

## Palaeo-Math 101

### The Centre Cannot Hold I: Z-R Fourier Analysis

Likely you were a bit dissatisfied with the *Sassia* example I closed the last column with. I know I was. At best editing an image in order to prevent it from having double, triple, or multiple valued regions seems a last-ditch compromise born of necessity. At worst it's a kluge, fit only for those who haven't thought very deeply about the geometric problem and/or unaware of the other mathematically reasonable options. In either case this technique — which you still see covered in some textbooks (e.g., Davis 2002), is far from the elegant solution mathematicians constantly strive to attain. Worse still, it only works with a reasonably conservative subset of possible shapes. Consider the positively pathological shape characteristic of the calcareous benthic foraminifera *Ramulina globulifera* (Fig. 1). Obviously, for this shape there are no simple cut planes available for us to use to subdivide the image and then reconstruct it in a manner that would prevent the occurrence of multi-valued segments.



Figure 1. Shape of the spinose benthic foraminifer species *Ramulina globulifera*.

The fundamental problem in both the *Sassia* and *Ramulina* outlines is our reliance on some landmark at or near the outline's centroid to supply a fixed point from which to calculate its form or shape function. Actually, when you think about it, this whole idea of using a central point as the basis for form/shape function calculation has caused us nothing but problems. It's all well and good when we're dealing with simple, regular outlines. But even in those cases the centroid is a calculated point, of no special status in terms of the parts of the form we're really interested in from a biological point-of-view. Also, as we saw in the last column, we can't even rely on the centroid staying put when it comes to interpolating the outline down to the level where it can be described by a reasonable number of harmonic amplitudes and phase angles. If we could just get rid of having to find and then rely on the outline's centre all of this geometric analysis of outlines business would be much easier. Fortunately, some clever people down the (metaphorical) hall have worked up some procedures that allow us to do precisely that.

Think of the *Ramulina* outline as a path; a (admittedly tortuous) way of getting from a convenient starting point, round the outline and back to the start. This pathway is complicated when looked at in its entirety. But it can be simplified if broken into small pieces.

To draw a practical analogy, when someone stops you on the street and asks for directions, the route they need to take to arrive at their destination could well be geometrically complex when drawn on a map. Still, that complexity can be broken down into a set of very simple sequential directions. The set of instructions you'd likely provide would run something like this.

Go north, down this street until you get to the second stoplight.  
Turn right onto Wabash.  
Go east on Wabash for three blocks.  
Turn left onto Cimarron.  
Go north on Cimarron until you come to a Stop sign, two blocks.  
Turn left onto Beaumont.  
Stay on Beaumont for four blocks and you're there, it's on the corner of Beaumont and Eastwood.

Assuming that each city block represents a unit distance — let's call it a 'step' — this set of directions is mathematically equivalent to the following.

Take 2 steps.  
Turn 90° clockwise.  
Take 3 steps.  
Turn 90° anticlockwise.  
Take 2 steps.  
Turn 90° anticlockwise.  
Take 4 steps.

This type of street-direction procedure was used by Charles Zahn and Ralph Roskies (1972) to develop a way to transform any closed curve no matter how complex into a single-valued, periodic, mathematical function that could be decomposed using Fourier analysis.

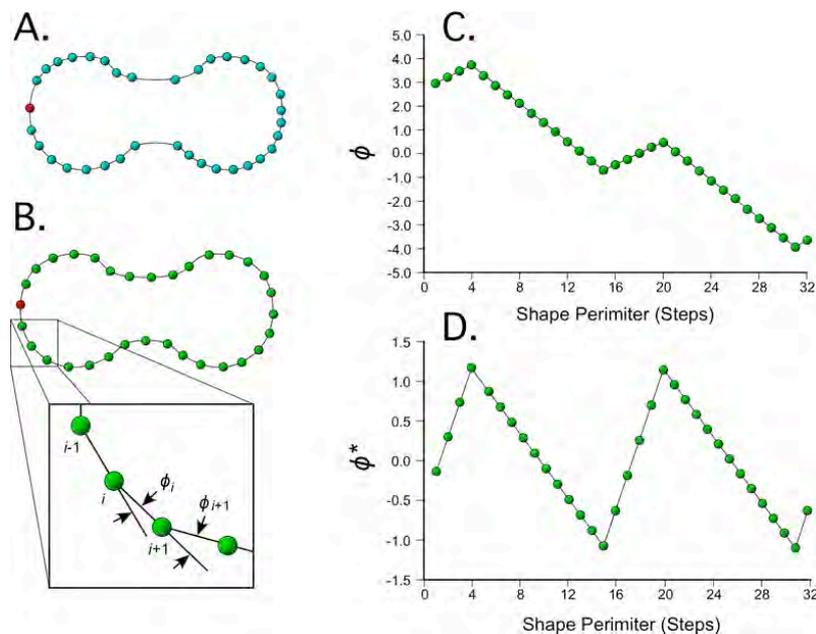


Figure 2. Steps in calculating the Zahn and Roskies (Z-R) shape function. A. original set of semilandmark data points placed on the periphery of a hypothetical shape. The red landmark represents the starting point for digitization. Ideally this point should be placed on a topologically homologous landmark. 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 as a series of net angular deviations (see text for discussion). C. the  $\phi$  form of the Z-R shape function with a typical ramp that denotes a closed curve. D. the  $\phi^*$  form of the Z-R shape function with represents the shape residual after removal of the ramp of circularity.

