Palaeo-Math 101
Going Round the Bend II: Extended Eigenshape Analysis

In the last essay I explained the basics of eigenshape analysis and compared/contrasted that method to both Fourier and Procrustes approaches to shape characterization and analysis. The mathematics of eigenshape analysis has much more in common with Procrustes-based approaches (*e.g.*, relative warps, Procrustes PCA) than with Fourier-based approaches despite the fact that its canonical shape functions — the Zahn & Roskies (1972) Φ and Φ^* shape functions — were developed originally to solve a problem in Fourier shape characterization (see MacLeod 2011).

While there is no question that eigenshape analysis *can* be used to represent the form of boundary outlines, questions have been raised regarding the extent to which eigenshape-like approaches in particular, and outline-based approaches more generally, portray the forms to which they are applied accurately. Interestingly, this criticism raises an important set of issues that pertain as much to landmark-based morphometric methods as they do to outline-based methods. Even more importantly, addressing some of the concerns that have been raised in this area has driven the development of eigenshape analysis to become a much more comprehensive and flexible tool than the one proposed originally by Lohmann (1983, see also Lohmann and Schweitzer 1990).

The crux of the problem many morphometricians have with outline-based analysis methods in general has been well described by Bookstein (1991) and is illustrated in Figure 1. In the absence of information about which points on a boundary outline curve in one specimen correspond to which points on a boundary outline curve in another specimen, all boundary outline curve sampling protocols contain a degree of ambiguity with regard to how forms or shapes within a sample are to be matched. This matters because the degree to which two forms or shapes are judged to be similar to, or different from, one another is controlled entirely by the manner in which the points used to represent their forms or shapes are matched.
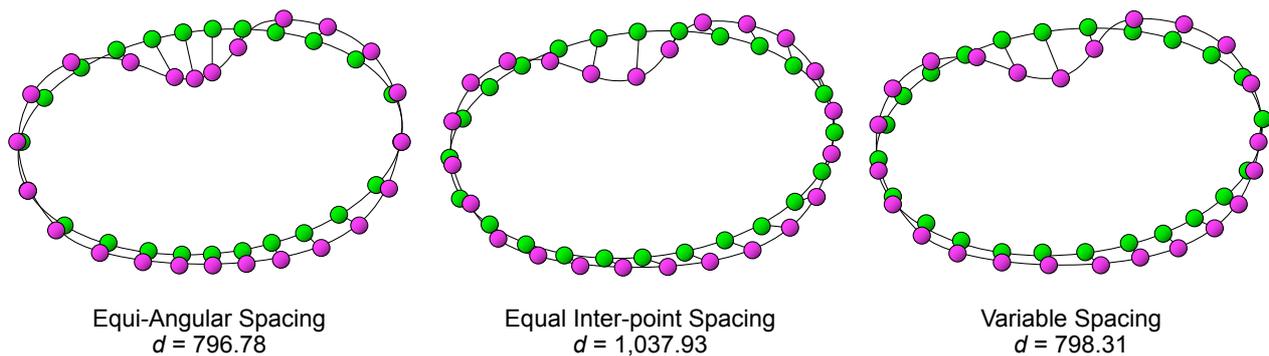

Equi-Angular Spacing	Equal Inter-point Spacing	Variable Spacing
d = 796.78	*d* = 1,037.93	*d* = 798.31

Figure 1. Variations in the shape 'distance' estimates (*d*) for the same forms under different semilandmark sampling schemes.

In terms of landmark analysis the question of how best to match forms also arises, but in a somewhat different context. Figure 2 shows the point-level correspondences between two fish morphologies characterized using a landmark-based measurement system. Landmarks are usually regarded as being more definite and so less subject to disagreements among investigators about how they should be matched. However, it is simply not the case that, in all instances, it is clear exactly how landmark locations on one specimen correspond to landmark locations on another specimen.
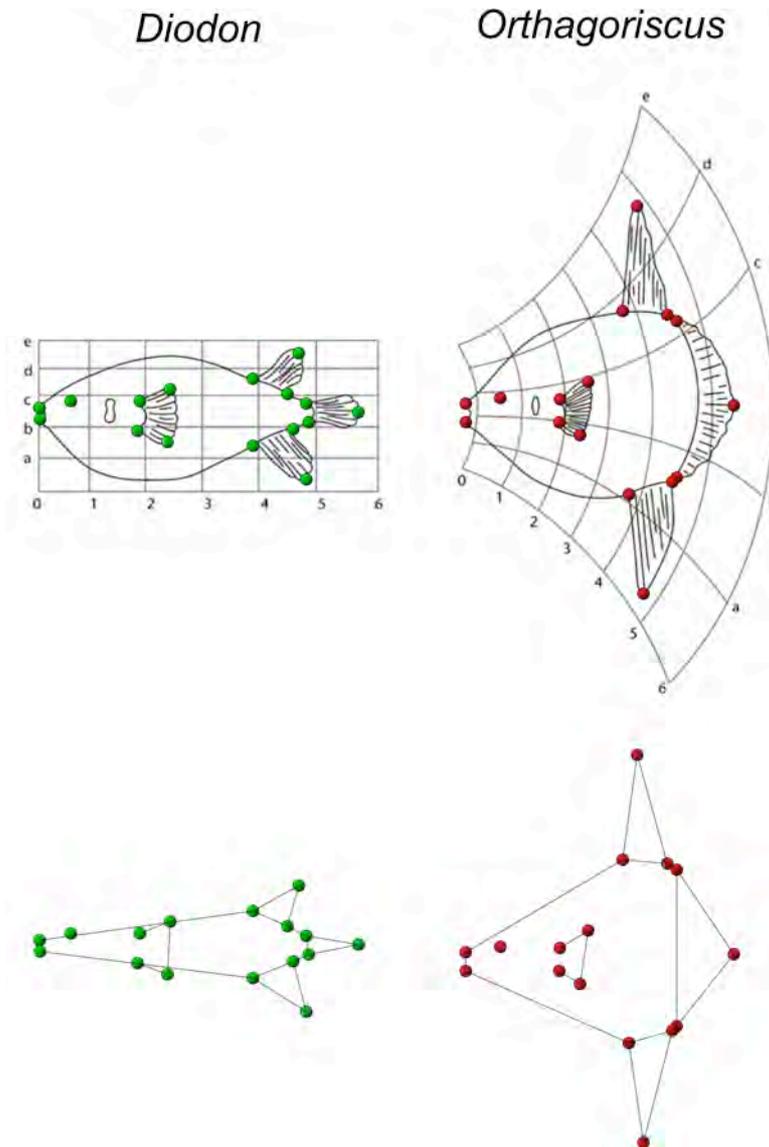
Figure 2. Landmark representations of the two actinopterygian fish morphologies used by d'Arcy Thompson (1917) to illustrate his transformation grid approach to shape comparison. While a comparison of topologically corresponding landmarks between forms is sufficient to infer the general geometric character of the implied shape transition, problems arise as a result of (1) lack of specific morphological information regarding the biological character of landmark point placements and (2) the concomitant abbreviation of the shape engendered by using only the few topologically relocatable landmark points common to both forms fails to capture critical aspects of the morphology (e.g., depth of the body, shape of the dorsal and anal fins, shape of the tail) leading to errors in both the overall and localized assessment of shape similarity across the form. These errors are of comparable magnitude, and comparable biological importance, to the errors induced by not knowing how to match semi-landmark points located along the outlines of two forms (see Fig. 1).

