Palaeo-Math 101

Centroids, Complex Outlines & Shape Functions

In the last column we learned how to used a powerful mathematical technique — the Fourier series — to characterize the shape of any single-valued outline no matter how complex by breaking it down into a set of consistently defined geometric descriptors (= shape variables) that we could then use to both analyze and model patterns of shape variation within any sample. However, as powerful as Fourier analysis is, the classic or 'radial' approach has several built-in disadvantages. Chief among these is the requirement that all outlines included in the sample be 'single-valued' (Fig. 1).

Figure 1. Fossil specimens exhibiting single-valued outlines.

Geometers call a closed curve 'single-valued' when any radius vector drawn from the outline's centre crosses the curve in one and only one location. For this class of outlines the radius-vector sampling scheme we discussed and illustrated last time effectively transforms the outline into an empirically defined mathematical function.[1] Once a set of outlines has been re-described in terms of their function-equivalent geometries, it's possible to use the Fourier series to tease apart their forms/shapes and assess the sample for patterns of form/shape similarity and difference. However, many biological forms are characterized by multi-valued outlines, in which at least some radius vectors cross the boundary at more than one location or in which the very idea of an outline centre is problematic for one reason or another (e.g., the mean *x,y* coordinate location falls outside the object's boundary, see Fig. 2).

Figure 2. Fossil specimens exhibiting multi-valued outlines.

These curves cannot be analysed using a standard radial Fourier sampling scheme because they cannot be transformed into valid mathematical functions. In these cases the trick is to find a way of converting the complex outline into a configuration that (1) preserves as much of the geometric information of relevance to the scientific question at hand as possible and (2) has the form of a mathematical function. Before we begin our discussion of non-radius vector-based shape functions though, we need to take care of an ugly little detail left over from our previous discussion of radial Fourier analysis.

This detail focuses on calculation of a single-valued outline's centre or centroid. Radial Fourier analysis requires location of the centre because it is from that point that set of radius vectors used to describe the

---

[1] In mathematics a function is a relation in which any input value (x) has exactly one output value (y). Hence the expression $x + 2 = y$ is a function whereas the expression $x + 2 = 3$ is not.

outline emanate. As you may recall, a basic assumption of radial Fourier analysis is that the set of adjacent radius vectors subtend equal angles as they move around the outline. This ensures that the form or shape has been sampled evenly and — more importantly — that the mathematical representation of the form/shape has not been biased by inconsistencies in the placement of the radius vectors relative to each specimen's geometry.

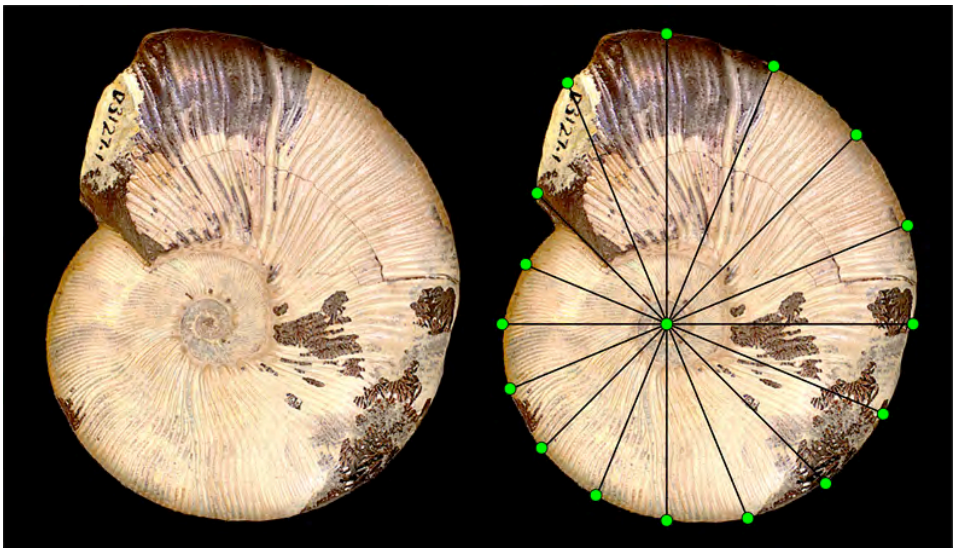

Figure 3. Fossil maoritid ammonite sampled using an equi-angular radius vector sampling scheme with the scheme's centroid being placed at the position of the specimen's proloculus. Note this centroid location is not synonymous with the outline's geometric centroid. Nevertheless, this location has the advantage of being able to be located on (virtually) every maoritid specimen and represents a point of unquestionable biological and geometric significance.

