IMPROVEMENTS TO THE METHOD OF FOURIER SHAPE ANALYSIS AS APPLIED IN MORPHOMETRIC STUDIES

by A. JOHN HAINES *and* JAMES S. CRAMPTON

ABSTRACT. Fourier outline shape analysis is a powerful tool for the morphometric study of two-dimensional form in organisms lacking many biologically homologous landmarks. Several improvements to the method are described herein; these modifications are incorporated into the new computer programs HANGLE, HMATCH and HCURVE. First, automated tracing of outlines using image capture software, although desirable, results in high frequency pixel 'noise' that can corrupt the Fourier analysis. Program HANGLE eliminates this noise using optional and variable levels of outline smoothing. Secondly, a widely used Fourier technique, elliptic Fourier analysis (EFA, Kuhl and Giardina 1982), yields coefficients that are not computationally independent of each other, a condition that hampers and compromises statistical analysis. In addition, EFA increasingly downweights successively more detailed features of the outline. Program HANGLE solves both of these problems. Lastly, Fourier methods in general are sensitive to the placement of the starting position of the digitized trace. This problem is acute when the organisms under study have no unambiguously defined, homologous point on the outline from which to start the trace. Program HANGLE allows the user to normalize for starting position using various properties of individual outlines. Alternatively, HMATCH takes a new approach and can be used to normalize using properties of the entire population under study.

KEY WORDS: Fourier shape analysis, morphometric studies, new computer programs, foraminiferal outlines.

IN recent years palaeontologists have been making increasing use of morphometric methods in taxonomic and evolutionary research. Furthermore, the ready availability of powerful personal computers means that these morphometric methods are becoming increasingly diverse and sophisticated (for example, see reviews in Rohlf and Bookstein 1990; Temple 1992; Marcus *et al.* 1996). Fortunately, many computer programs to perform these analyses are widely accessible, together with published accounts of their theoretical underpinnings. Consequently, with relatively little effort, even weakly numerate palaeontologists are able to take advantage of, and benefit from, advances in the field of morphometrics. The potential benefits of morphometric analyses are well documented and great: they can yield independent, objective and repeatable tests of taxonomic, phylogenetic and evolutionary hypotheses that are based on qualitative studies; they can specify, localize and quantify morphological change through evolution or ontogeny; and they allow taxonomic separation and evolutionary rates to be measured. Perhaps most importantly, however, morphometric analyses simply force the palaeobiologist to view their material from a very different perspective, to step outside the confines of their subjective biases.

One widely used family of morphometric techniques employs the Fourier transform to examine populations of outline shapes and is the subject of this paper. Fourier shape analysis has been used in studies of diverse organisms such as miospores (Christopher and Waters 1974), leaves (White *et al.* 1988; McLellan and Endler 1998), foraminifera (many papers, e.g. Healy-Williams and Williams 1981; Healy-Williams 1983, 1984; Belyea and Thunell 1984), ostracodes (Burke *et al.* 1987; Foster and Kaesler 1988), bryozoans (Anstey and Delmet 1973), bivalves (Ferson *et al.* 1985; Crampton 1996; Mehlhop and Cifelli 1997; Innes and Bates 1999; Crampton and Maxwell in press), echinoderms (Waters 1977), trilobites (Foote 1989; Crônier *et al.* 1998), insect wings (Rohlf and Archie 1984), cambroclaves (Conway Morris *et al.* 1997), and vertebrate skeletal elements (O'Higgins and Williams 1987; Renaud *et al.* 1996).

Recently, the method has been extended to the analysis of the internal morphology of coccoliths (Garratt and Swan 1992, 1997).

The relative merits of different morphometric analyses of form have been discussed widely in the literature and these arguments are not repeated here (e.g. recent reviews in Rohlf and Bookstein 1990; Temple 1992; MacLeod 1999). Suffice it to say that Fourier outline shape analysis is, without doubt, an extremely powerful tool for the analysis of biological form in taxa lacking many homologous landmarks (e.g. Crampton 1995, 1996; Conway Morris *et al.* 1997; Mehlhop and Cifelli 1997). It is worth noting some recent and exciting developments towards the integration of landmark- and outline-based techniques, advances that hold great promise for a general synthesis of hitherto disparate methodologies (Bookstein and Green 1993; Bookstein 1995, 1996*a*, *b*; MacLeod 1999).

This paper addresses four problems with existing methods of Fourier shape analysis:

1. Fourier methods in general are sensitive to the use of unsmoothed outline data captured using modern image analysis software. Such data contain a large amount of spurious, high frequency 'noise' that can distort or corrupt the analysis.
2. One particular Fourier technique that has been applied in many recent studies is elliptic Fourier shape analysis (EFA, Kuhl and Giardina 1982). Although this approach is powerful and has a number of advantages over other methods (Ferson *et al.* 1985; Crampton 1995), it yields a relatively large number of Fourier coefficients that are not computationally independent of each other and are, in part, redundant. This redundancy hampers and compromises statistical analysis of the coefficients.
3. During statistical analysis, EFA assigns undue weight to gross features of the outline and increasingly downweights progressively finer elements of shape. In populations of outlines that vary in fundamental ways this may not be a problem, but in the case of shapes that differ in detail, discriminatory power may be lost.
4. Fourier methods in general are rather sensitive to placement of the starting position of the digitized trace, and thereby orientation of the trace, and this can unduly influence the interpretation of results. This problem is particularly significant when there is no unambiguously defined, biologically homologous point on the outline from which to start the trace. Previous studies have used properties intrinsic to each outline to normalize for starting position (Ferson *et al.* 1985). The method described herein takes a new approach and allows the user to normalize for starting position and orientation using either properties of individual outlines, or properties of the entire population under study.

These problems are discussed in this order through the paper. The Fourier method described here solves them and some other, more general, problems whilst retaining the strengths of earlier approaches. The software to perform these computations, programs HANGLE, HMATCH and HCURVE, are available with and described in a complementary publication (Crampton and Haines 1996)[1]. These programs are compiled for both Macintosh and IBM compatible computers. They require as input a population of digitized, two-dimensional outlines that can be captured using any standard digitizing or image analysis software. The programs output a set of Fourier coefficients suitable for multivariate statistical analysis.

THE BASIC METHOD

Fourier shape analysis takes an outline contour, described as a polygon of digitized *xy*-coordinates, and 'decomposes' this into a spectrum of harmonically related trigonometric (i.e. sine and cosine) curves. The Fourier coefficients that are produced, two per harmonic, describe the size ('amplitude') and angular offset relative to the starting position ('phase angle') of each harmonic curve. In this way, and using some appropriate number of harmonics, it is possible to describe even extremely complex shapes. The present method employs the 'Fast Fourier Transform' (FFT) to compute the harmonic spectrum (e.g. Davis 1986). The FFT does not work directly with the raw *xy*-coordinates, as in elliptic Fourier shape analysis (Kuhl and Giardina 1982), but operates on the tangent angle as a function of arc-length connecting the coordinates. During processing, outlines are standardized for size using perimeter length. The FFT is performed here

[1] This publication is available from the Publications Officer, Institute of Geological and Nuclear Sciences, PO Box 30368, Lower Hutt, New Zealand; fax + 64-4-570-4679; e-mail j.wright@gns.cri.nz

using the program HANGLE (see Crampton and Haines 1996); computational details are given in the Appendix.

Given a set of Fourier coefficients, it is possible to invert the Fourier transform, producing an outline contour from its independent coefficients (see Appendix). This property has great utility in morphometric studies and can be used, for example, to represent graphically particular morphologies and to generate a suite of synthetic outlines, either average or extreme (e.g. Crampton 1996; see also below). The inverse Fourier transform is performed here using the program HCURVE (see Crampton and Haines 1996). In general, only low order Fourier coefficients are needed to accurately reproduce an outline (Crampton 1995), and in this paper only coefficients up to the tenth harmonic are used.