Imprecision and inconsistency in the placement of landmark points is be especially problematic in the case of type 2 landmarks (e.g., extremal points, maximum of curvature) and type 3 landmarks (semilandmarks) which often comprise the bulk of the points used in a landmark-based morphometric analysis (see Fig. 2 for examples). In the case of type 2 landmarks, despite the precise definition of this category, in most cases the positions of type 2 landmarks are judged 'by eye'. Also, as the definition of this type of landmark is logically bound up with an assessment of the positions of other points on the form, type 2 landmarks are actually a special case of type 3 landmarks, which is the same category used to define the locations of boundary outline landmarks. Accordingly, the problem of landmark matching is not so different in principle among 'landmark-based' and 'outline-based' approaches to shape characterization. In practice, more of these arbitrary matching decisions are necessary in the latter simply because more (semi)landmarks are used in the analysis. But is this a deficiency or an advantage?

It should be noted at this point that the number of landmarks available for use in a landmark-based investigation is usually so small relative to the amount of shape information available in the specimen or image as a whole that there is the ever-present danger of gross under-representation of the true shape of the specimen and so the pattern of similarity existing across a sample of forms (see Fig. 2). This issue does

not arise in those instances where there is a clear biological reason to track spatial changes in a few specific points across a set of specimens (*e.g.*, investigations of functional morphology). Nevertheless, in the vast majority of morphometric investigations what is required is a measure of overall geometric similarity for the shapes under consideration. It is in these (common) instances that representation of shape similarity via comparisons between a small number of landmark points may be misleading in terms of representing the overall amount and overall character of shape similarities or differences in a sample; especially so if a substantial proportion of the landmarks are located in regions of the form that have no particular significance with respect to the biological hypotheses under evaluation. In criticising outline-based approaches to shape analysis I can't help but feel many morphometricians have failed to take a critical look at their own (landmark) data in terms of their realized ability to represent shapes that are pertinent to the biological problems they are trying to solve. This situation is improving, though, as number of morphometricians interested in boundary outline analyses grows, as the software tools for undertaking outline-based analyses improve (e.g., Bookstein 1996, 1997; Green 1996), and as the morphometrics community becomes more open to extending morphometric procedures to new types of morphological data (*e.g.*, Gunz *et al.* 2005, MacLeod 2008, Polly 2008, Gunz *et al.* 2009) and to new fields of inquiry (*e.g.*, MacLeod *et al.* In press).
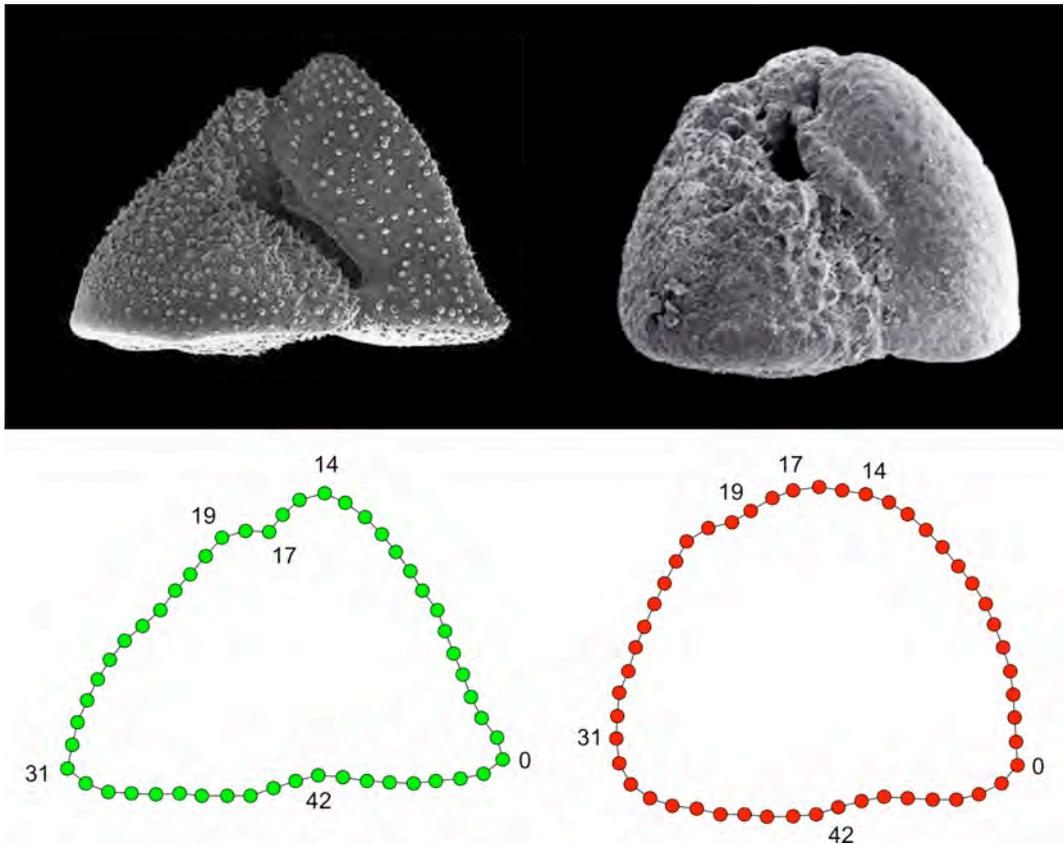

Figure 3. Two specimens of the planktonic foraminifer species *G. truncatulinoides* (upper) whose outlines in apertural view have been quantified geometrically by 50 intervals between semilandmark points. For both specimens digitization began at a landmark (point 0) that was judged to be topologically homologous on both specimens. As digitization proceeded however, other, equally homologous locations are represented by semilandmarks whose position varies in the point sequence. This lack of sequence correspondence introduces error into shape similarity calculations based on these data.

In the context of eigenshape analysis the outline shape registration problem is easy to visualize. Figure 3 shows a sequence of 51 semilandmark points that have been used to quantify the outline shapes of two specimens of the planktonic foraminifer *Globorotalia truncatulinoides*, each specimen having been oriented in apertural view. Even though the boundary outline digitization process began at the same point on each specimen's outline (point 0), many points in the semilandmark sequence fall on different biological structures. Accordingly, in such data there exists an artificial mismatch between points of equivalent topological position in these sequences of semilandmark points. As a result, the geometric difference estimated between these shapes can be divided into two factors: a factor arising from genuine differences in the boundary outline shapes and a factor arising from the mismatch between semi-landmark points with respect to localized biological structures. It is biologically valid to be interested in the true shape differences that characterize the former category (say the form of the ultimate chambers of different foraminifer shells), but less so if the semilandmark sequences being used to represent the spatial position of complex biological structures do not

enforce a strict and biologically comparable matching between semilandmark sequences at least in the regions where the correct matching pattern is known.