In the simplest of situations there will be some sort of landmark point that lies relatively close to the form's center can be found on all specimens in the sample. In this case the data analyst is perfectly justified in using this landmark point as the shape's 'center' from which a set of coordinate points can be located such that the angles between adjacent radius vectors are equal (Fig. 3). Since this landmark point is defined by a consistently and universally relocatable point defined and accepted *a priori* as the reference point for the geometric description of each shape, the equi-angular sampling criterion will always be true.

But what happens if we don't have an objectively locatable landmark point in this region of the shape that can be used as the reference? The fallback convention is to calculate the geometric centroid of the outline as the mean of all *x*-coordinate values and the mean of all *y*-coordinate values.

$$\bar{x} = \sum_{i=1}^{n} x_i \Big/ n \tag{23.1}$$

$$\bar{y} = \sum_{i=1}^{n} y_i \Big/ n \tag{23.2}$$

Where:

x_i = i^{th} *x*-value

y_i = i^{th} *y*-value

n = total number of specimens in sample

Once this centroid has been obtained it can be used, first to mean-centre the outline and then to calculate an initial estimate of the raw set of radius vectors by converting the x_i,y_i coordinate values into their r_i,Θ_i polar coordinate equivalents.

$$c_i = \sqrt{(x_i^2 + y_i^2)} \tag{23.3}$$

$$\theta_i = \tan^{-1}\left(y_i / x_i\right) \tag{23.4}$$

Next, a set of new radius vectors is calculated such that angle subtended between adjacent vectors is equal. The maximum number of Fourier amplitude and phase angle coefficients (= harmonics) that can be calculated from any given collection of boundary outline coordinates is set by the following relation.

$$k = (n-1)/2 \text{ , if } n \text{ is odd}$$
$$k = n/2 \text{ , if } n \text{ is even} \tag{23.5}$$

In these equations k is the number of Fourier harmonics and n the number of x,y points used to describe the outline. This relation is often referred to as the Nyquist frequency. If the Fourier series is expanded beyond the limit set by the Nyquist frequency errors will result due to aliasing of the spatial signal.

From a practical point of view the problem the Nyquist frequency limit imposes on Fourier calculations is one of interpolation. These days it's almost always the case that digitizers collect boundary outline coordinates that are not arranged in an equiangular series with respect to any central point. Conversion of a sequence of outline coordinates to an equiangular series usually amounts to working through the following procedure.

1. Deciding how many harmonics are necessary to describe the form(s) under consideration adequately
2. Calculating the angle between successive radius vectors as $\theta = 360/2k$
3. Determining the lengths of the $2k$ equiangular radius vectors by searching the original data that have been converted to polar coordinate form, locating empirical radius vectors that lie on either side of the desired radius vector, and estimating the length of the desired radius vector via linear interpolation

The radius vectors calculated as a result will be equiangular relative to the initial outline centroid, the position of which was estimated using all the coordinate points in the digitized outline (equations 23.1 and 23.2). Unfortunately, this does not mean these $2k$ radius vectors will be equiangular with respect to their own centroid. As the Fourier series equations we used in the last essay assume strict equi-angularity among the radius vectors, any deviation from this condition will introduce error into the calculation of the harmonic amplitudes and phase angles.

Schwarcz and Shane (1969), Full and Ehrlich (1982) and Boon *et al.* (1982) were the first to bring this problem to the attention of the geological community, originally in the context of the analysis of sedimentary particle shape. To resolve this problem they recommended comparing the centroid of the set of radius vectors that will be used to calculate the Fourier harmonics (= harmonic spectrum centroid) to the initial centroid used to calculate the total set of radius vectors. If the harmonic spectrum centroid lies within a tolerance envelope about the initial centroid no adjustment need take place. Full end Ehrlich (1982) recommend this tolerance envelope have a value of '0.007 pixel values', which seems to be an empirically determined limit based on their experience with sand grain shape analyses. My own experiments with radial Fourier centroid estimation suggest that a tolerance envelope about the initial centroid of 1.0 percent of the outline's maximum x, or maximum y dimension (whichever is longest) delivers approximately the same level of consistency.