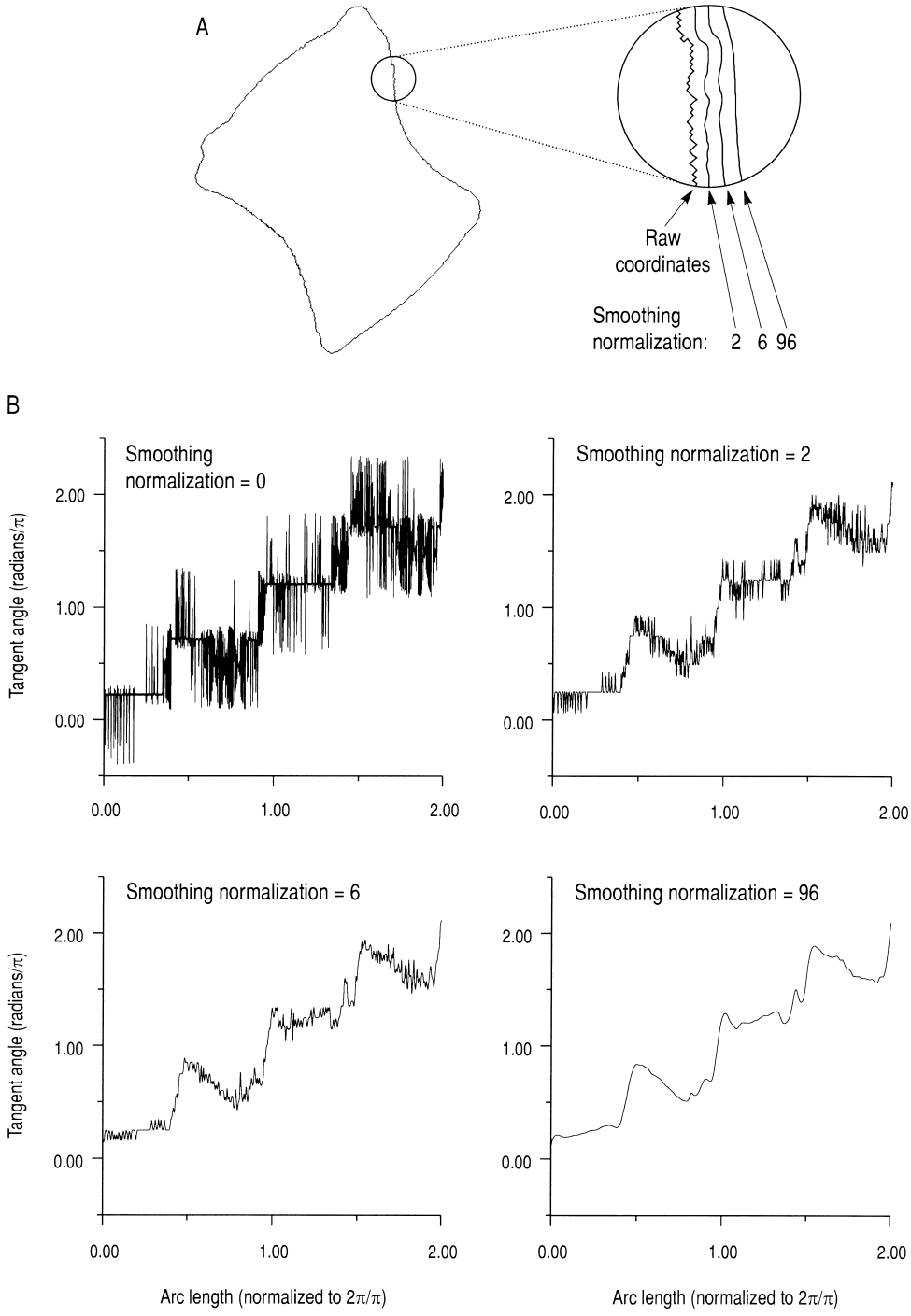
Methods described herein have been tested and evaluated using four rather different datasets: a subset of the Miocene foraminifera outlines of Scott (unpublished data); a subset of the Cretaceous bivalves studied by Crampton (1996); the Cambrian cambroclaves described by Conway Morris *et al.* (1997); and a suite of synthetic outlines describing a graded, 'evolutionary' continuum of morphologies. Results and conclusions from all these examples are essentially similar and, in the interests of brevity, the present discussion is illustrated primarily using just one dataset, the foraminiferal outlines of Scott. These data comprise two Tortonian (Upper Miocene) foraminifera, *Bolivinita compressa* Finlay and *B. pliobliqua* Vella, that are to be published electronically as part of a study of New Zealand *Bolivinita*. They were selected here because they are already well studied and define two *a priori* taxonomic groups, because they are representative of a large class of fossil organisms, and because these outlines are rather 'nondescript', lacking many clearly defined, unambiguous, biologically homologous landmarks. The outlines were traced automatically from electronically stored scanning electron micrograph images using image analysis software.

SMOOTHING NORMALIZATIONS

Increasingly, palaeontologists and biologists are making use of image analysis software to automatically trace and digitize outlines for use in morphometric studies (e.g. Ferson *et al.* 1985). Automated outline capture is rapid, potentially highly accurate, and desirable. A digitized outline is not, however, a continuous, smooth curve, but proceeds as a series of discrete, straight-line, pixel 'steps' parallel to the x and y-axes of the digitizing grid (Text-fig. 1A). In the case of an automatically and finely digitized outline this results in a considerable amount of high frequency 'noise' (Text-fig. 1B) that can distort or corrupt the Fourier analysis (see below). Hence, if an outline has been finely digitized it is necessary to smooth it prior to computation of the FFT. Smoothing is possible within many image analysis programs or, alternatively, within program HANGLE.

Smoothing is accomplished here by taking a weighted moving average over three successive coordinate points: $(x,\ y)_i^{new} = \frac{1}{4}(x,\ y)_{i-1}^{old} + \frac{1}{2}(x,\ y)_i^{old} + \frac{1}{4}(x,\ y)_{i+1}^{old}$. Varying degrees of smoothing can be achieved by subjecting an outline to several smoothing iterations (Text-fig. 1). The position of any particular point through many successive smoothing iterations has a Gaussian-like (i.e. normal) weighting in the dependence on the positions of the original points, with the standard deviation $\delta = \sqrt{\frac{N}{2}}$, where N is the number of smoothing iterations. It is desirable to have $\frac{Nsamp}{NFFT} \leq \delta$, where *Nsamp* is the original number of sampled xy-coordinates and *NFFT* is the number of re-sampled points that are used in the FFT (see below). This helps to ensure that the adjusted outline is smooth between the re-sampled points, which are evenly spaced with respect to arc-length around the outline. Hence, if *NFFT* is known, it is possible to calculate the *absolutely minimum* satisfactory number of smoothing iterations for any particular outline. In the case of program HANGLE, *NFFT* has been set to be equal to 1024. If outline smoothing is selected by the operator, then the program automatically calculates and performs the minimum desirable number of smoothing iterations whenever this is more than the number of iterations specified (Crampton and Haines 1996).

There are no hard-and-fast rules regarding the level of smoothing to be used. The degree to which any outline should be smoothed will vary from study to study and will also depend upon the level of image magnification relative to the digitization grid. If high frequency detail is likely to have some taxonomic significance, then lower levels of smoothing may be appropriate. Alternatively, in the foraminiferan example of Text-figure 1A, relatively high levels of smoothing result in outlines closest to the true


A
Raw
coordinates
Smoothing
normalization: 2 6 96
B
Smoothing
normalization = 0
Smoothing normalization = 2
Smoothing normalization = 6
Smoothing normalization = 96
Tangent angle (radians/π)
Arc length (normalized to 2π/π)
0.00
1.00
2.00

specimen shape (i.e. > 6 smoothing iterations). This can only be judged by comparison of actual specimens or high-quality images with their digitized outlines.

As noted above, a large amount of high frequency pixel noise, if not eliminated by smoothing, can 'corrupt' any subsequent Fourier decomposition and statistical analysis. This is illustrated in Text-figure 2, in which the raw coordinates and four reconstructed contours of a single outline are shown superimposed. The reconstructed outlines were generated by initial processing in program HANGLE, using different degrees of smoothing, and subsequent inversion of the Fourier transform using program HCURVE and the coefficients for the first 10 harmonics. It is clear that the reconstructed shape of the unsmoothed outline (i.e. smoothing normalization 0) has been distorted in comparison to the raw coordinates and the smoothed outlines. This distortion is a result of increased arc-length in regions of high frequency pixel noise, the magnitude of which is itself determined by the orientation of any particular margin segment with respect to the digitization grid. Since the FFT operates on the tangent angle as a function of arc-length, any increase in arc-length that is an artefact of digitization will have a spurious effect upon the FFT by, in effect, transferring information from low order to high order harmonics. When only low order harmonics are used during inversion of the FFT, this results in distortion of the reconstructed outline. This problem will affect most methods of Fourier analysis, although one commonly used method, elliptic Fourier analysis, is relatively insensitive to pixel noise for reasons outlined below.

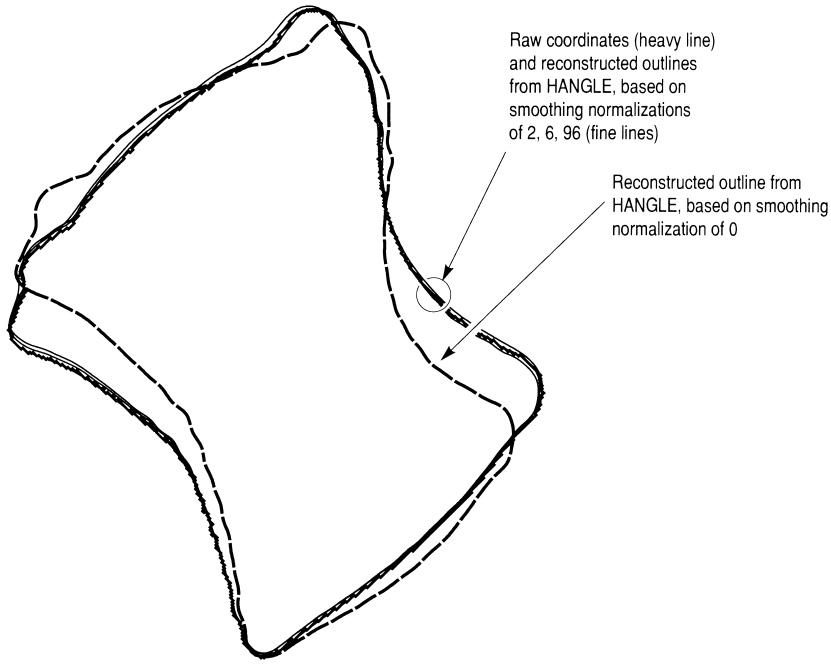
The effects of pixel noise upon statistical interpretation of Fourier coefficients is illustrated in Text-figure 3. This diagram shows four principal component analyses (PCAs) of the same dataset, comprising two *a priori* groups of foraminifera, and the effects of different degrees of smoothing. All plots separate the two *a priori* groups, but the separation is greatly diminished in the data that have not been smoothed. Similarly, there is some loss of discrimination in data that have been highly smoothed (96 iterations), with contraction of the first principal component.

