

Palaeo-Math 101

Going Round the Bend: Eigenshape Analysis I

Elliptic Fourier Analysis (EFA) has proven to be popular among morphometricians who need a technique that can deliver analyses of curves defined by sets of semilandmarks. In no small part this has been due to the provision of software for, and general championing of, the method, by F. James Rohlf (see Ferson *et al.* 1985; Rohlf 1986,1990). However, when examined in detail EFA has several limitations that constrain the domain of shape analysis problems it can be used to address. Foremost among these is the standard Fourier analysis constraint of requiring the curve to conform to the concept of a periodic function, a function that repeats its values at regular intervals or periods. In terms of outline analysis this effectively means that EFA — like all Fourier techniques — is designed primarily to analyze closed curves with the shape function's period representing one complete trip around the outline (Fig. 1). But not all biologically interesting curves are closed.¹

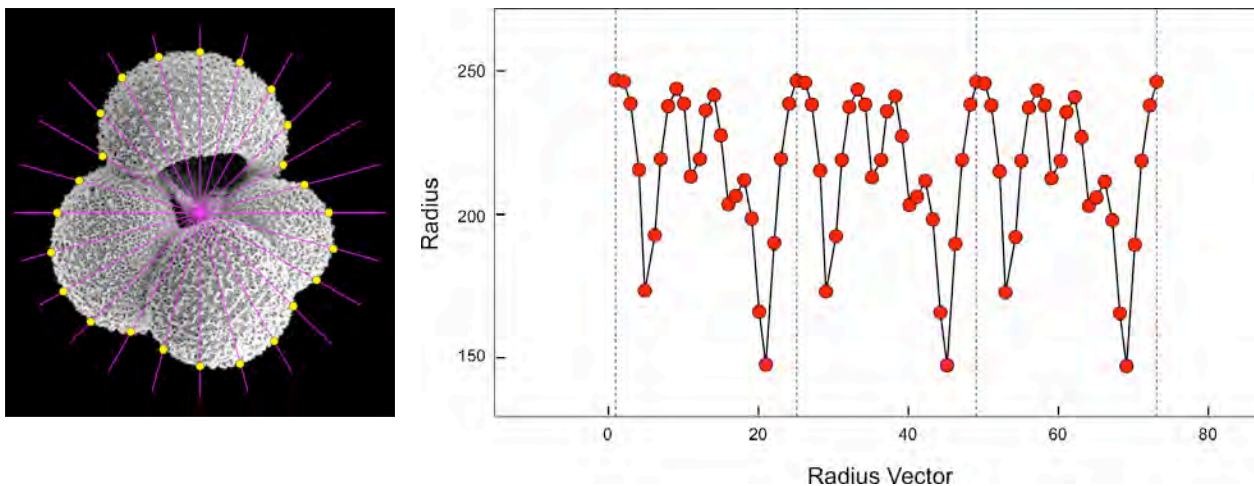


Figure 1. *Globigerina bulloides* image with 24 superimposed equiangular radius vectors associated boundary outline points (left). The periodic boundary outline function plotted over three cycles (right, with cycle boundaries marked by dashed lines). In order to apply a Fourier approach to the characterization of semilandmark sampled boundary outline the implied shape function must be periodic. This constraint applies equally to Z-R Fourier and EFA representations of specimen outlines though, in these cases the constraint of equal semilandmark spacing does not necessarily apply.

In addition, Fourier analysis simply redescribes the shape of a boundary outline curve by decomposing the positions of the set semilandmark points that represent the curve as an infinite series of harmonic amplitudes and phase angles (Fig. 2). This exercise achieves little in terms of shape analysis *per se*, but is typically used to prepare the semilandmark data for shape analysis. In this sense, transforming semilandmark data into a mathematical shape space via specification of a series of normalized radius vectors, Zahn and Roskies (Z-R) shape functions, or the separate *x* and *y* (and *z*) functions of EFA prior to Fourier decomposition represents the conceptual equivalent of *Procrustes* alignment, see MacLeod 2009a) while the redescription of such data following conversion to a shape function format is the equivalent of principal warps analysis (see Bookstein 1991; MacLeod 2010a,b) for landmark data. In standard geometric morphometrics this procedure was once thought useful in its own right, but has now been largely abandoned because it is recognized as being unnecessary for the computation of shape space ordinations via a principal components analysis (PCA) or singular value decomposition (SVP) of datasets consisting of *Procrustes* residuals (see MacLeod 2009a,b).

¹ In some cases Fourier analysis has been used to analyze open curves (*e.g.*, dental arcades, craniofacial profiles, see articles in Lestrel 1997), but in all cases the mathematics of Fourier analysis is applied to the data as if it constituted a periodic function. Also it is well known that the application of Fourier analysis to forms that do not represent periodic functions introduces inaccuracies that must be handled by various ad hoc strategies (*e.g.*, discrete Fourier transform, discrete-time Fourier transform, Hamming windowing, see Oppenheim *et al.* 1999; Jacobsen 2003). Interestingly, these discrete signal-correction strategies have rarely (if ever) been applied in morphometric analyses.

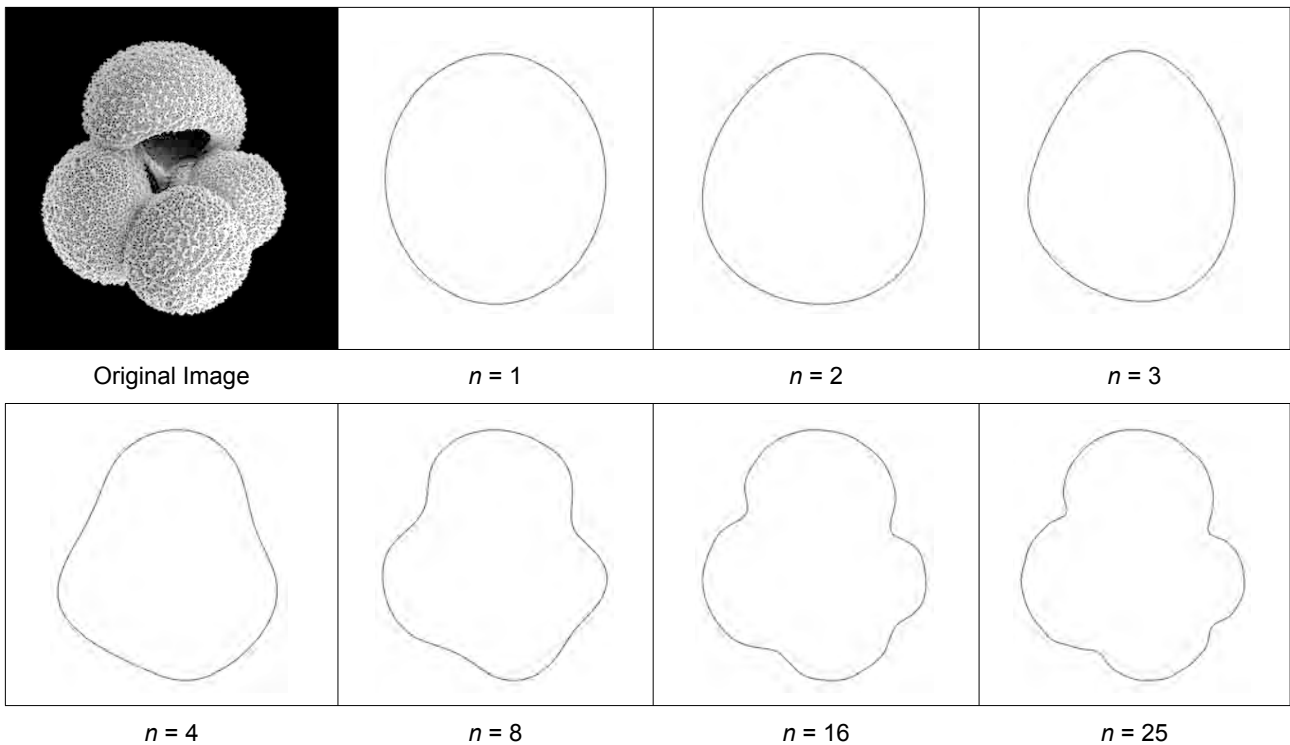


Figure 2. Reconstruction of the *G. bulloides* specimen outline using different numbers of elliptical Fourier harmonic amplitudes (n). These Fourier harmonics constitute terms or variables that describe features of the form ordered by steadily decreasing spatial detail.

