Palaeo-Math 101

Semilandmarks & Radial Fourier Analysis

The last column completed our extended discussion of landmarks and landmark analysis methods. You now have at your disposal the entire array of mathematically sophisticated and commonly used tools that originally were placed under the general heading of 'geometric morphometrics'. But just as there's more to analytic geometry than plotting discrete points in a Cartesian coordinate system, most morphometricians have come to realize that there must be more to geometric morphometrics than the simple analysis of sets — I often refer to them as 'constellations' — of landmarks.

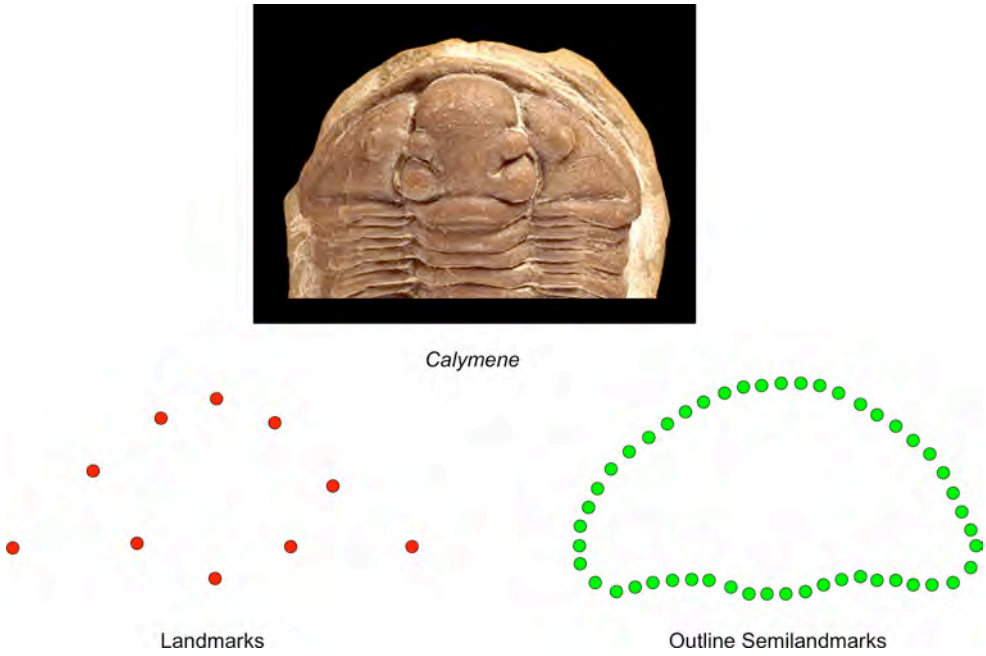

Figure 1. Alternative morphometric representations of *Calymene* cranidium form. Original image (top), landmark-based representation (lower left), semilandmark-based representation (lower right). Both representations are 'correct' insofar as both accurately express the positions of various cranidial features. However, both are also very different in terms of their geometric form and information content. Which representation is the more appropriate for a particular investigation depends on the objectives of the investigation.

The basic problem with landmarks is illustrated in Figure 1. Landmarks are specific point locations on a biological form or image of a form located according to some rule. Dependence relations among the rule sets used to locate landmarks give rise to the landmark type classification system.

> **Type I landmark** - a mathematical point whose [topological] homology is provided by biologically unique patterns on the form (e.g., juxtaposition of tissue types, small patch of some unusual histology).
>
> **Type II landmark** - a mathematical point whose [topological] homology is provided only by geometric, not biological or histological, criteria (e.g., point of maximum curvature along a boundary).
>
> **Type III landmark** - a mathematical point having at least one coordinate that's 'deficient' in the sense that its location is logically dependent on the location of other landmarks and/or the orientation of the specimen as a whole (e.g., either end of a longest diameter, or the bottom of a concavity).

Despite their inherent biological ambiguity these are, nevertheless, fairly stringent definitions. Even in the best of circumstances it is typically the case that a relatively small number of such points can be located on any set of forms. Landmark identification is maximized when all forms of interest represent the same or closely related species whose morphologies are composed of the same parts. Regardless, and as will be well appreciated by those with even a cursory experience of systematics and taxonomy (not to mention ecology, biogeography, phylogenetics), this imposes a rather sever constraint on the range of problems that can be considered under a solely landmark-based morphological sampling scheme.

Prior to the advent of the geometric morphometric paradigm a school of morphometrics developed a set of form-sampling protocols and data-analysis tools that provided all morphometricians with the ability to assess variation in the outlines of biological forms. This approach to biological form/shape analysis was developed alongside the inter-landmark distance-based approach that is usually referred to as multivariate morphometrics (Blackith & Reyment 1971; Pimentel 1979). Owing to (1) the necessary and compelling limitation of multivariate morphometric datasets to comparisons of features that could be regarded as being comparable in some meaningful biological sense and (2) the comparative ease by which the landmarks used to define the inter-landmark distances could be collected, multivariate morphometrics became the dominant approach to the quantitative analysis of biological form throughout the 1960s, 1970s, and 1980s. Nevertheless, owing to its inherently more geometric character, outline morphometrics developed a stronger tradition of shape modeling and a wider range of applications than did multivariate morphometrics.

Although the utility of analyzing the geometry of boundary outlines has long been recognized in a variety of applied geometric contexts, this approach has remained a sparsely populated sub-domain of morphometrics. This lack of popularity occurred for two reasons. First, some prominent morphometricians raised theoretical issues with regard to the idea of of comparing sequences of outline coordinates that were non-comparable in any biologically meaningful sense (see the Bookstein *et al.* 1982). The second, and much more practical, reason many found outline morphometrics difficult to apply was that most morphometricians lacked access to the types boundary outline sampling systems that were needed to pursue such analyses. To an extent, outline morphometrics suffers from the former criticism still and, as a result, this approach is not considered an option for morphometric analysis designs, even in situations where it would clearly be advantageous to do so.

Irrespective of this somewhat contentious history, during the 1990s recognition of the contributions boundary outlines can make to the resolution of a number of outstanding biological problems forced both their reassessment as a primary morphometric data type and the development of tools that effectively brought the analysis of outlines into the corpus of the morphometric synthesis. At the same time, sophisticated and comparatively low-priced software programs designed to extract boundary outline coordinates from digital images of specimens appeared and began to be used by a greater range of morphometricians and students of morphological variation. As a result, outline morphometrics now represents not only a fully justifiable choice for a wide range of morphometric analysis situations, but in many instances the only reasonable choice for evaluating a wide array of complex and information-rich biological forms. Over the next several essays I intend to explore the data, tools, and concepts that lie behind this highly useful, but presently underutilized, branch of morphometrics.

