Palaeo-Math 101
The Centre Cannot Hold II: Elliptic Fourier Analysis

The Zahn and Roskies (1972) shape function represents a comprehensive, generalized, and highly adaptable method for avoiding the limitations of center estimation when conducting a closed-form outline analysis via Fourier decomposition. However, this approach never really caught on with Fourier analysis practitioners and/or morphometricians. To a large extent this was probably a matter of not catching on with the authors of textbooks on Fourier analysis and the teachers of Fourier analysis methods. Indeed, most of the canonical numerical taxonomy and morphometrics textbooks (*e.g.*, Blackith and Reyment 1971, Sneath and Sokal 1973, Pimentel 1976, Bookstein *et al.* 1985, Zelditch *et al.* 2004) scarcely discuss outline-based methods at all while others have been highly critical of the very idea of trying to analyze outlines *(e.g.*, Bookstein 1991). A lack of easily accessible, public-domain software that could be used for implementing a Fourier analysis based on Zahn & Roskies (ZR) shape functions also hampered its application and so its acceptance. I will return to a discussion of the ZR shape function in the next column when I take up the topic of eigenshape analysis. However, I'd like to close this section on Fourier-based outline semilandmark analysis with a presentation of the most popular current method of implementing a Fourier-based outline analysis, elliptic Fourier analysis (EFA).[1]

Although I've never seen it acknowledged in any description of the EFA method, there is quite a suspicious similarity between EFA and ZR-based Fourier analysis that has always struck me as being suggestive of a conceptual link. In order to illustrate this let's return to the pathological outline of the benthic foraminiferal species *Ramulina globulifera* (Fig. 1).
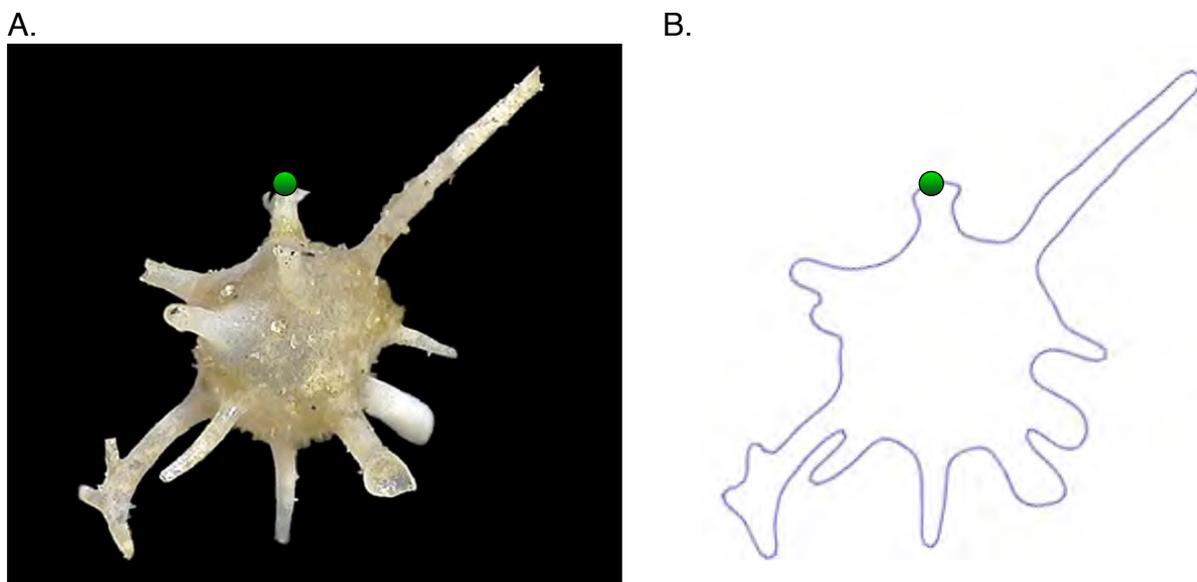
A.

B.

Figure 1. Shape of the spinose benthic foraminifer species *Ramulina globulifera.* A. Image of *R. globulifera* specimen, B. implied outline defined by an interpolation of 300 equally spaced boundary outline semilandmarks. The starting point for digitization is shown by the green landmark.

As you will recall from the last column, the problem with this outline is that it's very far from exhibiting a single-valued character. No matter what center is selected the outline cannot be represented accurately by a shape function based on the lengths of radius vectors drawn from the center to the periphery such that the angle between the radius vectors is constant and all radius vectors cross the outline at one, and only one, point. The ZR shape function solves this problem by interpolating a set of semilandmark points around the outline such that the inter-point distance is constant and the shape represented by a series of net angular deviations from a starting segment. This operation transforms the outline of any curve, no matter how complex, into a pattern that conforms to the definition of a mathematical function (Fig. 2).[2]

---

[1] This technique was originally named elliptic Fourier analysis (Giardiana and Kuhl 1977), but has also been referred to as elliptical Fourier analysis (*e.g.*, Lestrel 1997).

[2] In mathematics a function is a statement or argument in which every input value is associated with one and only one output value.
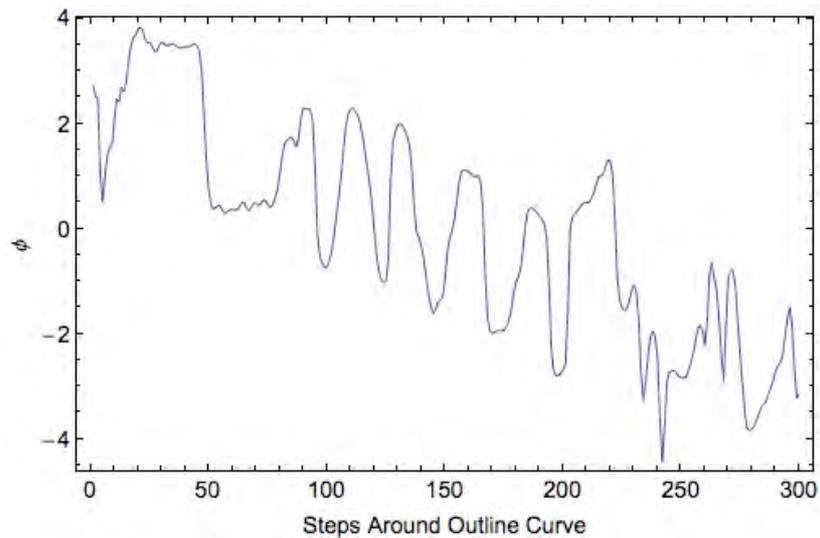
Figure 2. Zahn and Roskies shape function for a 300-point interpolation of the *R. globulifera* outline semilandmark data expressed as the net deviation of angles drawn between successive semilandmark around the form (ϕ).

The Z-R technique transforms a complicated outline into an alternative format that loses none of the outline's geometric information content, but is tractable to handle in the context of a Fourier analysis. It works fine. But this isn't the only solution to complex outline representation.