Of course, this effect is only noticeable when we have subordinate biological structures that provide evidence of shape correspondence *along the outline*. In some cases such structures will be lacking. When this occurs the analyst has no choice but to accept that some degree of potential mismatch between the semilandmark digitization sequence and the underlying biology will exist and contribute to the overall estimate of shape difference. However, absence of evidence for biological correspondence among outline segments is just that … absence of evidence. If there is no way to determine how points along a boundary outline sequence 'should' match up in terms of the underlying biology, the task of the morphometrician becomes one of representing the curve or curve segment in a manner that minimizes the number of *ad hoc* semilandmark matching hypotheses. In particular, when such situations arise it is not appropriate, in my opinion, to pretend such information exists or, even worse, to force semilandmark data to conform to some esoteric matching pattern mandated by (say) employing non-biological constraints to mask this fundamental lack of biological evidence. [Note: I will return to this topic in the next column when I discuss the sliding semilandmark procedure.] Nonetheless, if evidence for subordinate structural correspondences along boundary outlines exists — at it almost inevitably does in the context of most biological and/or palaeontological investigations — these correspondences can and should be used in designing the measurement schemes that sample the outline and quantify its biological structure across the sample. Of the outline analysis procedures available to date, only eigenshape has developed a procedure that allows this type of information to be accessed and utilised in a shape analysis investigation.

The solution to the dilemma of mismatch between semilandmark sequences and the underlying biological structure of organismal outlines is simple in principle. The form of each specimen's boundary outline can be represented by collecting a sequence of semilandmark points along the outline's trace. Although these points do not need to be equally spaced along the outline, calculations are simplified greatly if they are and this convention has become a useful standard (see Fig. 3). Next a series of landmarks representing points of corresponding or equivalent locations on the boundary outline are designated. These points are used to subdivide the outline into topologically equivalent segments. Each outline segment consists of three parts. The two ends of the segment are defined by landmark points at which strict biological and (semilandmark) sequence conformity is enforced. Between these points lies a region of uncertainty with regard to the correct matching of semilandmark point locations across the sample, but one that derives from a genuine lack of the biological information necessary to realize a specific inter-point matching scheme. In this region the most reasonable matching system to employ — the one that requires the fewest *ad hoc* hypotheses to justify — is one that samples the boundary evenly, using a set of equally spaced semilandmark points that are matched simply, according to their position in the outline segment sequence. Once these three sets of data have been collected it is a relatively simple matter to interpolate any given number of semilandmark points within each landmark-bounded outline segment in order to represent the form of the boundary outline curve in that region to any given accuracy standard. Extending this procedure to all segments into which the boundary outline has been subdivided results in (1) all outlines across the sample being represented by the same number of corresponding segments and (2) each corresponding segment of each outline across all objects in the sample being represented by the same number of semilandmark points.

To illustrate this procedure outlines for the two *G. truncatulinoides* specimens shown in Figure 3 were digitized to a resolution of 200 semilandmark points. Then, in addition to the starting landmark (point 0 in Fig. 3, henceforth Landmark 1), the following landmarks were chosen to subdivide the boundary outline into segments: the umbilical tip of the ultimate chamber (Landmark 2), the peripheral margin of the antepenultimate chamber along the test outline (Landmark 3), the angular bend (= imperforate keel) representing the intersection of the umbilical and spiral sides along the left-hand margin of the outline (Landmark 4), and the intersection between the coiled chambers of the lateral margin of the penultimate chamber of the spiral side test periphery (Landmark 5). These five landmarks were then used to divide the outlines of both specimens into five topologically corresponding outline segments. Finally, a sequence of ten semilandmark point segments was used to represent the form of the inter-landmark boundary in each outline segment. Results of this sampling procedure are shown in Figure 4. Treating the set of Φ shape function coefficients derived from this semilandmark sampling scheme as a column vector, it is a simple matter to obtain an overall measure of shape distinction between any two outlines as either a covariance or a Euclidean distance which, in the latter case for these two *G. truncatulinoides* outlines, is 2.3560. This should be compared with the distance calculated between these two outlines prior to shape registration (see Fig. 3), which is 2.3086.
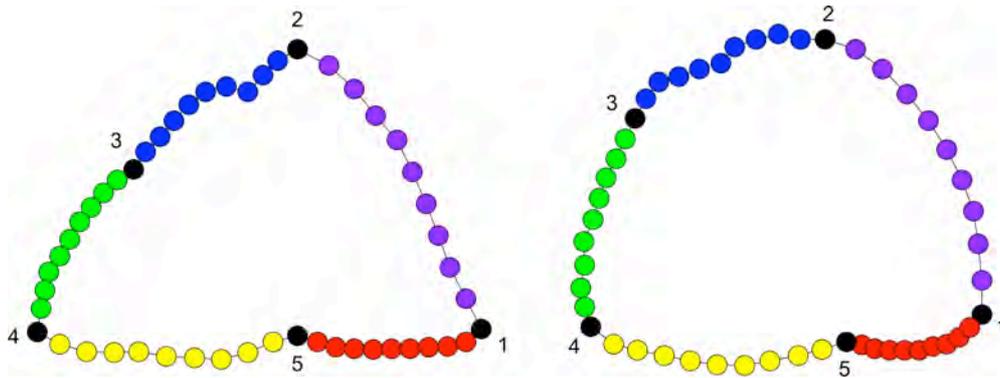
Figure 4. Results of the outline segment approach to characterizing shape variation in two planktonic foraminifer specimens. Colored symbols represent corresponding (= topologically homologous) outline segments whose end-points are defined by landmarks (black symbols). The numbers refer to landmark identifiers (see text). In this scheme an equal number of semilandmarks was used to quantify shape variation in each outline segment.

Note that in the case of these *G. truncatulinoides* specimens, outline registration using landmark data resulted in the two curves increasing their shape-based distinctiveness. This is typically the case when using sampling scheme that respects landmark matching information where that is available as well as one that is highly desirable in terms of maximizing the sensitivity of a morphometric analysis to accurate representation of the structure of form/shape differences among forms.

In the foregoing example an arbitrary decision was made to quantify the form of each of the outline segment using an equal number of semilandmark point locations. While it might be argued that this sort of outline sampling scheme is justified insofar as each outline segment is accorded the same degree of influence in determining the result, it can be appreciated from inspection of Figure 4 that some outline segments are more complex than others (*e.g.*, compare the nature of the boundary outline curves between landmark points 2 and 3 to the outline segment that lie s between landmarks 3 and 4). Enforcement of an equal weighing scheme such as that represented by Figure 4 ensures that some outline segments will be over-represented in terms of the number of semilandmark points required to quantify the segment's form while others may be under-represented.