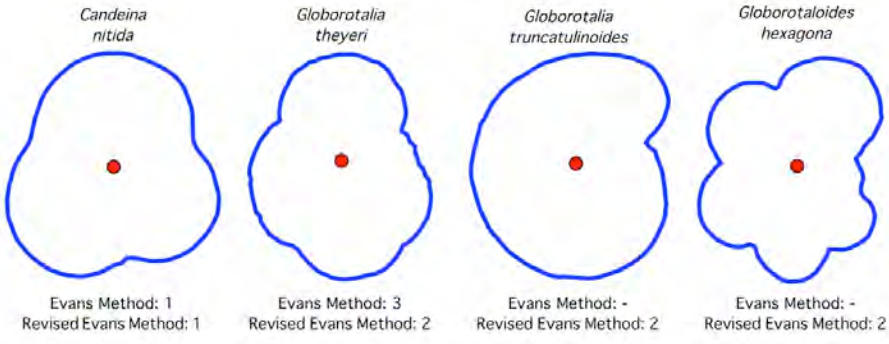

Figure 4. Outlines of four planktonic foraminifer species with statistics on how many adjustment iterations were required to find a stable centroid location using both the Evans and Revised Evans methods. Note that the centroid of *Globorotalia truncatulinoides* and *Goborotaloides hexagona* did not converge even after 40 iterations under the Evans Method.

Obviously, if there is no need for adjustment of the initial centroid, radial Fourier analysis can proceed as outlined in the previous column (see also Fig. 4). In some instances though, the positions of the initial and harmonic spectrum centroids will differ by a value greater than this tolerance envelope. For these cases Full and Ehrlich (1982) offer two iterative estimation procedures.

The first is termed the 'Evans Method' after David Evans who devised the solution originally (see also Boon *et al.* 1982). This method involves drawing a chord from the harmonic spectrum centroid back to the initial centroid and locating a new centroid at a position equal to twice the deviation between these two centroids but in the opposite direction. Algorithmically, this new centroid value can be found as follows.

$$\hat{x}_h = \bar{x}_h - 2\bar{x}_i \tag{23.6}$$

$$\hat{y}_h = \bar{y}_h - 2\bar{y}_i \tag{23.7}$$

Where:

$\bar{x}_{\text{initial}}, \bar{y}_{\text{initial}}$ = coordinates of the initial centroid

$\bar{x}_{\text{hsc}}, \bar{y}_{\text{hsc}}$ = coordinates of the (old) harmonic spectrum centroid

$\hat{x}_{\text{hsc}}, \hat{y}_{\text{hsc}}$ = coordinates of the (new) harmonic spectrum centroid

Once calculated, the new estimate of the harmonic spectrum centroid can be used to recalculate the polar-coordinate transformation of the original *x,y* outline data and the harmonic spectrum radius vectors. The tolerance envelope test is then repeated. If the new initial and harmonic spectrum centroids fall within the tolerance envelope, the estimation procedure is terminated and the radial Fourier spectrum calculated. If not, the centroid is re-estimated again using equations 23.6 and 23.7, after which all calculations are repeated.

A number of empirical studies have reported that this procedure is usually sufficient to stabilize the centroid locations for the majority of single-valued, closed curve outlines, usually within ten centroid-estimation iterations or less (Full and Ehrlich, 1982, Healy-Williams 1983, Pharr and Williams 1987, Healy-Williams et al. 1997). For those outlines whose centroid does not converge using the Evans Method, Full and Ehrlich offered a 'Revised Evans Method' which locates the new estimate of the harmonic spectrum centroid as the point mid-way between initial and (old) harmonic spectrum centroids. In terms of calculations, the Revised Evans Method can be implemented as follows.

$$\hat{x}_h = \bar{x}_h - 0.5\bar{x}_i \tag{23.8}$$

$$\hat{y}_h = \bar{y}_h - 0.5\bar{y}_i \tag{23.9}$$

These authors claim that the Revised Evans Method can find stable centroids for approximately half of the single-valued outlines whose centroids failed to converge under the Evans Method (see Fig. 4). Still, a rump of outlines is left whose centroids fail to converge under either method.