The Zahn and Roskies procedure begins with the collection of a set of  $x,y$  coordinates (or  $x,y,z$  coordinates of a three-dimensional analysis is required, see MacLeod 1999) along an outline or curve of interest (Fig. 2A). These points need not be equally spaced at the time they are collected and, if you are using a digitizer that requires you to place the points on the outline by hand, they won't be evenly spaced. Regardless, provided the set of collected points provides sufficient resolution in the parts of the outline where the bend is tight, it is a simple matter to search out and estimate a set of equally spaced points along the sampled outline via linear interpolation (Fig. 2B, see the *Palaeo-Math 101-2* spreadsheet for an example of these calculations).

Once a set of equally spaced semi-landmark coordinates has been estimated, the outline can be regarded as an  $n$ -sided polygon where  $n$  is the number of semilandmark points. Since the distance between each point is the same we only need to 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 represent the length of the outline. This length will often be a convenient size metric.<sup>1</sup> Size may be removed from consideration (and display) by setting the steplength to a unit value (e.g., 1.0) for all outlines in the dataset.

With size quantified and under our control, the shape of the outline can be represented in the ‘street direction’ manner alluded to above — as a series of angular turns that need to be executed in order to find the location of the next semi-landmark point and which is always exactly one steplength away. Mathematically this involves determining the angle between the adjacent sides of the polygon. There are many ways to calculate this set of angles, the most common of which is to regard adjacent polygon segments as vectors and use the vector dot product to find the angle.

The first step in this procedure is to calculate the displacements or distances of between adjacent semilandmark points in the  $x$  and  $y$  directions. Since these segments are defined by sets of three semilandmark points,  $(x_{i-1}, y_{i-1})$ ,  $(x_i, y_i)$ , and  $(x_{i+1}, y_{i+1})$  (see Fig. 2B, inset), the displacements in question are calculated as follows.

$$\begin{aligned} dx_1 &= x_i - x_{i-1} \\ dy_1 &= y_i - y_{i-1} \\ dx_2 &= x_{i+1} - x_i \\ dy_2 &= y_{i+1} - y_i \end{aligned} \quad (24.1)$$

Once these displacements have been obtained the dot product can be calculated.

$$c_i = ((dx_1 \cdot dx_2) + (dy_1 \cdot dy_2)) / \sqrt{((dx_1)^2 + (dy_1)^2) \cdot ((dx_2)^2 + (dy_2)^2)} \quad (24.2)$$

In order to get the orientation of this angle right the following rules must be applied.

$$\begin{aligned} \text{If } c_i^2 &\geq 1.0 & s_i &= 0.0 \\ \text{If } c_i^2 &< 1.0 & s_i &= \sqrt{1.0 - c_i^2} \end{aligned} \quad (24.3)$$

Once  $c_i$  and  $s_i$  have been calculated they can be used to calculate the angle between adjacent polygon segments using the arctangent function.

$$\phi_i = \tan^{-1}(s_i, c_i) \quad (24.4)$$

To ensure accuracy this value is estimated using high-precision numerical methods. Finally, in order to take advantage of the tangent function’s ability to locate angles in any of the four Cartesian quadrants, the sign of the  $\phi_i$ -value must be adjusted using a vector cross-product test.

$$\text{If } ((dx_1 \cdot dy_2) - (dx_2 \cdot dy_1)) < 0.0, \phi_i = -\phi_i \quad (24.5)$$

These calculations result in a set of angles —expressed in radians — the cumulative sum of which quantifies the net angular change around the perimeter of a shape from the user-chosen sequence starting point. This shape function can be expressed in its raw form ( $\phi$ , Fig. 2C) in which the cumulative radian values are used to represent the shape, or as a normalized shape function that expresses the degree and location of deviation from strict circularity ( $\phi^*$ , Fig. 2D). In the case of the latter the negative ramp of the former is removed by subtracting a cumulative constant term corresponding to the radian-equivalent value of 360° (6.2832) divided by number of steps used to subdivide the outline.