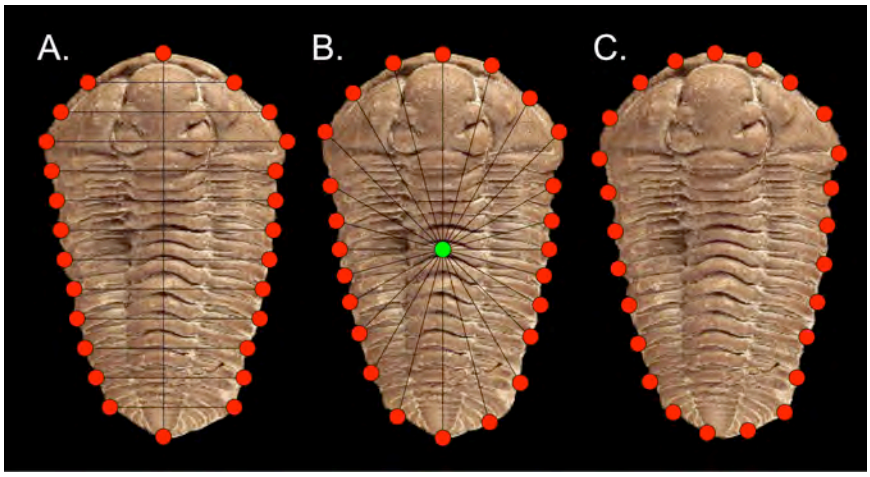

Figure 2. Alternative boundary outline semilandmark placement protocols. A. Equally spaced transverse chords along a maximum diameter. B. Equal angles from the object centroid. C. Equal spacing around the object periphery.

At the heart of the analysis of boundary outlines under the geometric morphometric paradigm is the concept of the semilandmark (Fig. 2). Bookstein (1996a,b) defined this term as referring to a series, or sequence, of landmark coordinates used to represent the form of a curve. Individually, semilandmark points usually conform to the definition of Type 3 landmarks insofar as their positions reference a series of criteria external to the form itself (e.g., orientation of the specimen, placement of the starting point and/or end-point of the boundary). However, the reason why semilandmarks are treated as being ontologically distinct from Type 3 landmarks has more to do with their sequential nature than with their definitional dependance relations.

Semilandmarks are always sequences of mathematical points, usually defined by some sequence-based criterion in addition to the form-based criteria used to define Type 3 landmarks. These criteria include equal

spacing along the length of the curve, equal spacing of horizontal or vertical grid lines along the length of a maximal diameter at the points that intersect the boundary of a curve, or intersections between a set of equiangular radius vectors emanating from a closed form's centroid at points that intersect the boundary of interest. In this essay we're going to take a close look at the oldest of the boundary outline analysis methods, classic radial Fourier analysis (RFA) which utilizes the equi-angular sequence-based sampling criterion.

Fourier analysis and the Fourier series are fundamental mathematical concepts that lie at the nexus of calculus and number theory. While working in the early 1800s on a physical problem relating to the flow of heat between two bodies, the French mathematician Joseph Fourier solved a complex differential equation by using the sine and cosine functions[1] to decompose the equation into a series of simpler components. On the basis of his success Fourier postulated that any continuous or discontinuous function can be represented to any desired level of accuracy by an infinite series of sine and cosine functions. Unexpectedly, this postulate over the nature of functions has not only remained controversial from Fourier's day to our own, but has gone on to inspire research in areas of mathematics well beyond the (already large) domain for which it was formulated originally.

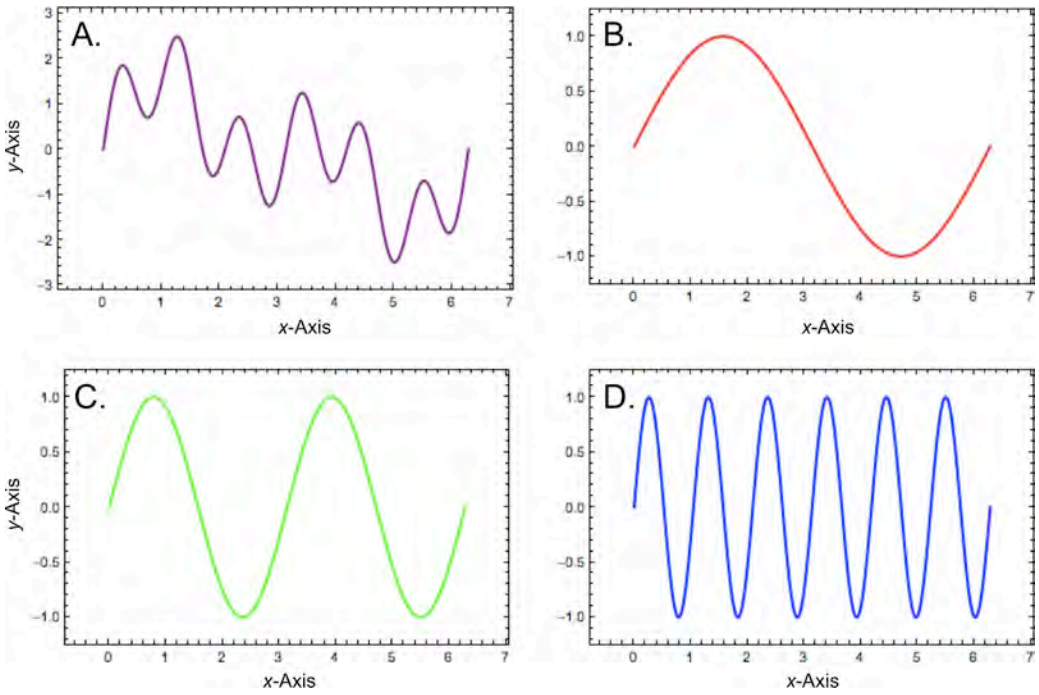

Figure 3. Decomposition of a complex boundary curve in a Cartesian coordinate space (A) into a set of component curves based on the sine function: $y = sin(1t)$, (B) $y = sin(2t)$ (C), and $y = sin(6t)$ (D) where t is allowed to vary continuously between 0.0 and 2π.

The basic method of Fourier analysis is simple enough to demonstrate. Take a complex mathematical function such as that shown in Figure 3A. This curve was formed by adding the three simple sine functions shown in figures 3B-D together. Fourier's contribution was to devise an infinite series of functions that could be used to represent or 'decompose' any complex empirical curve like this into a set of simpler component curves that resemble those shown in figures 3B-D.

Note that, in order to describe the empirical curve, it is not necessary for the decomposition to recover the same curves that were used to construct the complex curve. Indeed, it will usually be the case that the empirical curve is either not constructed out of a set of known curves and/or that the set of curves that contribute to its assembly are very large. From the standpoint of curve description, the recovery of the original generating curves — even if that were possible — is irrelevant. All we need is a method to find a set of component curves that can be used to describe and quantify aspects of the complex curve of interest and that, when combined with other descriptors of the curve, can reproduce the complex original to any desired level of accuracy. Of course, it also helps with interpretability if the series of curves used to describe the complex curve have known and simple geometric relations with one another. This is precisely what Fourier's series provides.

[1] A function is a mathematical relation that uniquely associates the members of one set with the members of another set.

Before we dive into the mathematics of Fourier analysis, let's take care of a little relevance detail. It probably hasn't escaped your notice that the curves shown in Figure 3 don't look very much like the outlines of organisms. Why are these curves - and by implication why is Fourier analysis - of any interest to people like us? The answer is to be found in altering the manner in which we draw the curves shown in Figure 3.