COMMENTS ON SOME EARLIER FOURIER TECHNIQUES

Zahn and Roskies (1972) described a Fourier technique similar to the method used here, although there are a number of computational differences. Their technique has two major disadvantages. First, it does not ensure that reconstructed outlines are closed, making meaningful comparisons between outlines impossible. In program HANGLE, on the other hand, closure is determined by the dependence of the first harmonic on all others (explained in Appendix). Secondly, the Zahn and Roskies method defines a relatively large number of 'phase angle Fourier descriptors' that are not computationally independent of each other and are, in part, redundant. For example, in the case of ten harmonics, they define 45 phase angle descriptors. Subsequent statistical analyses of such data will be compromised by correlations that result from this redundancy and are spurious with respect outline shape *per se*. In addition, this problem hampers statistical analysis when dealing with small populations, as is commonly the case in palaeontological studies, because the number of variables (i.e. Fourier coefficients) must be less than the number of specimens, and should be substantially less. In contrast to the method of Zahn and Roskies, program HANGLE produces two coefficients per harmonic that describe amplitude and phase angle; these coefficients are computationally independent (see Appendix).

Another Fourier technique, Elliptic Fourier analysis (EFA, Kuhl and Giardina 1982), has been applied

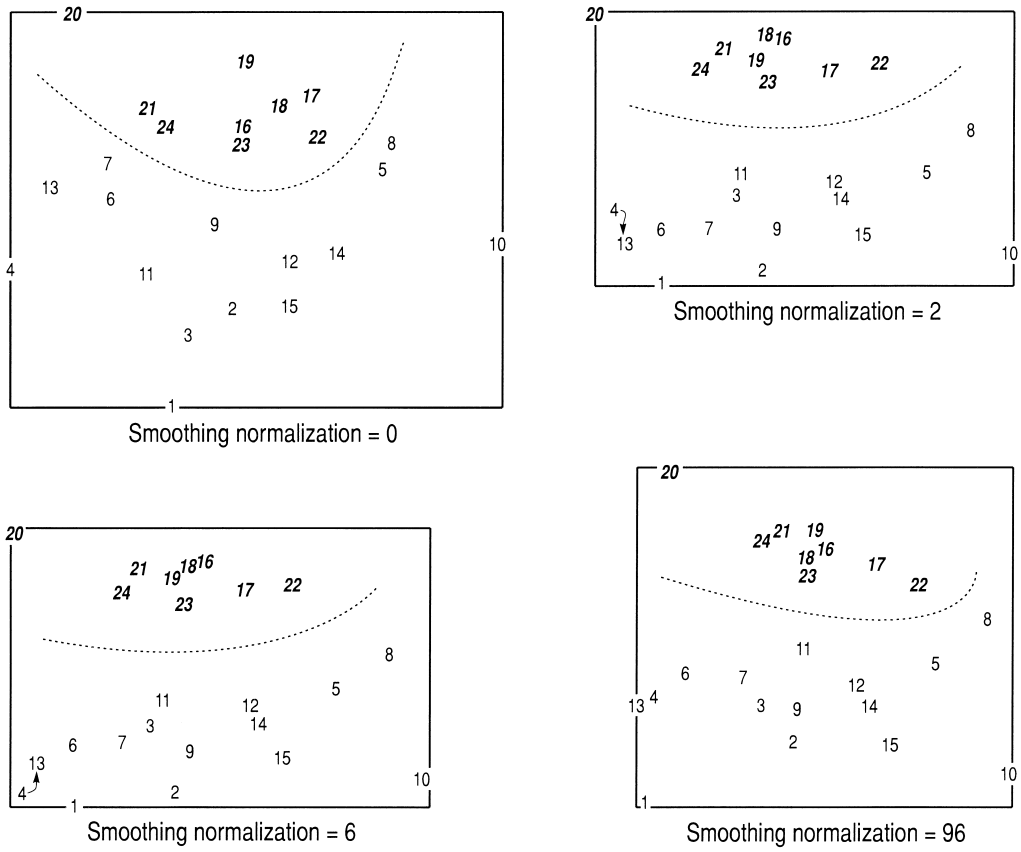
TEXT-FIG. 1. A, representative digitized outline of the foraminifer *Bolivinta pliobliqua*, spiral view (from Scott, unpublished data). This outline has been traced automatically from an electronically stored scanning electron micrograph image using image analysis software. Enlarged region shows detail of the outline as represented by the raw *xy*-coordinates, and following smoothing normalizations of 2, 6 and 96 in program HANGLE (note, for clarity each line has been offset, see also Text-fig. 2). The value of the smoothing normalization in HANGLE corresponds to the minimum number of iterations through the smoothing routine in the case of smoothing normalization 2 and increasing degrees of line smoothing in the other cases (see text for further discussion). B, the same data shown as plots of tangent angle versus arc-length around the outline contour. It is this function, tangent angle versus arc-length, that the FFT operates upon. The pixel 'noise' of the raw *xy*-coordinates is very clearly visible in the first plot (smoothing normalization 0) and is progressively removed through successive smoothing iterations.



TEXT-FIG. 2. The raw coordinates and four reconstructed outline contours of a single foraminifer, *Bolivinita pliobliqua* (the same individual as that shown in Text-fig. 1). The reconstructed outlines were generated by initial processing in program HANGLE, using different degrees of smoothing, and subsequent inversion of the Fourier transform using program HCURVE and the first 10 harmonics. Note that the raw coordinates and three of the reconstructed outlines are largely overlapping (segments of the corresponding original outlines, separated for clarity, are shown in Text-fig. 1A). The reconstructed outline that has not been smoothed (smoothing normalization 0) is significantly distorted because of local variations in arc-length that are artefacts of digitization. These variations are the pixel 'noise' of Text-figure 1B. See text for further explanation.

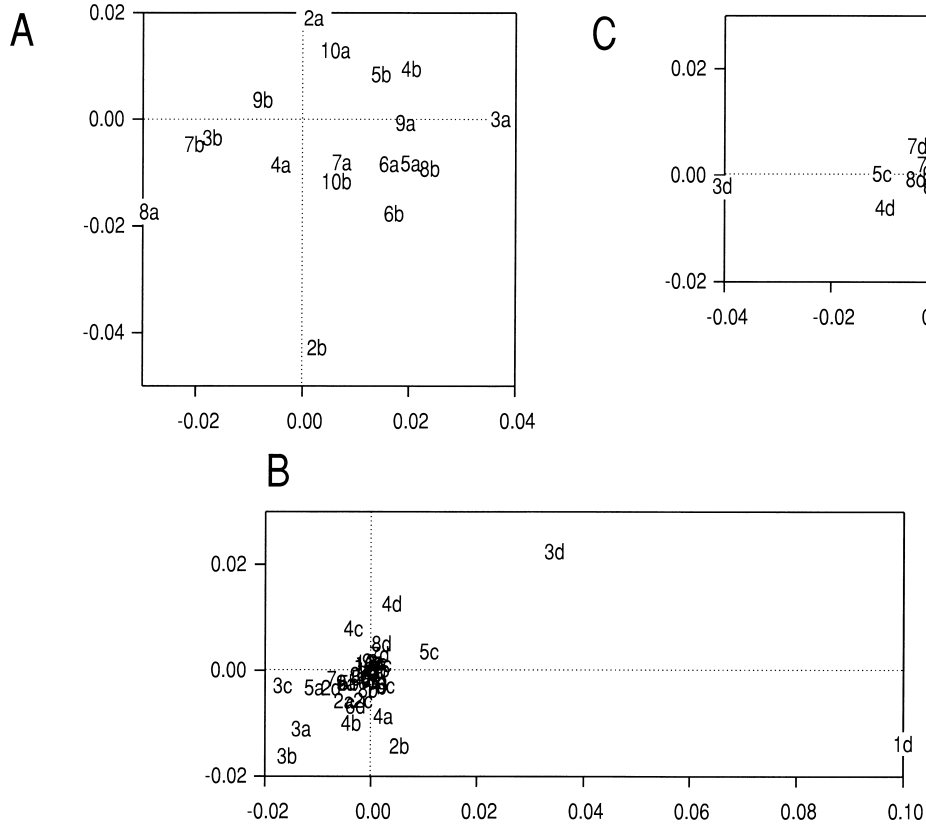
widely in morphometry and is a powerful technique (e.g. Ferson *et al.* 1985; Crampton 1996; Mehlhop and Cifelli 1997; McLellan and Endler 1998). In particular, it allows the user to normalize for size, location, starting position and rotation of the outline trace (Kuhl and Giardina 1982; Ferson *et al.* 1985). As noted in the introduction, however, the method also has two disadvantages. The first of these is analogous to that described above: EFA yields four Fourier coefficients per harmonic that are not computationally independent of each other, resulting in spurious correlations that will compromise statistical analysis. This is explained arithmetically in the Appendix.