<sup>1</sup> If the curve of interest is the outline of the entire specimens the steplength may be used as an index of size. However, if the curve of interest represents some spatially restricted feature of the specimen the steplength may represent a biased size index if interpreted in a naïve manner though it will still represent and unbiased measure of feature size.

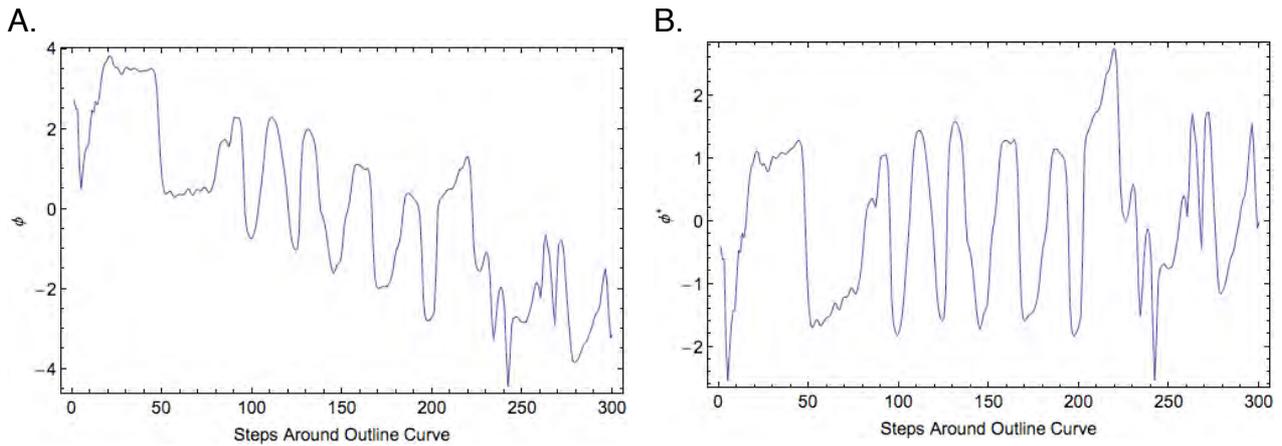


Figure 3. Zahn and Roskies shape functions for a 300-point interpolation of the *Ramulina globulifera* outline semilandmark data expressed in the  $\phi$  (A) and  $\phi^*$  (B) formats.

Figure 3 shows the Zahn & Roskies shape functions for a 300-point interpolation of the *R. globulifera* outline semilandmarks. As with the example shown in Figure 2, the broad negative ramp in the raw ( $\phi$ ) form of these data (Fig. 3A), along with the near equivalence of the absolute value of the starting and ending points ( $\phi$  and  $\phi^*$ ), identify this as the shape function of a closed outline. Local reversals in the trend of the shape function locate high-angle bends along the periphery of the shape (e.g., base of tubular spines where they join the spherical test body). From the standpoint of shape characterization and analysis, however, by far the most important aspect of these shape functions is the simple fact that, despite the complexity of the *R. globulifera* outline, the equivalent Zahn & Roskies shape function is a true mathematical function, with a single  $\phi$  value associated with every step along the entire outline. It is this property that allows Zahn & Roskies shape functions to be subjected to the Fourier (and other) data analysis procedure(s).

Before we move on to demonstrating the Fourier analysis part of Z-R Fourier analysis, I want to provide the equations that will allow you to transform any Z-R shape function back into its equivalent Cartesian form and say a word about spatial resolution. The back-transformation equations are trivially easy to apply.

$$\begin{aligned} x_i &= x_{i-1} - (\text{steplength} \cdot (\sin \phi_i)) \\ y_i &= y_{i-1} - (\text{steplength} \cdot (\cos \phi_i)) \end{aligned} \quad (24.6)$$

Applying these equations to the *R. globulifera* data shown in Figure 3 results in production of the following reconstructions (Fig. 4).

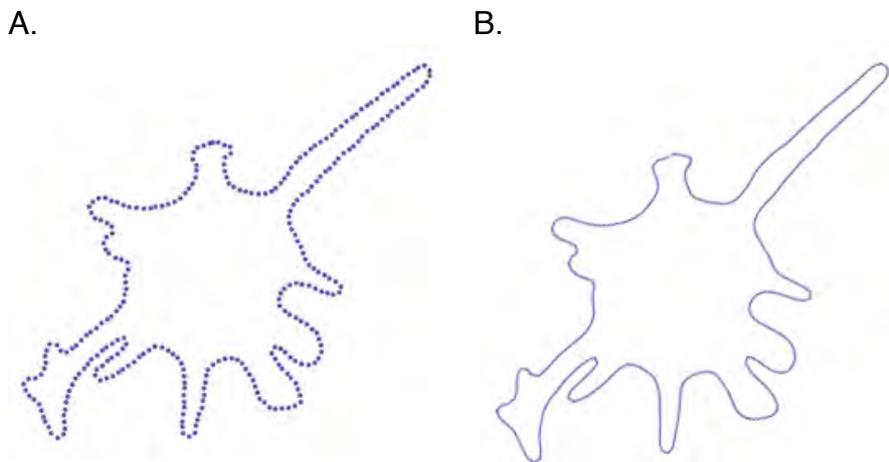


Figure 4. Reconstruction of the *Ramulina globulifera* outline from the  $\phi^*$  form of the Z-R function shown in Fig. 3B drawn as a series of 300 semilandmark points (A) and as a joined point series (B).