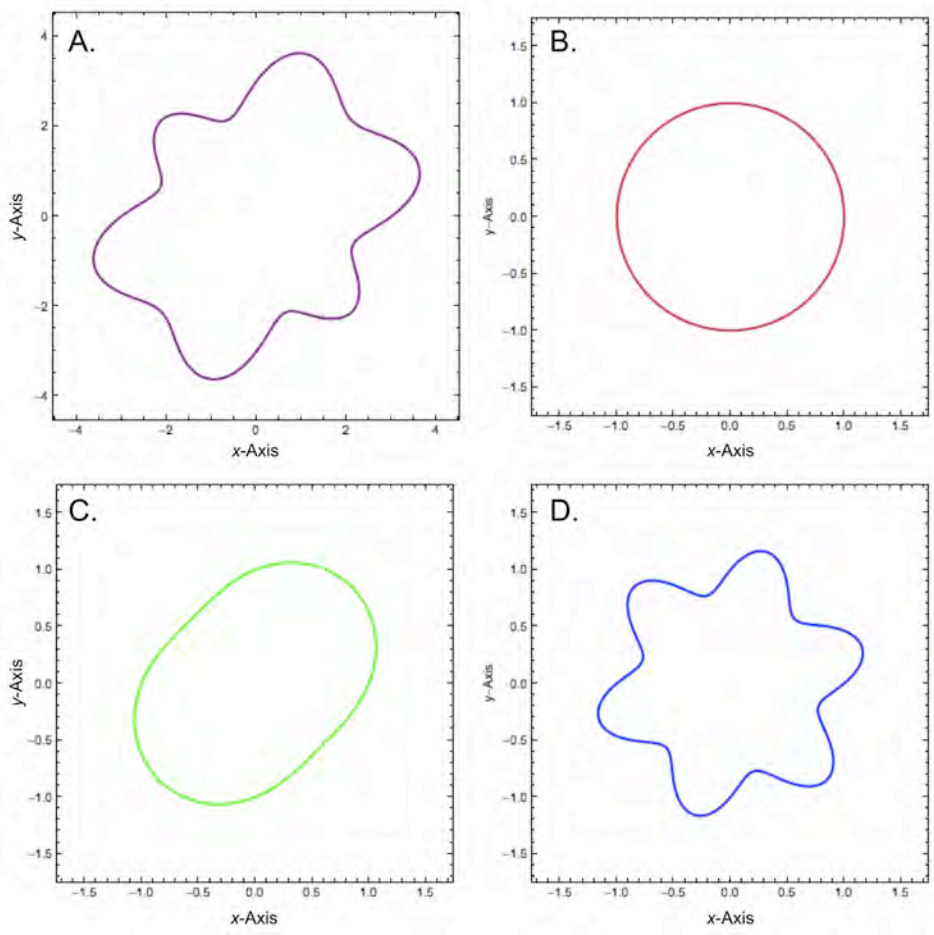

Figure 4. Decomposition of a complex boundary curve in a polar coordinate space (A) into a set of component curves based on the sine function: $y = sin(1t)$, (B) $y = 1+0.5\ sin(2t)$ (C), and $y = 1+0.5\ sin(6t)$ (D) where t is allowed to vary continuously between 0.0 and 2π.

Take a look at Figure 4. These are the equivalents of the equations used to plot the curves in Figure 3 produced this time as polar (rather than Cartesian) coordinates. Recall polar polar coordinates use two numbers to locate points within a linear coordinate system in the same way as Cartesian coordinates. But polar coordinates achieve this location by specifying the angle of a vector from a reference axis and the distance of the point from the coordinate system's origin. The curves in figures 3 and 4 are equivalent representations of the same sine and cosine functions. Fourier analysis can operate on the complex curves show in figures 3A and 4A in exactly the same manner. However, it's obvious that the polar form of a boundary outline can be used to represent a wide variety of biologically interesting outline shapes.

The canonical expression of the Fourier series is usually given as follows.

$$r(\beta) = \overline{r} + \sum_{j=1}^{k}\left[a_j \cos(j \cdot \beta) + b_j \sin(j \cdot \beta)\right] \qquad (22.1)$$

Where:

r = length of radius vector in polar coordinate system

β = angle of radius vector in polar coordinate system

$\overline{r}$ = average of all radius vectors

j = Fourier harmonic number

k = total number of harmonics in Fourier series

a_j = amplitude of the cosine term for the j^{th} harmonic

b_j = amplitude of the sine term for the j^{th} harmonic

This equation can be used to calculate the position of any point in the polar coordinate space that satisfies the equation's conditions. Each term in the summation specifies a geometric figure of whose size is determined by the mean radius vector term ($\overline{r}$) and whose shape is determined by the a_j and b_j coefficients. These control the amplitude of the sine and cosine functions. We can have a look at the shapes that comprise the radial (= polar coordinate) Fourier series by ignoring the summation, holding a_j and b_j constant, specifying different values for the harmonic number (j), evaluating the resulting expression for a continuous set of β-values between 0.0 and 2π (= 360°), and plotting the result. Plots of the first six harmonic shapes for amplitude coefficient values of 0.1 and a mean radius of 1.0 are shown in Figure 5.

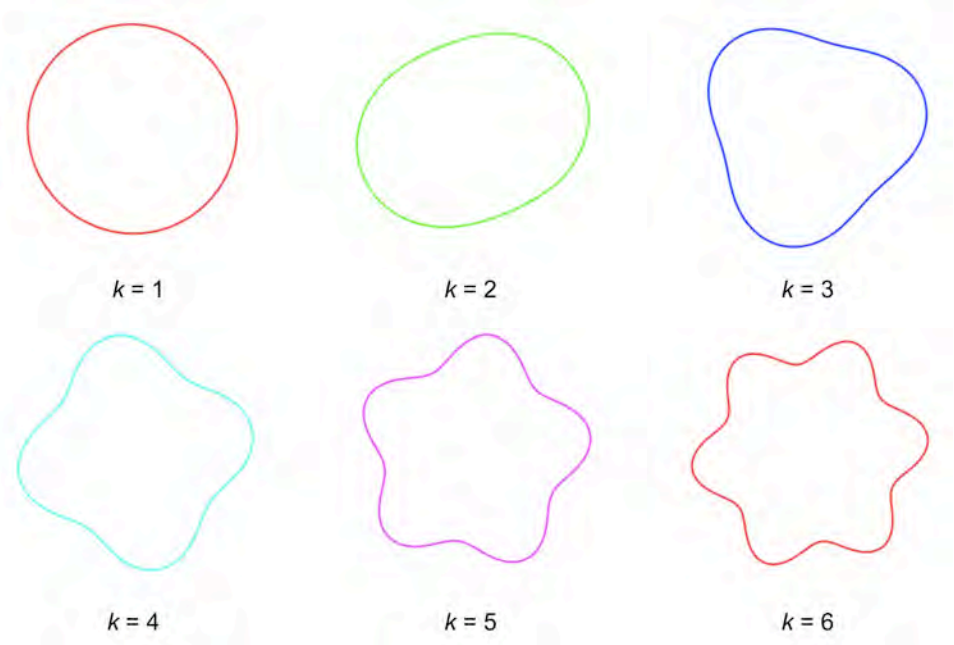

Figure 5. the first six radial Fourier harmonic shapes for the series defined by a mean radius value of 1.0 and amplitude coefficient values of 0.1.

There are several things to note about this set of figures. First, the harmonic number refers to the number of 'lobes' in the shape. The amplitude coefficients control the degree of lobe differentiation. High amplitude values specify deep indentations, low values shallow indentations. These values also control the orientation of the lobes as they can be varied in the manner of weighting coefficients to 'pull' the figure in the direction of the sine or of the cosine components of the curve. Also note that, since k can be given any integer value, the scope of the series is infinite. A k = 60 Fourier harmonic figure will have 60 lobes that (for this set of amplitude coefficient values) would be arranged to 'wave' back and forth within a deviation envelope of 10 percent about the unit circle (Fig. 6).