The second problem with EFA results from the relative downweighting of all harmonics above the first. [Note that numbering conventions for harmonics vary depending on how the Fourier coefficients are derived: the 'second harmonic' as used in this paper corresponds primarily to the 'first harmonic' of EFA as described by Kuhl and Giardina (1982) and Ferson *et al.* (1985)]. This is because EFA effectively divides the kth harmonic by k-1, i.e. HANGLE's second harmonic is divided by 1, the third harmonic is divided by 2, the fourth harmonic is divided by 3, and so on (explained arithmetically in the Appendix). This unduly reduces the influence of higher order harmonics in any statistical analysis based on the variance-covariance matrix. Successive harmonics describe progressively finer details of the outline. Hence, EFA tends to diminish the discriminatory power of outline details, even though such details may have taxonomic or other biological significance. The problem affects even relatively 'large' outline elements such as 'quadrate' features (i.e. HANGLE's fourth harmonic). This is illustrated in Text-figure 4 showing PCAs of the Scott dataset based on Fourier coefficients from HANGLE and from EFA. The plots ordinate variables (i.e. Fourier coefficients) against the first two principal component axes and identify



TEXT-FIG. 3. Four principal component analyses of a dataset comprising 24 specimens of the foraminifera *Bolivinita compressa* (specimens 1–15, normal text) and *B. pliobliqua* (specimens 16–24, bold italic text). The dashed line separates the two *a priori* groups (classification and data from Scott, unpublished data). In each plot the *x*-axis corresponds to the first principal component and the *y*-axis corresponds to the second principal component. Principal component values are shown using the same isotropic scale on all plots. In all four analyses, outlines were processed using starting normalization 0 followed by matching and principal components were computed using the variance-covariance matrix. Whereas all plots separate the two groups of specimens, this separation is greatly diminished in the data that have not been smoothed (smoothing normalization 0). With varying numbers of smoothing iterations the distribution of specimens in morphospace is stable, although there is contraction of the first principal component and a corresponding loss of resolution following high numbers of smoothings (96).

those variables that contribute most discriminatory power to each axis. In the first plot, based on HANGLE, all harmonics contribute substantially to both principal components. This indicates that even the high order harmonics have some discriminatory power and, potentially, some taxonomic utility. In the second plot based on EFA, however, the first principal component is dominated by the first harmonic and only the second to fifth harmonics contribute any significant variance to either axis. The remaining variables cluster around the origin and have little influence on the analysis. The same is true even when the first harmonic is eliminated from the PCA, as shown in Text-figure 4C: again the analysis is dominated by the same few low order harmonics. This demonstrates that the statistical analysis based on EFA downweights or effectively ignores medium to high frequency information from the outlines, information that does have discriminatory power. It is worth noting that the downweighting of higher order harmonics in EFA makes it relatively insensitive to the pixel noise discussed in the previous section.



TEXT-FIG. 4. The first two principal component axes based on the Scott dataset (see caption to Text-fig. 3) and 10 harmonics. The plots ordinate the variables (i.e. the Fourier coefficients) against the principal component axes and identify those variables that contribute most discriminatory power to each axis. Each variable is numbered according to its harmonic. All analyses utilise the variance-covariance matrix. A, analysis based on HANGLE, with smoothing normalization 2 and starting normalization 0 followed by matching. There are two coefficients per harmonic, denoted a and b. Note that the first harmonic has no significance here and is not output by HANGLE (see Appendix). All harmonics contribute substantially to the analysis and have some discriminatory power. B, analysis based on EFA, normalized for location, size, starting position and rotation. There are four coefficients per harmonic, denoted a–d. Note that normalization results in loss of coefficients 1a–1c. The first principal component is dominated by the first harmonic (variable 1d) and only the second to fifth harmonics contribute significantly, though comparatively weakly, to either axis (variables 2b, 3a–c, 4c–d, 5c). C, analysis as for B, with the first harmonic eliminated. Again, the analysis is dominated by low order harmonics, compared to the analysis based on coefficients from HANGLE. Differences between the analyses based on HANGLE and EFA are a consequence of the progressive downweighting of harmonics relative to the first in EFA. See text for further discussion.

STARTING NORMALIZATIONS

As noted in the introduction, the starting position of a digitized trace, and thereby its orientation, has a marked influence on the Fourier transform. Small variations in starting position can unduly affect interpretation of Fourier coefficients. Although it is desirable to digitize outlines in a standard manner, beginning at a homologous landmark and using a consistent orientation, in many studies this is difficult to achieve with adequate precision. For example, in a study of Mesozoic bivalves, Crampton (1995, 1996) found that the only available landmark, the umbo, is too broadly rounded to allow precise identification of

a homologous landmark. Similarly, in a study of enigmatic Cambrian cambroclaves, Conway Morris *et al.* (1997) were unable to exactly identify homologous orientations or starting positions for the traces.

Previous studies have addressed this problem by using features intrinsic to each outline to normalize for starting position and orientation. In particular, Kuhl and Giardina (1982) and Ferson *et al.* (1985) described normalizations that are based on alignment of the second harmonic, or best-fitting ellipse. Although this approach is appropriate in some instances, in other cases it does not result in close alignment of outlines or starting positions (Text-figure 5).

In the present study, normalization for starting position and orientation of outlines is approached in two ways (further details are given in the Appendix):

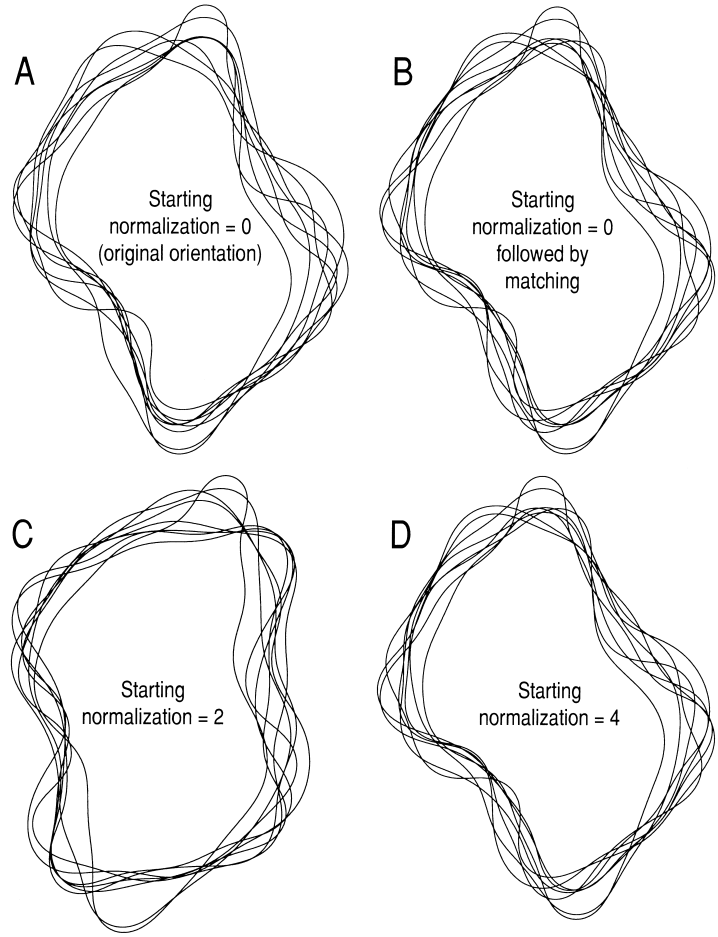
1. Program HANGLE will normalize using properties of designated harmonics of individual outlines (Crampton and Haines 1996). For example, normalization based on the second harmonic will align outlines using their best-fitting ellipses, as described above; normalization based on the third harmonic will align using 'triangular' features of the outlines; normalization based on the fourth harmonic will align using 'quadrate' features of the outlines; and so on. For these normalizations to work, it is important that the starting positions of traces, as digitized, correspond approximately to some biological landmark.
2. Alternatively, it is possible to normalize for starting position and orientation using properties of the entire population of outlines under study, here termed 'matching'. This is achieved here using program HMATCH (see Crampton and Haines 1996). This program adjusts the starting positions of the entire population of outlines under study such that reconstructed outlines are as closely aligned as possible in terms of their tangent angle curves (i.e. the curves shown in Text-fig. 1B).

The choice of which starting normalization to use will vary from study to study, depending on the presence/absence of a well-constrained landmark in all specimens, the morphological disparity of the dataset, the general form of the outlines, and the presence/absence of well-defined and consistent axes of symmetry in all outlines. In most studies some experimentation may be necessary to decide upon the most satisfactory approach to normalization.

