

Slice DE. 2001. Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. Syst Biol 50:141–149.

Spielman RS, Da Rocha FJ, Weitkamp LR, Ward RH, Neel JV, Chagnon NA. 1972. The genetic structure of a tribal population, the Yanomama Indians. VII. Anthropometric differences among Yanomama villages. Am J Phys Anthropol 37:345–356.

Valeri CJ, Cole TM III, Lele S, Richtsmeier JT. 1998. Capturing data from three-dimensional surfaces using fuzzy landmarks. Am J Phys Anthropol 107:113–124.

Viðarsdóttir US, O'Higgins P, Stringer C. 2002. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. J Anat 201:211–229.

White TD, Folkens PA. 2000. Human osteology, 2nd ed. San Diego: Academic Press.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004. Geometric morphometrics for biologists: a primer. San Diego: Elsevier Academic Press.