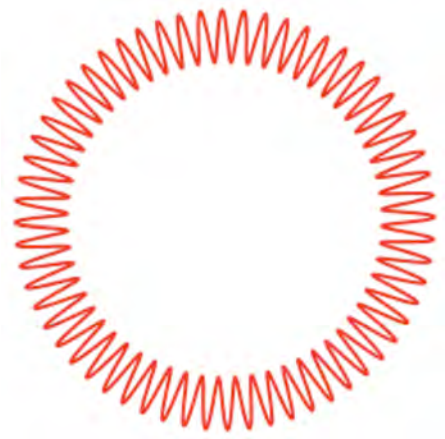

Figure 6. The 60th Fourier harmonic shape for the series defined by a mean radius value of 1.0 and amplitude coefficient values of 0.1.

These fancy mathematical graphics are all well and good. But the real purpose of the Fourier series lies in use the Fourier harmonic figures as variables that describe aspects of shape. Thus, the k = 1 harmonic can be thought of as an index of circularity, the k = 2 harmonic regarded as an index of 'elongatedness', the k = 3 harmonic an index of triangularity, and so forth.

How do we describe shapes using the radial Fourier harmonic series? Essentially we fit — in the sense of a regression analysis — the empirical shape to the set of harmonic figures one-by-one, by adjusting the values of the amplitude coefficients a_j and b_j. This is accomplished using the following equations.

$$a_j = \frac{2}{n}\sum_{i=1}^{n} r_i \cos(j \cdot \beta_i)$$
$$b_j = \frac{2}{n}\sum_{i=1}^{n} r_i \sin(j \cdot \beta_i) \qquad (22.2)$$

Where:

n = total number of points in an empirical curve

r_i = distance between i^{th} point and the curve centroid

j = Fourier harmonic number

β_i = angle of the i^{th} radius vector in polar coordinate system

The resulting amplitude coefficients measure the closeness of fit between the outline as expressed by the n points used to sample the outline and each of the j Fourier harmonic figures. Since each figure in the Fourier series is independent of, and uncorrelated with, all other figures in the series, a set of Fourier harmonics represents a set of j independent and uncorrelated shape variables. In this sense the set of Fourier harmonics are similar to the shape variables produced by an eigenanalysis of the covariance matrix calculated for a distance or landmark data set. However, since the Fourier series is derived from the sine and cosine functions which, unlike covariances, are the same for any sample, Fourier shape descriptors are invariant. In other words, each shape has its own, unique set of Fourier harmonic descriptors — often termed a Fourier spectrum — that can be compared with one another irrespective of what sample they belong to or what other shapes they are associated with in any analysis.

Two last points need to be made regarding Fourier analysis computations before we move on to an example computation. The first is the observation that the maximum number of Fourier harmonics that can be used to describe a shape is set by the number of coordinate points used to describe or sample the shape. This number (k) is determined by the following relation.

$$k_{Maximum} = \frac{n}{2} - 1 \tag{22.3}$$

Since *k* must be an integer, this means that the maximum number of non-zero harmonic amplitudes that can be extracted from any curve represented by *n* points is one less than half of *n*. It is perfectly acceptable to calculate the maximum number of Fourier harmonic amplitudes possible for a set of shapes, but then only use a subset of these for further analysis. The only way to increase the number of Fourier shape descriptors above the maximum for a particular dataset, however, is to resample the boundary outline shapes at a higher resolution.

The second point is that it is often inconvenient to work with two separate amplitude-based shape descriptors. Fortunately, there's no need to do this since Johnson (1944) determined that the radial Fourier series can also be expressed in the following manner.

$$r(\beta) = \overline{r} + \sum_{j=1}^{k} \left[c_j \cos\left(j \cdot \beta + \phi_j \right) \right] \tag{22.4}$$

Where:

$$c_j = \sqrt{\left(a_j\right)^2 + \left(b_j\right)^2} \tag{22.5}$$

$$\phi_j = \tan^{-1}\left(b_j / a_j\right) \tag{22.6}$$

The c_j term includes all the harmonic scaling information present in the separate a_j and b_j terms of equations 22.1 and 22.2. The ϕ_j term (known as the 'phase angle') contains information about the best-fit rotation of the harmonic figure to a position of maximum correspondence with the observed boundary outline. Because of the formulation of the arctangent (= $\tan^{-1}$) function, the following rules should be used to determine the true phase angle.

$$\begin{aligned}
&\text{If } a_j > 0.0 \text{ and } b_j > 0.0, \ \phi_j = \phi_j \\
&\text{If } a_j < 0.0 \text{ and } b_j > 0.0, \ \phi_j = \phi_j + 180° \\
&\text{If } a_j < 0.0 \text{ and } b_j < 0.0, \ \phi_j = \phi_j + 180° \\
&\text{If } a_j > 0.0 \text{ and } b_j < 0.0, \ \phi_j = \phi_j + 360° \\
&\text{If } a_j = 0.0 \text{ and } b_j > 0.0, \ \phi_j = 90° \\
&\text{If } a_j = 0.0 \text{ and } b_j < 0.0, \ \phi_j = 270°
\end{aligned} \tag{22.7}$$

Care must be taken when programming these correction rules as some computer systems and/or programming code compilers report the arctangent value in different ways. Let's now turn our attention from the developing to using this mathematical tool. Figure 7 shows three foraminiferal species that exhibit very different outlines.

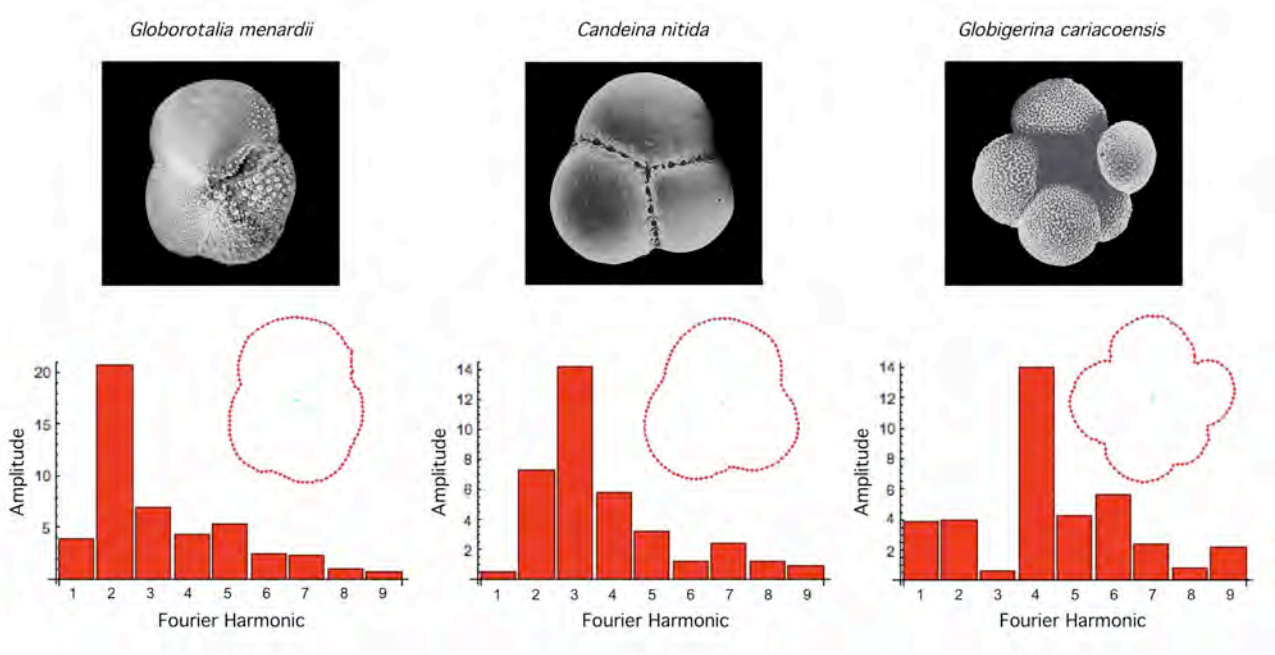

Figure 7. Images, boundary outlines, and Fourier harmonic spectra for three planktonic foraminifer species. SEM images modified from Saito *et al.* 1981.