Previously Rohlf (1986,1990) has argued that a Fourier analysis can be useful for smoothing boundary outline data prior to shape analysis. This smoothing is accomplished by using a subset of harmonic amplitudes and phase angles (e.g., the first few, first 10, first 20) to represent the curve (Fig. 2). But while signal analysts and electronic engineers often use Fourier calculations to construct electronic filters for precisely this purpose, the logic for undertaking this operation seems inconsistent with Rohlf's advice to always use all principal warps when the principal warps redescription of landmark data is used as the basis for shape analysis (see Rohlf 1993). Note here I'm not drawing attention to the infinite character of the Fourier series. The number of unique Fourier harmonics that can be used to is set by the number of semilandmark points the data analyst chooses. Consequently, it is possible to always use the maximum number of unique Fourier harmonics that can be calculated for any given dataset. Moreover, of smoothing is what you're after the same sort of outline smoothing that can be accomplished by specification of a Fourier-based filter is accomplished quickly, easily, and routinely at the point of boundary outline data collection when the total number of pixel coordinate points the represent an object's outline is interpolated down to a much lower number of (usually) equally spaced boundary outline semilandmark coordinate points (Fig. 3, see also Lohmann and Schweitzer 1990; MacLeod 1999).

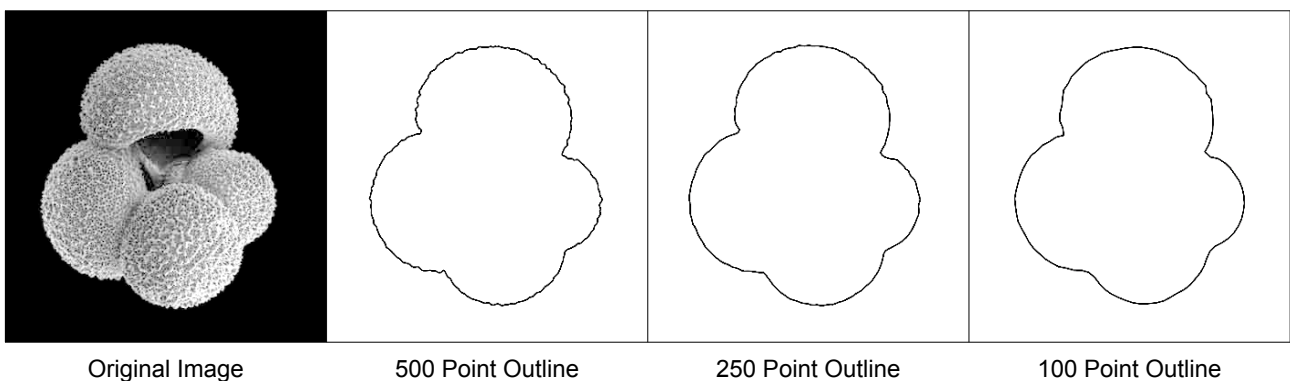


Figure 3. Example of outline smoothing achieved by interpolating a digitized representation of a specimen's boundary outline to a smaller number of equally spaced semilandmark points. The outline of the *G. bulloides* image (left) was originally digitized using 647 points.

Leaving these issues aside, there is also the fundamental objection to the application of boundary outline analysis strategies to biological morphometric problems first raised by Bookstein *et al.* (1982) in the context

of Fourier analysis, but later extended to all outline data sets (Bookstein 1990). This argument involves the general nature of achieving biologically meaningful comparisons between shapes and, in particular, the role the principle of biological homology plays in informing such comparisons (Fig. 4).

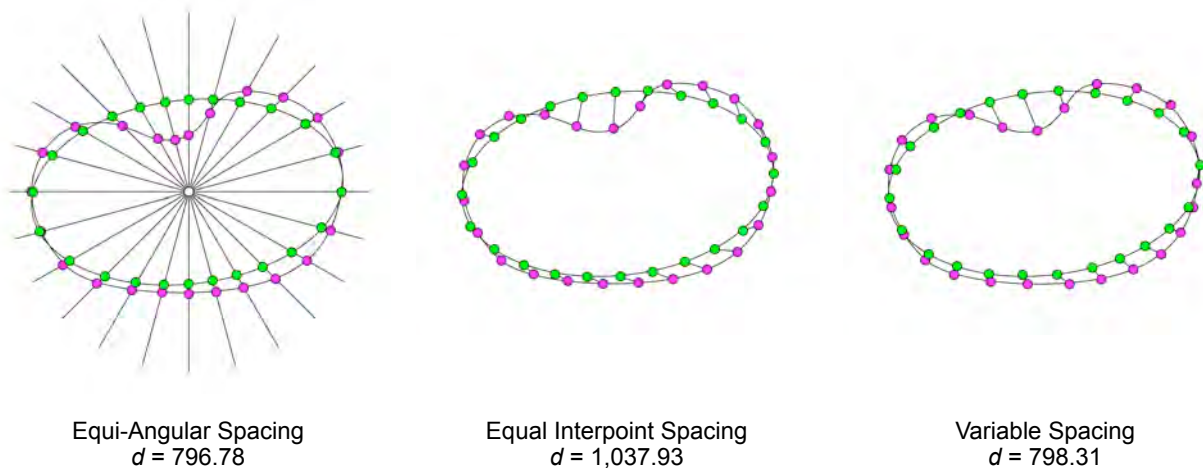


Figure 4. Variations in the shape 'distance' estimates for the same forms under different semilandmark sampling schemes.

Having worked in the area of mathematical outline analysis for most of my entire professional career I can say with a certain degree of authority that much confusion exists in the technical morphometric literature, and in the minds of many morphometricians, not to mention students and lay practitioners, regarding all these issues. In the final sequence of essays for this column I want to take this opportunity to offer a personal perspective on these matters by drawing together relevant arguments I have made in various technical articles over the years, but which are scattered across time and (literary) space. In the spirit of full disclosure I will unashamedly admit that my purpose in this series of essays will be to convince you that, if you have read anything about outline morphometrics before much of what you have read is incorrect and/or out of date, including a number of my own previous publications. But regardless of whether you have or haven't considered these arguments or even thought much about outline morphometrics before I hope you'll come away understanding more about the role the analysis of outlines — and their 3D extensions, surfaces — can play in contributing to the future of morphological data analysis in biological and palaeontological contexts.