The effects of different starting normalizations are illustrated in Text-figures 5–7. In Text-figure 5, eight randomly chosen foraminifera from the Scott dataset are shown in their original orientation and re-aligned using different starting normalizations. During image capture, these specimens were positioned with the outer wall of the penultimate chamber parallel to the base of the scanning electron microscope display. If the wall was convex, a tangent through the turning point was made parallel with the reference. Left-coiled specimens were mirrored. During digitization, a starting point for the trace was selected at the left end of the wall, a location of sharp curvature and commonly the site of a peripheral keel (G. H. Scott, Institute of Geological and Nuclear Sciences, pers. comm. 1998). Following subsequent processing in HANGLE and HMATCH, visual examination suggests that alignment using the best-fitting ellipse (starting normalization 2) is relatively poor, whereas alignments by matching and by using the fourth harmonic (starting normalization 4) are relatively tightly constrained, similar, and satisfactory. Alternatively, an objective measure of the 'goodness' of alignment is given by the sum of eigenvalues derived from a PCA: eigenvalues will be minimised for the best-aligned outlines. The sums of eigenvalues for each alignment are reported in the caption to Text-figure 5 and confirm that alignments by matching and using the fourth harmonic are indeed the most satisfactory. Use of the fourth harmonic succeeds here because of the essentially quadrate shape of these foraminifera. PCAs corresponding to these four alignments are shown in Text-fig. 7.

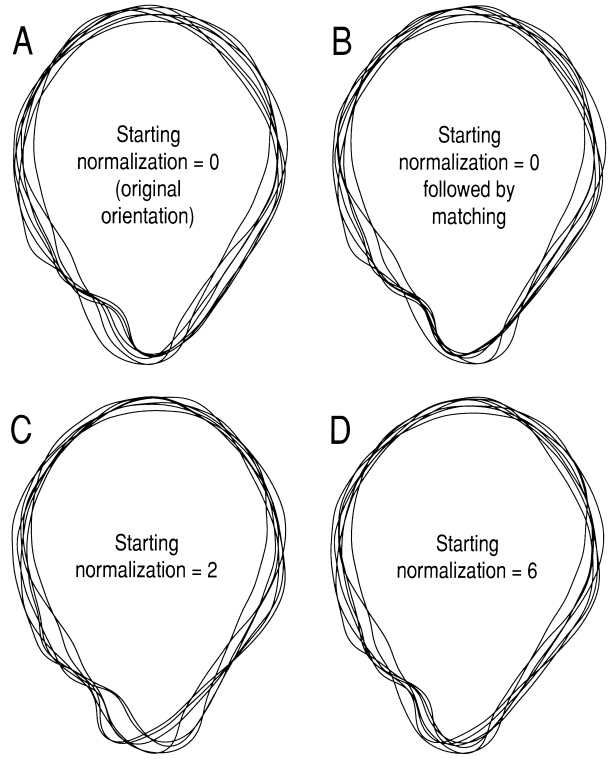
Trials using other datasets have indicated that matching is a consistently robust approach to normalization for starting position. This is illustrated in Text-figure 6 using a subset of the outlines of Cretaceous bivalves described in Crampton (1996). During digitization, these outlines were orientated with their hingelines parallel and traces were initiated at the umbones. As with the Scott example given above, starting normalization based on the best-fitting ellipse (starting normalization 2) results in a relatively poor alignment of outlines, whereas normalization by matching yields a comparatively good alignment. Because many of these bivalves display a very crudely six-faceted shape, normalization using the sixth harmonic (starting normalization 6) also results in a reasonable alignment of outlines.

PCAs based on the Scott data demonstrate the impact of starting position normalizations (Text-fig. 7). The analysis of outlines that have not been re-aligned (starting normalization 0) displays a high degree of



TEXT-FIG. 5. Outlines of eight randomly chosen foraminifera from the Scott dataset (see text), showing effects of different starting normalizations. During digitization these eight specimens were oriented by eye with their bottom right-hand margins aligned, and the trace was initiated at the bottom corner. The specimens were then aligned using program HMATCH (matching) and starting normalizations 2 and 4 in program HANGLE. 'Goodness' of alignment can be gauged visually or by summing the eigenvalues derived from PCA, the lowest eigenvalues corresponding to the best alignments. The sums of eigenvalues, based on PCA for the full dataset and each of the given alignments, are: A, $2{\cdot}64\times10^{-2}$; B, $2{\cdot}13\times10^{-2}$; C, 3.62×10^{-2}; D, 2.18×10^{-2}. Starting normalization 2, which aligns on the best-fitting ellipse, results in a conspicuously poor match between specimens. Alignments using matching and starting normalization 4 are relatively tightly constrained and yield similar, satisfactory results. Starting normalization 4 employs quadrate features of the outline and succeeds here because of the basically quadrate shapes of these foraminifera.

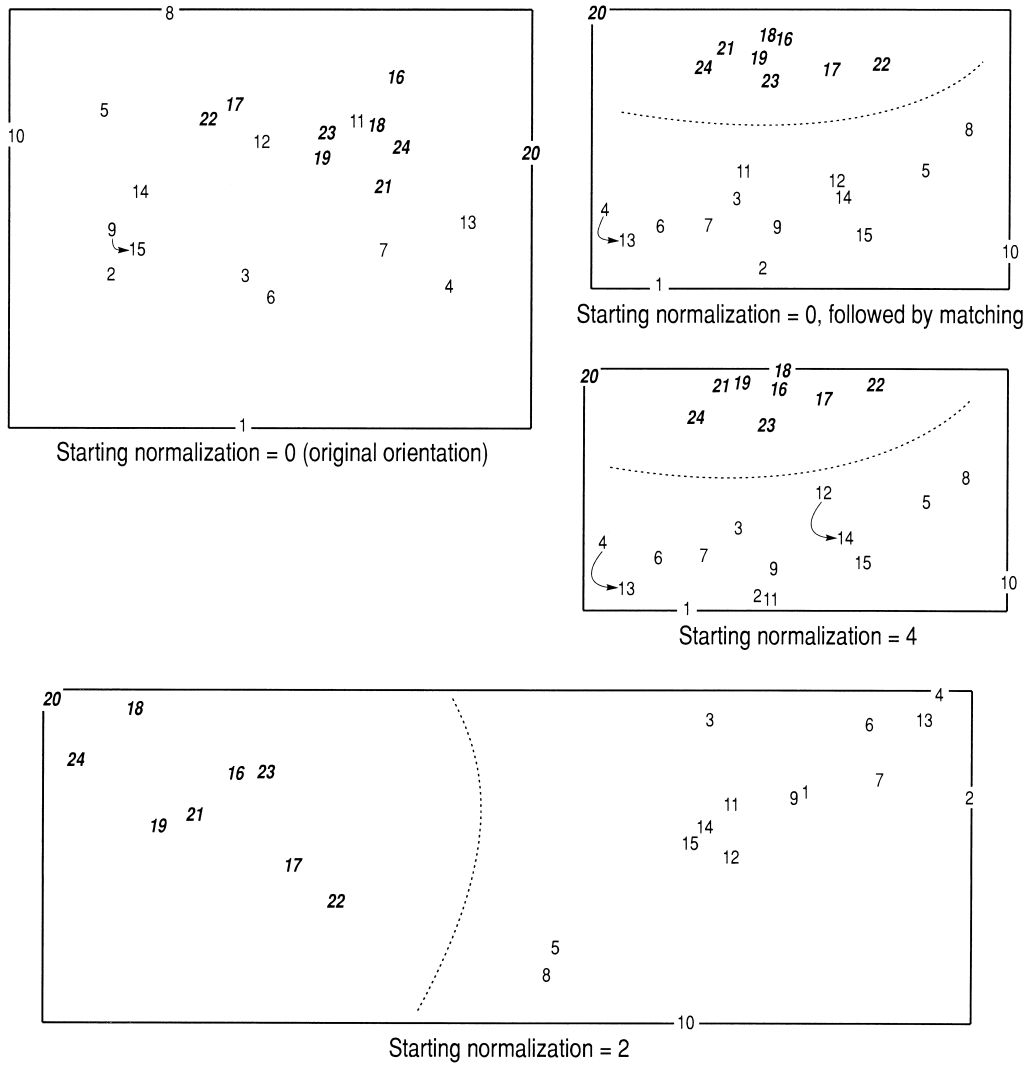
overlap between the two groups and could be misleading in any taxonomic or evolutionary study. PCAs based on re-aligned specimens, on the other hand, all separate clearly the *a priori* groups. The two analyses based on matching and on starting normalization 4 are very similar; the analysis based on starting normalization 2 relates to these as a reflection about the line $y = -x$ (i.e. the two principal component axes are interchanged), but otherwise indicates similar relationships between specimens. The reason for this difference is that many of the outlines have relatively small second harmonics; consequently, normalization based on the second is a poor choice. (In this case, however, the relatively small second harmonic