Note that this outline is distinctly smoother than the raw outline (see Fig. 1), with many of the small spatial irregularities absent and sharp angular bends transformed into more even curves. This smoothing is a by-product of the interpolation procedure. If a larger number of interpolated semilandmark points are used to represent the outline the steplength will decrease and small-scale features will be captured with greater accuracy. If a lesser number are specified the degree of smoothing will increase, sometimes to the point where important aspects of the outline are represented in a distorted manner or, depending on their relative size, not represented at all.

When performing any analysis that involves interpolation of raw data care must be taken to ensure that the spatial resolution selected captures those aspects of geometry that are important for shape characterization accurately and adequately across all specimens in the sample. Usually it's a good idea to reconstruct, plot, and inspect all the outlines you intend to use before you perform any subsequent analyses. That way any unanticipated sampling problems can be identified and corrected before time is spent attempting to interpret what may be artifactual results.

Once you've interpolated your outlines to a common number of equally spaced semilandmark points and transformed those into the angular-deviation format of their equivalent Z-R shape functions you are ready to submit the set of outlines to Fourier analysis. Zahn and Roskies (1972) recommended use of the  $\phi^*$  function for subsequent Fourier analysis as this 'rampless' function more closely approximates the form of a periodic function formed from radius vectors emanating from a central point. In principle either the  $\phi$  or  $\phi^*$  functions can be used as the subject of a Fourier analysis. However, the large difference between the values of the two endpoints of the  $\phi$  function usually means a larger Fourier harmonic spectrum must be used to accurately estimate the shape of a set of  $\phi$  functions than is needed to accurately estimate the forms of a set of  $\phi^*$  functions. Whichever form of the Z-R function is used the Fourier part of the procedure is performed in a manner identical to the one we used to analyse the *Sassia* outline in the last column. Results of using 25 Fourier harmonics to model the *R. globulifera* Z-R ( $\phi^*$ ) shape function are shown in Figures 5.

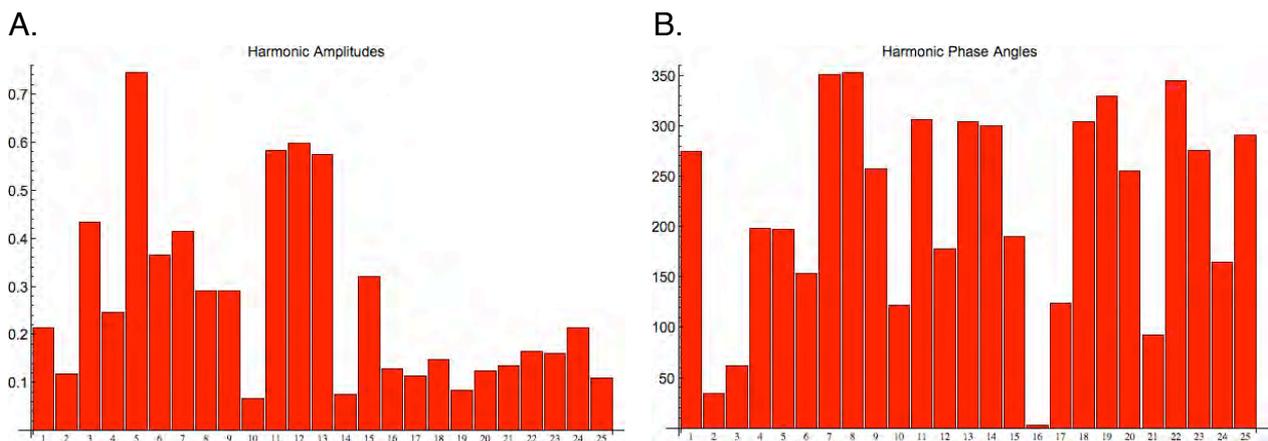


Figure 5. Fourier harmonic amplitudes (A) and phase angles (B) for a 25-term decomposition of the 300-point Z-R ( $\phi^*$ ) shape function of *Ramulina globulifera*

The overall quality of this modelling result can be assessed by using the 25 harmonic amplitudes and phase angles to estimate the original ( $\phi^*$ ) shape function (Fig. 6).