To show how it is possible to undertake an outline analysis without going through an initial Fourier redescription, and the advantages inherent in doing so, let's go back to a consideration of the Z-R shape function. As you will recall this function was developed for use with Fourier analysis as a way of representing a closed form outline as a periodic function without having to specify a center from which a series of radius vectors emanate. Using the Z-R shape function a Fourier analysis can be used to decompose any boundary outline curve, no matter how complex (Fig. 5), into a series of harmonic amplitudes and phase angles; even multi-valued curves that cross themselves. Interestingly, the shape functions used as the basis for EFA have this same property (see MacLeod 2012) as does Bookstein's (1978) tangent angle approach to outline characterization. For now, however, let's use the Z-R shape function as a place to begin developing an alternative to Fourier analysis for the study of boundary outlines that's more in keeping with the spirit, and the mathematical letter, of geometric morphometrics.

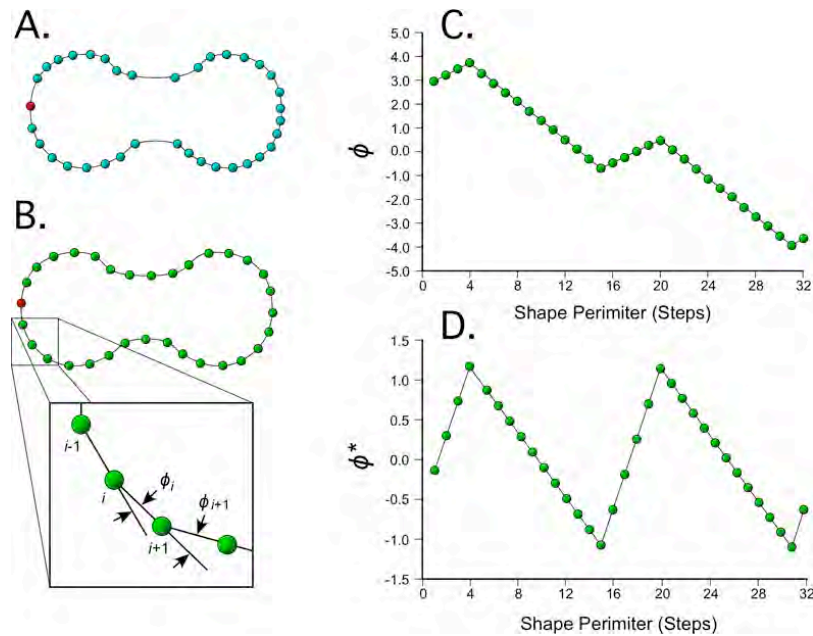


Figure 5. Steps in calculating the Zahn and Roskies (Z-R) shape function. A. original set of semilandmark data points placed on the periphery of a hypothetical shape. The red landmark represents the starting point for digitization. Ideally this point should be placed on a topologically homologous landmark. Note the uneven interlandmark spacing. B. Adjustment of original data (via interpolation) to a set of equally spaced semilandmark points. Again, the red landmark represents the starting point for digitization. The inset illustrates the expression of the shape of the outline as a series of net angular deviations (see text for discussion). C. the ϕ form of the Z-R shape function with a typical ramp that denotes a closed curve. D. the ϕ^* form of the Z-R shape function with represents the shape residual after removal of the ramp of circularity.

The Zahn and Roskies procedure (usually) begins with the collection of a set of equally spaced x,y coordinates (or x,y,z coordinates if a three-dimensional analysis is required, see MacLeod 1999) along an outline or curve of interest (Fig. 5A). If the curve has a closed form it can be regarded as being an n -sided polygon where n is the number of semilandmark points used to represent the curve's geometry. Since the distance between each point is the same we need only remember one distance value for the entire outline. This is termed the 'steplength'. For curves that have been sampled to the same number of semilandmark points the steplength will be proportional to the length of the outline, which is to say its size. Size may be removed or retained in an analysis by eliminating or including the steplength for each boundary outline curve in the sample data matrix.

Once control over size has been gained in this manner, the shape of the outline can be represented in a 'street direction' manner of the Z-R shape function: as a series of angular turns that need to be executed in order to travel around the outline in steps of equal length and (if the curve is closed) arrive back at the starting point. Mathematically it is convenient to express these angles as a series of net angular deviations from the direction taken in the previous step, and to express them in radians rather than in degrees.² This operation effectively removes differences in the rotational orientation between the specimens. Since we're expressing the shape of the curve as a set of angles, differences in the position of specimens within the system of semilandmark coordinate values are automatically rendered irrelevant. Accordingly, calculation of the Z-R shape function of the original semilandmark data, in addition to redescrbing the form of the outline exactly, also accomplishes the three tasks of a *Procrustes* alignment: removal of positional, rotational, and scaling differences between specimens. To be sure, the Z-R shape transformation does not accomplish this task using the same mathematics as *Procrustes* alignment. But the result is largely the same irrespective of the calculations employed (Fig. 6).

² A radian is the ratio of an angle's arc to its radius. It's used to express the value of an angle as a dimensionless distance rather than as a number of degrees of a circle.

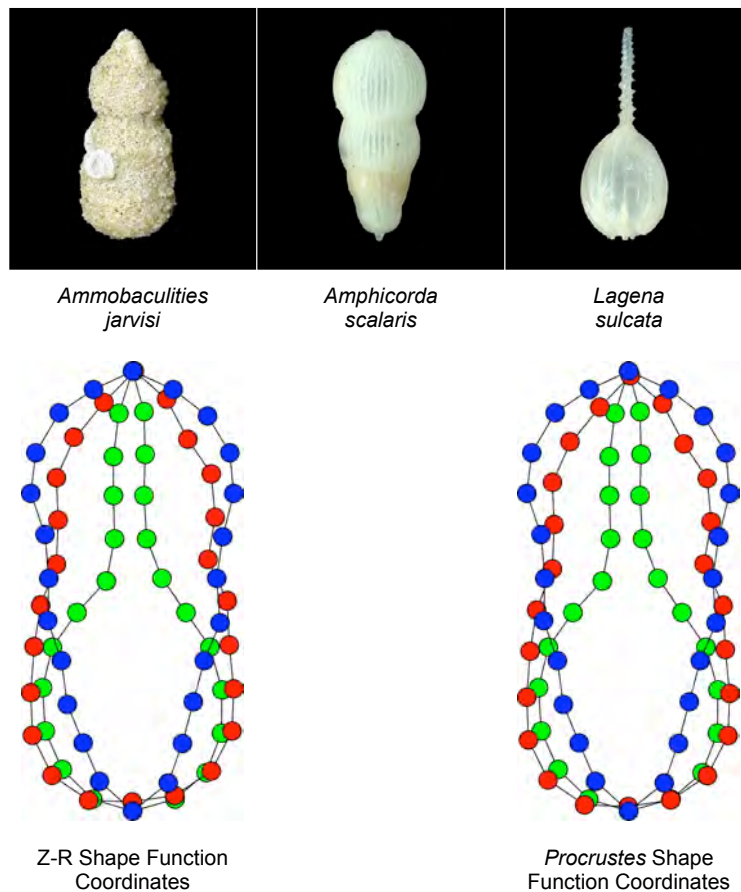


Figure 6. Comparison of shape coordinates calculated on the basis of the Z-R and *Procrustes* procedures for a set of 24 equally spaced semilandmark points around the peripheries of three benthic foraminifer species.