Note that the abscissa (traditionally the '*x*-axis') of the plot in Figure 2 is simply the ordinal number of intervals round the the outline. There are *n* such intervals in any curve where *n* is the number of semilandmarks used to represent the curve's form or shape. Obviously the ZR shape function is employing variations in the coordinate positions of both the *x* and *y* axes in order to determine the next angular direction to subsequent boundary outline locations (ϕ). But what if we considered these *x* and *y* deviations between successive semilandmark locations separately?

Figure 3 shows the form functions that result from plotting the *x* coordinates and the *y* coordinates of successive semilandmark points separately.
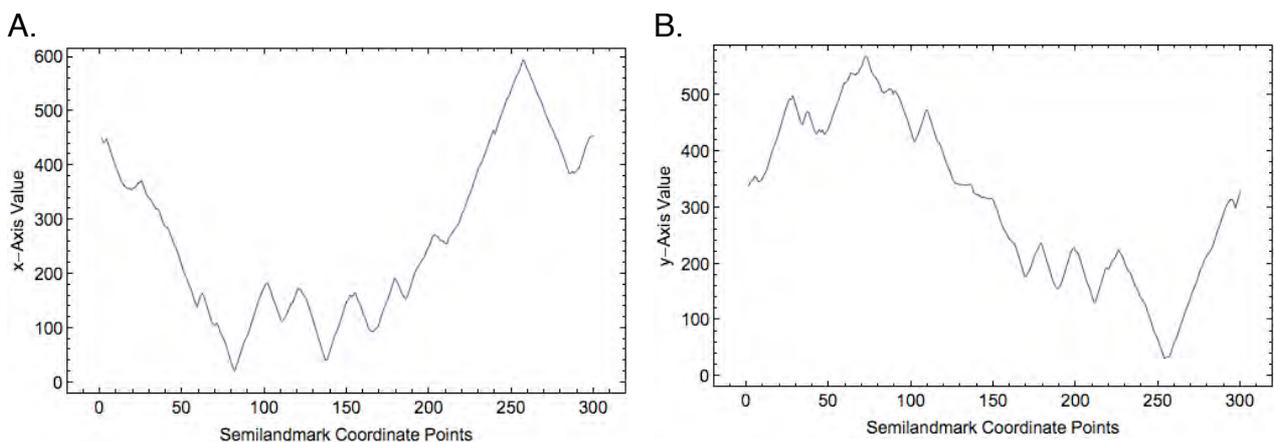

Figure 3. Form functions for separate *x* (A) and *y* (B) coordinate values of the *R. globulifera* outline.

These are form functions rather than shape functions because, as the raw coordinate values were used, the *x* and *y* values retain their scale. Since only one form is shown here the distinction between form and shape is not important. Regardless, it should be noted that, since this representation can utilize the raw coordinate values, it is possible to use this shape function to perform a form analysis or, if the semilandmark points are rescaled to eliminate size differences among a set of outlines, a shape analysis.

Like the ZR shape function, all the geometric information present in the original semilandmark coordinate data is present in these plots. Also as with the Z-R shape function, the advantage of this format is that the resulting data conform to the definition of mathematical functions, and so can be subject to methods of

analysis designed to extract information from function-type datasets. This will always be the case for any conceivable outline; even open outlines and/or those that intersect themselves. Moreover, because these shape/form functions are simply direct readings of variation in the x and y aspects of the outline boundary coordinates, the constraint of equi-angularity (standard radial Fourier analysis) or equal inter-coordinate spacing (ZR Fourier analysis) is relaxed. Forms can be represented by any number of semilandmark coordinate points the analyst deems appropriate and inter-landmark spacings can assume any value(s) necessary to ensure comparable portions of the outline are aligned properly in the semilandmark sequences. The only requirements are that the same number of semilandmarks be used to represent the form/shape of the outlines of all specimens in the sample and (for EFA) that all outlines in the sample be closed (= end at the same landmark coordinate position at which they began). In practice, the ability to vary the inter-semilandmark spacing as a strategy for ensuring biological correspondence between outline segments has rarely been used in any EFA application. But this does not mean it could not be used to boost biologically important signals in the data and improve the interpretability of EFA results.

If an alternative form of outline shape/form function representation was all that was involved with an EFA it would simply represent a minor variant of the ZR shape function approach. However, the developers of this method went further than Zahn and Roskies (1971) and used the separation of x and y form/shape functions as the beginning of a new approach to the Fourier analysis of closed-form outlines. Their key insight in this regard was to think of the total set of length and orientation steps around any outline (t) as being the sum of apparent displacements in x and y directions (see Kuhl and Giardina 1982, Ferson *et al.* 1985, Lestrel 1997).

$$x(t) = \sum_{n=1}^{N} \left[A_n \cos\left(\frac{2\pi nt}{T} \right) + B_n \sin\left(\frac{2\pi nt}{T} \right) \right] \tag{25.1}$$

$$y(t) = \sum_{n=1}^{N} \left[C_n \cos\left(\frac{2\pi nt}{T} \right) + D_n \sin\left(\frac{2\pi nt}{T} \right) \right] \tag{25.2}$$

Where: n = the harmonic number
N = the maximum harmonic number
t = displacement along outline
T = total displacement

This allowed Kuhl and Giardina (1982) to derived a set of parametric equations that could be used to fit a set of Fourier harmonic amplitudes to the outline semilandmark data directly. The elliptic Fourier coefficients for the x aspect of the form/shape function are:

$$A_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} \left[\cos\left(\frac{2\pi nt_p}{T} \right) - \cos\left(\frac{2\pi nt_{p-1}}{T} \right) \right] \tag{25.3}$$

$$B_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} \left[\sin\left(\frac{2\pi nt_p}{T} \right) - \sin\left(\frac{2\pi nt_{p-1}}{T} \right) \right] \tag{25.4}$$

Where: k = the total number of steps around the outline
n = the harmonic number
Δx = the displacement along the x axis between point p and p+1
Δt = the length of the step between point p and p+1
t_p = accumulated length of step segments at point p
T = sum of lengths of all steps around outline

Similarly, the elliptic Fourier coefficients for the y aspect of the form/shape function are:

$$C_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta y_p}{\Delta t_p} \left[\cos\left(\frac{2\pi nt_p}{T} \right) - \cos\left(\frac{2\pi nt_{p-1}}{T} \right) \right] \tag{25.5}$$

$$D_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta y_p}{\Delta t_p} \left[\sin\left(\frac{2\pi n t_p}{T}\right) - \sin\left(\frac{2\pi n t_{p-1}}{T}\right) \right] \tag{25.6}$$

Giardina and Kuhl (1977) showed that equations 25.1 and 25.1 specify an outline-specific series of elliptical shapes of monotonically increasing spatial detail that, when summed together, represent an approximation of the form. This approximation minimizes of the deviation of the empirical outline with respect to the estimated model in a least-squares sense. The EFA outline estimation process can be illustrated by reconstructing the form of the *R. globulifera* outline using different numbers of EFA harmonics (Table 1). All harmonics calculated in this manner are independent from each other. Therefore, the sets of harmonic coefficients representing each can be used as empirical form-descriptors in their own right, either singly or in combination.