In order to ensure uniform standard of shape representation is applied to each outline segment it is possible to use an iterative procedure to estimate the number of equally spaced semilandmark points required to quantify the form of a boundary outline curve or curve segment to any specified degree of precision. This procedure can be approached in a number of ways, the simplest of which is to use the curve's perimeter as an index of geometric accuracy.
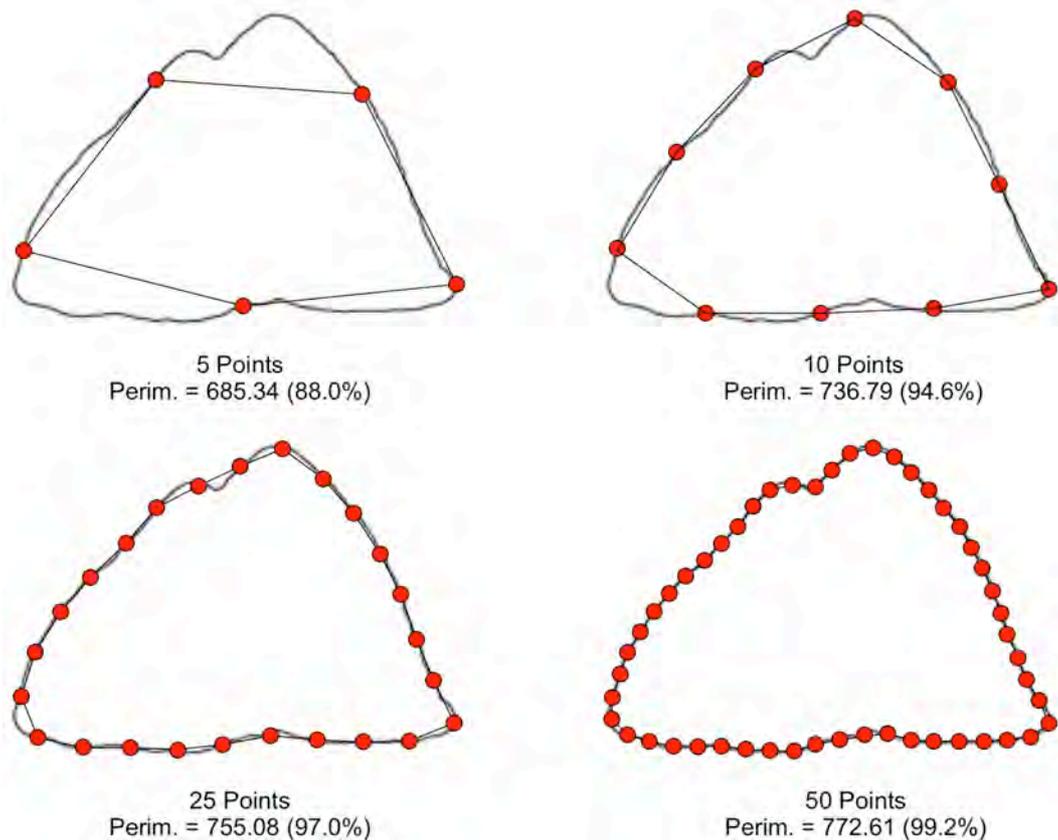
Figure 5. Results of results of a perimeter-based, iterative search for the optimal number of equally-spaced semilandmark points to use to represent a complex curve — in this case a G. truncatulinoides outline — to a qualifiable accuracy level. Percentage values indicate the proportion of the total perimeter (based on n = 200 raw coordinate points) that is represented by the estimated perimeter where n < 200. Note that even comparatively small numbers of semilandmark points can achieve a remarkably faithful representation of the form.

Figure 5 shows a simple illustration of the perimeter-based semilandmark-based curve estimation procedure. Taking an entire *G. truncatulinoides* outline as represented by 200 raw coordinate values as a starting point, the perimeter of this complex curve can be calculated as the sum of Euclidean distances between adjacent landmarks. By increasing the number of equally spaced semilandmarks used to estimate the curve the overall accuracy of the representation increases as the estimated perimeter approaches the value of the measured perimeter. The rate of accuracy increase is surprisingly rapid and a point is usually reached quite quickly after which the inclusion of additional points makes little difference to the overall accuracy of the boundary outline form estimate.

While Figure 5 provides an example of semilandmark point estimation applied to an entire outline — and while this application is eminently appropriate and useful for situations in which an entire outline needs to be analysed as a single segment — it is possible to apply this same procedure to the geometric representation of separate outline segments (see Fig. 4). In both cases all outline curves or corresponding curve segments across a sample would be assessed in the manner shown in Figure 5 and the number of semilandmark points required to represent the most complex shape in the sample to the desired accuracy level (*e.g.*, 95.0%, 97.5%, 99.0%) determined. Then, all outline curves or corresponding curve segments included in the sample are re-estimated using this semilandmark resolution. The re-estimation step is necessary because the same outline curve sampling scheme must be applied to all specimens in the sample and the resolution required by the most shape-rich outline curve or curve segment in the sample employed in order to ensure that the fidelity with which geometric information is represented in the dataset conforms to a uniform minimum standard.

Of course, relatively straight curves will require fewer semilandmarks to represent their forms and more complex curves will require a greater number. Thus, the number of semilandmark points used to represent different outline segments will vary across the set of shapes with a greater proportion of data coming from portions of the shapes that show the greatest geometric complexity across the sample. This variation has the effect of allowing the more geometrically complex regions of the form to exert a differential influence on the results of subsequent multivariate analyses of these data by virtue of the fact that more data are included

from more geometrically complex regions than from geometrically simpler regions; an influence MacLeod (1999) termed 'complexity weighting'.

While the imposition of any weighting scheme may strike some readers as being undesirable, from a geometric point-of-view the differential weighting of some regions of the form relative to others is unavoidable. In Figure 3 note that the 'lower' region of the *G. truncatulinoides* shape between semilandmarks 31 and 0 (= the sharply angled periphery where the spiral and umbilical sides of the test meet) is rather straight when seen in apertural view. However, when strict equality in semilandmark spacing is enforced, fully a third of the shape data sampled comes from this relatively featureless part of the form. Representing what is little more than a straight line using a large number of semilandmark points is, in effect a weighting scheme that differentially accentuates shape similarity. A complexity weighted sampling scheme applied to these data (see below) would reduce the number of semilandmark points assigned to the representation of this curve segment relative to the segments that represent the other two limbs of the test, which is where the overwhelming majority of systematically and taxonomically important components of the shape variation within this species resides. Still, and as we have seen above, complexity weighting is an option for any eigenshape analysis; one that can be taken up, ignored, or modified as the investigator deems most appropriate to address the scientific problem at hand. My point is that, in the context of eigenshape analysis, and unlike current implementation of any form of Fourier analysis, these options exist. Judicious use of complexity or equal weighting schemes can be used by the morphometrician, in effect, to 'tune' an outline analysis to be either sensitive to shape differences or sensitive to shape similarities among objects in a sample.