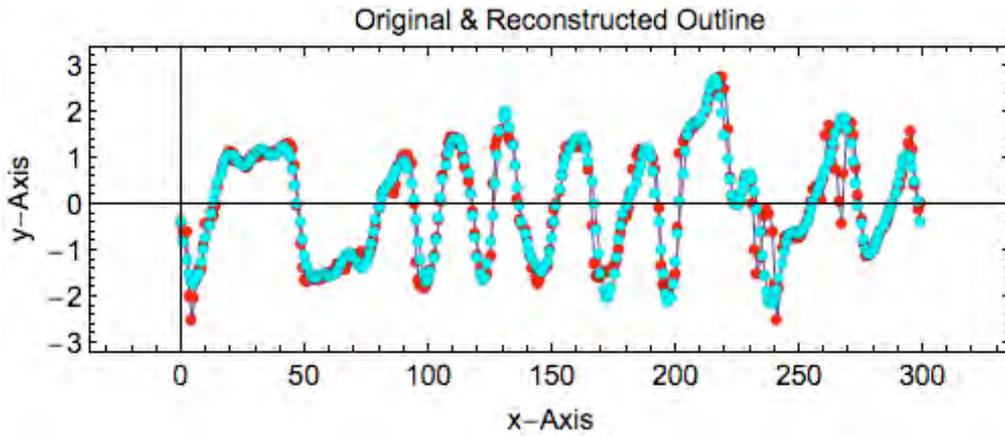


Figure 6. Overlay plot of estimates (red) and reconstructed versions of the *Ramulina globulifera*  $\phi^*$  shape function based on amplitude and phase angle data from a 25-term Fourier harmonic spectrum.

Even using this relatively small number of harmonic descriptors, the form of the original 300-point shape function is represented to a remarkably high degree of accuracy. For a more direct assessment of shape variation the estimated shape function could be transformed into its Cartesian equivalent using equations 24.6 and compared to the original image. Of course, the level of accuracy could be improved further through expansion of the Fourier harmonic spectrum. Given a 300-point Z-R shape function the number of potential harmonics that could be used to describe the *R. globulifera* outline is 149.

We are now ready to review an example of how Z-R shape functions can be used to analyze real palaeontological data. Consider the set of benthic foraminifer shapes presented in Figure 7.

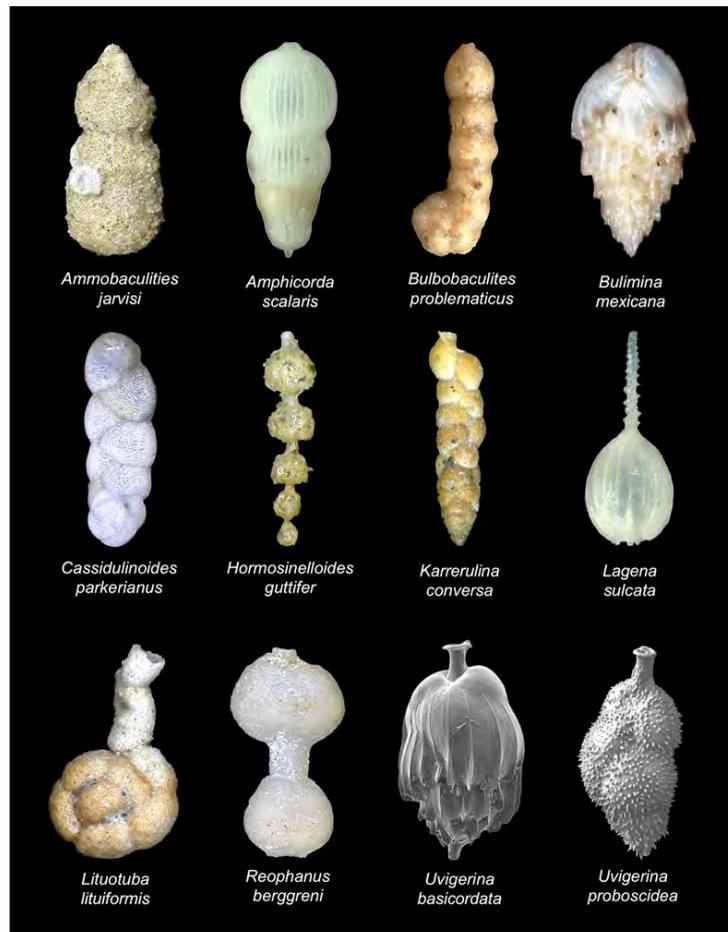


Figure 7. Benthic foraminiferal shapes used in the example analysis. Note preponderance of multi-valued object outlines.