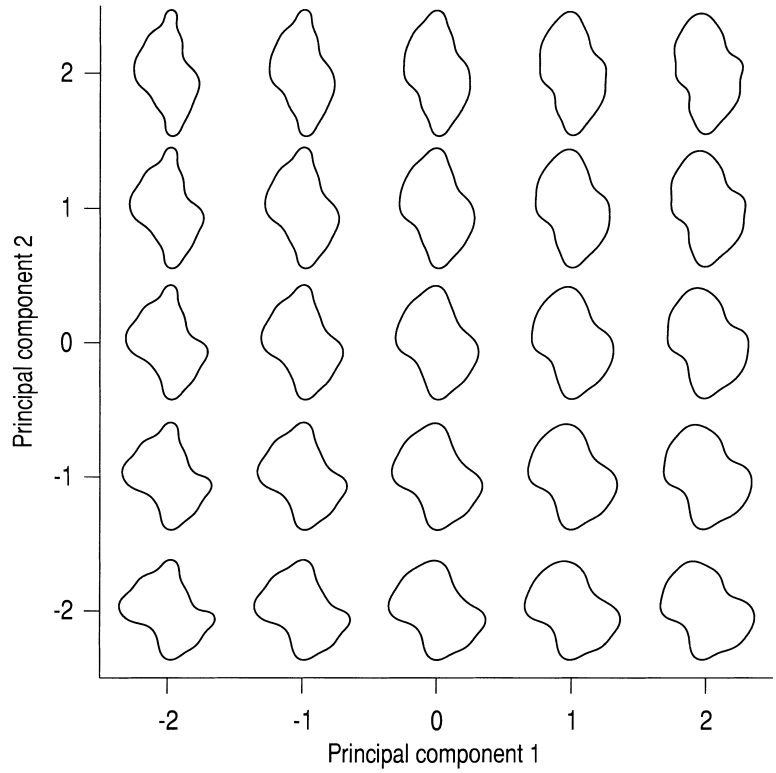
TEXT-FIG. 6. Outlines of seven randomly chosen bivalves from the dataset of Crampton (1996), showing the effects of different starting normalizations. During digitization these specimens were oriented by eye with their hingelines parallel and the traces were initiated at the umbones. As with Text-figure 5, starting normalization based on the best-fitting ellipse (starting normalization 2) results in a relatively poor alignment of outlines, whereas normalization by matching yields a comparatively good alignment. Because many of these bivalves display a very crudely six-faceted shape, normalization using the sixth harmonic (starting normalization 6) also yields a reasonable alignment of outlines. Sums of eigenvalues, based on the full dataset and each of the given alignments, are: A, $7{\cdot}37\times10^{-3}$; B, $6{\cdot}55\times10^{-3}$; C, $1{\cdot}34\times10^{-2}$; D, $7{\cdot}93\times10^{-3}$ (see explanation in text and in caption to Text-figure 5).

appears to correlate well with other differences between the two populations, therefore a satisfactory separation is obtained). Consideration of Text-figure 5, discussed above, also suggests that analyses based on matching or on starting normalization 4 are to be preferred. Irrespective of this, the general correspondence between the three PCAs based on re-aligned specimens implies that the groupings and morphospace distributions are 'real' and robust.

A very useful aid to the interpretation of PCAs is illustrated in Text-figure 8. Given any set of principal component axes, it is possible to generate a series of wholly synthetic, model shapes corresponding to the mean shape and to arbitrarily chosen positions along each axis, measured in units of standard deviations. Some of these shapes will approximate real outlines, others will represent fictional or extreme morphologies that do not occur in the natural population. All of the shapes represent 'pure' shapes for each specified point in the given morphospace, free from any other sources of variation or random 'noise'. Clearly, such a plot can be used to interpret the multivariate morphospace defined by the principal component axes, and can assist in the identification of trends and patterns in the data. In this way, morphometric data are presented within a geometric framework that can be related to previous qualitative studies, an important requirement of any morphometric study (MacLeod 1999). The synthetic shapes were generated in the following way. First, Fourier coefficients of the 'mean shape', positioned at the centre of



TEXT-FIG. 7. Four principal component analyses of 24 foraminifera from the Scott dataset (see caption to Text-fig. 3), illustrating the effects of different starting normalizations. The dashed line separates two *a priori* groups (classification from Scott, unpublished data). In each plot the x-axis corresponds to the first principal component and the y-axis corresponds to the second principal component. Principal component values are shown using the same isotropic scale on all plots. All four analyses utilise smoothing normalization 2 and were computed using the variance-covariance matrix. The analysis based on the original orientation of outlines (starting normalization 0) is poor, with a high degree of overlap between the two groups and comparatively few of the differences between specimens resolved in the first two principal components. Analyses based on re-aligned specimens, on the other hand, all yield reasonable results with clear separations between groups. As predicted by Text-figure 5, the two analyses based on matching and on starting normalization 4 are very similar. The analysis based on starting normalization 2 relates to these as a reflection about the line $y = -x$ (i.e. with the first two principal components transposed) and in other respects suggests a similar distribution of specimens in morphospace. The very large variance of the first principal component for starting normalization 2 is a consequence of the mismatch of the outlines illustrated in Text-figure 4.



TEXT-FIG. 8. The first two principal component axes based the Scott dataset (see caption to Text-fig. 3), using smoothing normalization 2 and starting normalization 0 followed by matching (compare with Text-fig. 7, upper right hand plot). Ordinated against these axes are a series of wholly synthetic, model outline shapes corresponding to the positions of unit standard deviation steps along each of the two axes. These shapes facilitate easy interpretation of morphospace distributions of the real data. Generation of the synthetic shapes is explained in the text.

the plot, were derived by averaging the coefficients from all sample outlines. To generate a suite of closely spaced *xy*-coordinates around the mean shape, these averaged Fourier coefficients were simply processed using the inverse Fourier transform program HCURVE. Fourier coefficients for the remaining synthetic shapes follow from a useful property of the eigenvectors determined from PCA. These eigenvectors, for each of the principal component axes, correspond to units of standard deviation; principal component loadings on each of the axes correspond effectively to Fourier coefficients in principal component space. The Fourier coefficients of the synthetic shapes, therefore, were derived by the appropriate simple vector additions and subtractions of the eigenvectors to/from the Fourier coefficients of the mean shape. For example, coefficients for the shape at position $(2, -1)$ are equal to the vector sum: (mean coefficients) + (2 × first eigenvector) − (second eigenvector). Each of the resulting sets of Fourier coefficients was inverted using program HCURVE to create the synthetic outlines shown.

SUMMARY

1. Fourier shape analysis of digitized outlines is a powerful tool for the morphometric study of organisms lacking many homologous landmarks.

2. Automated digitization of specimen outlines, using image analysis software, is potentially rapid, accurate and expedient. Outlines captured in this way, however, typically include spurious high frequency

'noise' that can corrupt subsequent Fourier analysis and compromise statistical interpretation. Program HANGLE solves this problem by optionally smoothing an outline to the minimum appropriate level using a weighted moving average of the original coordinate points. In addition, it allows the user to specify greater levels of smoothing if so desired.

3. Some other Fourier techniques yield relatively high numbers of Fourier coefficients that are not computationally independent of each other. This condition hampers and compromises statistical analysis. In addition, one widely-used technique, elliptic Fourier analysis (EFA, Kuhl and Giardina 1982) increasingly downweights successively higher order harmonics. This reduces or suppresses altogether the discriminatory power of outline details even though such details may have some biological significance. Program HANGLE solves these problems and produces two, computationally independent coefficients per harmonic. The program employs the 'Fast Fourier Transform' and operates upon the tangent angle as a function of arc-length around a spline curve fitted to the (optionally) smoothed outline.

4. Fourier methods in general are sensitive to placement of the starting position of the digitized trace and this can severely compromise the statistical interpretation of data. This problem is particularly acute when the organisms under study have no unambiguously defined, homologous point on the outline from which to start the trace. Previous studies have used properties intrinsic to each outline to normalize for starting position. This approach works when the outlines under study have a well-developed, consistent and biologically homologous shape component that is described by one of the Fourier harmonics. Program HANGLE allows the user to normalize in this way, using any harmonic shape component. Program HMATCH takes a different approach and normalizes using properties of the entire population under study, rotating each one to minimise differences between their tangent angle curves. Tests with a variety of data suggest that 'matching' is consistently a robust approach to normalization for starting position.

5. Program HCURVE inverts the Fourier transform, producing a tangent angle curve and an outline for any set of Fourier coefficients. This can be used to create synthetic outline shapes that are a great aid in the interpretation of multivariate statistical data (e.g. Text-fig. 8).

Acknowledgements. This is Institute of Geological and Nuclear Sciences contribution 1411. Earlier drafts of this paper were greatly improved by the thorough and thoughtful reviews of George Scott and Ian Raine, both of the Institute of Geological and Nuclear Sciences, and an anonymous referee. We acknowledge the use of unpublished data from George Scott.

REFERENCES

ANSTEY, R. L. and DELMET, D. A. 1973. Fourier analysis of zooecial shapes in fossil tubular bryozoans. *Bulletin of the Geological Society of America*, **84**, 1753–1764.

BELYEA, P. R. and THUNELL, R. C. 1984. Fourier shape analysis and planktonic foraminiferal evolution: the *Neogloboquadrina-Pulleniatina* lineages. *Journal of Paleontology*, **58**, 1026–1040.

BOOKSTEIN, F. L. 1995. The morphometric synthesis for landmarks and edge-elements in images. *Terra Nova*, **7**, 393–407.

—— 1996*a*. Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, **58**, 313–365.