In the early 1980s George (Pat) Lohmann, who was (and still is) a Woods Hole Oceanographic Institution micropalaeontologist stumbled onto the Z-R shape function while looking to develop a method to organize the outlines of microfossils quickly, easily, accurately, and as simply as possible. The Z-R shape function is well suited to the job Pat had in mind for not all microfossil outlines are single valued and often the mathematical center of a microfossil's outline does not correspond closely to its anatomical center. But unlike Zahn and Roskies (1972), Pat didn't see any need to redescribe the redescription of these outline shapes using Fourier harmonics and then analyse sets of harmonic amplitude values using a multivariate ordination technique such as principal components analysis (PCA) or singular value decomposition (SVD). Instead, he felt it would be more efficient to regard the values of the n angular terms of the Z-R shape function as a set of valid shape variables in their own right.

Lohmann dubbed his direct approach to the analysis of specimen outlines by means of the Z-R shape function 'eigenshape' analysis (Lohmann 1983). This name signifies the two critical aspects of his procedure (1) complete representation of the set of outlines as sets of geometrically equivalent shape functions and (2) assessment of the major directions of observed and measured shape variation in a dataset by means of eigenanalysis. However, in addition to these procedures Pat adopted several conventions early in the development of eigenshape analysis that, with the benefit of hindsight, I feel have tended to limit the scope of its application and obscure links between his eigenshape procedure and what came later to be known as geometric morphometrics.

In particular, Pat followed Zahn and Roskies' (1972) recommendation to use a 'normalized' version of the raw Z-R function as his preferred form of the shape function. The factor Zahn and Roskies recommended be removed from shape data was the form of a circle which they described as 'the most shapeless closed form' (Zahn and Roskies 1972, p. 270). Mathematically, this operation means that normalized Z-R shape functions express patterns of deviation from circularity.

It should be appreciated that this suggestion is entirely in keeping with the Fourier-based aesthetic of Zahn and Roskies' original work. After all, the 0th harmonic of a radial Fourier series is a circle and all subsequent harmonics in the series express patterns of deviations from this circular ideal. Also, removal of the ramp that denotes constant angular deviation in the raw Z-R shape function (see figs 5C and 5D) makes the function

appear to fit the ideal of a periodic function to a greater extent than the typical form of the raw shape function (compare Fig. 5D with Fig. 1) — another nod to the exigencies of applying a Fourier decomposition to such representations of shape.

Strictly speaking, however, use of this normalization procedure is, at best unnecessary and at worst detrimental, from the standpoint of shape analysis. The raw Z-R shape function data is an exact description of the outline's geometry all by itself. Indeed, the raw Z-R shape function is a more complete representation of the boundary outline's geometry than the normalized version because it contains all the information necessary to reconstruct the measured shape. By removing the factor of circularity from the raw function the normalized form, in a sense, 'hides' the circular nature of the curves' geometry from view (and from subsequent analysis). But most importantly from the standpoint of shape theory, arbitrary selection of a circle as the reference shape means that the linear plane(s) tangent to the Kendall shape space onto which the outline data will be projected by the PCA and/or SVD procedures in order to represent patterns of similarity and difference within a sample of outline shapes will always be located in a suboptimal orientation relative the data of any given sample (see Kendall 1984; Bookstein 1991; MacLeod, 2009c). This, in turn, means that the resulting ordinations in PCA/SVD-determined geometric subspaces will contain a systematic bias in the placement of shapes the severity of which will be proportional to the difference between the samples' true mean shape and that of a circle. To be fair, the problems inherent in arbitrarily selecting a shape to use for shape normalization were not known in the early 1980s, much less the early 1970s. In this regard Zahn and Roskies' and Pat's failure to appreciate the effect this type of normalization would have on subsequent shape analyses is perfectly understandable. But these issues are well understood now and need to be kept in mind when evaluating classical eigenshape analysis as well as subsequent developments in the formulation, as well as options for application, of the eigenshape procedure.

The other aspect of the original eigenshape procedure that can be questioned legitimately is the manner in which biologically common features are matched across a sample of outlines by eigenshape analysis. In standard radial, Z-R, and elliptical Fourier analysis the issue of feature mapping does not arise as the coefficients of the Fourier amplitudes are insensitive to the starting point for outline digitization. Indeed, it is for this very reason that most Fourier representations of outline shape employ only the amplitude terms as shape descriptors. This is fine for a wide variety of physical shapes (e.g., sand grains). But the outlines of biological specimens differ from the outlines of most natural physical objects. Most biological outlines include combinations of discrete anatomical regions (e.g., head, trunk, appendages), structures (e.g., eye, nose, mouth) and substructural characters that exhibit a polarities of various sorts (e.g., proximal, distal). Ideally, discrete subsets of semilandmark points in the outline sequence should fall on biologically comparable parts of the form across all specimens in the sample. Fourier analysis finesses this critical issue because such distinctions don't exist in terms of harmonic amplitude-based representations of outline shape. But exist they do in the real worlds of biology and palaeontology. Morphometricians who decide to throw this information away do so at their peril for any single set of Fourier harmonic amplitudes, when taken in isolation from their associated phase angles, is non-unique. Such data actually describe an infinity of shapes.

Lohmann (1983) approached this issue in the context of eigenshape analysis in two ways. First, if a landmark could be identified on the outline that was common to all specimens in the sample it was recommended this be used as a common starting point for outline digitization. By using a common point of reference for sampling the outline, and by sampling the outline using a constant number of equally spaced semilandmark points, the outline is 'homologized' in a topological sense irrespective of which biological structures individual semilandmark points fell on across the sample. In this way outlines on which truly comparable point locations are few could be matched in terms of their computed geometries. In cases where the specimen outlines included no landmark that could be used as a starting point for outline digitization, Pat recommended that a reference specimen be selected and the Z-R shape functions be rotated to positions of maximum correspondence with this reference. Again, the homology is topological and is computed rather than interpreted, but only because the biological information necessary match outlines using other criteria is lacking.

In no case was any pretence made that this method of computing topological homology maps between specimens was preferable to the location of genuine landmarks provided these were available. Rather, the eigenshape strategy was justified as simply being preferable to pretending that landmark point locations existed on a structure when they clearly did not or were subject to a great deal of uncertainty with regard to their exact positions. Eigenshape approaches the the outline analysis problem are regarded by their practitioners as an efficient and pragmatic solution that while far from perfect, is undeniably preferable to giving up and foregoing the quantitative, geometric analysis of a large number important biological structures that taxonomists, palaeoecologists, palaeogeographers, biostratigraphers, *etc.* have been comparing qualitatively for (literally) centuries. Indeed, those with direct experience of how taxonomists actually make qualitative comparisons between differing sets of morphologies in the absence of the biological signposts provided by valid landmarks know that most use an approach essentially identical to the computation of topological homology maps.