Inspection of Figure 4 also suggests some rough guidelines that could be useful for determining whether an outline is likely to run afoul of the centroid-estimation problem. Based on this analysis, as well as my own experience, outlines composed of two or more unequal lobes (e.g., *Globorotalia truncatulinoides*, *Goborotaloides hexagona*) are often problematic. This is because the number of radius vectors falling into each of the two lobes can differ with the local size differential between the lobes often accentuating the effect of that difference. In these cases the centroid estimate often settles into a quasi-stable oscillatory pattern outside the tolerance envelope. Somewhat counter-intuitively three-lobed (e.g., *Candeina nitida*) or four-lobed (e.g., *Goborotalia theyeri*) outlines don't seem to suffer from centroid instability problems to anywhere near as great an extent as two-lobed and some multi-lobed forms. Also, based on my experience, the Revised Evans Method does indeed turn in a better performance in finding a stable centroid than the standard Evans Method, especially if relatively small numbers of harmonic amplitudes are being used to characterize the shape.

What difference does it make to a radial Fourier analysis if you don't get the centroid right? Figure 5 shows the result of using the initial and tolerance envelope-adjusted centroid for *Globorotalia truncatulinoides* to calculate the harmonic amplitude spectrum.

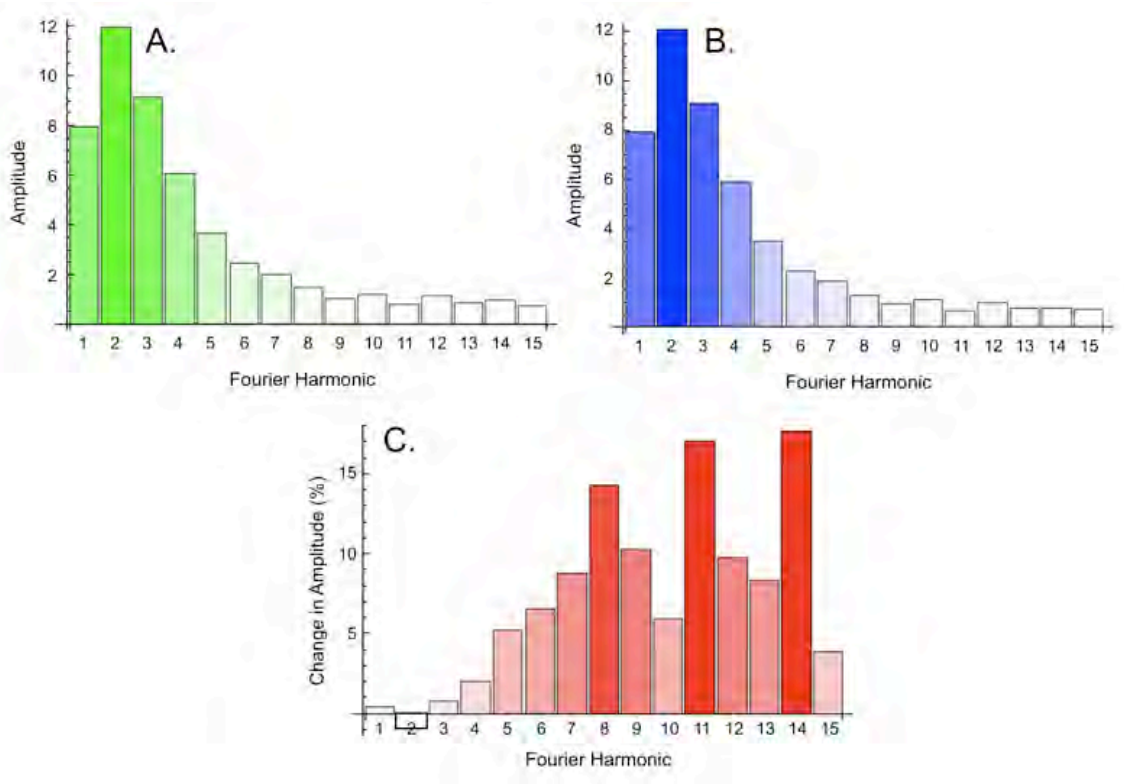

Figure 5. Harmonic amplitude spectra for *Globorotalia truncatulinoides* using the raw outline centroid (A) and the Revised Evans Method adjusted outline centroid. Although the spectra for these two analysis may appear superficially similar. Calculation of the percentage difference of the harmonic amplitude values (C) shows that the shift in centroid location had both a significant and a highly unpredictable effect of the Fourier amplitude values, with three harmonics exhibiting a greater then 10% change.