—— 1996*b*. Landmark methods for forms without landmarks: localizing group differences in outline shape. 279–289. *In* AMINI, A., BOOKSTEIN, F. L. and WILSON, D. (eds). *Proceedings of the workshop on mathematical methods in biomedical image analysis*. IEEE Computer Society, San Francisco, 340 pp.

—— and GREEN, W. D. K. 1993. A feature space for edgels in images with landmarks. *Journal of Mathematical Imaging and Vision*, **3**, 213–261.

BURKE, C. D., FULL, W. E. and GERNANT, R. E. 1987. Recognition of fossil freshwater ostracodes: Fourier shape analysis. *Lethaia*, **20**, 307–314.

CHRISTOPHER, R. A. and WATERS, J. A. 1974. Fourier series as a quantitative descriptor of miospore shape. *Journal of Paleontology*, **48**, 697–709.

CONWAY MORRIS, S., CRAMPTON, J. S. and CHAPMAN, A. 1997. Lower Cambrian cambroclaves (*incertae sedis*) from Xinjiang, China, with comments on the morphological variability of sclerites. *Palaeontology*, **40**, 167–189.

CRAMPTON, J. S. 1995. Elliptic Fourier shape analysis of fossil bivalves: some practical considerations. *Lethaia*, **28**, 179–186.

—— 1996. Biometric analysis, systematics and evolution of Albian *Actinoceramus* (Cretaceous Bivalvia, Inoceramidae). *Monograph of the Institute of Geological and Nuclear Sciences*, **15**, 1–80.

—— and HAINES, A. J. 1996. Users' manual for programs HANGLE, HMATCH, and HCURVE for the Fourier shape analysis of two-dimensional outlines. *Science Report of the Institute of Geological and Nuclear Sciences*, **96/37**, 1–28.

—— and MAXWELL, P. A. in press. Size: all it's shaped up to be? Evolution of shape through the lifespan of the Cenozoic bivalve *Spissatella* (Crassatellidae). *In* HARPER, E. M., TAYLOR, J. D. and CRAME, J. A. (eds). Evolutionary biology of the Bivalvia. *Geological Society, London, Special Publication.*

CRONIER, C., RENAUD, S., FEIST, R. and AUFFRAY, J.-C. 1998. Ontogeny of *Trimerocephalus lelievrei* (Trilobita, Phacopida), a representative of the Late Devonian phacopine paedomorphocline: a morphometric approach. *Paleobiology*, **24**, 359–370.

DAVIS, J. C. 1986. *Statistics and data analysis in geology*. Second edition. Wiley and Sons, New York, 646 pp.

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

FOOTE, M. 1989. Perimeter-based Fourier analysis: a new morphometric method applied to the trilobite cranidium. *Journal of Paleontology*, **63**, 880–885.

FOSTER, D. W. and KAESLER, R. L. 1988. Shape analysis. Ideas from the Ostracoda. 53–69. *In* MCKINNEY, M. L. (ed.). *Heterochrony in evolution: a multidisciplinary approach*. First edition. Topics in Geobiology, 7, Plenum Press, New York, 348 pp.

GARRATT, J. A. and SWAN, A. R. H. 1992. Morphological data from coccolith images. *In* HAMRSMID, B. and YOUNG, J. R. (eds). Nannoplankton research, **1**. *Knihovnicka ZPZ*, **14a**, 11–34.

—— —— 1997. Fourier transforms of image-derived data: application to Albian coccoliths. *Micropaleontology*, **43**, 303–317.

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

—— 1984. Quatitative image analysis: application to planktonic foraminiferal palaeoecology and evolution. *Geobios, Mémoire Spéciale*, **8**, 425–432.

—— and WILLIAMS, D. F. 1981. Fourier analysis of test shape of planktonic foraminifera. *Nature*, **289**, 485–487.

INNES, D. J. and BATES, J. A. 1999. Morphological variation of *Mytilus edulis* and *Mytilus trossulus* in eastern Newfoundland. *Marine Biology*, **133**, 691–699.

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

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

MARCUS, L. F., CORTI, M., LOY, A., NAYLOR, G. J. P. and SLICE, D. E. 1996. *Advances in morphometrics*. Plenum, New York, 587 pp.

MCLELLAN, T. and ENDLER, J. A. 1998. The relative success of some methods for measuring and describing the shape of complex objects. *Systematic Biology*, **47**, 264–281.

MEHLHOP, P. and CIFELLI, R. L. 1997. A comparison of morphometric techniques to distinguish sympatric mussel species (family Unionidae) with similar shell morphology. 249–258. *In* YATES, T. L., GANNON, W. L. and WILSON, D. E. (eds). *Life among the muses: papers in honor of James S. Findley*. The Museum of Southwestern Biology, University of New Mexico, Albuquerque, 290 pp.

O'HIGGINS, P. and WILLIAMS, N. W. 1987. An investigation into the use of Fourier coefficients in characterizing cranial shape in primates. *Journal of Zoology, London*, **211**, 409–430.

RENAUD, S., MICHAUX, J., JAEGER, J.-J. and AUFFRAY, J.-C. 1996. Fourier analysis applied to *Stephanomys* (Rodentia, Muridae) molars: nonprogressive evolutionary pattern in a gradual lineage. *Paleobiology*, **22**, 255–265.

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

—— and BOOKSTEIN, F. L. (eds) 1990. *Proceedings of the Michigan Morphometrics Workshop*. First edition. Special publication, **2**, The University of Michigan Museum of Zoology, Ann Arbor, Michigan, viii+380 pp.

TEMPLE, J. T. 1992. The progress of quantitative methods in palaeontology. *Palaeontology*, **35**, 475–484.

WATERS, J. A. 1977. Quantification of shape by use of Fourier analysis: the Mississippian blastoid genus *Pentremites. Paleobiology*, **3**, 288–299.

WHITE, R. J., PRENTICE, H. C. and VERWIJST, T. 1988. Automated image acquisition and morphometric description. *Canadian Journal of Botany*, **66**, 450–459.

ZAHN, C. T. and ROSKIES, R. Z. 1972. Fourier descriptors for plane closed curves. *Institute of Electrical and Electronic Engineers, Transactions on Computers*, **C-21**, 269–281.

A. JOHN HAINES

Bullard Laboratories
Department of Earth Sciences
University of Cambridge
Madingley Rise, Madingley Road
Cambridge CB3 OEZ, UK
e-mail haines@esc.cam.ac.uk

JAMES S. CRAMPTON

Institute of Geological and Nuclear Sciences
PO Box 30368
Lower Hutt, New Zealand
e-mail j.crampton@gns.cri.nz

Typescript received 3 June 1998
Revised typescript received 17 April 2000

APPENDIX

Program HANGLE

The digitized outline of a fossil consists of a set of discrete points (x_i, y_i) that define a closed curve. The initial specification of the closest curve is completed by constructing a continuous function $(x(s), y(s))$ through these points using a spline technique that is designed to be convenient for performing Fourier analysis on the associated tangent angle $\theta(s)$ with respect to the arc-length s. These functions are related:

$$\frac{dx}{ds} = \cos\,\theta(s),\ \frac{dy}{ds} = \sin\,\theta(s),$$

and the tangent to the curve is given by

$$\frac{dy}{dx} = \tan\,\theta(s).$$

Thus, once $\theta(s)$ is known, $x(s)$ and $y(s)$ can be determined by integration. The condition that the curve is closed requires that

$$\int_0^L \frac{dx}{ds}\,ds = \int_0^L \cos\,\theta(s)ds = 0$$

and

$$\int_0^L \frac{dy}{ds}\,ds = \int_0^L \sin\,\theta(s) = 0$$

where L is the total arc-length around the curve. That is, after integrating around the entire curve one gets back to the starting point. When it comes to performing the Fourier analysis on $\theta(s)$, it is seen that these two constraints determine the values of one pair of Fourier coefficients in terms of the other coefficients. Otherwise, there is no other inherent constraint on the Fourier coefficients of $\theta(s)$.

[In contrast, half the corresponding Fourier coefficients of $x(s)$ and $y(s)$ are redundant in the elliptic Fourier shape analysis technique (EFA) of Kuhl and Giardina (1982). Knowledge of any one of the following three functions is equivalent to knowing the other two functions for the cases we are dealing with, which involve smooth curves with continuous derivatives of the tangent angles:

$$\theta(s),\ \frac{dx}{ds} = \cos\,\theta(s),\ \frac{dy}{ds} = \sin\,\theta(s).$$

The HANGLE coefficients can be determined from $\theta(s)$ and, conversely, $\theta(s)$ can be determined from the HANGLE coefficients. Similarly, the EFA coefficients for $x(s)$ can be determined from $\frac{dx}{ds}$, and $\frac{dx}{ds}$ can be determined from the EFA coefficients for $x(s)$. In the same manner, the EFA coefficients for $y(s)$ can be determined from $\frac{dy}{ds}$, and $\frac{dy}{ds}$ can be determined from the EFA coefficients for $y(s)$. Therefore, any one of the three sets of coefficients

HANGLE, EFA for $x(s)$, EFA for $y(s)$

is sufficient to calculate the other two sets. In particular, the EFA set for $x(s)$ can be used to calculate the EFA set for $y(s)$, and vice-versa. This means that the EFA coefficients are not independent and precisely half the full set (i.e. the set for both $x(s)$ and $y(s)$) are redundant.]