Many of these species possess multi-valued outlines. Accordingly, this sample cannot be analyzed using the radial Fourier method. But all of these shapes possess an outline that can be converted to a single-valued mathematical function using the Zahn and Roskies (1972) technique. Please note the Zahn and Roskies procedure works as well for specimens with true single-valued outlines as it does for specimens whose outlines are multi-valued. As a result it is usually a good generalized choice for an outline shape-characterization strategy, useful for analysing samples consisting of known single-valued outlines, known multi-valued outlines, mixed single-valued and multi-valued outlines, or samples whose outline-value states you are either unsure of or can't be bothered to assess.

The complete set of  $\phi^*$  shape functions for the set of 12 species shown in Figure 7 is plotted in Figure 8.

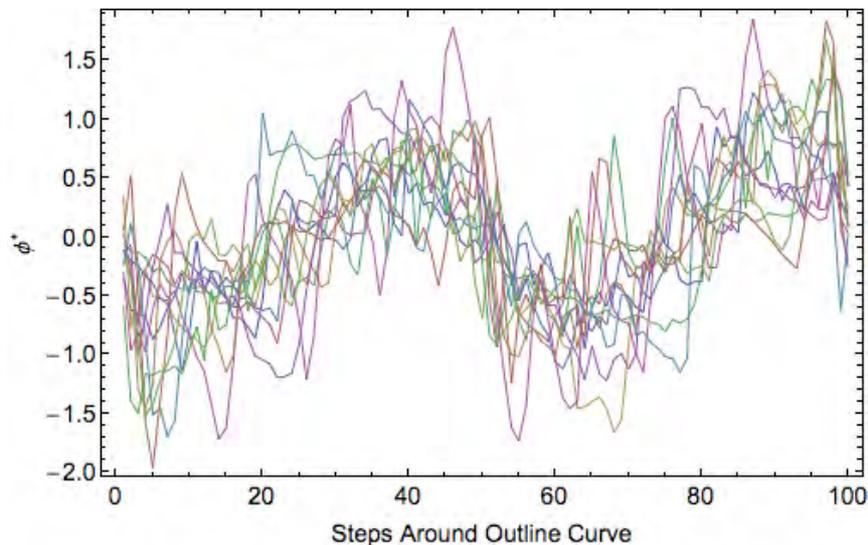


Figure 8. Swarm of 12 Z-R ( $\phi^*$ ) shape functions for the benthic foraminifer test outline shapes.

This type of overlay diagram is analogous to the plots of *Procrustes*-aligned landmark configurations I've shown you in previous columns (see MacLeod 2008a, 2009a-c). Just as with *Procrustes* alignment, the semilandmark data have been projected from a form space into a shape space in which translation, scaling, and rotational differences between specimens have all been removed from the system of observations. At each of the 100 interpolated steps around the outline of each specimen (*x*-Axis) 12 angular rotations specify the direct to the next semilandmark in the sequence which is equally distant from the preceding landmark on a per specimen basis. The mean shape can be calculated as the average of the angular values across all specimens in the sample at each step.

As you can see from Figure 8, all these outlines are broadly similar with angular deviations beginning just below the 0.0 value and descending to a diffusely defined nadir between steps 5 and 10. These deviations then increase to a local maximum in the vicinity of step 40. This trend reverses itself as we continue around the outline where the angular deviations descend the  $\phi^*$  scale to the a local minimum located between steps 55 and 80. Positive angular deviations then predominate, rising to a complexly structured peak between steps 75 and 97 after which they fall sharply back to the origin of the  $\phi^*$  axis. About this sample-level trend, exist a complex suite of interesting shape variation patterns. Recall also that Figure 8 presents us with a picture not of the geometries of the Figure 7 shapes themselves, but how those shapes differ from a circle.

As with radial Fourier analysis, we can use the Fourier expansion to define a set of geometrically comparable and mutually independent reference shapes (= the Fourier harmonic series) that can be used as a set of variables with which we can describe the complex patterns of shape variation we see in Figure 8. A set of example amplitude and phase angle spectra for a 25 harmonic solution is shown in Figure 9. Close comparison of barcharts for both the amplitude and phase angle descriptors shows that they are indeed picking up the similarities and the differences that exist between these shapes.