Table 1. Sequential reconstruction of the *R. globulifera* form from the four elliptic Fourier terms calculated over different sets of harmonic amplitudes (*n*).

| *n* = 1 | *n* = 3 | *n* = 5 | *n* = 7 | *n* = 9 |

| *n* = 13 | *n* = 15 | *n* = 20 | *n* = 30 | *n* = 40 |
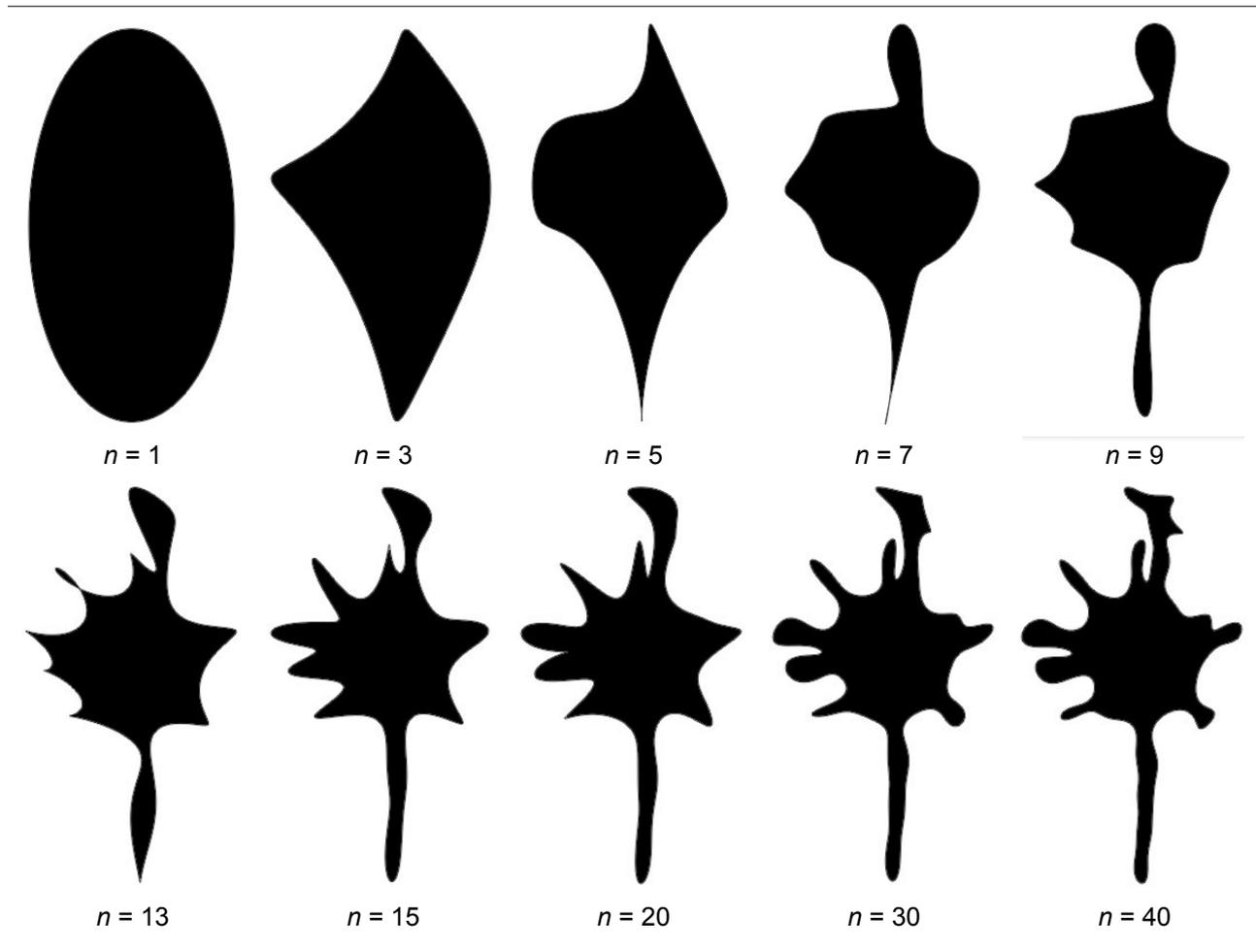
As can be appreciated from Table 1, the four terms of the elliptic Fourier spectrum, calculated over a series of harmonic amplitudes (*n*), are sufficient to represent the form of any outline shape to any desired level of accuracy. Generally speaking, an empirical assessment of the number of harmonics necessary to represent the most complex shape in the sample is made either by eye or by reference to some tolerance criterion (see MacLeod 1999). Each specimen in the sample is then re-described by using this number of EFA harmonics.

Mistakes are often made by inexperienced EFA practitioners (indeed by inexperienced outline data analysts in general), by using arbitrary criteria to decide how many Fourier harmonics to use to quantify the shapes present in a sample. This is especially important in the context of EFA analysis because four terms are required to quantify each harmonic: two terms for the *x* series and two for the *y*. If reduction of the dimensionality of a shape representation problem is the purpose of undertaking an EFA, the most accurate representation of the *R. globulifera* outline presented in Table 4 (*n* = 40) would require 160 terms to represent. While this is a distinct improvement over the 300 *x,y* coordinates used to quantify the outline of

the original shape, it remains a high-dimensional dataset. However, if the number of harmonics used to represent this form was cut back to a lower number — say 10 — in order to achieve substantial dimensionality reduction (from 600 variables to 40), the quality of the form's representation would suffer. Of course, the *R. globulifera* outline is an extreme example. In many cases relatively small numbers of harmonics will be able to represent biological forms of interest (see Ferson *et al.* 1985 for an example). Nevertheless, it is always advisable to make this critical decision on the basis of an inspection of reconstructed forms rather than simply plucking a value for *N* out of thin air.

In addition to containing information about the shape of an an outline, the A_n, B_n, C_n, and D_n coefficients contain information about the location, size, and rotational orientation of the outline. In some cases this information may be of interest. But in most biological morphometric analyses shape variation is the primary target. Although these extraneous variables could always be corrected for through various *ad hoc* normalization procedures or by subjecting the semilandmark sets to generalized least-squares *Procrustes* alignment (see Rohlf and Slice 1990, MacLeod 2009a) prior to EFA, Kuhl and Giardina (1982) specified a direct procedure for normalizing the EFA coefficients calculated from raw outlines for these factors. The basic equation for achieving this normalization is as follows.

$$\begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} = \frac{1}{E^*} \begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \cdot \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} \cdot \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix} \tag{25.7}$$

Obviously several of the terms of this expression need to be explained and defined. The E^* coefficient is the magnitude of the best fitting ellipse. This is, what Kuhl and Giardina (1982) regarded as, effectively, the size term, the influence of which is removed from the data by reciprocal scaling. This E^* coefficient is calculated using the following expressions.