Circles are, of course, the most nondescript form of closed curve, for which

$$\theta(s) = \theta(0) + \frac{2\pi}{L}\, s.$$

The spline technique preserves circle-like characteristics as much as possible. At each of the points (x_i, y_i) the tangent angle θ_i is determined by initially fitting a circular arc through that point and the two adjoining points, (x_{i-1}, y_{i-1}) and (x_{i+1}, y_{i+1}). Then, between each pair of neighbouring points, (x_i, y_i) and (x_{i+1}, y_{i+1}), the final function $\theta(s)$ is constructed using the quadratic expansion

$$\theta(s) = \theta_i(1-t) + \theta_{i+1}t + \frac{1}{2}\Delta_i(1-t)t$$

in terms of

$$t = \frac{s - s_i}{s_{i+1} - s_i}.$$

There are two unknowns in this, $s_{i+1} - s_i$ and Δ_i, which are determined from the relationships

$$x_{i+1} - x_i = \int_{s_i}^{s_{i+1}} \cos\,\theta(s)ds = (s_{i+1} - s_i)\int_0^1 \cos\,\theta(t)dt.$$

$$y_{i+1} - y_i = \int_{s_i}^{s_{i+1}} \sin\,\theta(s)ds = (s_{i+1} - s_i)\int_0^1 \sin\,\theta(t)dt.$$

If the coefficient Δ_i of the nonlinear term in $\theta(t)$ is zero, then the arc segment between (x_i, y_i) and (x_{i+1}, y_{i+1}) is circular. This will occur only if the four points (x_{i-1}, y_{i-1}), (x_i, y_i), (x_{i+1}, y_{i+1}) and (x_{i+2}, y_{i+2}) lie on the same circular arc.

In practice, any digitized outline contains digitization 'noise' and in general it is necessary to smooth this out before constructing the function $\theta(s)$. The smoothing method is explained in the body of the paper.

The final step in HANGLE is to perform the Fourier analysis of $\theta(s)$. First, the arc-length s is normalized to be between 0 and 2π, by dividing by the total length L determined in constructing $\theta(s)$ and multiplying by 2π. In terms of the normalized arc-length, the Fourier expansion of $\theta(s)$ is taken to be of the form

$$\theta(s) = s + \sum_{k=0}^{\infty} A_k \cos(k(s - s_k))$$

in which A_k and s_k are the amplitude and phase terms for the kth harmonic. Apart from the linear term s at the start of the expansion, this is a standard Fourier series. The linear term is there so that $\theta(s)$ defines a loop. For a circle, all the harmonics are zero, except possibly the 0th harmonic term A_0, which is an arbitrary rotation of the contour as a whole. In HANGLE and the other two programs the 0th harmonic A_0 is removed so that there is a common reference orientation for all specimens.

Of the remaining harmonics, the first is uniquely determined for any value of A_1 not much larger than 1 by the requirement that the loop be closed. The following two mathematical results illustrate in particular how small values of A_1 depend on the amplitudes of the other harmonics:

(i) if $A_k = 0$ for all $k \geq 2$, except at most one such k, then $A_1 = 0$;

(ii) if $\Sigma_{k=2}^{\infty} A_k = 0(\epsilon)$, then $A_1 = 0(\epsilon^2)$.

In non-mathematical terms, the second of these results states that whenever the other harmonics are small, which is usually true for simple-shaped contours, the first harmonic will be very much smaller that the largest of the other harmonics. Since the first harmonic can be reconstructed from the others, using the requirement that the curve be closed, it is not output by HANGLE.

Rather than A_k and s_k, for $k \geq 2$ HANGLE outputs the real and imaginary parts, a_k and b_k, of the complex-valued Fourier coefficients

$$a_k + ib_k = \frac{1}{2}\, A_k \cos ks_k - \frac{i}{2}\, A_k \sin ks_k = \frac{1}{2}\, A_k\, e^{-iks_k}.$$

Associated with $a_k + ib_k$ there is the coefficient of the corresponding negative harmonic in the complex Fourier expansion of $\theta(s)$

$$a_k - ib_k = \frac{1}{2}\ A_k\ e^{+iks_k}.$$

Combined together with their Fourier exponents, these give

$$(a_k + ib_k)e^{iks} + (a_k - ib_k)e^{-iks} = \frac{1}{2}\ A_k(e^{+ik(s-s_k)} + e^{-ik(s-s_k)})$$

$$= A_k\ \cos(k(s - s_k))$$

which is the value of the kth harmonic in the real Fourier expansion. The complex coefficient $a_k + ib_k$ is the term actually calculated in the FFT of $\theta(s)$. A_k and s_k can be determined from a_k and b_k using the inverse relationships

$$A_k = 2\sqrt{a_k^2 + b_k^2},\ ks_k = -\tan^{-1}\left(\frac{b_k}{a_k}\right).$$

To illustrate how the Fourier expansion of the tangent angle $\theta(s)$ is related to the corresponding elliptic Fourier expansion of the position functions $x(s)$ and $y(s)$ (Kuhl and Giardina 1982) we will make two simplifications. First, we assume that only one of the amplitude terms A_k is non-zero, so that apart from the linear term s, the expansion of $\theta(s)$ is a single harmonic:

$$\theta(s) = s + A_k\ \cos(k(s - s_k)).$$

From the mathematical result above, we know that A_1 is zero and that k is greater than or equal to 2. Next, we will assume that A_k is small, which means that to first order in A_k we can write

$$\frac{dx}{ds} = \cos\ \theta(s) = \cos(s) - \sin(s)A_k\ \cos(k(s - s_k))$$

$$= \cos(s) - \frac{1}{2}A_k\{\sin((k+1)s - ks_k) - \sin((k-1)s - ks_k)\},$$

$$\frac{dy}{ds} = \sin\ \theta(s) = \sin(s) + \cos(s)A_k\ \cos(k(s - s_k))$$

$$= \sin(s) + \frac{1}{2}A_k\{\cos((k+1)s - ks_k) + \cos((k-1)s - ks_k)\},$$

Integrating these expressions with respect to the arc-length s (and requiring that the average values of x and y around the outline are zero) gives

$$x(s) = \sin(s) + \frac{1}{2}A_k\left\{\frac{1}{(k+1)}\ \cos((k+1)s - ks_k) - \frac{1}{(k-1)}\ \cos((k-1)s - ks_k)\right\}$$

$$y(s) = -\cos(s) + \frac{1}{2}A_k\left\{\frac{1}{(k+1)}\ \sin((k+1)s - ks_k) + \frac{1}{(k-1)}\ \sin((k-1)s - ks_k)\right\}.$$

Thus, we see that $x(s)$ and $y(s)$ both consist of, in general, three harmonics, the first, $(k-1)$th and $(k+1)$th, with amplitudes 1, $\frac{A_k}{2(k-1)}$ and $\frac{a_k}{2(k+1)}$ respectively. In the special, elliptic case, k = 2, the first and $(k-1)$th combine to form a single first harmonic. Consequently, from a single harmonic in the expansion of $\theta(s)$ there are either 4 or 6 separate harmonics in the expansions of $x(s)$ and $y(s)$, 4 of which have amplitudes determined by the value of A_k. Whenever, as in general, there are multiple harmonics in the expansion of $\theta(s)$, the level of redundancy reduces from being four-fold to generally being two-fold. This occurs because the second and fourth harmonics, for example, in the expansion of $\theta(s)$ both contribute to the third harmonic in the expansions of $x(s)$ and $y(s)$, rather than to completely separate harmonics.

HMATCH and HCURVE

The output from HANGLE is usually specified to be a restricted number of harmonics: in the examples presented here we have used up to the tenth harmonic. Program HCURVE takes any such set of Fourier coefficients and reconstructs not only the tangent angle function $\theta(s)$, but also the position functions $x(s)$ and $y(s)$ which define the outline contour. This is done by first using the inverse FFT and determining the values of the two parts of the first harmonic in the expansion

of $\theta(s)$ such that the outline is a closed curve. Then the arbitrary (x, y) origin for the outline is specified so that the average values of $x(s)$ and $y(s)$ are both zero. The resulting outline has a normalized arc-length of 2π, and the minimum and maximum values of $x(s)$ and $y(s)$ typically are about ± 1.

HMATCH is an intermediate program designed to adjust the outputs from HANGLE for any given set of outlines, prior to multivariate statistical analysis of the Fourier coefficients. When comparing the functions $\theta(s)$ and the associated Fourier coefficients for a variety of outlines, the starting point at which the arc-length s is set to zero for each outline is critical. In HANGLE itself, two options are available. The starting point of the original digitization can be selected as the starting point, which is appropriate if this represents a homologous landmark for all outlines. Alternatively, the starting point can be based on a particular dominant harmonic, say the kth. In this case, the starting points are chosen such that the differences between the complex Fourier coefficients $a_k + ib_k$ (see program HANGLE above) for the different outlines are minimised for the chosen harmonic. Much the same is done in program HMATCH, except that the sums of the squares of all differences between the outlines are collectively minimised for all harmonics output by HANGLE. This has the effect of minimising the total variance between the outlines and removing spurious differences due to poorly selected starting points. This can be very important during subsequent multivariate statistical analysis (see body of text).