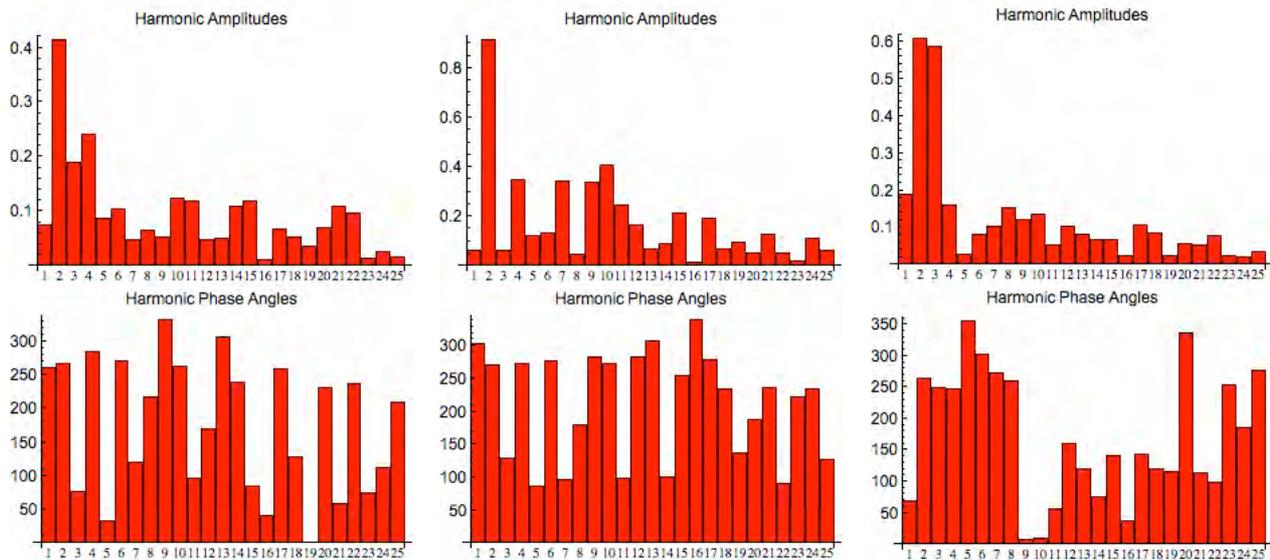
*Bulimina mexicana**Hormosinelloides guttifer**Lituotuba lituiformis*

Figure 9. Images (upper row), harmonic amplitudes (middle row) and harmonic phase angles (lower row) for a 25-term Fourier description of a 100-point Z-R ( $\phi^*$ ) shape functions for three example benthic foraminifer species. Note both similarities and differences in the harmonic amplitude and phase angle spectra, which reflect similarities and differences in the outline shapes of these species.

Once we have described the set of foraminifer test shapes in terms of their component Fourier harmonic shape variables we can compare and contrast the amplitude and/or phase angle data using an appropriate multivariate data analysis procedure. If we are interested in assessing the major patterns of variation in our shape data we can use principal components analysis (PCA, see MacLeod 2005). If we are interested in obtaining an image of shape similarities and differences among these species we could also use PCA, but an arguably better and certainly more interesting choice would be multi-dimensional scaling (MDS, see MacLeod 2008b). And if we have subsidiary groupings in our data and want to understand how this group-level structure is reflected in the shape data we have collected we can apply canonical variates analysis (CVA, see MacLeod 2007). The important point to remember here is not to simply apply a procedure because you've seen someone else apply it to their data or because you've heard that this procedure is 'accepted' by this or that school of thought. Take the time to understand how your problem, and how your data, match up to the capabilities and assumptions of the different data-analysis procedures, then go with the one you believe to be appropriate after researching the subject for yourself. If you haven't got time — or aren't interested — in doing this research yourself, the safest course of action is always to bring in an advisor or a collaborator who is interested in this aspect of the investigation and let them guide you.

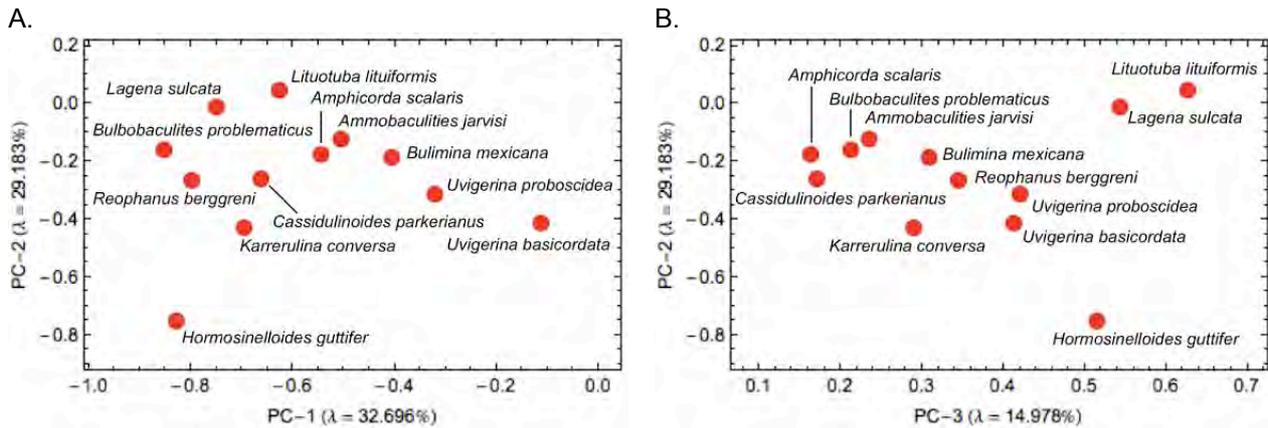


Figure 10. Ordinations of the 12 benthic foraminiferal outlines in the subspace formed by the first three principal components (covariance-based) of the 25 term Fourier amplitude characterization of the 100-point Z-R ( $\phi^*$ ) benthic foraminifer shape functions. See text for discussion.