The subdivision of biologically complex outlines into segments at landmarks located on the boundary outline, along with specification of a intra-boundary segment representation scheme has been termed 'extended eigenshape analysis' by MacLeod (1999) in order to distinguish it from Lohmann's original procedure under which the entire object boundary was treated as a single segment and an arbitrary number of semilandmarks (usually 200) used to quantify the form of this outline curve. Extended eigenshape analysis is best thought of as a hybrid procedure that makes use of the information contained in both landmarks and boundary outline semilandmarks to inform a shape analysis-based investigation. This form-characterization strategy respects the inherent strengths of both landmark and outline data and achieves a geometrically detailed representation of each specimen's entire geometry insofar as that can be expressed meaningfully by the form of the boundary outline.
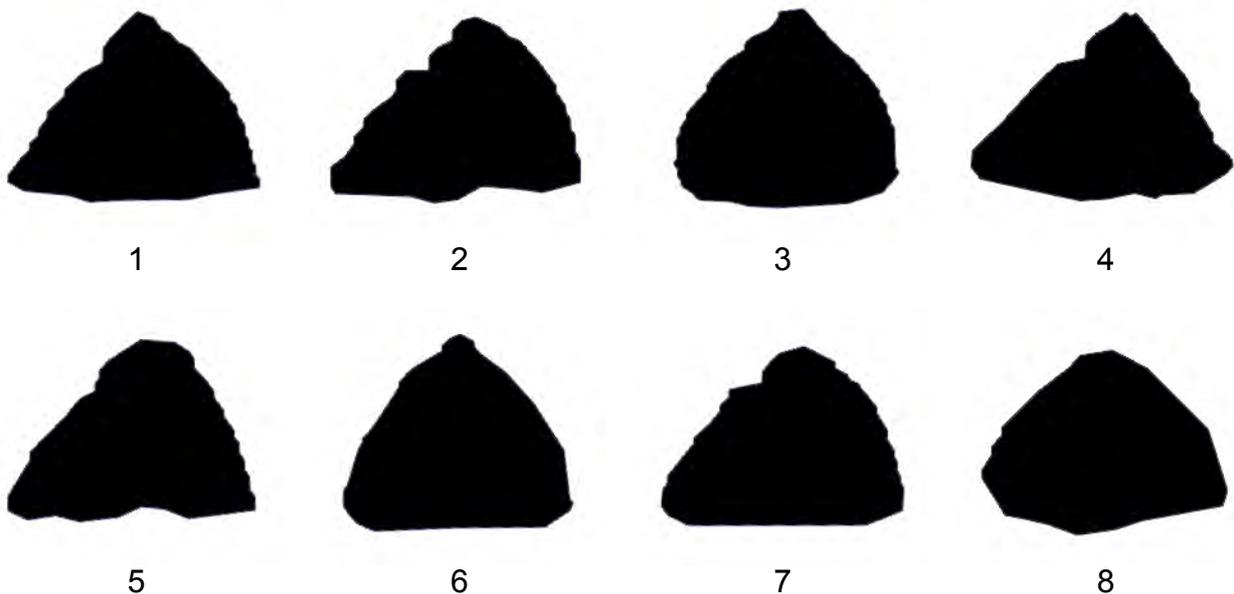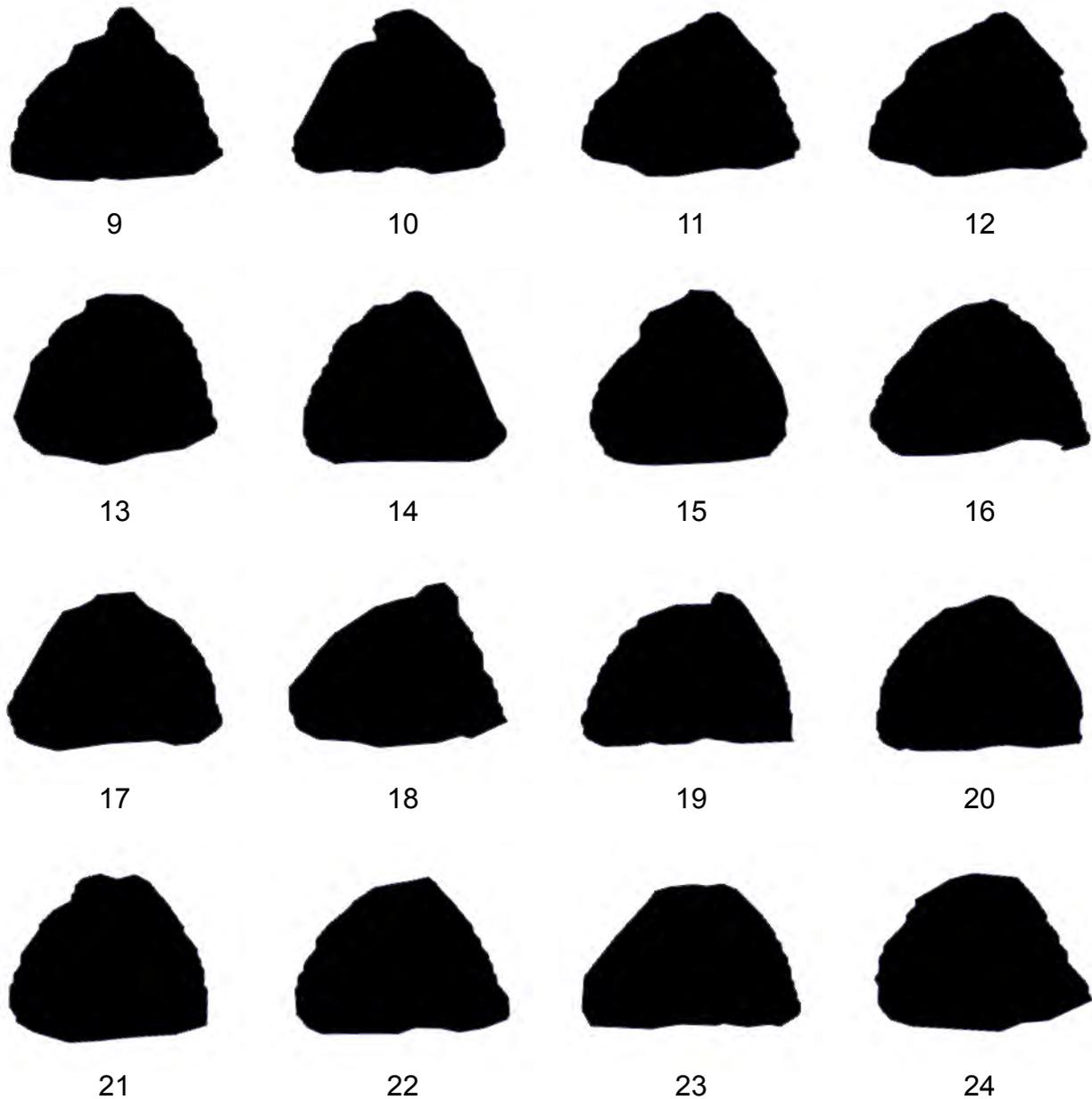
Figure 6. Silhouettes of *G. truncatulinoides* specimens used to illustrate the extended eigenshape analysis procedure. These represent scans of the specimens used to Lohmann (1983) to illustrate use of the original eigenshape method of shape analysis.

In order to illustrate use of the extended eigenshape procedure a selection of 24 outlines of *G. truncatulinoides* specimens oriented in apertural view were obtained from Lohmann (1983, see Fig. 6). These outlines were digitized at an initial resolution of 100 semilandmark points per outline in order to quantify their form. As these outlines lack evidence for the positions of landmarks 3 and 5 that were used quantify the two example *G. truncatulinoides* specimens in Figure 4 for all specimens in the sample, these two landmarks were dropped from the analysis. This, of course, illustrates another practical disadvantage of adopting a strictly landmark-based approach to form or shape characterization — the requirement that all landmarks be visible and able to be located on all specimens in the sample; a requirement that often results in a very small number of valid landmarks that can be used to quantify shape variation across all specimens in a sample (see also Figure 2).