Once the outlines for a set of specimens had been quantified via specification of equal series of semilandmarks, redescribed using the Z-R shape function, and assembled into an $n \times m$ data matrix (where n = the number of specimens in the sample and m = the number of semilandmarks collected from each outline, Lohmann (1983) advocated description of the structure of relations among the semilandmarks by calculating an $m \times m$ pairwise correlation matrix. Selection of the correlation matrix as the basis for structural comparison seems an odd choice as all the values in the data matrix cells are angles (expressed as radians) and so represent the same types of both quantities and magnitudes. In most instances the covariance matrix would be chosen to represent data of this type. However, use of the covariance matrix would mean that some parts of the outline — specifically the parts characterized by more angular bends — would have a differential influence in determining the orientation of the eigenvectors that are used to assess patterns of shape variation. Pat made the decision that he did not want certain regions of the outline to ‘pull’ the eigenvectors toward themselves in an orientational sense, and so opted to represent structural relations in a manner that ensured all regions of the outline would count equally in determining the final result. This decision is contrary to what has become standard practice in geometric morphometrics of employing the covariance matrix to represent structural relations among landmark variables and simply accepting that, within such a system, landmarks whose relative positions are more variable across the sample will be more highly weighted in the result than more conservative landmarks.

After calculation of the covariance matrix Lohmann recommended using SVD to decompose the correlation matrix. If X is the $n \times m$ data matrix of n specimens and m shape function values the basis matrix of structural relations among variables can be provided by either of two matrices.

$$Z_r = XX' \quad (26.1)$$

$$Z_Q = X'X \quad (26.2)$$

Where: X' = transpose of X

Z_r = matrix of covariances/correlations
between semilandmarks

Z_Q = matrix of distances/correlations between
specimens (shapes)

If each shape function is normalized to have a zero mean and unit variance (= row normalization) Z_Q will contain the pairwise correlations between specimens, otherwise these values will be distances. Similarly, each term of the shape function is normalized to have zero mean and unit variance (column normalization) Z_r will contain the pairwise correlations between shape variables; otherwise these values will be covariances.

The Eckart-Young Theorem tells us that any matrix can be expressed as the product of three matrices.

$$X = VWU' \quad (26.3)$$

Where: V = eigenvectors of Z_r

W = diagonal matrix of singular values (= square
roots of the eigenvalues of V and of W)

U' = transpose of eigenvectors of Z_Q

If Z is a symmetrical, square matrix the sets of eigenvectors contained in V and U will be identical. These m eigenvectors will coincide with the major directions of variable-normalized shape variation present in the data subject to the constraint that all eigenvectors be oriented at right angles to one another (= orthogonality). The eigenvalues represent the lengths of these eigenvectors which, when added together will be equal to the sum of the variances of each of the original (shape) variables. Because the eigenvectors are aligned with the maximal directions of variation in the set of variables as a whole — taking account of inter-shape variable covariances/correlations — the first few eigenvectors will represent a greater proportion of the observed shape variation than any single shape variable can represent; often a dramatically greater proportion. Geometrically the m eigenvectors contain m values each of which is a covariance or correlation between the eigenvector and each of the m original variables, so long as $n \geq m$. If $n < m$ (which is often the case in an eigenshape analysis) only n eigenvectors and with n positive eigenvalues will be extracted.

In standard eigenshape terminology these eigenvectors are termed ‘eigenshapes’ though this is somewhat confusing insofar as the eigenvectors do not represent singular shapes. Rather, these coefficients (or loadings or weights) represent patterns of association between the orientation of the eigenvector and the positions of the original variables in the space defined by between variable covariances/correlations. In effect, each eigenvector represents a hypothetical trend or pattern of outline shape deformation with some

regions of the outline being more directly aligned with a particular eigenshape axis than other areas. The geometric signature of this alignment takes the form of the positively and negatively aligned regions becoming more differentiated from one another at the positive and negative extremes of shape variation seen in the sample and less differentiated from one another near the center of the observed shape distribution (see below for a graphical example).

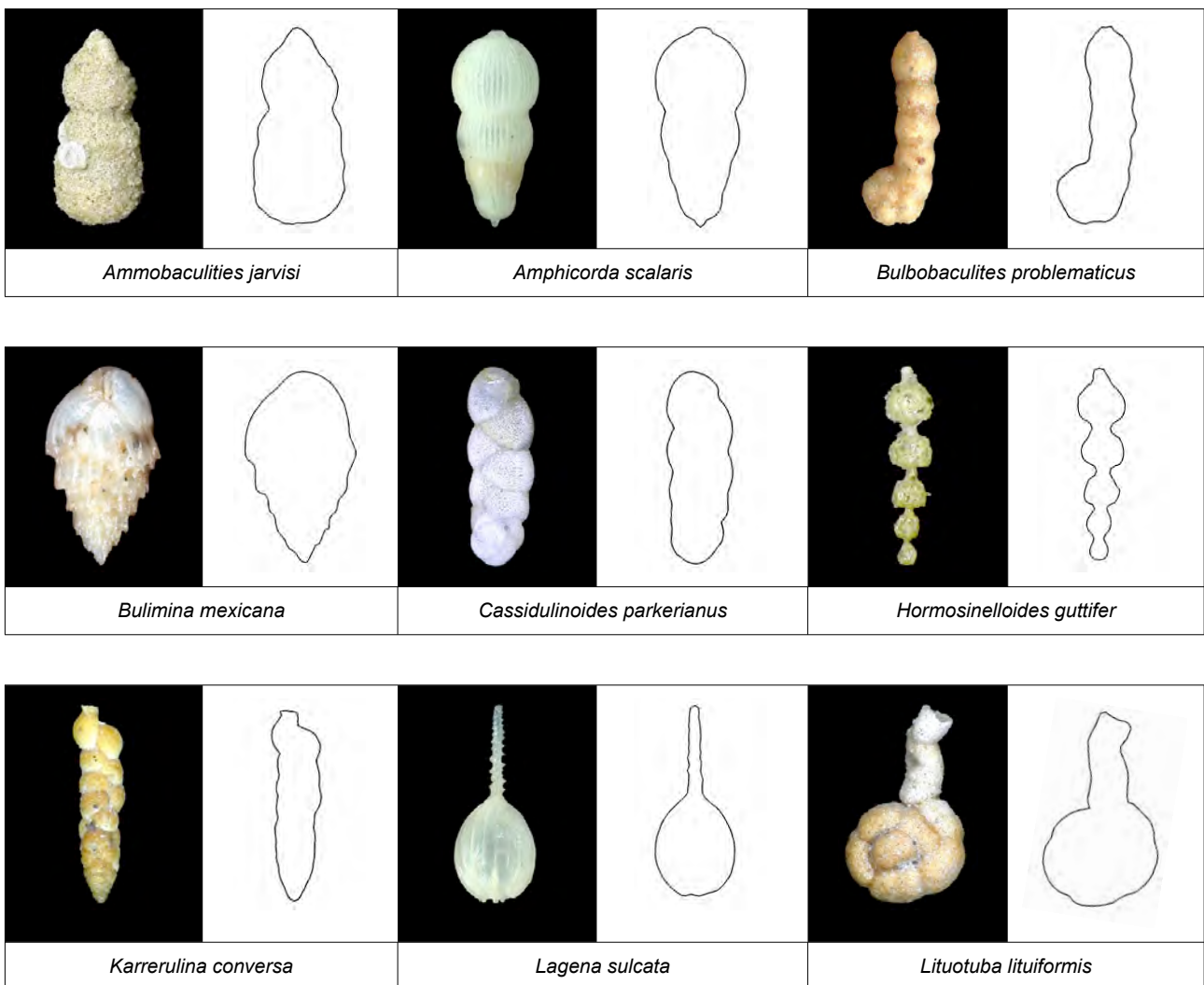
As with principal components analysis and/or principal coordinates analysis individual outlines the covariance or correlation of Z-R shape functions of equivalent dimensionality with each of the m eigenshapes (= eigenvectors) can be determined either by using the standard covariance/correlation equations or by using their matrix algebraic equivalents, either:

$$\text{scores} = XV \quad (26.4)$$

or

$$\text{scores} = UX' \quad (26.5)$$

Now that we have the basics of a standard eigenshape analysis down let's take a look at the results of a typical analysis by applying the Lohmann (1983) procedure to our sample of foraminifer outlines (Fig. 7).