Full and Ehrlich (1982) provide a theoretical discussion of the effect centroid repositioning has on calculation of the radial Fourier harmonic spectrum. For the purpose of this discussion it is sufficient to point out the magnitude and non-linearity of the deviations in the harmonic amplitude spectrum, Since these amplitudes represent the independent 'characters' used by Fourier analysis to summarize and model shape variation, instabilities on the order of 10 percent in the values of these parameters — due entirely to geometric inconsistencies in centroid placement — should be avoided wherever possible.

In addition to this issue of instability of the harmonic spectrum, one must also consider the fact that it is impossible to obtain centroid convergence for some outlines. When this occurs two options present themselves. Either the unstable outline must the eliminated from the dataset, or some manner of representing the geometry of the objects' outlines that does not require location of each outline's center must be employed. Fortunately, a number of strategies have been developed to describe outline shape variation without having to find the outline's center, not only for case of pathological single-valued outlines, but additionally for the far more common situation in which the objects under consideration (or some subset thereof) are characterized by multi-valued outlines. It is to these more generalized shape-characterization approaches that we will now turn our attention.

Oddly enough, the oldest of these procedures involves a form of image processing that strikes many data-analysts as rather extreme. For those objects or images in which one axis is markedly longer than the other giving rise to multiple-value outline issues as a result of pathological variation in the outline in regions of the form remote from the centre, it is often possible to solve the problem by slicing the image into two halves along the long axis and pivoting one of the halves so that its *x*-pixel coordinate values are reversed (Fig. 6). This has the effect of 'unfolding' the outline along the specimen's long axis, and in so doing transforming the closed outline curve into a periodic waveform. Such periodic data are exactly the sort that Fourier analysis was developed to analyse originally. Accordingly, analysis of these wave-form data proceeds in a straight-forward manner. The curve is digitized at equally-spaced intervals at a resolution that corresponds to twice the number of Fourier terms desired in the harmonic spectrum and the locations of these points along the *y*-axis (= equivalents to the lengths of the radius vectors) recorded.