Accordingly, using landmark points corresponding to the the right-most intersection of the umbilical and spiral test faces, the umbilical termination of the ultimate chamber, and he left-most intersection of the umbilical and spiral test faces of each specimen's outline, each outline was subdivided into three segments (Fig. 7). Applying complexity weighting to the sampling and representation of each of these three segments at an accuracy standard of 97.5 percent of the raw semilandmark perimeter value resulted in the umbilical trace of the ultimate chamber periphery (shown in red in Fig. 7) being represented by 21 semilandmark points (31.8%

of the total), the trace of the spiral side topography (shown in blue in Fig. 7) being represented by 14 semilandmark points (21.2%), and the complex curve representing aspects of the shape of the umbilicus and the contributions of pre-ultimate chambers in the final test whorl (shown in green in Fig. 7) being represented by 31 semilandmark points (47.0%). Despite the fact that the green outline segment is complex both geometrically and biologically, the form of the periphery in this region of the *G. truncatulinoides* test is a key feature used by systematists routinely to characterise different populations of of theses species and, in some cases to place specimens into subspecies categories (see Kennet and Srinivasan 1983). In this respect the ability of extended eigenshape analysis to represent the form of shape relations in morphologically and biologically complex regions of the specimens under consideration in a manner that mimics the way a human taxonomists would analyse such regions should be seen as an advantage of the employment of eigenshape analysis for practical morphometric analyses. It is also worth pointing out here that shape variation in the spiral side topography of these tests is much less complex than shape variation in the two umbilical regions and so has been down-graded in terms of its influence in subsequent analyses and interpretations (see below) by virtue of its smaller proportion of representation in the dataset.
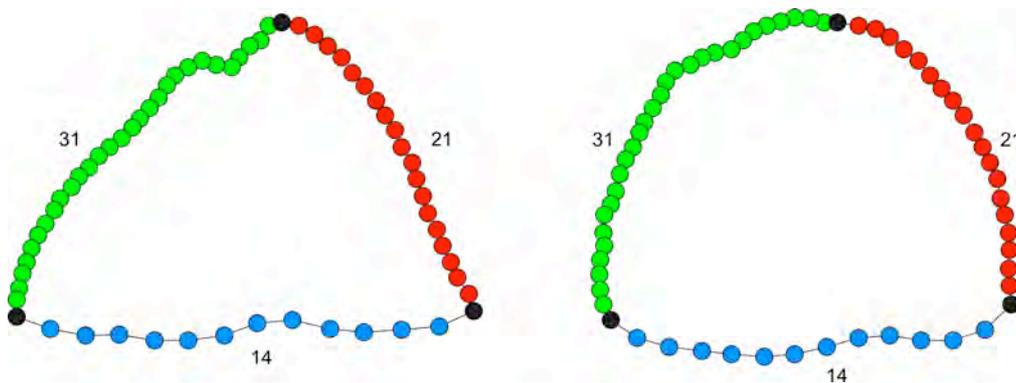

Figure 7. Outline semilandmark sampling scheme used for quantifying the sample of *G. truncatulinoides* test shapes shown in Figure 6. Under this scheme the outline is subdivided into three homologous segments at three landmarks visible in the outline when the specimen is oriented in apertural view (see Fig. 3): the right-most intersection of the umbilical and spiral test faces, the umbilical termination of the ultimate chamber, and he left-most intersection of the umbilical and spiral test faces. Based on the geometric complexity of the these outline curve segments across the sample (Fig. 6) 21, 31, and 14 equally spaced semilandmarks were required to represent the shape of these three segment to a consistent accuracy, 97.5% of perimeter length. This sampling scheme emphasizes shape variation (as opposed to shape similarity) in the subsequent multivariate analysis of these data.

Once the form of each specimen in the dataset had been sampled in the manner shown in Figure 7 the resultant semilandmark coordinate data were used to calculate each specimen's Φ shape function. This operation corrects the raw semilandmark data for extraneous differences in position and rotation as well as sequestering size information from shape information. Careful inspection of Figure 7 shows that, in addition to the shape-related changes in net angular change in the vectors between adjacent semilandmark points, each outline segment across the sample also differs in characteristic spacing between adjacent landmarks. If these inter-landmark spacing data are included in the dataset subjected to subsequent numerical analysis it is the form (= shape + size) of the specimens that is subject of comparison. Alternatively, if these inter-landmark spacing data are excluded from the dataset subjected to subsequent numerical analysis it is shape of the specimens that is subject of the comparison. [Note: in the former care the correlation matrix should be used to estimate the form similarity structure for the sample, in the latter the covariance matrix should be used to estimate the shape similarity structure for the sample. For this demonstration the pure shape analysis option was selected. Nevertheless, the ease with which extended eigenshape analysis — and indeed standard eigenshape analysis — support both types of comparisons is an inherent (and underexploited) feature of the eigenshape approach.

Once the set of Φ shape functions had been obtained for all 24 *G. truncatulinoides* specimens these were used to calculate a shape covariance matrix which was then subjected to principal component analysis (PCA) in the same manner as a standard eigenshape analysis (see MacLeod 2012). This procedure organizes the representation of shape variation in the sample in a series of orthogonal vectors (= eigenshapes) expressing the predominant trends in shape variation present in the sample. The scatter of shapes projected in the subspace formed by the first three extended eigenshapes is shown in Figure 8 along with the equivalent result for a standard eigenshape analysis of the same outlines.
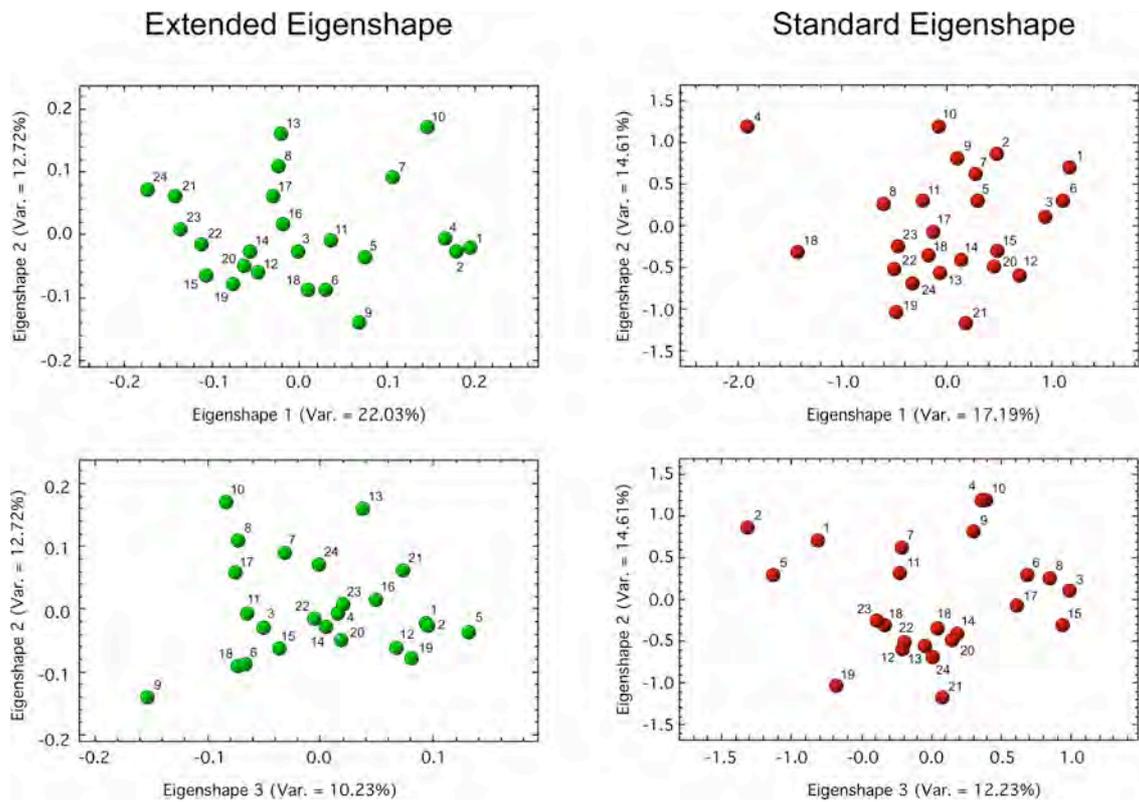
Figure 8. Scatterplots of shapes samples using the extended (left) and standard (right) eigenshape protocols that have been projected into the subspaces formed by the first three eigenvectors (= eigenshapes) of the shape covariance matrices calculated from the Zahn & Roskies Φ shape functions. See text for discussion.