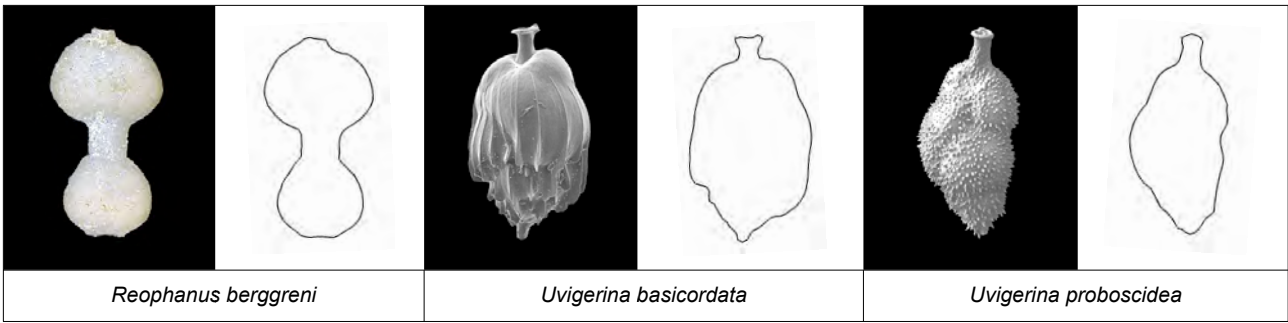


Figure 7. Zahn & Roskies shape function representations of the the outline shapes of 12 benthic foraminifer species. The outline of each specimen was interpolated to 100 equally spaced semilandmark points with outline digitization beginning at the center of the aperture in each case. Note highly diagnostic character of the outline shapes along with the lack of consistently identifiable landmark points (other than the aperture) on the peripheries of these specimens.

As has been the case typically with classic Lohmann-style eigenshape analysis the resolution of the boundary coordinate outlines for this dataset was set arbitrarily to a value of 100 semilandmark points (see Lohmann 1983). This figure is based on experience with eigenshape analyses and seems to result (in most cases) in representation of an outline’s geometry to a level of accuracy such that the form of most taxonomically important morphological substructures are recognizable while, at the same time, suppressing the incidental variation associated with surface texture, minor imperfections in structure, adhering sediment particles and/or dust, etc. These x,y coordinate points were transformed into their equivalent normalized Z-R shape functions (ϕ^*) and the values of those functions used to construct a 12 x 100 data matrix of outlines and shape function coefficients.

Eigenanalysis decomposition of the pairwise correlation matrix resulted in the extraction of 12 eigenshapes (= eigenvectors) of which the first nine represented > 95 percent of the observed shape variation (Table 1). By way of comparison, an eigenanalysis of a matrix of 50 Z-R Fourier harmonic amplitudes and phase angles also resulted in the extraction of 12 eigenvectors of which the first ten represented > 95 percent of the observed shape variation. While the saving of a single eigenvector may not sound terribly significant, remember this is a very small example dataset. When larger datasets are considered the dimensionality reduction that can be achieved by eigenshape is often more impressive, Still, even with these data it is clear that Lohmann’s (1983) eigenshape approach results in a more efficient analysis than the equivalent Fourier procedure; more information relevant to the characterization of shape variation in the sample is loaded onto the first few eigenvectors which, in terms of the qualitative interpretation of major shape trends, are typically the only shape variables that are inspected in any detail.

Table 1. Comparison between eigenvalues extracted from the eigenshape and ZR-Fourier analysis of the benthic foraminifer data.

Eigenvectors	Eigenvalues	
	Eigenshape	Z-R Fourier
1	24.7607	24.8702
2	43.4817	39.6529
3	61.8433	52.9220
4	71.1037	62.9865
5	77.9684	71.9990
6	84.1625	78.8567
7	88.4970	84.8566
8	92.5691	89.7922
9	95.7202	93.9723
10	97.9810	97.8552
11	100.0000	100.0000
12	100.0000	100.0000

This diagram included in Table 1 represents a tabular and graphic demonstration of one of Lohmann’s main practical arguments for the eigenshape approach over that of Fourier analysis. Greater efficiency is gained by allowing the data to specify the modes of shape variation most suitable for its own characterization than

by forcing this characterization to be routed through an arbitrary — though highly elegant — set of idealized shape descriptors: the Fourier series. In addition, if the pairwise covariance rather than the correlation matrix is used to represent geometric relations within the data, eigenshape analysis has the advantage of allowing regions of pronounced shape variation within the boundary outline to attract the eigenvector axes to themselves in an orientation sense.³ This improves the interpretability of the eigenshapes in contexts that are useful for testing biological hypotheses. The distribution of foraminifera outline shapes within the space of the first three eigenshapes is shown in Figure 8.

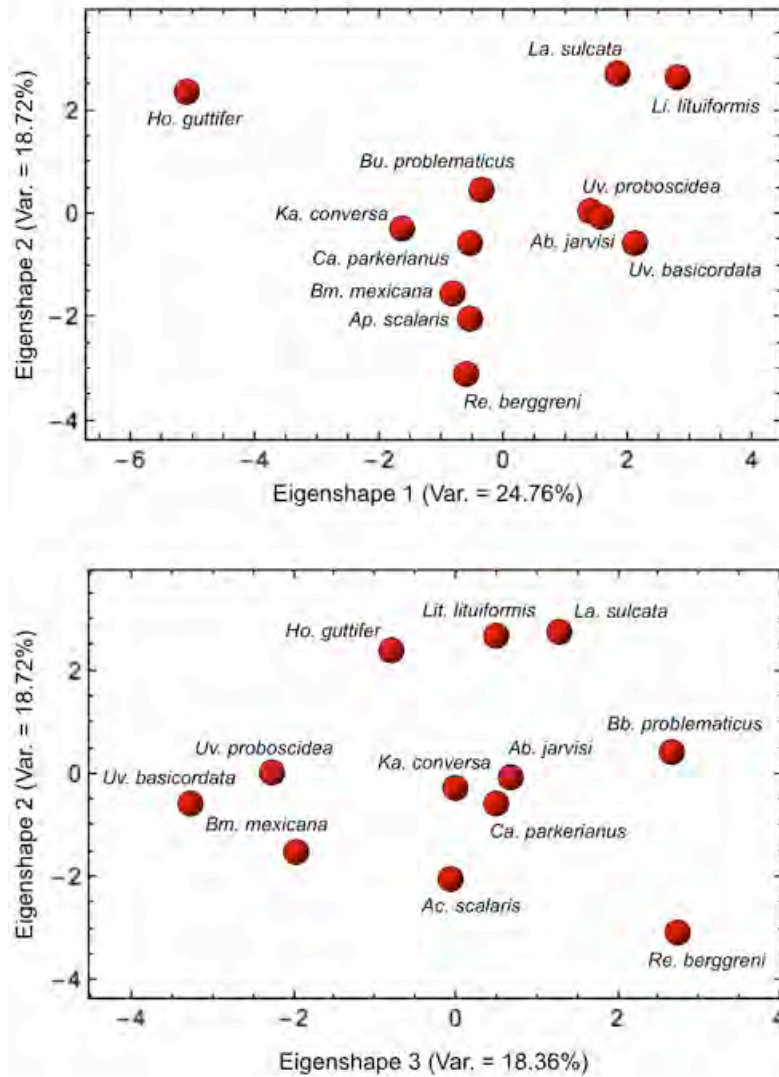


Figure 8. Distribution of benthic foraminifer shapes in the subspace formed by the first three eigenshape axes. See text for discussion.