For the purposes of this column let's take a look at the major trends in Fourier descriptor-based shape variation using PCA. Because there are only 12 specimens in the dataset there are only 12 principal components with positive eigenvalues, the first seven of which account for over 95 percent of the observed shape variation. Inspection of the ordinations of the first three of these axes (76.86% of shape variation represented, see Fig. 10) suggests that they are indeed reasonable. Along PC-1 species characterized by thin, uniserial tests (e.g., *Hormosinelloides guttifer*, *Lagena sulcata*) are separated from species characterized by wide, flaring tests (e.g., *Uvigerina basicordata*, *Uvigerina proboscidea*). The second PC axis separates forms with a large, ovate proximal and narrow distal shape (e.g., *Lagena sulcata*, *Lituotuba lituiformis*) from species with a uniformly narrow shape throughout (e.g., *Hormosinelloides guttifer*). The third axis separates species with the most pronounced body-neck discrepancy (e.g., *Lagena sulcata*, *Lituotuba lituiformis*) from those serial forms whose outline is either modestly flaring (e.g., *Amphicorda scalaris*) or sub-uniform throughout (e.g., *Cassidulinoides parkerianus*). Overall, this analysis is analogous to either a *Procrustes* PCA or a relative warps analysis performed on landmark data.

The Zahn and Roskies version of Fourier analysis has not been popular among either biologists or palaeontologists because it's been superceded by other approaches to outline characterization and analysis. I've presented it in its original guise in order to clarify the origins of the Z-R shape function and to provide space for a fairly detailed discussion of its two variants,  $\phi$  and  $\phi^*$ . Here we've been mostly concerned with the  $\phi^*$  variant. I'll deal with the  $\phi$  variant in detail in a future discussion of eigenshape analysis. Regardless, there no reason not to use the Z-R version of Fourier analysis as a more generalized approach to the Fourier-based analysis of outlines than the traditional radial Fourier approach.

With respect to software, other than my own programs and software based on the algorithms I use (e.g., Morpho-tools, <http://www.morpho-tools.net/>) the only public domain implementation of the Z-R shape function of which I'm aware is included in the PAST data analysis package (<http://folk.uio.no/ohammer/past/>, see also Hammer and Harper 2006). Claude (2008) includes a listing of R code to calculate interpolations and the Z-R shape function, but a knowledge of R programming is required to implement this code. The *Palaeo-Math* 101-2 spreadsheet contains a simple example of the calculations involved. However the equal point-spacing interpolation algorithm is a bit tricky — involving successive approximation to the idealized configuration — and the Z-R transformation algorithm requires the serval passes through the data in order to adjust the geometric details appropriately. Irrespective of these complications, the pay-off for surmounting them is possession of a very generalized algorithm that, when employed properly, provides users with the ability to quantify a large number of comparisons systematists make on a routine basis, albeit qualitatively, imprecisely, and inconsistently.

**Norman MacLeod**  
 Palaeontology Department, The Natural History Museum  
[N.MacLeod@nhm.ac.uk](mailto:N.MacLeod@nhm.ac.uk)

## REFERENCES

CLAUDE, J. 2008. *Morphometrics in R*. Springer, Amsterdam 318 pp.

- DAVIS, J. C. 2002. *Statistics and data analysis in geology (third edition)*. John Wiley and Sons, New York 638 pp.
- HAMMER, Ø. and HARPER, D. 2006. *Paleontological data analysis*. Blackwell Publishing, Oxford, UK 351 pp.
- MacLEOD, N. 1999. Generalizing and extending the eigenshape method of shape visualization and analysis. *Paleobiology*, **25**, 107–138.
- MacLEOD, N. 2005. Regression 4: Going Multivariate. *The Palaeontological Association Newsletter*, **58**, 44–53.
- MacLEOD, N. 2007. Groups II. *Palaeontological Association Newsletter*, **65**, 36–49.
- MacLEOD, N. 2008a. Size and shape coordinates. *Palaeontological Association Newsletter*, **69**, 26–36.
- MacLEOD, N. 2008b. Multidimensional scaling and ordination. *Palaeontological Association Newsletter*, **67**, 26–44.
- 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.
- 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](http://www.palass.org/modules.php?name=palaeo_math&page=1)