The effect of achieving a more biologically constrained matching between semilandmark points is evident on these plots in two ways. First, the amount of shape variation represented on the first few eigenshape axes differs strongly, especially on eigenshape axis 1. In the standard eigenshape result the first eigenshape accounts for 17.19% percent of the total shape variation observed for the sample whereas, in the extended eigenshape result, this figure stands at 22.03 percent. However, while shape variation was maximized along the first eigenshape in the extended analysis, variation along extended eigenshapes 2 and 3 is lower than for the comparable standard eigenshape results. These patterns are consistent with the purpose of extended eigenshape, which is to focus the analysis on biologically justified comparisons between shapes and minimize the 'shape leakage' that occurs when information regarding biologically appropriate semilandmark point matchings is lacking. As a consequence the extended eigenshape procedure typically results in more shape variation being loaded onto the first shape variation axis and less onto subsequent axes.

Second, the ordination of shapes within the two spaces are very different. Subordinate shape groups present under one shape sampling protocol (*e.g.*, shapes 1, 2, and 4 in the extended eigenshape result) do not cluster together in the other. Shape outliers present in one analysis (*e.g.*, 7, 9 and 10 in the extended eigenshape result) are not outliers in the standard eigenshape result. These distinctions reflect fundamental differences in the shape trends being expressed in the two analyses. Both are 'correct' in that the are both accurate reflections of the major trends in shape variance quantified under the two different sampling schemes. But their differences serve to underscore the importance of selecting the right sampling scheme for the biological problem at hand. For the *G. truncatulinoides* shape sample in Figure 6 the extended eigenshape result should be regarded as the more accurate because it incorporates of more of the biological information available than the standard eigenshape result and because there is a consensus among planktonic foraminiferal taxonomists regarding the validity and importance of the three landmark points used to subdivide the outlines.

While it is possible to get a sense of what these shape variation trends might be by comparing the shape of specimens that plot at the extreme ends of the along-axis distributions shown in Figure 8 with the shapes themselves (Fig. 6), this comparison can be achieved in a more accurate and intuitive manner by modelling along-axis shape variation for both results and comparing the model sequences (tables 1 and 2).

Table 1. Along-axis shape models through the subspace defined by the first three principal components (= eigenshapes) of the *G. truncatulinoides* extended eigenshape shape covariance matrix calculated from the Zahn and Roskies *Φ* shape functions. These models were calculated for five equally spaced coordinate locations within the extended eigenshape space (Fig. 8 [left]) that ranged from the most extreme negative (-2) to the most extreme positive (+2) projected location of a sample specimen on each axis. The middle model in this sequence falls near, but not at, the mean outline shape for the sample. Plots of an overlay of these models along each axis are included to facilitate geometric interpretation of the along-axis shape trends.

Axis	-2	-1	'Mean'	+1	+2	Overlay
Extended Eigenshape 1						
Extended Eigenshape 2						
Extended Eigenshape 3						

Table 2. Align-axis shape models through the subspace defined by the first three principal components (= eigenshapes) of the *G. truncatulinoides* standard eigenshape shape covariance matrix calculated from the Zahn and Roskies *Φ* shape functions. These models were calculated for five equally spaced coordinate locations within the standard eigenshape space (Fig. 8 [right]) that ranged from the most extreme negative (-2) to the most extreme positive (+2) projected location of a sample specimen on each axis. The middle model in this sequence falls near, but not at, the mean outline shape for the sample. Plots of an overlay of these models along each axis are included to facilitate geometric interpretation of the along-axis shape trends.

Axis	-2	-1	'Mean'	+1	+2	Overlay
Standard Eigenshape 1						
Standard Eigenshape 2						
Standard Eigenshape 3						

Looking at the extended and standard eigenshape along-axis models the it is clear that the shape transformation represented by extended eigenshape 1 is very similar to the shape transformation represented by standard eigenshape 2, albeit reversed in polarity.[1] In both cases shapes characterized by a

---

[1] Recall that polarity directions for eigenvectors are arbitrary.

pointed umbilical regional created by a differentially developed ultimate chamber, and an acutely angled (keeled) spiral periphery, occupy one end of these axes and are readily distinguished from those characterized a relatedly flat umbilical area as a result of under-development of the ultimate chamber with a more broadly rounded left-side spiral periphery, which occupy the other extreme. Outline shapes that occupy the central region of these axes exhibit the broadly equilateral triangular form of the sample mean shape. Interestingly, the third eigenshape axis represents very similar shape transformations in both the extended and standard datasets. Here, shapes that exhibit a differentially developed ultimate chamber whose umbilical ends occupy a central position, and whose spiral peripheries are broadly rounded project to the low ends of both axes whorl shapes with more acutely angles spiral peripheries and ultimate chambers whose umbilical ends verge toward the right, project to the high ends of both axes. However, the shape transformation that characterize standard eigenshape 1 and extended eigenshape 2 are unique to each result. In the case of the former, shapes that are relatively elongate parallel to the tests' spiral axis project to positions low on standard eigenshape 1 and those that exhibit a degree of test compression in the direction of the spiral axis project to positions high on standard eigenshape 1. In terms of extended eigenshape 2, the shape distinction represent by this axis separates tests whose ultimate chambers exhibit pointed umbilical ends (low projected positions) from those whose umbilical ends is relatively flattened (high projected positions). In addition, tests that project high on this axes are slightly more inflated than those that project to low positions.

These distinctions are fairly straightforward and easy to appreciate via use of the along-axis shape models. However, the real story, in terms of characteristic differences between the standard and extended eigenshape outline sampling protocols, is more readily understood by examining the overlaid along-axis models (= last column in tables 1 and 2). Note that, in the set of standard eigenshape results, along-axis shape distinctions occupy broad regions of the outline whereas, in the case of the extended eigenshape results, along-axis shape distinctions are much more localized and differentially focused in extended eigenshape 1. This difference in the general character of these two geometric subspaces arises because, in the case of the standard eigenshape results, true biological shape differences are being combined with apparent shape differences due to the mismatching of biological structures. This situation comes about because no control is exerted on the nature of semilandmark matching beyond that of starting the outline digitization sequence at a topologically equivalent landmark across all specimens in the sample. The extended eigenshape outline sampling procedure mitigates against the gross mismatching of semilandmarks throughout the outline sequence by periodically re-calibrating the matching sequence at multiple landmark positions along the boundary outline. The effect of this recalibration is to emphasize differences in shape that have a biological cause and to force these to be expressed on the first few eigenshape axes. The overall result of employing the extended eigenshape procedure is both a more efficient and a more specific summary of shape deviation patterns contained within a sample that can be interpreted with greater biological confidence.