$$E^* = \sqrt{a^{*2} + c^{*2}} \tag{25.8}$$

Where: $$a^* = A_1 \cos\theta + B_1 \sin\theta \tag{25.9}$$

$$c^* = C_1 \cos\theta + D_1 \sin\theta \tag{25.10}$$

Of course, the A_1, B_1, C_1, and D_1 terms in equations 25.9 and 25.10 are the four coefficients associated with the first EFA harmonic which is, in all cases, an ellipse (see Table 1). The θ term is the angle between the starting point of the outline digitization sequence and the major axis of the ellipse. This angle is given by the following expression.

$$\theta = 0.5 \arctan\left[\frac{2A_1 B_1 + C_1 D_1}{A_1^2 + C_1^2 - B_1^2 - D_1^2} \right] \tag{25.11}$$

The θ correction is factor is used in equations 25.7, 25.9, and 25.10 to align the EFA harmonic sequence to an internal standard set by the first harmonic. The final set of EFA coefficients are adjusted to bring the estimated outline model into a standing rotational orientation via calculation of the ϕ coefficient, as follows.

$$\phi = \arctan\left(\frac{c^*}{a^*} \right) \tag{25.12}$$

When these normalizations are applied the values of $a_1 = 1.0$ and those and b_1 and $c_1 = 0.0$ signalling the loss of information inherent in the normalization procedure. This procedure is formally equivalent to the normalizations applied by *Procrustes* superposition. A comparison of the raw and normalized EFA coefficients for the first ten EFA harmonic amplitudes for the *R. globulifera* outline is shown in Figure 4.
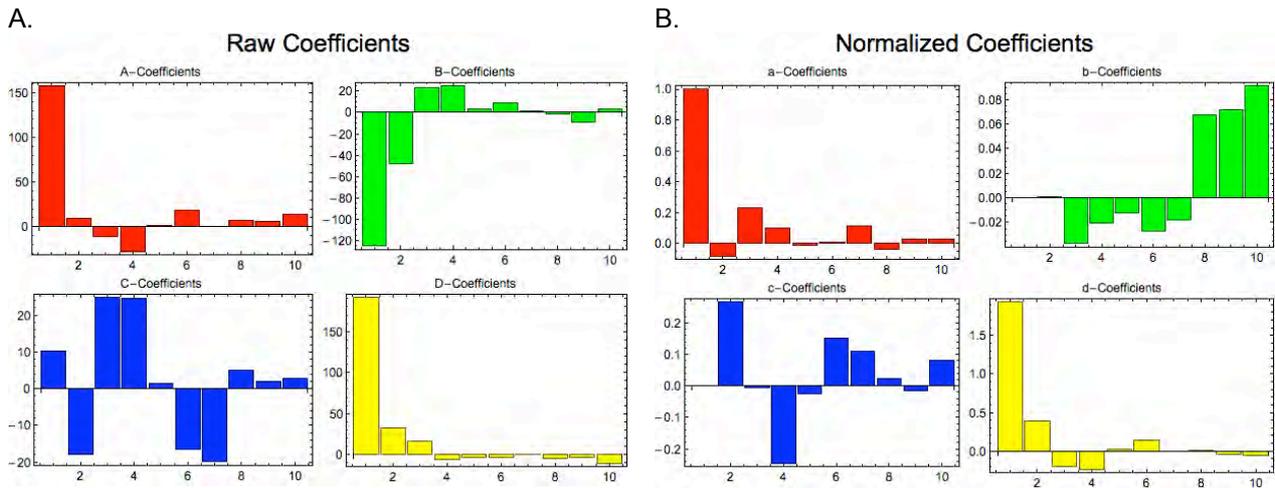
A.

B.

Figure 4. Raw (A) and normalized (B) EFA coefficients for the first 10 EFA harmonic amplitudes of the *R. globulifera* outline. Note rescaling of all harmonic spectra as well as the setting of a_1 to 1.0 and both b_1 and c_1 to 0.0 as a result of the normalization procedure.

While most applications of EFA in morphometric contexts will require implementation of all three normalizations, there may be occasions in which size, location of the starting point for digitization, and/or rotation are parameters of interest. In these cases the form of equation 25.7 can be adjusted so that one or more of these factors remains present in the harmonic spectra.

Once a set of normalized harmonic spectra have been obtained from a sample of shapes these are most typically used as input into a secondary multivariate data analysis procedure (*e.g.*, PCA, PCoord, MDS, CVA, PLS) in order to extract the major dimensions of shape variation, represent patterns of shape variation in a low-dimensional ordination space, investigate questions of group separation or distinctiveness, and/or document the association of particular aspects of outline shape variation with variables external to the morphological data *per se* and, in so doing, test causal hypotheses. As I have noted in previous discussions of these procedures, results generated by their application can be inspected and/or used as input into tertiary statistical procedures designed to test specific patterns via reference to a probability or likelihood models. They can also be used to obtain models of shape variation patterns at particular points along the axes of the ordination spaces so defined, or along any trajectory through the hyper-dimensional ordination space that the data analyst cares to specify.

As an example of the use of EFA in the context of a morphometric investigation I will apply it to the same benthic foraminiferal test outlines I used in the last column to illustrate use of the ZR shape function with standard Fourier analysis. The images from which these outlines were derived are shown in Figure 5.
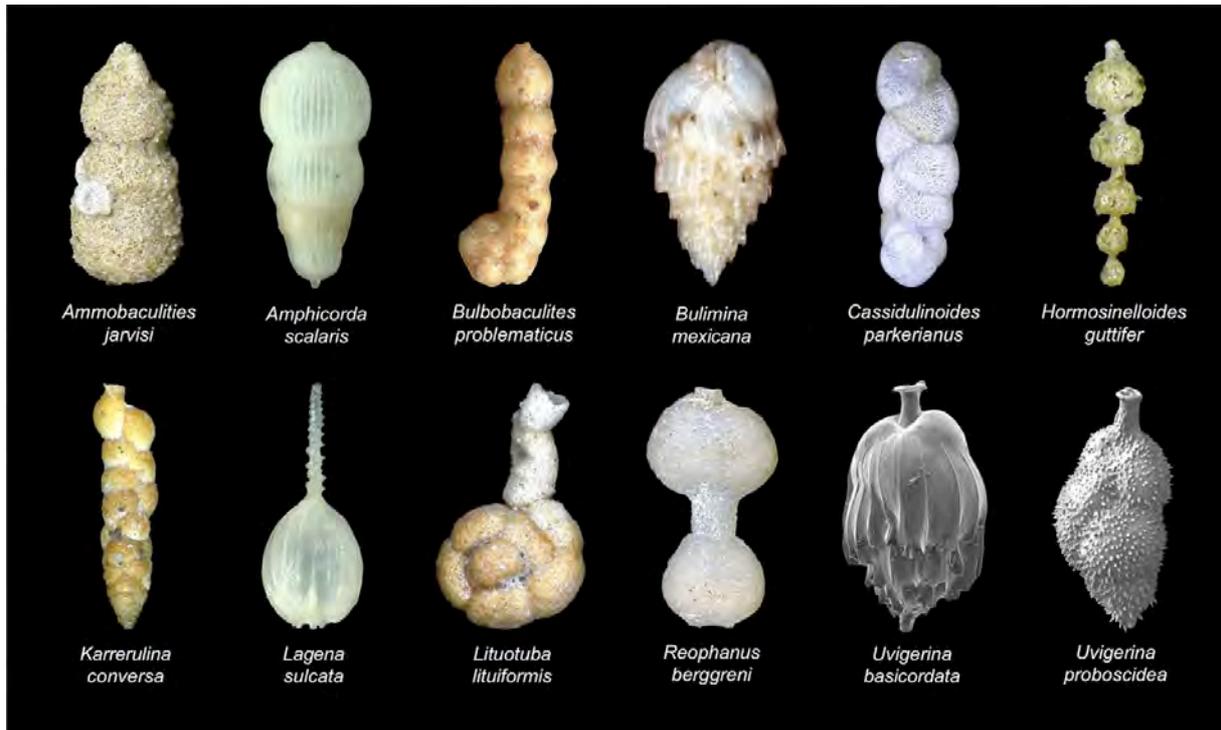
Figure 5. Benthic foraminiferal shapes used in the example analysis. Note the presence of shapes with multi-valued outlines.