While this plot may seem superficially similar to those we have seen for this dataset before, the outline shape grouping we see recorded there in are actually rather remarkable and certainly quite a bit different picture of patterns of shape similarity and difference for this sample from that offered by elliptical Fourier analysis (EFA, compare with Fig. 6 in the previous *Palaeo-Math 101* column, MacLeod 2012). This plot also shows nicely why you need to develop skill in visualizing point distributions in (at least) three dimensions in order to interpret these data correctly.

There are three obvious groups of outline shapes along the first eigenshape axis (ES-1). *Hormosinelloides guttifer* projects to the lower end of ES-1 which seems appropriate as it is the only species exhibiting inflated, spherical, uniserially arranged chambers. At the other extreme of this axis *La. sulcata*, *Li. lituiformis* the two *Uvigerina* species, and *Ab. jarvisi* form a heterogeneous group whose unifying characteristics appear to be

³ This feature was not taken advantage of in the example analysis included here because I want to begin the discussion of eigenshape with an example presentation of its original form. An equivalent covariance-based analysis for this dataset results in additional efficiencies in eigenanalysis over the results presented above.

common possession of a pronounced apertural neck or, in the case of the latter species, pointed apertural constriction. This group is further subdivided along the second eigenshape axis by the relative length of the neck/constriction with relatively short features plotting low along ES-3 and relatively long features plotting high. In the middle of ES-1 a heterogeneous grouping of species if gathered together that possess neither of these (for this sample) extreme morphologies.






Interestingly while accounting for a smaller proportion of the observed shape variation the ordination of shapes along the third eigenshape axis (ES-3) is as informative if not more so. Here shapes whose outlines are pinched at either end and inflated in the middle (the two *Uvigerina* species and the bulimulid) are contrasted with shapes that are narrow along their long axis, but inflated at either one (*Bu. problematicus*) or both (*Re. berggreni*) ends. Again, this seems quite a natural distinction given the set of shapes present in the dataset, but one that is far from obvious as the third most important shape trend in these data from a simple visual inspection of Figure 7. Also, far from obvious in Figure 7 is the fact that these major shape groupings are quite well structured within this dataset. The uvigerinid and bulimulid species form a distinct subgroup within this subspace that does indeed reflect their distinctive shapes, as do the ‘long-necked’ species *La. sulcata* and *Li. lituiformis*. There are no intermediates occupying the theoretical shape space between these well-defined regions. Uniquely shaped species such as *Ho. guttifer*, *Re. berggreni*, and *Bu. problematicus* are also identified as such in this subspace, along with unanticipated — and rather charming — underlying organizational similarities (e.g., the geometric link between *Bu. problematicus* and *Re. berggreni* in the context of this small sample of shapes).

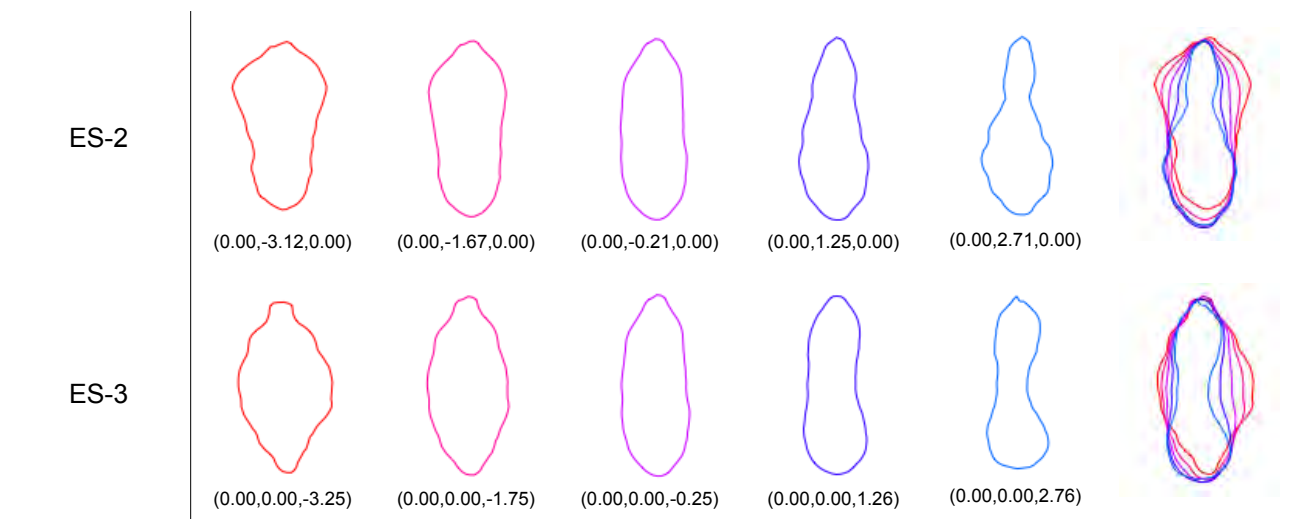
Some, but by no means all, of the structure we see in the eigenshape results as present in the ordination spaces created as a result of the PCA analysis of EFA amplitude coefficients extracted from of the same empirical data (compare Fig. 8 with Fig. 6 of the previous *Palaeo-Math 101* column, MacLeod 2012). But with the exception of a few of the extreme shapes (e.g., *La. sulcata*, *Li. lituiformis*, *Re. berggreni*) the same level of clarity in the recognition of outline shape based subgroupings achieved by eigenshape analysis is simply not present in the EFA-based shape space ordinations. Presumably this is because of the intermediate step taken by EFA of decomposing and redescribing outline shape variation as a series of Fourier harmonic amplitudes.

It’s also worth noting here that, while the EFA analysis was conducted using 97 variables (and so was comparable to the eigenshape analysis in terms of overall dimensionality) only 25 EFA harmonics were used to characterize each shape. It could be the case that these 25 harmonically-defined shapes were insufficient to capture all of the salient morphological features present in the outlines of these sample shapes. If so, this a deficiency that could be addressed by simply increasing the harmonic resolution of the EFA analysis. However, this would increase the dimensionality of the data analysis and, as we have already seen (e.g., Bellman 1957; MacLeod 2007) increasing the dimensionality of a dataset often has unanticipated consequences for a data analysis and usually requires dramatic increases in the sample size in order to be confident in the results. But even if we accept this as a potential strategy for EFA analysis, it still does not change the fact that eigenshape analysis was able to sense and represent accurately the structure of shape relations in this small dataset in the context of a dimensionality that was comparable to that of an EFA of the same empirical data to an extent that the latter procedure was not. Neither analysis is wrong. But the result produced by eigenshape analysis is the more biologically informative.