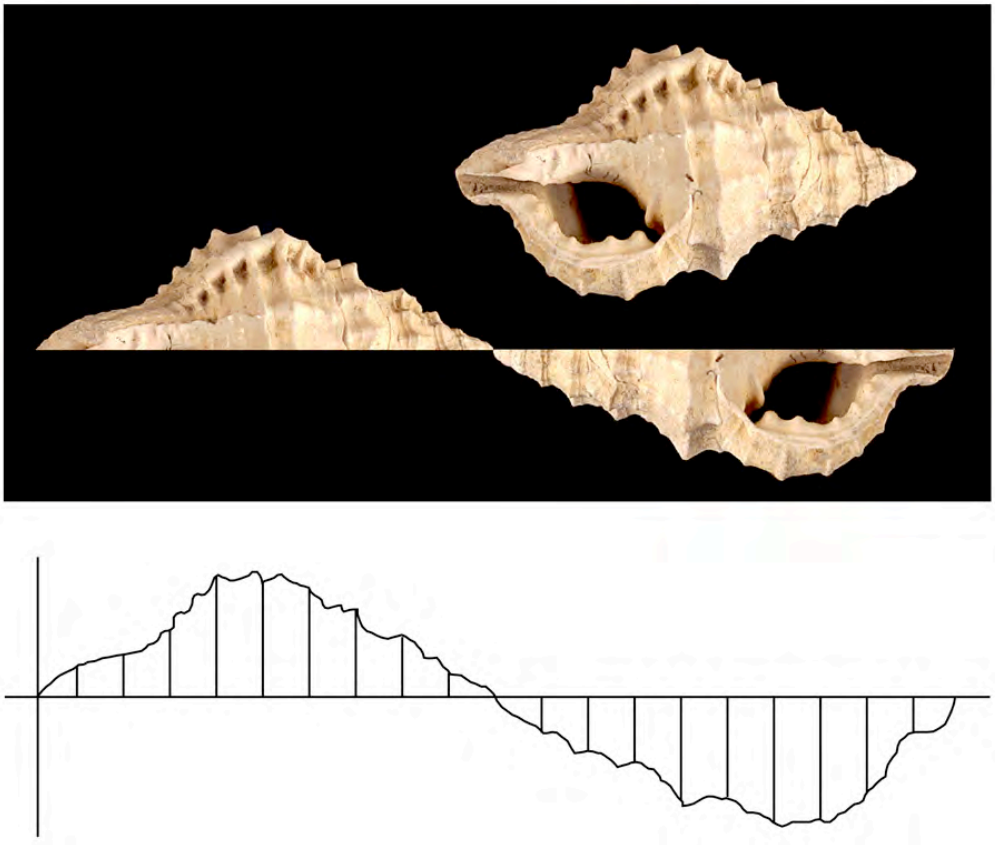

Figure 6. Transformation of the boundary outline curve characterizing the fossil gastropod *Sassia* to a periodic waveform by slicing the image in half along the specimen's long axis and pivoting or reflecting the lower (= left) half at the position of the proloculus such that the outline forms a continuous, single-valued curve. See text for discussion.

Revising the notation we developed for the radial Fourier series calculations, we can analyse the curve presented in the lower portion of Figure 6 using the following (standard-form) Fourier series equations.

$$y_j = \bar{r} + \sum_{j=1}^{k}\left[a_j \cos\left(j \cdot \beta\right) + b_j \sin\left(j \cdot \beta\right)\right] \qquad (23.10)$$

Where:

r = length of a sampled (radius) vector along the *y*-axis

β = angle of sampled vector in radians

$\bar{r}$ = average of all sampled (radius) vectors

j = Fourier harmonic number

k = total number of harmonics in Fourier series

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

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

The amplitudes of the sine and cosine terms for equation 23.10 can be calculated using the following expressions.

$$a_j = \frac{2}{n}\sum_{i=1}^{n} r_i \cos(j \cdot \beta_i)$$
$$b_j = \frac{2}{n}\sum_{i=1}^{n} r_i \sin(j \cdot \beta_i) \qquad (23.11)$$

Where:

n = total number of sampled points along empirical curve

r_i = distance between i^{th} point and y-axis

j = Fourier harmonic number

β_i = angle of the i^{th} radius vector in radians

Finally, the values of the amplitude and phase angles for each terms in the harmonic spectrum can be calculated using these standard expressions (equations 23.3 and 23.4_.

The harmonic spectrum for the first 15 terms of the *Sassia* Fourier series calculated on the basis of the waveform curve shown in Figure 6. A comparison of original and reconstructed outlines based on these 15 harmonics, are shown as figures 7 and 8.

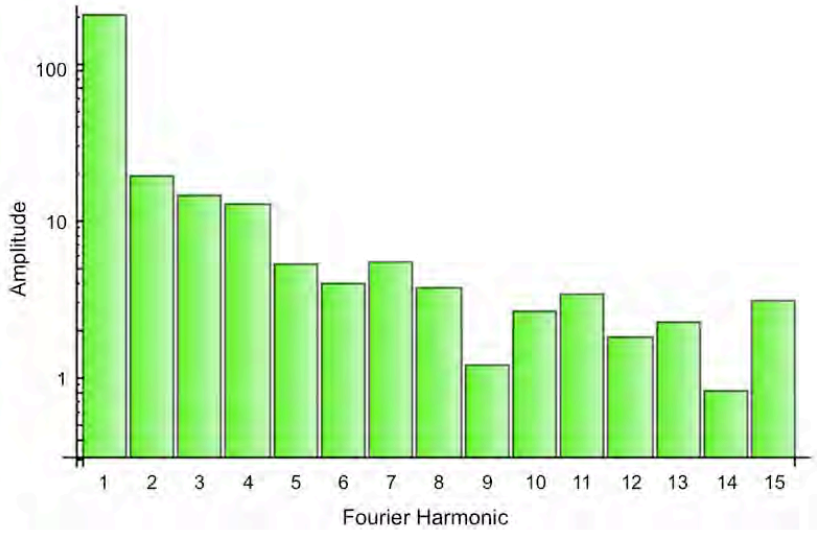

Figure 7. First 15 amplitude (c_j) values in the harmonic spectrum of the *Sassia* processed outline. Note logarithmic scale indicating that the overwhelmingly predominant shape component is that of a single sinusoidal waveform of length 1.0. See text for discussion.