Each outline was quantified using 200 equally-spaced semilandmark points with the landmark from which digitization started being located at the center of the aperture. Inspection of outline morphologies reconstructed using different numbers of EFA harmonic amplitudes indicated that, for this dataset, the major features of each outline could be captured using the first 25 normalized elliptic Fourier harmonics (see Table 2).

Table 2. Outline silhouettes for benthic foraminiferan shapes reconstructed from normalized EFA coefficients using 25 harmonic amplitudes. See Figure 4 for species names.
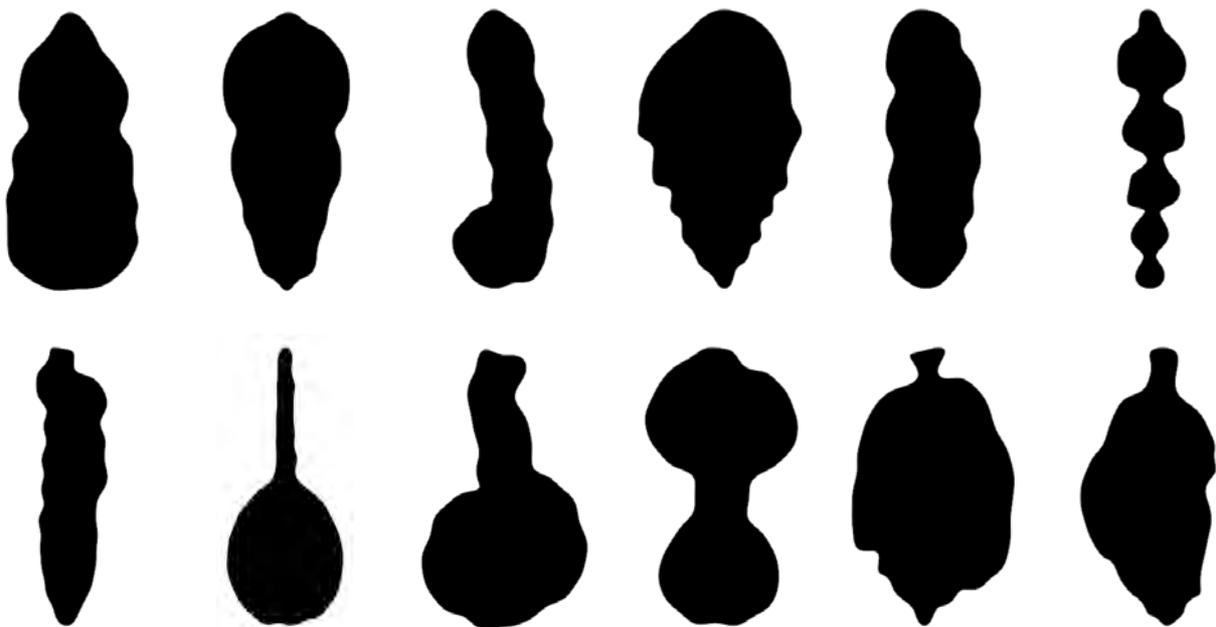
These EFA coefficients were then assembled into a data matrix. This matrix did not include the a_1, b_1, and c_1 amplitude coefficients as these were set to constant values of 1.0, 0.0, and 0.0 respectively as a result of the size, position, and rotation normalization procedure (see above). Overall, the EFA procedure resulted in

these outlines being described by (25 x 4) - 3, or 97 variables. These were used as input to a covariance-based principal components analysis (PCA) whose purpose was to assess the major modes of shape variation in the sample and establish a low-dimensional ordination space within which patterns of shape similarity and difference could be assessed. Results of the PCA analysis indicate that over 90 percent of the observed outline shape variation can be represented on three orthogonal principal component axes. Ordinations of the outline shapes within this three-dimensional subspace are shown in Figure 6.
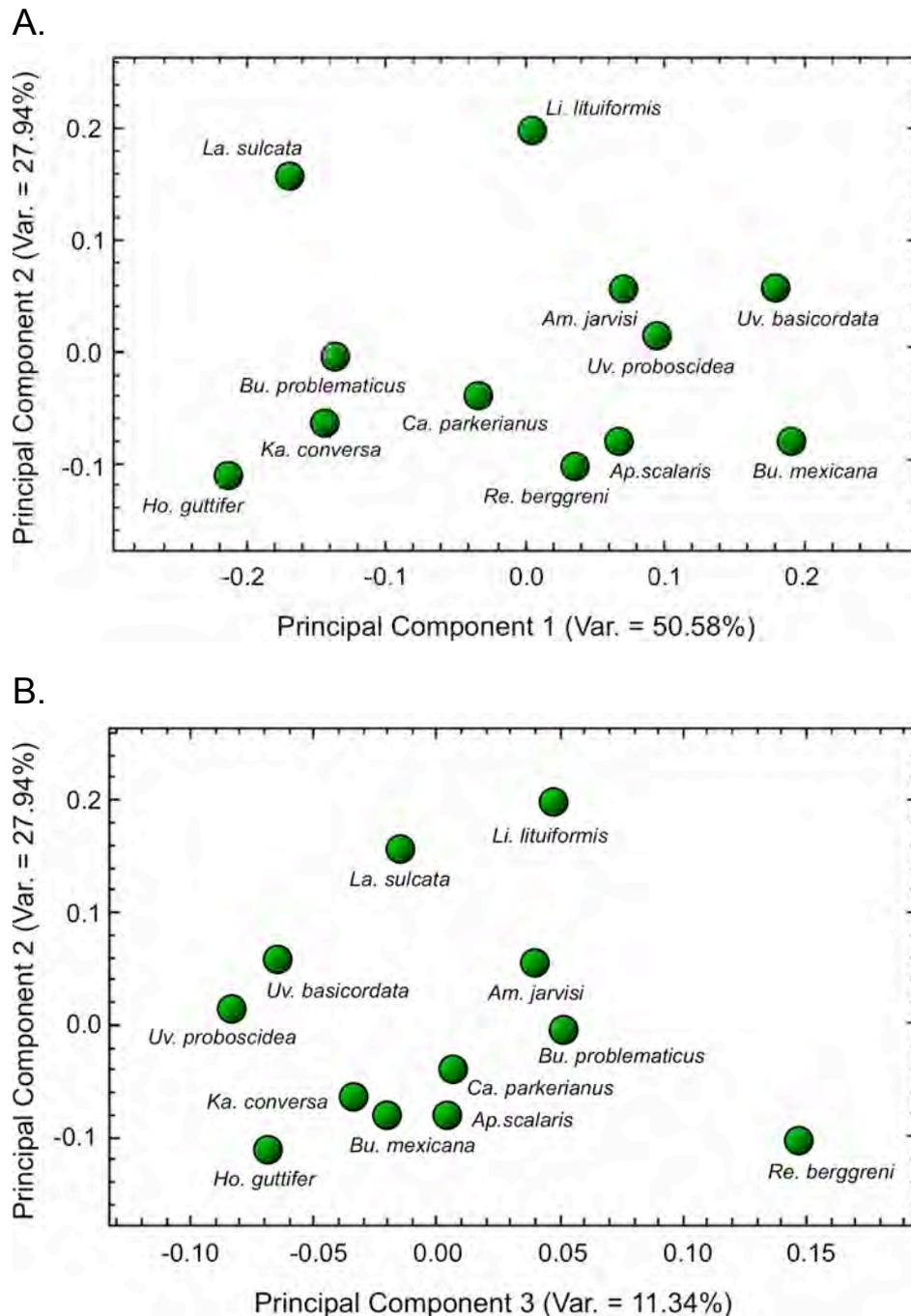
A.