Last, but by no means least, once the lessons of combining the strengths of landmarks and boundary outline semilandmarks have been learned, the last step in the generalization of the eigenshape procedure, and its formal linkage to standard geometric morphometric methods, is for the data analyst to free themselves from the implicit restriction of eigenshape analysis to problems that only consider the form of boundary outlines. In principle we would like to be able to include any geometric data from any structure in our morphometric investigations irrespective of (1) how many boundary outlines are of interest and/or (2) whether all landmarks lie on boundary outlines. Conceptually we would like to compare the equivalents of line drawings of complex biological structures consisting of isolated point locations continuous, closed curves and discontinuous curves terminated at specific landmark points. Such complex assemblages of geometric data cannot be accommodated under the standard or extended eigenshape procedures because both are tied to representation of the shape through use of the Zahn and Roskies Φ and Φ^* shape functions or some equivalent (*e.g.*, Bookstein's, 1978, tangent angle shape function). However, the outline sampling protocols described above do not depend on the use of any particular outline-based shape function in order to be applied. Indeed, transformation of these semilandmark data into the format of a shape function takes place after these data have been interpolated and assembled into topologically corresponding outline segments. Moreover, there exists a fully generalized procedure for transforming any geometric data that can be represented by any type of landmark — including sets of semilandmarks — into its equivalent shape coordinate space: the method of Procrustes superposition and alignment.

Use of Procrustes shape coordinates as the basis for an eigenshape-style analysis was first explored explicitly by Sampson *et al.* (1996) and later by MacLeod (2001). More recently this approach has been used to analyse evolution and adaptation in carnivore crania (see Figueirido *et al.* 2011). The combination of extended eigenshape-like sampling strategies with Procrustes shape coordinate transformations represents a true synthesis between strictly landmark-based and strictly outline-based approaches by regarding these as conceptual end-members of complete spectrum of data combinations and alignment strategies. Owing to this spectrum's inherent flexibility, it is possible to address morphometric problems of any complexity required by the scientific hypotheses and the data at hand while allowing the data analyst to retain complete control over the level of data resolution required to perform hypothesis tests, control over the data used to realize shape coordinate alignment, and control over which data are allowed to participate in which phases

of the analysis (e.g., some data might be carried through the alignment stage analysis passively) so long as the overall strategy/procedure can be justified in terms of the morphological under consideration and the scientific questions being asked.

Software for performing an extended eigenshape analysis is readily available. I maintain stand-alone, public-domain applications for Windows and Apple Macintosh platforms and have written new *Mathematica™* notebook scripts to perform all the analyses described in this essay. In addition, public domain standard and extended eigenshape routines are available in the Morpho-Tools website (http://www.morpho-tools.net)

Norman MacLeod
Palaeontology Department, The Natural History Museum
N.MacLeod@nhm.ac.uk

REFERENCES

BOOKSTEIN, F. L. 1978. *The measurement of biological shape and shape change*. Springer, Berlin 191 pp.

BOOKSTEIN, F. L. 1991. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge 435 pp.

BOOKSTEIN, F. L. 1996. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *In* A. Amini, et al. (eds). *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*. IEEE Computer Society Press, San Francisco, 279–289 pp.

BOOKSTEIN, F. L. 1997. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *Medical Image Analysis*, **1**, 225–243.

FIGUEIRIDO, B., MACLEOD, N., KRIEGER, J., DE RENZI, M., PÉREZ-CLAROS, J. A. and PALMQVIST, P. 2011. Constraint and adaptation in the evolution of carnivoran skull shape. *Paleobiology*, **37**, 490–518.

GREEN, W. D. K. 1996. The thin-plate spline and images with curving features. *In* K. V. Mardia, et al. (eds). *Proceedings in image fusion and shape variability techniques*. Leeds University Press, Leeds, 79–87 pp.

GUNZ, P., MITTEROECKER, P. and BOOKSTEIN, F. L. 2005. Semilandmarks in three dimensions. *In* D. E. Slice (ed). *Modern Morphometrics in Physical Anthropology*. Kluwer Academic/Plenum Publishers, New York, 73–98 pp.

GUNZ, P., BOOKSTEIN, F. L., MITTEROECKER, P., STADLMAYR, A., SEIDLER, H. and WEBER, G. W. 2009. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proceedings of the National Academy of Science*, **106**, 6094–6098.

KENNETT, J. P. and SRINIVASAN, S. 1983. *Neogene planktonic foraminifera: a phylogenetic atlas*. Hutchinson Ross, Stroudsbourg, PN 263 pp.

LOHMANN, G. P. 1983. Eigenshape analysis of microfossils: A general morphometric method for describing changes in shape. *Mathematical Geology*, **15**, 659-672.

LOHMANN, G. P. and SCHWEITZER, P. N. 1990. On eigenshape analysis. *In* F. J. Rohlf and F. L. Bookstein (eds). *Proceedings of the Michigan morphometrics workshop*. The University of Michigan Museum of Zoology, Special Publication No. 2, Ann Arbor, 145-166 pp.

MacLEOD, N. 1999. Generalizing and extending the eigenshape method of shape visualization and analysis. *Paleobiology*, **25**, 107–138.

MacLEOD, N. 2001. Landmarks, localization, and the use of morphometrics in phylogenetic analysis. *In* G. Edgecombe, et al. (eds). *Fossils, phylogeny, and form: an analytical approach*. Kluwer Academic/Plenum, New York, 197–233 pp.

MACLEOD, N. 2008. Understanding morphology in systematic contexts: 3D specimen ordination and 3D specimen recognition. *In* Q. Wheeler (ed). *The New Taxonomy*. CRC Press, Taylor & Francis Group, London, 143–210 pp.

MacLEOD, N. 2011. The cannot hold I: Z-R Fourier analysis. *Palaeontological Association Newsletter*, **78**, 35–45.

MacLEOD, N. 2012. Going round the bend: eigenshape analysis I. *Palaeontological Association Newsletter*, **80**, 32–48.

MacLEOD, N., KRIEGER, J. and JONES, K. E. in press. Geometric morphometric approaches to acoustic signal analysis in mammalian biology. *Hystrix*.

POLLY, P. D. and MACLEOD, N. 2008. Locomotion in fossil Carnivora: an application of the eigensurface method for morphometric analysis of 3D surfaces. *Palaeontologia Electronica*, **11**, 13p.

SAMPSON, P. D., BOOKSTEIN, F. L., SHEEHAN, F. H. and BOLSON, E. L. 1996. Eigenshape analysis of left ventricular outlines from contrast ventriculograms. *In* L. F. Marcus, et al. (eds). *Advances in Morphometrics*. Plenum Press, New York, 211–234 pp.

THOMPSON, D. W. 1917. *On growth and form*. Cambridge University Press, Cambridge 793 pp.

ZAHN, C. T. and ROSKIES, R. Z. 1972. Fourier descriptors for plane closed curves. *IEEE Transactions, Computers*, **C-21**, 269-281.

Don't forget the *Palaeo-math 101-2* web page, now at a new home at:
http://www.palass.org/modules.php?name=palaeo_math&page=1