It probably should go without saying at this point, but all the shape modelling tools I have introduced you to and illustrated the utility of in previous columns are also available for eigenshape analysis. Their use greatly improves the interpretability of the ordination spaces in which eigenshape data are often portrayed (e.g., Fig. 8). Along-axis shape models for the first three eigenshape axes of the benthic foraminifer outline dataset, along with accompanying model overlay or ‘strobe’ plots, are shown in Table 2.

Table 2. Along axis models existing at coordinate locations along the first three eigenshape axes of the benthic foraminifer dataset. The specific coordinate position at which each model was calculated is shown below each model (in parentheses).

Eigenshape Axis	Model 1	Model 2	Model 3	Model 4	Model 5	Overlay
ES-1	 (-5.05,0.00,0.00)	 (-3.09,0.00,0.00)	 (-1.13,0.00,0.00)	 (0.84,0.00,0.00)	 (2.80,0.00,0.00)	



Comparing these models with the equivalent EFA shape space models (see Table 3 of the previous newsletter's *Palaeo-Math 101* column, MacLeod 2012) is also instructive. The eigenshape models look decidedly rougher, more asymmetric; on occasion virtually pathologic (e.g., ES-1, Model 1). This rough look may strike many as disquieting compared to the overt symmetries that Fourier shape models usually display. But this rough look underscores the fundamental strength of eigenshape analysis and the reason it delivers better results in the vast majority of instances than radial Fourier, Z-R Fourier, or elliptical Fourier analyses. The outline shapes present in the dataset are also rough, asymmetric and full of relatively small irregularities. In some cases these are nothing more than idiosyncrasies of the specimen chosen for analysis; part of the noise that is present in any shape analysis. But in others these roughnesses, asymmetries and irregularities are part of the fundamental geometry, not only of the specimen, but part of the group the specimen represents; part of the signal the data analyse is seeking. Fourier analysis passes the representation of these geometrically 'difficult' features through the filter of a set of highly structured, smooth, symmetrical shape variables. Accordingly, it often takes quite a number of Fourier harmonics to represent these aspects of organismal outlines accurately.

Eigenshape analysis, on the other hand, is not troubled in the least by roughness, asymmetry or irregularity. All eigenshape responds to is the collection of shapes at whatever level of spatial resolution the data analyst as chosen to represent them at. All it does is deliver an efficient representation this observed shape variance. Eigenshape analysis zeros in on precisely those features of the outline shapes that are responsible for shape variation in the sample and not concern itself with the elegance of the shape variables it uses for this purpose. Rohlf (1986) assumed these rough sorts of features are more likely to be part of the shape noise than part of the shape signal and so would lead to the production of spurious and difficult-to-interpret results in an eigenshape analysis. I must say that after almost 30 years of personal involvement performing eigenshape analyses in a wide variety of contexts, just the opposite has been my experience. In the vast majority of cases eigenshape analysis does a better job recognizing the geometric structure of the distribution of shapes present in a sample than Fourier (and other forms of) outline analysis because real specimens exhibit a variety of shape-based similarity and difference patterns at a variety of scales and because these highly complex, geometrically 'difficult' patterns. These are the very stuff of biological shape variation; the aspects of that variation biologists are interested in, the aspects that comprise the subjects morphological taxonomy, morphological ecology, morphological biogeography, morphological function, etc. Best of all, the eigenshape approach to outline analysis I've described and demonstrated here is just the starting point for a set of variations on the eigenshape theme that — as we'll see in the next column — can (i) expand the utility of eigenshape analysis beyond the assessment of closed curves, (ii) improve the link between topological and biological homology in the representation of boundary curves, (iii) combine the analysis of landmarks with the analysis of outlines, and (iv) align this technique with the basis of geometric morphometrics in a formally mathematical (rather than simply a conceptual) sense.

As for software, since classical eigenshape analysis amounts to little more than a PCA of Z-R shape function data, and since the Z-R shape function is quite easy to calculate from normal x,y coordinate point data (see the section in the *Palaeo-Math-2 Spreadsheet* for this column and for MacLeod 2011), with a little ingenuity this methods can be implemented by anyone using resources available to them in the public domain. I have made available my personal eigenshape routines for eigenshape analysis as compiled applications for both Mac and PC operating systems. Øyvind Hammer's *Past* (<http://folk.uio.no/ohammer/past/>) programme package implements a form of eigenshape analysis. Both standard and extended version of eigenshape analysis based on my algorithms are also available for use as web-based applications from the Morpho-Tools web side (<http://www.morpho-tools.net/>). Finally, the *Mathematica*[™] routines I have developed for the

implementation of eigenshape analysis, and that I used to perform the analyses I reported here, are available for users of that software computing system. I also am aware that R-based eigenshape routines are included in Claude (2008).

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

REFERENCES

- BELLMAN, R. E. 1957. *Dynamic programming*. Princeton University Press, Princeton 340 pp.
- 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., CHERNOFF, B., ELDER, R., HUMPHRIES, J., SMITH, G. and STRAUSS, R. 1985. *Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes*. Academy of Natural Sciences of Philadelphia, Philadelphia 277 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*, **31**, 85–92.
- FERSON, S., ROHLF, F. J. and KOEHN, R. K. 1985. Measuring shape variation of two-dimensional outlines. *Systematic Zoology*, **34**, 59–68.
- JACOBSEN, E. and LYONS, R. 2003. The sliding DFT. *Signal Processing Magazine*, **20**, 74–80.
- KENDALL, D. G. 1984. Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, **16**, 81–121.
- LESTREL, P. E. 1997. *Fourier descriptors and their applications in biology*. Cambridge University Press, Cambridge 466 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. 2007. Groups II. *Palaeontological Association Newsletter*, **65**, 36–49.
- MacLEOD, N. 2009a. Who is Procrustes and what has he done with my data? *Palaeontological Association Newsletter*, **70**, 21–36.
- MacLEOD, N. 2009b. Shape theory. *Palaeontological Association Newsletter*, **71**, 34–47.
- MacLEOD, N. 2009c. Form & shape models. *Palaeontological Association Newsletter*, **72**, 14–27.
- MacLEOD, N. 2010a. Principal & partial warps. *Palaeontological Association Newsletter*, **74**, 35–45.
- MacLEOD, N. 2010b. Principal warps, relative warps and Procrustes PCA. *Palaeontological Association Newsletter*, **75**, 22–33.
- MacLEOD, N. 2011. The cannot hold I: Z-R Fourier analysis. *Palaeontological Association Newsletter*, **78**, 35–45.

MacLEOD, N. 2012. The center cannot hold II: elliptic Fourier analysis. *Palaeontological Association Newsletter*, **79**, 29–42.

OPPENHEIM, A., V., SCHAFER, R. W. and BUCK, J. R. 1999. *Discrete-time signal processing*. Prentice Hall, Upper Saddle River, N.J. 1120 pp.

ROHLF, F. J. 1986. Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. *Mathematical Geology*, **18**, 845–854.

ROHLF, F. J. 1993. Relative warp analysis and an example of its application to mosquito wings. In L. F. Marcus, et al. (eds). *Contributions to Morphometrics*. Museo Nacional de Ciencias Naturales **8**, Madrid, 131–160 pp.

ROHLF, F. J. 1996. Introduction to outlines. In L. F. Marcus, et al. (eds). *Advances in Morphometrics*. Plenum Press, New York, 209–210 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