The bar graphs at the bottom of this figure illustrate the values of c_j for a nine-harmonic Fourier analysis of these species' boundary outlines. Note how each harmonic spectrum reflects the specimen's outline shape. Although the terminal whorl of *Globorotalia menardii* is composed of three chambers, the general aspect of its outline is that of an ellipse with three prominent subsidiary lobes and two smaller, lobes located to the right of the aperture. The geometry of this species' shape is reflected in the very high amplitude of harmonic 2 and relatively high amplitudes for harmonics 3 and 5. The *G. menardii* spectrum contrasts strongly with that of *Candeina nitida*, which is characterized by a high amplitude value for harmonic 3 — reflecting the strong triangularity of this species' test in umbilical view — with subordinate components of elongation (harmonic 2) and quadrateness (harmonic 4). Finally, the shape of *Globigerina cariacoensis* strongly reflects the final four chambers of the ultimate whorl (harmonic 4) with subordinate and subequal components of circularity (harmonic 1), hexagonality (harmonic 6), septagonality (harmonic 7), and nonagonality (harmonic 9). The point, of course, is that the radial Fourier series can be used to describe, characterize, and distinguish between different classes of boundary outline shapes with very high levels of precision.

Note also that these amplitude spectra remain the same regardless of the rotational orientation of the specimen. Information about this aspect of each specimens' pose is contained in the phase angle spectra (not shown). However, if rotational orientation is not a parameter you happen to be interested in, in a rotation-invariant description of shape a geometrically valid analysis can be achieved by ignoring the phase angle data altogether. To be sure, the decision as to which data are needed to answer any biological question must always be made with care. But this useful and mathematically elegant option is a characteristic feature of Fourier analysis.[2] Naturally, the conversion of the original *x,y* coordinate values to polar coordinate values has already taken care of outline translation via centroid superposition while sequestering of the objects size as the mean radius vector effectively normalizes each outline for size differences.

Once a set of boundary outline shapes have been represented, or 'decomposed', into a series of amplitude values, the matrix of these shape descriptors can be subjected to standard multivariate analysis. Figure 8, shoes a series of images of 18 Recent planktonic foraminifer species. Table 1 contains the amplitude spectra for the first nine Fourier harmonic decompositions of the species' boundary outlines. Finally, Figure 9 shows a scatterplot of the distribution of these shapes in the space of the first two principal components of the Fourier harmonic-shape covariance matrix.

[2] If phase angles are to be included in a secondary analysis of a set or sample of Fourier harmonic spectra (see below) it's usually a good idea to represent these angles as radians ranter than degrees to make their magnitudes more comparable to those of the harmonic amplitude term c_j.

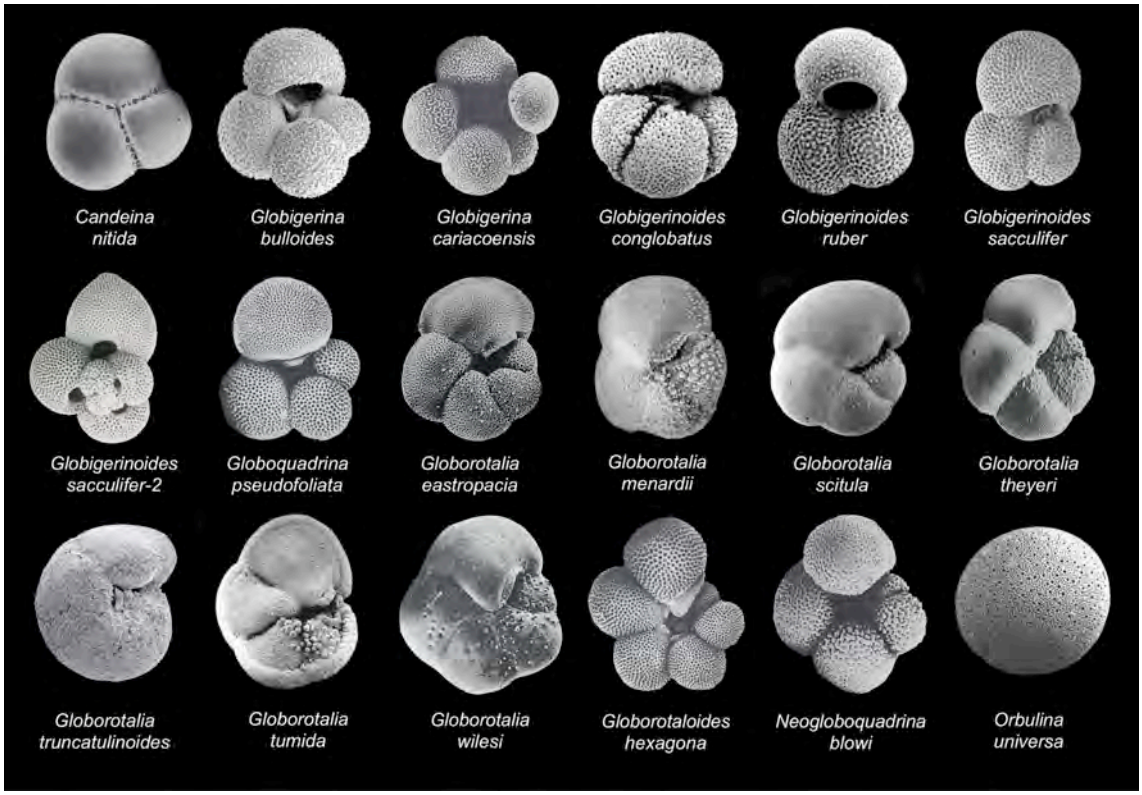

Figure 8. Images of planktonic foraminifer species used in the example radial Fourier outline analysis (modified from Saito *et al.* 1981).

Table 1. Harmonic amplitude values for the boundary outlines of the foraminiferal species shown in Figure 8.