B.

Figure 6. Distribution of the 12 benthic foraminiferan test outlines shown in Figure 5 and Table 2 in the shape space formed by the first three principal components of the covariance matrix calculated between normalized EFA coefficients. See text for discussion.

Inspection of the distribution of species along these three principal component axes suggests that PC-1 captures the distinction between tests characterized by uniserially arranged globular chambers (*e.g.*, *Hormosinelloides guttifer*) and flaring tests with multi-serially arranged chambers (*e.g.*, *Bulimina mexicana*), PC-2 the distinction between tests whose chambers increase in size with development (*e.g.*, *Hormosinelloides guttifer*) as opposed to those whose tests are characterized by a relatively thin, tubular neck (*e.g.*, *Lituotuba lituiformis*), and PC-3 the distinction between tests with complex, lobulate peripheries

that are thick in the middle (*e.g.*, *Uvigerina proboscidea*) relative to those characterized by a medial constriction (*e.g.*, *Reophanus berggreni*). It's also important to note that, in terms of outline shape variation, for this sample *Lagena sulcata*, *Lituotuba lituiformis* and *Reophanus berggreni* all represent shape outliers.

While sufficient for a interpretation of the generalized geometry of the principal component spaces, the qualitative trends described above fail to represent the detailed geometric contrasts that exist along the primary PCA vectors, information which has been captured by the original outline data and employed by the eigenvector equations that define PC-1, PC-2, and PC-3. Fortunately, we are not restricted to such qualitative approaches to EFA shape space interpretation.

The eigenvector loading coefficients that define PC-1, PC-2 and PC-3 represent a series of cosines that quantify the angular relationship between these eigenvectors and the original variables in a manner that takes note of covariance structure of these data (see MacLeod 2005). Since this is the same information we used previously to construct geometric modes of PCA form/shape spaces for landmark (MacLeod 2009b, 2010a) and outline data (MacLeod 2011), it should be possible to create similar models for the EFA-defined PCA ordination spaces shown in Figure 4. It's possible to calculate these models using either the direct landmark or thin plate spine approaches (see MacLeod 2009b and 2010a respectively), though the former seems the more natural and informative choice for these data. The models calculated by this method will be expressed in terms of the original variables submitted to the EFA-PCA analysis. These may be converted back to sets of *x,y* coordinate data using the following equations.

$$x_n = \sum_{n=1}^{N} a_n \cos(nt) + b_n \sin(nt) \qquad (25.13)$$

$$y_n = \sum_{n=1}^{N} c_n \cos(nt) + d_n \sin(nt) \qquad (25.14)$$

Where: n = harmonic amplitude
N = maximum no of harmonic amplitudes used in the construction
t = evaluation angle (varies from 0 to 2π)

Application of equations 25.13 and 25.14 to sets of EFA amplitude coefficients calculates at a series of coordinate point locations along the PC-1, PC-2 and PC-3 axes resulted in construction of the along-axis shape models shown in Table 3.

Table 3. Elliptic Fourier amplitude coefficient models existing are coordinate locations along the first three principal component axes of the outline shape covariance matrix. The specific coordinate position at which each model was calculated is shown below each model (in parentheses).

PC Axis	Model 1	Model 2	Model 3	Model 4	Model 5	Overlay
PC-1	(-0.20,0.00,0.00)	(-0.09,0.00,0.00)	(0.01,0.00,0.00)	(0.11,0.00,0.00)	(0.21,0.00,0.00)	
PC-2	(0.00,-0.20,0.00)	(0.00,-0.12,0.00)	(0.00,-0.04,0.00)	(0.00,0.03,0.00)	(0.00,0.11,0.00)	

PC-3					
(0.00,0.00,-0.15)	(0.00,0.00,-0.09)	(0.00,0.00,-0.03)	(0.00,0.00,0.03)	(0.00,0.00,0.08)	
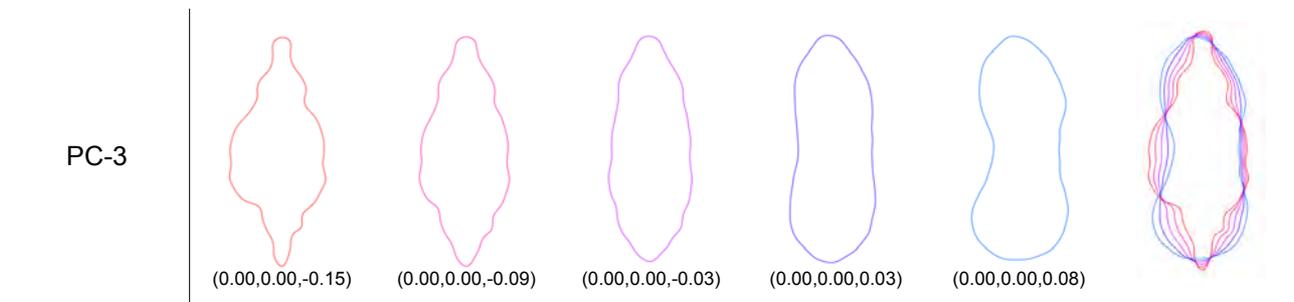
As can be seen graphically in the table — especially when the individual along-axis shape models are overlain to create a 'strobe' effect — the geometric trend being expressed along each axis is a function of (1) the degree of inflation (PC-1) and lobateness PCs 2 & 3) of the test peripheries on either side of the test's long axis. Based on these results the distinction between low scores and high scores along PC-1 reflects the distinction between broadly elliptical shapes with concave to convex lateral margins respectively. The PC-2 axis captures aspects of spatial changes in lateral lobateness with respect to the test's long axis, with low scores representing lobe formation confined to late stages of test development and high scores representing patterns of lateral lobe formation that occur at earlier developmental stages. Finally, PC-3 expresses distinctions between the lobateness of the medial portion of the test with respect to the two ends. Here, tests characterized by strongly convex medial outlines sandwiched between distinctly pinched terminal regions project to low positions along this axes while those characterized by bulbous terminal regions separated by a narrow, concave, medial lobe project to high positions. The dramatic increase in geometric insight into the natural of EFA ordination spaces that can be achieving through use of models to query and interpret the EFA-PCA is (I hope) obvious.