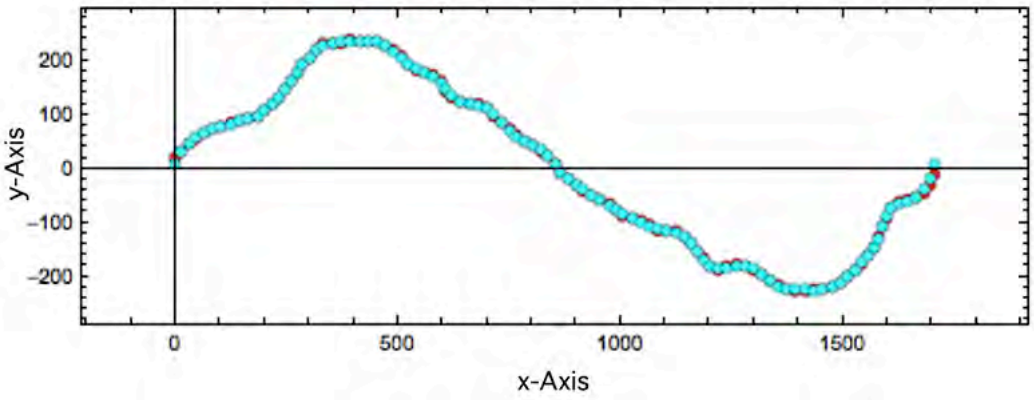

Figure 8. Original (red) and reconstructed (cyan) *Sassia* outline curves calculated based on a 15 harmonic Fourier amplitude and phase angle spectrum.

Although we're still at the beginning of our discussion of outline-based data analysis methods, I hope you can already appreciate the power of techniques such a Fourier analysis for describing, quantifying, and modelling organic forms in a way that has meaning in a wide range of biological and systematic contexts. Use of these methods has been somewhat eclipsed by the understandable enthusiasm with which landmark-based approaches to shape analysis were embraced by the morphometric community, aided and abetted by an inflexible — almost ideological — stance on Fourier analysis taken by several early proponents of geometric morphometrics. Over the next several essays we'll work our way through a set of increasingly more sophisticated and generalized approaches to outline analysis until finally arriving a true syntheses between these two (supposedly) separate approaches to form/shape characterization.

In terms of software, virtually all higher-level statistical data analysis packages for personal computers implement one of more Fourier analysis routine. While discrete-form, radial Fourier analysis is rarely included in these packages, a little work understanding what their Fourier routines are designed to do usually results in the identification of a procedure or modification of the data format that should allow you to implement Fourier analysis yourself. Unfortunately, the problems inherent in the centroid stabilisation issue are unique to radial Fourier analysis and so are not covered by any pre-programmed package with which I am familiar. Regardless, it's an easy matter to program a simple Excel spreadsheet routine that will allow you to check radius vector datasets to determine whether any of your specimens have a problem with unstable centroid location (see the *Palaeo-Math* 101 spreadsheet). This having been said, the fact that so few data analysts make use of landmark points that lie at or close to the forms central region has always struck me as odd. If such a landmark is available not only is the centroid stability issue easily and elegantly avoided, the biological interpretability of the shape analysis as a whole is often improved dramatically.

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

REFERENCES

BOON, J. D., III, EVANS, D. A. and HENNIGAR, H. F. 1982. Spectral information from Fourier analysis of digitized grain profiles. *Mathematical Geology*, **14**, 589-605.

FULL, W. E. and EHRLICH, R. 1982. Some approaches for location of centroids of quartz grain outlines to increase homology between Fourier amplitude spectra. *Mathematical Geology*, **14**, 43–55.

HEALEY-WILLIAMS, N. 1983. Fourier shape analysis of Globorotalia truncatulinoides from late Quaternary sediments of the southern Indian Ocean. *Marine Micropaleontology*, **8**, 1–15.

HEALEY-WILLIAMS, N., EHRLICH, R. and FULL, W. 1997. Closed-form Fourier analysis: a procedure for extracting ecological information from foraminiferal test morphology. *In*). *Fourier descriptors and the applications in biology*. Cambridge University Press, Cambridge, 129–156 pp.

PHARR, R. B. and WILLIAMS, D. F. 1987. Shape changes in *Globorotalia truncatulinoides* as a function of ontogeny and paleobiogeography in the southern ocean. *Marine Micropaleontology*, **12**, 343–355.

SCHWARCZ, H. P. and SHANE, K. C. 1969. Measurement of particle shape by Fourier analysis. *Sedimentology*, **13**, 213–231.

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