	Fourier Harmonic Amplitudes								
Species	1	2	3	4	5	6	7	8	9
Candeina nitida	0.49	7.31	14.20	5.83	3.19	1.20	2.38	1.25	0.86
Globigerina bulloides	4.63	14.00	12.99	5.06	8.01	5.40	1.20	3.96	0.51
Globigerina cariacoensis	3.86	3.97	0.64	13.96	4.29	5.66	2.41	0.88	2.22
Globigerinoides conglobatus	6.49	13.39	4.28	5.32	6.83	1.54	3.87	2.44	3.51
Globigerinoides ruber	2.31	22.25	12.77	8.75	5.04	1.20	2.56	1.79	1.52
Globigerinoides sacculifer-2	1.89	23.55	21.14	10.11	1.32	2.29	1.71	3.03	2.51
Globigerinoides sacculifer	0.46	25.06	6.09	7.67	2.05	2.65	1.36	0.25	0.63
Globoquadrina pseudofoliata	4.03	17.54	17.94	6.55	6.45	5.42	3.71	2.80	0.63
Globorotalia eastropacia	2.20	14.69	8.56	3.33	6.94	2.91	0.58	2.07	0.93
Globorotalia menardii	3.88	20.65	6.88	4.36	5.27	2.35	2.30	0.97	0.67
Globorotalia scitula	4.63	10.64	5.07	6.11	1.14	3.71	1.57	1.50	1.05
Globorotalia theyeri	6.03	17.43	10.69	4.01	3.37	1.49	2.90	0.70	2.13
Globorotalia truncatulinoides	7.82	12.30	9.45	6.22	4.58	2.86	2.42	1.79	1.66
Globorotalia tumida	4.11	15.65	7.01	2.14	4.14	0.46	0.61	1.18	0.80
Globorotalia wilesi	3.69	8.86	8.67	2.49	4.55	1.88	2.29	0.93	0.99
Globorotaloides hexagona	11.13	26.59	14.82	17.50	8.18	11.31	4.34	0.76	4.08
Neogloboquadrina blowi	3.96	14.56	9.77	10.97	5.19	1.71	1.51	2.84	1.52
Orbulina universa	2.57	5.39	1.12	0.78	0.40	0.06	0.18	0.33	0.48

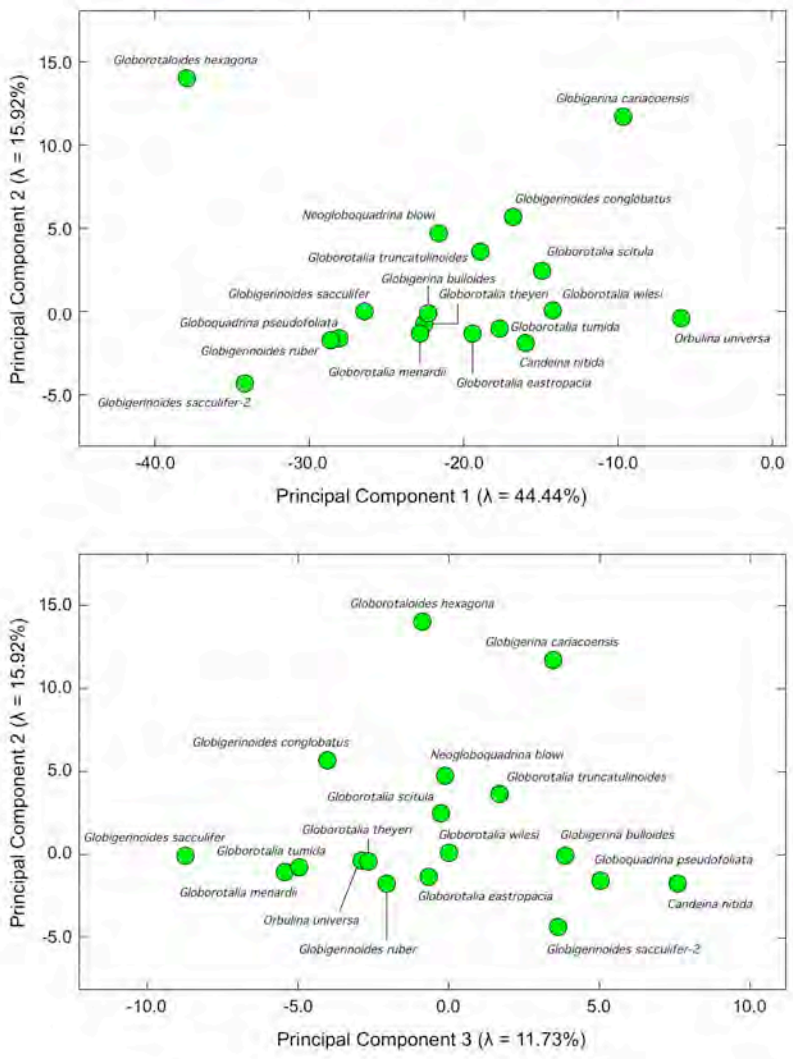

Figure 9. Scatterplots of the projections of the planktonic foraminifera species' boundary outline shapes into the planes formed by the first three principal components of the Fourier harmonic-shape covariance matrix. See text for discussion.

Despite the fact that they represent a subordinate shape variation trend in the example foraminifer dataset, *Globorotaloides hexagona* and *Globigerina cariacoensis* clearly represent shape outliers with respect to this particular species set. A quick inspection of Figure 8 shows these are the two species in the dataset with the most lobulate peripheries. In terms of their Fourier amplitude spectra, both are characterized by relatively high amplitudes for intermediate-level harmonics (e.g., harmonics 5-7, see Fig. 10). This contrasts strongly with the amplitude spectra for species that plot low on the second principal component of the combined amplitude and phase angle spectra (e.g., *Globigerinoides sacculifer*, *Candeina nitida*, *Globigerinoides ruber*), all of which are characterized by quite low amplitudes in this intermediate region of the harmonic spectrum.

Interestingly, the species whose shapes are most intermediate along this axis (e.g., *Globigerinoides conglobatus*, *Neogloboquadrina blowi*) exhibit amplitude spectra that are markedly higher in the lower portion of this intermediate harmonic range (harmonics 3-5). Thus, along this secondary shape variance axis we appear to be seeing the effect of a progressive drift from low values in the intermediate harmonic levels, to a raising of the amplitudes in the low range of this intermediate level that finally culminates in either a broadly elevated, plateau-like spectrum (*Globorotaloides hexagona*) or a distinctly bimodal spectrum *Globigerina cariacoensis*. Of course, the difference between these end-member harmonic geometries is responsible for the separation of these two shape outliers in the PC-1 vs. PC-2 space.