Like Z-R Fourier analysis, elliptic Fourier analysis effectively addresses the need for a numerical method that can be used to represent outline form/shape variation in a rigorously quantifiable, geometric manner while overcoming the limitations of radial Fourier analysis in terms of its ability to treat single and multivalued outlines and its freedom from being subject to the centroid-location problems which plague the latter, at least with respect to some outlines. Elliptic Fourier analysis is fully extensible to the treatment of 3D outlines (see Lestrel 1997). Owing to its dependence on ellipses EFA is technically restricted to the analysis of closed outlines. However, Kuhl and Giardina (1982) include example analysis of forms that come very close to being open curves in their discussion of EFA applications, though the ability of EFA to reconstruct such curves seems decidedly suboptimal to me.

Owing to the availability of public-domain software for performing EFA, the technique has proven reasonably popular with morphometricians despite the fact that much of the early geometric morphometric literature was rather hostile to outline analysis methods in general and Fourier-based methods in particular. There were two reasons for this. The first has to do with non-correspondence between the biological features that sequences of semi-landmark points may fall on when an outline is sampled. This is a well-known problem that I'll have more to say about in my discussion of eigenshape analysis. In terms of sampling the form, EFA can circumvent this issue rather easily because it does not assume that the spacing between semi-landmark points will be equal. Accordingly, it is a relatively simple matter to subdivide the outline into segments at true landmarks and use semi-landmark sequences to 'join' the landmarks in a matter that accurately represents for form/shape of of the inter-landmark boundary curves (see MacLeod 1999 for an example of this procedure). Data collected in this way can be submitted to an EFA analysis without any further processing. However, this sampling protocol has rarely been used with EFA despite the clear advantages of doing so.

The second criticism that has been made of the EFA (and all Fourier-based) procedures is that they unnecessary. The point of all Fourier analyses is to estimate the form of a signal — in the case of morphometrics a morphological signal. The Fourier approach provides sophisticated and mathematically elegant tools with which to do this. But that's all they do. Fourier harmonic amplitudes and phase angles provide a means to describe outline form/shape variation in quantitative terms. But sample ordinations, the major directions of sample form/shape variation, discrimination between sample sub-groups, comparison of patterns of co-variation, form/shape modelling, statistical testing of various sorts, etc. all must be performed on Fourier-based re-descriptions of morphological variation by other procedures. This raises the question of why to take one's morphological data through a Fourier analysis in the first place as opposed to working directly with the semilandmark outline coordinates themselves as these can easily be aligned and normalized for various sources of extraneous information using other methods (e.g., Bookstein shape coordinates, *Procrustes* shape coordinates).

As an experiment to address this question empirically we can perform a quick analysis of the same benthic foraminiferal outline data using the *Procrustes* PCA technique. Results of the projection of the outline shape

coordinate data into the spaces formed by the first three eigenvectors of the shape-coordinate covariance matrix the are shown in Figure 7. Close comparison of figures 6 and 7 reveals a few differences (e.g., position of *Am. jarvisi* along PC-3) which are to be expected considering the differences between the EFA and *Procrustes* PCA datasets. However, both the gross and detailed similarities between the two ordinations are even more striking. It is doubtful that routine interpretations of these two spaces would differ substantially.
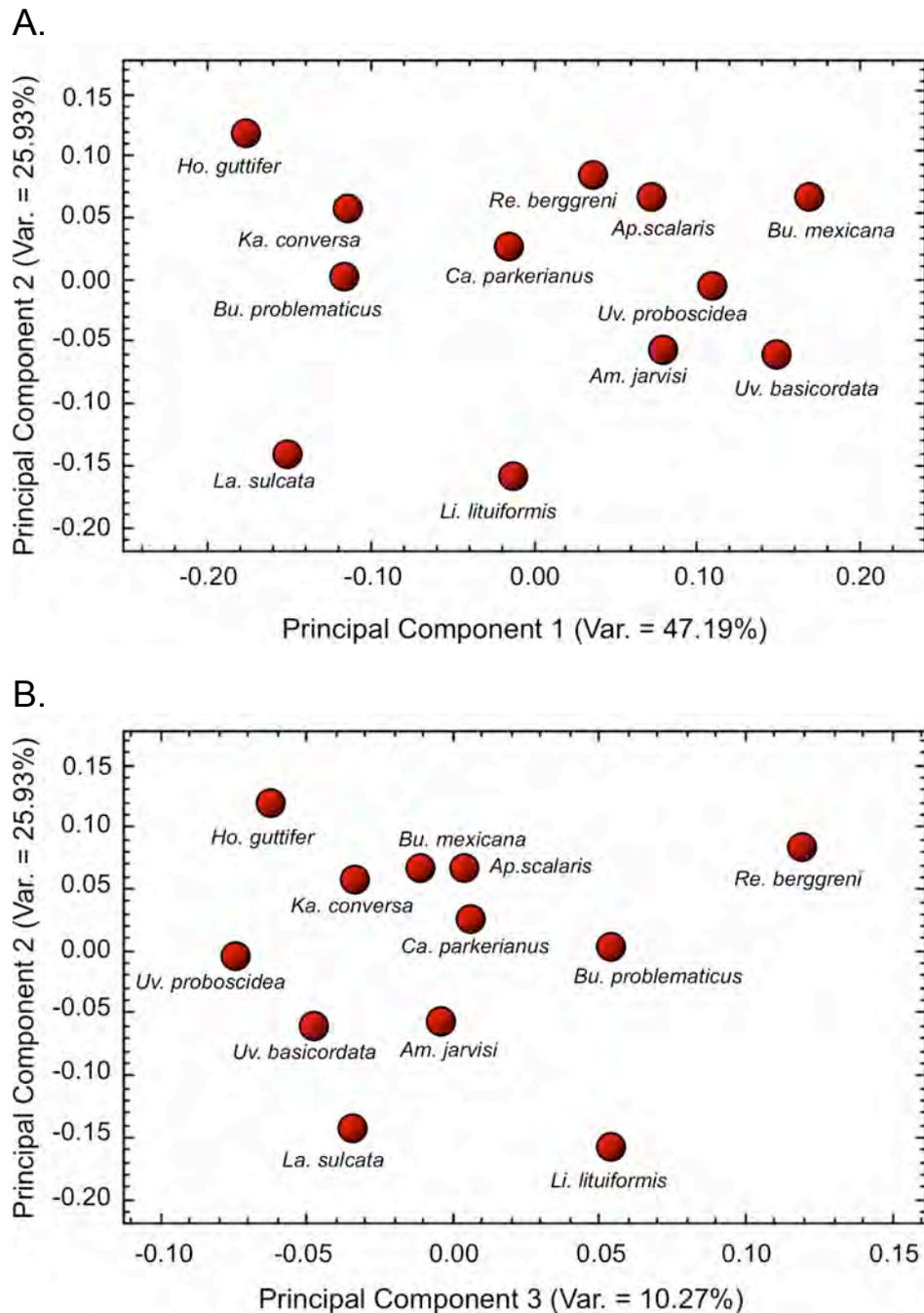
A.