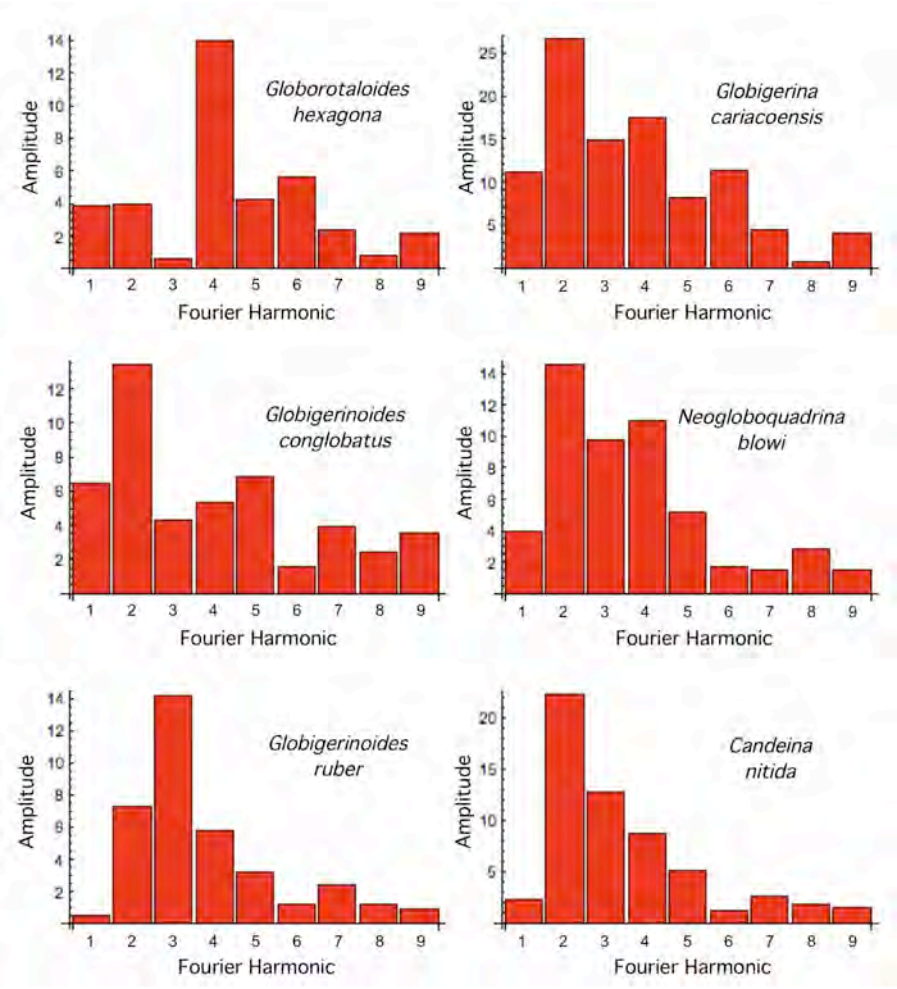

Figure 10. Fourier amplitude spectra for selected foraminifer taxa in the example dataset. Note the drift of middle-range harmonic amplitudes for those shapes that plow at lower (bottom plots), intermediate (middle plots), and high (upper plots) positions along PC-2. See text for additional discussion.

Turning now to the major, but far more subtle, shape variation trend in our example dataset, the difference between *Globigerinoides sacculifer*-2 (a commonly seen variant morph of the typical *G. sacculifer* characterized by a large, but incompletely inflated, ultimate chamber) and *Orbulina universa* is both clear and compelling (see Fig. 8). In terms of the Fourier amplitude spectra for these two taxa the contrast is equally striking (Fig. 10). The other non-outlier species form a quasi-continuous distribution of shapes between these two end-member taxa along PC-1, forming a complex, but quasi-continuous shape variation sequence.

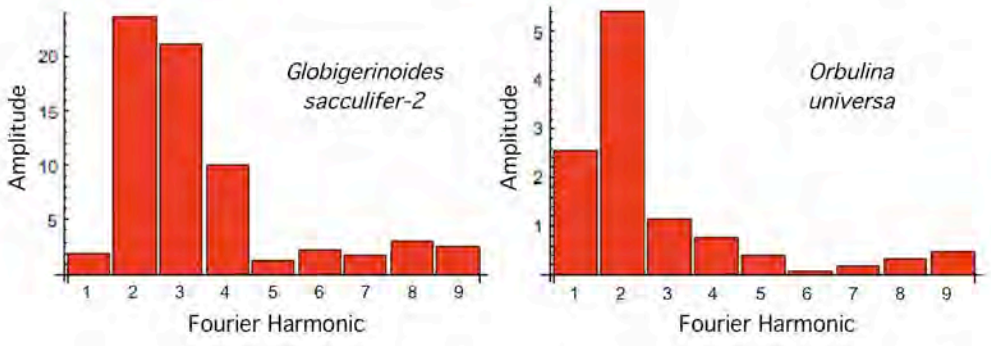

Figure 11. Fourier amplitude spectra for two foraminifer shapes that plot at the extremes along PC-1 and at broadly comparable positions along PC-2. Note the virtually reciprocal pattern of amplitude values characterizing all Fourier harmonics except harmonic 2. Note also the broad discrepancy between the amplitude scales for the two shapes. See text for additional discussion.

While it is possible to make interpretations of the geometric meaning of a Fourier harmonic-based principal component space using a qualitative analysis of extreme taxa, as we have seen above, this task is made much easier and much more precise now that we have access to, and can compare, the harmonic spectra of individual shapes. These shape variables are complex in the sense of being able summarize a much greater amount of geometric information than inter-landmark distance variables or *Procrustes* shape coordinate variables. But at the same time the geometric regularity of the radial Fourier series makes fine interpretation both tractable and intuitive.

Still, we don't need to stop or be satisfied with solely qualitative, or even semi-quantitative, interpretations of these Fourier harmonic-based principal component spaces. In the same way that we were able to use simple matrix algebra to develop a tool that allowed us to obtain precise geometric models of any point in a PCA ordination plot in terms of the original variables (see MacLeod, 2009), we can apply that same technique to provide us with the means to visualize any point, or model any transformation series, within the Fourier harmonic-based principal component spaces. This is a very under-exploited approach to the interpretation of Fourier analysis and one that, because of the richly geometric nature of the radial Fourier series, yields insights that simply cannot be matched by most landmark-based approaches to morphometric analysis, much less those based on inter-landmark distance data. In order to operationalize this technique the data matrix submitted to the PCA must consist of either the a_j and b_j coefficients of the standard Fourier decomposition both harmonic amplitudes and phase angles (see equations 21.1 and 21.2) or the composite harmonic amplitude c_j and associated phase angle ϕ_j (see equations 22.4, 22.5, 22.6). Either will do since equations 22.1 and 22.4 are equivalent.

In order to demonstrate the utility of PCA-based radial Fourier shape models a transect of five shape models was calculated along the mean transects of each of the first three PC axes, with the extremes placed at the extreme positions of the data point cloud and the remaining three models placed as equally spaced positions along each transect. Figure 11 shows the three sets of back-projected shape variation models that, together, express the geometry of shape variation along these axes within the PC-1 - PC-2 - PC-3 subspace in a precise geometric manner. Figure 12 overlays each of the three model sets in the manner of a 'strobe plot (see MacLeod 2009) so that the regions and directions of shape variation along each transect can be assessed and compared in an easy-to-understand manner. Note how much more definite, information-rich, and biologically informative the shape modeling approach to the interpretation of these morphometric data are as well as how much easier it is to us them to communicate the results of such an analysis to non-quantitative colleagues and students.

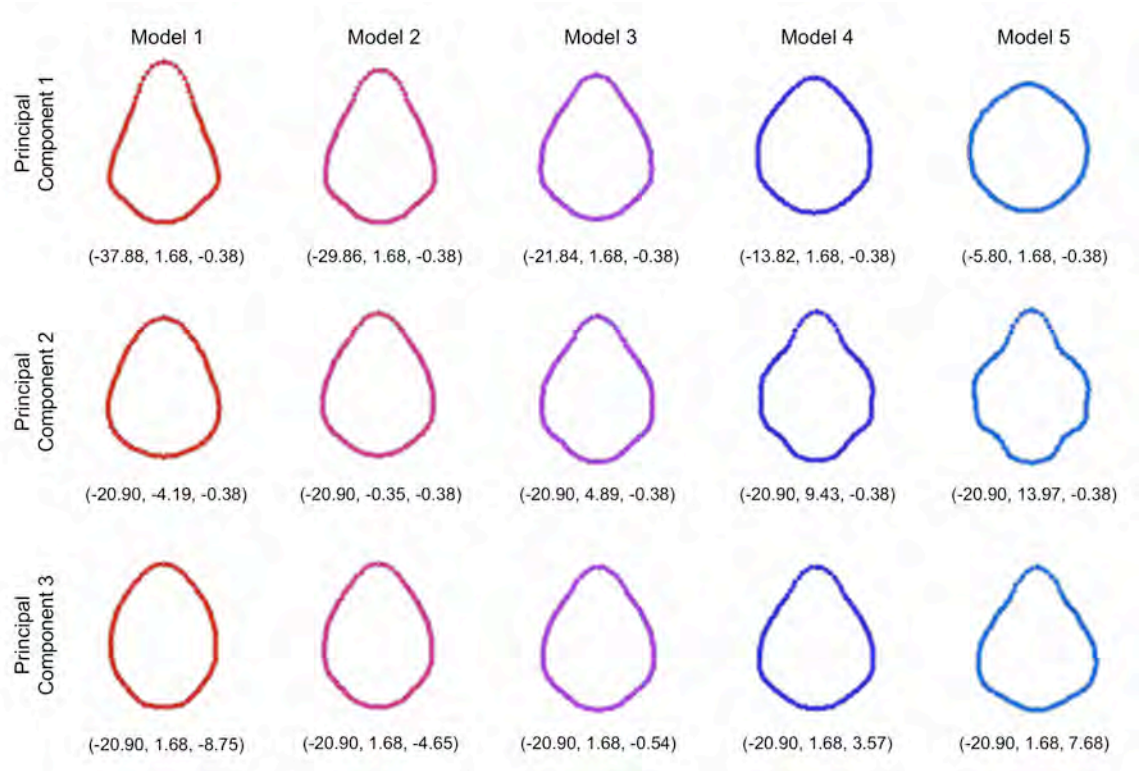

Figure 12. Shape models that corresponding to a series of mean transects through the data point cloud shown in Fig. 9. Coordinates of the modelled positions in the PC-1 - PC-2 - PC-2 subspace are shown below each model. These models were determine using the back-projection method described in MacLeod 2009. Note how much more geometrically precise explicit shape models of the PCA space can be than the (standard) qualitative approach based on extreme objects. See text for further discussion.

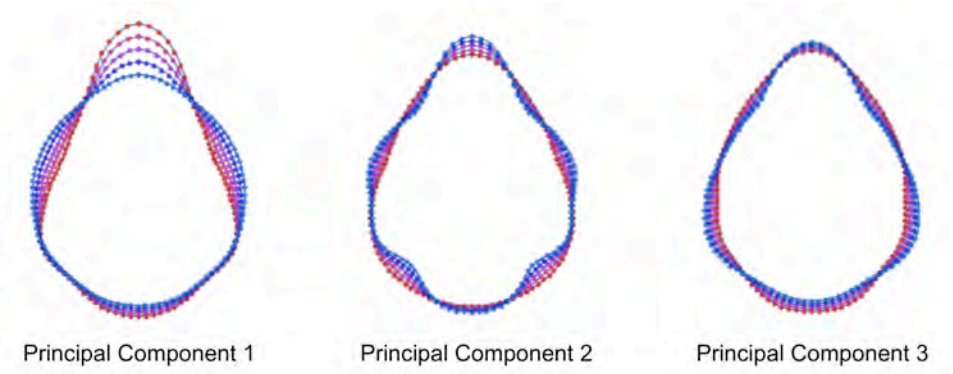

Figure 13. Overlay or 'strobe' plots of the along-axis shape models shown in Fig. 12. These plots are useful for identifying which parts of the outline are moving relative to which other parts as positions within the ordination space change as well as the relative directions of movement. Color codes as in Fig. 13.

In the past many have criticized the use of semilandmark points in any sort of biologically valid morphometric analysis because, with the exception of the starting and/or ending points of the curve, their definition rests wholly on geometric rather than biological criteria. But this stance was always plainly inconsistent with the fact that these same authors advocate the use of Type 3 landmarks which are defined on to-all-intents-and-purposes the same criteria. Moreover, semilandmark-defined boundary curves provide a strategy for quantitatively representing the form and shape of curves that systematists, taxonomists, ecologists, biogeographers, phylogeneticists, etc. use routinely to both characterize and distinguish between individuals and groups in the normal course of their qualitative investigations. Indeed, many taxa are defined — in whole or in part — on the basis of precisely such curves. It has always struck me as odd to regard the analysis of such geometries as perfectly acceptable so long as the analysis remains qualitative, but to disallow it if any attempt is made to quantify the results despite the fact that the mathematical tools required to support such quantification and quantified analysis are not only available, but have been proven time and time again to yield interesting and important results; results unachievable by any other qualitative or quantitative analytic approach.

This is not to say that the use of RFA should be advocated in all instances and/or blindly. Far from it. Radial Fourier analysis is inherently limited, at least in its classic formulation. It can only be used to analyze closed curves and is dependent on accurate location of the curve centroid. The latter limitation is more complex than it appears due to the fact that the calculation of k harmonics requires the specification of $2k+1$ equiangular radius vectors, the estimation of which requires an initial estimate of the outline's centroid. Unfortunately, the centroid of the outline points specified by the resulting interpolated radius vectors may, or may not, coincide with the initial estimate of the raw outline's centroid depending on geometric interactions between the number of boundary outline points used to quantify the curve and the shape of the boundary outline curve itself. In those cases in which the original data estimated, and radius vector estimated, centroids do not coincide, imprecision will be introduced into the Fourier calculations. There are strategies to correct this problem but these will not lead to centroid convergence in all cases.

Even more worryingly, classic RFA can only be applied to single-valued outlines, which are those in which all radius vectors cross the boundary curve just once. Many interesting biological outlines violate this single-value condition. Indeed, this problem is what required substitution of foraminifer for our standard trilobite dataset as the subject of the example calculations. Since there are even many foraminiferal species that exhibit multi-valued outlines, classic RFA could not even be used as a generalized approach to the analysis of the geometry of planktonic foraminifera. We'll be taking a look at the solutions to some of these problems in the next column. Regardless, an understanding of RFA is a very efficient tool to use if your curves of interest are single valued and it represents an obvious starting point for our discussion of the morphometric analysis of outlines.

As for software, surprisingly few computer applications are available for performing this procedure. A fully worked example of all the calculations presented above is provided in the *Palaeo-Math 101-2* spreadsheet. This can be used for simple problems. I have developed Wolfram *Mathematica*™ notebooks for radial Fourier analysis and would be happy to share those with any *Mathematica*™ users. Øyvind Hammer's PAST program package for PC computers (http://folk.uio.no/ohammer/past/) includes a radial Fourier analysis routine. It is likely that public-domain radial Fourier routines also exist for MATLAB, R, and Maple software platforms, but I am unaware of any particularly appropriate or widely used examples. Fortunately, the calculations involved are quite simple and straightforward. If you need a 'quick-and-dirty' analysis, or just

want to play around with the equations to confirm your understanding of the technique, the spreadsheet should suffice.

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

REFERENCES

BLACKITH, R. E. and REYMENT, R. A. 1971. *Multivariate morphometrics*. Academic Press, London 412 pp.

BOOKSTEIN, F., L., STRAUSS, R. E., HUMPHRIES, J. M., CHERNOFF, B., ELDER, R. L. and SMITH, G. R. 1982. A comment on the uses of Fourier methods in systematics. *Systematic Zoology*, **32**, 202–206.

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. 1996. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, **1**, 225–243.

JOHNSON, W. C. 1944. *Mathematical and physical principles for engineering analysis*. McGraw-Hill Book Company, New York 346 pp.

MACLEOD, N. 2009. Form & shape models. *Palaeontological Association Newsletter*, **72**, 14–27.

PIMENTEL, R. A. 1979. *Morphometrics: the multivariate analysis of biological data*. Kendall/Hunt, Dubuque, IA 275 pp.

SAITO, T., THOMPSON, P. R. and BREGER, D. 1981. *Systematic index of Recent and Pleistocene planktonic foraminifera*. University of Tokyo Press, Tokyo 190 pp.

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