B.

Figure 7. Distribution of the 12 benthic foraminiferan test outlines shown in Figure 5 and Table 2 in the shape space formed by the first three principal components of the covariance matrix calculated between *Procrustes* superposed shape coordinates. See text for discussion.

Rohlf (1986) argued that Fourier analysis might be preferred to the analysis of coordinates because use of only the first few Fourier harmonics, in effect, causes the outline to be smoother — thus eliminating small scale variation that may not (or may) be of importance in any given analysis. If a sub-set of EFA harmonics are selected for analysis a smoothing of the represented outline would certainly be the case. But Rohlf's argument strikes me as one that misses the point. If smoothing of an outline is acceptable or important there are many approaches to this operation that could be employed as an initial data-processing step. The most

obvious of these is to reduce the number of coordinate points used to represent the boundary form/shape to the minimum needed to achieve an objective geometric tolerance (see MacLeod 1999 for an example algorithm). When this is done it is often surprising how few boundary outline semilandmarks are really necessary to replicate even complex outlines. But once this is done the difference between raw coordinates, Z-R Fourier and EFA become apparent. Taking (say) 50 semilandmarks as an example, the raw analysis of coordinates would imply the analysis of 50 *x* and 50 *y* coordinates, 100 variables in total. Under the Z-R Fourier approach 25 amplitudes and 25 phase angles or, more commonly, just the 25 unique harmonic amplitudes would be required, totalling to either 50 or 25 variables respectively. But under EFA each of the 25 unique Nyquist harmonics would require four Fourier coefficients to represent the series, specifying either 100 or 97 variables in total, depending on whether these coefficients were normalized. Of course, the data analyst could simply throw all these EFA variables into a PCA analysis and discard all but the first few principal components as being unimportant. But this standard dimensionality-reduction technique applies equally well to PCA results from raw coordinates, Z-R Fourier shape functions, and indeed to radial Fourier data.

Quite clearly the jury remains out on the question of whether any real advantages are gained by performing a Fourier analysis as a processing step that re-describes the shape in and alternative quantitative format. In this way Z-R Fourier and EFA are reminiscent of principal warps analysis (see MacLeod 2010b). If you are considering using EFA for an outline study I encourage you to explore some of its more unique, yet little-used, variants (*e.g.*, the ability to analyse non-uniformly spaced semilandmarks) in order to add useful and distinctive features to your investigation's design. But more than this I would recommend you defer the analysis until we have had an opportunity to consider a far more flexible outline sampling - data analysis procedure in detail: eigenshape analysis.

In terms of software, a number of different outline processing, outline digitization, and EFA applications for 2D and 3D data are available from the SUNY Morphometrics web site: http://life.bio.sunysb.edu/morph/. The industry standard in this area, however has to be Jim Rohlf's EFA application for Windows operating systems. Among the more comprehensive software packages only Jim Rohlf's *NTSYS* (http://www.exetersoftware.com/cat/ntsyspc/ntsyspc.html) and Øyvind Hammer's *Past* (http://folk.uio.no/ohammer/past/) programme packages have implemented forms of the EFA procedure. All the analyses and graphics used in this column (including the EFA models) were calculated using custom software I have written for Wolfram Researcher's *Mathematica™* software, any or all of which I'd be happy to share on request.

Norman MacLeod
N.MacLeod@nhm.ac.uk

Palaeontology Department, The Natural History Museum
Department of Earth Sciences, University College London
Nanjing Institute of Geology & Palaeontology, Chinese Academy of Sciences

References

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

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

BOOKSTEIN, F., CHERNOFF, B., ELDER, R., HUMPHRIES, J., SMITH, G. and STRAUSS, R. 1985. *Morphometrics in evolutionary biology*. The Academy of Natural Sciences of Philadelphia, Philadelphia 277 pp.

FERSON, S., ROHLF, F. J. and KOEHN, R. K. 1985. Measuring shape variation of two-dimensional outlines. *Systematic Zoology*, **34**, 59–68.

GIARDINA, C. R. and KUHL, R. K. J. 1977. Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Computer Graphics and Image Processing*, **6**, 277–285.

KUHL, R. K. J. and GIARDINA, C. R. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, **18**, 236–258.

LESTREL, P. E. 1997. Introduction and overview of Fourier descriptors. *In* P. E. Lestrel (ed). *Fourier Descriptors and Their Applications in Biology*. Cambridge University Press, Cambridge, 22–44 pp.

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

MacLEOD, N. 2005. Principal components analysis (eigenanalysis & regression 5). *Palaeontological Association Newsletter*, **59**, 42–54.

MacLEOD, N. 2009a. Who is Procrustes and what has he done with my data? *Palaeontological Association Newsletter*, **70**, 21–36.

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

MacLEOD, N. 2010a. Shape models II: the thin plate spline. *Palaeontological Association Newsletter*, **73**, 24–39.

MACLEOD, N. 2010b. Principal & partial warps. *Palaeontological Association Newsletter*, **74**, 35–45.

MacLEOD, N. 2011. Semilandmarks and radial Fourier analysis. *Palaeontological Association Newsletter*, **76**, 25-42.

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

ROHLF, F. J. 1986. Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. *Mathematical Geology*, **18**, 846–854.

ROHLF, F. J. and ARCHIE, J. W. 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Zoology*, **33**, 302–317.

ROHLF, F. J. and SLICE, D. 1990. Extensions of the *Procrustes* method for optimal superposition of landmarks. *Systematic Zoology*, **39**, 40–59.

SNEATH, P. H. A. and SOKAL, R. R. 1973. *Numerical taxonomy: the principles and practice of numerical classification*. W. H. Freeman, San Francisco 573 pp.

ZELDITCH, M. L., SWIDERSKI, D. L., SHEETS, H. D. and FINK, W. L. 2004. *Geometric morphometrics for biologists: a primer*. Elsevier/Academic Press, Amsterdam 443 pp.